



HAL
open science

Etudes structurales du complexe de réplication des Rhabdoviridae et des Paramyxoviridae. Les interactions entre la phosphoprotéine et la nucléoprotéine

Filip Yabukarski

► **To cite this version:**

Filip Yabukarski. Etudes structurales du complexe de réplication des Rhabdoviridae et des Paramyxoviridae. Les interactions entre la phosphoprotéine et la nucléoprotéine. Médecine humaine et pathologie. Université de Grenoble, 2013. Français. NNT : 2013GRENV033 . tel-01424101

HAL Id: tel-01424101

<https://theses.hal.science/tel-01424101>

Submitted on 2 Jan 2017

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée par

Filip YABUKARSKI

Thèse dirigée par **Marc JAMIN**

préparée au sein de l'Unité Interactions Virus-Cellule hôte
dans l'École Doctorale Chimie et Sciences du Vivant

Etudes structurales du complexe de réplication des *Rhabdoviridae* et des *Paramyxoviridae*. Les interactions entre la phosphoprotéine et la nucléoprotéine.

Thèse soutenue publiquement le **27 Septembre 2013**,
devant le jury composé de :

Dr. Noël TORDO

Institut Pasteur, Paris, France (Rapporteur)

Dr. Félix REY

Institut Pasteur, Paris, France (Rapporteur)

Dr. Annabelle VARROT

Centre de Recherche sur les Macromolécules Végétales, Grenoble, France
(Examinatrice)

Pr. Ioan LASCU

Université Bordeaux Segalen, Bordeaux, France (Examineur)

Pr. Franck FIESCHI

Université Joseph Fourier, Grenoble, France (Président)

Pr. Marc JAMIN

Université Joseph Fourier, Grenoble, France (Directeur de thèse)

Université Joseph Fourier / Université Pierre Mendès France /

Université Stendhal / Université de Savoie / Grenoble INP



Remerciements

Je ne remercierai jamais assez le **Pr. Marc JAMIN** de m'avoir accepté comme étudiant de thèse dans son groupe, d'avoir été d'esprit ouvert, d'avoir dirigé mes recherches avec patience et intérêt, pour sa disponibilité, sa bonne humeur, sa compétence et d'avoir été un Directeur de thèse que chaque étudiant souhaiterait avoir. Grand MERCI Marc !

Je remercie le directeur adjoint de l'UVHCI, le **Pr. Rob RUIGROK** de m'avoir initié dans le monde de la virologie pendant mes cours de Master 1, de m'avoir accueilli au sein de son institut et son groupe et d'avoir témoigné de l'intérêt pour mon travail.

Je remercie le directeur de l'EMBL Grenoble et de l'UVHCI, le **Dr. Stephen CUSACK**, de m'avoir accepté au sein de son institut.

Je tiens à remercier le **Pr. Ioan LASCU** d'avoir accepté de faire partie de mon jury de thèse et de m'avoir initié dans la recherche il y a environ 7 ans maintenant.

Je remercie le **Dr. Félix REY** et le **Dr. Noël TORDO** d'avoir accepté d'être rapporteurs de mon travail de thèse ici présent. Je remercie également le **Dr. Annabelle VARROT** et le **Pr. Franck FIESCHI** d'avoir participé à mon jury de thèse.

Je serai toujours reconnaissant au **Dr. Julien PERARD** d'avoir partagé une partie de sa grande connaissance des techniques de laboratoire et notamment en biologie moléculaire et purification des protéines et pour sa bonne humeur.

Je remercie également le **Dr. Ivan IVANOV** de m'avoir coaché à mon arrivé au laboratoire et pendant ma thèse, d'avoir partagé sa connaissance de biologie moléculaire et purification des protéines, pour sa gentillesse, sa camaraderie et pour sa bonne humeur pendant les moments de pause-café.

Je tiens à remercier le **Dr. Nicolas TARBOURIECH** de m'avoir enseigné tous les aspects de la cristallographie des protéines, et en particulier de m'avoir appris la pêche dans les lacs micromillilitriques, la collecte des données de diffraction et la résolution et affinement des structures.

Je remercie le **Dr. Cédric LEYRAT** de m'avoir coaché à mon arrivée au laboratoire et pour sa contribution aux différents projets.

Je remercie le **Dr. Jean-Marie BOURHIS** pour son aide au laboratoire et sa contribution au projet « X », d'avoir partagé avec moi ses connaissances de biochimie, de cristallographie des protéines et diffusion aux petits angles, pour sa patience et bonne humeur.

Je remercie également nos collaborateurs de l'équipe « Flexibilité et dynamique des protéines » de l'IBS, **Guillaume COMMUNIE**, le **Dr. Malene Ringkjøbing JENSEN** et le **Dr. Martin BLACKLEDGE**.

Merci à **Diego CARRIEL**, étudiant en M2, pour sa motivation et sa contribution au projet « X ».

Je remercie le **Dr. Andrew THOMPSON**, le **Dr. Beatriz GUIMARAES** et le **Dr. Ulrich ZANDER** pour leur précieuse aide lors de la collecte des données de diffraction à la ligne PROXIMA-1 à SOLEIL.

Merci également le **Dr. Andrew MCCARTHY** et le **Dr. Pauline MACHBOEUF** pour toutes les discussions autour de la cristallographie des protéines.

Je tiens à remercier en particulier le **Dr. Nicolas MARTINEZ**, le **Dr. Eric THIERRY**, le **Dr. Aurélien DORDOR**, **Nick ASHMAN** et le **Dr. Pascal FENDER** pour leur bonne humeur et d'avoir créé une bonne ambiance au laboratoire et **Elena SLANICKOVA**, **Sandrine VIGNON** et **Pierre MATAIS** d'avoir simplifié ma vie au laboratoire.

Je tiens à remercier également tous les membres du laboratoire.

Je ne pourrai jamais suffisamment remercier ma famille, mes parents **Elena** et **Filcho**, ma sœur **Milena** et ma compagne **Anne-Claire** pour leur soutien, patience et encouragement.

Abréviations

A : Angstrom (10^{-10} m)
aa : amino acid/acide aminé
ADN : Acide DésoxyriboNucléique
ARN : Acide RiboNucléique
ARNm : ARN messenger
ATP : Adénosine triphosphate
CB : tampon classique 20mM Tris-HCl pH 7.5, 150mM NaCl
CD : Dichroïsme Circulaire
CK-II : Caséine Kinase II
CTD : Domaine C-terminal
Da/kDa : Dalton/kiloDalton
E.coli: *Escherichia coli*
EDTA: acide éthylène diamine tetraacétique
ESRF : European Synchrotron Radiation Facility
F : protéine de fusion
G : glycoprotéine
GE: gene end
GS : gene start
GTP : Guanosine triphosphate
HeV : virus Hendra
HSQC : Heteronuclear Single Quantum Coherence
IBS : Institut de Biologie Structurale
ICTV : International Committee on Taxonomy of Viruses
IDP: Intrinsically Disordered Protein
kb: kilobase
L: grand sous unité de l'ARN polymérase ARN dépendante virale qui porte les activités catalytiques
M/mM : Molaire/miliMolaire
M : protéine matrice
MAD : Multiwavelength anomalous diffraction/diffraction anormale à plusieurs longueurs d'onde
MALLS : Multi Angle Laser Light Scattering/ diffusion de lumière laser à angle multiple
MeV: measles virus/virus de la rougeole
MIR: Multiple Isomorphous Replacement/Remplacement Isomorphe Multiple
MM: Masse Moléculaire / MW : molecular weight
MNV : Mononegavirales
MoRE : Molecular Recognition Element/ élément de reconnaissance moléculaire
MuV : Mumps virus/virus des oreillons
NiV : virus Nipah
nm: nanometer
nt: nucléotide
N_{CORE} : domaine globulaire N-terminal de la Nucléoprotéine chez les *Paramyxovirinae*
N_{CTA} : bras C-terminal de la Nucléoprotéine (sous domaine C-terminal)
NC : nucléocapside
N_{NTA} : bras N-terminal de la Nucléoprotéine (sous domaine N-terminal)
N_{TAIL}: domaine désordonné C-terminal de la Nucléoprotéine chez certains *Paramyxovirinae*
NTD : domaine N-terminal
N : nucléoprotéine
OMS : Organisation Mondiale de la Santé

P : phosphoprotéine
 P_{CD}/P_{MD} : domaine centrale/domaine de multimérisation de la phosphoprotéine
 P_{CED} : partie (région) C-terminale de la P, contenant le domaine central d'oligomérisation (P_{CD}) et le domaine C-terminal (P_{CTD})
 P_{CTD} : domaine C-terminal de la phosphoprotéine chez les *Rhabdoviridae**
 P_{NTR} : région N-terminale de la phosphoprotéine
 P_{XD} : PX domain ; domaine C-terminal de la phosphoprotéine, appelé encore domaine P_X, chez les *Paramyxoviridae**
 PAGE : PolyAcrylamideGEl
 PCR : Polymerase Chain Reaction
 PDB : Protein Data Bank
 PEG : polyethylene glycol
 pH: potentiel Hydrogène
 pI: point Isoélectrique
 PID : protéine intrinsèquement désordonnée
 RaV : Rabies virus (encore RV, RAV ou RABV)
 RE50 : glutamate/arginine 50mM
 RI : Refractive Index
 Rg : Rayon de giration
 Rh : Rayon hydrodynamique
 Rs : Rayon de Stockes
 RMN : Résonance Magnétique Nucléaire
 RNase A : Ribonucléase A
 RNP : RiboNucléoProtéine
 RMSD : Root Mean Square Deviation
 rpm : rounds per minute
 RSV: Respiratory Syncytial Virus/virus respiratoire syncytial
 SAD : single wavelength anomalous diffraction/diffraction anormale à une longueur d'onde
 SANS : Small Angle Neutron Scattering/diffusion des neutrons aux petits angles
 SAXS : Small Angle X-ray Scattering/diffusion des rayons X aux petits angles
 SEC: Size Exclusion Chromatography/chromatographie d'exclusion de taille
 SeV : virus Sendai
 SIR : Single Isomorphous Replacement/remplacement isomorphe unique
 T : température
 TMAO : TriMéthyleAmine N-Oxide
 UV : ultraviolet
 VSV : Vésicular Stomatitis Virus/virus de la stomatite vésiculaire

* Pour des raisons historiques le domaine C-terminal de la phosphoprotéine s'appelle P_{CTD} chez les *Rhabdoviridae* et domaine P_X ou P_{XD} chez les *Paramyxoviridae*.

Table des matières

Chapitre 1 : Introduction	9
1. Préambule – Les virus sont parmi nous.....	10
2. La mosaïque révèle la nature des virus	10
3. Classification des virus.....	12
4. Les virus à ARN négatif	13
4.1. La famille des <i>Rhabdoviridae</i>	14
4.1.1. Le virus de la stomatite vésiculaire (VSV).....	14
4.1.2. Le virus de Chandipura (CHPV).....	15
4.1.3. Le virus de la rage (RaV)	15
4.2. La famille des <i>Paramyxoviridae</i>	17
4.2.1. Le virus de la rougeole (MeV) et le virus respiratoire syncytial (RSV).....	18
4.2.2. Les <i>Henipavirus</i>	18
5. Homogénéité structurale et fonctionnelle des <i>Mononegavirales</i>	21
5.1. Les protéines virales	23
5.2. Architecture de la particule virale	24
6. Le cycle de réplication viral	25
6.1. Entrée	26
6.2. Les glycoprotéines	27
6.3. Interaction avec le cytosquelette et les moteurs moléculaires	29
6.4. Multiplication virale.....	30
6.5. Assemblage et bourgeonnement	30
6.6. La protéine de matrice (M).....	31
7. Machinerie de transcription/réplication et son fonctionnement	32
7.1. Machinerie moléculaire.....	32
7.2. Transcriptase et transcription	33
7.3. Réplicase et réplication	34
8. Organisation structurale des nucléocapsides	35
8.1. Structure de l'homopolymère nucléoprotéique	35
8.2. Structure désordonnée de l'appendice N _{TAIL} des <i>Paramyxovirinae</i>	43
9. La polymérase virale.....	44
9.1. Architecture de la protéine L.....	44
9.2. Mécanismes de formation de la coiffe et de polyadénylation.....	46
10. La phosphoprotéine est une protéine modulaire oligomérique.....	46

10.1. Présence de régions intrinsèquement désordonnées.....	47
10.2. Présence d'un MoRE dans P _{NTR}	47
10.3. Le domaine d'oligomérisation.....	49
10.4. Le domaine C-terminal de fixation à N-ARN.....	50
10.5. La connexion flexible entre P _{CD} et P _{CTD} /P _{XD}	51
10.6. Phosphorylation de la P.....	51
11. Structure du complexe N-ARN-P.....	52
11.1. Complexe N-ARN-P des <i>Rhabdoviridae</i>	53
11.2. Complexes N _{TAIL} -P _{XD} des <i>Paramyxovirinae</i>	54
11.3. Interaction N-ARN-P chez les <i>Pneumovirinae</i>	55
12. Déplacements de la polymérase sur la matrice N-ARN.....	55
12.1. Modèle impliquant le déplacement de P sur la matrice N-ARN.....	55
12.2. Modèle d'une progression par saltation de L sur plots fixes de P.....	56
13. Mécanismes d'encapsidation et le complexe N ⁰ -P.....	57
13.1. Le processus d'encapsidation.....	57
14. Comment caractériser la structure des protéines modulaires?.....	59
15. Réponses immunitaires et mécanismes de défense contre des virus.....	61
15.1. Immunité innée et immunité acquise.....	61
15.2. La réponse interféron.....	61
15.4. Les divers produits du gène P.....	63
Chapitre 2 : Matériels et Méthodes.....	69
2.1. Clonage.....	70
2.2. Expression et Purification.....	70
2.3. Analyse bioinformatique.....	72
2.4. Dichroïsme circulaire (CD).....	74
2.5. Chromatographie d'exclusion de taille couplée à la diffusion de lumière statique et à la réfractométrie (SEC-MALLS-RI).....	75
2.6. Diffusion aux petits angles (SAS).....	77
2.7. Cristallographie.....	85
2.7.1. Cristallogénèse.....	85
2.7.2. La diffraction et la densité électronique.....	87
2.8. Calorimétrie de titrage isotherme (ITC).....	98
Résultats.....	99
Chapitre 3 : Structure de la phosphoprotéine.....	100
3.1. Introduction.....	101

3.1.1. Analyse bioinformatique de la séquence de VSV P	104
3.2. Structure de la phosphoprotéine du virus de la stomatite vésiculaire	105
3.2.1. Article: Ensemble structure of the modular and flexible full-length vesicular stomatitis virus phosphoprotein	106
3.2.2. Caractérisation d'un fragment tronqué de la partie C-terminale, P ₁₋₁₇₇	139
3.3 Structure de la phosphoprotéine du virus Nipah	141
3.3.1. Analyse bioinformatique de la phosphoprotéine du virus Nipah	141
3.3.2. Le domaine P _X d'interaction avec les nucléocapsides	144
3.3.2.1. Structure tridimensionnelle du domaine P _X du virus Hendra	144
3.3.2.2. Structure tridimensionnelle du domaine P _X du virus Nipah.....	146
3.3.2.3. Comparaison des structures cristallographiques de NiV et HeV P _{XD}	149
3.3.2.4. Détails sur la cristallisation et résolution des structures des domaines P _{XD} des virus Nipah et Hendra	161
3.3.3. La phosphoprotéine de NiV forme des tétramères.....	164
3.3.3.1. Etat d'oligomérisation de la P entière en solution	164
3.3.3.2. Structure cristalline du domaine de tétramérisation.....	166
3.3.4. La partie C-terminale de la phosphoprotéine du virus Nipah (aa 471-709) est globalement structurée mais contient des régions flexibles.....	175
3.3.5. La partie N-terminale (P _{NTR}) de la phosphoprotéine de NiV est globalement désordonnée mais contient des structures secondaires fluctuantes.....	179
3.3.6. Modélisation de la phosphoprotéine de NiV	184
3.4. Discussion	186
Chapitre 4 : Interaction entre la phosphoprotéine du virus Nipah et la protéine STAT1	196
4.1. Introduction.....	197
4.2. P ₉₁₋₁₉₀ est intrinsèquement désordonné mais contient des structures secondaires fluctuantes	198
4.3. Interaction entre P ₉₁₋₁₉₀ et STAT1	201
4.4. Discussion	204
Chapitre 5 : Le complexe N⁰-P	206
5.1. Introduction.....	207
5.2. Article: Structure of the Vesicular Stomatitis Virus N ⁰ -P Complex.....	211
5.3. Le Complexe N ⁰ -P entière de VSV	234
5.3.1. Stratégie de production des complexes N ⁰ -P de VSV.....	234
5.3.2. Le complexe N ⁰ -P n'a pas une stœchiométrie fixe	236
5.3.3. Etude en solution des complexe N _{Δ21} ⁰ -P ₁₋₁₇₇ et N _{Δ21} ⁰ -P entière	238
5.3.4. Modélisation du complexe N _{Δ21} ⁰ -P entière	242

5.3.5. Discussion	243
5.4. Article: Crystal structure of the N ⁰ -P complex of Nipah virus	246
Chapitre 6: Discussion	273
ANNEXES.....	285
Review: Structural insights into the rhabdovirus transcription/replication complex.....	286
Review: Transcription et réplication des <i>Mononegavirales</i> : une machine moléculaire originale	299
Article: Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus	333
Références Bibliographiques.....	366

Chapitre 1 : Introduction

1. Préambule – Les virus sont parmi nous

De nombreux virus infectent l'Homme et sont responsables de nombreuses maladies et de graves épidémies (poliomyélite, variole, rage, SIDA, grippe etc.). Certains virus sont impliqués dans le développement de cancers. Plus récemment, la séquence du génome humain a révélé la présence massive de séquences d'origine virale, soulignant des relations antérieures très intimes entre des virus et leur hôte humain qui pourraient avoir contribué à l'évolution de notre espèce. Dans certains cas, les virus peuvent aussi avoir un effet bénéfique pour l'hôte qui les abrite (Roossinck 2011).

Les virus sont des parasites intracellulaires obligatoires qui établissent des relations étroites avec les organismes qu'ils infectent. Parce qu'ils n'ont pas la capacité de se reproduire eux-mêmes, ils détournent les ressources et les machines de la cellule hôte à leur profit. Au cours de l'évolution, ils ont développé une large gamme de mécanismes leur permettant de recruter, d'adapter, de modifier ou d'usurper les machines cellulaires dont ils ont besoin. En contre partie, les cellules ont développé divers mécanismes de défense permettant de limiter la prolifération virale. Ceux-ci vont des mécanismes moléculaires de l'immunité innée, qui peuvent être rapidement mis en action, aux mécanismes systémiques de l'immunité acquise. En réponse, les virus ont développé un véritable arsenal de contre-mesures qui masquent leur présence ou bloquent les réponses cellulaires, y compris en utilisant certains composés de la défense antivirale. Dans cette confrontation entre virus et cellule hôte, les virus ont l'avantage de se reproduire avec une grande efficacité et de muter rapidement ; ils ont ainsi la capacité à s'adapter aux conditions environnementales à l'intérieur de leur cellule hôte.

2. La mosaïque révèle la nature des virus

La description de diverses maladies dont l'origine virale est aujourd'hui connue, en particulier dans le cas de la rage, peut être trouvée dans des écrits anciens remontant à l'Antiquité. Déjà ARISTOTE et CELSUS avaient compris l'origine infectieuse de ces maladies. En 1879, le chimiste allemand Adolf Eduard MAYER, alors directeur de la station d'agriculture expérimentale de Wageningen aux Pays-Bas, entreprend, à la demande des cultivateurs, l'étude d'une maladie particulière affectant le tabac. Celle-ci se manifeste par l'apparition de plages de décoloration plus ou moins intense dans la feuille qui lui donne un aspect de mosaïque. Les folioles gaufrés deviennent filiformes et ont tendance à s'enrouler, les

rendant impropres à la commercialisation (**Figure 1A**). MAYER va démontrer que cette maladie est transmissible entre les plants de tabac à partir de la sève d'un plant contaminés. Il donne à cette maladie le nom de « mosaïque du tabac » et en décrit les symptômes.

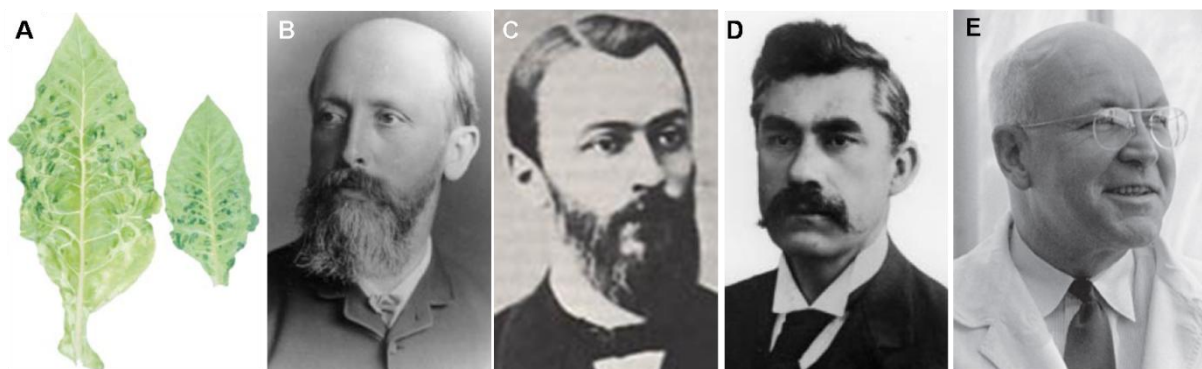


Figure 1 : La découverte du virus. (Figures d'après (Bos 1999) avec modifications) A) Des feuilles infectées par un agent inconnu. B) Adolf Eduard MAYER (1843-1942). C) Dmitri Iossifovitch IVANOVSKI (1864-1920). D) Martinus BEIJERINCK (1851-1931). E) Wendell Meredith STANLEY (1904-1971).

Il faut toutefois attendre 1888, pour que Dmitri Iossifovitch IVANOVSKI, un jeune étudiant de l'Université de St. Petersburg, mette pour la première fois en évidence les propriétés de l'agent responsable de cette maladie du tabac. Dans sa thèse, il montre que la sève extraite des plantes malades contient un agent infectieux, qui est transmis même après filtration à travers une bougie Chamberland. Cette bougie est constituée d'une colonne de fine porcelaine à travers laquelle le liquide est filtré en appliquant une pression. Ses pores sont de très petit diamètre et retiennent les bactéries. IVANOVSKI ne repère aucune bactérie dans la sève récoltée, ni avant ni après filtration. Son interprétation est qu'une toxine, produite par des bactéries, est responsable de la maladie – comme la toxine produite par le bacille de la diphtérie, découverte en 1888 par Emile ROUX et Alexandre YERSIN. En 1898, l'ingénieur chimiste hollandais Martinus BEIJERINCK répète les expériences d'IVANOVSKI et note qu'après la dilution de la sève et des passages en séries sur des plantes de tabac, le pouvoir infectieux persiste. Pour BEIJERINCK il ne s'agit ni d'une bactérie, car on ne la voit pas au microscope, ni d'une toxine, car son activité demeure après des dilutions successives. Il dénomme ce principe actif, à la fois vivant et fluide, « *contagium vivum fluidum* ». La controverse qui s'ensuivit au sujet de la nature réelle de l'agent de la mosaïque du tabac se termina avec le développement en 1917 du test par plaque de lyse par d'HERELLE et par la caractérisation biochimique et structurale dans les années 1930s.

En 1935 l'américain Wendell Meredith STANLEY réussit à « purifier » le virus de la mosaïque du tabac et à obtenir un « simple cristal protéique » qui conserve ses propriétés infectieuses. Ces prouesses valent à STANLEY de recevoir le prix Nobel de Chimie en 1946

et ouvrent la porte à la caractérisation structurale des virus. Il découvre un an plus tard que la protéine cristallisée n'est pas « pure » - elle contient des acides nucléiques. Le virus de la mosaïque du tabac n'est donc pas une simple protéine autorépliquative mais un complexe ribonucléoprotéique. Les premiers clichés de microscopie électronique sont obtenus en 1939 par Gustav KAUSCHE, Edgar PFANKUCH et Helmut RUSKA (le frère d'Ernst RUSKA).

3. Classification des virus

En 1957 LWOFF propose la définition suivante du virus :

« Les virus sont infectieux et potentiellement pathogènes ; ce sont des entités nucléoprotéiques possédant un seul type d'acide nucléique (ADN ou ARN) ; ils sont reproduits à partir de leur matériel génétique ; ils sont incapables de croître ou de se diviser ; ils sont dépourvus de métabolisme propre. »

En effet, on distingue trois grandes classes de virus selon la nature de leur génome et leur mode de répllication. Une première classe comprend les virus dont le génome est constitué d'ADN, simple ou double brin, et est répliqué directement par copie de l'ADN en ADN (virus de l'herpès, virus de la variole etc.). Une seconde classe regroupe les virus dont le génome est constitué d'ARN, qu'il soit simple brin de polarité positive (virus de l'hépatite C, virus de la fièvre jaune, virus de la poliomyélite) ou négative (virus de la rage, virus de la rougeole, virus de Nipah etc.), ou qu'il soit double brin, et dans les deux cas ils sont répliqués directement par copie de l'ARN en ARN. Finalement, une troisième classe renferme des virus dont le génome est constitué d'ARN mais qui se réplique en passant par un stade ADN par l'intermédiaire de la transcriptase inverse, un enzyme encodé par le virus. Aujourd'hui, l'International Committee on Taxonomy of Viruses (ICTV) est chargé de classer et de nommer les virus. Les virus sont classés en espèces, regroupées en genre, famille et éventuellement ordre selon qu'ils partagent des propriétés communes. Une « espèce » virale est constituée par un ensemble de virus constituant une lignée répliquative et occupant une niche écologique particulière.

4. Les virus à ARN négatif

De nombreux virus ont un génome constitué d'ARN simple brin de polarité négative. Ils sont responsables de nombreuses maladies chez les plantes et les animaux dont l'Homme (Ball 2007). Ils sont divisés en deux grands groupes. Ceux dont le génome est constitué d'une seule molécule d'ARN ont été regroupés au sein d'un même ordre, celui des *Mononegavirales* (*MNV*) par l'International Committee on Taxonomy of Viruses (ICTV) en raison de leur grande homogénéité structurale et fonctionnelle (voir ci-dessous) (Pringle 1997). D'autres ont un génome constitué de plusieurs segments. Ainsi, le virus de la grippe compte 8 segments, alors que celui des arenavirus en compte 2 et celui des bunyavirus (virus Hantaan) en compte 3. Bien qu'officiellement, l'ICTV n'ait pas créé un ordre, il est tentant sur la base des données structurales récentes de les regrouper dans un second ordre, celui des *Multinegavirales*.

Au sein de l'ordre des *MNV*, les virus sont actuellement classés en quatre familles selon des critères d'immunoréactivité et des analyses phylogénétiques basées sur la séquence en acides aminés de plusieurs protéines. Il s'agit des familles *Bornaviridae* (virus de la maladie de Borna), *Filoviridae* (virus Ebola et Marburg), *Paramyxoviridae* (virus de la rougeole, des oreillons, respiratoire syncytial et autres) et *Rhabdoviridae* (virus de la rage). Certains virus sont prévalent chez l'Homme, comme les virus de la rougeole (MeV), des oreillons (MuV), parainfluenza, respiratoire syncytial (RSV) ou le metapneumovirus, et ont un impact majeur en santé publique. D'autres *MNV* émergent ou ré-émergent à partir de réservoirs animaux, en particulier la chauve-souris, provoquant des épidémies de maladies graves ou mortelles comme les virus de la rage (RaV), de Chandipura (CHPV), d'Ebola (EBOV), d'Hendra (HeV), et Nipah (NiV) (Wong, Lau et al. 2007; Menghani, Chikhale et al. 2012). De même peut-on évoquer les pathologies vétérinaires telles la maladie de Carré chez les carnivores, la maladie des bovins provoquée par le virus rinderpest dont l'éradication vient d'être annoncée, diverses maladies des poissons provoquées par des rhabdovirus redoutés dans les piscicultures et ainsi de suite.

Dans ce travail, nous nous sommes intéressés à des virus appartenant aux familles des *Rhabdoviridae* et des *Paramyxoviridae* et nous limiterons cette introduction à ces familles. Ces deux familles renferment des pathogènes importants de l'Homme. Pour certains de ces virus un vaccin est disponible (RaV, MeV, MuV etc.), mais pas pour d'autres (RSV, NiV, HeV etc.) et nous ne disposons d'aucun médicament spécifique et efficace qui permette de bloquer ou de ralentir les infections causées par ces virus. Cependant, les populations humaines sont constamment exposées au risque de voir émerger une nouvelle souche où un

nouveau virus contre lequel la couverture vaccinale serait inefficace. Ces faits justifient l'importance d'essayer de mieux comprendre les mécanismes moléculaires de fonctionnement de ces virus dans l'espoir d'utiliser cette information pour développer des agents antiviraux ou de nouvelles stratégies de lutte contre ces pathogènes. Dans ce contexte, l'originalité et la conservation de la machinerie de synthèse d'ARN chez les *MNV* justifie son étude approfondie avec la perspective de développement de composés antiviraux ciblant sélectivement leur mécanisme de réplication.

4.1. La famille des *Rhabdoviridae*

Aujourd'hui, plus de 200 virus sont classés dans la famille des *Rhabdoviridae*. Ils infectent des plantes, des vertébrés et des invertébrés. Officiellement, cette famille comporte six genres, mais de nombreux virus ne sont pas classés. Les Vesiculovirus, Lyssavirus, Ephemerovirus et Novirhabdovirus infectent des animaux alors que les Cytorhabdovirus et les Nucleorhabdovirus infectent des plantes. Nous avons travaillé principalement avec le virus de la stomatite vésiculaire qui est considéré comme le prototype de cette famille et depuis de nombreuses années, sert de modèle au virus plus emblématique qu'est le virus de la rage.

4.1.1. Le virus de la stomatite vésiculaire (VSV)

La stomatite vésiculaire est une maladie contagieuse, affectant principalement les bovins, les porcins et les équidés. Elle est provoquée par le virus de la stomatite vésiculaire (VSV) qui est un arbovirus endémique aux Amériques. Il existe deux sérotypes majeurs de VSV dénommés New Jersey (NJ) et Indiana (IND). L'infection se manifeste par une inflammation au niveau de la muqueuse buccale mais également des pis et des sabots qui peuvent s'accompagner de l'apparition d'une ulcération. Les animaux atteints éprouvent des difficultés à se nourrir et peuvent présenter des signes d'anorexie et de léthargie, ainsi que de la fièvre. Bien que la maladie disparaisse au bout de quelques semaines sans laisser de séquelles, elle a des conséquences économiques importantes dans les élevages. La stomatite vésiculeuse est une zoonose, mais la plupart des infections humaines restent asymptomatiques. Chez les patients présentant des manifestations cliniques, le symptôme initial est une forte fièvre, suivi d'un état de type « grippal » (Letchworth, Rodriguez et al. 1999).

4.1.2. Le virus de Chandipura (CHPV)

D'autres virus du genre vesiculovirus, tel que le virus de Chandipura (CHPV), infectent l'Homme en provoquant des pathologies plus graves. Le CHPV a été isolé pour une première fois en 1965 dans le village de Chandipura, en Inde (Bhatt and Rodrigues 1967). Il ressemble fortement au VSV mais il s'en distingue par sa capacité à infecter efficacement l'Homme. C'est aussi un arbovirus, transmis par des moustiques du genre *Aedes* et des mouches de sable des genres *Phlebotomus* et *Sergentomyia* (Rao, Singh et al. 1967). CHPV est présent sur tout le territoire indien et a été récemment signalé en Afrique de l'ouest. En 2003, il a attiré l'attention en provoquant une épidémie dans l'état d'Andhra Pradesh, infectant 329 enfants dont 183 sont décédés (Rao, Basu et al. 2004). Depuis lors, d'autres épidémies se sont régulièrement déclarées dans diverses parties de l'Inde, toujours associées à un taux élevé de mortalité. Les premiers symptômes ressemblent à ceux de la grippe classique associés à des maux de ventre, des vomissements et des pathologies neurologiques. Ces symptômes sont souvent suivis d'inflammation cérébrale – encéphalite. Il n'existe pas de traitement spécifique, ni de vaccin contre l'infection et seuls les symptômes sont traités et les complications contrôlées.

4.1.3. Le virus de la rage (RaV)

La rage est une zoonose largement répandue dans le monde, connue depuis l'antiquité et dont on trouve des traces notamment dans les écrits mésopotamiens (XXIII^{ième} siècle avant J.C.), chinois (VI^{ième} siècle avant J.C.), égyptiens (III^{ième} siècle après J.C.) ou grecs (IX^{ième} siècle avant J.C.) (Theodorides). La nature infectieuse de la maladie fut démontrée au XIX^{ième} siècle par George G. ZINKE (1804), qui réussit à transmettre la maladie en utilisant la salive de chiens infectés, et plus tard par Pierre Victor GALTIER (1879), qui transmit la rage entre des lapins. En 1903, Adelchi NEGRI démontra la connexion entre le virus et le système nerveux, mettant en évidence la présence d'inclusions cytoplasmiques dans les cellules neuronales infectées. Louis PASTEUR produisit une forme atténuée du virus qu'il utilisa ensuite pour développer le premier vaccin protectif contre la rage. Le 6 juillet 1885, il administra ce vaccin à titre de prophylaxie post-exposition à un jeune garçon, Joseph Meister, qui venait d'être mordu par un chien enragé, et le garçon survécut.

RaV infecte les vertébrés, principalement les mammifères et occasionnellement l'Homme ; le plus souvent la transmission à l'Homme se fait par la morsure d'un chien infecté. RaV est un virus neurotrope. Après avoir été inoculé, le virus commence par se

multiplier à son point d'entrée dans le tissu musculaire, et ensuite pénétre les terminaisons nerveuses des neurones périphériques par l'intermédiaire desquels il se dirige vers les corps cellulaires et la moelle épinière. Chez l'Homme, la période entre la contamination et l'apparition des signes de la maladie (incubation), varie en moyenne entre 30 et 45 jours mais peut dépasser un an. Sa durée dépend de la dose de virus inoculée, de la localisation de la morsure et de la gravité de la plaie. L'incubation est d'autant plus courte que le siège de la contamination est proche du cerveau. L'invasion par le virus se manifeste par un prurit (démangeaisons) ou des douleurs dans la région mordue. Lorsqu'il atteint le cerveau, il se réplique activement et provoque une encéphalite dont la pathogénèse complexe conduit à différentes formes de la maladie, associée à différents troubles du comportement. L'insomnie est fréquente chez le malade qui devient anxieux, anormalement triste, sujet à des crises de larmes et souvent cherche à s'isoler. La maladie peut alors évoluer soit vers une forme paralytique (30% des cas) soit vers une forme dite « furieuse ». Deux signes sont caractéristiques de la maladie : une peur pathologique de l'air (aérophobie) et, surtout, de l'eau (hydrophobie). Par la suite le virus se répand vers les organes périphériques tels que les yeux et les glandes salivaires. L'encéphalite est incurable et entraîne toujours la mort de 4 à 6 jours après l'apparition des premiers symptômes de la maladie. Il n'existe pas de traitement curatif et la vaccination représente le seul moyen efficace de protection contre le virus. Cependant la vaccination pré-exposition n'est utilisée que pour des personnes en contact potentiel avec le virus. Le plus souvent le vaccin est administré post-exposition en combinaison avec des immunoglobulines antirabiques. La maladie est aujourd'hui sous contrôle en Europe et en Amérique du Nord notamment grâce à des campagnes de vaccination de la faune sauvage et grâce à l'utilisation prophylactique du vaccin et du traitement post-infection. Cependant, selon un rapport de l'Organisation Mondiale de la Santé (WHO, 2010) la rage tue aujourd'hui encore plus de 50.000 personnes par an de part le monde, principalement en Afrique et en Asie.

La rage est une zoonose et ses réservoirs naturels sont principalement les carnivores comme le renard ou le raton laveur et les chauves-souris (McColl, Tordo et al. 2000). Ces dernières constituent l'ordre des chiroptères qui représente un quart à un tiers des espèces de mammifères de la planète. De part cette grande diversité, les chauves-souris sont un « terreau » fertile pour le développement de nouvelles souches virales et de nouveaux virus, et les populations humaines sont constamment exposées au risque de voir émerger de nouveaux virus. Ainsi, diverses souches du virus de la rage circulent chez les chauves-souris, parmi lesquelles certaines sont moins sensibles au vaccin que celle de la rage classique (Hanlon,

Kuzmin et al. 2005). L'adaptation de ces souches à des animaux capables de transmettre l'infection à l'homme pourrait conduire à la réémergence de cette maladie et à de nouvelles épidémies.

4.2. La famille des *Paramyxoviridae*

La famille des *Paramyxoviridae* comporte deux sous-familles, celle des *Paramyxovirinae* et celle des *Pneumovirinae*, chacune comprenant des pathogènes humains proéminents (**Figure 2**). Elle inclut également des virus émergents tels que les virus Hendra et Nipah qui provoquent des épidémies chez l'Homme avec un taux de mortalité élevé et chez les animaux d'élevage avec des conséquences économiques majeures liées aux abattages ordonnés pour limiter la propagation de l'épidémie.

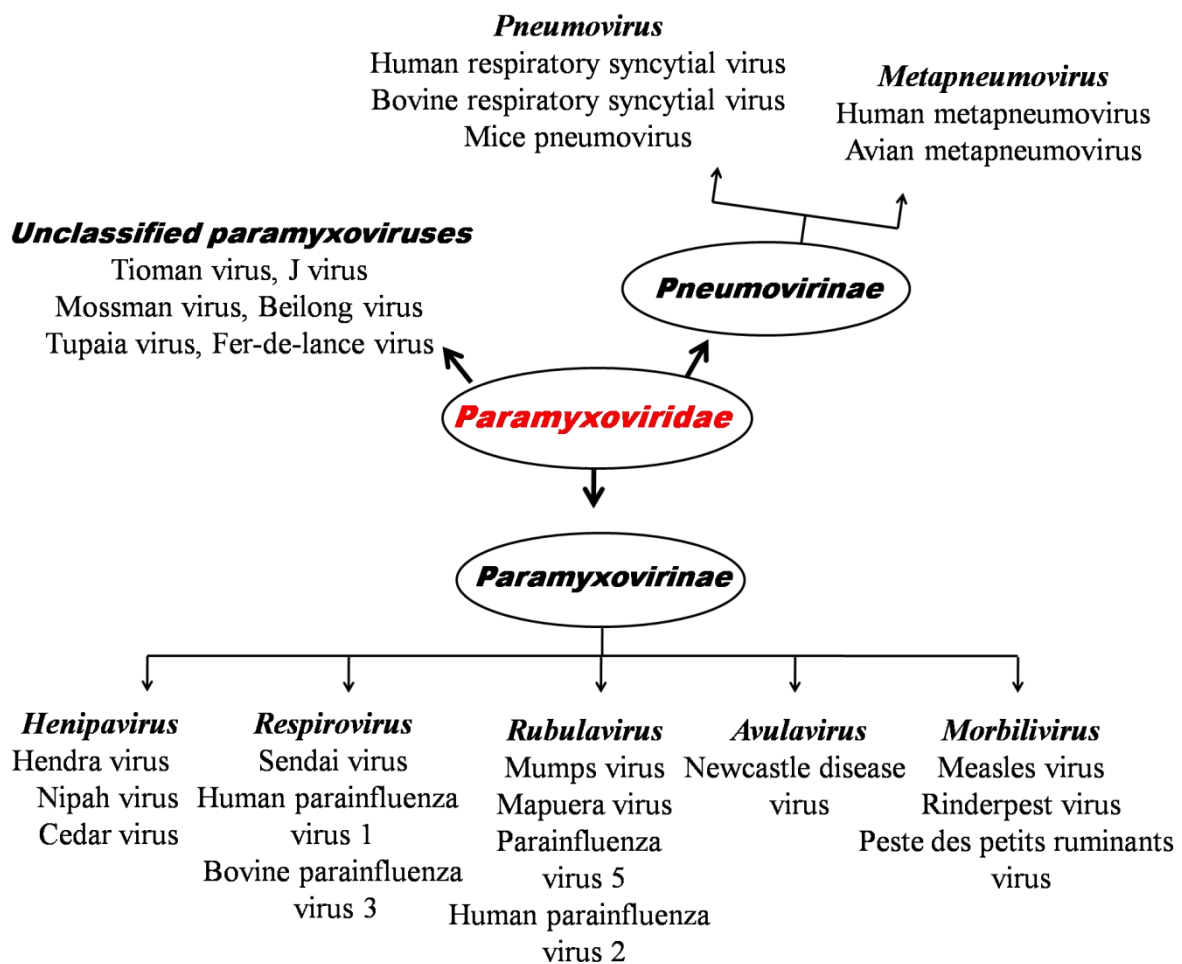


Figure 2 : Classification des *Paramyxoviridae*.

4.2.1. Le virus de la rougeole (MeV) et le virus respiratoire syncytial (RSV)

Le RSV est le principal responsable des bronchiolites et pneumonies sévères du nourrisson, et est le premier agent viral provoquant la mort chez les enfants. Il est également responsable de pneumonies et de pathologies respiratoires chez les personnes âgées et les receveurs d'une transplantation de moelle osseuse (<http://www.cdc.gov/rsv/>) qui entraînent des taux élevés d'hospitalisation et de décès. Aucun vaccin n'est actuellement disponible et seul un traitement prophylactique à base d'anticorps (Prolivizumab, Motavizumab) est utilisé dans le cas d'enfants prématurés. Deux composés antiviraux sont actuellement étudiés : le composé RSV604, qui cible la Nucléoprotéine de RSV, et le composé ALN-RSV01 qui est un siARN, dirigé contre l'ARN viral. La rougeole est une maladie ancienne, qui reste une cause majeure de mortalité chez les jeunes enfants (2011) et est sujette à une réémergence massive à l'instar de l'épidémie survenue en France de 2009 à 2011 consécutive à la négligence accrue de couverture vaccinale. Un vaccin très efficace est disponible, néanmoins il n'existe pas de traitement spécifique et après le déclenchement de la maladie seuls les symptômes sont traités.

4.2.2. Les *Henipavirus*

Les virus Nipah (NiV) et Hendra (HeV) ont été récemment identifiés, (Nipah en Malaisie en 1998 et Hendra en Australie en 1994) et ont été nommé d'après les noms des endroits où ils ont été primitivement isolés. Le premier cas de HeV, décrit en 1994, a touché un cheval de course et son propriétaire à Queensland, dans le quartier Hendra (Murray, Selleck et al. 1995), mais les expériences réalisées par la suite ont montré que ce virus est capable d'infecter une large gamme d'hôtes - chevaux, chiens, chats, lapins, ainsi que d'autres animaux de laboratoire (Murray, Rogers et al. 1995; Murray, Selleck et al. 1995; Westbury, Hooper et al. 1995; Westbury, Hooper et al. 1996; Hooper, Westbury et al. 1997). NiV a été identifié pour la première fois en 1998 comme le responsable d'une épidémie de troubles respiratoires et d'encéphalites chez des cochons en Malaisie du nord (Chua, Goh et al. 1999; Amal, Lye et al. 2000). La maladie s'est transmise à des humains en contact avec les cochons malades, causant plusieurs cas d'encéphalites fatales. D'autres épidémies se sont par la suite déclarées en Malaisie, puis à Singapour, au Bangladesh (2001, 2003, 2004, 2005, 2007 et 2008) et en Inde (2001 et 2007) (Epstein, Field et al. 2006). L'épidémie en Malaisie a tué 106 personnes sur 276 infectées (38.5%), celles au Bangladesh et en Inde ont eu respectivement des taux de mortalité supérieurs, de 67 et 92%, potentiellement liés aux mauvaises conditions

sanitaires auxquelles les malades étaient exposés (Wild 2009). Des virus apparentés ont été récemment isolés en Afrique de l'ouest (Hayman, Suu-Ire et al. 2008) et un nouveau membre du genre *Henipavirus*, le virus Cedar, moins virulent pour les animaux de laboratoire que ces cousins NiV et HeV en raison de l'absence de production de la protéine V (voir ci-après), a été récemment découvert chez des chauves-souris en Australie (Marsh, de Jong et al. 2012). Ces virus apparentés aux *Paramyxovirinae*, diffèrent néanmoins suffisamment pour justifier la création par ICTV d'un nouveau genre, les *Henipavirus*.

Les hôtes naturels de ces virus sont les chauves-souris du genre *Pteropus*, de l'ordre *Chiroptera* (**Figure 3**). Des virions d'HeV et de NiV ont été détectés dans leur salive, urine, sécrétions nasales et placenta, suggérant qu'ils sont transmis par l'intermédiaire de ces fluides (Halpin, Young et al. 2000). Il est intéressant de noter que les chauves-souris ne présentent pas de signes de maladie suggérant une adaptation entre ces animaux et ces virus. Des anticorps dirigés contre NiV et HeV sont détectés chez plusieurs espèces du genre *Pteropus* et sont très probablement transmis de la mère à la progéniture, qui serait ainsi protégée au début de sa vie.

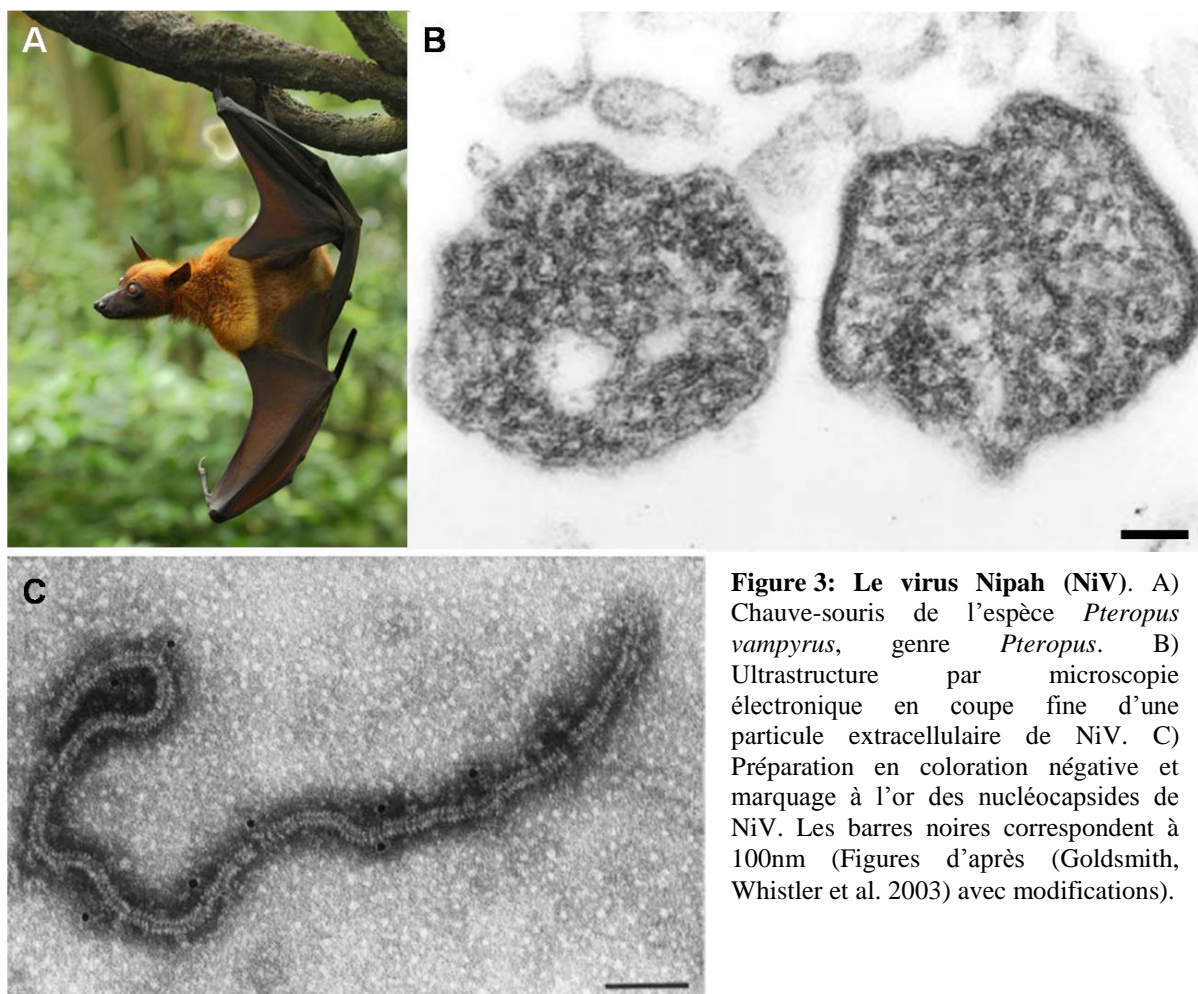


Figure 3: Le virus Nipah (NiV). A) Chauve-souris de l'espèce *Pteropus vampyrus*, genre *Pteropus*. B) Ultrastructure par microscopie électronique en coupe fine d'une particule extracellulaire de NiV. C) Préparation en coloration négative et marquage à l'or des nucléocapsides de NiV. Les barres noires correspondent à 100nm (Figures d'après (Goldsmith, Whistler et al. 2003) avec modifications).

L'émergence récente de ces virus comme responsables d'épidémies de maladie chez l'Homme peut être associée à la destruction de l'habitat naturel des chauves-souris qui mène à des rencontres de plus en plus fréquentes entre l'Homme, les animaux domestiques et ces animaux. Les animaux domestiques sont infectés lorsqu'ils ingèrent des fruits souillés par les chauves-souris ((Ksiazek, Rota et al. 2011) pour revue).

Les modes d'infection par HeV et NiV ne sont pas identiques et ils diffèrent par rapport à leur hôte « d'amplification », les chevaux pour HeV et les cochons pour NiV. Hev a été démontré pour causer une infection respiratoire accrue chez les chevaux mais sa transmission aérienne du cheval à l'Homme ou de cheval à cheval n'a pas été démontré (Williamson, Hooper et al. 1998). L'Homme devient infecté vraisemblablement par le contact direct avec des tissus infectés. Contrairement à HeV, NiV a infecté plusieurs espèces – cochons, chèvres, chats et chiens (CDFA) (Middleton, Westbury et al. 2002). La transmission de l'Homme à l'Homme a été documentée (Luby, Gurley et al. 2009; Sazzad, Hossain et al. 2013) mais d'autres moyens de transmission ont été également proposés. Il a été montré que l'infection récente d'un grand nombre d'individus au Bangladesh n'est peut être pas du à une transmission interhumaine mais plutôt à la consommation de lait de palme, dans lequel ont bu des chauves-souris infectées (Stone 2011).

Les pathologies et les symptômes associés aux infections par des *Henipavirus* sont très similaires même si, à cause de sa plus forte incidence, ils ont été mieux caractérisés pour NiV (Tan and Chua 2008). Chez l'Homme, les symptômes apparaissent entre 7 à 10 jours après la contamination même si des délais plus longs (plus de 50 mois) ont été constatés. Ils comprennent des méningites aseptiques, des encéphalites accompagnées de fièvre, de maux de tête, de vomissements et des épisodes de conscience réduite (Ksiazek, Rota et al. 2011). Chez les patients atteints d'une encéphalite, l'imagerie par résonance magnétique (IRM) révèle la présence de lésions multiples, principalement dans la matière blanche du cerveau (Goh, Tan et al. 2000). Les infections par des *Henipavirus* sont de nature systémique avec des évidences d'infection dans plusieurs organes. NiV cible de préférence les cellules endothéliales microvasculaires provoquant leur fusion (formation de syncytia) dans plusieurs organes (Wong, Shieh et al. 2002).

Une stratégie utilisée pour bloquer l'infection des cellules ciblées consiste à administrer la protéine G ou son récepteur cellulaire sous forme soluble, mais ce traitement a des effets indésirables. En 2012, un vaccin dirigé contre HeV s'est montré efficace dans la protection des singes contre NiV (Bossart, Rockx et al. 2012), mais il n'est pas encore

disponible pour l'Homme et il n'existe pas de traitement spécifique des infections par ces virus.

5. Homogénéité structurale et fonctionnelle des *Mononegavirales*

Très variables dans leur spécificité d'hôtes, leur mode de transmission et les maladies qu'ils provoquent, les *Mononegavirales* (MNV) partagent une organisation similaire de leur génome et de leur virion, et des stratégies de transcription et de réplication de l'ARN viral très proches et très originales dans le monde du vivant qui suggèrent une origine évolutive commune et ont justifié la création d'un ordre dans la classification virale (**Figure 4**).

Ces virus possèdent les caractéristiques communes suivantes :

- (1) un génome, dont la taille, comprise entre 9 et 20 kb, comprend entre 5 et 12 gènes et est codant sur plus de 90% de sa longueur (les génomes de HeV et NiV font partie des plus longs génomes de la famille avec 18,2kb, tandis que le génome de VSV est parmi les plus courts, autour de 11kb),
- (2) une organisation linéaire du génome avec dans l'ordre en allant de l'extrémité 3' vers l'extrémité 5', une région non-traduite en 3' (3'-UTR), des gènes codant pour les protéines de capsid, des gènes codant pour les protéines de surface et le gène codant pour l'ARN polymérase ARN-dépendante (RdRp) aussi appelée protéine L (Large),
- (3) une complémentarité entre les extrémités 3' et 5' du génome,
- (4) une ARN polymérase ARN-dépendante associée au virion,
- (5) une nucléocapside hélicoïdale qui est utilisée comme matrice par la polymérase virale,
- (5) un processus de transcription de 6 à 10 ARN messagers par un mécanisme de synthèse interrompue (mécanisme STOP-START) à partir d'un promoteur unique situé à l'extrémité 3' du génome,
- (6) un processus de réplication passant par la synthèse d'une copie antigénomique intermédiaire de polarité positive,
- (7) une maturation par bourgeonnement à partir d'une membrane cellulaire, formant des particules contenant une nucléocapside hélicoïdale en un ou plusieurs exemplaires.

Il faut noter que les *Bornaviridae* se distinguent des virus des trois autres familles par un processus d'épissage de leur ARN messenger qui utilisent la machinerie cellulaire.

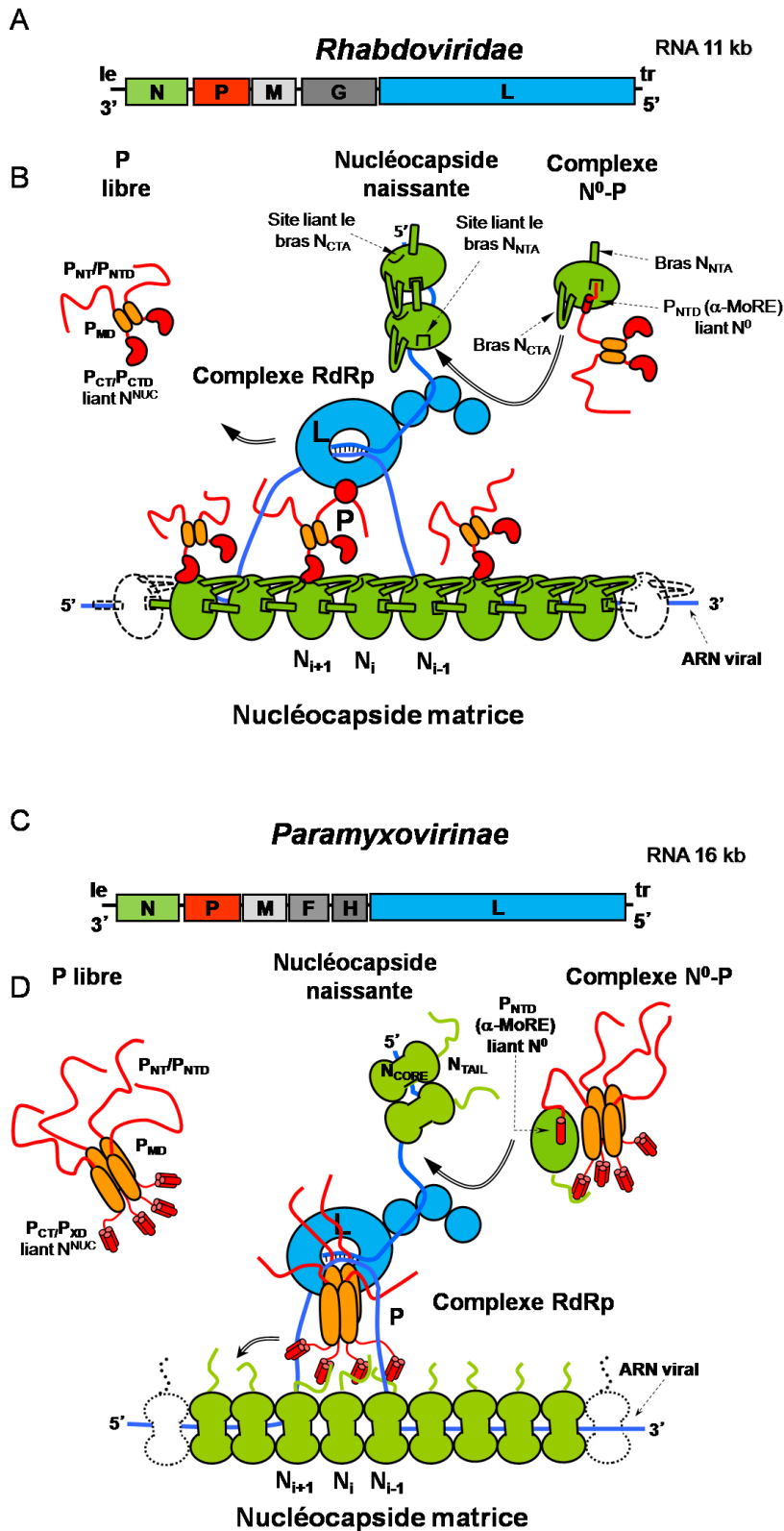


Figure 4 : Machinerie de transcription et répliation des *Rhabdoviridae* et *Paramyxovirinae* (figure d'après Blocquel, D. et al. et al avec modifications). **A**) et **C**) Organisation des génomes avec la nucléoprotéine N, la phosphoprotéine P, la protéine de matrice M, la protéine de fusion F et la protéine H (chez les *Paramyxovirinae*), la glycoprotéine G (chez les *Rhabdoviridae*) et la polymérase virale L. **B**) et **D**) Représentation schématique des complexes de répliation/transcription. La N (en vert) encapside l'ARN viral et forme la matrice de répliation/transcription pour la polymérase virale, composée de deux sous-unités – la L (bleu) et la P (orange-rouge). P forme des dimères (B) ou tétramères (D) et contient des régions flexibles (rouge) et un domaine central d'oligomérisation (orange).

5.1. Les protéines virales

Le génome des *MNV* code pour des protéines structurales, requises pour former la particule virale, et des protéines non-structurales, essentiellement garantes de la propagation chez l'hôte. Cinq du premier groupe sont communes à tous les *MNV* avec un agencement de leurs gènes identique le long du génome comprenant de l'extrémité 3' à l'extrémité 5' : ceux de la nucléoprotéine (N), de la phosphoprotéine (P), de la protéine de matrice (M), de la glycoprotéine de surface (G) et de l'ARN polymérase ARN-dépendante (RdRp) aussi appelée la protéine L car elle constitue la grande sous-unité de la polymérase fonctionnelle (**Figure 4**). N encapside l'ARN, formant un complexe N-ARN qui sert de matrice pour la transcription et la réplication. P et L constituent la polymérase RdRp à deux sous-unités (Kolakofsky, Le Mercier et al. 2004; Lamb and Parks 2007). Plusieurs copies du complexe polymérasique et des molécules de P s'associent à l'homopolymère de nucléoprotéine recouvrant l'ARN génomique pour former une ribonucléoprotéine (RNP) qui constitue la partie infectieuse du virus et est suffisante pour permettre sa reproduction (Szilagyi and Uryvayev 1973). M joue un rôle structural dans la particule virale, participe au processus d'assemblage des nouveaux virions dans le cytoplasme de la cellule hôte ((Solon, Gareil et al. 2005) et références citées) et est un régulateur négatif de la transcription (Iwasaki, Takeda et al. 2009). Chez les *Rhabdoviridae*, la glycoprotéine de surface G assure reconnaissance et fixation du virus sur la cellule hôte et induit la fusion de l'enveloppe virale avec la membrane cellulaire, préalablement à la pénétration du virus dans le cytoplasme de la cellule infectée (Albertini, Baquero et al. 2012). Chez les *Paramyxoviridae*, ces fonctions sont dissociées et assurées par deux glycoprotéines transmembranaires, la glycoprotéine d'attachement (G ou H) et la glycoprotéine de fusion (F) (Lamb and Parks 2007).

A ce socle de gènes communs, viennent s'ajouter pour certaines familles virales, d'autres gènes insérés à différents endroits du génome et codant pour des protéines structurales et/ou non-structurales. Ainsi, les rhabdovirus de plantes possèdent un gène supplémentaire spécifique codant pour la protéine 4b impliquée dans le transport du virus au travers des plasmodesmes et du système vasculaire. Le génome du RSV comprend deux autres cofacteurs, M2-1 et M2-2, impliqués dans la machinerie de réplication du génome. Le premier augmente la processivité de la polymérase lors de la transcription et le second favorise la réplication de l'ARN génomique en fin de cycle (Collins and Crowe 2007). D'autre part, chez de nombreux *MNV*, le gène P code pour d'autres protéines : des isoformes de P tronquées, des isoformes issues d'un mécanisme d'édition de l'ARNm, (protéines V, W),

et des protéines provenant de la traduction de l'ARNm à partir d'un codon d'initiation alternatif (protéines C, Y) (Eaton, Mackenzie et al. 2007; Lamb and Parks 2007; Lyles and Rupprecht 2007). Ces divers produits du gène P sont notamment impliqués dans la régulation de la réplication virale et dans l'inhibition de la réponse immunitaire innée intrinsèque de la cellule hôte.

5.2. Architecture de la particule virale

Les virions des *Paramyxoviridae* sont pléiomorphes, de forme grossièrement sphérique (150-350 nm) ou filamenteuse (Lamb and Parks 2007). La forme des *Rhabdoviridae* est plus régulière, en balle de fusil ou en bacille (Lyles and Rupprecht 2007; Desfosses, Ribeiro et al. 2013). De façon générale, la particule des *MNV* est constituée de trois couches superposées. La couche externe, ou enveloppe, est une bicouche lipidique dérivée de la membrane plasmique de la cellule hôte riche en glycoprotéines qui forment des spicules visibles en surface. La couche intermédiaire, constituée de l'assemblage en un réseau tridimensionnel d'une partie des protéines M, assure la cohésion entre les couches interne et externe. La couche interne est la nucléocapside (NC) composée de protéines N polymérisées sur l'ARN génomique (N-ARN) auxquelles sont associées les protéines L, P et, dans le cas du RSV, la protéine M2-1. Cette couche n'est visible en microscopie électronique que pour les virus ayant une forme géométrique bien régulière comme les *Rhabdoviridae* (**Figure 5**). Pour VSV, à pH acide et à faible force ionique, le complexe N-ARN isolé s'auto-assemble en une particule en forme de balle, morphologiquement similaire à la couche interne du virion, mais dotée d'une plus grande flexibilité (Desfosses, Ribeiro et al. 2013). La protéine N est donc suffisante pour déterminer la forme de la particule virale. Toutefois, l'addition de protéine M réduit la variabilité observée dans les dimensions des particules de N-ARN, suggérant que M régule la compaction de la particule dans le virion (Desfosses, Ribeiro et al. 2013). La reconstitution de la partie cylindrique du virion de VSV par cryo-microscopie électronique a révélé une géométrie hélicoïdale des couches intermédiaire et interne (Ge, Tsao et al. 2010). La localisation des protéines P et L reste indéterminée, et il est proposé que ces protéines, associées avec le complexe N-ARN, soient localisées dans la cavité centrale de la particule en forme de « cigare » définie par l'enroulement hélicoïdal de la NC (Barge, Gaudin et al. 1993). Plus récemment, une reconstruction du cône sommital (Desfosses, Ribeiro et al. 2013) suggère que la nucléation de l'assemblage de cette particule commence par la formation d'un tour de dix sous-unités comparable à celui observé dans les anneaux cristallisés (Green, Zhang et al. 2006). En cryomicroscopie électronique, la structure hélicoïdale de N-ARN de

MeV est visible à l'intérieur d'un virion, et elle est recouverte, au moins partiellement, d'un manchon de protéine M (Liljeroos, Huiskonen et al. 2011) (**Figure 5**).

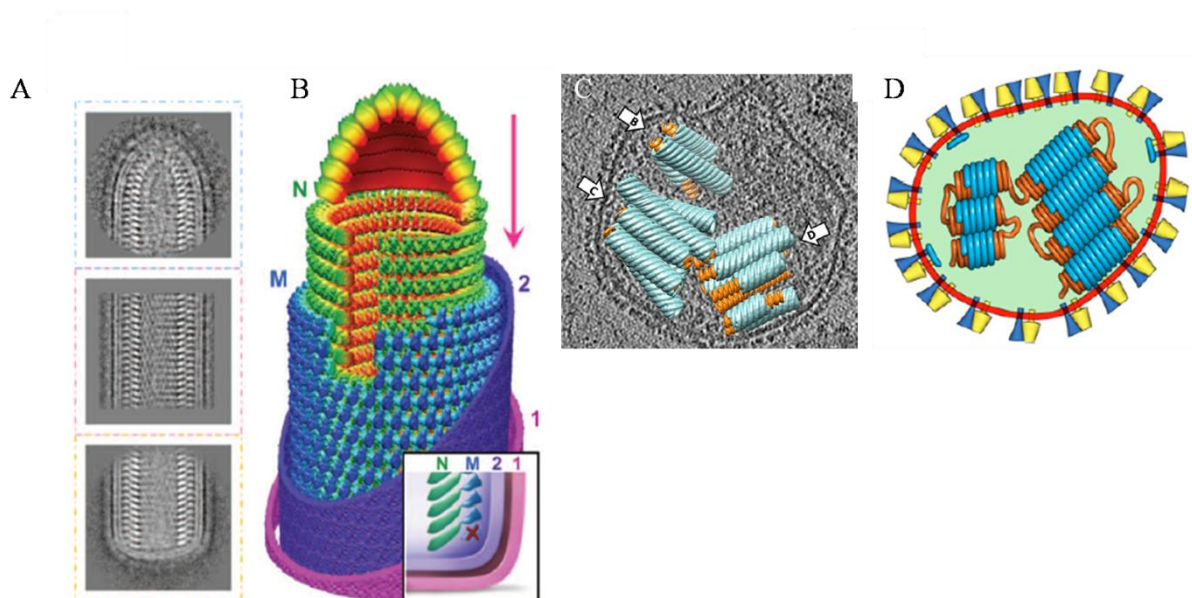


Figure 5: Architecture des virions de VSV et MeV. **A)** Moyenne des classes 2D représentative de l'extrémité conique, le tronc et la base du virion de VSV. **B)** Modèle de l'extrémité conique et du tronc obtenu par reconstruction cryo-microscopique (N en vert-rouge, M en bleu, et les faces interne et externe sont en violet et en rose respectivement) (figure d'après (Ge, Tsao et al. 2010) avec modifications). **C)** Structure moyenne des filaments de M (bleu) et N (orange) superposée sur la carte de densité du virion de MeV. **D)** Présentation schématique de l'ultrastructure du virion de MeV (figure d'après (Liljeroos, Huiskonen et al. 2011) avec modifications).

6. Le cycle de réplication viral

Le cycle de réplication viral peut être divisés en trois parties majeures: (1) l'attachement du virus à son récepteur, la pénétration dans la cellule hôte et le transport jusqu'au site de réplication; (2) la multiplication proprement dite du virus, incluant la production d'ARN messagers et de protéines virales et la réplication du génome; (3) l'assemblage de nouvelles particules virales et leur bourgeonnement à partir de la cellule hôte (**Figure 6**). En plus de ces processus conduisant à la génération de nouvelles particules virales, les virus interagissent avec les composants de la cellule hôte afin notamment de réguler la production de protéines cellulaires et d'interférer avec les systèmes de défense immunitaire. Pour les *Rhabdoviridae* (à l'exception des nucléorhabdovirus) et les *Paramyxoviridae*, la réplication s'effectue dans le cytoplasme de la cellule hôte, même si certaines protéines sont transportées dans le noyau où interagissent avec des partenaires cellulaires notamment impliquées dans la réponse immunitaire (voir plus loin).

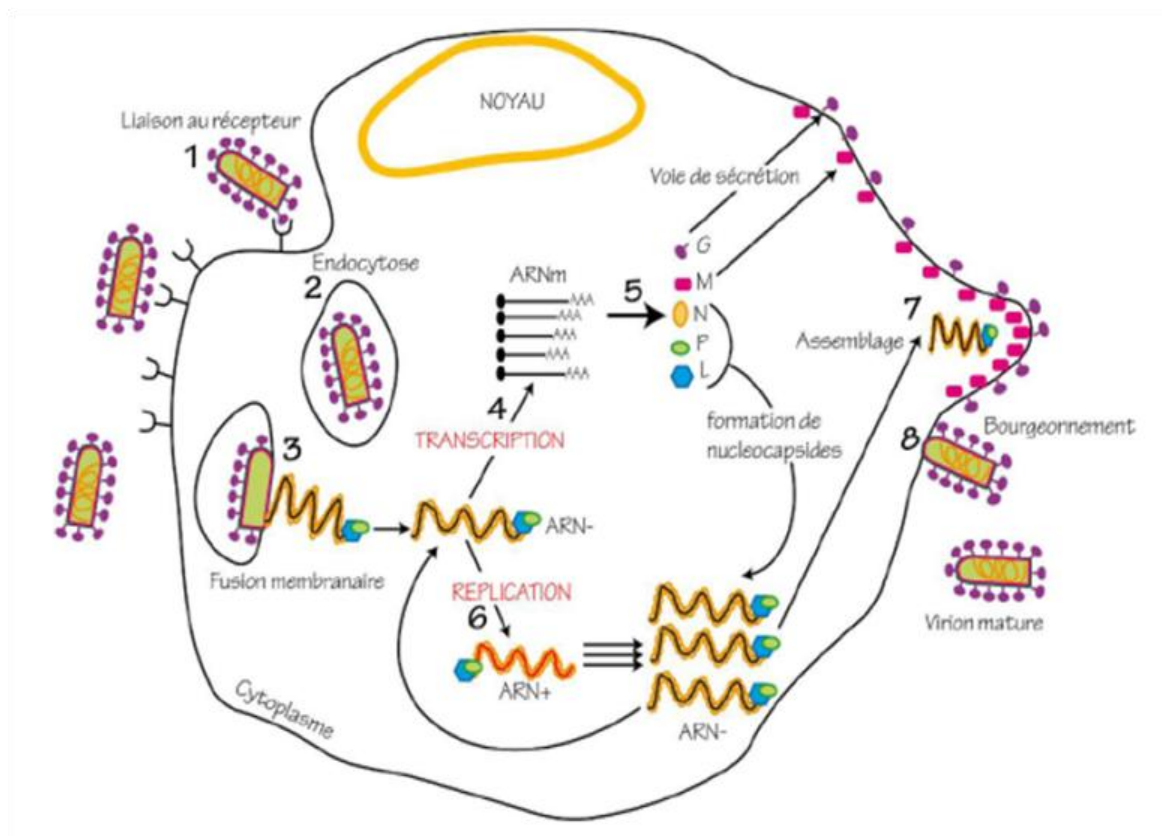


Figure 6: Le cycle viral. La première phase comporte la fixation au récepteur (1). La fusion des membranes peut avoir lieu à la surface de la cellule ou après internalisation (2) ce qui dans tous les cas amène à la libération dans le cytoplasme du contenu du virion (3). La deuxième phase est la phase de multiplication virale – la transcription (4) la synthèse des protéines virales (5) et la réplication (6). Dans la dernière phase les nouvelles particules virales sont assemblées (7) et bourgeonnent de la cellule hôte (8).

6.1. Entrée

Le cycle de réplication du virus nécessite l'introduction dans la cellule hôte du génome accompagné de la machinerie virale de synthèse d'ARN. Celui-ci débute donc par l'attachement du virus sur la membrane cellulaire, qui se produit par l'interaction spécifique entre les spicules de glycoprotéines, présentes à la surface du virus, et un récepteur cellulaire. La nature du récepteur cellulaire détermine le tropisme du virus. Ainsi, RaV qui est un virus neurotrope, utilisent différents récepteurs présents à la surface des neurones, notamment le récepteur nicotinique à l'acétylcholine (AChR) (Broughan and Wunner 1995), la molécule d'adhésion des cellules nerveuses (NCAM ou CD56) (Thoulouze, Lafage et al. 1998) et le récepteur à basse affinité du facteur de croissance neuronal (p75NTR) (Tuffereau, Benejean et al. 1998). Toutefois, RaV comme VSV semblent également utiliser des molécules réceptrices telles que des glycolipides (gangliosides) qui sont présentes sur différents types cellulaires même si l'affinité pour la protéine G est plus faible.

6.2. Les glycoprotéines

Les rhabdovirus possèdent une seule protéine de surface, la protéine G, alors que les paramyxovirus en possèdent deux, la protéine G (ou encore H ou NH) et la protéine F. Chez les rhabdovirus, la fusion des membranes se fait après l'endocytose du virus et l'acidification du milieu qui provoque un changement conformationnelle de la G. Cette protéine, particulièrement bien étudiée chez VSV, semble conservée parmi les différents genres des *Rhabdoviridae* (Roche, Albertini et al. 2008). Après le clivage du peptide signal N terminal, la G mature existe sous la forme de trimères, et est constituée d'un ectodomaine et d'une partie transmembranaire et peut se trouver au moins sous trois conformations différentes: état pré-fusion, état activé, et état post-fusion. Tous ces états existent en état d'équilibre dépendant du pH (Albertini, Baquero et al. 2012). Récemment des modèles cristallographiques de l'état pré- et post-fusion ont été obtenus dans lesquels l'ectodomaine est composé de quatre domaines distincts (Roche, Bressanelli et al. 2006; Roche, Rey et al. 2007). Ces deux structures ont montré que le changement de conformation entre l'état pré- et post-fusion nécessite une réorganisation profonde de la glycoprotéine (**Figure 7**).

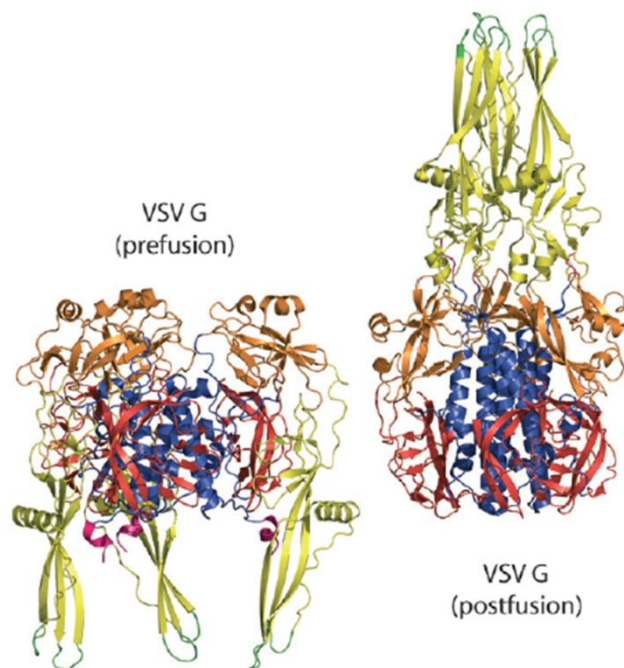


Figure 7 : Structure trimérique de l'ectodomaine de la glycoprotéine du VSV dans ses conformations pré- et post-fusion. Pré-fusion à gauche (PDB : 2J6J), post-fusion à droite (PDB : 2CMZ). Domaine I en rouge, domaine II en bleu, domaine III en orange, domaine IV en jaune, les boucles responsables de la fusion en vert et le C-terminal en magenta. Le domaine II a la même orientation dans les deux conformations (figure d'après (Roche, Albertini et al. 2008))

Il a été proposé qu'une telle réorganisation ne soit possible que par le passage de la G par une forme intermédiaire monomérique (Albertini, Merigoux et al. 2012). La plasticité conformationnelle de la protéine G des rhabdovirus est une propriété unique parmi les machineries virales de fusion dépendante du pH, différente de ce qui est observé avec les

protéines de fusion de type I et II, et une nouvelle classe de protéines de fusion de type III à été proposée (Albertini, Baquero et al. 2012). Il a été récemment montré que la chaperonne endoplasmique gp96 est nécessaire pour l'infection par VSV mais le(s) récepteur(s) cellulaire(s) reste inconnu(s) (Bloor, Maelfait et al. 2010).

Tous les *Paramyxoviridae* possèdent deux glycoprotéines de surface et certains rubulavirus et tous les pneumovirus en possèdent trois. Une des protéines (HN, H ou G) est responsable de la fixation alors que l'autre glycoprotéine (F) est responsable de la fusion indépendante du pH de l'enveloppe virale avec la membrane plasmique de la cellule hôte. Certains virus, les respirovirus et les rubulavirus, se fixent sur des récepteurs contenant de l'acide sialique et ont une glycoprotéine d'attachement qui possède les propriétés de fixer et d'agglutiner les globules rouges (H, hemagglutination) et a une activité neuraminidase (N). Les morbillivirus se fixent aux cellules hôtes par un mécanisme indépendant de l'acide sialique et n'ont pas d'activité neuraminidase associée à leur protéine de surface malgré la présence d'une activité d'hemagglutination. La glycoprotéine d'attachement des henipavirus ne possède aucune de ces deux activités (Yu, Hansson et al. 1998).

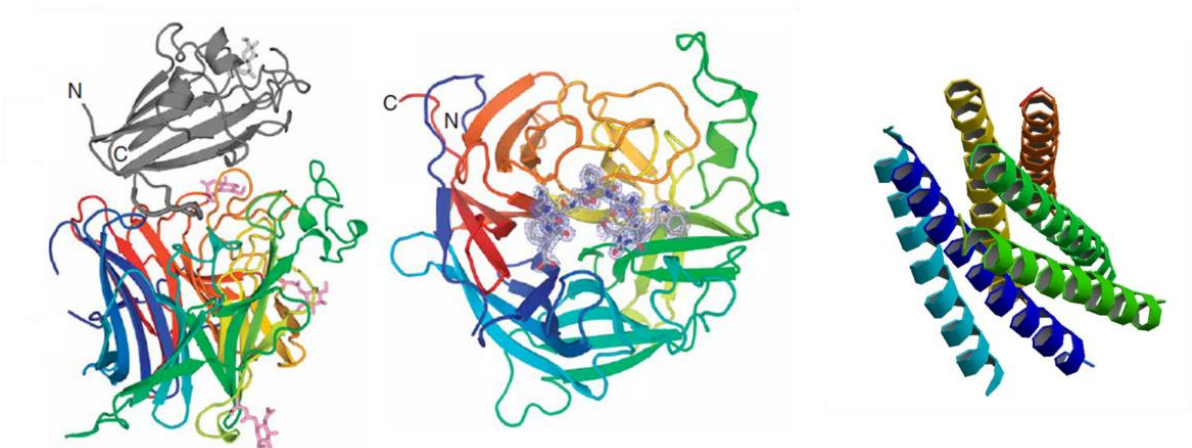


Figure 8 : Structures du complexe entre la glycoprotéine de NiV/HeV et son récepteur l'ephrinB2 et la protéine de fusion (F). Les structures des complexes entre les glycoprotéines de NiV et HeV et leur récepteur ephrinB2 sont pratiquement identiques. A gauche la structure du complexe entre la glycoprotéine de NiV et ephrinB2 ; la glycoprotéine est colorée du bleu (N-terminal) au rouge (C-terminal), ephrinB2 en gris (PDB : 2VMS pour NiV et 2VSK pour HeV ; figure d'après (Bowden, Aricescu et al. 2008) avec modifications). Au milieu la vue du haut. A droite le faisceau de six hélices de la protéine de fusion F (PDB : 1WP7 pour NiV et 1WP8 pour HeV).

La protéine G des Henipavirus est constituée d'une partie cytoplasmique, d'une région transmembranaire, d'une tige et d'une tête globulaire. La séquence en acide aminés du domaine globulaire est relativement peu conservée chez les Paramyxovirus, mais sa structure tertiaire en forme de hélice de bateau (propeller) est conservée (**Figure 8**).

Le récepteur cellulaire des HeV et NiV est l'ephrinB2 et/ou l'ephrinB3. Ces protéines appartiennent à la famille des ligands des glycoprotéines de surface qui fixent les récepteurs Ephrin. Ces molécules sont omniprésentes chez les vertébrés en particulier à la surface des cellules endothéliales des artères, des muscles lisses de la *tunica media* et à la surface des neurones, ce qui explique le spectre large d'hôtes et l'infection systémique observés (Bonaparte, Dimitrov et al. 2005; Negrete, Levroney et al. 2005).

La protéine F des Henipavirus est une protéine de fusion de type I qui contient deux domaines alpha hélicoïdaux (domaines dis de répétition de sept). Ces domaines forment un faisceau de six hélices pendant ou immédiatement après la fusion (**Figure 8**). Cette protéine F est aussi une protéine membranaire de type I et la forme biologique active est composée de deux sous-unités (F1 et F2) liés par un pont disulfure. La forme active est obtenue après le clivage du site RxxR du précurseur inactif F0 (x étant un aa quelconque). Dans le cas des Henipavirus la protéase impliquée dans la maturation de la F est la cathepsin I (Pager and Dutch 2005). Le peptide hydrophobe libéré du côté N-terminal de F1 est considéré comme le responsable de l'insertion de la F dans la membrane cellulaire. Comme pour la plupart des paramyxovirus, les protéines G et F sont nécessaires pour conduire la fusion virus - cellule hôte indépendamment du pH, contrairement à la fusion VSV et cellule hôte. En fait la protéine F est une protéine métastable dont le changement conformationnel, nécessaire pour la fusion membranaire, est irréversible. D'après un modèle récent la conformation pré-fusion de la F est stabilisée par l'interaction avec la G (Lee and Ataman 2011). C'est après l'attachement de la G sur son récepteur cellulaire que la protéine F peut changer sa conformation et entraîner le processus de fusion membranaire.

6.3. Interaction avec le cytosquelette et les moteurs moléculaires

Une fois que le virus a pénétré dans la cellule, il doit être transporté vers le site de réplication dans le cytoplasme ou le noyau. Etant donné l'encombrement moléculaire à l'intérieur de la cellule, les virus détournent les systèmes de transports intracellulaires pour circuler dans la cellule. Certains utilisent les mécanismes de transport le long des microtubules ou des filaments d'actine, d'autres exploitent les mécanismes d'import et d'export au travers les pores nucléaires.

La phosphoprotéine de RaV possède un motif linéaire d'interaction avec une chaîne légère de la dynéine (LC8) qui pourrait être impliquée dans un processus de transport (Raux, Flamand et al. 2000; Poisson, Real et al. 2001). Les phosphoprotéines de VSV et de RaV

possèdent également des motifs d'import et d'export nucléaire, reconnus par les karyophérines (importines et exportines) qui permettent la localisation de certains produits du gène P dans le noyau (Pasdeloup, Poisson et al. 2005; Oksayan, Wiltzer et al. 2012).

6.4. Multiplication virale

Le cycle se poursuit par la multiplication du virus. Dans un premier temps, la polymérase virale transcrit l'ARN viral sous la forme d'ARNm maturés correspondant à chaque gène – on parle de la transcription primaire. Chez les rhabdovirus et les paramyxovirus, la plupart des gènes conduisent à la production de protéines uniques, à l'exception du gène P qui est à l'origine de la production de différentes protéines, dont certaines sont impliquées dans la transcription et la réplication, alors que d'autres sont destinées à combattre le système immunitaire cellulaire (voir plus loin).

Dans un deuxième temps a lieu la réplication du génome. Par un mécanisme qui n'est pas encore compris, la polymérase virale se convertit de « transcriptase » en « réplacase » et génère des copies du génome entier, sans s'arrêter aux extrémités des gènes. Elle génère des copies de polarité positive (ou anti-génomes), qui sont par la suite utilisées comme matrice pour la synthèse de nouveaux génomes de polarité négative. Contrairement aux ARNm, les génomes et anti-génomes ne sont pas maturés mais ils sont encapsidés par la protéine N en même temps qu'ils sont synthétisés (Patton, Davis et al. 1984; Moyer, Smallwood-Kentro et al. 1991).

Finalement, une seconde phase de transcription, dite secondaire, est observée, vraisemblablement due à l'apparition de nouveaux génomes et de nouvelles molécules de polymérase dans la cellule hôte. Pendant cette étape on observe une production importante de protéines virales.

6.5. Assemblage et bourgeonnement

L'assemblage et le bourgeonnement de nouveau virus constituent la dernière étape du cycle viral. Pour l'assemblage des nouvelles particules virales, les génomes viraux néo-synthétisés et encapsidés par la protéine N, auxquels sont attachés la polymérase virale et la protéine P pour former la ribonucléoprotéine (RNP), sont exportés vers la membrane plasmique. Les glycoprotéines sont transportées par l'appareil de Golgi et se positionnent dans la membrane plasmique. Les protéines matrice M interagissent avec la partie cytoplasmique des protéines G ainsi qu'avec la RNP, conduisant à la condensation des RNPs.

La phase finale du cycle viral est le bourgeonnement de nouvelles particules virales à partir de la membrane cellulaire (Harty, Paragas et al. 1999; Harty, Brown et al. 2001).

6.6. La protéine de matrice (M)

La protéine de matrice joue un rôle structural dans l'organisation de la particule virale. Chez les *Rhabdoviridae*, elle assure la cohésion entre les glycoprotéines G insérées dans la membrane lipidique et la nucléocapside interne (Newcomb and Brown 1981; Mebatsion, Weiland et al. 1999). Chez certains paramyxovirus, l'interaction entre la M et la partie cytoplasmique des glycoprotéines virales a été mise en évidence par différentes méthodes (Schmitt, He et al. 1999; Moll, Klenk et al. 2002) tandis que chez d'autres virus une interaction entre la N et la G est observée (Iwasaki, Takeda et al. 2009). D'autre part la protéine M est essentielle pour l'assemblage et le bourgeonnement de nouvelles particules virales dans la cellule hôte. Dans certains cas, il a été démontré que l'expression cytoplasmique de cette protéine en absence d'autres protéines virales conduit à l'assemblage de particules ressemblantes à des virus – des VLPs (Virus-like Particles) (Ciancanelli and Basler 2006; Pohl, Duprex et al. 2007).

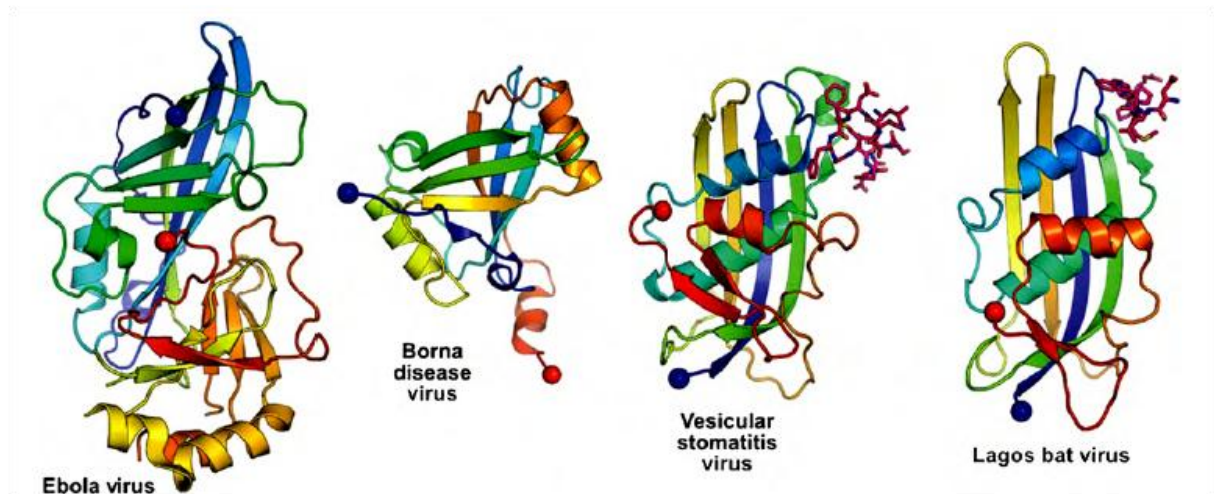


Figure 9 : Structures des protéines de matrice de différents MNV (figure d'après (Assenberg, Delmas et al. 2010)). De gauche à droite : Les structures des protéines de matrice des virus Ebola (VP40, Filoviridae, PDB : 1ES6), Borna (Bornaviridae, PDB : 3F1J), VSV et Lagos Bat (Rhabdoviridae, PDB : 1LG7 et 2W2R respectivement). Les structures sont colorées du bleu (N-terminal) au rouge (C-terminal).

Chez les *Rhabdoviridae*, la protéine M serait également impliquée dans la régulation de différent processus, notamment de la balance entre les activités de transcription et de réplication (Finke, Mueller-Waldeck et al. 2003), du bourgeonnement (Harty, Paragas et al. 1999), de la modulation de la traduction et de l'apoptose de la cellule hôte (von Kobbe, van

Deursen et al. 2000; Finke and Conzelmann 2005; Komarova, Real et al. 2007) (voir plus loin).

La structure de la M de plusieurs *MNVs* est connue (**Figure 9**). Elle consiste en un ou deux domaines de repliement similaires en sandwich de brins bêta avec une des faces chargée positivement ce qui facilite l'interaction avec les têtes polaires des lipides de la membrane cellulaire.

7. Machinerie de transcription/réplication et son fonctionnement

La réplication se déroule exclusivement dans le cytoplasme (Lamb and Parks 2007; Lyles and Rupprecht 2007) et conduit à la formation d'inclusions cytoplasmiques (Carlos, Young et al. 2009; Lahaye, Vidy et al. 2009; Heinrich, Cureton et al. 2010). Cependant, certains *Rhabdoviridae* de plantes se répliqueraient en partie dans le noyau (Goodin, Chakrabarty et al. 2007) et les *Bornaviridae*, dont les particularités structurales ne sont pas traitées ici, ont un cycle nucléaire (Lipkin and Briese 2007).

7.1. Machinerie moléculaire

La machinerie responsable de la transcription et de la réplication est un complexe ribonucléoprotéique (RiboNucléoProtéine, RNP) qui comprend l'ARN viral et les protéines L, P et N auxquelles s'ajoutent les protéines M2-1 et M2-2 des *Pneumovirinae* ou la protéine VP30 des *Filoviridae*. Ce complexe constitue le module infectieux minimal capable de produire de nouveaux virions à l'intérieur d'une cellule hôte (**Figure 4**) (Szilagyí and Uryvayev 1973). Au sein de ce complexe, l'ARN est encapsidé par un homopolymère non covalent de N parfaitement régulier couvrant le génome sur son intégralité. A l'entrée dans le cytoplasme, le complexe N-ARN, associé à des protéines P isolées indispensables à l'activité polymérase, sert de matrice pour le complexe polymérasique L/P ayant seulement une activité transcriptase (Kolakofsky, Le Mercier et al. 2004; Lamb and Parks 2007; Lyles and Rupprecht 2007). En effet, l'ARN génomique ou antigénomique nu et isolé n'est pas infectieux car il ne peut être ni transcrit ni traduit (Huang and Wagner 1966; Emerson and Yu 1975; Morin, Rahmeh et al. 2012). *In vitro*, sur un ARN nu, la protéine L purifiée est capable d'amorcer la synthèse d'un ARN complémentaire sur une longueur de quelques dizaines de nucléotides, alors que le cofacteur P et l'ARN sous forme encapsidé par de la N polymérique sont requis pour transcrire et répliquer le génome entier (Morin, Rahmeh et al. 2012; Morin and Whelan 2012; Rahmeh, Morin et al. 2012).

7.2. Transcriptase et transcription

Le promoteur transcriptionnel des *MNVs* est situé à l'extrémité 3' du génome en chevauchant le début du premier gène (Kolakovsky, Le Mercier et al. 2004; Lamb and Parks 2007; Lyles and Rupprecht 2007). *In vitro*, la polymérase des *Rhabdoviridae* et des *Paramyxoviridae* initie la transcription à l'extrémité 3' du génome et transcrit un court fragment d'ARN non-maturé correspondant à la région non-traduite « Leader ». Elle initie ensuite la transcription du premier gène et transcrit les gènes séquentiellement dans l'ordre N, P, M, G et L par un mécanisme d'arrêt/démarrage, produisant des ARNs messagers (ARNm) coiffés et polyadénylés et donc traduisibles par la machinerie cellulaire (Barr and Wertz 2001; Ogino and Banerjee 2011). *In vivo*, certains arguments expérimentaux suggèrent qu'elle puisse initier la synthèse d'ARN au début du gène N (Whelan and Wertz 2002). Dans ce cas, la polymérase éviterait la synthèse de l'ARN « Leader » et balayerait (« scan ») l'ARN jusqu'à trouver le signal d'initiation du premier gène. Lorsqu'elle atteint la première région intergénique et rencontre un signal de fin de gène, la polymérase fait une pause. Puis, suite à un balayage en aval - ou même amont pour certains gènes de *Filoviridae* et de *Pneumoviridae* ayant leur GS (gene start) en amont du GE (gene end) du gène qui le précède - sur au maximum une centaine de nucléotides, elle reconnaît le signal d'initiation GS du gène suivant et initie sa transcription (Kolakovsky, Le Mercier et al. 2004; Plattet, Strahle et al. 2007). La ré-initiation échoue dans une proportion variable selon les séquences intergéniques et la transcriptase devient inactive (**Figure 10**). Des arguments cinétiques en cellules infectées indiquent que cette inactivité est transitoire jusqu'à ce que la transcriptase se réamorce en retrouvant l'unique promoteur transcriptionnel à l'extrémité 3' du génome. Y-a-t-il migration non productive de la polymérase le long de la NC ou bien détachement puis ré-attachement à distance à l'extrémité 3' du génome ? Aucune donnée expérimentale ne permet de trancher (Plumet, Duprex et al. 2005).

Ce mécanisme de régulation, bien caractérisé chez plusieurs virus, en particulier VSV et SeV, se répète à chaque région intergénique et conduit à un gradient d'abondance des ARNm de chacun des gènes décroissant avec l'éloignement de l'extrémité 3' du génome en relation avec les besoins en protéines pour l'assemblage de nouvelles particules virales (**Figure 10**). L'ARNm codant pour la nucléoprotéine, dont le gène correspondant occupe la première place du génome, est produit de manière plus abondante que celui du deuxième gène qui est lui-même produit de manière plus abondante que celui du troisième gène et ainsi de suite. La nucléoprotéine, qui permet d'encapsider les antigénomes intermédiaires et les

nouveaux génomes, est la protéine virale qui doit être produite en plus grande quantité. Par contraste, la RdRp, dont le gène se situe à l'extrémité 5', doit être produite en moins grande proportion, puisque seulement quelques dizaines de copies sont incorporées dans chaque virion. Les ARNm ne sont pas encapsidés par la nucléoprotéine, mais ils sont coiffés et polyadénylés au moment de leur synthèse par la polymérase virale de sorte qu'ils sont confondus avec les ARNm cellulaires et sont donc reconnus et traduits par la machinerie de traduction cellulaire (**Figure 10**).

7.3. Réplicase et réplication

La complétion de la phase de transcription primaire conditionne le démarrage de la réplication. Dans cette seconde phase, la polymérase reconnaît le promoteur de réplication *sis* dans la région leader en amont du promoteur de transcription et elle synthétise une copie complémentaire complète du génome, appelée antigénome en ignorant les signaux de début et de fin de gène (Kolakofsky, Le Mercier et al. 2004). Puis, l'antigénome sert à son tour de matrice pour la synthèse de nouvelles copies du génome à partir du promoteur antigénomique *sis* dans la région trailer. Dans une cellule infectée, et malgré leur complémentarité parfaite, les ARNs génomique et antigénomique ne forment pas d'ARN double brin, car ils sont toujours étroitement encapsidés sous forme d'ARN simple brin par un homopolymère ininterrompu de N formant la nucléocapside (NC). En effet, la réplication se distingue de la transcription par l'encapsidation concomitante de l'ARN génomique (ou antigénomique) naissant grâce à l'assemblage de monomères de N, reconstituant ainsi le complexe N-ARN qui sert de matrice pour la polymérase. Ainsi, la réplication du génome nécessite la production continue de nouvelles molécules de N (Kolakofsky, Le Mercier et al. 2004). En vertu de son rôle cardinal dans l'encapsidation du génome, N est la protéine virale la plus abondante. La disponibilité de quantités suffisantes de molécules de N contrôlerait le basculement entre la transcription et la réplication (Kolakofsky, Le Mercier et al. 2004; Plumet, Duprex et al. 2005). Ainsi, dans les cellules infectées par les *MNV*, la protéine N existe soit sous forme monomérique soluble (appelée N^0) soit sous forme polymérique complexée à l'ARN viral (N-ARN). Comme la protéine N a une forte affinité pour l'ARN et une forte propension à s'auto-assembler, la production de N monomérique nécessite, dès sa biosynthèse, l'intervention d'un chaperon, la protéine P. En s'associant à N^0 sous la forme du complexe soluble N^0 -P, elle empêche l'interaction de N avec un ARN cellulaire et son auto-assemblage illégitime. Le complexe N^0 -P sert de substrat d'encapsidation des ARN génomiques et antigénomiques durant la réplication (voir pour revue (Lamb and Parks 2007;

Lyles and Rupprecht 2007) et références citées). La décision de produire des ARNm ou des génomes entiers doit donc être prise au moment de la synthèse de l'ARN complémentaire du « leader » et implique *a priori* la participation de la protéine N selon des mécanismes non élucidés. Chez RSV, le basculement entre les activités de transcription et de réplication impliquerait également la protéine M2-2. Chez les *Rhabdoviridae* la phosphorylation de P (VSV) ou de N (RAV) jouerait ce rôle (Pattnaik, Hwang et al. 1997; Toriumi, Eriguchi et al. 2004).

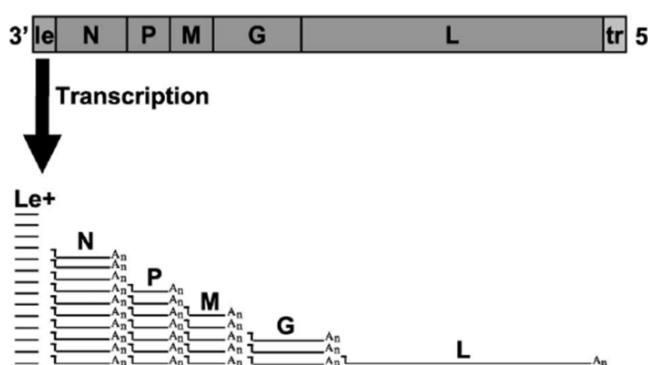


Figure 10 : Schéma de la transcription chez les MNV ; exemple des *Rhabdoviridae*. (figure d'après (Barr, Whelan et al. 2002) avec modifications). De 3' vers 5' la région leader (le) et les gènes codant pour la nucléoprotéine (N), la phosphoprotéine (P), la protéine de matrice (M), la glycoprotéine (G), la polymérase virale (L) et la région trailer (tr). L'abondance relative des produits de la transcription, le leader et les 5 ARNm maturés et polyadénylés sont indiqués. Les ARNm viraux ne sont pas encapsidés par la N.

8. Organisation structurale des nucléocapsides

8.1. Structure de l'homopolymère nucléoprotéique

L'ARN génomique des *MNV* est encapsidé par la nucléoprotéine sous la forme d'une nucléocapside hélicoïdale et le même complexe N-ARN sert de matrice à la fois pour la transcription et pour la réplication. Dans ces complexes, l'ARN est fortement lié à la N, résistant à la centrifugation en gradient de CsCl à froid (Lynch and Kolakofsky 1978). La fixation de N permet à la fois de maintenir l'ARN viral dans une forme simple brin, évitant la formation d'ARN double brin qui constituerait un MAMP (Microbe Associated Molecular Pattern) reconnu par les récepteurs de l'immunité innée (Chen, Short et al. 2010), et de protéger l'ARN viral contre les nucléases et les ARN interférant (Iseni, Baudin et al. 2000; Bitko, Musiyenko et al. 2005). En microscopie électronique, les nucléocapsides de *MNV*, extraites de virus ou de cellules infectées ou exprimées sous forme recombinante, apparaissent comme de longues hélices irrégulières. Les nucléocapsides de *Rhabdoviridae* sont très souples et ressemblent à un câble de combiné téléphonique surenroulé. Celles des *Paramyxoviridae* forment des tubes creux avec une morphologie plus régulière en forme d'épi

(Albertini, Schoehn et al. 2005) (**Figure 11**). L'apparence striée de ces structures vues en microscopie électronique correspond à l'arrangement côte-à-côte des nucléoprotéines le long de l'ARN.

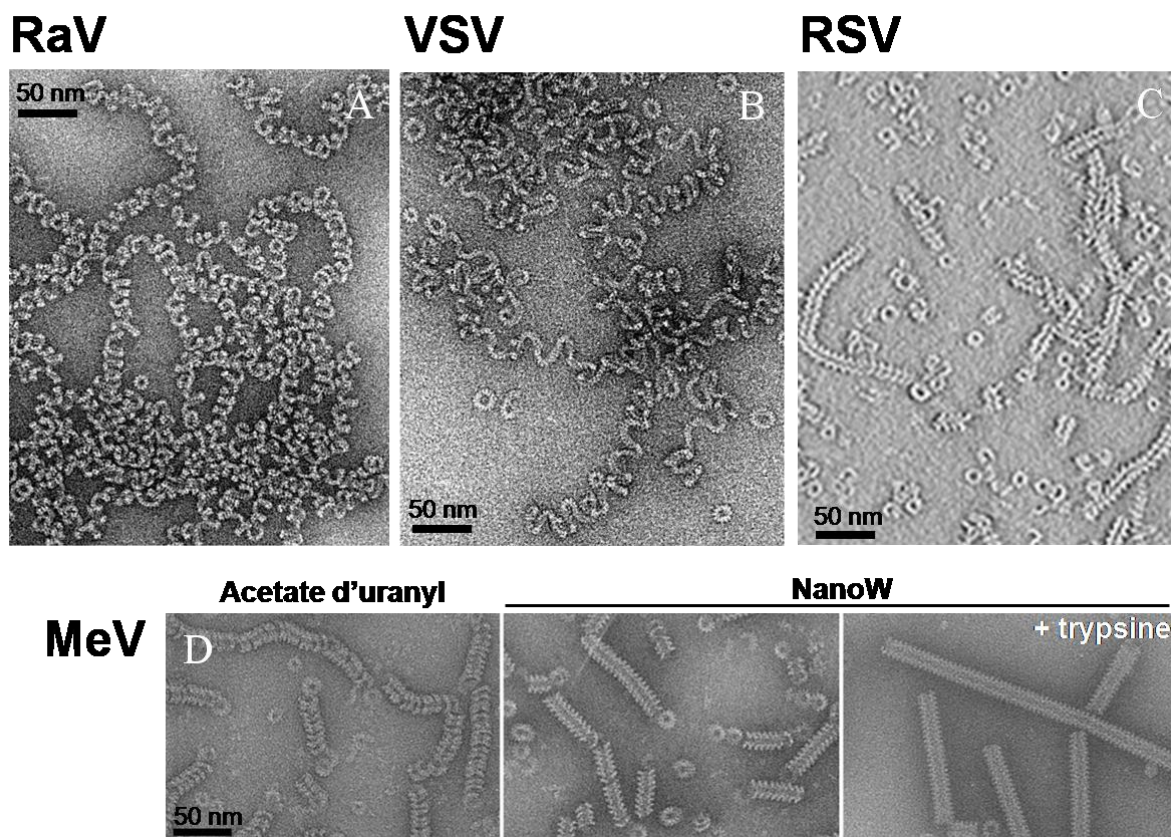


Figure 11: Les nucléocapsides (NC) des *MNV* (figure d'après Blocquel, D. et al. Virologie E). Images de microscopie électronique en coloration négative. **A**) NC virales du virus de la Rage (RaV). **B**) NC virales du virus de la stomatite vésiculaire (VSV). **C**) NC recombinantes du virus respiratoire syncytial (RSV). **D**) NC du virus de la rougeole (MeV) dans différents colorants ; NC intactes (gauche) et trypsinées (droite).

La nucléocapside peut adopter différentes conformations plus ou moins compactes caractérisées par différents pas d'hélice (Egelman, Wu et al. 1989; Schoehn, Iseni et al. 2001; Bhella, Ralph et al. 2002; Bhella, Ralph et al. 2004; Schoehn, Mavrikakis et al. 2004). Dans la conformation la plus étendue de la nucléocapside du SeV (pas hélicale de 37.5 nm contre 5.3 nm pour la forme la plus compacte), l'ARN viral est probablement plus accessible à la polymérase. Il est tentant d'imaginer qu'un tel phénomène de "désenroulement" pourrait intervenir au moment de la synthèse d'ARN pour permettre à la polymérase d'accéder à l'ARN génomique enfoui entre les mâchoires des protomères de N.

Chez les *MNV*, chaque monomère de N interagit avec un nombre précis de nucléotides, caractéristique de chaque famille : 6 pour les *Paramyxovirinae* (Calain and Roux 1993), 7 pour les *Pneumovirinae* (Tawar, Duquerroy et al. 2009), 9 pour les *Rhabdoviridae* (Thomas, Newcomb et al. 1985; Iseni, Barge et al. 1998) et entre 12 et 15 pour les *Filoviridae*

(voir (Albertini, Schoehn et al. 2005) et références citées). Ainsi, la longueur du génome des *Paramyxovirinae* obéit à une "règle de six" stricte selon laquelle le génome doit être constitué d'un multiple entier de six nucléotides pour assurer une réplication efficace (Roux 2005). Cependant, même si chez les *Rhabdoviridae*, le nombre de molécules de N dans le virion correspond approximativement aussi au nombre théorique nécessaire pour couvrir le génome entier (Thomas, Newcomb et al. 1985), les génomes des *Pneumovirinae* et des *Rhabdoviridae* n'ont pas nécessairement une longueur multiple exact du nombre de nucléotides couvert par un protomère de leur N.

Dans les cellules infectées, N se fixe exclusivement sur l'ARN génomique viral. Cependant, exprimées isolément en bactéries ou en cellules eucaryotes, les N des *Rhabdoviridae* et des *Paramyxoviridae* s'auto-assemblent sur des ARNs cellulaires formant des structures nucléocapsidiques de taille hétérogène mais de forme identique à celles contenant l'ARN viral (Spehner, Kirn et al. 1991; Warnes, Fooks et al. 1995; Iseni, Barge et al. 1998; Green, Macpherson et al. 2000; Schoehn, Iseni et al. 2001; Bhella, Ralph et al. 2002; Tan, Ong et al. 2004; Kerdiles, Cherif et al. 2006; Albertini, Schoehn et al. 2008). Fort opportunément, la N de certains virus forme également des anneaux composés d'un petit nombre de protomères de N, de 9 à 13 chez RaV avec majorité d'anneaux à 10 sous-unités (Iseni, Barge et al. 1998; Albertini, Wernimont et al. 2006; Albertini, Clapier et al. 2007), essentiellement des anneaux de 10 protomères pour VSV (Green and Luo 2006; Green, Zhang et al. 2006) et 10 ou 11 protomères pour RSV (Tawar, Duquerroy et al. 2009). Chez les *Rhabdoviridae*, ces anneaux ont un diamètre et un nombre de sous-unités comparables au premier tour de l'extrémité conique des particules virales (Albertini, Wernimont et al. 2006; Ge, Tsao et al. 2010; Ivanov, Yabukarski et al. 2011; Desfosses, Ribeiro et al. 2013) et chez RSV, ils ont le même diamètre et le même nombre de sous-unités que les nucléocapsides hélicoïdales, suggérant qu'un anneau correspond à un tour d'hélice refermé sur lui-même par des interactions N-N (Tawar, Duquerroy et al. 2009). En revanche, la N des *Paramyxovirinae* ne semble pas former d'anneaux (Bhella, Ralph et al. 2002).

Des anneaux homogènes en taille de RaV, VSV et RSV ont été cristallisés, révélant ainsi les détails de la structure polymérique de N et des interactions entre N et l'ARN (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006; Tawar, Duquerroy et al. 2009) (**Figure 12**). La protéine N des *Rhabdoviridae* et des *Pneumovirinae* possède deux domaines globulaires, un domaine N-terminal (N_{NTD}) et un domaine C-terminal (N_{CTD}) qui sont connectés entre eux par une région charnière. La molécule d'ARN est fixée dans un sillon

délimité par l'interface entre les deux domaines qui semblent agir comme les mâchoires d'une pince emprisonnant entièrement la molécule d'ARN (**Figure 12**).

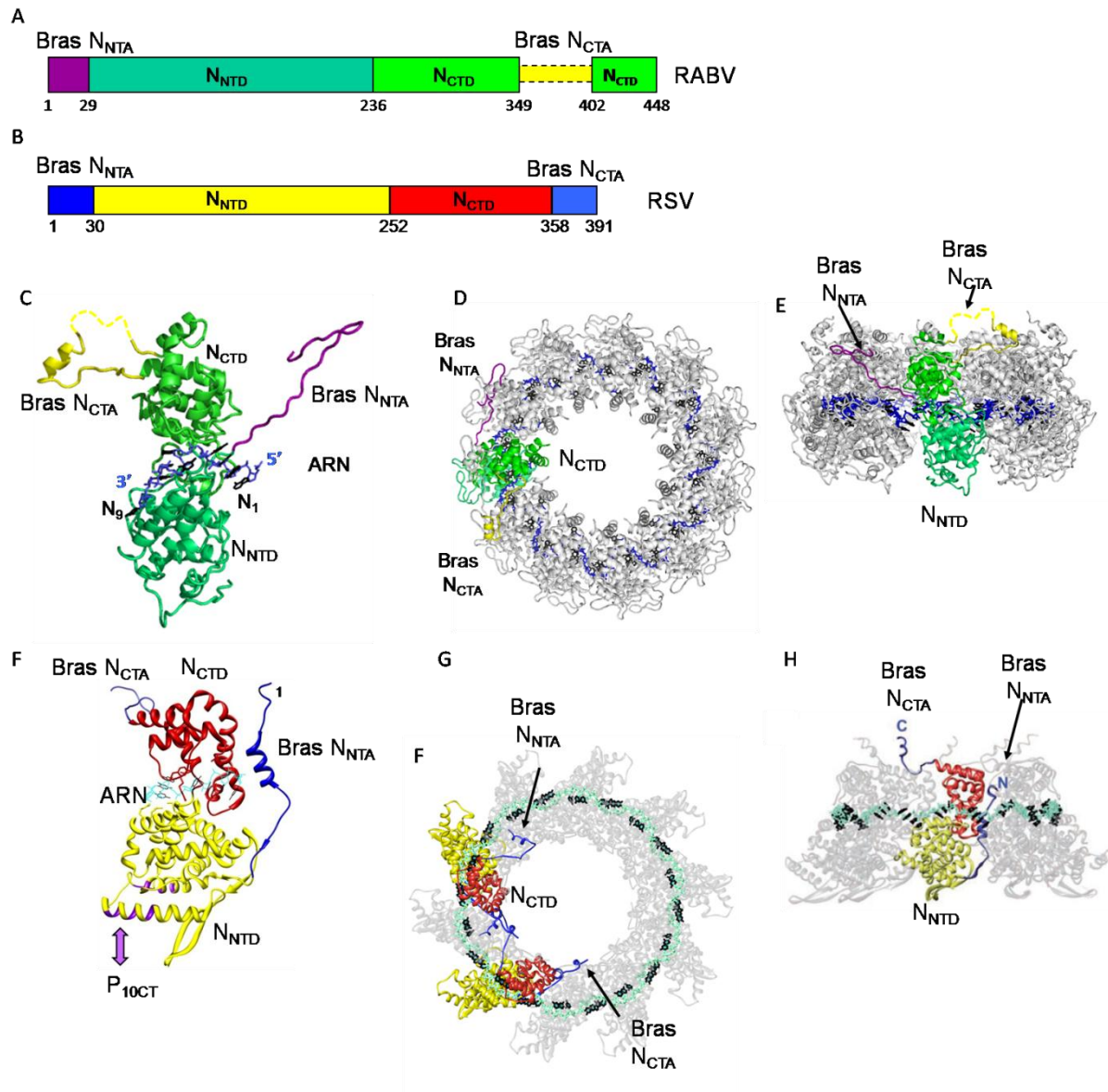


Figure 12: Comparaison de l'organisation et la structure de la nucléoprotéine et des anneaux de RaV et RSV (figures d'après (Tawar, Duquerroy et al. 2009) et Blocquel, D. et al. *Virologie E* 2012) Schéma de l'organisation des domaines et sous-domaines de **A)** RaV et **B)** RSV. La structure de la N de **C)** RaV et **F)** RSV représentée en cartoon. Les anneaux de **D)** et **E)** RaV et **G)** et **H)** RSV vue du haut et du côté respectivement. Le même code couleur est utilisé pour les figures A, C, D et E et B, F,G et H respectivement. Dans figure F) les résidus de la N qui interagissent avec la partie C-terminale de la P sont colorés en violet.

La région de N en contact avec l'ARN est riche en résidus basiques dont certains sont conservés au sein des différentes familles (Luo, Green et al. 2007; Ivanov, Yabukarski et al. 2011; Ruigrok, Crepin et al. 2011), et la fixation de l'ARN implique la formation de ponts salins entre les résidus chargés positivement de la protéine et les groupes phosphates chargé négativement du squelette de l'ARN (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006; Tawar, Duquerroy et al. 2009). Ces structures ont confirmés la fixation de 9 nucléotides

par les protéines N de RaV et VSV et de 7 nucléotides par la protéine de RSV. Il faut noter que les N de VSV et RaV possèdent des structures très similaires malgré une faible identité de séquence d'environ 15% (**Figure 13**). Néanmoins les résidus responsables de l'interaction avec l'ARN sont très conservés. La fixation par N protège essentiellement le squelette sucre-phosphate de l'ARN, expliquant la résistance aux nucléases, et laisse certaines bases accessibles au solvant (Iseni, Baudin et al. 2000). Chez les *Rhabdoviridae*, sur les 9 nucléotides fixés à chaque protomère de N, les bases des nucléotides (nt) 3, 4 et 6 sont enfouies dans la protéine. Les autres sont exposées au solvant (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006), avec celles des nt 7, 8 et 9 empilées avec les bases 1 et 2 des nt fixés par le protomère aval. La base du nt 5 est libre.

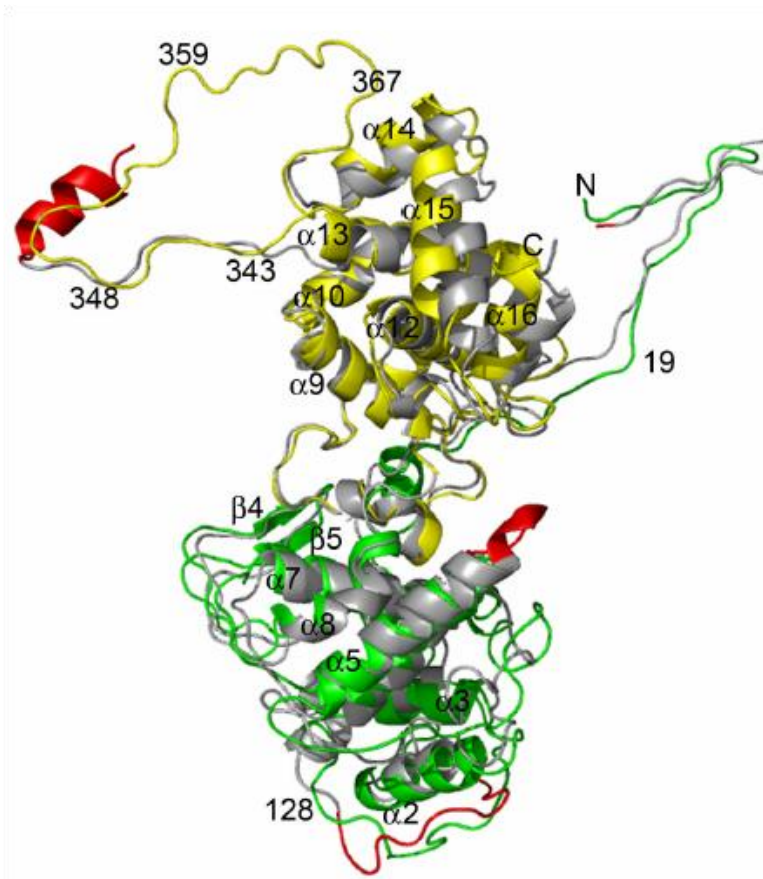
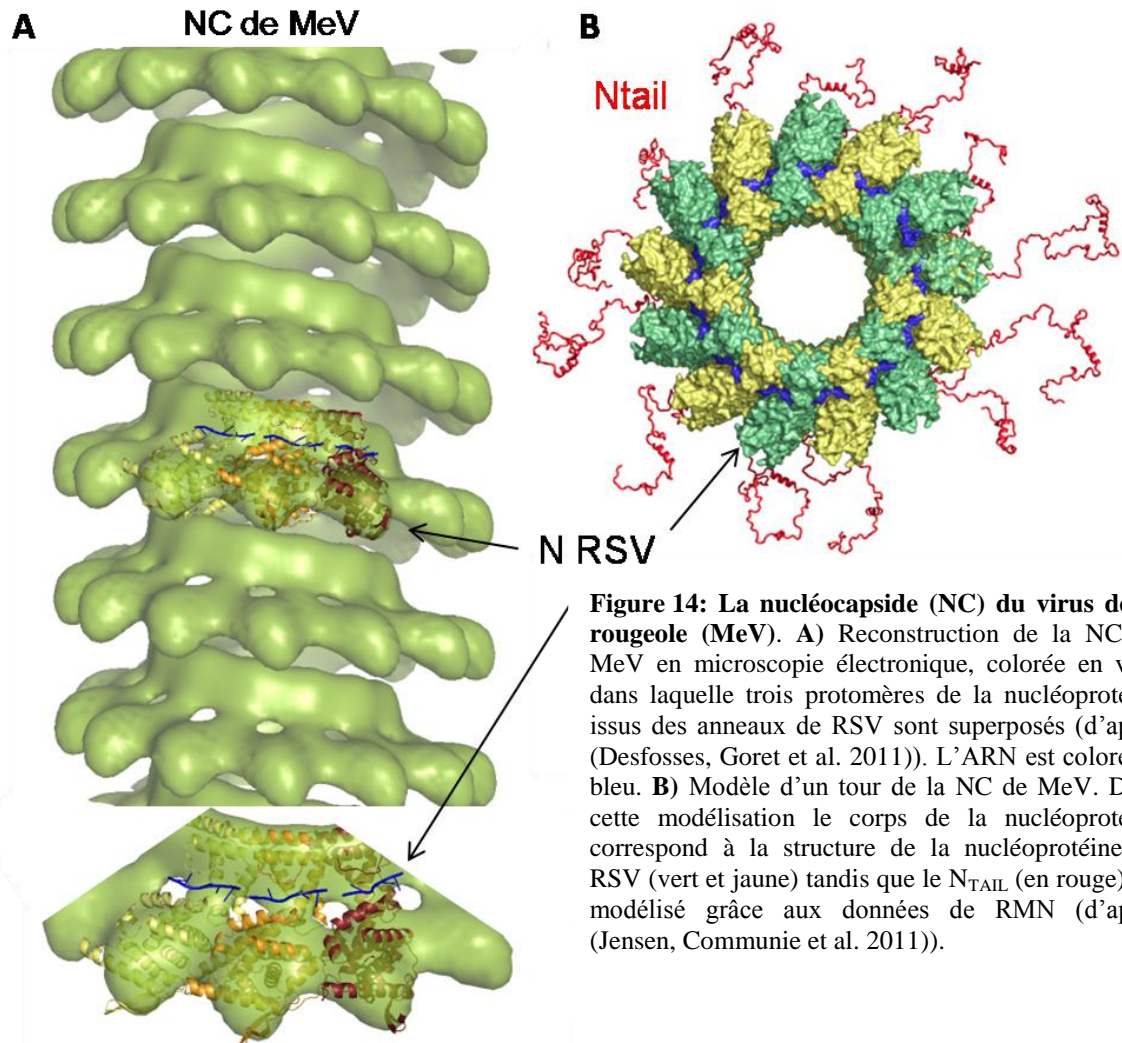


Figure 13 : Comparaison des structures des nucléoprotéines de VSV e RaV (figure d'après (Luo, Green et al. 2007)). Les structures de la nucléoprotéine de VSV (PDB code : 2GIC) et RaV (PDB code : 2GTT) sont superposées. Pour VSV N le domaine N-terminal est coloré en vert et le domaine C-terminale en jaune et les hélices sont numérotées. RaV N est coloré en gris et les différences structurales en rouge.

Chez RSV, l'organisation est similaire avec les bases des nt 2, 3 et 4 enfouies dans la protéine et celles des nt 5, 6 et 7 empilées avec la base 1 fixée sur le protomère aval (Tawar, Duquerroy et al. 2009). La modélisation de la structure atomique de N de RSV dans la carte de densité de cryo-microscopie électronique de MeV (**Figure 14**) prédit comme attendu une organisation similaire avec toutefois six nucléotides par protomère de N (Desfosses, Goret et al. 2011). Les bases orientées vers l'extérieur de la structure pourraient constituer des signaux, notamment pour l'encapsidation ou la polymérase virale, même si les analyses de séquences ne sont pas en faveur d'un tel schéma. Ces structures suggèrent que N doit s'ouvrir au

passage de la polymérase de manière à rendre l'ARN accessible (Albertini, Wernimont et al. 2006). L'ARN peut être déplacé ou échangé de N sans affecter la structure polymérique à 37°C et au-delà (Green, Rowse et al. 2011). La formation de l'homopolymère de N ne nécessite pas la présence d'ARN comme le montre la structure du mutant S290W qui est dépourvue d'ARN (Zhang, Green et al. 2008). Donc, cela est compatible avec une persistance de la forme polymérique de N lors du passage de la polymérase et à une ré-encapsulation consécutive du brin d'ARN ayant servi de matrice.



Outre les deux domaines globulaires, les protéines N de *MNV* possèdent un sous-domaine N-terminal, appelé le bras N-termina (N_{NTA}), et un sous-domaine C-terminal, appelé le bras C-terminal (N_{CTA}), qui se structurent partiellement ou entièrement dans le contexte du polymère de N (**Figures 12, 13, 15**). N_{NTA} et N_{CTA} assurent l'assemblage et la cohésion du polymère de N par des contacts avec les protomères amont et aval adjacents. Chez les *Rhabdoviridae*, N_{CTA} consiste en une boucle plus ou moins longue et plus ou moins flexible située à l'intérieur du domaine C-terminal (**Figures 12, 13, 15**). Chez la N de RaV, une

grande partie de cette boucle n'est pas visible dans la structure cristallographique du complexe N-ARN indiquant une grande flexibilité. Chez VSV, la boucle N_{CTA} est plus courte et apparaît entièrement structurée dans certains protomères. Des mutations perturbant l'agencement de cette boucle accroissent la réplication sans modifier la transcription (Harouaka and Wertz 2009). Le bras N_{CTA} du protomère N_{M+1} se fixe sur le domaine C-terminal du protomère aval N_M et le bras N_{NTA} du protomère N_{M-1} se fixe sur le domaine C-terminal du protomère amont N_M . Ainsi le bras N_{CTA} du protomère N_{M+1} et le bras N_{NTA} du protomère N_{M-1} sont en contact sur la surface du N_{CTD} du protomère N_M (**Figures 12 et 15**). La mutation F348A prédite pour rompre ce contact entre bras N_{NTA} et N_{CTA} contigus abolit sélectivement la transcription sans affecter la réplication (Harouaka and Wertz 2009). Ainsi donc, les contacts à distance entre protomères de N semblent mis en jeu de manière différentielle lors des étapes de transcription et de réplication. Dans les anneaux de RSV (*Pneumovirinae*), le bras N_{NTA} se fixe d'une manière comparable à celle observée chez les *Rhabdoviridae*, mais le bras N_{CTA} reste localisé à proximité du N_{CTD} , bien que les 12 à 20 derniers résidus C-terminaux soient désordonnés dans la structure cristallographique (**Figures 12, 15**).

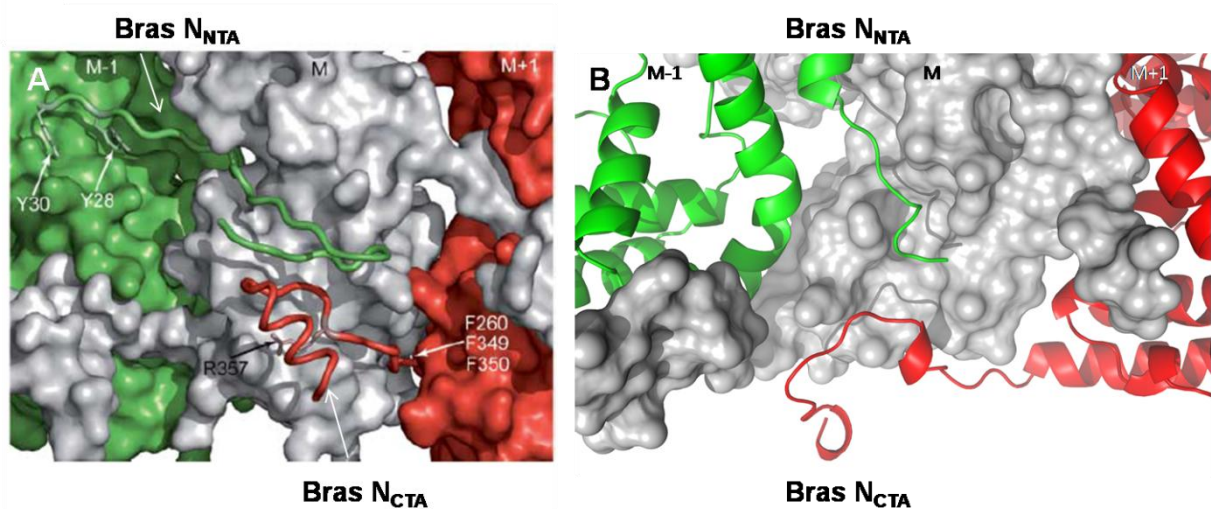


Figure 15: Les sous-domaines de la nucléoprotéine responsables de son oligomérisation. A) Rav et B) RSV. Dans les anneaux le sous-domaine N-terminal (N_{NTA}) du protomère M-1 et le sous domaine C-terminal (N_{CTA}) du protomère M+1 se fixent sur le domaine C-terminal (N_{CTD}) du protomère M. Le protomère M-1 est coloré en vert et le protomère M+1 est coloré en rouge. Pour RaV les sous domaines N-terminal de protomère M-1 et C-terminal du protomère M+1 sont en contact entre eux à la surface de protomère M (figure à gauche d'après (Albertini, Wernimont et al. 2006) avec modifications; figure à droite code PDB 2WJ8).

A côté du rôle joué par les bras N_{NTA} et N_{CTA} , les N_{CTD} adjacents participent à la cohésion du polymère N-ARN grâce à une surface d'interaction de type mixte associant des contacts hydrophobes et polaires. Par contre, les N_{NTD} n'interagissent pas, d'où une certaine souplesse dans le positionnement des sous-unités N les unes par rapport aux autres (Albertini,

Wernimont et al. 2006; Green, Zhang et al. 2006). Chez RSV, les interfaces entre N_{NTD} et N_{CTD} adjacentes sont très hydratées et n'impliquent que des contacts de van der Waals intermittents (Tawar, Duquerroy et al. 2009). Les interactions limitées aux interfaces entre N_{CTD} chez RaV et VSV ou très lâches entre les protomères de N chez RSV autorisent par conséquent des grandes variations des pas d'hélice et des tors des nucléocapsides comme observé en microscopie électronique (Schoehn, Iseni et al. 2001; Bhella, Ralph et al. 2004; Albertini, Clapier et al. 2007).

Une différence majeure entre la structure de la nucléocapside des *Rhabdoviridae* et celle des *Paramyxoviridae* (**Figure 12**) réside dans le positionnement de la molécule d'ARN. Dans la structure cristallographique des anneaux de RaV (Albertini, Wernimont et al. 2006) et de VSV (Green, Zhang et al. 2006), l'ARN se situe sur la face concave du complexe à l'intérieur de l'anneau, alors que dans celle des anneaux du RSV (**Figure 12**) (Tawar, Duquerroy et al. 2009) et dans la reconstruction par cryo-microscopie électronique de la nucléocapside du MeV (**Figure 14**) (Desfosses, Goret et al. 2011), l'ARN est situé sur la face convexe à l'extérieur de l'anneau. Ces structures suggèrent donc *a priori* une organisation différente de la nucléocapside, bien que l'on ne puisse exclure chez l'un des groupes une possible inversion de conformation au moment du passage du complexe polymérisique. Un tel scénario est cependant peu plausible compte tenu des contraintes structurales.

Une autre différence réside dans l'organisation de N et notamment de la partie C-terminale de N qui participe à la fixation de la P sur la forme polymérique N-ARN. Chez les *Rhabdoviridae*, hormis le bras N_{CTA}, la N est parfaitement structurée (Schoehn, Iseni et al. 2001; Albertini, Wernimont et al. 2006), alors que chez les *Paramyxovirinae*, la nucléoprotéine est constituée de deux parties : la partie N-terminale, dénommée N_{CORE}, (approximativement résidus 1 à 400 chez MeV, SeV, et chez les *Henipavirus*) est prédite structurée en deux domaines N_{NTD} et N_{CTD} responsables de l'auto-assemblage et de la liaison à l'ARN (Buchholz, Spehner et al. 1993; Curran, Homann et al. 1993; Bankamp, Horikami et al. 1996; Liston, Batal et al. 1997; Myers, Pieters et al. 1997; Myers, Smallwood et al. 1999; Karlin, Longhi et al. 2002; Kingston, Hamel et al. 2004), et la partie C-terminale, dénommée N_{TAIL} (résidus 401 à 525 pour MeV, 401 à 524 pour SeV et 400 à 532 pour les *Henipavirus*), est désordonnée et fixe P. Chez les *Pneumovirinae*, l'appendice C-terminal est court (16 résidus) et n'intervient pas dans la fixation de P (Galloux, Tarus et al. 2012).

8.2. Structure désordonnée de l'appendice N_{TAIL} des *Paramyxovirinae*

N_{TAIL} des *Paramyxovirinae* présente les caractéristiques typiques de régions désordonnées, à savoir : (i) une hyper-sensibilité à la protéolyse (Karlin, Longhi et al. 2002; Habchi, Blangy et al. 2011), (ii) une variabilité de séquence au sein des membres phylogénétiquement proches (Habchi, Mamelli et al. 2010), (iii) un biais compositionnel en acides aminés (Karlin, Ferron et al. 2003; Longhi, Receveur-Brechot et al. 2003; Bourhis, Johansson et al. 2004; Habchi, Mamelli et al. 2010) avec notamment une richesse en résidus promoteurs de désordre et une pauvreté en résidus promoteurs d'ordre, (iv) l'impossibilité à être visualisé en microscopie électronique (Bhella, Ralph et al. 2004). D'après des approches hydrodynamiques et spectroscopiques, la nature désordonnée des domaines N_{TAIL} isolés est confirmée, bien qu'ils ne soient pas complètement étendus en solution et qu'ils conservent une certaine compacité résiduelle associée à un certain degré de structure secondaire fluctuante (Longhi, Receveur-Brechot et al. 2003; Bourhis, Johansson et al. 2004; Houben, Marion et al. 2007; Habchi, Mamelli et al. 2010).

D'après des reconstructions par microscopie électronique (Bhella, Ralph et al. 2004; Schoehn, Mavrakis et al. 2004), l'élimination de la région N_{TAIL} partiellement exposée à la surface de la nucléocapside (Heggeness, Scheid et al. 1980; Heggeness, Scheid et al. 1981; Karlin, Longhi et al. 2002), conduit à une augmentation de la rigidité de la nucléocapside (**Figure 11D**) avec des changements significatifs dans le pas de l'hélice et dans le tors (Longhi, Receveur-Brechot et al. 2003; Bhella, Ralph et al. 2004; Schoehn, Mavrakis et al. 2004; Desfosses, Goret et al. 2011). Bien que les domaines N_{TAIL} désordonnés ne soient pas résolus dans la reconstruction de la nucléocapside N-ARN du MeV, l'origine de l'appendice N_{TAIL} désordonné est positionné à l'intérieur de la nucléocapside (**Figure 14A**) (Desfosses, Goret et al. 2011). L'analyse par RMN combinée à la diffusion de rayons X aux petits angles (SAXS) a confirmé que N_{TAIL} de MeV reste désordonné *in situ* dans le contexte des nucléocapsides intactes, d'où la proposition d'un modèle de l'organisation de N_{TAIL} dans la nucléocapside (Ringkjøbing Jensen, Communie et al. 2011). La région N_{TAIL} conserve sa flexibilité au sein des nucléocapsides intactes, à l'exception des 50 premiers acides aminés de N_{TAIL} qui sont conformationnellement restreints (**Figure 14B**).

La région N_{TAIL} comprend un ou plusieurs éléments de reconnaissance moléculaires ou MoREs (Molecular Recognition Elements), qui sont de courtes régions au sein d'une région désordonnée ayant une propension à former des structures secondaires et à subir un repliement induit lors de la fixation à un partenaire (Garner, Romero et al. 1999). Pour

différents paramyxovirus, un MoRE constitué d'hélices α transitoires (α -MoRE) et impliqué dans l'interaction avec le domaine C-terminal X de P (domaine PX ou encore P_{XD}) et dans le repliement α -hélical de N_{TAIL} a été localisé (Bourhis, Johansson et al. 2004) (Habchi, Blangy et al. 2011; Blocquel, Habchi et al. 2012). Cet α -MoRE est partiellement pré-configuré en solution en l'absence de P_{XD} (Morin, Bourhis et al. 2006; Belle, Rouger et al. 2008; Gely, Lowry et al. 2010) (Martinho, Habchi et al. 2012). D'après des approches quantitatives combinant des mesures de couplages résiduels dipolaires et des méthodes d'optimisation d'ensembles (Ringkjøbing Jensen, Communie et al. 2011), cette région existe en tant qu'ensemble conformationnel comprenant une forme étendue et quatre conformères hélicaux de différentes longueurs.

9. La polymérase virale

La polymérase virale est constituée de deux sous-unités, une grande sous-unité appelée protéine L (Large) qui porte différentes activités catalytiques et une petite sous-unité non-catalytique, la phosphoprotéine (P), qui assure la fixation de la polymérase sur la matrice N-ARN. La synthèse de longs ARN complémentaires nécessite que l'ARN soit encapsidé par un homopolymère de nucléoprotéine. Toutefois, chez VSV il a été récemment montré, que la L seule peut utiliser de l'ARN nu comme matrice et synthétiser de très courts brins d'ARN complémentaires et que la présence de P favorise l'initiation de la synthèse d'ARN et augmente la processivité de la L (Morin, Rahmeh et al. 2012; Rahmeh, Morin et al. 2012). Ces effets seraient liés à un changement de conformation de la L induit par la P qui conduirait à la dimérisation de la polymérase (Rahmeh, Morin et al. 2012). La région de la P, responsable de cette réorganisation de la L, réside dans la partie N terminale, entre les aa 40 à 106. Ainsi, en plus de son rôle d'amener la L sur sa matrice N-ARN, la P jouerait un deuxième rôle de facteur de processivité. Chez RaV, il a été démontré par des tests de co-immunoprécipitation que la même région N-terminale de P interagit avec la L (Chenik, Schnell et al. 1998).

9.1. Architecture de la protéine L

L est une grande protéine d'environ 2.200 acides aminés (~250 kDa), dans laquelle six régions conservées ont été identifiées (RC I à VI) (Poch, Blumberg et al. 1990). La protéine L est une protéine multienzymatique qui possède les différentes activités catalytiques

nécessaires pour la synthèse de nouvelles copies du génome et pour la synthèse d'ARNm et leur maturation avec synthèse et méthylation de la coiffe à leur extrémité 5' et polyadénylation de leur extrémité 3'. Certaines fonctions ont pu être cartographiées dans la séquence en acides aminés. La région conservée III (RC III) comprend la triade catalytique GDNQ responsable de la coordination d'un ou deux ions divalents essentiels pour l'activité polymérasique. Chez les Henipavirus cette séquence est GDNE (Harcourt, Tamin et al. 2001). Chez VSV, la substitution de l'aspartate central résulte en l'inactivation de la polymérase (Sleat and Banerjee 1993). Cette région est aussi nécessaire pour la polyadénylation des ARNm, qui résulte d'un mécanisme itératif dans lequel la polymérase « dérape » sur une courte séquence poly-U à l'extrémité 5' terminale de chaque gène ((Morin and Whelan 2012) pour revue). Les régions V et VI sont responsables de la formation et de la méthylation de la coiffe. Chez RaV la partie C-terminale de la L serait responsable de l'interaction avec le cofacteur P (Chenik, Schnell et al. 1998).

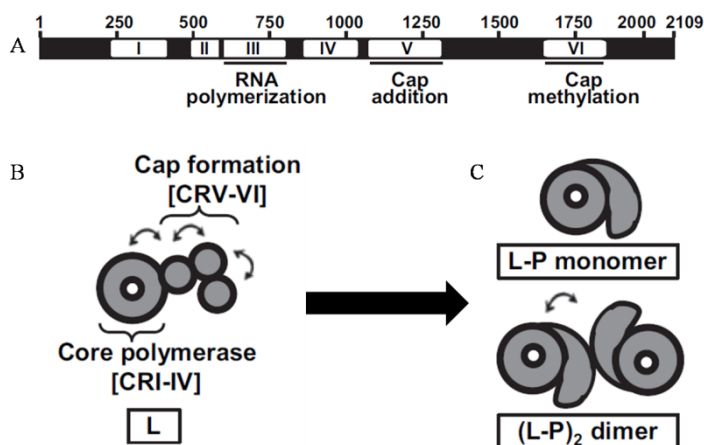


Figure 16: Architecture moléculaire de la polymérase (L) de VSV. A) Les domaines de I à VI représentent les régions conservées. B) Organisation de la L en anneau contenant l'activité polymérase et trois appendices flexibles responsables de la formation de coiffe maturée. C) Réorganisation moléculaire de la L après interaction avec la phosphoprotéine (P). La L prend la forme du nombre six et également observé comme monomère ou dimère (d'après (Rahmeh, Schenk et al. 2010)).

Les images de microscopie électronique ont montré que la L de VSV est constitué d'un domaine en forme d'anneau de diamètre externe de 9-10 nm, correspondant aux RC I à IV, et de trois appendices globulaires de diamètre allant de 4.5 à 5.0 nm, correspondant aux RC V et VI. Ces appendices globulaires semblent être accrochés à l'anneau, mais occupent des positions variables dans les différentes particules isolées de L (**Figure 16**) (Rahmeh, Schenk et al. 2010). Par homologie avec d'autres polymérases, l'anneau pourrait être impliqué dans le guidage de l'ARN matrice, toutefois ces dimensions ne sont pas compatibles avec le passage de l'ARN couvert par la nucléoprotéine, suggérant que, durant la synthèse, la protéine doit se décrocher de l'ARN. *In vitro*, l'addition de P induit un changement de conformation de la L de VSV et conduit à une dimérisation partielle qui pourrait être la forme fonctionnelle de l'enzyme (Rahmeh, Schenk et al. 2010; Rahmeh, Morin et al. 2012), comme semble être le

cas également chez les Paramyxoviridae (Smallwood, Cevik et al. 2002; Smallwood and Moyer 2004).

9.2. Mécanismes de formation de la coiffe et de polyadénylation

Afin d'être pris en charge par la machinerie cellulaire de traduction, les ARNm viraux doivent posséder une coiffe de type m⁷GpppNm. Chez VSV, la coiffe est ajoutée aux ARNm naissant par un mécanisme non-conventionnel, différent de celui utilisé par la machinerie de la cellule hôte. La synthèse de la coiffe par la machinerie de VSV s'effectue par le transfert d'une molécule de guanosine diphosphate (GDP) plutôt que par une molécule de guanosine monophosphate (GMP) (Ogino and Banerjee 2007). La protéine virale L reconnaît l'extrémité 5' d'un ARN dont la séquence correspond à la séquence « gene start » (GS) conservée en début de chaque gène, pppApApcpApG, et lorsque la chaîne d'ARN atteint une longueur d'au moins 31 nucléotides, elle catalyse la réaction de transfert de GDP à partir de GTP vers l'extrémité 5' phosphate de l'ARN (GDP:polyribonucleotidyltransferase) en passant par la formation d'un intermédiaire covalent avec la chaîne latérale de l'His1227 localisée dans le domaine V (Ogino, Yadav et al. 2010). La coiffe est ensuite méthylée en position 2'-O et en position N-7 en deux étapes successives par le même site actif du domaine VI de la polymérase (Rhodes, Moyer et al. 1974; Rahmeh, Li et al. 2009).

A la fin de chaque gène, la polymérase rencontre une séquence de terminaison riche en uridines (AUACUUUUUUU chez VSV) sur laquelle elle « dérape » de manière itérative, produisant la polyadénylation de l'ARN viral (Schubert, Keene et al. 1980; Iverson and Rose 1981; Lamb and Kolakofsky 2001).

10. La phosphoprotéine est une protéine modulaire oligomérique

P est une protéine multifonctionnelle essentielle de la machinerie virale. L'existence de régions fonctionnelles indépendantes, d'abord proposée d'après le phénotype de mutants de délétion (Curran, Pelet et al. 1994; Das, Pattnaik et al. 1997; Bowman, Smallwood et al. 1999), a été confirmée par la démonstration d'une organisation structurale modulaire associant des régions intrinsèquement désordonnées et deux domaines structurés (Karlin, Longhi et al. 2002; Karlin, Ferron et al. 2003; Gerard, Ribeiro Ede et al. 2009; Leyrat, Gerard et al. 2010; Leyrat, Ribeiro et al. 2011) (**Figure 17**). P agit comme un cofacteur non-catalytique de la polymérase (Emerson and Yu 1975) et comme un chaperon de N⁰ monomérique (Arnheiter, Davis et al. 1985; Peluso and Moyer 1988; Leyrat, Jensen et al.

2011) et de L (Canter and Perrault 1996; Cevik, Holmes et al. 2004). De plus, comme les autres protéines dérivées du gène P (V, W, C), elle participe à la régulation de la réplication virale et à l'inhibition de la réponse immunitaire innée chez certains *MNV* (voir ci-dessous). La séquence en acides aminés des P de *Rhabdoviridae* et de *Paramyxoviridae* n'est pas conservées, mais ces protéines sont oligomériques et partagent une organisation modulaire similaire avec une distribution des fonctions répartie de manière identique entre les régions désordonnées et structurées. En conséquence, la P peut être divisée en deux régions distinctes (**Figure 17**) : une région N-terminale (P_{NTR}), commune chez les *Paramyxovirinae* à P, V et W, et une région C-terminale (P_{CED}), propre à la protéine P. P_{NTR} est globalement désordonnée, alors que P_{CED} contient un domaine d'oligomérisation et un domaine C-terminal structurés et reliés entre eux par un lien flexible plus ou moins long selon les virus.

10.1. Présence de régions intrinsèquement désordonnées

Des prédictions de désordre ont révélé l'organisation modulaire et la présence de longues régions intrinsèquement désordonnées chez la P des *Paramyxovirinae* (Karlin, Longhi et al. 2002; Karlin, Ferron et al. 2003), *Henipavirus* (Habchi, Mamelli et al. 2010), *Pneumovirinae* (Llorente, Garcia-Barreno et al. 2006) et *Rhabdoviridae* (Ribeiro, Favier et al. 2008; Gerard, Ribeiro Ede et al. 2009). Cependant, individuellement, ceux-là divergent quant à la localisation précise des régions désordonnées dans la séquence. Afin de faciliter l'interprétation de ces données, il a été proposé un calcul de score simple, appelé D-score, permettant d'intégrer les résultats de multiples prédictions de désordre en une méta-prédiction consensus. Ce D-score a permis la localisation correcte des régions désordonnées en accord avec les limites des domaines structurés de P antérieurement définies par la résolution de leur structure chez RaV et SeV (Ribeiro, Favier et al. 2008; Gerard, Ribeiro Ede et al.) (**Figure 17**). De même, la prédiction correcte des domaines P_{CD} de RaV et P_{CTD} de VSV a permis de définir les bornes de constructions compatibles avec une purification de qualité et avec la détermination de leur structure atomique (Ribeiro, Favier et al. 2008; Ivanov, Crepin et al. 2010) (**Figure 17**). De plus, des analyses biophysiques, dont notamment des expériences hydrodynamiques de SAXS et de RMN, ont confirmé l'état majoritairement désordonné des régions prédites comme telles (Gerard, Ribeiro Ede et al. 2009; Leyrat, Gerard et al. 2010).

10.2. Présence d'un MoRE dans P_{NTR}

La partie N-terminale désordonnée, de longueur variable selon le virus, comprend des régions de reconnaissance pour différentes protéines virales et cellulaires. Elle comprend

notamment un MoRE conservé responsable de l'activité chaperon sur N⁰ qui contient des hélices α transitoires (**Figure 17**) (Leyrat, Jensen et al. 2011). Chez certains virus, cette région désordonnée comprend également des sites de phosphorylation qui participeraient à la régulation du cycle répliatif viral (Pattnaik, Hwang et al. 1997). Enfin, chez les *Rhabdoviridae*, cette région comprendrait également un site de fixation à la protéine L (Emerson and Schubert 1987; Chenik, Schnell et al. 1998; Castel, Chteoui et al. 2009).

La caractérisation par RMN et SAXS du peptide correspondant aux 60 premiers résidus de P liant N⁰ de VSV confirme son caractère globalement désordonné. Cependant, elle montre la présence de deux hélices α dont la structuration alterne continuellement à partir de, et vers, des conformations désordonnées de la chaîne peptidique (Leyrat, Jensen et al. 2011). D'où l'adoption d'un modèle de cette région sous la forme d'un ensemble de conformères qui s'échangent sous l'influence de l'agitation thermique plutôt que sous la forme d'une structure unique. Cette représentation structurale d'ensemble rend bien compte des propriétés hydrodynamiques du peptide et met en évidence la diversité des conformations de la protéine et le volume d'espace qui peut être échantillonné par cette dernière. Cette modélisation révèle l'existence d'une distribution bimodale des valeurs du rayon de giration (Rg) suggérant la coexistence de deux sous-populations de conformères, les uns compacts, les autres étendus, comme observé avec d'autres PIDs (Paz, Zeev-Ben-Mordehai et al. 2008; Boze, Marlin et al. 2010).

Dans le cas des *Paramyxoviridae*, la partie N-terminale de P est en grande partie désordonnée (Karlin, Longhi et al. 2002; Habchi, Mamelli et al. 2010), mais deux régions prédites ordonnées au sein des premiers 40 résidus des protéines P (appelée Soyuz1 et Soyuz2) sont conservées parmi tous les membres de la sous-famille des *Paramyxovirinae* (Karlin and Belshaw 2012) (**Figure 17**). Chez SeV, comme chez les *Rhabdoviridae*, la partie N-terminale de la P est responsable de l'activité chaperon de la N⁰ (Watanabe, Kawano et al. 1996). Par extension, on peut spéculer que ces régions N-terminales conservées serait impliquées dans la fixation à N⁰ chez tous les *Mononegavirales* (Karlin and Belshaw 2012).

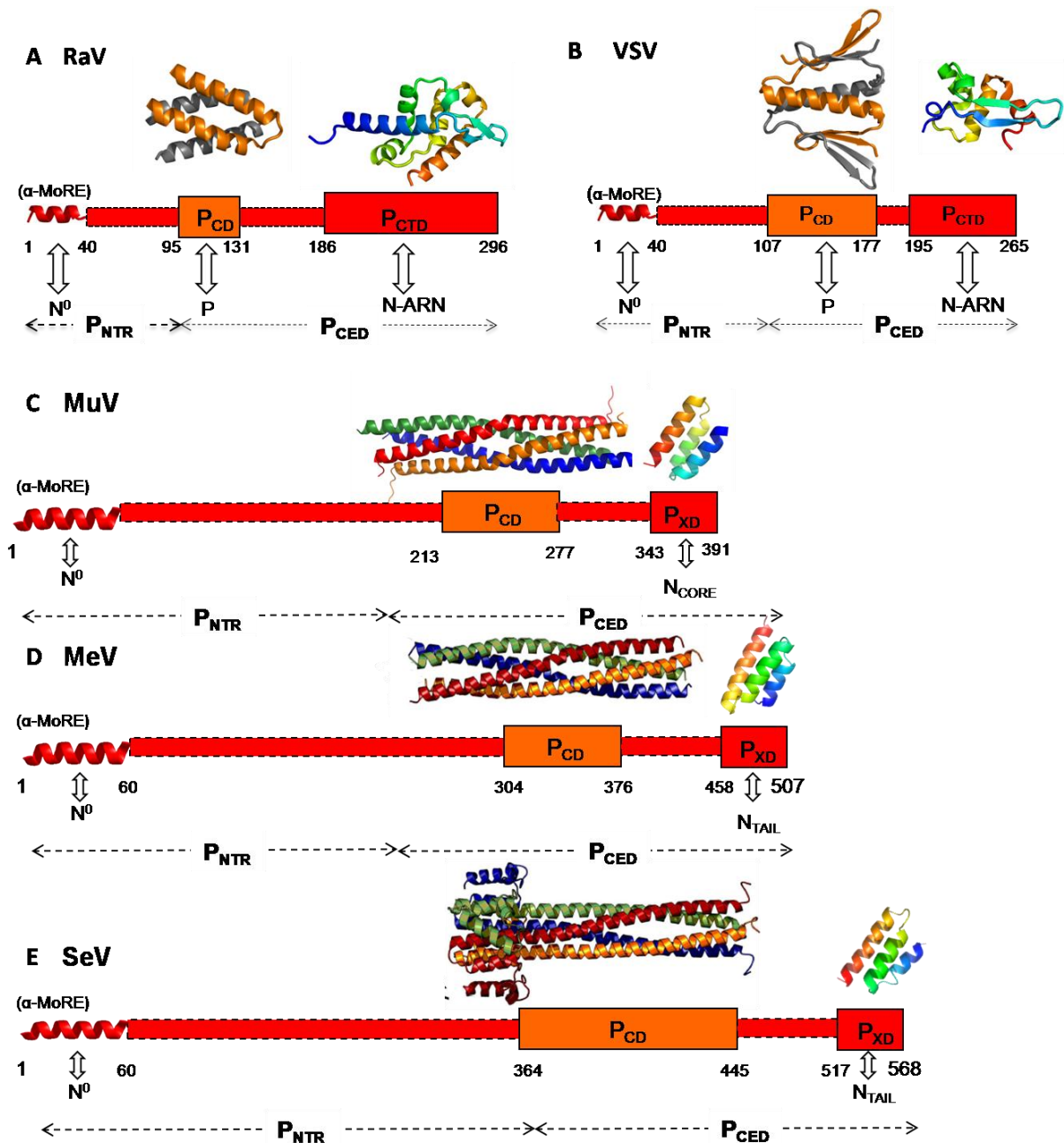


Figure 17: Organisation modulaire de la phosphoprotéine (P) (figure d'après Blocquel, D. et al Virologie E avec modifications). Les différents domaines de la P sont annotés comme suit : P_{CTD} est le domaine C-terminal d'interaction avec les nucléocapsides. P_{CD} est le domaine central d'oligomérisation. P_{NTR} est la région N-terminale dans laquelle se trouve la partie responsable de la formation du complexe N^0 -P. **A)** RaV PDB : 3L32 (P_{CD}) et 1VYI (P_{CTD}). **B)** VSV PDB : 2FQM (P_{CD}) et 2K47 (P_{CTD}). **C)** MuV PDB : 2FYZ (P_{CD}) et 3BBZ (P_{XD}). **D)** MeV PDB : 3ZDO (P_{CD}) et 1OKS (P_{XD}). **E)** SeV PDB : 1EZJ (P_{CD}) et 1R4G (P_{XD}).

10.3. Le domaine d'oligomérisation

P possède un domaine d'oligomérisation appelé P_{CD} (ou encore P_{MD}), dimérique chez les *Rhabdoviridae* (Gérard, Ribeiro et al. 2007; Ivanov, Crepin et al. 2010) (**Figure 17A et B**) et tétramérique chez les *Paramyxovirinae* (**Figure 17C, D et E**) et les *Pneumovirinae* (Tarbouriech, Curran et al. 2000; Tarbouriech, Curran et al. 2000; Castagne, Barbier et al. 2004; Llorente, Garcia-Barreno et al. 2006; Communie, Crepin et al. 2013). Le P_{CD} dimérique

des *Rhabdoviridae* est assez court comme l'indique les structures pour RaV (Ivanov, Crepin et al. 2010) et VSV (Ding, Green et al. 2006) (**Figure 17A et B**). Le P_{CD} tétramérique de SeV, MeV et MuV comprend un faisceau de quatre hélices α parallèles surenroulées (coiled-coil) qui semble être conservée au sein des *Paramyxovirinae*, mais qui est également prédit pour les *Pneumovirinae*. Par contre, dans la structure de SeV, ce faisceau de 4 hélices est complété par une collerette (couronne) d'autres hélices α qui semblent stabiliser la partie N-terminale du faisceau (**Figure 17E**) (Tarbouriech, Curran et al. 2000; Tarbouriech, Curran et al. 2000), mais est absente chez MeV (Communie, Crepin et al. 2013) et MuV (Cox, Green et al. 2013) (**Figure 17C et D**). L'organisation parallèle ou antiparallèle des chaînes α au sein de l'oligomère, le degré d'oligomérisation et les différences structurales ne permettent pas de postuler une origine commune pour les P_{CD} des *Rhabdoviridae* et des *Paramyxovirinae*. De plus, chez les *Paramyxovirinae*, ce domaine est impliqué dans la fixation de la polymérase L (Curran, Pelet et al. 1994; Bowman, Smallwood et al. 1999; Cevik, Holmes et al. 2004).

10.4. Le domaine C-terminal de fixation à N-ARN

Chez tous les *MNV*, la région de fixation au complexe polymérique N-ARN est localisée dans la partie C-terminale de P. Celle-ci est constituée soit d'un domaine structuré, connecté au domaine central d'oligomérisation par un lien flexible plus ou moins long selon les virus (*Paramyxovirinae*, *Rhabdoviridae*), soit d'une région désordonnée (*Pneumovirinae*). Chez les *Paramyxovirinae*, la structure conservée du domaine C-terminal (domaine P_X ou P_{XD}) est un faisceau de trois hélices α (**Figure 17**). Chez les *Rhabdoviridae*, le domaine C-terminal (appelé P_{CTD}), plus long, comprend 4 (VSV, **Figure 17A**), 5 (LNYV, (Martinez, Ribeiro et al. 2013)) ou 6 hélices α (RaV, **Figure 17B**) et un court feuillet β à deux brins. Au sein d'une même famille, la superposition des structures fait apparaître une parenté évidente en dépit de l'absence de conservation de la séquence et de la présence d'insertions et de délétions (comparer les P_{XD} **Figure 17** et (Ribeiro, Favier et al. 2008; Martinez, Ribeiro et al. 2013)).

En accord avec les analyses bioinformatiques (Karlin, Ferron et al. 2003; Habchi, Mamelli et al. 2010), les domaines P_X des protéines P de MeV et de SeV (résidus 459-507 et 516-568, respectivement) sont organisés en trois hélices α (**Figure 17C, D et E**) (Johansson, Bourhis et al. 2003; Blanchard, Tarbouriech et al. 2004; Gely, Lowry et al. 2010). De même, d'après des études par CD et RMN, les domaines P_{XD} des *Henipavirus* (résidus 660-709 pour NiV et 657-707 pour HeV), très similaires avec ~90% de similarité et ~80% d'identité, adoptent également une structuration α -hélicale prédominante (Habchi, Mamelli et al. 2012).

Chez ces virus, le domaine P_{XD} est le domaine de liaison à la nucléocapside via l'interaction avec N_{TAIL} (Johansson, Bourhis et al. 2003; Kingston, Hamel et al. 2004; Houben, Marion et al. 2007; Habchi, Mamelli et al. 2012).

Bien que le domaine P_{XD} (résidus 343-391 de P) du MuV adopte en solution une structure dite en "molten globule" et soit donc dépourvu de structure tertiaire stable (Kingston, Hamel et al. 2004), il a cristallisé et sa structure a été déterminée par cristallographie aux rayons X (Kingston, Gay et al. 2008). Il s'agit d'un bel exemple où l'empilement cristallin a permis de "capturer" une conformation stable. Corrélativement, ce domaine possède des propriétés distinctes par rapport aux domaines P_{XD} des *Morbillivirus*, *Respirovirus* et *Henipavirus* car il n'interagit pas avec N_{TAIL}, mais avec la région structurée, N_{CORE} (Kingston, Walter et al. 2004).

La comparaison de la structure des domaines C-terminaux des protéines P de RaV, VSV, de MokV (Mavrakis, McCarthy et al. 2004; Ribeiro, Favier et al. 2008; Assenberg, Delmas et al. 2010) et de LNYV (Martinez, Ribeiro et al. 2013) avec celle des domaines P_X de MeV (Johansson, Bourhis et al. 2003), SeV (Houben, Marion et al. 2007) et MuV (Kingston, Gay et al. 2008), indique des différences significatives qui ne permettent pas de conclure à une homologie. Toutefois, l'observation de ces structures suggère une topologie conservée entre *Paramyxovirinae* et *Rhabdoviridae* (Delmas, Assenberg et al. 2010) et un arbre phylogénétique construit sur la base de l'alignement structurale reproduit le même schéma que les alignements de séquence de protéines mieux conservées telles que la L ou la N (Martinez, Ribeiro et al. 2013).

10.5. La connexion flexible entre P_{CD} et P_{CTD}/P_{XD}

Chez tous les *Paramyxovirinae*, P_{CD} et P_{XD} sont séparés par une région flexible prédite comme faiblement ordonnée (Karlin, Ferron et al. 2003). D'après des études par RMN de la région 474-568 de P de SeV, la région en amont de P_{XD} (474-515) est désordonnée (Bernado, Blanchard et al. 2005; Houben, Blanchard et al. 2007). Chez les *Morbillivirus* et les *Henipavirus*, une région flexible supplémentaire est prédite en amont de P_{CD} (**Figure 17**) (Karlin, Ferron et al. 2003; Habchi, Mamelli et al. 2010).

10.6. Phosphorylation de la P

La P des MNVs est phosphorylée dans différentes régions de la séquence et par des différentes protéines kinases. Il a été proposé que chez VSV, la phosphorylation des aa 60, 62

et 64 (VSV-IN) par la créatine kinase II stimulerait la transcription du génome viral (Barik and Banerjee 1992; Takacs, Barik et al. 1992) tandis que la phosphorylation des résidus 226 et 227 dans le domaine C-terminal de la P par une autre kinase cellulaire stimulerait la réplication (Das and Pattnaik 2004). Une étude plus récente propose que la phosphorylation de la région N-terminale soit indispensable pour empêcher la N de fixer de l'ARN d'une façon non-spécifique et permettre l'encapsidation spécifique du génome par la N (Chen, Zhang et al. 2013). Le mécanisme moléculaire de ces régulations reste néanmoins mal connu. Il a été proposé que la phosphorylation dans la région N terminale induise l'oligomérisation de la protéine et que ce changement régule la balance entre les activités de transcription et de réplication (Gao and Lenard 1995). Des travaux récents dans notre laboratoire, indiquent que la protéine non-phosphorylée forme exclusivement des dimères et que la mutation en aspartate des résidus sérines de la région N-terminale, qui mime *in vivo* l'effet de la phosphorylation, n'a aucun effet sur l'état d'oligomérisation (Gérard, Ribeiro et al. 2007). Par contre, ce travail a démontré qu'*in vitro*, en solution, l'incubation de la P à 37 °C provoquait son agrégation, suggérant que les effets observés d'oligomérisation induite par la phosphorylation pourraient être liés aux conditions dans lesquelles la protéine était incubée pour sa phosphorylation par la créatine kinase II (CKII) (Gao and Lenard 1995). D'autre part dans une étude récente le groupe de Whelan n'a observé aucun effet de la phosphorylation de la P sur la synthèse de courts fragments d'ARN par la L (Morin, Rahmeh et al. 2012).

Chez les henipavirus, la protéine P est également phosphorylée, mais les sites de phosphorylation et les mécanismes de régulations impliqués sont également mal connus. En raison de la longueur de leur séquence, les P de NiV et HiV contiennent de nombreux sites potentiels de phosphorylation (NiV : 63 sérines et 39 thréonines ; HeV : 58 sérines et 41 thréonines), avec approximativement 50% des sérines présentant un score de prédiction élevé pour un site de phosphorylation consensus (Shiell, Gardner et al. 2003). La phosphorylation des résidus Ser224 et Thr239 de HeV et Ser240 et Ser472 de NiV par la PDPK et la CKII a été confirmée expérimentalement, mais combien de ces sites ont un rôle fonctionnel et/ou structural reste une question ouverte (Shiell, Gardner et al. 2003).

11. Structure du complexe N-ARN-P

P est un cofacteur essentiel de la polymérase, car elle assure l'attachement de la protéine L sur la matrice N-ARN. En effet, la polymérase seule se fixe avec une faible affinité

sur le complexe N-ARN, et son activité en absence de P est peu processive (Emerson and Yu 1975; Mellon and Emerson 1978; Morin, Rahmeh et al. 2012). En possédant un site de fixation pour le complexe N-ARN dans sa région C-terminale et un site de fixation pour la polymérase dans sa partie N-terminale ou dans le domaine central d'oligomérisation, P assure le recrutement de L sur le complexe homopolymérique N-ARN. De plus, P interviendrait dans le déplacement de la polymérase le long de sa matrice au cours des phases de transcription et de réplication.

11.1. Complexe N-ARN-P des *Rhabdoviridae*

Chez les *Rhabdoviridae* comme chez les *Paramyxovirinae*, l'attachement de P sur la matrice N-ARN implique l'interaction entre le domaine C-terminal de P et l'extrémité C-terminale de N, mais le mode d'interaction est différent. Chez les *Rhabdoviridae*, la région C-terminale de P forme un domaine globulaire P_{CTD} (**Figure 17A et B**). La région C-terminale de N, N_{CTD}, est constituée d'une boucle suivie de trois hélices qui forment la surface d'interaction au sommet de la molécule de N. La structure d'un complexe entre un anneau N-ARN de VSV contenant 10 protomères de N et 10 domaines P_{CTD} a été résolue par cristallographie (Green and Luo 2009) (**Figure 18C**).

Chez RaV, d'après des données de titration, seulement deux P_{CTD} peuvent se fixer simultanément sur un anneau de 10 ou 11 protomères de N. Cette fixation est non-coopérative avec une constante de dissociation de 160 ± 20 nM (Ribeiro Ede, Leyrat et al. 2009). Un modèle structural de ce complexe a été construit par une méthode d'arrimage moléculaire flexible sur la base des données de SAXS (Ribeiro Ede, Leyrat et al. 2009) et en utilisant des contraintes de contact entre P_{CTD} et certains résidus de N identifiés par mutagenèse chez MokV (Jacob, Real et al. 2001) (**Figure 18A et B**).

Dans le modèle de RaV et dans la structure cristallographique de VSV, le domaine P_{CTD} est fixé sur le sommet du N_{CTD} d'un protomère N et à l'extrémité C-terminale du bras N_{CTA}. L'hélice C-terminale de P_{CTD} se positionne sur l'hélice α 14 de N_{CTD} et l'extrémité N-terminale de P_{CTD} pointe vers l'intérieur de l'anneau. Dans les deux complexes, les boucles C-terminales du bras N_{CTA} de deux protomères adjacents (N_i et N_{i+1}) se fixent sur les faces latérales opposées du P_{CTD} à la manière des mâchoires d'une pince (**Figure 18A, B et C**). La cohérence architecturale de l'hélice α 14 de N_{CTD} avec son environnement est critique avec des mutations ponctuelles ayant un effet délétère soit sur l'encapsidation de l'ARN soit sur la transcription, et ce indépendamment de la liaison à P_{CTD} (Heinrich, Morin et al. 2012).

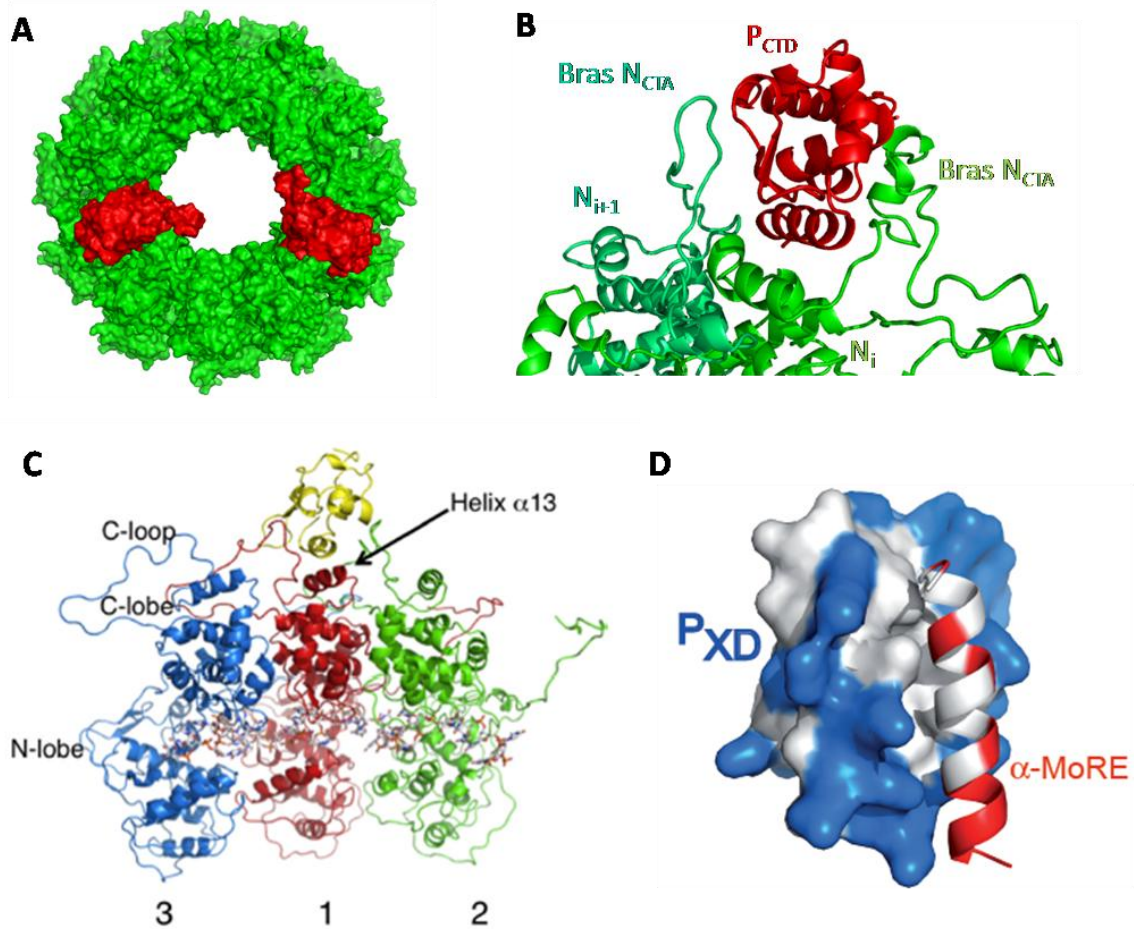


Figure 18 : Complexes entre le domaine C-terminal de la phosphoprotéine et N-ARN (figures d'après Blocquel, D. et al. Virologie E et (Green and Luo 2009) avec modifications). **A)** Vue du haut et **B)** vue de face du complexe entre P_{CTD} et N-ARN de RaV, obtenu par arrimage moléculaire (Ribeiro, Leyrat et al. 2009). Les N sont colorées en vert, la P_{CTD} en rouge. **C)** Le même complexe pour VSV obtenu par cristallographie (PDB code : 3HHW). Les N sont colorées en bleu (N_{M+1}), rouge (N_M) et vert (N_{M-1}) ; le P_{CTD} est coloré en jaune. **D)** Complexe entre le domaine C-terminal (P_{XD}) de la P et le domaine C-terminal (N_{TAIL}) de la N de NiV. Ce complexe a été modélisé (Habchi, Blangy et al. 2011) par homologie avec le complexe correspondant du virus de la rougeole (PDB code : 1T6O).

11.2. Complexes N_{TAIL} - P_{XD} des *Paramyxovirinae*

La fixation de la P sur le complexe N-ARN des *Paramyxovirinae* repose sur une stratégie moléculaire très différente. En effet, la partie de N fixant la P est désordonnée en l'absence de cette dernière et se replie en hélice α lors de la fixation au domaine P_{XD} de la protéine P (Johansson, Bourhis et al. 2003; Bourhis, Johansson et al. 2004; Bourhis, Receveur-Bréchet et al. 2005; Habchi, Blangy et al. 2011). L' α -MoRE en hélice α vient s'insérer dans un large sillon délimité par les hélices $\alpha 2$ et $\alpha 3$ de P_{XD} avec une constante de dissociation (K_D) faible d'environ 100 nM (Bourhis, Receveur-Bréchet et al. 2005; Shu, Habchi et al. 2012) (**Figure 18D**). Néanmoins une affinité plus faible d'environ une dizaine

de μM a été également documentée pour le couple $\text{P}_{\text{XD}}/\text{N}_{\text{TAIL}}$ de MeV (Kingston, Baase et al. 2004).

Les complexes $\text{N}_{\text{TAIL}}\text{-P}_{\text{XD}}$ du MeV et du SeV partagent une organisation structurale remarquablement similaire, mais l'interface est dominée par des résidus hydrophobes chez le premier et par des résidus chargés chez SeV (Houben, Marion et al. 2007). Des études par SAXS ont fourni un modèle à basse résolution du complexe $\text{N}_{\text{TAIL}}\text{-P}_{\text{XD}}$ du MeV où une grande partie de N_{TAIL} reste désordonnée au sein du complexe (Bourhis, Receveur-Bréchet et al. 2005).

11.3. Interaction N-ARN-P chez les *Pneumovirinae*

Rappelons que la situation est encore différente chez RSV. Le site d'interaction de P sur l'homopolymère de N complexé à l'ARN est situé dans le domaine globulaire N-terminal de N (N_{NTD}) (Galloux, Tarus et al. 2012) (**Figure 12F**). Il est constitué d'un sillon hydrophobe entouré de résidus basiques. La partie C-terminale de P ($\text{P}_{10\text{CT}}$), hydrophobe et acide, vient se loger dans ce sillon situé à la périphérie de la nucléocapside hélicoïdale (Tran, Castagne et al. 2007). La structure du complexe n'est pas établie, ni sa dynamique.

12. Déplacements de la polymérase sur la matrice N-ARN

Deux modèles ont été proposés pour expliquer comment la polymérase virale se déplace le long de sa matrice lorsqu'elle synthétise de l'ARN.

12.1. Modèle impliquant le déplacement de P sur la matrice N-ARN

Un premier modèle propose que P (associée à L) avance le long du complexe N-ARN «en faisant la roue» (“cartwheeling”) par un mécanisme dans lequel les domaines P_{XD} s'attachent et se détachent alternativement de la matrice nucléocapsidique hélicoïdale à la manière d'un tétrapode «faisant la roue» en descendant chaque marche (c'est-à-dire chaque N_{TAIL}) d'un escalier en colimaçon (Curran 1998; Curran and Kolakofsky 1999; Kolakofsky, Le Mercier et al. 2004) (**Figure 4**). Chez SeV, ce modèle est inféré (i) de la faible affinité du P_{XD} pour le complexe N-ARN ($60 \pm 20 \mu\text{M}$), (ii) du caractère indispensable du domaine central de tétramérisation pour permettre la fixation de P_{XD} dans des expériences de « pull-down » et (iii) pour une synthèse efficace de l'ARN à partir de minigénomes (Curran 1998).

A travers la fixation simultanée de plusieurs P_{XD} , c'est-à-dire en augmentant l'avidité de P pour le complexe N-ARN, le domaine tétramérique permettrait la synthèse d'ARN.

Ce modèle prédit qu'une stabilité accrue du complexe $N_{TAIL}-P_{XD}$ ralentit ou même bloque le cheminement du complexe polymérase le long de la matrice nucléocapsidique (Curran and Kolakofsky 1999; Kolakofsky, Le Mercier et al. 2004). Cependant, une forte affinité entre les protéines N et P n'est probablement pas incompatible avec une bonne processivité de la polymérase d'après les faits expérimentaux suivants. Malgré une différence apparente d'un facteur 600 entre le K_D du couple $N_{TAIL}-P_{XD}$ chez SeV (60 μ M (Houben, Marion et al. 2007)) et chez MeV (100 nM) (Bourhis, Receveur-Brechot et al. 2005; Shu, Habchi et al. 2012), la vitesse d'élongation apparente de leur polymérase est du même ordre de grandeur avec 1,7 nt/s chez SeV (Gubbay, Curran et al. 2001) et 3,3 nt/s chez MeV (Plumet, Duprex et al. 2005). Ces observations sont également en accord l'affinité d'environ une dizaine de μ M documentée pour le couple P_{XD}/N_{TAIL} de MeV (Kingston, Baase et al. 2004). De plus, cette vitesse reste relativement constante quelle que soit l'affinité entre N_{TAIL} et P_{XD} : en effet des mutations chez ce dernier virus entraînant une baisse de l'affinité envers P_{XD} pouvant aller jusqu'à un facteur 30 ne semblent pas modifier la vitesse d'élongation (Shu, Habchi et al. 2012). Cependant à ce jour aucune étude n'a pas permis de mesurer directement l'affinité et les constantes d'association/dissociation entre des NC et une phosphoprotéine entière chez les *Paramyxoviridae*.

Comment réconcilier le modèle d'une progression de P-L «faisant la roue» avec, éventuellement, une forte affinité du couple $N_{TAIL}-P_{XD}$ du MeV, si tant est qu'il soit correct ? Peut-être faut-il faire intervenir un troisième acteur tel qu'un facteur cellulaire capable de moduler la force de l'interaction $N_{TAIL}-P_{XD}$. Hsp70 est l'un des candidats possibles (Zhang, Glendening et al. 2002; Zhang, Bourhis et al. 2005; Oglesbee 2007). Des études fonctionnelles visant à élucider le rôle de la multimérisation de la protéine P du MeV et/ou l'importance de l'expression d'Hsp70 par l'utilisation de cellules infectées appauvries en Hsp70 (et/ou en hsc70, une protéine apparentée) apporteront sans doute des éléments de compréhension. Esistent-ils d'autres mécanismes de modulation de l'affinité du couple $N_{TAIL}-P_{XD}$?

12.2. Modèle d'une progression par saltation de L sur plots fixes de P

Chez les *Rhabdoviridae*, les données expérimentales ne sont pas en faveur du modèle de progression «en faisant la roue». (i) L'affinité du domaine P_{CTD} monomérique de RaV pour

le complexe N-ARN est forte (K_D de 160 ± 20 nM) (Ribeiro Ede, Leyrat et al. 2009). (ii) La lente cinétique de dissociation mesurée par résonance plasmonique de surface semble incompatible avec un mécanisme dans lequel la P progresserait en dissociant continuellement sa paire de P_{CTD} de la N-ARN. (iii) L'ablation du domaine de dimérisation P_{CD} rend la P monomérique (Gerard, Ribeiro Ede et al. 2009) et n'empêche pas une synthèse efficace d'ARN dans un essai fonctionnel de transcription d'un minigénome *in cellula* (Jacob, Real et al. 2001).

Un mécanisme alternatif a donc été proposé selon lequel la polymérase L progresse par saltation entre les P fixées à intervalles réguliers tout le long de la matrice N-ARN (Ribeiro Ede, Leyrat et al. 2009; Ivanov, Yabukarski et al. 2011) (**Figure 4B**). Comme les particules de VSV contiennent ~ 1 dimère de P pour 5 N (Thomas, Newcomb et al. 1985), on imagine une L séjournant temporairement sur un même dimère de P le temps de copier l'ARN encapsidé par les 5 monomères de N contigus entre chaque saut de P_i à P_{i+1} . Selon ce modèle, et pour reprendre notre langage imagé, les marches de l'escalier N-ARN en colimaçon ne seraient plus les N_{TAIL} , mais des dimères de P fixés tous les 5 protomères de N sur lesquels L, au lieu de P-L, se déplacerait.

13. Mécanismes d'encapsidation et le complexe N^0 -P

13.1. Le processus d'encapsidation

L'encapsidation des antigénomes et des génomes néo-synthétisés est une étape essentielle dans le processus de la multiplication du virus dans la cellule infectée car la polymérase utilise exclusivement de l'ARN encapsidé comme matrice pour une transcription et une réplication efficace (**Figure 4**).

Le processus d'encapsidation se déroule de manière concomitante à la synthèse d'ARN. Il requiert donc la production continue de N monomérique et soluble dépourvue d'ARN cellulaire (Masters and Banerjee 1988; Peluso and Moyer 1988). Lors de la synthèse de N en présence de P se forme un complexe soluble N^0 -P (**Figure 4**). Il sert de substrat pour la polymérase en assurant le transfert de N^0 sur le brin naissant d'ARN pour former la nucléocapside par homopolymérisation dirigée, unidirectionnelle et régulière depuis l'extrémité 5' de l'ARN jusqu'à son extrémité 3' (Blumberg, Leppert et al. 1981; Dillon and Gupta 1988). Cette exigence expliquerait qu'un peptide comprenant l' α -MoRE N-terminal de

P et le site de fixation de N^0 puisse inhiber la réplication d'un mini-génome ou du virus (Castel, Chteoui et al. 2009).

Comment s'effectue le transfert de N^0 -P vers la nucléocapside naissante et comment est-il régulé ? Lors des processus de transcription et de réplication, la synthèse d'ARN débute à l'extrémité 3' de la matrice d'ARN. Donc, la décision de synthétiser des ARNm en cascade ou de produire une copie complète de l'ARN génomique doit être prise avant, ou au moment, où la polymérase atteint la fin de la séquence « leader ». D'après les rares observations expérimentales disponibles, et intuitivement, la concentration du complexe N^0 -P devrait jouer un rôle dans le basculement entre les activités de transcription et de réplication (Curran and Kolakofsky 2008), ne serait-ce que pour éviter l'initiation d'une réplication finissant de manière abortive suite à une pénurie de substrat d'encapsidation. La réplication est toujours différée d'un laps de temps (en heures) suffisant pour une production détectable de N et P, ce qui est compatible avec un effet seuil d'une concentration minimale de substrat N^0 -P. Corrélativement, dans une cellule infectée, l'arrêt de la synthèse protéique abroge la réplication sans affecter la transcription primaire. Lorsque cet arrêt intervient en cours de réplication, celle-ci s'interrompt rapidement et les polymérases en mode réplisase semblent basculer en mode transcriptase, comme si le complexe N^0 -P, ou un cofacteur cellulaire clef, avait une demi-vie courte (Plumet, Duprex et al. 2005). Chez VSV, la phosphorylation de P serait impliquée dans cette régulation (Pattnaik, Hwang et al. 1997). Le transfert de N^0 vers l'ARN pourrait être contrôlé par une affinité plus élevée de N^0 pour des signaux d'encapsidation présents dans les régions leader (et trailer) en 3' de l'ARN génomique (et antigénomique) comparativement à celle vis-à-vis de P (Blumberg, Leppert et al. 1981). Alternativement, en s'associant à P-L, le complexe N^0 -P pourrait commuter le complexe polymérasique de transcriptase à réplisase (Qanungo, Shaji et al. 2004).

14. Comment caractériser la structure des protéines modulaires?

La protéine P des *MNV* est une protéine modulaire, c'est-à-dire qu'elle est constituée par la concaténation de plusieurs régions structurellement indépendantes les unes des autres. Certaines sont des régions autonomes de repliement qui adoptent une structure tridimensionnelle bien définie – on parle alors de domaines structurés – alors que d'autres sont désordonnées et donc très flexibles. La structure de telles protéines ne peut être déterminée par cristallographie car la grande flexibilité ne permet généralement pas la cristallisation. La microscopie électronique s'avère également inadaptée car les domaines structurés sont généralement trop petits pour permettre une reconstruction tridimensionnelle, et les régions flexibles ne sont pas visibles. Par opposition, la résonance magnétique nucléaire est un outil très puissant pour étudier les protéines intrinsèquement désordonnées ou dénaturées en fournissant de l'information à niveau atomique. Des avancées importantes sont faites dans ce domaine (pour revue (Schneider, Huang et al. 2012) et références citées). Par exemple la comparaison des déplacements chimiques observés avec ceux prédits pour une protéine complètement désordonnée et une protéine complètement structurée est informative pour la présence de structures secondaires flottantes (Wishart and Sykes 1994; Marsh, Singh et al. 2006). Un autre exemple est la possibilité de détection de conformations très faiblement peuplées par des différentes mesures de relaxation (pour revue (Cloue 2011) et références citées). Les méthodes de couplages dipolaires résiduels et la résonance paramagnétique électronique sont aussi utilisées pour étudier des protéines désordonnées. Néanmoins l'étude des domaines structurés par cette technique est rapidement limitée par la taille de la protéine. Des méthodes de marquages spécifiques ont démontré leur utilité pour l'étude de très grands complexes protéiques (Gans, Hamelin et al. 2010) mais en pratique, la détermination de structure par RMN se limite à des protéines qui ne dépassent pas les 30 kDa. La diffusion (des neutrons ou rayons X) aux petits angles est une méthode bien adaptée pour étudier les paramètres globaux tels que le rayon de giration (R_g), la dimension maximale de la molécule étudiée et son état d'oligomérisation et il n'existe pas de limite quant à la taille des objets étudiés. Néanmoins la courbe expérimentale obtenue est une moyenne – c'est la somme des contributions de toutes les conformations existantes en solution – et la technique ne fournit que de l'information de basse résolution ne dépassant pas les 15-20 Å de résolution (Putnam, Hammel et al. 2007).

En conséquence la caractérisation structurale d'une protéine modulaire requiert la combinaison de plusieurs techniques de la biologie structurale : après une analyse de

bioinformatique appropriée, les domaines structurés peuvent être isolés et cristallisés et leur structure déterminée par la diffraction des rayons X, les régions flexibles peuvent être isolées et étudiées par la RMN. La combinaison de ces deux techniques permet la construction de modèles atomiques de conformations uniques. Une protéine modulaire ne peut pas être décrite par une seule conformation et au lieu de cela une représentation dite « d'ensemble » nous paraît plus appropriée. Cet ensemble contient quelques à quelques dizaines de modèles de la protéine qui sont physiquement possibles et très probablement existent en solution. La collecte de données de diffusion aux petits angles permet de sélectionner de tels ensembles de modèles dont la courbe de diffusion théorique correspond bien à la courbe expérimentale en utilisant par exemple l'Ensemble Optimization Method (Bernado, Mylonas et al. 2007).

15. Réponses immunitaires et mécanismes de défense contre des virus

15.1. Immunité innée et immunité acquise

Au cours de leur évolution, les organismes ont développés différents mécanismes pour se défendre contre les infections par divers pathogènes qui sont désignés sous le terme de réponse immunitaire. On distingue généralement deux systèmes : l'immunité innée et l'immunité acquise (Janeway 2001). Les mécanismes de l'immunité innée constituent une première ligne de défense. Ils sont non-spécifiques, se mettent en action immédiatement après qu'un agent pathogène a pénétré dans l'organisme et ne dépendent pas d'une exposition antérieure à cet agent. Il s'agit notamment des processus d'activation du complément, de la phagocytose ou du développement d'un état antiviral déclenché par la production d'interférons (voir ci-dessous). Les mécanismes de l'immunité acquise sont hautement spécifiques, nécessitent un temps de réaction, mais surtout la vitesse et l'ampleur de leur action dépend de l'exposition préalable à l'agent pathogène, d'une « immunisation ».

15.2. La réponse interféron

Chez les mammifères, la production d'interféron (IFN) fait partie de la première ligne de défense du système immunitaire inné contre les pathogènes, et en particulier de défense contre les virus. On distingue deux types d'IFN : les IFN de types I et II. Les IFN de type I, IFN α ou IFN β , sont typiquement produits lors d'une infection virale ou bactérienne. Ils induisent l'expression de divers gènes qui vont permettre aux cellules de devenir résistantes aux virus et jouent un rôle important en stimulant la mise en action des mécanismes de l'immunité acquise. L'IFN de type II, IFN γ , est produit seulement par quelques types de cellules du système immunitaire, les cellules T activées et les cellules tueurs naturels (lymphocytes NK), et est impliqué dans la régulation des réponses immunitaires et inflammatoires.

La première étape dans l'activation de la réponse immunitaire dépend de la reconnaissance de signaux de « non-soi » porté par un agent extérieur. En terme moléculaire, ces signaux sont appelés « Pathogen Associated Molecular Pattern » ou PAMPs. Ce sont des motifs moléculaires généralement présents en multiples copies à la surface d'un pathogène, qui sont spécifiques de l'organisme et peuvent être reconnus par le système immunitaire inné.

Parmi les PAMPs les plus étudiés, nous trouvons les lipopolysaccharide, le peptidoglycane, les acides téichoïques, les acides nucléiques modifiés mais aussi les protéines de surface virales.

Des récepteurs dont le rôle est de reconnaître et de fixer les PAMPs et de déclencher la réponse immunitaire innée sont présents à la surface des cellules mais aussi sous forme soluble à l'intérieur des cellules. Dans le cas d'infection par un virus, les récepteurs cytosoliques Retinoic acid-Inducible Gene 1 (RIG-I) et Melanoma Differentiation-Associated protein 5 (MDA5) ou les récepteurs à la surface cellulaire de type Toll-Like Recepteur (TLR) (Andrejeva, Childs et al. 2004; Yoneyama, Kikuchi et al. 2004) détectent des molécules d'ARN double brin ou d'ARN simple brin possédant un groupement triphosphate en position 5' qui apparaissent de manière intermédiaire au cours de la transcription ou de la réplication virale (Jacobs and Langland 1996). Les récepteurs TLR détectent très probablement l'ARN db libéré par les cellules infectées (Beutler 2004). La reconnaissance d'un PAMP par l'un de ces récepteur déclenche une cascade d'activation de facteurs de transcription déjà présents dans la cellule (e.g. IFN-regulatory factor 3 et NF- κ B) qui conduit à la production et l'excrétion d'IFN de type I (Goodbourn, Didcock et al. 2000; Conzelmann 2005).

Les interférons sont des cytokines qui vont stimuler l'expression de divers gènes (Interferon Stimulated Genes) dans d'autres cellules et de cette manière permettent de développer un état de résistance contre les virus dans lequel les virus ne sont plus capables de se répliquer ou se répliquent de manière moins efficace. Les IFN sont reconnus par des récepteurs à la surface de leurs cellules cibles (IFNAR1 et IFNAR2) conduisant à l'activation de voies de signalisation complexes dans lesquels les facteurs de transcription STAT (Signal Transducer and Activator of Transcription) jouent un rôle central. Parmi cette famille de protéines, les protéines STAT1 et STAT2 sont impliquées dans la réponse à l'interféron. Dans des cellules non-infectées STAT1 est partagé entre le cytoplasme et le noyau tandis que STAT2 est principalement cytoplasmique. Après leur activation STAT1 et STAT2 sont dirigés vers le noyau ou ils s'associent à d'autres facteurs et activent la transcription de gènes importants pour le contrôle de l'infection virale (de Weerd, Samarajiwa et al. 2007) et notamment des gènes codant des protéines possédant une activité antivirale (e.g. la protéine kinase R, la protéine M_X etc.) (Sadler and Williams 2008). Il faut également noter que la production des IFNs de type I agit aussi en stimulant la réponse immunitaire adaptative, notamment en augmentant la production des protéines du Complexe Majeur d'Histocompatibilité I (MHC I) (Stetson and Medzhitov 2006).

15.3. Les mécanismes de contre-défense

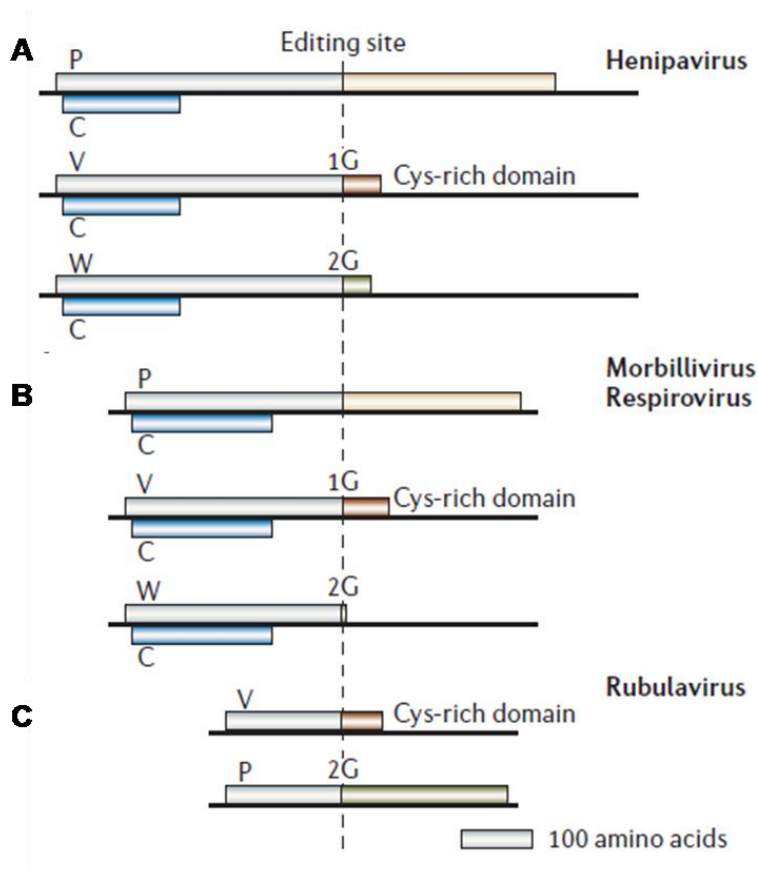
Au cours de la coévolution des virus et des hôtes qu'ils infectent, ces derniers ont donc développés des moyens de défense, et les virus ont, en contre partie, développé des mécanismes leur permettant d'y échapper afin de persister dans leurs hôtes et de continuer à s'y reproduire. De nombreux virus ont ainsi développé des mécanismes leur permettant d'échapper à la détection par le système IFN cellulaire ou de minimiser cette réponse immunitaire : l'inhibition de la transcription et réplication cellulaire, inhibition de la signalisation par l'ARN double brin et antagonisme des protéines effectrices impliquées dans la signalisation de la réponse immunitaire, pour en nommer quelques uns. Chez les *MNV*, différents virus, dont les virus de la rage (RaV) et le virus Nipah (NiV), agissent sur les voies de signalisation de l'interféron, par l'intermédiaire de protéines exprimées à partir du gène P.

15.4. Les divers produits du gène P

De manière générale, le gène P des *MNV* code pour plusieurs protéines, même si les mécanismes d'expression sont différents selon les familles. Chez les *Paramyxovirinae*, l'expression de plusieurs protéines à partir du gène P implique à la fois un mécanisme d'édition des ARNm et l'utilisation d'un second cadre de lecture. Ainsi, chez certains paramyxovirus, l'insertion dans l'ARNm par la polymérase virale d'un G non codé permet la production de la protéine V et dans certains cas, l'insertion de deux Gs permet la production de la protéine W (Lamb and Parks 2007). Dans d'autres cas, la protéine V est produite en absence d'édition de l'ARN et l'insertion d'un G supplémentaire est nécessaire pour la production de la P entière (Ohgimoto, Bando et al. 1990). Comme chez beaucoup d'autres *MNV*, les protéines V et W des *Henipavirus* sont plus courtes que la protéine P, partageant avec celle-ci une partie N-terminale commune mais possédant un domaine C-terminal spécifique (Kulkarni, Volchkova et al. 2009) (**Figure 19**). Le domaine C-terminal de la V est très conservé chez tous les *Paramyxoviridae* et contient sept cystéines, qui forment un domaine en doigt de zinc, dont la fonction reste inconnue. Le domaine C terminal de la W des henipavirus, composé de 43 aa, est différent de celui des autres *Paramyxoviridae* et contient un signal d'import au noyau (Shaw, Cardenas et al. 2005). Chez les paramyxovirus, le gène P code également pour une petite protéine basique, appelée protéine C, dans un cadre de lecture différent de celui de la P mais proche de son extrémité 3' (**Figure 19**).

Le gène P de VSV et RaV code également pour des protéines supplémentaires. Chez VSV deux petites protéines basiques, C et C', sont codées dans un cadre de lecture différent

de celui de la P (Spiropoulou and Nichol 1993). Ces protéines sont bien conservées chez les vesiculovirus mais pas chez les lyssavirus. La fonction de ces deux protéines est mal connue mais elles ne semblent pas être importantes pour la réplication virale (Kretzschmar, Peluso et al. 1996). De plus, une phosphoprotéine plus courte est produite à partir du gène P en utilisant le même cadre de lecture mais à partir d'un AUG situé en aval de celui utilisé pour produire la P entière mais sa fonction n'est pas connue (Herman 1986). Chez RaV plusieurs isoformes de la phosphoprotéine existent. Elles sont produites à partir du même ARNm, en utilisant le même cadre de lecture que la P mais par un mécanisme de traduction alternatif à partir de différents codons AUG initiateurs. Ainsi, en plus de la phosphoprotéine entière (299 aa), qui est la plus abondante, quatre formes tronquées du côté N terminal sont exprimées dans les cellules infectées: P2 (279 aa), P3 (246 aa), P4 (230 aa) et P5 (216 aa) (Chenik, Chebli et al. 1995). Tous ces produits du gène de la phosphoprotéine sont présents dans le cytoplasme des cellules infectées mais les produits tronqués P3, P4 et P5 sont majoritairement transportés vers le noyau grâce à leur signal d'import. Il est intéressant de noter que ces versions tronquées de la P sont absentes des virions maturés.



15.5. Stratégies d'inhibition de la réponse immunitaire par les *henipavirus* et les *rhabdovirus*

Tandis que pour la plupart des paramyxovirus la protéine V seule inhibe l'induction de l'IFN par l'ARN double brin, chez les henipavirus cette fonction est assurée de différentes manières par les protéines V et W, mais aussi par la protéine P (Shaw, Garcia-Sastre et al. 2004). V reste cytoplasmique et cible MDA5 et RIG-I en les empêchant d'activer les promoteurs des gènes des IFN de type I (Andrejeva, Childs et al. 2004; Childs, Andrejeva et al. 2009; Childs, Randall et al. 2012) tandis que W a une localisation nucléaire et bloque l'activité du TLR3 (Shaw, Cardenas et al. 2005) (**Figure 20**). En plus de cette inhibition de la reconnaissance, les protéines P/V/W bloquent également la voie de signalisation induite par l'interféron. La protéine V empêche l'activation de STAT1/STAT2 en les séquestrant dans des complexes de haut poids moléculaire dans le cytoplasme (Rodriguez, Parisien et al. 2002) (**Figure 20**). La protéine W assure l'inhibition dans le noyau, en séquestrant les protéines STAT dans le noyau. La région d'interaction entre les produits du gène P et la protéine STAT1 a été localisée dans le segment allant du résidu 111 au résidu 141 dans la région N terminale de V/W/P (Ciancanelli, Volchkova et al. 2009). Il est intéressant de noter que la mutation du résidu en position 125 suffit pour compromettre l'interaction entre V et STAT1 (Hagmaier, Stock et al. 2006). La région d'interaction avec STAT2 est moins bien définie, mais il semble que la fixation de STAT1 soit nécessaire pour la fixation de STAT2.

La protéine C des paramyxovirus une protéine multifonctionnelle, qui est impliquée dans la pathogénèse virale, l'inhibition de la réplication virale (Cadd, Garcin et al. 1996), l'inhibition des réponses de l'immunité innée cellulaire (Gotoh, Komatsu et al. 2002), l'inhibition de l'apoptose cellulaire induite par le virus (Koyama, Irie et al. 2003) et le bourgeonnement des particules virales (Sakaguchi, Kato et al. 2005). Chez les henipavirus c'est une protéine basique de 166 aa pour laquelle peu de choses sont connues concernant sa structure. Le mécanisme par lequel C inhibe les effets antiviraux des IFNs α et β n'est également pas bien compris mais il semble que la protéine C régule la réponse précoce proinflammatoire des cellules infectées (Mathieu, Guillaume et al. 2012; Mathieu, Guillaume et al. 2012).

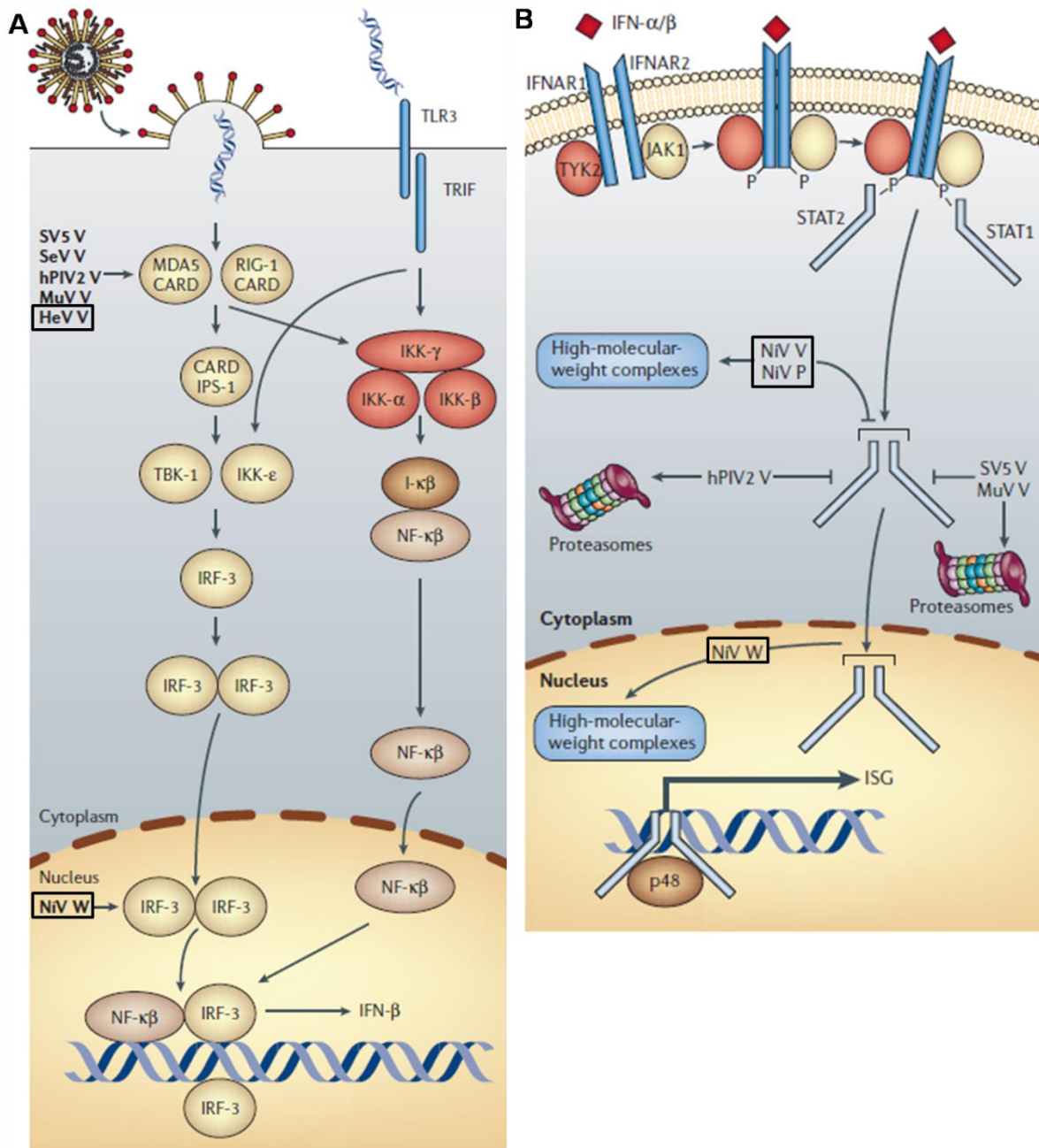


Figure 20 : Inhibition du système immunitaire cellulaire par les henipavirus (figure d'après (Eaton, Broder et al. 2006) avec modifications). A) Mécanismes d'inhibition de la signalisation et l'activation de la production des interférons par la cellule pendant l'infection par les henipavirus. Les protéines V de NiV et HeV bloquent MDA5 et RIG-1 dans le cytoplasme tandis que la protéine W bloque la signalisation induite par l'infection dans le noyau. B) Inhibition de la réponse à l'interféron. Les protéines V et moindre degré la P des henipavirus bloquent les activateurs de transcription STAT1 et STAT2 dans le cytoplasme tandis que cette activité est exercée par W dans le noyau. Les protéines STAT ne sont pas dégradées (comme chez les rubulavirus) mais sont séquestrées dans des gros complexes moléculaires.

Les *Rhabdoviridae* ont également développé des stratégies pour interférer avec le système immunitaire cellulaire. Dans cette famille les vesiculovirus et les lyssavirus utilisent des stratégies très différentes. Les vesiculovirus se répliquent rapidement et en nombreuses copies ce qui provoque une réponse immunitaire massive. Pour contrer celle-ci, ces virus ont

développé des moyens très efficaces pour réduire ou arrêter l'expression de presque tous les gènes de l'hôte, en agissant sur la transcription des ARNm et leur transport du noyau vers le cytoplasme, ainsi que sur la traduction des ARNm en protéines. Deux molécules virales participent à ces activités – l'ARN leader et la protéine M. Le rôle de l'ARN leader n'a pas été clairement établi, par contre, il a été montré que l'expression de la protéine M seule en culture cellulaire bloque l'expression d'autres gènes (Black and Lyles 1992; Ahmed and Lyles 1998), et des virus possédant un gène M défectif, ne peuvent pas inhiber efficacement l'expression des gènes de l'hôte (Black, Rhodes et al. 1993). La protéine M bloquerait l'export du noyau vers le cytoplasme des ARNm de l'hôte en interagissant avec la protéine Rae I (Faria, Chakraborty et al. 2005). L'initiation de la traduction des ARNm de l'hôte est également inhibée par la protéine M (Lodish and Porter 1980). La plupart des effets cytopathiques d'une infection par VSV sont causés par l'induction de l'apoptose (Kopecky, Willingham et al. 2001). Il est généralement accepté que l'apoptose cellulaire a pour but de limiter la production du virus mais des expériences d'infection par VSV ont montré que même en retardant l'apoptose, par la surexpression de protéines antiapoptotiques, il n'y a pas de différence dans le nombre de particules virales récupérés à partir des cellules lysées (Kopecky and Lyles 2003). Ces résultats suggèrent que le cycle viral de VSV est complété longtemps avant que la cellule meure par apoptose.

Contrairement aux vesiculovirus, les lyssavirus se répliquent plus lentement et en conséquence induisent des réponses immunitaires cellulaires de moindre importance. De plus les lyssavirus inhibent efficacement le système immunitaire cellulaire grâce aux différentes isoformes de la phosphoprotéine. Ces protéines sont capables d'interagir avec la protéine PML, une protéine induite par l'interféron et dont l'absence induit une augmentation de l'infection virale (Brzozka, Finke et al. 2005), provoquant une réorganisation des corps nucléaires PML. Ceci suggère que la phosphoprotéine joue un rôle important dans l'inhibition de l'immunité cellulaire. D'autre part les différentes isoformes de la P interagissent avec le facteur de transcription STAT1 provoquant l'inhibition de la transduction du signal de réponse aux interférons (Vidy, Chelbi-Alix et al. 2005). Cette inhibition n'est pas liée à la dégradation du STAT1 (comme par le virus simien 5 ou MuV) ou à l'inhibition de sa phosphorylation mais plutôt à son accumulation sous une forme inactive dans le noyau (Brzozka, Finke et al. 2005). L'infection par RaV induit très peu d'effets cytopathiques dans les cellules infectées et peut établir une infection persistante (Fernandes, Wiktor et al. 1963).

Objectifs du travail

La phosphoprotéine de virus à ARN négatif non-segmenté est un composant essentiel de la machinerie virale de transcription/réplication, agissant comme cofacteur non-catalytique de l'ARN polymérase ARN-dépendante et comme chaperon de la N dans le complexe N⁰-P compétent pour l'encapsidation de nouveaux génomes viraux. Ce travail s'inscrit dans la démarche du laboratoire d'obtenir une description de cette machinerie virale à la meilleure résolution possible afin d'en comprendre les mécanismes d'action et de régulation. Le travail que nous avons entrepris avait deux objectifs principaux. Premièrement, obtenir une caractérisation structurale de la phosphoprotéine entière des virus de la famille des *Rhabdoviridae* et de la famille des *Paramyxoviridae* en utilisant comme systèmes modèles la stomatite vésiculaire (VSV) et de Nipah (NiV). Les études réalisées au préalable dans notre laboratoire et dans d'autres laboratoires avaient démontrés que cette protéine est modulaire, comprenant de longues régions désordonnées alternant avec des régions structurées en domaines, mais aucune représentation de la P entière n'avait été proposée (Gérard, Ribeiro et al. 2007; Gérard, Ribeiro et al. 2009). Pour la P de VSV, nous disposions des structures atomiques du domaine central (Ding, Green et al. 2006) et du domaine C-terminal (Ribeiro, Favier et al. 2008), mais pour la P du NiV, nous avons dû déterminer la structure atomique de ces deux domaines.

Notre deuxième objectif était de déterminer la structure du complexe N⁰-P, indispensable pour la multiplication virale, afin de comprendre les mécanismes moléculaires par lesquels P maintient N dans un état compétent pour l'encapsidation d'ARN viraux néo-synthétisés en empêchant la polymérisation de la N et l'encapsidation d'ARN. Bien qu'il soit possible de produire le complexe N⁰-P de VSV ou RaV en co-exprimant les deux protéines dans une même cellule (Mavrakis, Iseni et al. 2003), la purification en grande quantité s'est avérée difficile. Sur la base de la structure atomique du complexe N-ARN, nous avons choisi de développer une stratégie de reconstitution du complexe. En ciblant le complexe N⁰-P du NiV, nous avons également pour objectif d'obtenir la première structure atomique d'une nucléoprotéine de la sous-famille de *Paramyxovirinae*.

Chapitre 2 : Matériels et Méthodes

2.1. Clonage

Au cours de ce travail, nous avons réalisé de nombreuses constructions des gènes de la phosphoprotéine (P) et de la nucléoprotéine (N) du virus de la stomatite vésiculaire (VSV), du virus de la rage (RaV), du virus Nipah (NiV) et du virus Hendra (HeV) et ainsi que de la protéine STAT1h (human Signal Transducer and Activator of Transcription 1). Ces constructions ainsi que la purification des produits de leur expression sont résumées dans le **Tableau 1**. Les gènes de la P et N du virus Nipah, souche UMMC1, ont été fournis par le Dr. Viktor Volchkov, ENS Lyon. Le gène de STAT1 a été fourni par le Dr. Daniel Panne, EMBL Grenoble.

L'expression de ces protéines et fragments de protéines a été réalisée dans des cellules d'*Escherichia coli* (Rosetta2) en utilisant les plasmides d'expression pETM40 et pET28. Les séquences d'intérêt ont été amplifiées par PCR et purifiées. Après digestion par les enzymes de restriction appropriés (NcoI/XhoI ou BamHI/SalI, Fastdigest® de Fermentas) les inserts ont été ligaturés (T4 Ligase, Fermentas) dans les vecteurs, digérés et déphosphorylés (CIAP phosphatase, New England Biolabs) au préalable. L'amplication des vecteurs a été réalisée dans des cellules *E. coli* TOP10 et les séquences ont été vérifiées par séquençage.

2.2. Expression et Purification

Pour la production des différentes protéines les bactéries sont cultivées à 37°C. Lorsque la densité optique à 600nm (DO_{600}) atteint une valeur de 0.5, la température est réduite à 18°C, l'expression est induite en ajoutant de l'Isopropyl β -D-1-thiogalactopyranoside (IPTG) à concentration finale de 1 mM et la culture est poursuivie pour une durée de 12 à 18h. Les cellules sont collectées par centrifugation à 6000 g à 4°C. Ensuite, le culot est solubilisé dans un tampon de lyse (Tableau 1) et les cellules sont cassées par sonication. Le lysat cellulaire est centrifugé à 40000 g, et la fraction soluble est récupérée pour les étapes de purification. Toutes les protéines ont été purifiées en deux étapes. La première étape est une purification par chromatographie affinité (Tableau 1). Le lysat cellulaire obtenu est chargé sur la résine en deux passages. Ensuite, la résine est lavée et les protéines sont éluées avec le même tampon, contenant 300 mM imidazole pour les protéines contenant une étiquette hexa-histidine ou 50 mM maltose pour les protéines contenant une étiquette « maltose binding protein » (MBP). Dans les deux cas la solution d'éluion est supplémentée avec un cocktail d'inhibiteurs de protéase (Complete, Roche). Après vérification de la pureté par SDS-PAGE, les fractions contenant la protéine d'intérêt sont

concentrées en utilisant des concentrateurs Sartorius ou Millipore (cut-off de 3 kDa à 10 kDa en fonction de la protéine), et la solution est injectées sur la colonne de chromatographie d'exclusion de taille (Tableau 1). Les fractions contenant la protéine sont regroupées et concentrées. La pureté de chaque protéine est vérifiée par SDS-PAGE et la concentration est déterminée par intégration du pic de chromatographie d'exclusion de taille en utilisant la mesure de réfractométrie différentielle (RI). Le clivage de l'étiquette MBP N-terminale est effectué en utilisant un mutant du domaine catalytique de la protéase TEV (Tobacco Etch Virus) que nous avons purifié au préalable. Toutes les étapes sont effectuées à 4°C ou dans la glace.

La production de la protéase TEV a été réalisée avec le plasmide pRK793 contenant le gène de la TEV en fusion avec des étiquettes poly-histidine et MBP. Le plasmide est transformé dans des cellules *E. coli* (Rosetta2). L'expression est induite par ajout de 1 mM IPTG à 30°C pendant 5h. Après autoclivage, la TEV est purifiée sur colonne d'affinité Ni²⁺ (HisSelect, Sigma).

Le fragment 132-683 de la protéine STAT1h est purifié comme décrit dans (Vinkemeier, Cohen et al. 1996) avec des modifications apportées. Brivement après la précipitation au sulfate d'ammonium la protéine a été solubilisée dans un tampon de 20 mM Tris-HCl à pH 8.8 contenant 150 mM NaCl et éluée à travers une colonne SP sepharose équilibrée dans le même tampon. La protéine STAT1₁₃₂₋₆₈₃ n'est pas retenue sur la colonne, contrairement à divers contaminants qui sont ainsi éliminés. La solution contenant STAT1₁₃₂₋₆₈₃ est ensuite concentrée et éluée à travers une colonne d'exclusion de taille Superdex S200 (GE Healthcare) équilibrée dans un tampon 20 mM Tris-HCl à pH 7.5 contenant 150 mM NaCl. Pour les mesures d'ITC la solution tampon a été échangée en injectant la protéine sur la même colonne de gel filtration, pré-équilibrée dans 20 mM phosphate de sodium à pH 7.5 et 150 mM NaCl. Les étapes de purification sont effectuées à 4°C.

Pour les purifications des fragments désordonnés des phosphoprotéines nous avons utilisé une solution tampon, contenant comme additif 50 mM des acides aminés glutamate et arginine (50mM RE) (Golovanov, Hautbergue et al. 2004). Cet additif a un effet de stabilisation à long terme et diminue la dégradation des protéines. L'addition de 50 mM RE dans la solution tampon pour plusieurs fragments de la P de VSV et de NiV pendant leur purification et leur stockage à long terme a énormément facilité leur étude.

Nous avons également utilisé l'arginine seul comme additif. L'utilisation de cet acide aminé dans le repliement des protéines et l'inhibition de leur agrégation est largement décrite dans la

littérature ((Tsumoto, Umetsu et al. 2004) et références citées). L'utilisation des solutions tampon supplémentées avec de l'arginine (0.1 à 0.4 M) a permis la purification et la caractérisation structurale du domaine centrale P_{CD} et le fragment P₄₇₁₋₇₀₉ de la phosphoprotéine du virus Nipah.

Pour les études par spectroscopie de Résonance Magnétique Nucléaire (RMN) les protéines sont stablement marquées avec ¹⁵N et/ou ¹³C comme décrit dans (Marley, Lu et al. 2001). La dernière étape de purification (chromatographie d'exclusion de taille) est réalisée dans un tampon 20 mM Bis-Tris à pH 6.0, contenant 150 à 500 mM NaCl, 50 mM Glu, 50 mM Arg et 0.5 mM Tris 2-carboxyethyl phosphine (TCEP). Les spectres RMN des fragments des différentes protéines ont été enregistrés à 25°C sur des spectromètres Bruker 600 et 800 MHz. Les expériences de RMN sont effectuées en collaboration avec l'équipe du Dr. Martin Blackledge, IBS, Grenoble.

La phosphoprotéine de VSV deutérée à 80% a été produite par le laboratoire de deutération de l'ILL (Dlab, CIBB) et purifiée comme décrit dans (Gérard, Ribeiro et al. 2007).

2.3. Analyse bioinformatique

Afin de localiser les limites entre régions désordonnées et structurées à partir de la séquence en acides aminés des phosphoprotéines et des nucléoprotéines des virus VSV et NiV, nous avons utilisé seize algorithmes de prédiction et nous avons calculé un consensus sous la forme d'un score, appelé D-SCORE (Ribeiro, Favier et al. 2008; Gérard, Ribeiro et al. 2009). Pour chaque prédiction, un score de 1 est attribué à chaque acide aminé prédit structuré et un score de 0 est attribué à chaque acide aminé prédit désordonné. Pour chaque acide aminé, le score obtenu à partir des 16 prédicteurs est moyenné et porté en graphique en fonction de la séquence en acides aminés. Les régions caractérisées par des valeurs de D-score moyen supérieures à 0.5 (plus de la moitié des prédictions indiquent que cet acide aminé fait partie d'une région structurée) sont considérées comme structurées, les régions caractérisées par des valeurs de D-score moyen inférieures à 0.5 sont considérées comme désordonnées. Le programme COILS a été utilisé pour détecter des régions suspectées de former des faisceaux d'hélices surenroulées (Lupas, Van Dyke et al. 1991).

Tableau 1 : Constructions de NiV, VSV et RaV utilisées et leur purification

Protéine/ Construction	Virus	Etiquette	Colonne d'affinité	Tampon de lyse	Gel filtration	Tampon de gel filtration
P entière	VSV	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem +TCEP
P1-177	VSV	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem +TCEP
P1-105	VSV	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem +TCEP
P1-80	VSV	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem
P1-60	VSV	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S75	idem
P17-40	VSV	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S75	idem
N _{Δ21}	VSV	MBP Nter	Amylose	CB	-	-
P1-50	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S75	idem
P1-70	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S75	idem
P1-50	Nipah	MBP Nter	Amylose	CB RE50 AP	S75	idem
P1-100	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem +TCEP AP
P92-190	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem +TCEP AP
P92-250	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem +TCEP AP
P173-240	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S75	idem +TCEP
P222-319	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem +TCEP
P300-401	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem +TCEP
P387-479	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S200	idem +TCEP
P471-580	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 R400 AP	S200	CB RE50 +TCEP
P655-709	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB	S75	idem
P655-709	Nipah	MBP Nter	Amylose	CB	S75	idem
G683D	Nipah	MBP Nter	Amylose	CB	S75	idem
K687R	Nipah	MBP Nter	Amylose	CB	S75	idem
R661K	Nipah	MBP Nter	Amylose	CB	S75	idem
P588-650	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S75	idem
P471-709	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 R200 AP	Superose 6	CB RE50 +TCEP
P1-580	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 R100 AP	Superose 6	CB RE50 +TCEP
P entière	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 R100 AP	Superose 6	CB RE50 +TCEP
N entière	Nipah	-	-	CB AP	-	-
N30-532	Nipah	MBP Nter	Amylose	CB RE50 AP	-	-
N2-402	Nipah	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	-	-
N30-402	Nipah	MBP Nter	Amylose	CB RE50 AP	-	-
N30-384	Nipah	MBP Nter	Amylose	CB	-	-
P665-707	Hendra	His ₆ Cter	Ni ²⁺ HisSelect	CB	S75	idem
P665-707	Hendra	MBP Nter	Amylose	CB	S75	idem
P1-68	RAV	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S75	idem
P1-55	RAV	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S75	idem
P1-42	RAV	His ₆ Cter	Ni ²⁺ HisSelect	CB RE50 AP	S75	idem
N _{Δ23}	RAV	MBP Nter	Amylose	CB	-	
STAT1 132-683	human	-	Purifié comme décrit dans (Vinkemeier, Cohen et al. 1996)			

L'étiquette MBP en N-terminal est clivée à des étapes tardives. Résine Ni²⁺ HisSelect de Sigma ; résine Amylose de New England Biolabs (NEB). CB – 20 mM Tris-HCl 150 mM NaCl ; RE50 – 50mM Arginine/Glutamate ; R100/400 – arginine 100/400 mM ; AP – cocktail anti protéases sans EDTA de Roche ; TCEP – Tris Carboxyéthyl Phosphine HCl de Sigma ; S75/S200 – colonnes Superdex de GE Healthcare.

2.4. Dichroïsme circulaire (CD)

Le dichroïsme circulaire repose sur la capacité d'un échantillon possédant un chromophore chiral ou placé dans un environnement asymétrique, d'absorber différemment la lumière polarisée circulairement droite et la lumière polarisée circulairement gauche. Le spectre dichroïque correspond à la différence d'absorbance entre ces deux types de lumière, pour chaque longueur d'onde. Dans le cas des protéines, la liaison peptidique est principalement responsable de l'absorbance dans l'ultraviolet lointain (< 250 nm) et la mesure du dichroïsme circulaire renseigne sur le contenu moyen en structure secondaire.

Les mesures de dichroïsme circulaire ont été réalisées avec un spectrophotomètre JASCO J-810 équipé d'un système Peltier pour le contrôle de la température. Les spectres CD dans l'UV lointain ont été enregistrés à 20 °C dans une cuve de 1 mm en utilisant une concentration de protéine de l'ordre de 20 µM. Après soustraction du signal du tampon, l'ellipticité molaire par résidu ($[\Theta]$) est calculée avec l'équation :

$$[\Theta]_{222nm} = \frac{0.1 \delta}{C l N_{aa}} [deg.cm^2.dmol^{-1}]$$

où δ est le signal CD mesuré en mdeg, l est la longueur de la cuvette, N_{aa} est le nombre d'acides aminés de la protéine et C est sa concentration. Le contenu moyen en hélice α peut être déterminé à partir de la valeur de l'ellipticité molaire en utilisant l'équation suivante :

$$f_{\alpha} = \frac{[\Theta]_{exp}}{[\Theta]_{helix}}$$

où $[\Theta]_{helix} = -35,000 deg.cm^2.dmol^{-1}$ représente la valeur de l'ellipticité molaire par résidu pour 100% d'hélice α calculée à partir de l'équation suivante dérivée pour des peptides (Luo and Baldwin 1997):

$$[\Theta]_{helix} = (-44,000 + 250 T) \left(1 - \frac{n}{N_{aa}}\right)$$

où T est la température en degrés Celsius, n est le nombre de groupes CO libre dans le peptide et N_{aa} est le nombre d'acides aminés du peptide. Pour les protéines, nous utilisons une valeur de $n = 1$, en supposant que la majorité des groupes CO à l'extrémité N-terminale des hélices α sont coiffés et une valeur de $N_{aa} = 10$ en supposant une longueur moyenne 10 résidus pour les différentes hélices présentes dans la protéine.

Les dénaturations thermiques ont été enregistrées à 222 nm dans une cuvette de 1 cm de trajet optique en utilisant une concentration de protéine de l'ordre de 2 μM . La dénaturation et la renaturation sont suivies en variant la température de 4 à 95 °C à une vitesse allant de 1 à 4 °C par minute.

2.5. Chromatographie d'exclusion de taille couplée à la diffusion de lumière statique et à la réfractométrie (SEC-MALLS-RI)

L'intensité de la lumière diffusée par une solution monodisperse de particules est directement proportionnelle au produit de la masse moléculaire et de la concentration. La technique de la chromatographie d'exclusion de taille (*Size Exclusion Chromatography* : SEC) couplée à la détection par diffusion de lumière à plusieurs angles (*Multi-Angle Laser Light Scattering* : MALLS) et à la réfractométrie (*refractometry index* : RI) fournit une mesure absolue de la masse moléculaire des particules. L'utilisation d'une détection en ligne en sortie de colonne de chromatographie d'exclusion de taille permet d'une part d'éliminer les contaminants ou les agrégats qui, même en quantité limitée, pourraient perturber la mesure de l'intensité statique de lumière en cuvette et d'autre part de séparer plusieurs composants qui seraient présents en solution.

La calibration de la colonne de SEC avec des protéines de rayon hydrodynamique (R_h) connu (Uversky 1993) permet d'estimer le R_h pour la particule étudiée. La SEC est réalisé à l'aide d'une pompe HPLC et le signal de diffusion de lumière est mesuré à 18 angles à l'aide d'un détecteur DAWN-EOS (Wyat Technology Corp., Santa Barbara, CA) en utilisant un laser de longueur d'onde de 690nm. L'excès d'indice de réfraction est mesuré par un réfractomètre différentiel RI2000 (Schambeck). Les données sont analysées avec le logiciel ASTRA (Wyat).

Pour une solution diluée de particules, l'intensité de la lumière diffusée par les particules en solution est reliée au produit de la masse moléculaire des particules et de leur concentration par :

$$\frac{R_{\theta}}{KC} = MP(\theta) - 2A_2CM^2P^2(\theta)$$

où R_θ est le rapport de Rayleigh mesuré, C est la concentration en protéine, M est la masse moléculaire (g.mol^{-1}), $P(\theta)$ est le facteur de forme qui dépend de la structure des particules diffusant la lumière et décrit une dépendance angulaire, A_2 est le second coefficient virial et K est une constante optique qui dépend des caractéristiques de l'instrument et de la solution :

$$K = \frac{1}{N_A} \left(\frac{2\pi n_0}{\lambda^2} \right)^2 \left(\frac{dn}{dc} \right)^2$$

où N_A est le nombre d'Avogadro, n_0 est l'indice de réfraction du solvant à la longueur d'onde incidente (1.333 pour les tampons aqueux), dn/dc est l'incrément d'indice de réfraction du soluté (0.185 mL.g^{-1} pour les protéines) et λ est la longueur d'onde. Pour les particules de taille largement inférieure à la longueur d'onde de la lumière utilisée, la dépendance angulaire est négligeable et $P(\theta) = 1$. Pour une solution diluée, A_2 est également négligeable, et l'équation se simplifie, donnant :

$$\frac{R_\theta}{K} = C \times M$$

La mesure de l'excès d'indice de réfraction fournit une alternative très intéressante à la mesure de l'absorbance pour la mesure de la concentration de la molécule étudiée. Contrairement au coefficient d'extension molaire, l'incrément d'indice de réfraction varie très peu avec la composition en acides aminés d'une protéine ou même entre protéines et acides nucléiques. Une valeur générique de dn/dc pour les protéines ou les acides nucléiques peut être utilisée. Au cours de la séparation par SEC un chromatogramme est obtenu pour chaque détecteur et en utilisant l'équation décrivant R_θ il est possible d'obtenir la masse moléculaire pour chaque point du chromatogramme. Pour chaque tranche du chromatographe, la masse moléculaire peut-être déterminée à partir de l'intensité mesurée à chaque angle et de la concentration. La masse moléculaire moyenne pour chaque pic chromatographique est déterminée en moyennant les valeurs de masse moléculaire obtenue pour chaque tranche du chromatogramme. Le graphe de la masse moléculaire en fonction du volume d'éluion fournit une mesure de l'homogénéité de l'espèce éluant de la colonne. Ainsi, pour une espèce monodisperse, la masse moléculaire calculée est constante pour tous les points à travers le pic.

Toutes les protéines utilisées pour des mesures biophysiques ont été analysées par SEC-MALLS-RI en utilisant différentes colonnes selon la taille de la protéine (Superdex S75 et S200 HR10/30 (GE Healthcare) ; TSK G3000SW et G4000SW (TOSOH) ; Shodex KW-803). Typiquement, ces colonnes étaient équilibrées à 20°C dans un tampon 20 mM Tris HCl à pH 7.5 contenant 150 mM NaCl.

2.6. Diffusion aux petits angles (SAS)

La diffusion aux petits angles est une méthode utilisée pour l'analyse des macromolécules en solution. Elle n'est limitée ni en masse moléculaire, ni en taille ou en niveau de repliement. Elle fournit une mesure de la taille, masse moléculaire et rayon de giration, et de la forme de la molécule étudiée. Pendant une expérience l'échantillon est illuminé par un faisceau monochromatique de rayons X (généralement de longueur d'onde d'environ 1 Å) ou de neutrons (généralement de longueur d'onde entre 2 et 10 Å), et la lumière diffusée aux petits angles est enregistrée sur un détecteur. Pendant cette expérience on mesure l'intensité de la lumière diffusée en fonction du vecteur de diffusion \vec{q} défini par :

$$\vec{q} = \vec{k}_i - \vec{k}_f$$

ou \vec{k}_i et \vec{k}_f sont respectivement les vecteurs d'onde des faisceaux incident et diffusé (**Figure 1**). On relie le module de \vec{q} à l'angle de diffusion par :

$$q = \frac{4\pi \sin\theta}{\lambda}$$

où 2θ est l'angle de diffusion et λ est la longueur d'onde.

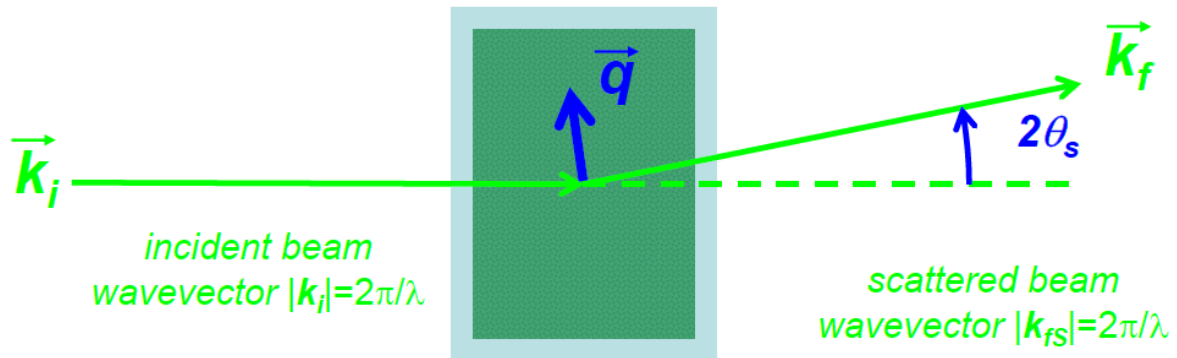


Figure 1: Illustration d'une expérience de diffusion aux petits angles. \vec{k}_i et \vec{k}_f sont les vecteurs d'onde du faisceau incident et diffusé, respectivement. \vec{q} est le vecteur de diffusion et 2θ est l'angle de diffusion.

La diffusion aux petits angles est une méthode de contraste et l'intensité mesurée est due au contraste entre les particules étudiées et le solvant dans lequel elles se trouvent. Lorsqu'on utilise des rayons X (SAXS) le signal de diffusion est extrait de la différence ($\Delta\rho$) entre la densité électronique moyenne de la molécule étudiée (ρ_M) et celle de la solution (ρ_S) :

$$\Delta\rho = \rho_M - \rho_S$$

Dans une expérience de diffusion par des particules en solution, le signal provient de des molécules orientées de manière anisotrope. La diffusion en solution est continue et radialement symétrique, et la courbe de la molécule étudiée est obtenue après réduction des données et soustraction du signal provenant du solvant seul.

Approximation de Guinier et rayon de giration R_g

Aux très petits angles la diffusion peut être décrite par l'approximation de Guinier :

$$\ln I(q) = \ln I(0) - \frac{1}{3} R_g^2 q^2$$

où $I(0)$ est la diffusion à angle nul et R_g est le rayon de giration (qui est donné par la racine carrée de la somme des distances entre le centre de gravité et tous les points de la particule). La masse moléculaire de la particule diffusant la lumière peut être déduite de l'ordonnée à l'origine, $I(0)$, et la valeur de R_g peut être déduite à partir de la pente ($R_g = \sqrt{\text{pente}/3}$). Pour une solution monodisperse de particules, le graphique de Guinier, $\ln I(q)$ en fonction de q^2 , est linéaire sur une gamme plus ou moins restreintes de valeurs de q selon la forme et la flexibilité de la molécule. L'approximation est valable sur une plus large gamme de q lorsque le R_g de la molécule est plus petit. Pour les protéines globulaires, le graphique de Guinier est linéaire pour des valeurs de q telles que le produit $qR_g < 1.3$, tandis que pour des protéines allongées ou désordonnées, le graphique est linéaire uniquement pour des valeurs de q telles que $qR_g < 1.1$. L'absence de linéarité dans ce graphique est caractéristique d'une agrégation. D'autre part, pour les échantillons qui se comportent bien, le R_g ne varie pas d'une façon significative avec la concentration de la molécule étudiée. Une variation significative avec la concentration peut indiquer une oligomérisation ou une agrégation de la molécule ou l'existence de d'interactions intermoléculaires.

« A cautionary word is in order regarding the use of the measurement of the radius of gyration of a particular protein as the sole criterion for random-coil behavior. Other conformations can have similar radii of gyration »

(Tanford, 1968)

La théorie des polymères, développée par Flory, permet de décrire les dimensions globales d'un polymère qui se comporte comme une chaîne aléatoire en solution (random coil). L'équation de Flory décrit la dépendance du rayon de giration moyen, R_g , vis-à-vis de la longueur de la chaîne N , selon l'équation :

$$R_g = R_0 N^\nu$$

où R_0 est une constante et ν est l'exposant de Flory. Pour un polymère de chaîne d'épaisseur infinitésimale, dans un « bon » solvant, la valeur de ν est de 0.5 (Flory 1960). Pour un polymère réel, dont la chaîne a une certaine épaisseur, cette valeur est de 3/5. Une valeur d'environ 0.598 (R_0 de 2.08) a été déterminée à partir de mesures réalisées avec des protéines chimiquement dénaturées (Kohn, Millett et al. 2004), et plus récemment, une valeur de 0.522 (R_0 de 2.54) à été proposée pour les protéines intrinsèquement désordonnées qui seraient donc plus compactes en solution que les protéines dénaturées chimiquement (Bernado and Blackledge 2009).

Une manière d'adresser la question du rapport entre dimensions et nombre de monomères dans la chaîne aléatoire est d'examiner directement les relations internes de mise à l'échelle (scaling relationships). Si la distance entre deux résidus dans une chaîne désordonnée est R_{ij} alors la distance moyenne (RMS) suit la relation :

$$\langle R_{ij} \rangle \propto N_{ij}^\nu$$

ou ν est l'exposant de Flory, à condition que le nombre de résidus N_{ij} soit suffisamment grand ((Johansen, Jeffries et al. 2011; Johansen, Trewhella et al. 2011) et références citées).

La valeur de ν peut être estimée directement à partir du profil de SAXS. Sur une gamme intermédiaire de valeurs de q , l'intensité diffusée suit la relation :

$$I(q) \propto q^{-D_m}$$

ou D_m est une constante (Glatter and Kratky 1982). Dans un graphique de $\log I(q)$ en fonction de $\log q$, la valeur de D_m est estimée à partir de la pente de la partie linéaire du graphique (**Figure 2A**) qui reflète la distribution des distance interatomiques à des échelles de longueurs différentes. Le paramètre D_m représente les relations entre la masse enfermée dans un volume et les dimensions du volume (Johansen, Jeffries et al. 2011; Johansen, Trewhella et al. 2011). Plus généralement la valeur de D_m pour un polymère est reliée à l'exposant de Flory :

$$D_m = \frac{1}{\nu}$$

Pour un polymère idéal qui peut se croiser cette valeur est 2 tandis que pour un polymère réel dans un bon solvant, qui ne se croise pas et du aux effets du volume exclu de la chaîne, Dm est 1.67 (Beaucage G. 1996 et références citées). Ainsi la valeur de Dm et donc de v peut être estimée directement d'une courbe de SAXS. Cette valeur ne dépend pas de la taille du polymère mais seulement de ces relations internes ce qui permet de comparer des chaînes de différents nombre de monomères. Dm est également sensible à la solvation de la chaîne.

Intensité extrapolée à angle zéro, $I(0)$

L'intensité à l'angle zéro, $I(0)$, est déterminée par extrapolation car l'angle zéro correspond à la direction du faisceau incident qui est bloqué par un masque (« beam stop ») afin de protéger le détecteur. Sur une échelle absolue, $I(0)$ dépend du carré du nombre d'électrons dans la particule diffusante et ne dépend pas de sa forme. $I(0)$, normalisé par la concentration c est proportionnelle à la masse moléculaire M de la molécule :

$$\frac{I0}{c} = \frac{N_A M}{\mu^2} (1 - \rho(s)\psi)$$

où N_A est le nombre d'Avogadro, $\rho(s)$ la densité électronique du solvant, ψ est le volume électronique spécifique moyen et μ est la masse moyenne par nombre d'électrons. En pratique l'expérience est calibrée avec une (ou plusieurs) protéine de masse moléculaire connue et de concentration connue. Très récemment une autre méthode pour le calcul de la masse moléculaire a été proposé (Rambo and Tainer 2013) qui présente l'avantage de ne pas dépendre de mesure avec des protéines standards.

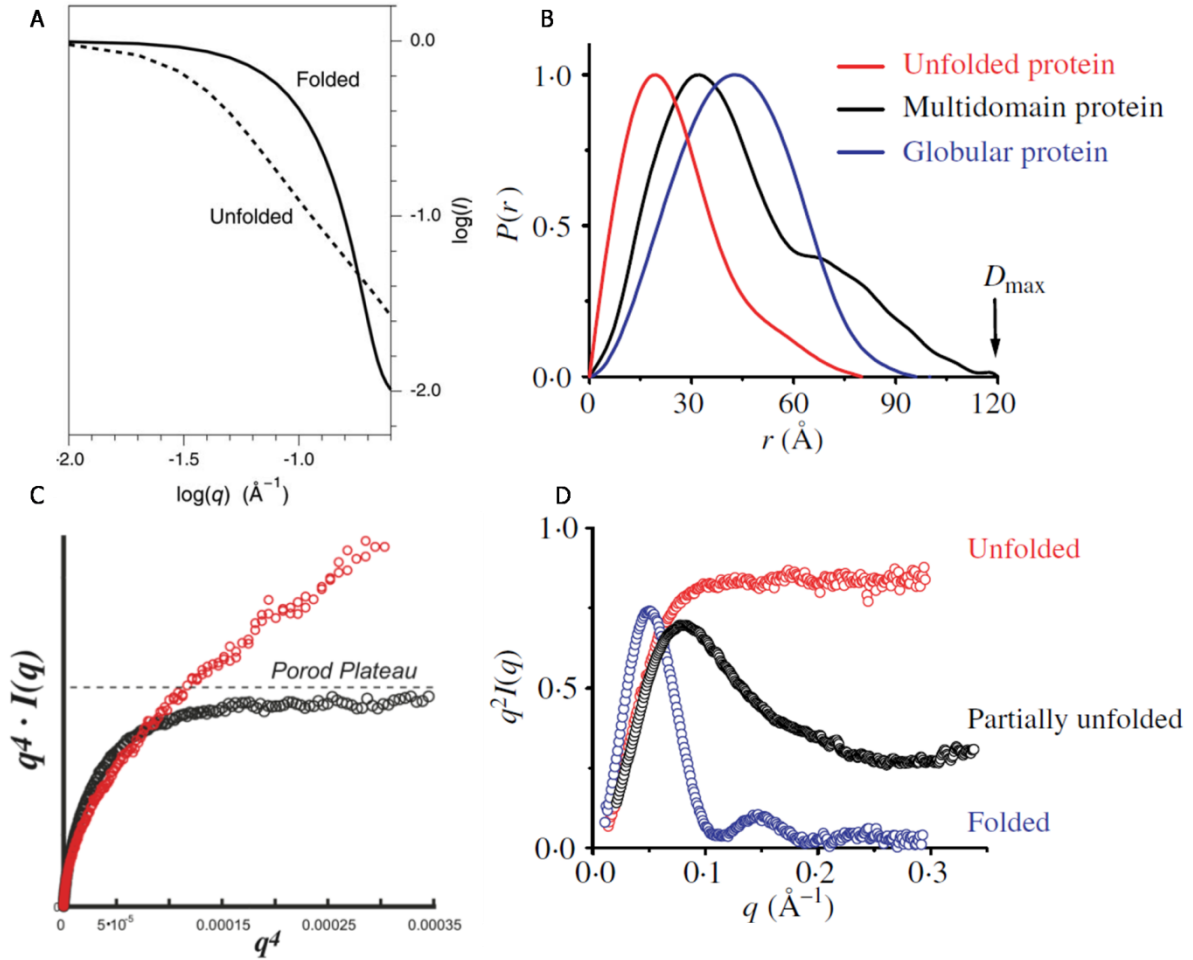


Figure 2 : Différentes représentations d'une courbe de diffusion aux petits angles. **A)** Graphique $\log I(q)$ en fonction de $\log q$. La pente négative du graphique aux angles moyens représente D_m . Les courbes solide et pointillé représentent une protéine globulaire et une protéine désordonnée respectivement. **B)** Fonction de distribution de paires $P(r)$ pour une protéine désordonnée (ligne en rouge), une protéine à plusieurs domaines majoritairement structurée (ligne en noir) et une protéine globulaire (ligne en bleu). **C)** Graphique de Porod-Debye pour une protéine majoritairement structurée (ronds en noir) et une protéine désordonnée (ronds en rouge). **D)** Graphique de Kratky pour une protéine désordonnée (ronds en rouge), une protéine à plusieurs domaines majoritairement structurée (ligne en noir) et une protéine globulaire (ronds en bleu). (Figures d'après (Putnam, Hammel et al. 2007; Johansen, Trehwella et al. 2011; Rambo and Tainer 2011))

Fonction de distribution des distances, $P(r)$

La fonction de distribution des distances, $P(r)$, est un histogramme de distribution des distances entre paires d'atomes dans la molécule étudiée :

$$P(r) = \frac{2}{\pi} r^2 \int_0^{\infty} q^2 I(q) \frac{\sin qr}{qr} dq$$

En pratique, il est impossible de mesurer $I(q)$ jusqu'à $q = 0$ et $q \rightarrow \infty$ et ces valeurs sont extrapolées. Le fonction est supposée égale à zéro pour les distances supérieurs à l'extension maximale de la particule, D_{max} . Cette fonction peut être calculée directement à partir de la courbe de diffusion en utilisant une transformée de Fourier (Svergun 1992). Les paramètres $I(0)$ et R_g peuvent être également déterminés à partir de cette fonction et un bon accord est attendu entre leurs valeurs déterminées à partir de la $P(r)$ et à partir de l'approximation de Guinier. La forme de la fonction de $P(r)$ en fonction de r (nm) permet d'évaluer quelques caractéristiques de la forme globale de la molécule étudiée (**Figure 2B**). Une particule parfaitement sphérique a une $P(r)$ en forme de cloche symétrique dont le maximum est situé à $D_{max}/2$. Pour les molécules désordonnées ou allongées dans une direction, la symétrie de la courbe est perdue et son maximum se trouve à $r < D_{max}/2$. Les protéines contenant plusieurs domaines sont caractérisées le plus souvent par la présence de plusieurs maxima dans leur $P(r)$. Des protéines dont les domaines sont rigides génèrent une $P(r)$ présentant divers maxima et minima, alors que les protéines très flexibles génèrent une courbe lisse (Bernado 2010).

Graphique de Kratky

La représentation de Kratky permet d'estimer qualitativement l'état de repliement de la molécule étudiée (Glatter 1982). Dans le graphique de $I(q) \cdot q^2$ en fonction de q , une molécule entièrement structurée et globulaire donne une courbe en cloche, alors qu'une molécule désordonnée et flexible donne une courbe croissante. Les molécules partiellement désordonnées présentent un profil intermédiaire. (**Figure 2D**). Il faut cependant noter que cette représentation est très sensible à la soustraction du solvant et au bruit expérimental.

Graphique de Porod-Debye

La loi de Porod-Debye décrit une approximation de puissance 4 entre q et l'intensité de diffusion observée ((Rambo and Tainer 2011) et références citées) :

$$I(q) \approx \Delta\rho^2 \frac{2\pi}{q^4} S$$

ou $\Delta\rho$ est le contraste de densité électronique entre la particule et le solvant et S est la surface de la particule étudiée. Similaire à l'approximation de Guinier, cette approximation est

valable aux très petits angles, typiquement $q < 0.15 \text{ \AA}^{-1}$, et suggère que la diffusion d'une molécule structurée décroît avec q^{-4} .

En pratique la transformation de la courbe de diffusion d'une molécule bien structurée en plot de $q^4 \cdot I(q)$ en fonction de q^4 doit générer une courbe hyperbolique qui approche un plateau lorsque l'analyse est limitée aux petits angles, du à l'existence d'un contraste $\Delta\rho$ homogène et bien défini entre la densité électronique de la molécule ρ_M et le solvant ρ_S . L'absence d'un tel plateau suggère que le contraste entre ρ_M et ρ_S est moins bien défini et cela est caractéristique d'une flexibilité au sein de la molécule étudiée (**Figure 2C**). Cette représentation est moins sensible à la soustraction du solvant car utilise seulement l'information aux très petits angles.

Diffusion de neutrons aux petits angles (SANS) et variation de contraste

Dans la méthode de diffusion aux petits angles en solution, la courbe de diffusion obtenue après la soustraction de la courbe de diffusion du solvant dépend du contraste entre les particules étudiées et le solvant dans lequel elles se trouvent. Le contraste peut être altéré en modifiant les propriétés de diffusion des atomes (les amplitudes de diffusion atomiques) et/ou les propriétés de diffusion du solvant (la densité de diffusion du solvant). Lorsque les neutrons sont utilisés comme rayonnement pour la diffusion, les amplitudes de diffusion atomiques et la densité de diffusion du solvant peuvent être modifiées en échangeant les isotopes des éléments qui les constituent. En effet, les amplitudes de diffusion de l'hydrogène et de son isotope, le deutérium, sont très différentes. En ajustant la densité de diffusion du solvant en variant le rapport D_2O/H_2O dans le solvant, il est possible d'obtenir un contraste nul entre la densité moyenne d'une particule et celle du solvant. Dans ce cas, la molécule d'intérêt ne contribue plus à la diffusion, elle est « invisible ».

Dans le cas de l'étude de complexes protéine-protéine, s'il est possible de produire un des partenaires dans une forme deutérée, cette technique permet d'étudier ce partenaire au sein du complexe indépendamment des autres. Un des partenaires est produit en remplaçant les atomes d'hydrogène par des atomes de deutérium (« protéine deutérée ») et le complexe est formé avec le second partenaire dans sa forme naturelle (« protéine hydrogénée »). Le rapport D_2O/H_2O dans le solvant est ajusté de sorte que la densité moyenne du solvant égale la densité d'un des deux partenaires, rendant celui-ci invisible. La protéine hydrogénée peut être égalée pour une solution contenant environ 40 % de D_2O . Si la protéine « deutérée » à un

taux de deutération de 85 %, il est également possible d'égaliser la densité de diffusion de la protéine deutérée avec une solution contenant environ 100% D₂O. Il est donc possible en choisissant les conditions de visualiser l'un ou l'autre partenaire dans le complexe, ou les deux. Le graphique de $\sqrt{I(0)/C}$ (C étant la concentration en protéine) en fonction du pourcentage de D₂O est linéaire si le complexe est monodisperse et dépourvu d'agrégation.

Collecte et traitement des données

Pour les expériences de SAXS, chaque protéine est injectée sur la colonne de chromatographie d'exclusion de taille et la fraction centrale du pic d'éluion est collectée et utilisée. Cette méthode permet de s'assurer de l'absence d'agrégats. De plus, elle permet d'échanger facilement le solvant et d'obtenir un échantillon de solvant de composition parfaitement identique à celui contenant la protéine. Cela est nécessaire pour enregistrer une courbe de SAXS qui, par soustraction, permet d'éliminer la contribution du solvant. L'utilisation du SEC-MALLS-RI permet également de vérifier la monodispersité de l'échantillon et de déterminer sa concentration par intégration du pic de réfractométrie.

Pour les expériences de diffusion de neutrons, la collecte et la réduction des données ont été réalisées sur la ligne D22 de l'Institut Laue Langevin avec l'aide du Dr. Phil Callow. Pour la collecte des données, les échantillons sont placés dans des cuves de quartz de 1 mm d'épaisseur et les courbes sont enregistrées à 6°C pendant 6 h (par échantillon) à deux distances échantillon-détecteur (2 et 6 mètres) successivement afin d'enregistrer des données aux très petits angles et aux angles moyens. Les données sont réduites et le solvant est soustrait en utilisant les programmes GRASP et IGOR (programmes ILL).

Les données de diffusion des rayons X aux petits angles ont été collectées sur la ligne de lumière BM29 à l'ESRF en injectant l'échantillon à travers le capillaire de mesure pendant l'irradiation aux rayons X. Typiquement dix courbes, chacune résultante de 2 secondes d'exposition, sont moyennées après une inspection pour s'assurer de l'absence de dégâts d'irradiation. Au minimum trois concentrations sont mesurées par protéine afin de s'assurer de l'absence d'une dépendance à la concentration qui signifierait une agrégation, une oligomérisation ou des phénomènes de répulsion ou attraction entre les molécules. Les données de SAXS sont ensuite moyennées de manière radiale et le signal du tampon est soustrait en utilisant le programme PRIMUS (Konarev, Volkov et al. 2003). Le rayon de giration R_g est calculé pour chaque concentration en utilisant l'approximation de Guinier, comme implémenté dans PRIMUS.

Les données de diffusion des rayons X aux petits angles ont été utilisées pour déterminer une enveloppe pour les complexes $N_{\Delta 21}^0$ -P₆₀ de VSV et N_{32-383}^0 -P₅₀ de NiV et également pour la protéine P₄₇₁₋₇₀₉ en utilisant les programmes DAMMIN et DAMMIF (Svergun 1999). Le point de départ de ces programmes consiste en une sphère de diamètre correspondant au *Dmax* déterminé précédemment avec le programme GNOM (Svergun 1992), remplie par un ensemble de billes figurant des pseudo-résidus. Chaque bille a des propriétés de diffusion des rayons X correspondantes à la moyenne pour les acides aminés. Une procédure de recuit simulé est utilisée afin de déterminer la combinaison pour laquelle la courbe de diffusion théorique correspond le mieux à la courbe de diffusion expérimentale. Parce que ce problème n'a pas une unique solution et que généralement plusieurs modèles peuvent reproduire la courbe expérimentale, la pratique couramment utilisée consiste à construire plusieurs modèles et à les comparer entre eux.

Le programme DAMAVER permet d'aligner les différents modèles dans l'espace à trois dimensions en déterminant un score qui qualifie la correspondance entre les modèles (NSD, Normalized Spatial Discepancy) (Volkov and Svergun 2003).

Différents programmes permettent de calculer la courbe de diffusion théorique à partir d'un modèle structural. Nous avons utilisé les programmes CRY SOL (Svergun, Barberato et al. 1995) et CRYSON (Svergun, Richard et al. 1998) qui transforment respectivement un modèle atomique en une courbe de diffusion de SAXS ou SANS.

2.7. Cristallographie

2.7.1. Cristallogenèse

A la différence de SAXS et comme le terme « Cristallographie aux rayons X » le suggère, l'échantillon examiné se trouve dans un état cristallin. De façon générale les méthodes de cristallisation des protéines nécessitent la dissolution de la protéine purifiée dans un tampon aqueux contenant un agent précipitant (polyéthylène glycol, sels, composés organiques). Trois phases sont communes au phénomène de cristallisation de toutes les molécules : la nucléation, la croissance cristalline et l'arrêt de la croissance. Pendant la nucléation les molécules s'associent afin de former un agrégat thermodynamiquement stable, un noyau critique, qui constitue la surface de croissance cristalline. L'étape de croissance se caractérise par la diffusion des molécules jusqu'à la surface de nucléation et leur assemblage en cristal ordonné. L'arrêt de croissance s'impose quand la concentration en protéine diminue drastiquement ou avec l'incorporation d'impuretés dans le cristal.

Le processus de croissance cristalline peut être visualisé sur un diagramme de phase qui décrit l'état stable en fonction des concentrations de protéine et précipitant (**Figure 3**). Quand la concentration de protéine dépasse ses limites de solubilité on parle de l'état de sursaturation. Si la sursaturation est importante, les molécules forment des agrégats amorphes, si elle est intermédiaire on observe la formation de centre de nucléation et la croissance cristalline et si elle est faible, seule la croissance de cristaux déjà nucléés est possible.

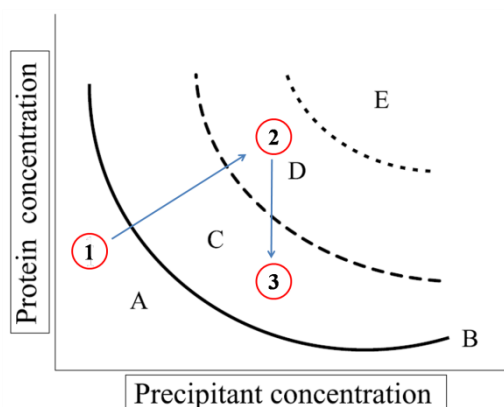


Figure 3 : Diagramme de phase. **A)** protéine en solution ; **B)** Courbe de solubilité délimitant une zone de sous saturation (A) et une zone de sursaturation (B-E) ; **C)** Zone métastable de sursaturation faible favorisant la croissance cristalline ; **D)** Zone labile de nucléation ou encore zone de sursaturation intermédiaire ; **E)** Zone de sursaturation forte, favorisant la précipitation.

Plusieurs techniques de cristallisation existent, dont les plus célèbres sont la diffusion de vapeur et le « micro batch ». Nous avons essentiellement utilisé la technique de diffusion de vapeur qui consiste à enfermer une goutte de la solution de macromolécule dans laquelle le précipitant et la protéine sont mélangés, (typiquement 0.1 à 5 μ L) avec un réservoir de précipitant de volume beaucoup plus important (typiquement 0.1 à 2ml) et de laisser le système s'équilibrer. Selon sa position dans le système, la goutte peut être assise ou suspendue. La diffusion de vapeur dans ce système fermé résulte en transfert net d'eau de la goutte contenant la macromolécule vers le réservoir de précipitant de manière à égaliser la concentration de précipitant dans les deux compartiments. A ce point la diffusion d'eau de la goutte provoque l'augmentation de [protéine] et [précipitant] ce qui déplace les conditions en diagonale vers la région de nucléation/croissance cristalline (sursaturation intermédiaire) ((1) \rightarrow (2) **Figure 3**). Après la formation des premiers cristaux la concentration de protéine diminue, provoquant le déplacement des conditions vers la zone de croissance (nommée encore zone métastable) ((2) \rightarrow (3) **Figure 3**).

Plusieurs variables influencent la formation de cristaux de macromolécules et notamment la méthode de cristallisation utilisée, le volume de la goutte, le rapport de volume goutte/réservoir, la pureté de la molécule, la concentration de la molécule et le précipitant, le

pH, la température, les vibrations, la source et l'âge de la protéine. Une approche couramment choisie aujourd'hui pour obtenir des conditions de cristallisation consiste à cribler un grand nombre de conditions. Pour cela, nous avons utilisé le robot de cristallisation de l'EMBL Grenoble (HTX Lab) qui utilise la méthode de la goutte assise. Une fois que des conditions avaient été identifiées, nous avons réalisé un affinement de ces conditions de manière manuelle en utilisant essentiellement la méthode de la goutte suspendue.

Ensemencement

Un des paradoxes de la cristallisation est que les conditions optimales pour la formation des cristaux ne sont pas optimales pour permettre leur croissance. La nucléation spontanée a plus de chance de se produire à une sursaturation intermédiaire (région D dans la **Figure 3**) tandis que la croissance lente, qui permet l'obtention des cristaux bien ordonnés, est favorisée à une sursaturation faible (région C dans la **Figure 3**). La méthode d'ensemencement (« seeding ») permet de séparer la phase de nucléation de la phase de croissance cristalline. Cette méthode consiste à utiliser les semences (seeds) obtenues précédemment et de les transférer dans des gouttes fraîches dans des conditions de sursaturation plus faible. Le transfert de semences microscopiques obtenues en cassant des cristaux est nommé « Microseeding » tandis que le transfert de cristaux entiers est nommé « Macroseeding ». Le « seeding » est utilisé afin d'optimiser la croissance ou la forme cristalline, mais il peut également être utilisé pour permettre la cristallisation d'une protéine qui diffère de celle utilisée pour former les premiers cristaux, qu'il s'agisse d'une protéine homologue, d'une protéine contenant un ligand ou d'une protéine modifiée chimiquement ou qui contient des mutations (« seeding hétérologue »).

Nous avons utilisé la technique de « microseeding » afin d'améliorer les cristaux du domaine P_{XD} de la phosphoprotéine du virus Hendra (voir **Chapitre 3**) ainsi que pour obtenir des cristaux du complexe N₃₂₋₃₈₃⁰-P₅₀ de NiV sélénié en utilisant du matériel cristallin du complexe N₃₂₋₃₈₃⁰-P₅₀ natif comme semence (voir **Chapitre 5**).

2.7.2. La diffraction et la densité électronique

Les rayons X diffractés à partir d'un cristal peuvent être traités mathématiquement comme s'ils étaient réfléchis sur un plan formant un angle θ avec le faisceau incident et donc les maximums de diffraction mesurés pendant l'expérience sont fréquemment appelés

« réflexions ». Chaque réflexion enregistrée sur le détecteur est le résultat d'un rayon diffracté et peut être décrite par un facteur de structure F_{hkl} . Chaque facteur de structure correspond à une somme de Fourier dans laquelle chaque terme représente la contribution de chaque atome à la réflexion :

$$F_{hkl} = \sum_{i=1}^n f_i e^{2\pi i(hx_i + ky_i + lz_i)}$$

où f_i est le facteur de diffusion de l'atome i et l'exponentiel représente une fonction périodique tridimensionnelle. x_i, y_i et z_i sont les coordonnées de l'atome i dans la maille élémentaire ; h, k et l , en plus d'être les fréquences, représentent également les indices de la réflexion et les phases α_{hkl} sont implicites dans la représentation exponentielle. Donc chaque rayon diffracté est une onde complexe qui résulte des contributions diffractives de tous les atomes dans la maille élémentaire. Chaque facteur de structure F_{hkl} peut être également représenté comme la somme des contributions de tous les éléments volumiques de densité électronique présents dans la maille élémentaire :

$$F_{hkl} = \int_V \rho(x, y, z) e^{2\pi i(hx + ky + lz)} dV$$

où V est le volume de la maille élémentaire et l'intégrale est calculée pour l'ensemble des valeurs de x, y et z dans la maille élémentaire. Chaque élément volumique contribue au facteur de structure F_{hkl} avec une phase déterminée par ses coordonnées. Après une transformée de Fourier on obtient :

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx + ky + lz)}$$

où F_{hkl} représente une réflexion et sa phases α_{hkl} est toujours implicite. F_{hkl} est un vecteur complexe qui peut être décomposé :

$$F_{hkl} = |F_{hkl}| \times e^{i\alpha_{hkl}}$$

où $|F_{hkl}|$ est l'amplitude du vecteur complexe F_{hkl} et α_{hkl} est sa phase par rapport à l'origine (ou encore l'angle entre le vecteur et l'axe des nombres réels sur un diagramme de Harker). Après remplacement et transformation l'équation du haut devient :

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i(hx + ky + lz - \frac{\alpha}{2\pi})}$$

Cette équation permet de calculer la densité électronique dans la maille élémentaire en utilisant les modules $|F_{hkl}|$ (les amplitudes des facteurs de structure représentant les réflexions enregistrées sur le détecteur) et les phases α_{hkl} pour la réflexion hkl .

Résoudre la structure d'une macromolécule par diffraction des rayons X consiste à obtenir une carte de cette densité électronique en trois dimensions qui permette de construire

un modèle atomique de la macromolécule étudiée. La mesure de l'intensité I_{hkl} de chacune des réflexions permet de déterminer l'amplitude de l'onde diffractée dans la direction correspondante, puisque l'intensité est proportionnelle au carré de l'amplitude. Or, pour reconstruire la densité électronique à partir de ces mesures, il faut connaître non seulement l'amplitude de l'ensemble des ondes diffractées, mais aussi leur déphasage par rapport au rayonnement incident, autrement dit leurs phases α_{hkl} .

L'amplitude de chacun des facteurs de structure, étant déterminée directement de manière expérimentale, le problème central de la cristallographie est d'obtenir la phase de chacun de ces facteurs de structure. Diverses méthodes permettent d'obtenir ces phases, utilisant uniquement des données expérimentales, la structure connue d'une autre protéine ou une structure prédite par modélisation.

Phasage expérimental

Chaque atome dans la maille élémentaire contribue à chaque réflexion enregistrée sur les images de diffraction. La perturbation des réflexions causée par l'addition d'un atome peut être utilisée afin d'estimer les phases initiales. L'ajout d'un atome lourd dans un cristal « natif » de protéine afin d'obtenir des phases s'appelle « remplacement isomorphe ». Le cristal ainsi obtenu s'appelle un « dérivé ». La perturbation induite dans les réflexions peut être utilisée afin d'estimer les phases initiales. Plusieurs conditions doivent être satisfaites pour le succès d'une telle approche. La première condition est que l'ajout de l'atome lourd ne perturbe pas les paramètres de la maille cristalline et plus précisément n'induit pas un changement de conformation de la protéine. La deuxième condition est que l'ajout de l'atome lourd provoque un changement d'intensité de plusieurs réflexions. La contribution du remplacement isomorphe à l'amplitude des facteurs de structure et des phases est bien illustrée sur le diagramme d'Argand (**Figure 4A**). L'amplitude d'une réflexion mesurée à partir du cristal natif et à partir du cristal dérivé sont respectivement nommées $|FP|$ et $|FPH|$. La différence isomorphe est donnée par :

$$|FH| \cong |FPH| - |FP|$$

Cette différence peut être utilisée dans l'estimation des amplitudes du facteur de structure de l'atome lourd afin de déterminer sa position en utilisant la méthode de Patterson ou des méthodes directes. A partir des coordonnées de l'atome lourd sa phase α_H peut être estimée permettant d'estimer la phase α_P de la protéine native. Une « ambiguïté de phase » subsiste

puisque deux solutions sont possibles ; les deux cercles de la **Figure 4C** se coupent en deux endroits. La méthode utilisant un seul dérivé ou SIR (Single Isomorphous Replacement) n'est donc, en théorie, pas suffisante pour résoudre le problème de phase. L'utilisation de plusieurs dérivés par la méthode MIR (Multiple Isomorphous Replacement) est donc nécessaire. Un deuxième dérivé lourd permet de lever l'ambiguïté en identifiant la solution commune aux deux situations (**Figure 5A**). Afin de résoudre l'ambiguïté de phase du premier dérivé, il est nécessaire que l'atome lourd du second dérivé se fixe sur la protéine à un endroit différent du premier.

Le facteur de diffusion (ou encore facteur de forme) atomique f dépend de la longueur λ d'onde et de l'angle α de diffraction et possède trois composants :

$$f(\alpha, \lambda) = f_0(\alpha) + f'(\lambda) + f''(\lambda)$$

où $f_0(\alpha)$ est le facteur de diffusion normal qui dépend seulement de l'angle, $f'(\lambda)$ et $f''(\lambda)$ correspondent respectivement au terme de dispersion et au terme d'absorption, qui dépendent seulement de la longueur d'onde pour un élément donné.

Lorsque la longueur d'onde est loin du seuil d'absorption de l'atome, le facteur de diffusion possède seulement le composant normal (dit aussi réel) et la loi de Friedel est respectée (loi indiquant que les paires de réflexions liées par le centre de symétrie de la maille réciproque ont des intensités identiques) ce qui veut dire que les réflexions $h k l$ et $-h - k - l$ ont la même amplitude et des phases complémentaires. A une longueur d'onde proche du seuil d'absorption atomique f' et f'' contribuent également au facteur de diffusion, créant une variation des intensités pour les réflexions $h k l$ et $-h - k - l$ et introduisant un décalage entre leurs phases, qui ne sont donc plus complémentaires (**Figure 4B**).

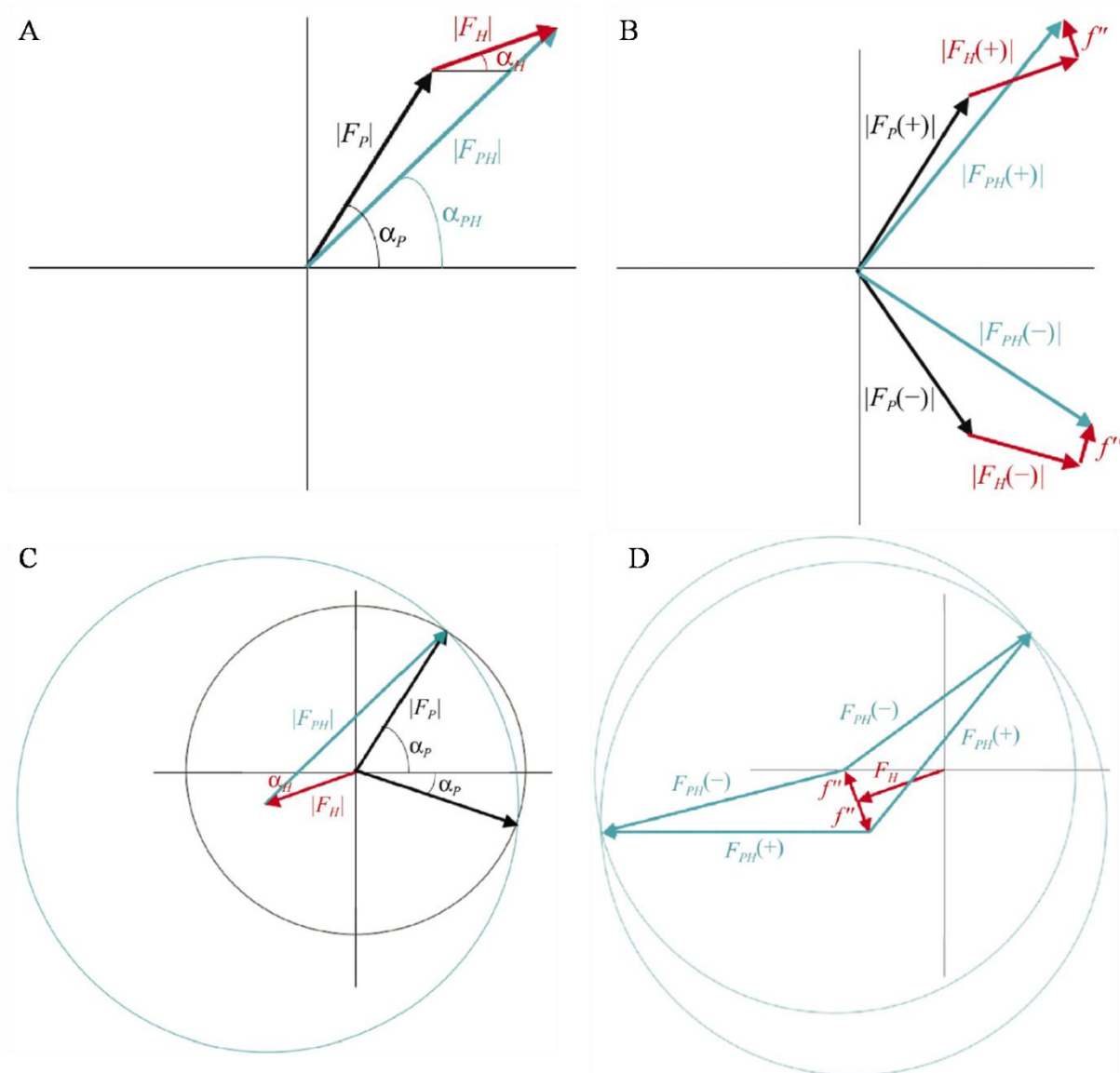


Figure 4 : Le phasage expérimentale (figures d'après (Taylor 2003) avec modifications) **A)** Diagramme d'Argand pour SIR. et $|F_P|$ est l'amplitude pour une réflexion du cristal natif, et $|F_{PH}|$ pour le cristal dérivé et $|F_H|$ est la différence isomorphe (ou encore facteur de structure de l'atome lourd). **B)** Diagramme d'Argand représentant la loi de Friedel. et $|F_P(+)|$ et $|F_P(-)|$ sont les paires de Friedel en absence de diffusion anormale, $|F_{PH}(+)|$ et $|F_{PH}(-)|$ sont les paires de Friedel lorsqu'une diffusion anormale est présente, et $|F_H(+)|$ et $|F_H(-)|$ et f'' représentent la contribution du diffuseur anormal. **C)** Construction de Harker pour SIR et **D)** pour SAD.

Ce phénomène de diffusion anormale à une longueur d'onde donnée ou SAD (Single-wavelength Anomalous Diffraction) peut être utilisé de la même façon que le remplacement isomorphe afin de localiser les atomes qui absorbent à la longueur d'onde choisie. La recherche des diffuseurs anomaux est réalisée en utilisant la méthode de Patterson ou des méthodes directes, telles que celles implémentées dans les programmes SOLVE (Terwilliger and Berendzen 1999) et SHELXD (Schneider and Sheldrick 2002). Toutefois, cette approche est également limitée par la même ambiguïté de phase (**Figure 4D**).

Une solution à ce problème est de combiner la méthode SIR avec la méthode de diffusion anormale SAD en utilisant un dérivé lourd par la méthode SIRAS (single Isomorphous Replacement with Anomalous Scattering) ou avec plusieurs dérivés lourds par la méthode MIRAS (Multiple Isomorphous Replacement with Anomalous Scattering) (**Figure 5 B**). Nous avons utilisé la méthode SIRAS pour résoudre la structure du domaine P_X de la phosphoprotéine du virus Nipah (voir **Chapitre 3**).

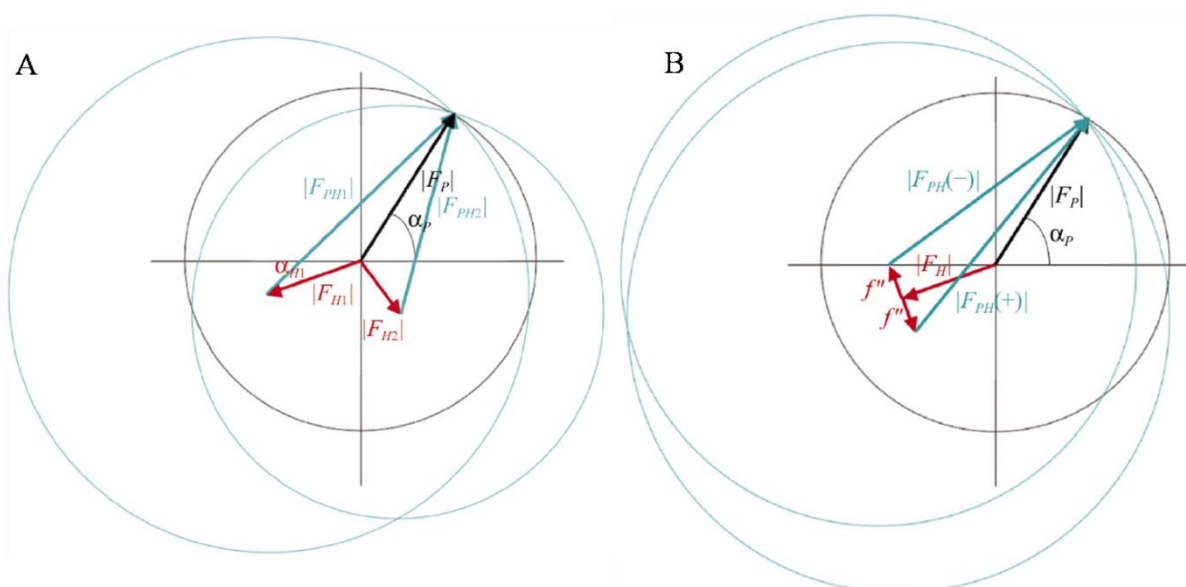


Figure 5 : Solution de l'ambiguïté de phase (figures d'après (Taylor 2003) avec modifications). Diagramme de Harker pour **A**) la méthode MIR avec deux dérivés d'atomes lourds et **B**) la méthode SIRAS.

Une autre manière de résoudre l'ambiguïté de phase avec des données de diffusion anormale à une seule longueur d'onde (SAD) consiste à utiliser des procédures de modification de la densité électronique, comme implémenté par exemple dans les programmes RESOLVE (Terwilliger 2000), SHELXE (Uson, Stevenson et al. 2007) et DM (Cowtan 1999). En modifiant la densité électronique initiale, ces méthodes permettent de générer de nouvelles phases qui, combinées avec les phases expérimentales par le moyen de lissage du solvant (solvent flattening), d'inversion du solvant (solvent flipping), de moyenne des copies de la molécule liées par une symétrie non cristallographique (NCS) ou d'autres méthodes, permettent également de lever l'ambiguïté de phase. Nous avons principalement utilisé les programmes SHELXE (comme implémenté dans le programme HKL2MAP (Pape and Schneider 2004)) et le programme DM pour les modifications de densité électronique.

« However, there is still a need for accurate data, successful phase improvement and a certain amount of luck »

(Dodson 2003)

Finalement, la collecte de données à plusieurs longueurs d'onde en présence de diffuseur anomal ou méthode MAD (Multi-wavelength Anomalous Diffraction) permet également de lever l'ambiguïté de phase.

Dérivation des cristaux

De nombreux composés peuvent être utilisés pour générer des dérivés à partir de cristaux natifs qui soient utilisables pour le phasage expérimental tels que les gaz nobles, les sels d'atomes lourds, les chélates de métaux, les halogénures ou les métaux alcalins. Deux grandes approches peuvent être utilisées pour introduire ces composés dans les cristaux protéiques : cristalliser la molécule en présence du composé ou bien en l'introduisant après l'obtention des cristaux.

La méthode la plus classique de dérivation consiste à tremper un cristal dans une solution de sels d'atomes lourds pour une période allant de quelques heures à quelques jours. Cette technique a permis de résoudre de nombreuses structures mais elle présente quelques inconvénients comme le large choix de composés à tester qui peut demander un temps considérable et beaucoup de cristaux. Récemment, il a été proposé que certains anions ou cations simples, comme les halogénures ou les métaux alcalins, peuvent être introduits dans les cristaux de protéine par des trempages rapides (quick soaks) (Dauter, Dauter et al. 2000; Nagem, Dauter et al. 2001). Cette technique qui consiste à tremper brièvement (10 à 300 secondes) le cristal natif dans une solution, contenant à la fois un cryoprotectant et un composé approprié à forte concentration, puis à congeler le cristal, combine en une étape la dérivation et la cryoprotection du cristal. Les bromures et iodures (anions de Br et I) sont des ions monoatomiques, mous et polarisables. Leur diffusion au sein du cristal est très rapide et ils se fixent sur la surface des protéines par des interactions de faible énergie à des groupes chargés positivement comme les chaînes latérales des arginines et lysines, à des donneurs de liaison hydrogène ou sur des surfaces hydrophobes. Les cations des métaux alcalins sont plus spécifiques, interagissant de préférence avec les groupements carboxyliques des aspartates et glutamates ou avec le groupement carbonyle de la liaison peptidique.

Dans ce travail, nous avons effectué des trempages rapides des cristaux du domaine P_{XD} du virus Nipah dans des solutions contenant les ions Br⁻, I⁻, Cs⁺ ou Rb⁺ à des concentrations de 0.5 à 1.0 M (voir **Chapitre 3**). Ces éléments ont un signal anomal

significatif aux longueurs d'onde accessibles sur les lignes de lumière de l'ESRF et peuvent être utilisés pour des phasages de type SAD/SIRAS (I^- et Cs^+) et MAD/MIRAS (Br^- et Rb^+).

Remplacement moléculaire

Le problème de phase peut être également résolu à partir de la structure connue d'une molécule structurellement proche de la molécule étudiée. Dans ce cas, les phases des facteurs de structure de la molécule connue, appelée « modèle de phasage » ou « modèle de remplacement moléculaire », sont utilisées pour le phasage de la nouvelle molécule. Le modèle de remplacement moléculaire peut correspondre à une partie de la molécule étudiée ou à une molécule très proche car homologue. Le modèle peut être une structure cristallographique mais peut également être un modèle dérivé de données de RMN (Mao, Guan et al. 2011), un modèle à basse résolution obtenu par reconstruction à partir de données de microscopie électronique ou de diffusion aux petits ou grands angles (Hao 2006), voir un modèle *ab initio* construit par modélisation moléculaire (Qian, Raman et al. 2007).

Les phases des facteurs de structure dépendent de la position des atomes dans la maille élémentaire et donc, afin d'utiliser les phases issues d'un modèle de remplacement moléculaire, il est nécessaire de positionner correctement celui-ci dans la maille. Pour cela les méthodes de Patterson sont utilisées afin de déterminer d'abord l'orientation du modèle dans la maille élémentaire et puis sa localisation exacte par une recherche translationnelle par rapport à l'origine. Ces méthodes sont implémentées dans les programmes PHASER (McCoy 2007) et MOLREP (Vagin and Teplyakov 2010). Nous avons utilisé la méthode de remplacement moléculaire avec le programme PHASER afin de résoudre la structure du complexe $N_{\Delta 21}^0$ -P₆₀ du virus de la stomatite vésiculaire (VSV).

Collecte et traitement des données de diffraction

Pendant ces travaux de nombreux jeux de données de diffraction ont été collectés sur les lignes de lumière BM14-4, ID23-1, ID23-2 et ID29 à l'ESRF, ainsi que sur la ligne PROXIMA-1 à SOLEIL.

Un phasage expérimental réussi repose sur la mesure très précise des petites différences entre les paires de Friedel, une opération qui peut être compromise par les dégâts provoqué par l'irradiation du cristal par les rayons X. En théorie, il est possible d'enregistrer les paires de Friedel sur la même image, néanmoins cela exige un alignement des axes du

crystal et du dispositif qui peut s'avérer fastidieux. Pour ces raisons, les données de diffractions pour le domaine central (P_{CD}) et les domaines C-terminaux (P_{XD}) de la phosphoprotéine des virus Nipah et Hendra ont été collectées en utilisant la stratégie du faisceau inversé (inverse beam). Dans cette approche, le cristal est tourné de 180° tous les 5 à 15° de collecte, assurant l'enregistrement successif dans le temps des paires de Friedel. Cette stratégie s'est avérée cruciale pour la résolution de la structure de NiV P_{CD} dont les cristaux subissaient d'importants dégâts lors de leur irradiation et qui n'avait pas pu être résolue par SAD ou MAD après une collecte normale.

L'objectif principal des programmes de traitements des données de diffractions est de générer une liste d'indices hkl avec leurs intensités I_{hkl} (et les erreurs associées σI_{hkl}) et une bonne estimation des paramètres de la maille cristalline. L'indexation se fait image par image en affinant plusieurs paramètres comme par exemple la distance cristal-détecteur, la position du centre du détecteur et les paramètres de maille. Pendant l'intégration les intensités des réflexions sont mesurées en appliquant un masque autour de chaque tâche. A l'intérieur de ces limites se trouve la surface à intégrer et l'extérieur permet d'estimer le bruit de fond. La géométrie de la maille, la symétrie de la diffraction et la présence d'absences systématiques permettent d'estimer le groupe d'espace. Pendant la collecte de données le cristal tourne et selon la forme et la position du cristal par rapport au faisceau, le volume du cristal, qui est exposés aux rayons X peut varier, conduisant à des variations d'intensité des réflexions. Une étape de mise à l'échelle des intensités est donc nécessaire afin de créer une cohérence interne.

Lors de la procédure de réduction des données, la qualité des données est estimée par le facteur R_{merge} :

$$R_{merge} = \frac{\sum_h \sum_l |I_{hl} - \langle I_h \rangle|}{\sum_h \sum_l \langle I_h \rangle}$$

Où h est l'indice pour les réflexions uniques, l et l'indice pour les réflexions symétriques redondantes et I_h est l'intensité moyenne. En pratique des facteurs R_{merge} normalisés par rapport à la multiplicité d'observation semblent plus appropriés (Evans 2011).

Au cours de ce travail, les programmes iMOSFLM (Battye, Kontogiannis et al. 2011) et XDS (Kabsch 2010) ont été utilisés pour l'intégration et l'indexation, les programmes SCALA (Evans 2011) et XSCALE (Kabsch 2010) pour et la mise à l'échelle, le programme POINTLESS (Evans 2011) pour la détermination du groupe d'espace et le programme XDSCONV (Kabsch 2010) pour la conversion des formats.

Construction et affinement des modèles

Après l'obtention d'une carte de densité électronique, qui typiquement permet de distinguer des éléments de structures secondaire, l'étape suivante consiste à construire un modèle atomique en se basant sur la séquence en acides aminés mais aussi sur des propriétés connues des protéines comme la structure des acides aminés, la longueur des liaisons covalentes, des angles entre les liaisons et ainsi de suite.

L'affinement des structures cristallographiques est un processus complexe qui est composé de différentes étapes. Chaque cycle d'affinement nécessite la sélection de paramètres du modèle, une cible d'affinement et une méthode d'optimisation. Les paramètres du modèle sont des variables qui décrivent le contenu du cristal tels que les coordonnées des atomes, paramètres de déplacements atomiques (abrégé en facteurs de température ou encore facteurs B), occupations des atomes, modèle du solvant et autres. Ces paramètres sont combinés et composent les facteurs de structure du modèle, F_{model} qui sont supposés être en bon accord avec les valeurs observées, F_{obs} . Et donc l'affinement de structure est un processus pendant lequel les paramètres du modèle sont altérés dans le but d'optimiser la fonction cible (d'affinement) de type:

$$E = E_{CHEM} + \omega E_{DATA}$$

ou E est la fonction à minimiser, E_{CHEM} représente les connaissances à priori et est une fonction qui décrit les interactions covalentes et non covalentes, E_{DATA} décrit la différence entre les données de diffraction observées F_{obs} et calculées F_{model} et ω donne le poids relatif. En pratique, cette fonction est beaucoup plus complexe (Pannu 1996). Les méthodes d'optimisation permettent d'améliorer l'accord entre les paramètres du modèle et les données expérimentales et donc la cible d'affinement. Les méthodes les plus utilisées sont (1) la construction manuelle dans la carte de densité électronique ou (2) le recuit simulé (Brunger and Adams 2002). La qualité du fit du modèle aux données de diffraction est donnée par le facteur R qui quantifie les différences entre les amplitudes des facteurs de structure observés F_o et calculés F_c :

$$R_{factor} = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$$

Néanmoins une valeur basse du facteur R ne correspond pas nécessairement à un bon modèle. Typiquement le facteur R peut être artificiellement abaissé en augmentant le nombre de paramètres ajustables ou autrement dit en diminuant le rapport « nombre d'observations /

nombre de paramètres ». Afin d'éviter cela les données de diffraction sont divisées en « réflexions de travail » (working set) qui comprends la majorité des réflexions et « réflexions de test » (test set) qui comprend 5 à 10% des réflexions. L'ajustement des facteurs de structures observés et calculés pour l'ensemble « de travail » est donné par le facteur R_{work} calculé selon l'équation ci-dessus. Les réflexions de l'ensemble « test » ne sont pas utilisées pour l'affinement de la carte de densité électronique, mais sont utilisées pour calculer un facteur R_{free} (Brunger 1992; Kleywegt and Brunger 1996). La comparaison des facteurs R_{work} et R_{free} permet de juger de la qualité de l'affinement.

La manière de présenter les chaînes latérales des acides aminés pour lesquelles, pour des raisons de flexibilité aucune densité électronique n'est observée dans la carte, reste une question ouverte. Trois possibilités existent : (1) couper la chaîne latérale créant ainsi une pseudo-alanine dans la séquence qui néanmoins est identifiée comme le vrai acide aminé, (2) affiner la structure en imposant une occupation zéro pour les atomes de la chaîne latérale qui ne sont pas visibles ou (3) construire une chaîne latérale dans le modèle et utiliser le facteur B pour indiquer sa grande flexibilité. Chacune de ces solutions à ses avantages et ses inconvénients. Nous avons choisi de modéliser les chaînes latérales en entier, même si leur densité électronique n'est pas claire, en affinant leur facteur B.

*« But at the end of the day, it is the driver's responsibility to learn the meaning of red light »
(ccp4bb)*

Dans ce travail, nous avons utilisé le programme Coot (Emsley and Cowtan 2004) pour construire et modifier manuellement le modèle, les programmes BUCCANEER ($N_{\Delta 21}^0$ - P_{60} de VSV, P_{XD} et P_{CED} de NiV) (Cowtan 2006), ARP/wARP (P_{XD} de HeV) (Langer, Cohen et al. 2008) et Autobuild ($N_{32-383}^0 P_{50}$ de NiV) (Terwilliger, Grosse-Kunstleve et al. 2008) pour construire automatiquement les modèles et les programmes d'affinement REFMAC5 (Murshudov, Skubak et al. 2011) et Phenix.refine (Afonine, Grosse-Kunstleve et al. 2012) pour affiner nos modèles.

2.8. Calorimétrie de titrage isotherme (ITC)

La calorimétrie de titrage isotherme est une méthode qui mesure l'échange de chaleur qui se déroule lors de l'interaction entre deux molécules (Wiseman, Williston et al. 1989). Dans une expérience, un des partenaires de l'interaction, le ligand, contenu dans une seringue, est titré dans une solution contenant l'autre partenaire, le récepteur, contenue dans une cellule. L'expérience est effectuée à température constante. La quantité de chaleur absorbée ou libérée lors de chaque addition d'une petite quantité de ligand dans la cellule de mesure contenant le récepteur est mesurée par rapport à une cellule de référence contenant le solvant. La quantité de chaleur mise en jeu est quantifiée comme la puissance électrique nécessaire pour maintenir la température constante entre les cellules de mesure et de référence, qui sont placées dans une enceinte adiabatique. Ainsi l'intégration de la différence de puissance électrique nécessaire pour maintenir la température par rapport au temps définit le changement de chaleur apparent $q_{i,app}$ qui correspond à la surface du pic i . Le changement de chaleur réel q_i est obtenu après soustraction du changement de chaleur non-spécifique $q_{i,ns}$ et du changement de chaleur dû à la dilution du ligand $q_{i,dill}$. La variation d'enthalpie ΔH et la constante d'association K_A sont directement obtenus à partir de:

$$q_i = n[M]_{tot} \times V_{cellule} \times \Delta H_{app} \times \mathbb{R}$$

ou $[M]_{tot}$ est la concentration totale du récepteur dans la cellule de mesure, $V_{cellule}$ est le volume et \mathbb{R} est la racine d'une équation quadratique résolue par rapport à la concentration de ligand et récepteur (Jelesarov and Bosshard 1999). A partir d'une expérience de titration ITC et en utilisant une régression non-linéaire à la base de l'équation ci-dessus on peut déterminer ΔH , K_A et n (le nombre de sites de fixation).

Des expériences de titrage calorimétrique isotherme ont été réalisées afin de caractériser l'interaction entre un fragment de la protéine STAT1 humaine (aa 132-683) et le peptide P₉₁₋₁₉₀ de la phosphoprotéine du virus de Nipah (**Chapitre 4**). Les mesures ont été effectuées dans une solution tampon 20 mM phosphate de sodium à pH 7.5, contenant 150 mM NaCl. Le tampon a été choisi en raison de sa faible enthalpie de protonation. Les mesures ont été effectuées à 20 °C sur les instruments ITC200 ou VP-ITC (Microcal). Le peptide P₉₁₋₁₉₀, utilisé comme ligand, était à une concentration de 1.5 mM et la protéine STAT₁₃₂₋₆₈₃ utilisé comme récepteur, était à une concentration de 70 µM. La chaleur de mélange observée lorsque du tampon est injecté dans la solution de STAT1 132-683 est négligeable et l'enthalpie de dilution du ligand dans une solution de tampon a été soustraite.

Résultats

Chapitre 3 : Structure de la phosphoprotéine

3.1. Introduction

La phosphoprotéine (P) est un composant essentiel de la machinerie de synthèse d'ARN responsable de la transcription et de la réplication du génome des virus à ARN négatif non-segmenté (Lamb 2007). Au sein de cette machinerie, la phosphoprotéine joue de multiples rôles au cours du cycle de multiplication virale. (1) Elle agit comme un cofacteur non-catalytique de la polymérase, formant avec la protéine L un complexe polymérasique à deux sous-unités. L'ARN génomique est complètement encapsidé par la nucléoprotéine qui s'assemble sous la forme d'un long homopolymère hélicoïdal, dénommé nucléocapside (NC). Ce complexe N-ARN et non l'ARN nu sert de matrice pour la polymérase virale, mais la protéine L seule ne peut pas se fixer sur la nucléocapside, alors que la protéine P est capable d'interagir à la fois avec la protéine L et avec le complexe N-ARN, permettant donc d'associer le complexe polymérasique avec la matrice N-ARN utilisée pour la transcription et la réplication (Chattopadhyay and Banerjee 1987; Paul, Chattopadhyay et al. 1988; Takacs, Das et al. 1993). (2) La protéine P agit également comme un chaperon de la nucléoprotéine monomérique, empêchant celle-ci de s'associer avec des ARN cellulaires et de polymériser en absence d'ARN viral. P constitue avec la forme soluble et sans ARN de N (N^0) un complexe essentiel pour la réplication du virus, qui permet l'encapsidation d'ARN néo-synthétisés – le complexe N^0 -P (Masters and Banerjee 1988; Curran, Marq et al. 1995; Gupta and Banerjee 1997; Mavrakis, Iseni et al. 2003; Mavrakis, Mehoulas et al. 2006). Toutefois, ni les mécanismes d'encapsidation de l'ARN viral, ni le rôle joué par la protéine L dans ces mécanismes ne sont actuellement connus. (3) A côté de ces fonctions dans le cycle de réplication viral, la protéine P ainsi que divers produits dérivés du même gène, telles que les protéines V, W, C des paramyxovirus, ou les protéines P1 à P5 du virus de la rage, sont responsables de l'antagonisme de la réponse interféron des cellules hôtes. Selon le virus, ces protéines interagissent avec différents facteurs impliqués dans les voies d'induction de la production d'interférons ou dans les voies de signalisation par l'interféron de la production de facteurs antiviraux et l'établissement d'un état antiviral. Cette protéine joue donc un rôle essentiel pour la réplication du virus dans la cellule hôte et dans le but de mieux comprendre la manière dont elle remplit ces fonctions, nous avons entrepris la caractérisation de ses propriétés structurales et dynamiques.

La séquence en acides aminés de la protéine P est faiblement conservée, et l'homologie n'est généralement pas identifiable à partir de la séquence au-delà du genre ou

de la famille. Par contre, des prédictions de désordre et de structure secondaire à partir de la séquence en acides aminés, ainsi que des caractérisations biochimiques et biophysiques de la protéine entière ou de fragment démontrent que la phosphoprotéine des MNV possède une organisation modulaire commune, avec une conservation de plusieurs régions fonctionnelles en allant de l'extrémité N- à l'extrémité C-terminale (**Figure 1**). La protéine P des *MNV* peut être divisée en deux grandes parties, une partie N-terminale globalement désordonnée (P_{NTR}) mais qui possède selon les virus des structures secondaires transitoires et une région C-terminale (P_{CTR} ou encore appelé P_{CED}) plus structurée, composée de deux domaines séparés par une connexion flexible (Karlin, Ferron et al. 2003; Gérard, Ribeiro et al. 2009; Habchi, Mamelli et al. 2012). P_{NTR} comprend les sites de reconnaissance de la nucléoprotéine N^0 et vraisemblablement de la protéine L. Le premier domaine structuré de P_{CTR} est responsable de l'oligomérisation de la protéine, formant des dimères chez les *Rhabdoviridae* et des tétramères chez les *Paramyxoviridae*. Le domaine structuré C-terminal est responsable de la fixation de P sur les nucléocapsides.

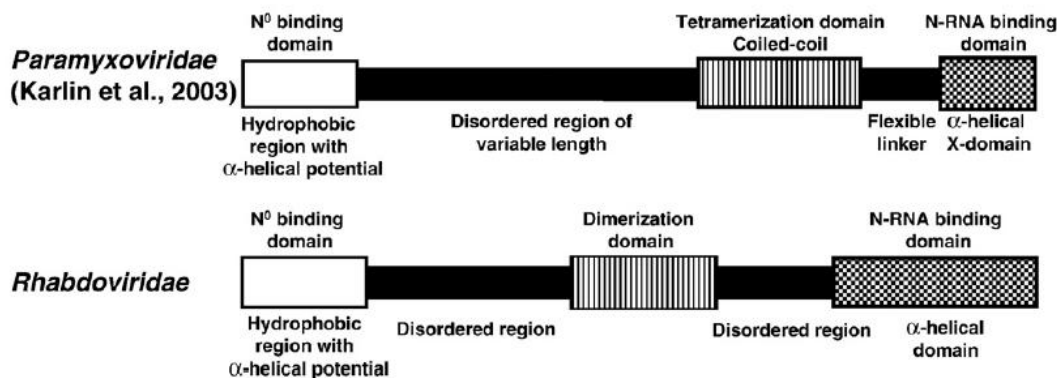


Figure 1 : Organisation modulaire des phosphoprotéines de *Rhabdoviridae* et *Paramyxoviridae* (figure d'après (Gérard, Ribeiro et al. 2009). Les rectangles représentent les domaines structurés.

Le caractère modulaire de la P et la présence de longues régions désordonnées sont deux facteurs qui compliquent sa caractérisation structurale, pour laquelle l'utilisation d'une combinaison de méthodes de biologie structurale s'avère être la solution la plus adaptée. A mon arrivée au laboratoire, un travail de prédiction et de caractérisation biochimique de la phosphoprotéine de VSV et de RaV avait été réalisé (Gérard, Ribeiro et al. 2007; Gérard, Ribeiro et al. 2009; Leyrat, Jensen et al. 2011). Les structures atomiques du domaine central d'oligomérisation (Ding, Green et al. 2006; Ivanov, Crepin et al. 2010) et du domaine C-terminal (Mavrakis, McCarthy et al. 2004; Ribeiro, Favier et al. 2008) étaient connues et un travail était en cours pour obtenir un modèle atomique de la P de VSV en combinant les structures atomiques du domaine central et du domaine C-terminal avec des données de

spectroscopie de résonance magnétique nucléaire (RMN) et de diffusion des rayons X aux petits angles (SAXS). Ma contribution dans ce travail a consisté en le clonage, l'expression et la purification du fragment 1 à 105 (P₁₋₁₀₅) désordonné, y compris sous une forme marquée avec des isotopes pour l'étude par RMN. J'ai également déterminé le rayon hydrodynamique (R_h) de plusieurs fragments par chromatographie d'exclusion de taille (SEC) sur la base d'une calibration de la colonne avec des protéines globulaires standards de R_h connu. Ces travaux sont présentés sous la forme d'un article publié en 2012 dans le Journal of Molecular Biology (voir plus loin).

Ce premier travail a donné l'image d'une phosphoprotéine très étendue, dont les dimensions, comparées à celles de la particule virale, indiquent qu'elle pourrait connecter entre eux des tours successifs de la nucléocapside. Lorsqu'elle est fixée sur la nucléocapside par l'intermédiaire de son domaine C-terminal, la phosphoprotéine échantillonne un espace très large autour qui pourrait lui permettre de capter des partenaires ou, en occupant le volume, d'empêcher d'autres molécules d'approcher de la nucléocapside.

Au-delà de la conservation d'une architecture modulaire propre, les phosphoprotéines des *MNV* sont très diverses en taille. Au moment où nous avons entrepris ce travail, la P des virus Nipah et Hendra étaient les plus longues des *MNV*, puisqu'elles comprennent respectivement 709 aa et 707 aa contre 507 pour la P du virus de la rougeole, 568 pour la P du virus Sendai, seulement 265 aa et 297 aa pour les P du VSV et du RaV, respectivement (**Chapitre 1, Figure 17**), alors que la P des virus de la sous-famille des *Pneumovirinae* est encore plus courte ne comportant que 241 aa chez le virus respiratoire syncytial (RSV). Néanmoins, en dépit de cette variabilité dans le nombre d'acides aminés, les prédictions à partir de la séquence indiquent une architecture modulaire comparable pour la P des hénipavirus (Karin, Ferron et al. 2003) (Blocquel, D. et al. 2013 dans ANNEXES). Nous avons entrepris une caractérisation structurale de la protéine NiV P dans le but d'obtenir, comme pour la P de VSV, un modèle atomique.

Lorsque nous avons entrepris ce travail, les informations structurales concernant la P des hénipavirus se limitaient à la démonstration que la partie N-terminale (aa 1 à 406) est globalement désordonnée et à la modélisation par homologie du domaine C-terminal d'interaction avec la partie N_{TAIL} de la nucléoprotéine (Habchi, Mamelli et al. 2010; Habchi, Blangy et al. 2011). Nous avons entrepris une analyse bio-informatique de la séquence en acides aminés. Afin de localiser les limites entre régions structurées et désordonnées, nous avons eu recours au calcul du D-score qui permet d'intégrer les résultats de multiples prédictions de désordre en une méta-prédiction consensus. Cette méthode avait précédemment

permis de déterminer avec précision la localisation des domaines structurés dans les protéines P du virus de Sendai, du VSV et du RaV (Ribeiro, Favier et al. 2008; Gérard, Ribeiro et al. 2009; Ivanov, Crepin et al. 2010). Sur cette base, nous avons cloné, exprimé, purifié et caractérisé le domaine central P_{CD} et le domaine C-terminal P_{XD} de la phosphoprotéine du virus Nipah. Nous avons obtenu des structures cristallographiques pour ces deux domaines ainsi que la structure du domaine P_{XD} de la P de HeV. Afin de construire un modèle atomique de la P entière, nous avons produit et caractérisé par RMN (collaboration avec le groupe du Dr. Martin Blackledge, IBS, Grenoble), par SAXS et par différentes méthodes biophysiques différents fragments d'une centaine d'acides aminés prédits comme intrinsèquement désordonnés ou partiellement structurés, ainsi que des fragments plus longs contenant à la fois des régions désordonnées et structurées et la P entière.

3.1.1. Analyse bioinformatique de la séquence de VSV P

Des alignements de séquence de la P des virus appartenant au genre des Vesiculovirus et des Lyssavirus indiquent l'existence de plusieurs régions hautement conservées séparées par des régions faiblement conservées (Gérard, Ribeiro et al. 2009). Pour la P de VSV le consensus des prédictions de désordre obtenu sous la forme du D-score indique la présence de quatre régions structurées, séparées par des régions flexibles (**Figure 2**). Les frontières prédites entre les régions désordonnées et domaines structurés sont en accord avec la structure cristalline du domaine central de dimérisation (Ding, Green et al. 2006) et ont prédit précisément les frontières du domaine C-terminale de fixation au complexe N-ARN ce qui a permis sa production est caractérisation structurale (Ribeiro, Favier et al. 2008). Le D-score a prédit deux autres domaines structurés dans la partie N-terminale et plus précisément les résidus qui sont impliqués dans l'interaction avec le N⁰ et la polymérase virale, respectivement (Chen, Ogino et al. 2007; Rahmeh, Morin et al. 2012). Néanmoins l'étude du peptide isolé correspondant aux 60 premiers aa du côté N-terminal a démontré son caractère désordonné (Leyrat, Jensen et al. 2011). Le désaccord entre les expériences et les prédictions dans ce cas peut refléter un état désordonné en solution et un repliement induit par l'interaction avec le partenaire moléculaire, et dans le cas de P₆₀ cela serait le N⁰ (Chen, Ogino et al. 2007).

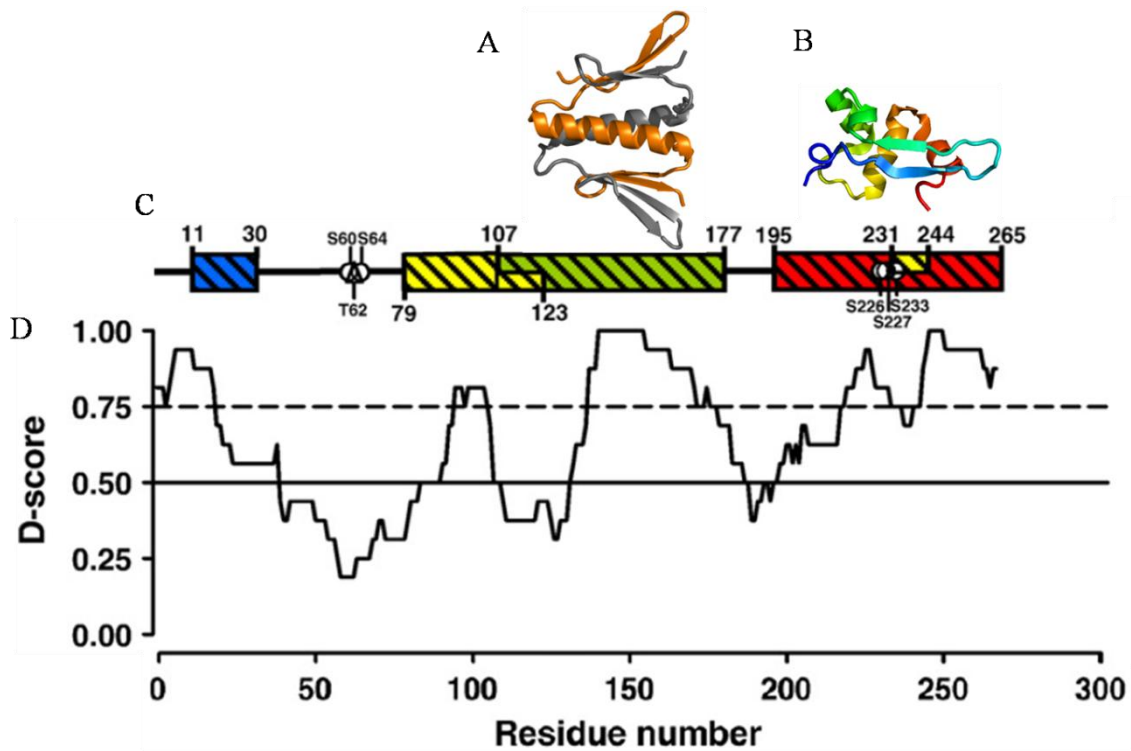


Figure 2: Prédiction de désordre pour la phosphoprotéine (P) du virus de la stomatite vésiculaire (VSV) (figure d'après (Gérard, Ribeiro et al. 2009) avec modifications). **A)** et **B)** Structures cristallines du domaine central (P_{CD}) de dimérisation (PDB code : 2FQM) et domaine C-terminal (P_{CTD}) d'interaction avec N-ARN (PDB code : 2K47), respectivement. **C)** Localisation des régions de fonction connue et notamment résidus 11-30 interaction avec N⁰, 79-107 interaction avec la polymérase virale, P_{CD} 107-177 et P_{CTD} 195-265. **D)** Le graph D-score qui représente la prédiction le désordre en fonction du numéro des résidus. La valeur de 1 correspond à un résidu « structuré » tandis que la valeur de 0 correspond à un résidu « désordonné ».

3.2. Structure de la phosphoprotéine du virus de la stomatite vésiculaire

3.2.1. Article: Ensemble structure of the modular and flexible full-length vesicular stomatitis virus phosphoprotein

Authors: Leyrat C, Schneider R, Ribeiro EA Jr, **Yabukarski F**, Yao M, Gérard FC, Jensen MR, Ruigrok RW, Blackledge M, Jamin M

Publié en 2012 dans *Journal of Molecular Biology*, Volume 423, pages 182-197



Ensemble Structure of the Modular and Flexible Full-Length Vesicular Stomatitis Virus Phosphoprotein

Cédric Leyrat¹, Robert Schneider², Euripedes A. Ribeiro Jr¹, Filip Yabukarski¹, Mingxi Yao², Francine C. A. Gérard¹, Malene Ringkjøbing Jensen², Rob W. H. Ruigrok¹, Martin Blackledge² and Marc Jamin^{1*}

¹Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS, 6 rue Jules Horowitz, BP 181, 38042 Grenoble Cedex 9, France

²Protein Dynamics and Flexibility, Institut de Biologie Structurale, UMR 5075 CEA-CNRS-UJF, 41 rue Jules Horowitz, 38027 Grenoble, France

Received 3 February 2012;
received in revised form
15 June 2012;
accepted 1 July 2012
Available online
9 July 2012

Edited by R. W. Kriwacki

Keywords:

phosphoprotein;
intrinsically disordered
regions;
vesicular stomatitis virus;
NMR;
SAXS

The phosphoprotein (P) is an essential component of the viral replication machinery of non-segmented negative-strand RNA viruses, connecting the viral polymerase to its nucleoprotein–RNA template and acting as a chaperone of the nucleoprotein by preventing nonspecific encapsidation of cellular RNAs. The phosphoprotein of vesicular stomatitis virus (VSV) forms homodimers and possesses a modular organization comprising two stable, well-structured domains concatenated with two intrinsically disordered regions. Here, we used a combination of nuclear magnetic resonance spectroscopy and small-angle X-ray scattering to depict VSV P as an ensemble of continuously exchanging conformers that captures the dynamic character of this protein. We discuss the implications of the dynamics and the large conformational space sampled by VSV P in the assembly and functioning of the viral transcription/replication machinery.

© 2012 Elsevier Ltd. All rights reserved.

Introduction

Vesicular stomatitis virus (VSV) is the prototype virus of the *Rhabdoviridae*, a viral family that also includes rabies virus.¹ The genome of these viruses is made of a single-stranded negative-sense RNA molecule. The family *Rhabdoviridae* is grouped within the order *Mononegavirales* (MNV) with three other families of non-segmented negative-sense

RNA viruses, which also contain important human pathogens, the *Paramyxoviridae* (measles virus, respiratory syncytial virus, and Nipah virus), the *Filoviridae* (Ebola virus and Marburg virus) and the *Bornaviridae* (Borna disease virus). These viruses have different morphologies and different modes of transmission, and they induce different pathologies in their hosts but share similar genomic and structural organizations and similar modes of RNA transcription and replication.² The genome of VSV contains five genes that are common to all MNV viruses and are organized in the same order along the RNA genome. They encode successively from the 3' end to the 5' end of the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the large subunit of the RNA-dependent RNA polymerase (L). The G and M

*Corresponding author. E-mail address: jamin@embl.fr.

Abbreviations used: VSV, vesicular stomatitis virus; SAXS, small-angle X-ray scattering; IDR, intrinsically disordered region; SEC, size-exclusion chromatography; HSQC, heteronuclear single quantum coherence; SSP, secondary structure propensity.

proteins are involved in the entry of the virus in its host cell and in the assembly and budding of new viral particles, respectively. The three other proteins together with the RNA genome constitute the infectious core of the virus and its transcription/replication machinery. The RNA genome is encapsidated by N,³ and the N-RNA complex rather than the naked RNA constitutes the actual template for the viral two-subunit polymerase complex comprising the L and P proteins.⁴ P is an essential component of the virus that plays multiple roles at different stages of the viral life cycle. First, P acts as a non-catalytic cofactor of the viral polymerase.⁵ The L subunit carries out the enzymatic activities required for RNA synthesis, mRNA capping, methylation and poly-adenylation^{6–11} but is unable to bind efficiently to the N-RNA template.¹² P is required for processive RNA synthesis.¹³ Since P possesses binding sites for L in its N-terminal part and for the N-RNA template in its C-terminal domain,¹⁴ it is assumed that P correctly positions L on its template and maintains it attached when the polymerase complex moves along the template during transcription and replication. Second, P forms a complex with nascent N named N⁰-P (the superscript 0 indicates the absence of RNA) that prevents N from polymerizing and from binding to cellular RNA and thus preserves N in a soluble form, available for encapsidating intermediate positive-sense RNA antigenomes and newly synthesized negative-sense RNA genomes.^{15–18}

VSV P forms non-globular homodimers.¹⁹ The protein has a modular organization, which was first suggested by functional studies^{14,20–22} and was recently evidenced by structural studies.^{23–26} The protein contains a long N-terminal intrinsically disordered region (IDR) (IDR_{NT}; amino acids 1–106) and two autonomous folding domains (P_{CED} and P_{CTD}) separated by a disordered linker (amino acids 178–194). The central P_{CED} domain (amino acids 107–177) constitutes the dimerization interface,^{23,25} and the C-terminal P_{CTD} domain (amino acids 195–265) binds to the N-RNA template.^{26,27} P_{CED} was identified by limited proteolysis,²⁴ and its structure was solved by X-ray crystallography.²³ P_{CTD} was located by a meta-prediction of disordered regions,²⁵ and its structure was solved by nuclear magnetic resonance (NMR) spectroscopy.²⁶ The NMR spectrum obtained with a slightly longer construct (amino acids 185–265) showed that the interdomain region separating P_{CED} and P_{CTD} is disordered²⁶ and therefore suggests that, in the homodimer, the two C-terminal domains are tethered to the central dimerization domain through flexible linkers. In addition, the meta-analysis of disorder predicted the presence of a structured region in the 40 N-terminal amino acids,²⁵ which contains the binding site for N⁰.¹⁷ A protein fragment corresponding to the first 60

residues of VSV P (P₆₀) appeared globally disordered in isolation, but two short regions (amino acids 2–12 and amino acids 25–38) transiently adopted α -helical conformations.²⁸ This N-terminal region constitutes a molecular recognition element (MoRE) that undergoes a disorder-to-order transition upon binding to its physiological partner, N⁰.¹⁶ On the basis of these studies, a model in which VSV P consists of a central dimeric core with two long N-terminal flexible arms and two C-terminal globular domains attached through flexible linkers was proposed.^{29,30} Nevertheless, a description at atomic resolution of the behavior of the entire protein in solution is necessary in order to develop a molecular understanding of the roles played by this protein in the process of viral replication.

Unlike folded proteins, intrinsically disordered proteins and IDRs lack a single stable tertiary structure and are thought to exchange between different conformations in stochastic processes driven by thermal fluctuations.^{31,32} These proteins cannot be described by a unique conformation in solution and are better represented by ensembles of molecular conformations.^{33–35} For multidomain proteins, the atomic structure of isolated well-folded domains can be determined by X-ray crystallography or NMR spectroscopy, while information about the disordered regions and the spatial organization of the domains in the entire protein may be obtained by NMR spectroscopy and small-angle scattering experiments. Various NMR observables, including chemical shifts, scalar couplings, ¹⁵N relaxation rates, residual dipolar couplings and paramagnetic relaxation enhancements, provide information about local conformational preferences and long-range order.^{36,37} Small-angle scattering of X-rays or neutrons allows the determination of a limited number of independent parameters that characterize the global size and shape of the molecule in solution^{38,39} and can be used to reconstruct three-dimensional low-resolution models.^{40–44} In combination with the development of computational tools for generating libraries of physically relevant conformations,^{45–48} different methods have been developed over the past few years for building or selecting conformational ensembles on the basis of experimental constraints from NMR,^{45,49–51} small-angle X-ray scattering (SAXS)^{33,47,52,53} or a combination of both methods.^{54–56}

Here, we use NMR spectroscopy and SAXS to characterize the spatial organization of the full-length dimeric VSV phosphoprotein. NMR experiments confirm the existence of folded domains and unfolded regions and, in combination with SAXS data, provide the biophysical basis for an ensemble model of the protein in solution. With this ensemble representation in hand, we discuss the possible roles played by the intrinsic disorder of VSV P in the assembly and functioning of the transcription/replication complex

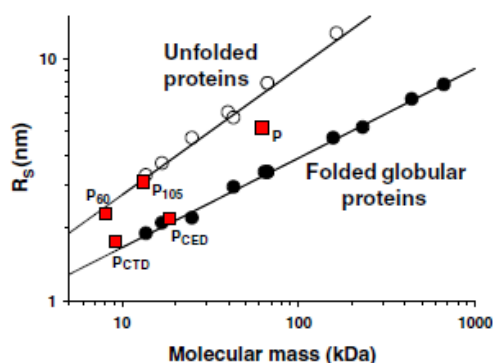


Fig. 1. Hydrodynamic properties of VSV P and its fragments. The plot displays the Stokes' radius (R_s) measured by SEC as a function of the molecular mass of the protein. The red squares are for P and its different fragments. The filled and open circles show data for standard globular proteins in their native or unfolded forms, respectively.⁵⁷

and consider the potential advantages conferred by this high level of molecular flexibility.

Results

Full-length VSV phosphoprotein and its fragments

Full-length VSV P, a short fragment (P_{60}) and a long fragment (P_{105}) of the IDR_{NT} corresponding to the first 60 and 105 residues, P_{CED} and P_{CTD} were produced as described below (see Materials and Methods). Each protein elutes as a single peak from the size-exclusion chromatography (SEC) column. The molecular mass determined from static light scattering is constant throughout the chromatographic peak, and the polydispersity index (M_w/M_n) is lower than 1.01, indicating that each sample is monodisperse.

As previously reported, full-length VSV P elutes from the SEC column as a dimer¹⁹ (Table S1). P_{60} , P_{105} and P_{CTD} are monomeric, whereas P_{CED} forms dimers, confirming that this domain contains the dimerization interface (Table S1). In a plot of hydrodynamic radius (R_s) as a function of molecular mass (Fig. 1), P_{CED} and P_{CTD} behave similar to folded globular proteins, while P_{60} and P_{105} behave similar to unfolded proteins. Full-length P appears at an intermediate position in accordance with the presence of both folded domains and IDRs in similar proportions.

NMR spectroscopy

Full-length VSV P and its domains were investigated by NMR (Fig. 2). The ^1H - ^{15}N two-dimensional

heteronuclear single quantum coherence (HSQC) spectrum of P_{105} exhibits poor chemical shift dispersion of the amide ^1H resonances (Fig. 2b) as is typical for a disordered protein.⁵⁸ Chemical shifts depend on the backbone ϕ/ψ dihedral angles, and in disordered systems, they are highly sensitive to the presence of transient secondary structure,^{50,59-61} commonly expressed in terms of a secondary structure propensity (SSP).^{50,60} The SSP score for P_{105} reveals the presence of two transient α -helices in the N-terminal moiety of the peptide (amino acids 2-16 and amino acids 25-31), in agreement with data previously recorded for P_{60} ²⁸ (Fig. 2a) and with the consensus predictions of disorder (Fig. S1a) and secondary structure²⁵ (Fig. S1b).

The NMR spectra of isolated P_{CED} (amino acids 107-177) and P_{CTD} (amino acids 194-265) confirm that both domains are folded in solution. The HSQC spectra of these fragments show well-dispersed amide ^1H resonances, characteristic of folded proteins⁵⁸ (Fig. 2c and d). For P_{CED} (79 residues including the eight-amino-acid histidine tag), the HSQC spectrum exhibits 68 clearly identifiable signals. For P_{CTD} , the backbone amide and side-chain resonances were assigned in a previous study and used to generate a high-resolution structure of this domain.²⁶

The HSQC spectrum of the full-length P dimer contains two sets of resonances of different intensities. A set of intense signals with limited chemical shift dispersion corresponds well to the spectrum recorded for P_{105} (Fig. 2e), indicating that the N-terminal part is flexible with the same conformational behavior as the isolated peptide and negligible interdomain contacts. Some additional sharp resonances that may correspond to residues in the flexible linker between P_{CED} and P_{CTD} are visible in the spectrum of full-length P. A second set of broader and less intense resonances corresponds to resonances found in the spectra of isolated P_{CED} and P_{CTD} , showing that these two regions remain structured in the context of the full-length protein (Fig. 2f and g). The presence of long IDRs in VSV P exerts a frictional drag that slows down the overall tumbling rate of the molecule^{62,63} and may explain the decrease of signal intensity observed for resonances in the globular domains as compared to the spectra of the isolated domains (Fig. 2c and d).

These results clearly indicate that (1) the N-terminal region of P exchanges between multiple conformers on a fast timescale and contains transient α -helical elements, that (2) P_{CED} and P_{CTD} are folded in the full-length protein and that (3) there are no significant interactions between the different regions of the molecule. These observations form the basis for subsequent ensemble modeling approaches.

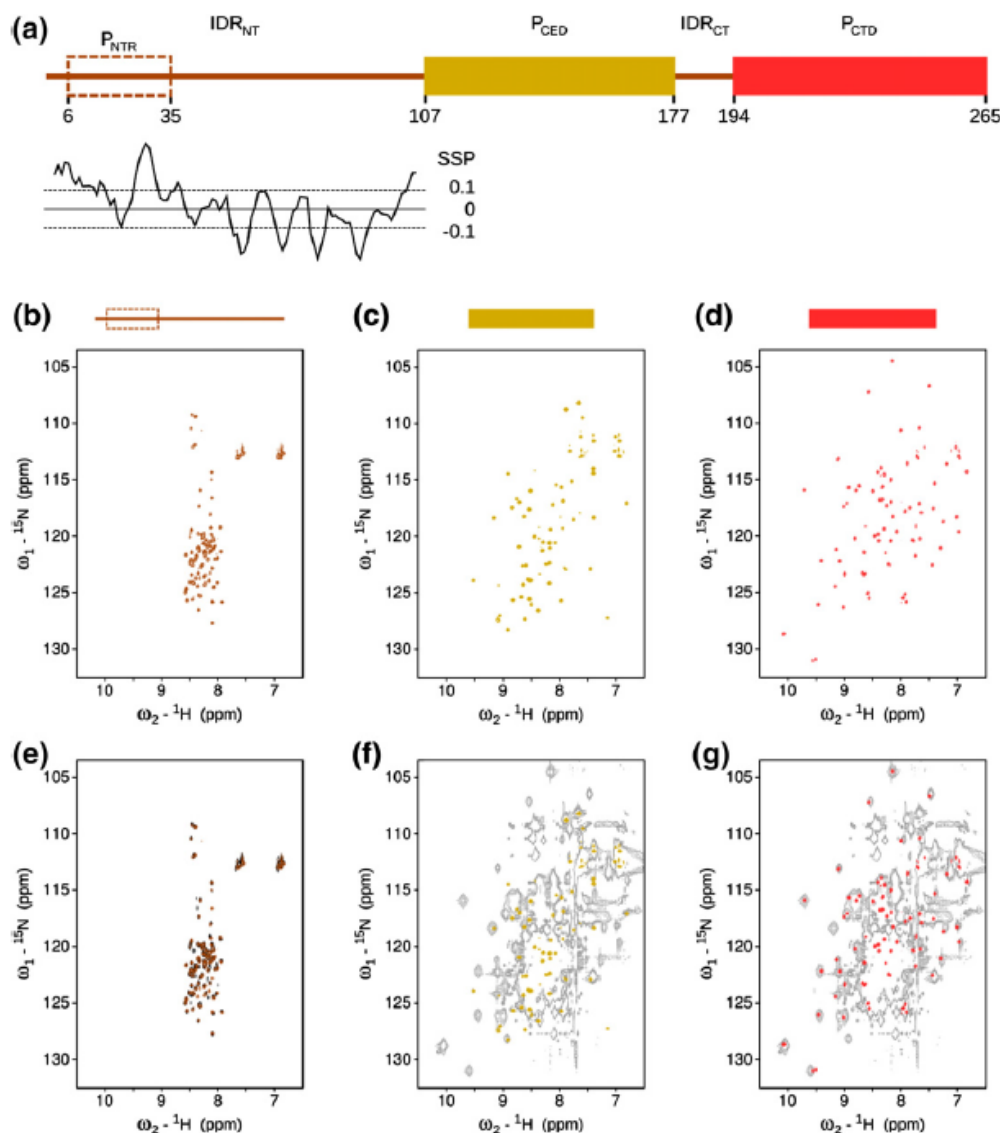


Fig. 2. NMR spectroscopy. (a) Schematic representation of VSV P and SSP score for P_{105} . The scheme shows the location of the folded domains, P_{CED} (orange box) and P_{CTD} (red box), and of the IDRs (in brown), as well as the P_{NTR} region that binds to N^0 (dashed box). The SSP score was calculated from experimental C^α and C^β chemical shifts. (b–d) 1H - ^{15}N HSQC spectra of P_{105} (b), P_{CED} (c) and P_{CTD} (d). NMR data were recorded at 600 MHz and 20 °C and pH 7.5. (e–g) The 1H - ^{15}N HSQC spectrum of full-length P overlaid with the spectra of P_{105} (e), P_{CED} (f) and P_{CTD} (g). The spectrum of full-length P is shown in black at a high contour level (e) or at a lower contour level (in gray) (f and g).

Small-angle X-ray scattering

The shapes of the SAXS curves (Fig. 3a) and of the Guinier plots (Fig. 3b) obtained of full-length P are independent of protein concentration, indicating the absence of significant aggregation. The model-free analysis of small-angle scattering curves provides information about the ensemble averaged dimensions and shape of the scattering particles.⁶⁴ The

radius of gyration, R_g , determined from the Guinier approximation is 5.4 ± 0.1 nm in the range $Q \cdot R_g < 1.3$ and is 5.8 ± 0.4 nm in the range $Q \cdot R_g < 1.0$ (Table S2), in agreement with the value of 5.6 ± 0.1 nm determined from the pair distribution function $[P(r)]$ (Fig. 3c) and with the value of 5.3 ± 0.1 nm measured previously by small-angle neutron scattering.¹⁹ This R_g value is significantly larger than that expected for a globular protein of the same

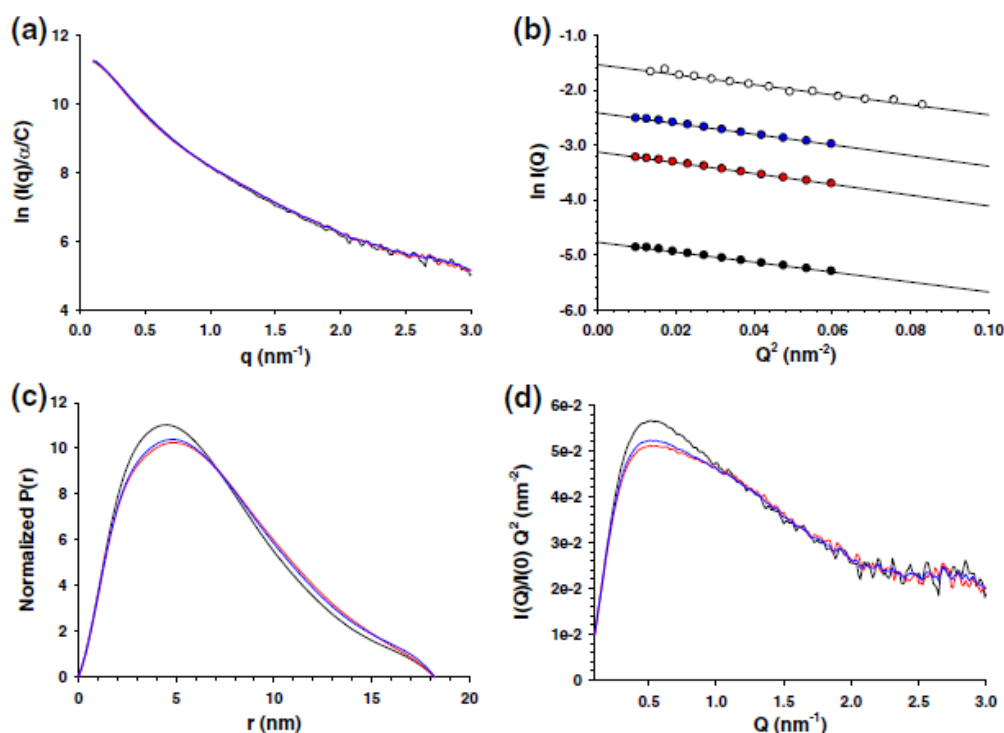


Fig. 3. Small-angle scattering experiments. (a) SAXS curves of VSV P at three concentrations. SAXS data were recorded for Q values in the range $0.1 \text{ nm}^{-1} < Q < 3.0 \text{ nm}^{-1}$. The curves obtained for three protein concentrations (1 mg ml^{-1} , 5.5 mg ml^{-1} and 11 mg ml^{-1}) are represented after correction for concentration by black, red and blue lines, respectively. (b) Guinier plot. Guinier analysis was performed for each concentration and is shown with the same colors as in (a). Open circles show previously reported small-angle neutron scattering data recorded at a concentration of 5 mg ml^{-1} .¹⁹ (c) Normalized pair distribution function $P(r)$. The pair distribution function was calculated with the GNOM program using the datgnom routine. The surface areas under the curves were normalized to account for the differences in protein concentration. (d) Kratky plot. The analysis is shown with the same colors as in (a).

molecular mass, in agreement with the presence of more extended conformations expected for a protein with extensive disorder. For elongated or disordered proteins, the linearity of the Guinier plot is restricted to a lower range of Q values than for globular proteins, and for a random-coil chain, it is recommended to use the Debye plot rather than the Guinier plot because its linearity extends to a larger range of Q values.⁶⁵ For VSV P, the linearity of this plot is restricted to a narrow range of $Q \cdot R_g$ values (< 1.5) and yields a value of 6.0 ± 0.1 (Fig. S2a and Table S2). These two approaches, thus, set limits on the estimation of the R_g value.

The pair distribution function obtained by Fourier transformation of the experimental curve exhibits an asymmetrical shape with a large D_{max} value (Fig. 3c), whereas the Kratky plot presents a maximum at 0.5 nm^{-1} and a flat region above 2 nm^{-1} (Fig. 3d), also supporting the presence of folded domains and disordered regions. However, despite the presence of several folded domains, both curves are smooth, lacking interdomain correlation

peaks as would be expected for a multidomain protein in which the folded domains are positioned at fixed distances from each other. The differences observed in the pair distribution functions (Fig. 3c) and Kratky plots (Fig. 3d) for the three protein concentrations arise from minor variations within the limits of experimental errors (Fig. S2b).

Bead models generated with the program DAMMIN³⁹ by *ab initio* reconstruction from the SAXS data appear as elongated cylinders (Fig. S3a and b) that could accommodate the structures of one P_{CED} dimer and two P_{CTD} domains (Fig. S3c). Models of similar shape were obtained with the program BUNCH⁶⁶ that combines atomic structures for P_{CED} and P_{CTD} and bead models for the unfolded strands (Fig. S3d and e). Such models may be considered as population-weighted representations of a potentially highly heterogeneous conformational ensemble. However, these programs, designed to model compact structures, are constrained by the large D_{max} value; therefore, modeling of highly disordered molecules such as VSV P as a single

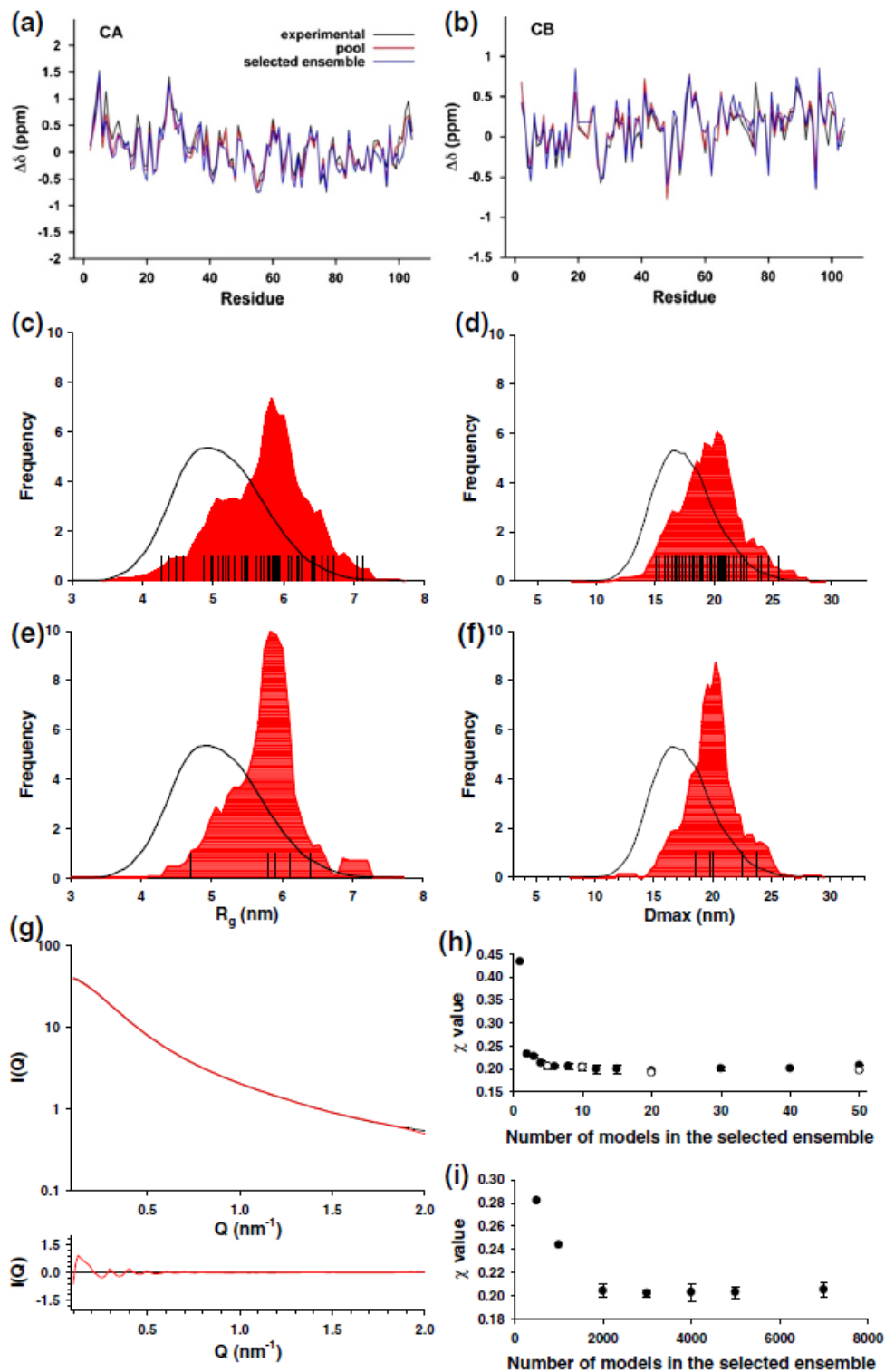


Fig. 4 (legend on next page)

conformer is biased toward elongated shapes that may not necessarily reflect the actual structural organization of the system.^{33,67} It should be noted, however, that modeling with DAMMIN or BUNCH yielded good fits to the scattering data with χ values of 0.056 and 0.293, respectively. Therefore, it must be stressed that the goodness of fit of the SAXS data cannot be used as the only criterion for discriminating between different possible structural models.⁴⁸

Modeling VSV P as an ensemble of conformations

To account for the mobility of the polypeptide chain observed by NMR spectroscopy, we modeled the structure of the P dimer as an ensemble of conformers rather than as a single structure. An ensemble of conformations of the VSV P dimer was built with the program flexible-meccano.^{46,68} This program generates physically relevant conformers of disordered proteins and uses a simple exclusion procedure to avoid clashes within the protein. For modeling the IDR_{NT} region of P, we used a database of backbone dihedral angles compiled by the ensemble selection algorithm ASTEROIDS⁵⁰ in order to match predicted ¹³C^α, ¹³C^β and ¹³C' chemical shifts for this region of P (calculated using the program SPARTA⁶⁹) with experimental values measured on P₁₀₅ (Fig. 4a and b; see Materials and Methods). The dimeric P_{CED} and the two P_{CTD} domains were constructed in their folded forms. In this way, an initial ensemble of 8000 conformers was built with flexible-meccano. This ensemble exhibits broad Gaussian distributions of R_g and D_{max} values as expected for a polypeptide chain obeying random-coil statistics (Fig. 4c–f); however, the average SAXS curve calculated for the 8000 models using CRYSOLOG⁷⁰ does not fit the experimental curve ($\chi=0.75$) (Fig. S4).

From this initial pool, sub-ensembles containing different numbers of conformers were selected with

the program GAJOE (Genetic Algorithm Judging Optimization of Ensembles) for which the average theoretical curve reproduces the experimental SAXS curve up to $Q=2.0 \text{ nm}^{-1}$ (Fig. 4g and Fig. S4).⁵³ The χ value⁶⁶ [Eq. (2)] used here to evaluate the quality of the fit decreases with increasing number of models in the selected ensemble and levels off for selected ensembles of more than five models (Fig. 4h). This result clearly shows that the SAXS curve is better represented by an ensemble of conformers than by a single conformer. Also, similar fits were obtained when the initial ensemble was reduced to 2000 conformers, suggesting that 8000 conformers are sufficient to reproduce the conformational diversity of the molecule at the resolution level accessible from the SAXS curve (Fig. 4i).

Selected ensembles of 50 conformers adequately reproduce the SAXS curve (Fig. 4g) and the NMR chemical shifts for the 105 N-terminal residues (Fig. 4a and b), providing an atomic representation of VSV P in agreement with both sets of experimental data. The different conformers obtained in independent selection processes exhibit a broad range of R_g and D_{max} values, and independent rounds of selection yielded similar results (Fig. S5 and Table S3). Figure 4c and d shows the distributions of R_g and D_{max} for a representative ensemble of 50 conformers, and Fig. 4e and f shows the distributions of R_g and D_{max} for a representative minimal ensemble of 5 conformers. The averaged R_g and D_{max} values of the selected ensembles are within the limits of values determined from the experimental curves (Table S3) but are significantly larger than those of the initial ensemble (mean R_g value of 5.1 nm with a standard deviation of 0.6 nm and mean D_{max} value of 17 nm with a standard deviation of 3 nm). The distribution functions of the selected ensembles are shifted to higher R_g and D_{max} values as compared to those of the initial ensembles (Fig. 4b–e). This suggests that P is more extended

Fig. 4. Modeling of VSV P as an ensemble of conformers with GAJOE. (a) Experimental and calculated C^α chemical shifts for VSV P ensembles. The panel shows C^α secondary chemical shifts. The black line shows experimental chemical shifts for P₁₀₅, and the red and blue lines show chemical shifts calculated using SPARTA for IDR_{NT} residues of the initial 8000-strong full-length VSV P ensemble (red) and of a selected ensemble of 50 conformers (blue). (b) Experimental and calculated C^β chemical shifts for VSV P ensembles. (c) R_g distributions for an ensemble of 50 conformers. The black curve shows the R_g distribution calculated for the initial pool of 8000 conformers. The red curve shows the R_g distribution for the selected ensemble of 50 conformers that fits the experimental SAXS data. Typically, the selection process performed by GAJOE involves 50 successive cycles, and the distribution shown in red corresponds to the distribution of R_g values for the models selected in the best ensemble of each of these successive selections. The black bars show the R_g values for the 50 conformers that form the best of all of these ensembles. (d) D_{max} distribution for an ensemble of 50 conformers. The same color scheme as in (a) is used for the D_{max} distributions. (e) R_g distributions for an ensemble of five conformers. (f) D_{max} distribution for an ensemble of five conformers. (g) Fitting of SAXS data. The experimental SAXS curve is shown in black, and the fitted line for a selected ensemble of 50 conformers is shown in red. The χ value is 0.207. The lower panel shows the residuals of the fit. (h) Variation of the goodness of the fit (χ) with the number of models in the selected ensemble. The selection with GAJOE was repeated for ensembles containing 1–50 conformers. The black circles are for selections performed without allowing multiple selections of the same conformer, whereas white circles are for selections allowing multiple selections. (i) Variation of the goodness of the fit (χ) with the number of models in the initial ensemble. χ values are for selection of ensembles of five conformers without allowing multiple selections of the same conformer.

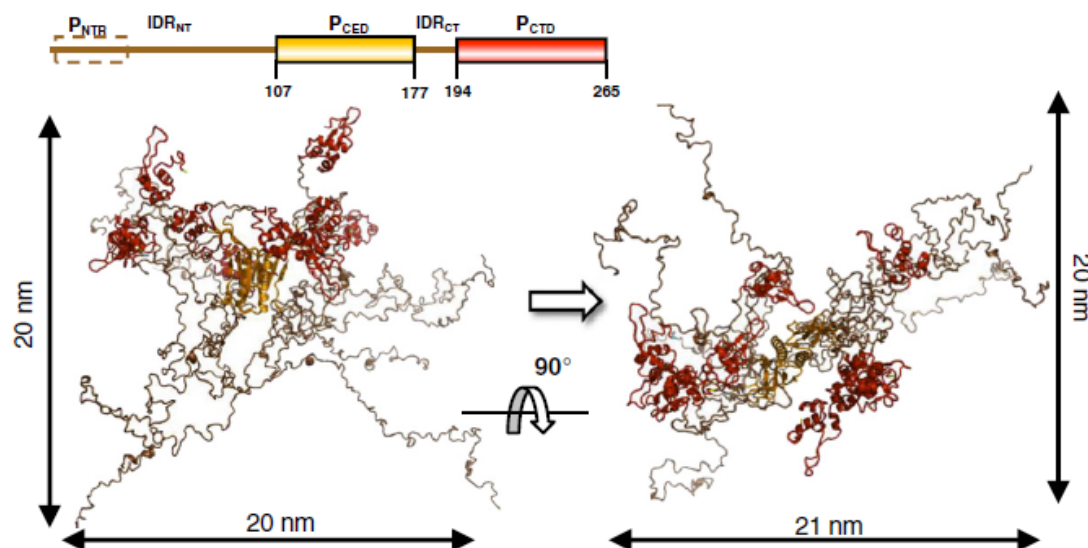


Fig. 5. Ensemble representation of VSV P. Representative ensemble of the five conformers selected by GAJOE. The conformers were aligned by superimposing their individual P_{CED} . The superposition is shown in two orientations. P_{CED} is shown in orange, P_{CTD} is shown in red and the flexible IDRs are shown in brown.

than expected based on pure random-coil sampling of the conformational space assigned to the flexible regions. The IDR_{NT} is rich in acidic residues; thus, charge repulsion most likely accounts for this more extended organization.⁷¹ The alignment of the different models from one selected ensemble obtained by superimposing their P_{CED} reveals that the P dimer samples a large conformational space in solution (Fig. 5), and the comparison of different ensembles indicates that this behavior is independent of the number of models in the selected ensemble (Figs. S5 and S6).

The next questions are to test whether a discrete number of conformers are sufficient for representing the conformational diversity of such a flexible molecule and whether fitting the experimental SAXS curve with an ensemble of conformers allows one to discriminate between different models of VSV P. To answer these questions, we performed simulations with reference ensembles. In this approach, a set of conformers is chosen from the first pool of 8000 conformers to constitute the reference ensemble, and a synthetic scattering profile is calculated for this reference ensemble using CRY SOL.⁷⁰ Then, the program GAJOE is used to select a subset of 1–50 conformers from the initial pool of conformers that minimize the χ value between the scattering curves calculated for this selected subset and those calculated for the reference ensemble (Fig. S7). In the first case considered, the reference ensemble consisted of a single conformer arbitrarily chosen from the initial pool of models. On the basis of its theoretical scattering curve, GAJOE is

able to correctly retrieve this same conformer from the initial pool (selected subset of one model). Attempting to represent this unique conformer with an ensemble of several conformers decreases the quality of the fit, as judged by the χ value, with increasing numbers of conformers. When multiple selection of the same model is allowed, the program GAJOE selects the same model several times for ensembles containing up to five models. This clearly indicates that the system is best represented by a unique conformer. In the second case, the reference ensemble consisted of a set of 10 different conformers. Here, the quality of the fit significantly improves with increasing number of models in the selected ensemble up to ensembles of 10 conformers. In a selected ensemble of 20 structures, none of the models selected by GAJOE corresponds to the models constituting the reference ensemble, and the lowest r.m.s.d. value between conformers of the selected and reference ensembles is 2.3 nm (r.m.s.d. values ranging between 2.3 and 5.9 nm) (Fig. S8). Thus, in this case, the conformational diversity of the reference system is best represented by an ensemble of conformers. In the third case, the reference ensemble consisted of 8000 structures. Again, the quality of the fit improves with increasing number of models in the selected ensemble but, as in the second example, exhibits little improvement above 10 conformers. Thus, in line with previous comparisons applied to completely disordered systems and multidomain proteins,^{33,53} these simulations demonstrate that, in the case of proteins containing folded and disordered domains, selection using

GAJOE clearly differentiates a situation where the protein adopts a unique structure from one where it is averaged over multiple conformers. However, because of the limited information content of the SAXS curve, discrete ensembles of 10 or 20 conformers appear to be sufficient for describing such a system of large conformational diversity.

Discussion

Characterizing the overall structure of multi-domain proteins that contain folded domains concatenated with long IDRs is complicated by the intrinsic conformational heterogeneity of the disordered segments. The size of the VSV P particle determined by SEC and SAXS indicates a non-globular character suggesting either an elongated molecule or an ensemble of rapidly interconverting conformers. The smooth profiles of the pair distance distribution function and Kratky plot derived from the SAXS data (Fig. 3c and d) could be explained either by a physically unrealistic organization in which the folded domains adopt fixed positions in space or, more likely, by conformational averaging. However, definite evidence of the conformational heterogeneity of VSV P emerged from NMR spectroscopy. The NMR data clearly confirm the presence of disordered regions in the full-length protein and show that the disordered and ordered regions behave independently. NMR is particularly suitable for studying the dynamic behavior of disordered polypeptide chains, providing amino-acid-specific identification of transient structural elements, motional correlations and long-range contacts in globally disordered proteins.^{37,49,72} For VSV P, NMR clearly shows that, in the context of the full-length P dimer, the N-terminal region of each P monomer interconverts between multiple conformations on a timescale faster than the millisecond range, while P_{CED} and P_{CTD} adopt the stable, folded structures that were determined from isolated domains. Although the slow tumbling of the full-length dimeric protein leads to resonance broadening and low signal intensity for the folded domains, nevertheless, we were able to exclude significant interactions between unfolded and folded domains. Only one set of resonances is observed in the NMR spectra, demonstrating that the chains of the VSV P dimer behave identically. We have incorporated this information into an ensemble description of the molecular behavior of VSV P that conveys the dynamic character of the molecule and its departure from pure random-coil behavior. The different models of the selected ensemble represent different possible conformations of VSV P that collectively reproduce the experimental scattering curve, as well as the local conformational behavior and chemical shifts of the N-terminal disordered region as

identified from NMR. In the selected ensembles, P dimers adopt a large variety of conformations with different dimensions and relative orientations between the N- and C-terminal arms. It should also be noted that there is not a unique subset of conformers that reproduce both SAXS and NMR data. The conformers of one selected ensemble are, thus, merely examples of the structure of the protein that are in agreement together with the experimental data.

Comparison with the phosphoprotein of other MNV viruses

VSV P shares a similar modular organization not only with the phosphoprotein of rabies virus, another member of the *Rhabdoviridae*,^{19,25} but also with the phosphoprotein of different viruses from the *Paramyxoviridae*^{73–77} and the *Bornaviridae*.⁷⁸ In all these proteins, long IDRs alternate with structured domains, and similar regions of the protein carry out similar functions during the replication cycle. The N-terminal region is involved in binding RNA-free N⁰, the central region of the protein that contains a homo-oligomerization domain, while the C-terminus binds the N–RNA template. All these proteins also interact with the L subunit of the RNA-dependent RNA polymerase, although the site of interaction is not clearly mapped, and specific functions or post-translational modifications have been located either in the IDRs or in the folded domains. Thus, the model proposed here for VSV P, in which N-terminal and C-terminal flexible extensions are tethered to a central homo-oligomeric core, may also apply to the P proteins from these other MNV viruses. Although evidence that the P protein contains disordered and ordered regions have been gathered for different viruses, this study provides the first structural description of a full-length oligomeric phosphoprotein obtained on the basis of experimental data.

Importance of size and flexibility of VSV P for its functions

VSV P plays multiple roles and interacts with different viral and cellular partners throughout the viral life cycle. Here, we endorse the view that the remarkable dimensions and flexibility of the protein dimer have important implications in the assembly of the virus and the functioning of the replication machinery. By sampling a large volume space, the VSV P dimer may increase the probability of encountering its partners in solution and, therefore, increase the on-rate for binding as proposed in the “fly casting” mechanism^{79,80} (Fig. 6). It is noticeable that the long N-terminal flexible arm of VSV P contains binding sites for multiple partners, including phosphorylation sites, a nuclear export signal, a

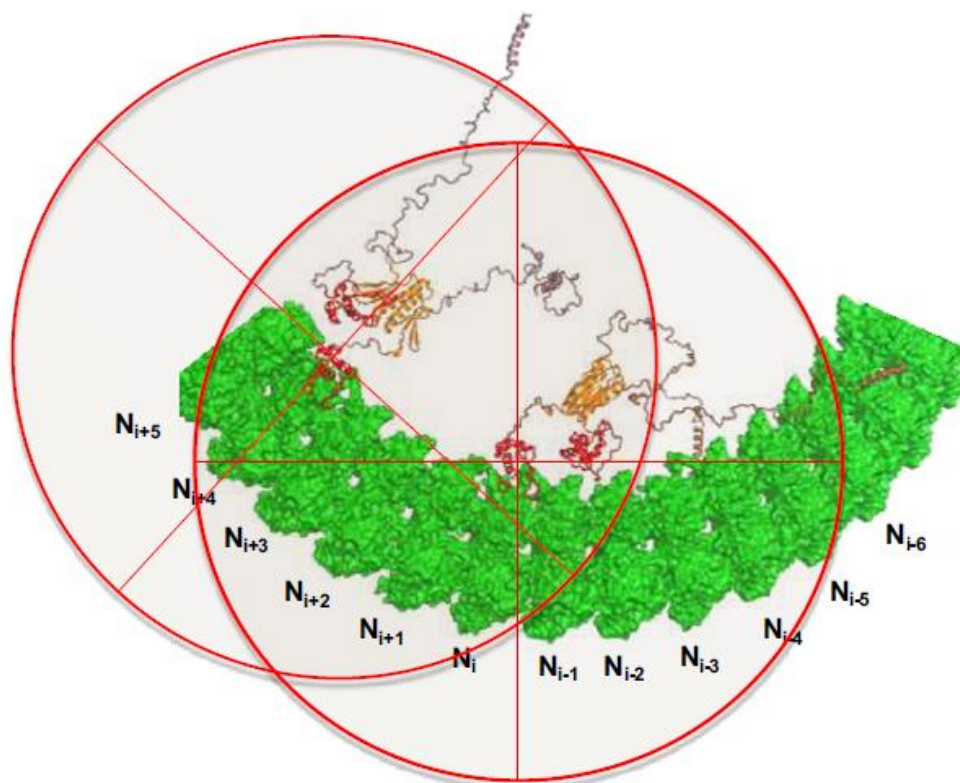


Fig. 6. Radius of capture of P attached to the N-RNA template. The figure represents a model generated by aligning the N-RNA template structure fitted into the cryo-electron microscopy density,⁸¹ the crystal structure of P_{CTD} bound to N-RNA⁸² and a representative model of VSV P (this study). The polymeric N-RNA complex is shown as a surface in green, while P molecules are shown as cartoons with the IDRs in brown, P_{CED} in orange and P_{CTD} in red. Arbitrarily, we chose to show the model with only one P_{CTD} per P dimer bound to the template. The circle representing the space sampled by the N-terminal region around the attachment point has a radius of about 17 nm corresponding to the experimentally determined D_{max} (Fig. 3c).

MoRE for binding N⁰ and a binding site for the L protein. A main function of P is to position the polymerase on the N-RNA complex and to keep it attached when the polymerase moves along the template, in particular, when it pauses at the intergenic regions during transcription. When P is bound to the N-RNA template in the infected cell, the N-terminal arms extend into solution and may serve to catch L and position it onto the template. Also, it has been proposed that P moves by cartwheeling along the N-RNA template, carrying the L subunit.^{83,84} This model was proposed on the basis of results obtained with Sendai virus (*Paramyxoviridae*) in which oligomerization of the protein is indispensable.⁸⁴ In contrast, the homodimeric domain of rabies virus P is dispensable for transcription,⁸⁵ and we propose an alternative mechanism for the progression of the polymerase of rhabdoviruses along its template. In this new model, P remains attached to the N-RNA template through its P_{CTD}. Currently, it is not clear whether

both P_{CTD} domains of a P dimer can bind simultaneously to the N-RNA complex, but our ensemble model shows that it is physically possible. In the 8000 model ensembles, which gives an idea of the physically accessible conformations of this protein, P_{CTD} can project up to 9 nm away from P_{CED}, and thus, it is possible that both P_{CTD} domains bind simultaneously to neighboring sites of the N-RNA complex.

By sampling a large conformational space around their attachment points on the N-RNA template, P dimers may act similar to polymer brushes⁸⁶ by entropic exclusion,⁸⁶ as demonstrated for disordered regions of nucleoporins, neurofilaments or microtubule-associated proteins.⁸⁷⁻⁸⁹ Once a P molecule is bound to the N-RNA complex through its C-terminal domains, its thermally driven motions may restrict access to the space around this molecule to other P molecules and thereby regulate the spacing between P molecules along the N-RNA complex (Fig. 7a). Entropic brushes have been

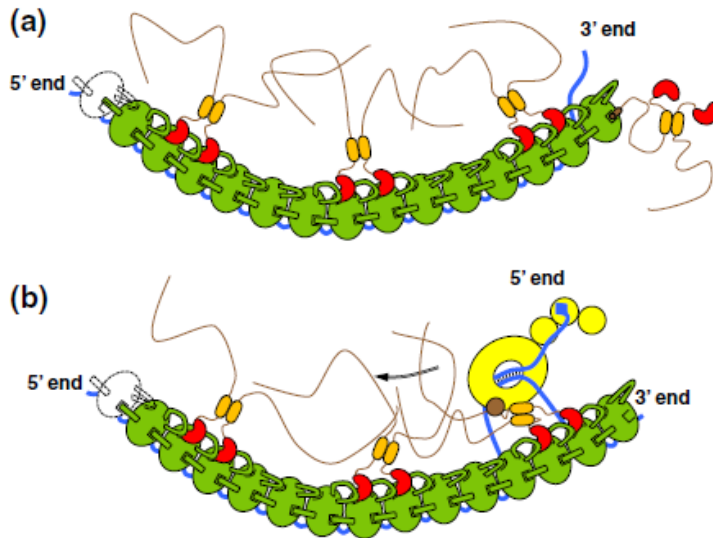


Fig. 7. A possible mechanism for the role of P in the initiation and processivity of the viral polymerase. (a) Initiation process. In the infected cell, the nucleoprotein (in green) coats the RNA (in blue), forming a long N-RNA complex that serves as template for the polymerase complex. The phosphoprotein (P) forms dimers that bind at regular intervals along the N-RNA complex, with about one dimer of P bound to every five N protomers. The N^0 -binding MoRE of a P molecule (shown as a short cylinder at the N-terminus of P) binds at the 3' extremity of the genome¹⁶ and triggers local RNA release from N to initiate RNA synthesis, allowing access of the polymerase to the viral RNA. (b) Possible role of P in the processivity

of the polymerase. The polymerase complex composed of the L subunit (in yellow) and of a dimer of P binds to the RNA template through P and initiates RNA synthesis. The polymerase complex moves along the template by jumping between adjacent P molecules. The P molecule bound downstream of the RNA polymerase complex can flex and catch the polymerase. Then, the polymerase moves along the template as it synthesizes RNA, and the flexibility of P allows the bound P molecule to remain attached.

involved in controlling the spacing between neurofilaments or microtubules.^{87,89} The estimated number of P and N proteins in the VSV virion results in one P dimer bound to every five N protomers.⁹⁰

If P molecules are distributed at regular intervals along the template, the capture radii of the N-terminal arms of adjacent P molecules overlap (Fig. 6), and we propose that the polymerase moves along the template by jumping from one P to the next (Fig. 7b). Also, in the generated ensemble of 8000 conformers, the distance separating the center of mass of P_{CTD} from that of IDR_{NT} in dimeric P can extend up to 18 nm, indicating that once P is bound to the N-RNA complex through its C-terminal domain, the search volume accessible to the N-terminal extremity of each monomer extends to sites located further than five N protomers from the binding site (Fig. 6). The long N-terminal arm of P could fetch the polymerase upstream, keep it attached to the template during the time it replicates a stretch of RNA encapsidated by roughly five N protomers (~45 nt), and then deliver it downstream to the next P molecule. A mechanism of folding upon binding may provide a means of specific recognition of P by L without the corollary of high affinity^{91,92} and may allow the L protein to exchange between bound P proteins. It now remains to be determined if the L protein possesses a single binding site for P that would imply that the incoming N-terminal arm exchanges with the previously bound one or if it possesses multiple binding sites that would allow the simultaneous attachment of multiple P N-terminal arms.

The dimensions and plasticity of P may also serve in the assembly of new N-RNA complexes. The encapsidation of the newly synthesized RNA genome requires the delivery of soluble RNA-free N^0 to the site of RNA synthesis. The N^0 -P complex can bind to the nucleocapsid,⁵³ but little is known about the reaction by which N^0 is transferred from P to the RNA. When an N^0 -P complex is bound to the N-RNA template through the C-terminal domain of P, the flexibility of the P molecule may correctly position and orient the N^0 molecule within the replication complex and deliver it to the site of encapsidation.

Finally, it is worth noting that the dimensions of the phosphoprotein are remarkable as compared to the dimensions of the virion. The VSV particle has a bullet shape of about 200 nm in length and 70 nm in diameter.⁸¹ In this structure, the N-RNA complex curls into a helix that constitutes a tubular inner leaflet and defines an internal cavity of length 150 nm and a radius of 15 nm. The average D_{max} of the P dimer in solution of 18 nm is longer than the radius of this inner cavity and about one-eighth of its length. However, in accordance with the partial specific volume of proteins, this cavity is large enough to accommodate 225 molecules of P and 55 molecules of L that are present in the virion, as well as 500 M proteins that are also present in virion but are not involved in the intermediate shell of the particle.^{90,94} It is not known how these proteins are packed within the cavity, but because the C-terminal domain of N that binds the C-terminal domain of P is oriented toward the internal cavity, it is possible

that P is bound to the N-RNA template. Also, since the rise of the N-RNA helix is 5.1 nm,⁸¹ P_{CTD} domains of one P dimer could bind to successive rungs of the N-RNA helix and stabilize the virion.

Materials and Methods

Sample preparation

VSV full-length P, P₆₀ and P_{CTD} were produced and purified as previously described,^{19,26,28} and similar procedures were set up for purifying P₁₀₅ and P_{CED}. The homogeneity of each sample was checked by SEC combined with detection by multi-angle laser light scattering and refractometry.^{19,95} SEC was performed with an S200 Superdex column (GE Healthcare) equilibrated with 20 mM Tris-HCl at pH 7.5 containing 150 mM NaCl. Separation was performed at 20 °C with a flow rate of 0.5 ml min⁻¹. We injected 50 µl of a protein solution at a concentration of 5.5 mg ml⁻¹. Online multi-angle laser light scattering detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting at 690 nm. Data were analyzed, and absolute molecular weights (*M_w*) were calculated using the ASTRA software (Wyatt Technology Corp.) as described previously by Gérard *et al.*¹⁹ The S200 Superdex column was calibrated with proteins of known Stokes' radii (*R_S*):⁵⁷ chymotrypsinogen (*R_S*=2.3 nm), RNase A (*R_S*=1.9 nm), ovalbumin (*R_S*=3.0 nm), albumin (*R_S*=3.4 nm), aldolase (*R_S*=4.7 nm), catalase (*R_S*=5.2 nm), ferritin (*R_S*=6.8 nm) and thyroglobulin (*R_S*=7.9 nm).

NMR spectroscopy

NMR experiments were performed on Varian spectrometers operating at ¹H frequencies of 600 and 800 MHz. ¹H-¹⁵N HSQC spectra of VSV P₁₀₅ and full-length P shown in Fig. 2 were recorded on samples in 20 mM Tris-HCl, 150 mM NaCl, 50 mM Glu and 50 mM Arg with 10% D₂O at 20 °C and pH 7.5 at a 600-MHz ¹H resonance frequency. Resonance assignment of VSV P₁₀₅ was carried out using a double-labeled (¹³C, ¹⁵N) sample of the peptide with the buffer adjusted to pH 6.0 to avoid loss of resonances due to fast amide proton exchange with the solvent protons. Assignment experiments were carried out at 25 °C and a ¹H resonance frequency of 600 MHz. The assignment was obtained from a series of BEST-type triple-resonance experiments⁹⁶: HNCO, intrasidue HN(CA)CO, HN(CO)CA, intrasidue HNCA, HN(COCA)CB and intrasidue HN(CA)CB. The ¹H-¹⁵N HSQC spectrum of the central P_{CED} domain was recorded at 20 °C and pH 7.5 at a ¹H resonance frequency of 600 MHz, using a buffer containing 20 mM HEPES, 150 mM NaCl and 1 mM DTT. The spectrum of the C-terminal P_{CTD} domain in Fig. 2 was recorded at a ¹H resonance frequency of 800 MHz, at 20 °C and pH 7.5 in 20 mM Tris-HCl buffer containing 150 mM NaCl. All spectra were processed with NMRPipe⁹⁷ and analyzed using SPARKY.⁹⁸ Automatic assignment of P₁₀₅ resonances on the basis of SPARKY peak lists was performed using the program Mars.⁹⁹

SAXS experiments

The scattered intensity of full-length VSV P was collected for scattering vector values ($Q=4\pi(\sin \theta)/\lambda$) ranging from 0.1 to 3.0 nm⁻¹ over a concentration range of 1–11 mg ml⁻¹. SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamline ID14-3. The sample-to-detector distance was 1 m, and the wavelength of the X-rays was 0.0995 nm. Samples were contained in a 1.9-mm-wide quartz capillary. The time of exposure was optimized for reducing radiation damage. Data acquisition was performed at 20 °C.

Data reduction was performed using the established procedure available at ID14-3, and buffer background runs were subtracted from sample runs. The radius of gyration and forward intensity at zero angle [*I*(0)] were determined with the program PRIMUS¹⁰⁰ according to the Guinier approximation at low *Q* values, in a *Q*·*R_g* range up to 1.3:

$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3} \quad (1)$$

The forward scattering intensity was calibrated using bovine serum albumin and lysozyme as references. The radius of gyration and pairwise distance distribution function, *P*(*r*), were calculated with the program GNOM.¹⁰¹ The maximum dimension (*D_{max}*) value was adjusted such that the *R_g* value obtained from GNOM agreed with that obtained from the Guinier analysis.

Modeling of VSV P

Modeling VSV P as a single conformer

The program DAMMIN was used to generate *ab initio* low-resolution models filled with spheres (dummy atoms) by fitting the calculated scattering curve to the experimental curve.^{39,66} This program uses a simulated annealing minimization procedure to find the optimum positions of the dummy atoms, starting from a spherical volume of defined radius. By imposing connectivity constraints, DAMMIN minimizes the interfacial area between the molecule and the solvent and therefore favors compact models. We performed 20 independent runs of DAMMIN with no symmetry restriction.

The program BUNCH⁶⁶ was used to perform rigid-body modeling of the full-length protein using the known structures of VSV P_{CED} and P_{CTD}. The regions of the protein for which no atomic structure was available were modeled by dummy residues. By a simulated annealing protocol, the optimal position and orientation of the folded domains and the conformation of the flexible parts were obtained by fitting the calculated scattering curve to the experimental curve. We performed 16 independent runs of BUNCH with a 2-fold symmetry restriction.

Modeling VSV P as an ensemble

An ensemble of 8000 full-length VSV P conformers was generated with the program flexible-meccano.⁴⁶ The models consisted of P dimers and included the high-resolution three-dimensional structures of the central

and C-terminal domains.^{23,26} To accurately model the disordered N-terminal region, we first applied the ASTEROIDS selection algorithm⁵⁰ to the P₁₀₅ fragment in order to obtain an ensemble of conformers of this region in agreement with experimental ¹³C^α, ¹³C^β and ¹³C' NMR chemical shifts. Briefly, in successive rounds of selection, 5×200 pairs of (φ, ψ) dihedral angles were selected for each residue from a pool of 10,000 P₁₀₅ conformers assembled by flexible-meccano such that the difference between ¹³C chemical shifts predicted by the program SPARTA⁶⁹ and corresponding experimental values was minimized. The selected dihedral angles were then used in generation of a new set of 10,000 conformers for the next round of selection. The (φ, ψ) database thus obtained after four rounds of selection was then used by flexible-meccano to build the IDR_{NT} of full-length VSV P. Residues in the linker region between P_{CED} and P_{CTD} were stochastically assigned amino-acid-specific backbone dihedral angles in agreement with random-coil statistics. In the disordered regions, side chains were constructed using the program SCCOMP.¹⁰² Scattering curves were calculated with CRY SOL,⁷⁰ and selections were performed with the program GAJOE.⁵³

We selected optimized ensembles comprising 1–50 conformers that minimize the discrepancy between the experimental and calculated curves according to the following equation:

$$x^2 = \frac{1}{K-1} \sum_{j=1}^K \left[\frac{\mu I(Q_j) - I_{exp}(Q_j)}{\sigma(Q_j)} \right]^2 \quad (2)$$

where K is the number of points in the experimental curve, σ is the standard deviation and μ is a scaling factor.

Both for the initial ensemble of 8000 conformers and for selected ensembles of 50 conformers, NMR chemical shifts were back-calculated using the program SPARTA⁶⁹ that was already used in derivation of the backbone dihedral angle database for IDR_{NT} of P.

Acknowledgements

This work was supported by grants from the French National Research Agency (ANR) [ANR-07-001-01 (ANRAGE) and ANR JCJC 2010 Protein Disorder (M.R.J.)], the FINOVI foundation and Lyonbiopôle. We thank the Fédération TGIR RMN Très Hauts Champs for providing access to the 800-MHz NMR spectrometer in Grenoble. We thank Dr. G. Zaccai (Institut Laue-Langevin, Grenoble) for helpful discussions. R.S. was supported by the Human Frontier Science Program (long-term fellowship LT000322/2011-L). E.A.R. was supported by postdoctoral fellowships from both ANR and FINOVI programs. C.L. and F.G. were supported by MENRT fellowships from the French government. We thank the Partnership for Structural Biology for the excellent structural biology environment.

Supplementary Data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2012.07.003>

References

- Lamb, R. A. (2007). Mononegavirales. In *Fields Virology* (Knipe, D. M. & Howley, P. M., eds), 5th Ed, 1, pp. 1357–1361. Lippincott Williams & Wilkins, Philadelphia, PA.
- Lyles, D. S. & Rupprecht, C. E. (2007). Rhabdoviridae. In *Fields Virology* (Knipe, D. M. & Howley, P. M., eds), 5th Ed, 1, pp. 1363–1408. Lippincott Williams & Wilkins, Philadelphia, PA.
- Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. (2006). Structure of the vesicular stomatitis virus nucleoprotein–RNA complex. *Science*, **313**, 357–360.
- Arnheiter, H., Davis, N. L., Wertz, G., Schubert, M. & Lazzarini, R. A. (1985). Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell*, **41**, 259–267.
- Emerson, S. U. & Yu, Y. (1975). Both NS and L proteins are required for *in vitro* RNA synthesis by vesicular stomatitis virus. *J. Virol.* **15**, 1348–1356.
- Whelan, S. P., Barr, J. N. & Wertz, G. W. (2004). Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr. Top. Microbiol. Immunol.* **283**, 61–119.
- Li, J., Wang, J. T. & Whelan, S. P. (2006). A unique strategy for mRNA cap methylation used by vesicular stomatitis virus. *Proc. Natl Acad. Sci. USA*, **103**, 8493–8498.
- Moyer, S. A. & Banerjee, A. K. (1975). Messenger RNA species synthesized *in vitro* by the virion-associated RNA polymerase of vesicular stomatitis virus. *Cell*, **4**, 37–43.
- Abraham, G., Rhodes, D. P. & Banerjee, A. K. (1975). The 5' terminal structure of the methylated mRNA synthesized *in vitro* by vesicular stomatitis virus. *Cell*, **5**, 51–58.
- Hercyk, N., Horikami, S. M. & Moyer, S. A. (1988). The vesicular stomatitis virus L protein possesses the mRNA methyltransferase activities. *Virology*, **163**, 222–225.
- Barr, J. N. & Wertz, G. W. (2001). Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAAC-5' tetranucleotide: implications for the mechanism of transcription termination. *J. Virol.* **75**, 6901–6913.
- Mellon, M. G. & Emerson, S. U. (1978). Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus. *J. Virol.* **27**, 560–567.
- Morin, B., Rahmeh, A. A. & Whelan, S. P. (2012). Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. *EMBO J.* **31**, 1320–1329.
- Emerson, S. U. & Schubert, M. (1987). Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of

- the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl Acad. Sci. U.S.A.* **84**, 5655–5659.
15. Peluso, R. W. & Moyer, S. A. (1988). Viral proteins required for the *in vitro* replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology*, **162**, 369–376.
 16. Leyrat, C., Yabukarski, F., Tarbouriech, N., Ribeiro, E. A., Jr, Jensen, M. R., Blackledge, M. *et al.* (2011). Structure of the vesicular stomatitis virus N⁰-P complex. *PLoS Pathogens*, **7**, e1002248.
 17. Chen, M., Ogino, T. & Banerjee, A. K. (2007). Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N-terminus of P that are involved in N⁰-P complex formation and encapsidation of viral genome RNA. *J. Virol.* **81**, 13478–13485.
 18. Masters, P. S. & Banerjee, A. K. (1988). Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J. Virol.* **62**, 2658–2664.
 19. Gérard, F. C. A., Ribeiro Ede, A., Jr, Albertini, A. A., Gutsche, I., Zaccari, G., Ruigrok, R. W. & Jamin, M. (2007). Unphosphorylated *Rhabdoviridae* phosphoproteins form elongated dimers in solution. *Biochemistry*, **46**, 10328–10338.
 20. Das, S. C. & Pattnaik, A. K. (2005). Role of the hypervariable hinge region of phosphoprotein P of vesicular stomatitis virus in viral RNA synthesis and assembly of infectious virus particles. *J. Virol.* **79**, 8101–8112.
 21. Das, T., Pattnaik, A. K., Takacs, A. M., Li, T., Hwang, L. N. & Banerjee, A. K. (1997). Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. *Virology*, **238**, 103–114.
 22. Takacs, A. M., Das, T. & Banerjee, A. K. (1993). Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. *Proc. Natl Acad. Sci. U.S.A.* **90**, 10375–10379.
 23. Ding, H., Green, T. J., Lu, S. & Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* **80**, 2808–2814.
 24. Ding, H., Green, T. J. & Luo, M. (2004). Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2087–2090.
 25. Gérard, F. C. A., Ribeiro Ede, A., Jr, Leyrat, C., Ivanov, I., Blondel, D., Longhi, S. *et al.* (2009). Modular organization of rabies virus phosphoprotein. *J. Mol. Biol.* **388**, 978–996.
 26. Ribeiro, E. A., Jr, Favier, A., Gerard, F. C., Leyrat, C., Brutscher, B., Blondel, D. *et al.* (2008). Solution structure of the C-terminal nucleoprotein–RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J. Mol. Biol.* **382**, 525–538.
 27. Green, T. J., Macpherson, S., Qiu, S., Lebowitz, J., Wertz, G. W. & Luo, M. (2000). Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J. Virol.* **74**, 9515–9524.
 28. Leyrat, C., Jensen, M. R., Ribeiro, E. A., Gérard, F. C. A., Ruigrok, R. W. H., Blackledge, M. & Jamin, M. (2011). The N⁰-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α -helices. *Protein Sci.* **20**, 542–556.
 29. Leyrat, C., Gérard, F. C., de Almeida Ribeiro, E., Jr, Ivanov, I., Ruigrok, R. W. & Jamin, M. (2010). Structural disorder in proteins of the rhabdoviridae replication complex. *Protein Pept. Lett.* **17**, 979–987.
 30. Leyrat, C., Ribeiro, E. A., Gérard, F. C. A., Ivanov, I., Ruigrok, R. W. H. & Jamin, M. (2011). Structure, interactions with host cell and functions of rhabdovirus phosphoprotein. *Future Virol.* **6**, 465–481.
 31. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208.
 32. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C. & Brown, C. J. (2000). Intrinsic protein disorder in complete genomes. *Genome Inf. Ser. Workshop Genome Inf.* **11**, 161–171.
 33. Bernado, P. (2010). Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur. J. Biophys.* **39**, 769–780.
 34. Bernado, P. & Blackledge, M. (2010). Structural biology: proteins in dynamic equilibrium. *Nature*, **468**, 1046–1048.
 35. Heller, W. T. (2005). Influence of multiple well defined conformations on small-angle scattering of proteins in solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **61**, 33–44.
 36. Schneider, R., Huang, J., Yao, M., Communie, G., Ozenne, V., Mollica, L. *et al.* (2012). Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol. Biosyst.* **8**, 58–68.
 37. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622.
 38. Moore, P. B. (1980). Small-angle scattering. Information content and error analysis. *J. Appl. Crystallogr.* **13**, 168–175.
 39. Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886.
 40. Koch, M. H., Vachette, P. & Svergun, D. I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* **36**, 147–227.
 41. Mertens, H. D. & Svergun, D. I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172**, 128–141.
 42. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285.
 43. Lipfert, J. & Doniach, S. (2007). Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 307–327.

44. Rambo, R. P. & Tainer, J. A. (2010). Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Curr. Opin. Struct. Biol.* **20**, 128–137.
45. Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M. R. & Blackledge, M. (2009). Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.* **131**, 17908–17918.
46. Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. H. & Blackledge, M. (2005). A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA*, **102**, 17002–17007.
47. Jamros, M. A., Oliveira, L. C., Whitford, P. C., Onuchic, J. N., Adams, J. A., Blumenthal, D. K. & Jennings, P. A. (2010). Proteins at work: a combined small angle X-ray scattering and theoretical determination of the multiple structures involved on the protein kinase functional landscape. *J. Biol. Chem.* **285**, 36121–36128.
48. Fisher, C. K., Huang, A. & Stultz, C. M. (2010). Modeling intrinsically disordered proteins with Bayesian statistics. *J. Am. Chem. Soc.* **132**, 14919–14927.
49. Jensen, M. R., Markwick, P. R., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S. *et al.* (2009). Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure*, **17**, 1169–1185.
50. Jensen, M. R., Salmon, L., Nodet, G. & Blackledge, M. (2010). Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J. Am. Chem. Soc.* **132**, 1270–1272.
51. Ganguly, D. & Chen, J. (2009). Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. *J. Mol. Biol.* **390**, 467–477.
52. Pelikan, M., Hura, G. L. & Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* **28**, 174–189.
53. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664.
54. Wells, M., Tidow, H., Rutherford, T. J., Markwick, P., Jensen, M. R., Mylonas, E. *et al.* (2008). Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl Acad. Sci. USA*, **105**, 5762–5767.
55. Marsh, J. A. & Forman-Kay, J. D. (2009). Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol.* **391**, 359–374.
56. Jensen, M. R., Communie, G., Ribeiro, E. A., Jr, Martinez, N., Desfosses, A., Salmon, L. *et al.* (2011). Intrinsic disorder in measles virus nucleocapsids. *Proc. Natl Acad. Sci. USA*, **108**, 9839–9844.
57. Uversky, V. N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*, **32**, 13288–13298.
58. Zhang, O. & Forman-Kay, J. D. (1995). Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. *Biochemistry*, **34**, 6784–6794.
59. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **222**, 311–333.
60. Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. (2006). Sensitivity of secondary structure propensities to sequence differences between α - and γ -synuclein: implications for fibrillation. *Protein Sci.* **15**, 2795–2804.
61. Spera, S. & Bax, A. (1991). Empirical correlation between protein backbone conformation and C^α and C^β ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**, 5490–5492.
62. Bae, S. H., Dyson, H. J. & Wright, P. E. (2009). Prediction of the rotational tumbling time for proteins with disordered segments. *J. Am. Chem. Soc.* **131**, 6814–6821.
63. Nicastro, G., Margiocco, P., Cardinali, B., Stagnaro, P., Cauglia, F., Cuniberti, C. *et al.* (2004). The role of unstructured extensions in the rotational diffusion properties of a globular protein: the example of the titin i27 module. *Biophys. J.* **87**, 1227–1240.
64. Glatter, O. & Kratky, O. (1982). *Small Angle X-ray Scattering*. Academic Press, London, UK.
65. Calmettes, P., Durand, D., Desmadril, M., Minard, P., Receveur, V. & Smith, J. C. (1994). How random is a highly denatured protein? *Biophys. Chem.* **53**, 105–113.
66. Petoukhov, M. V. & Svergun, D. I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250.
67. Bernado, P. & Svergun, D. I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. BioSyst.* **8**, 151–167.
68. Ozenne, V., Bauer, F., Salmon, L., Huang, J. R., Jensen, M. R., Segard, S. *et al.* (2012). Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, **28**, 1463–1470.
69. Shen, Y. & Bax, A. (2007). Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR*, **38**, 289–302.
70. Svergun, D., Barberato, C. & Koch, M. H. (1995). CRYSOLE—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.
71. Muller-Sp ath, S., Soranno, A., Hirschfeld, V., Hofmann, H., R uegger, S., Reymond, L. *et al.* (2010). Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl Acad. Sci. USA*, **107**, 14609–14614.
72. Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M. & Blackledge, M. (2010).

- NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* **132**, 8407–8418.
73. Jensen, M. R., Bernadó, P., Houben, K., Blanchard, L., Marion, D., Ruigrok, R. W. & Blackledge, M. (2010). Structural disorder within sendai virus nucleoprotein and phosphoprotein: insight into the structural basis of molecular recognition. *Protein Pept. Lett.* **17**, 952–960.
74. Habchi, J., Mamelli, L., Darbon, H. & Longhi, S. (2010). Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLoS One*, **5**, e11684.
75. Curran, J., Boeck, R., Lin-Marq, N., Lupas, A. & Kolakofsky, D. (1995). Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. *Virology*, **214**, 139–149.
76. Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R. W. & Burmeister, W. P. (2000). On the domain structure and the polymerization state of the sendai virus P protein. *Virology*, **266**, 99–109.
77. Karlin, D., Ferron, F., Canard, B. & Longhi, S. (2003). Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* **84**, 3239–3252.
78. Hock, M., Kraus, I., Schoehn, G., Jamin, M., Andrei-Selmer, C., Garten, W. & Weissenhorn, W. (2009). RNA induced polymerization of the Borna disease virus nucleoprotein. *Virology*, **397**, 64–72.
79. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl Acad. Sci. U.S.A.* **97**, 8868–8873.
80. Pontius, B. W. (1993). Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem. Sci.* **18**, 181–186.
81. Ge, P., Tsao, J., Schein, S., Green, T. J., Luo, M. & Zhou, Z. H. (2010). Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science*, **327**, 689–693.
82. Green, T. J. & Luo, M. (2009). Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc. Natl Acad. Sci. USA*, **106**, 11713–11718.
83. Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. (2004). Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, **318**, 463–473.
84. Curran, J. (1998). A role for the Sendai virus P protein trimer in RNA synthesis. *J. Virol.* **72**, 4274–4280.
85. Jacob, Y., Real, E. & Tordo, N. (2001). Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.* **75**, 9613–9622.
86. Milner, S. T. (1991). Polymer brushes. *Science*, **251**, 905–914.
87. Brown, H. G. & Hoh, J. H. (1997). Entropic exclusion by neurofilament sidearms: a mechanism for maintaining interfilament spacing. *Biochemistry*, **36**, 15035–15040.
88. Lim, R. Y., Huang, N.-P., Köser, J., Deng, J., Lau, K. H. A., Schwarz-Herion, K. *et al.* (2006). Flexible phenylalanine–glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl Acad. Sci. USA*, **103**, 9512–9517.
89. Mukhopadhyay, R. & Hoh, J. H. (2001). AFM force measurements on microtubule-associated proteins: the projection domain exerts a long-range repulsive force. *FEBS Lett.* **505**, 374–378.
90. Thomas, D., Newcomb, W. W., Brown, J. C., Wall, J. S., Hainfeld, J. F., Trus, B. L. & Steven, A. C. (1985). Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. *J. Virol.* **54**, 598–607.
91. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60.
92. Wright, P. E. & Dyson, H. J. (2009). Linking folding and binding. *Curr. Opin. Struct. Biol.* **19**, 31–38.
93. Gupta, A. K. & Banerjee, A. K. (1997). Expression and purification of vesicular stomatitis virus N–P complex from *Escherichia coli*: role in genome RNA transcription and replication *in vitro*. *J. Virol.* **71**, 4264–4271.
94. Barge, A., Gaudin, Y., Coulon, P. & Ruigrok, R. W. (1993). Vesicular stomatitis virus M protein may be inside the ribonucleocapsid coil. *J. Virol.* **67**, 7246–7253.
95. Wyatt, P. J. (1998). Submicrometer particle sizing by multiangle light scattering following fractionation. *J. Colloid Interface Sci.* **197**, 9–20.
96. Lescop, E., Schanda, P. & Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187**, 163–169.
97. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multi-dimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
98. Goddard, T. D. & Kneller, D. G. (2003). SPARKY 3. University of California, San Francisco, CA.
99. Jung, Y. S. & Zweckstetter, M. (2004). Mars—robust automatic backbone assignment of proteins. *J. Biomol. NMR*, **30**, 11–23.
100. Konarev, P. V., Volkov, V. V., Sokolova, A., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282.
101. Semenyuk, A. V. & Svergun, D. (1991). GNOM—a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* **24**, 537–540.
102. Eyal, E., Najmanovich, R., McConkey, B. J., Edelman, M. & Soboлев, V. (2004). Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *J. Comput. Chem.* **25**, 712–724.

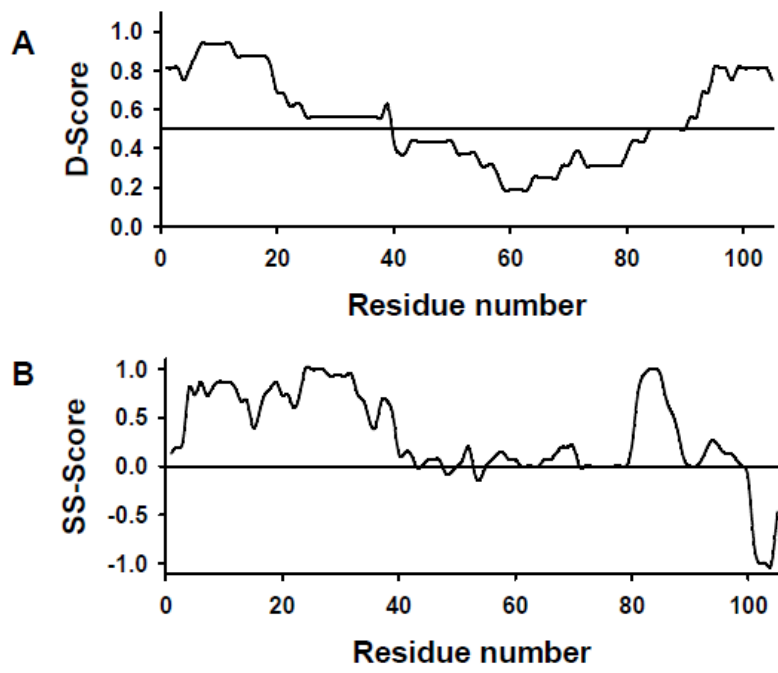


FIGURE S1.

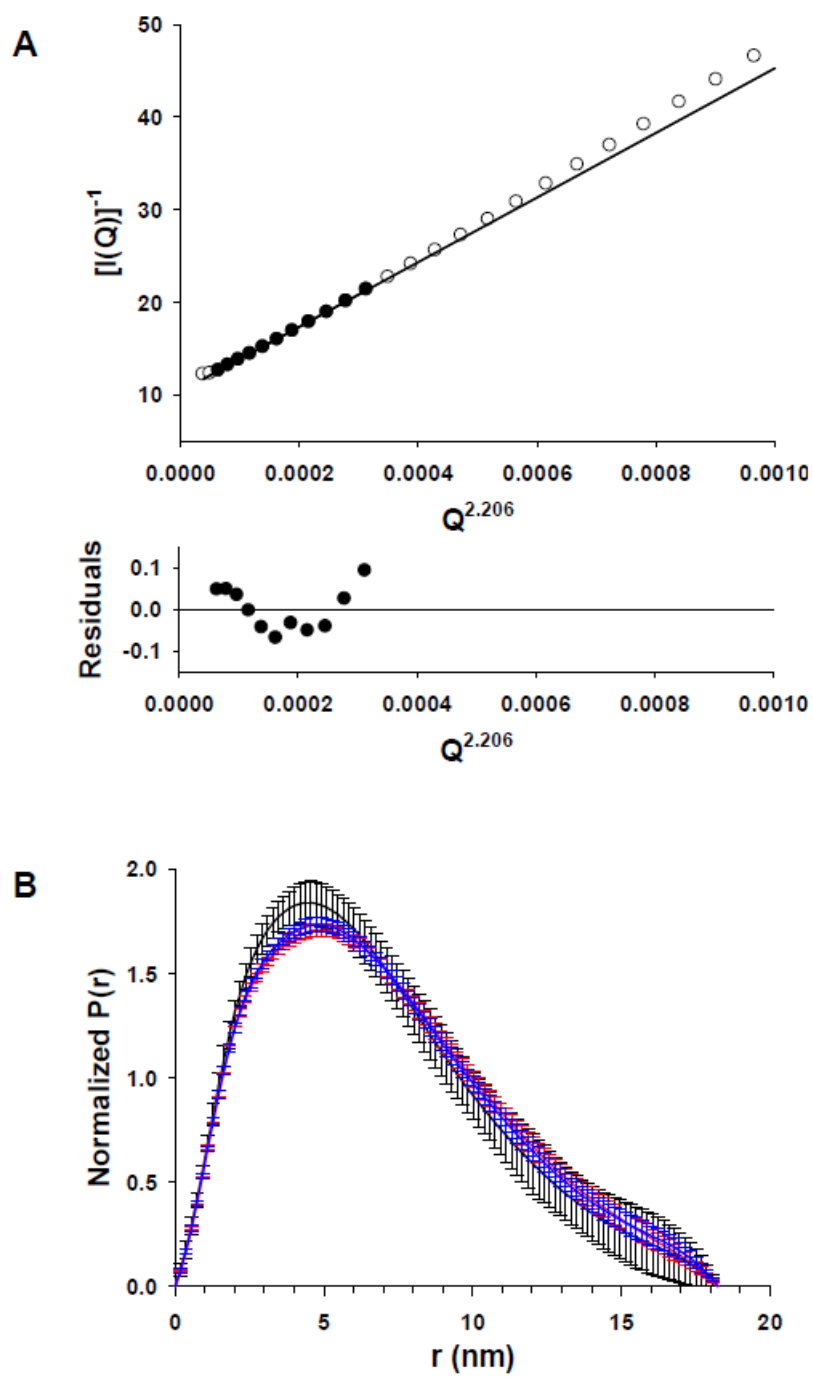


FIGURE S2.

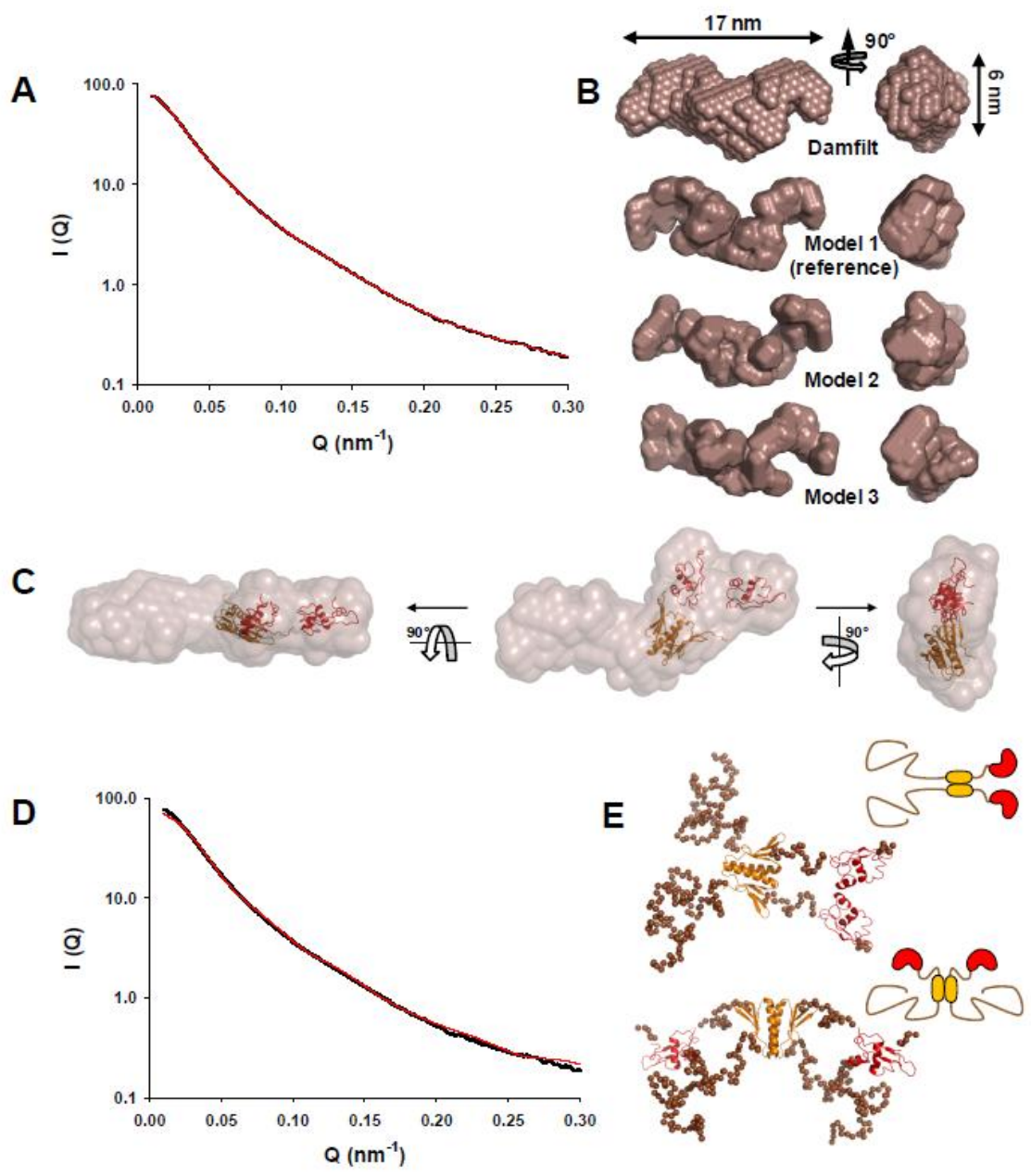


FIGURE S3.

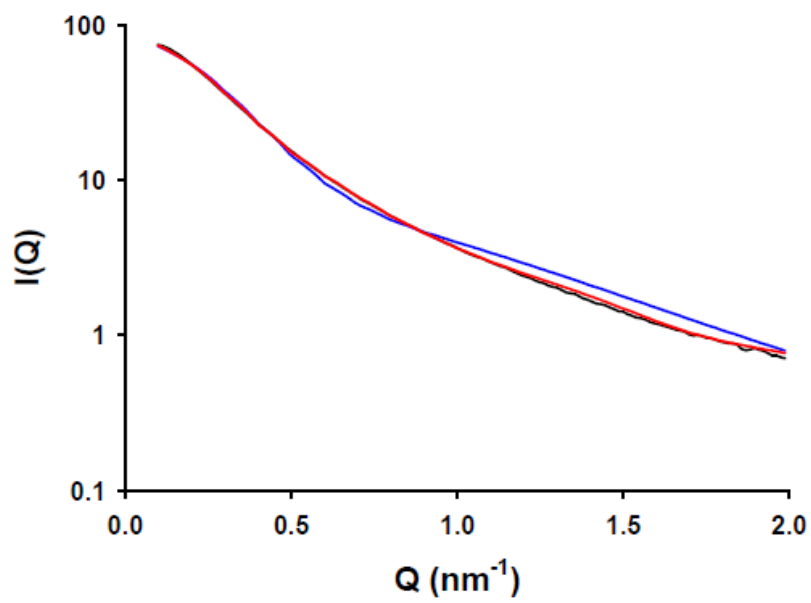


FIGURE S4.

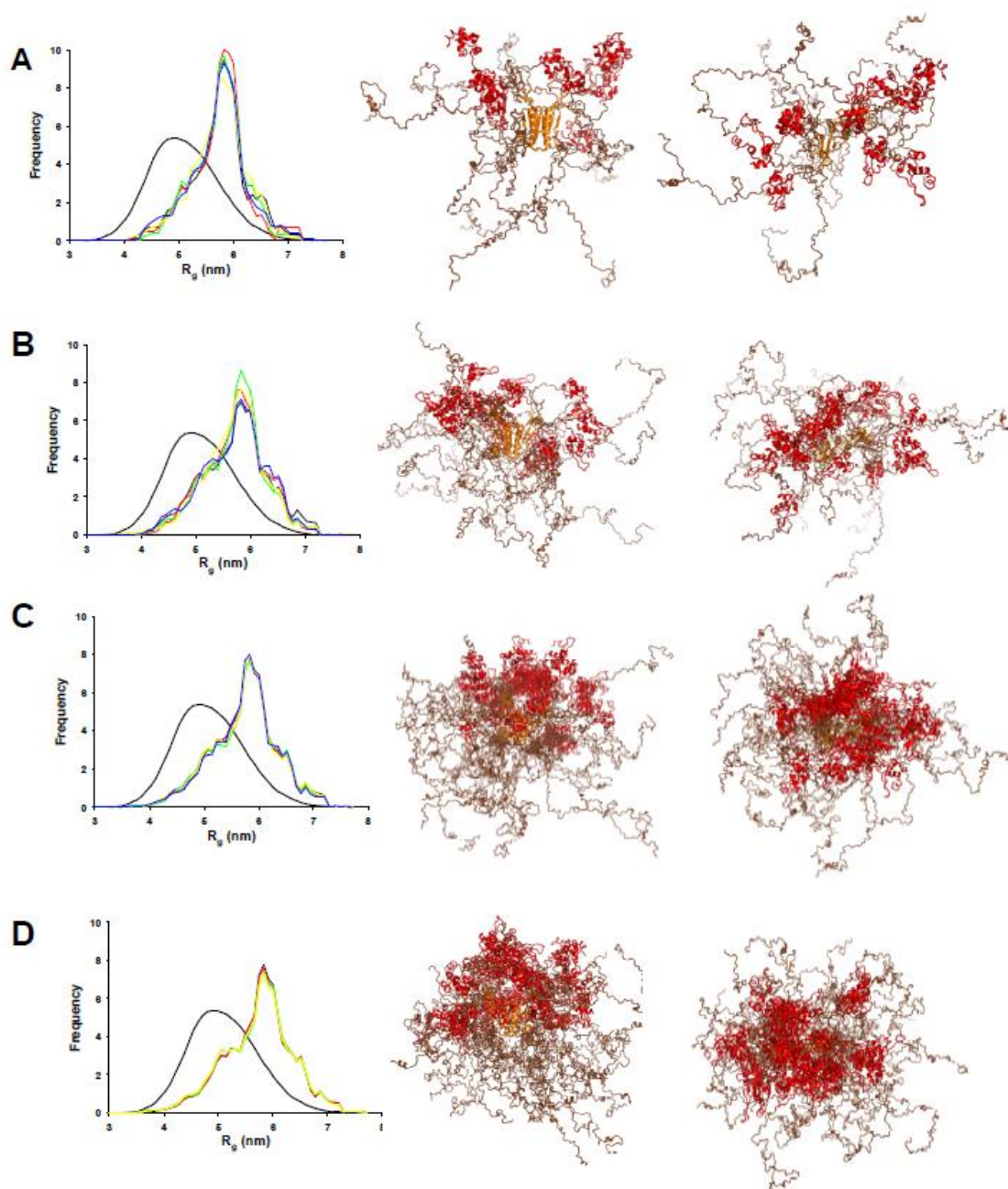


FIGURE S5.

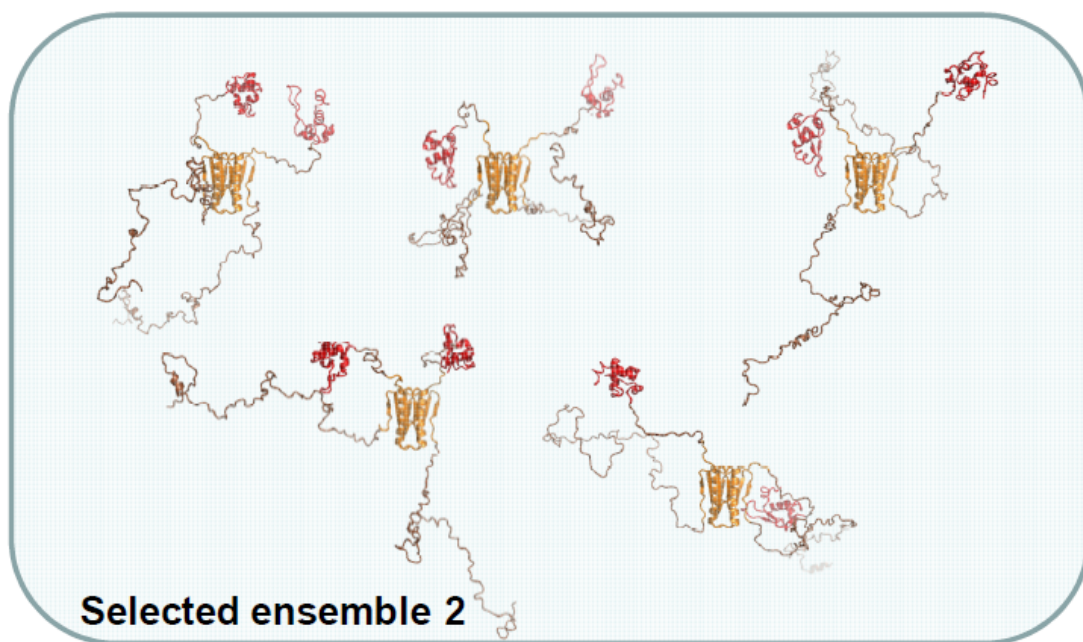
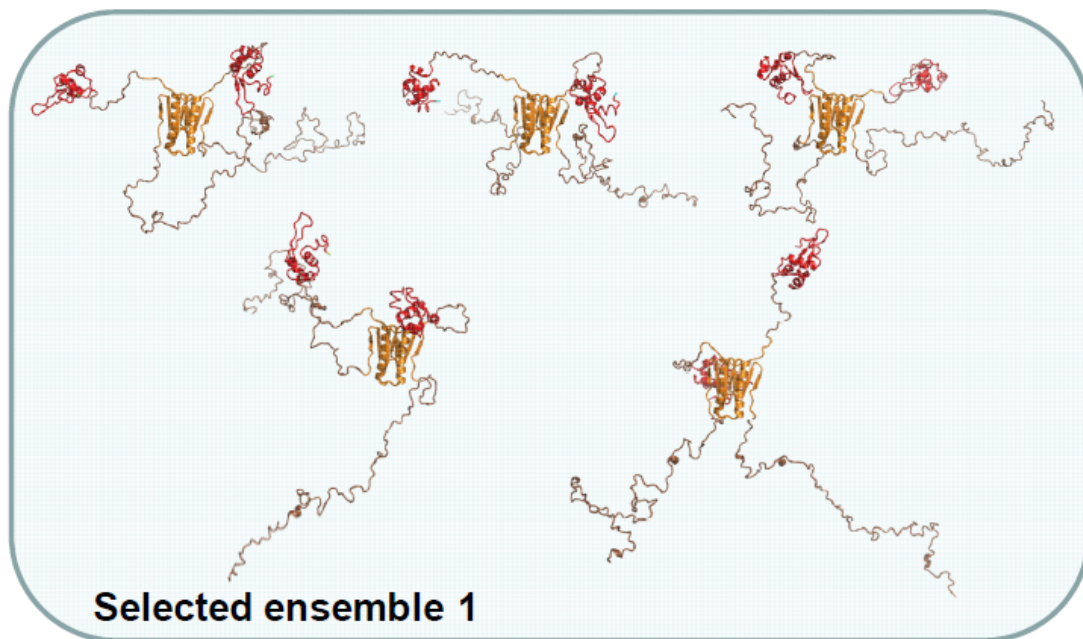


FIGURE S6.

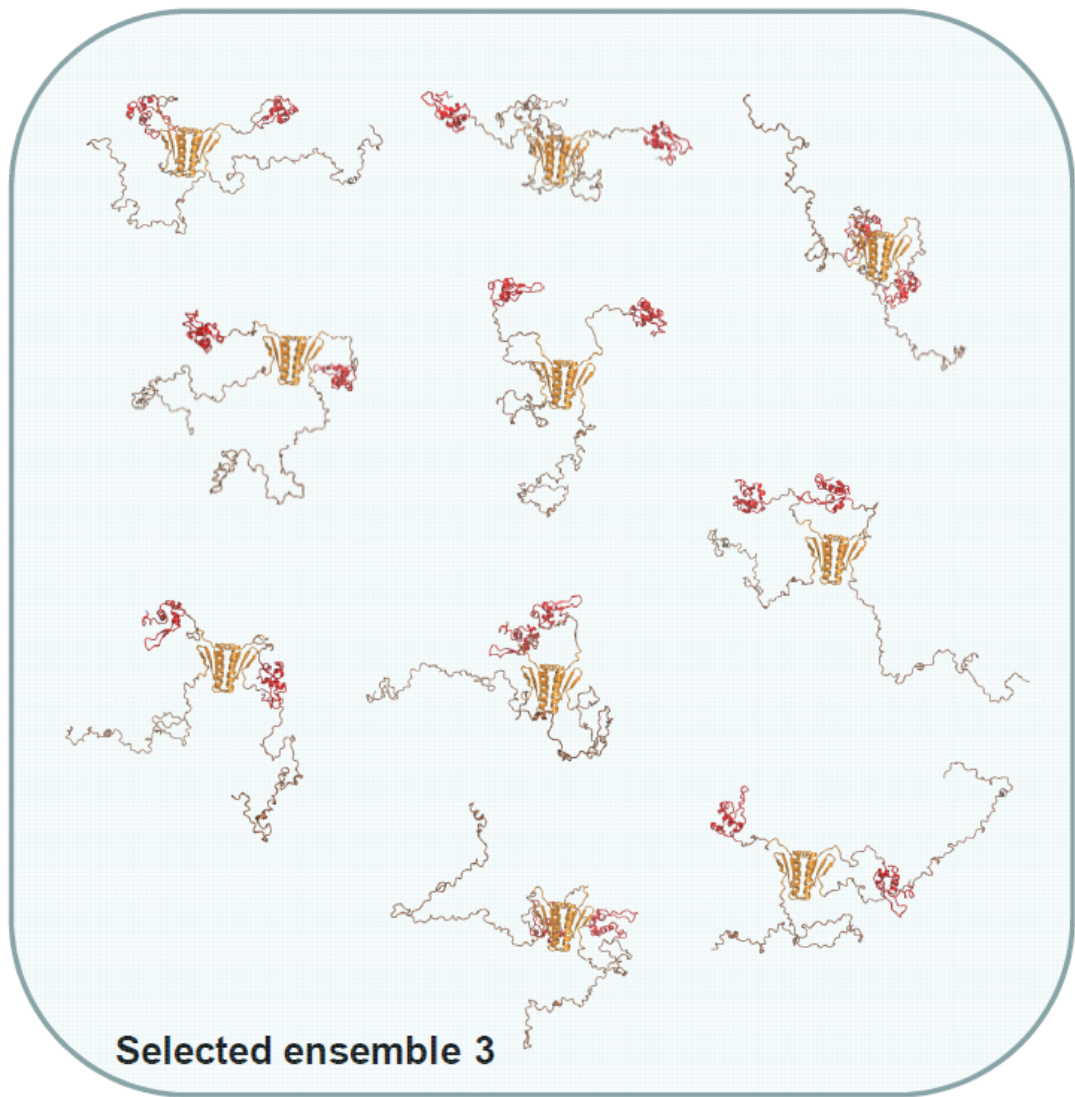


FIGURE S6.

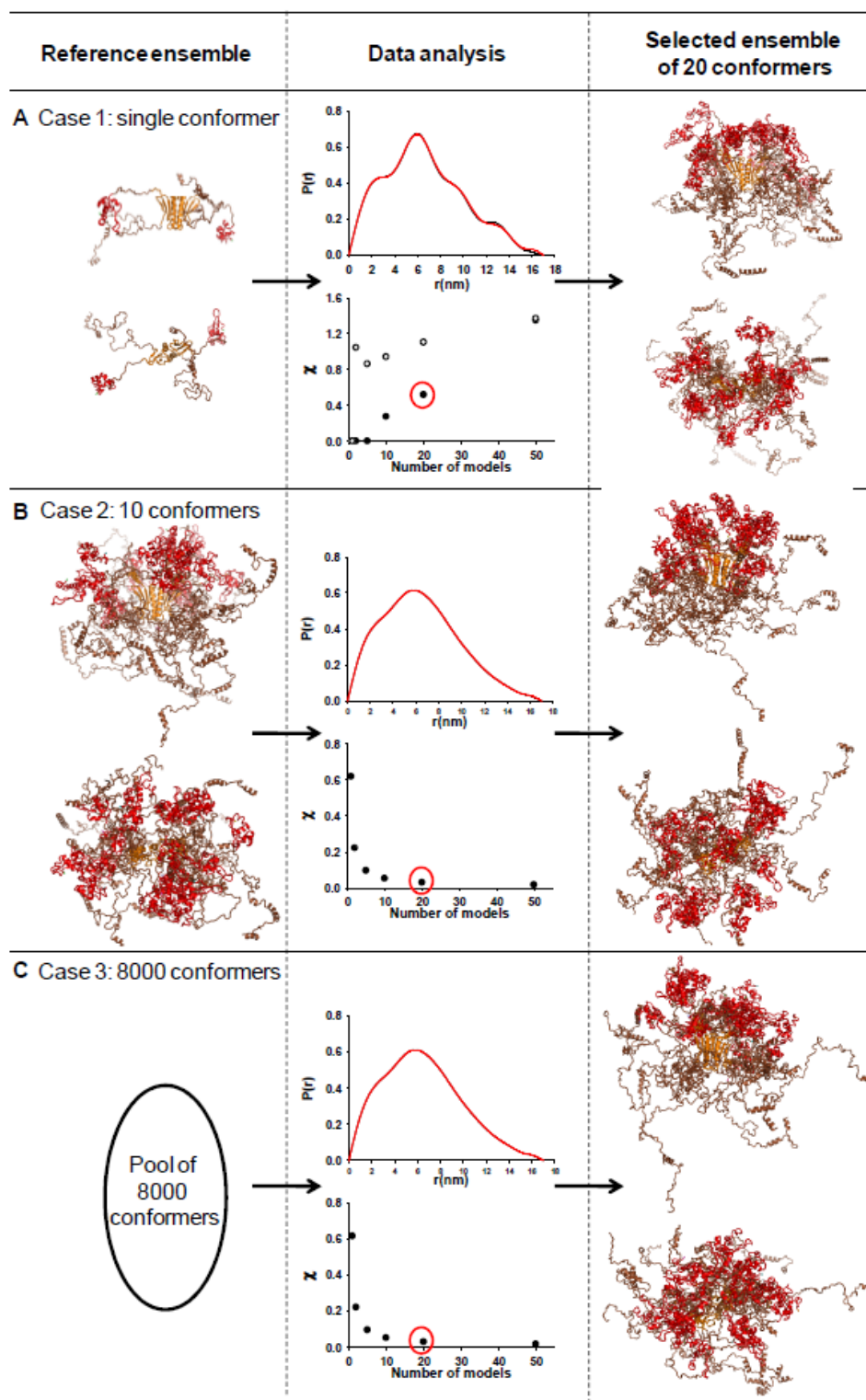


FIGURE S7.

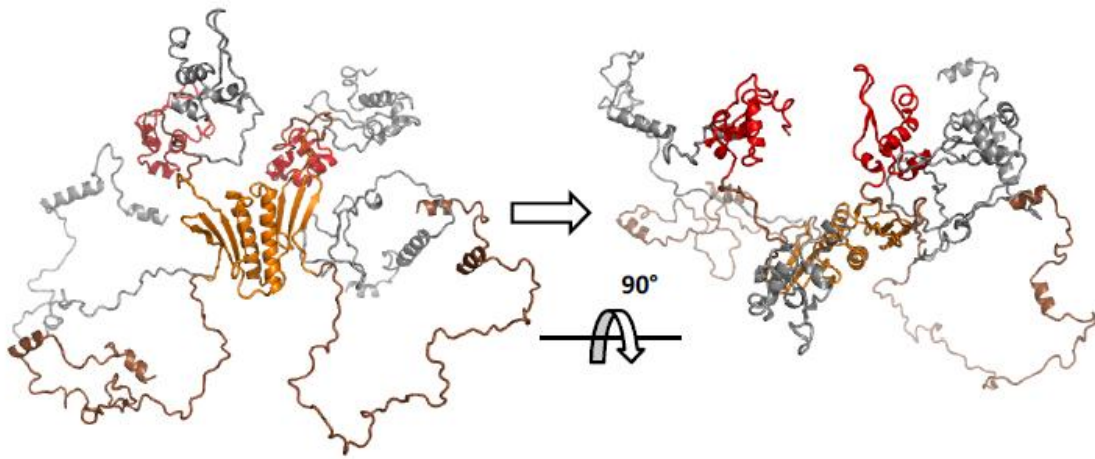


FIGURE S8.

Legends for the supplementary figures

Figure S1. Consensus prediction of disorder and secondary structure for P₁₀₅. The D-score and SS-score plots are reproduced from Gérard et al. ¹. **A. D-score.** D-score values higher than 0.5 for contiguous residues indicate a folded region, and D-score values lower than 0.5 indicate a disordered region. **B. SS-score.** Positive SS-score values for contiguous residues indicate α -helices, and negative SS-score values indicate β -structure or extended conformations.

Figure S2. Analysis of SAXS data. (A) Debye plot. According to the Debye theory, for a random coil polymer, the reciprocal scattering intensity as a function of $Q^{2.206}$ is linear at low Q value according to the following equation²:

$[I(Q)]^{-1} = [I(0)]^{-1} + 0.359 ([I(0)]^{-1} \cdot R_g^2)^{1.103} Q^{2.206}$. The black circles show the data points that are used for the linear regression. **(B) Limits of error on the calculated normalized pair distribution function $P(r)$.** The pair distribution function was calculated with the GNOM program. The curves calculated for the three different protein concentrations are shown as black (1 mg.ml⁻¹), red (5.5 mg.ml⁻¹) and blue (11 mg.ml⁻¹) lines with error bars. The surface areas under the curves were normalized to account for the differences in protein concentration.

Figure S3. Modeling of VSV P as a single conformer. A. Modeling with DAMMIN. The experimental SAXS curve is shown in black and the fitted line in red. For this typical fit, the χ value is 0.056. **B. *Ab initio* bead models obtained with DAMMIN.** Three independent models and the average model from 20 independent runs are represented as surface in two different orientations. The average dimensions of the model are indicated by double arrows. **C. Fitting high-resolution structures.** The structures of the dimeric P_{CED} (in orange) and of two P_{CTD} (in

red) were manually fitted into the bead model. **D. Modeling with BUNCH.** The experimental SAXS curve is shown in black and the fitted line in red. For this typical fit, the χ value is 0.293. **E Rigid body models obtained with BUNCH.** P_{CTD} is colored in red and P_{CED} in orange. The remainder of the structure is shown in brown. The two types of rigid body models are shown in the upper and lower panels.

Figure S4. Fitting SAXS data. The experimental SAXS curve is shown in black. The SAXS curve calculated as an average for the 7,000 models of the initial pool is shown in blue ($\chi = 0.75$). The fitted line for a representative ensemble of 5 conformers selected with GAJOE is shown in red ($\chi = 0.21$).

Figure S5. Ensemble selections. **A. Ensembles of 5 conformers.** The left panel shows the R_g distributions obtained in 5 independent selection processes. The black curve shows the R_g distribution of the initial pool of 7,000 conformers. The right panel shows the superposition of 5 conformers from a representative ensemble selected with GAJOE in two different orientations. **B. Ensembles of 10 conformers.** **C. Ensembles of 20 conformers.** **D. Ensembles of 50 conformers.**

Figure S6. Individual conformers of the selected ensemble. The figure shows the individual conformers of representative ensembles of 5 and 10 conformers.

Figure S7. Reference ensemble simulations. For each case, the left panel shows the reference ensemble. The middle panel shows theoretical data calculated from the reference ensemble and the results from the modeling of these data. The upper plot shows, in black, the pair distribution function calculated for the reference ensemble and, in red, the pair distribution function calculated for the selected ensemble of 20 conformers. The lower plot shows the variation of the

χ value as a function of the number of models in the selected ensemble. Open circles are for selection processes in which multiple selection of the same conformer was not allowed and filled circles are for selection processes in which it was allowed. The right panel shows the selected ensemble of 20 conformers in two orientations. **(A) Case 1: the reference ensemble contains a single conformer. (B) Case 2: the reference ensemble contains 10 different conformers. (C) Case 3: the reference ensemble contains 8,000 conformers.**

Figure S8. Superposition of models with a r.m.s.d. of 2.3 nm. The model in gray is from the reference ensemble of 10 conformers and the model in color is from a representative ensemble of 20 conformers selected with GAJOE. These models exhibited the lowest r.m.s.d. value between the reference and selected ensembles, highlighting the weak similarity between the models forming these two ensembles.

Table S1. Molecular mass and hydrodynamic radius of P and its fragments

Protein	Theoretical MM* (kDa)	Experimental MM (kDa)	Experimental R _s (nm)
P	30.976	61 ± 2	5.2 ± 0.1
P ₆₀	8.054	11 ± 1	2.3 ± 0.1
P ₁₀₅	13.052	14 ± 1	3.1 ± 0.1
P _{CED}	9.189	18 ± 1	2.2 ± 0.1
P _{CTD}	9.075	10 ± 1	1.8 ± 0.1

* The theoretical MM corresponds to the molecular mass calculated from the amino acid sequence for the protein including its His-tag

Table S2. Molecular dimensions of VSV P calculated from SAXS experiments^a

Protein concentration (mg/mL)	R _g (nm) (>Guinier) (q.R _g < 1.3)	R _g (nm) (>Guinier) (q.R _g < 1.0)	R _g (nm) (>Debye)	R _g (nm) (>P(r))	D _{max} (nm)
1.0	5.2 ± 0.1	5.4 ± 0.4	5.8 ± 0.1	5.4 ± 0.2	18.2
5.5	5.4 ± 0.1	5.7 ± 0.4	6.1 ± 0.1	5.6 ± 0.1	18.2
11.0	5.4 ± 0.1	5.8 ± 0.4	6.0 ± 0.1	5.6 ± 0.1	18.2
SANS	5.4 ± 0.2	5.5 ± 0.5	5.9 ± 0.2	n.d.	n.d.

^a The radius of gyration, R_g, was calculated from the initial slope of the Guinier plot, from the slope and intercept of the Debye plot and from the distance distribution function (P(r) *versus* r). D_{max} represents the optimal value used in the calculation of the distance distribution function and obtained with the datgnom routine of the Atsas package³.

Table S3. Molecular dimensions of VSV P calculated from SAXS experiments^a

Number of models in the selected ensemble	χ	R_g (nm)	D_{max} (nm)	R.M.S.F. (nm)	N.S.D.
5	0.207	5.8 ± 0.8	19.3 ± 3.1	4.4 ± 0.9	4.1 ± 0.5
5	0.198	5.8 ± 0.7	20.9 ± 2.2	3.4 ± 0.8	4.1 ± 0.4
5	0.207	5.8 ± 0.2	19.5 ± 1.2	3.9 ± 0.6	3.8 ± 0.2
5	0.201	5.8 ± 0.7	20.9 ± 1.2	4.3 ± 0.9	4.1 ± 0.2
5	0.214	5.9 ± 0.6	21.1 ± 2.0	4.4 ± 0.6	3.6 ± 0.2

References

1. Gérard, F. C. A., Ribeiro, E. A., Leyrat, C., Ivanov, I., Blondel, D., Longhi, S., Ruigrok, R. W. H. & Jamin, M. (2009). Modular organization of rabies virus phosphoprotein. *J. Mol. Biol.* **388**, 978-996.
2. Calmettes, P., Durand, D., Desmadril, M., Minard, P., Receveur, V. & Smith, J. C. (1994). How random is a highly denatured protein? *Biophys Chem* **53**, 105-13.
3. Konarev, P., Petoukhov, M., Volchkov, V. & Svergun, D. I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Cryst.* **39**, 277-286.

3.2.2. Caractérisation d'un fragment tronqué de la partie C-terminale, P₁₋₁₇₇

Afin d'étudier l'influence que peuvent avoir les différentes régions sur les dimensions globales de la molécule nous avons produit et caractérisé un variant tronqué de la P de VSV contenant les 177 résidus du côté N-terminal. Cette protéine ne contient ni les domaines C-terminaux ni les jonctions flexibles qui les lient au domaine central mais elle conserve néanmoins son domaine de dimérisation intact (**Figure 2**).

La masse moléculaire (MM) de 40 ± 4 kDa mesurée par SEC-MALLS-RI est proche de la MM théorique de 40.9 kDa confirmant l'état dimérique du fragment. Son volume d'éluion de la colonne est très proche de celui de la P entière (**Figure 3F**), confirmant qu'il s'agit également d'une protéine non-globulaire. Afin de déterminer les dimensions globales de cette protéine nous avons enregistré des courbes de SAXS à des concentrations allant de 2.6 à 1 mg.ml⁻¹. Les profils des courbes sont très similaires ce qui suggère l'absence d'agrégation (**Figure 3A**). Les valeurs de Rg calculées en utilisant l'approximation de Guinier ne changent pas avec la concentration confirmant l'absence d'agrégation et donnant un Rg moyen est de 5.0 ± 0.1 nm (**Figure 3B**). La fonction de distribution de distances, P(r), est très asymétrique et lisse, en accord avec ce qui est attendu pour une protéine désordonnée et flexible. La dimension maximale *Dmax* est de 17.5 ± 0.5 nm avec un maximum de distances intramoléculaire à environ 3 nm (**Figure 3C**). Le Rg de 5.0 ± 0.1 nm et le Dmax de 17.5 ± 0.5 nm sont très proches des valeurs obtenues pour la protéine entière qui sont respectivement de 5.3 ± 0.1 nm et 18 nm (Gérard, Ribeiro et al. 2007; Leyrat, Schneider et al. 2012), suggérant que les domaines C-terminaux contribuent très peu aux dimensions globales de la molécule. Le graphique de Kratky présente un pic très asymétrique caractérisé par une augmentation continue pour les valeurs de q élevées, tandis que le graphique de Porod-Debye n'atteint pas de plateau, ce qui est caractéristique d'une protéine majoritairement désordonnée (**Figure 3D et 3E**).

Ce fragment de P a été utilisé pour la reconstitution d'un complexe N⁰-P avec la nucléoprotéine tronquée du bras N-terminal (N_{Δ21}), et ce complexe N_{Δ21}⁰-P₁₋₁₇₇ a été utilisé dans des expériences de RMN pour la caractérisation de la stœchiométrie du complexe (voir **Chapitre 5**).

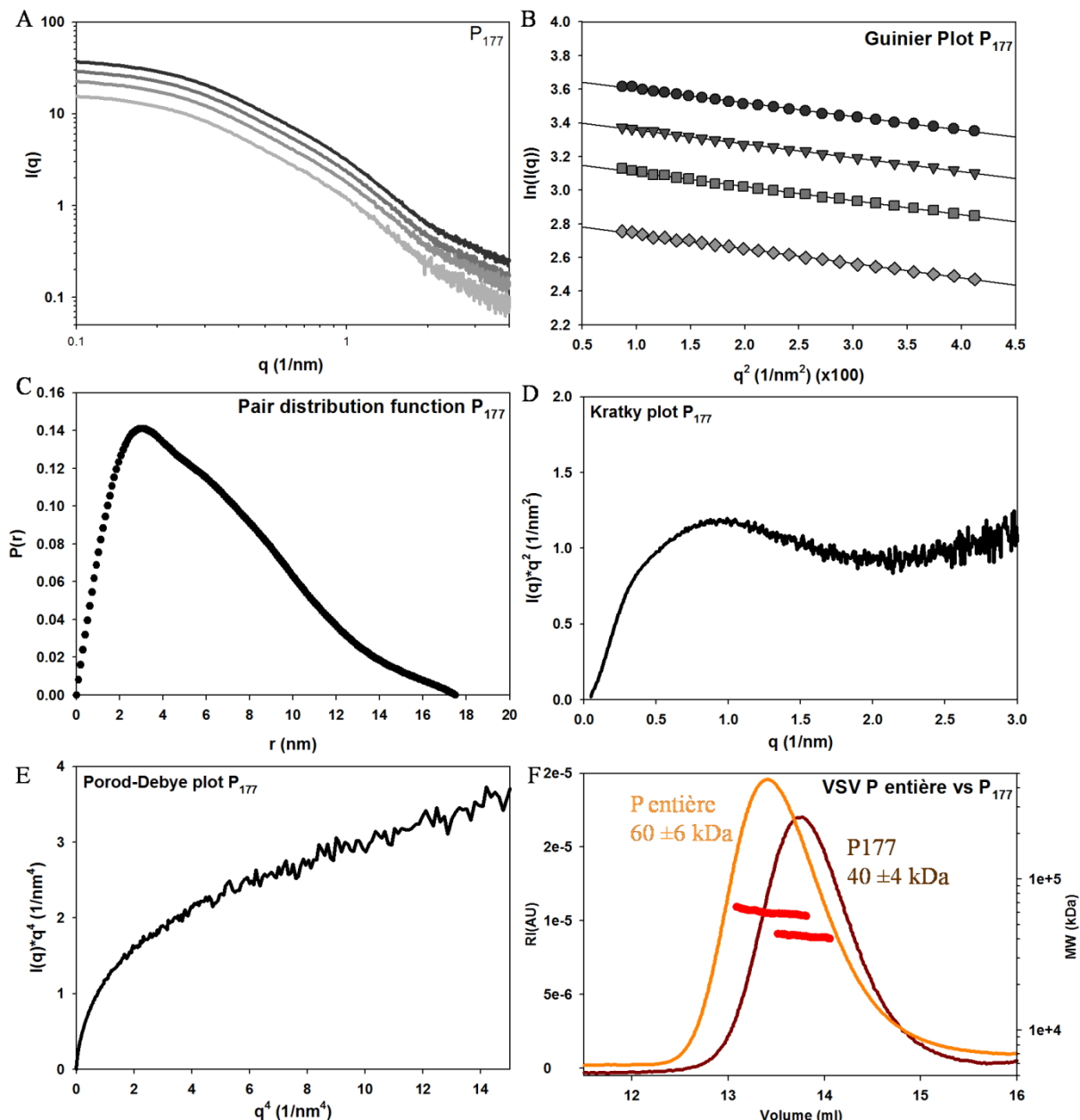


Figure 3 : Expériences de diffusion des rayons X aux petits angles (SAXS) et SEC-MALLS-RI sur la phosphoprotéine P₁₋₁₇₇ du virus de la stomatite vésiculaire (VSV). **A)** Les courbes de SAXS ont été enregistré pour des valeurs de q entre 0.05 nm^{-1} et 4 nm^{-1} à quatre concentrations de protéine ($2.6, 2.0, 1.5$ et 1 mg.ml^{-1}). **B)** Analyse de Guinier dans la région de $q.R_g < 1.1$. **C)** Fonction de distribution des distances $P(r)$, **D)** représentation de Kratky et **E)** représentation de Porod-Debye pour la courbe à 2.6 mg.ml^{-1} . **F)** Profiles d'élution de P₁₋₁₇₇ (ligne en marron) et de P entière (ligne en orange) de la colonne de gel filtration Superdex S200, équilibrée dans $20 \text{ mM Tris-HCl pH } 7.5, 150 \text{ mM NaCl}$. En rouge les masses moléculaires (MM) à travers les pics d'élution, calculées à partir de la diffusion statique de la lumière et la réfractométrie pour chaque protéine. Les valeurs numériques des MM ont le même code couleur que les profils d'élution.

3.3 Structure de la phosphoprotéine du virus Nipah

3.3.1. Analyse bioinformatique de la phosphoprotéine du virus Nipah

Un alignement de séquences en acides aminés de différents isolats de virus Nipah et Hendra indique la présence de régions bien conservées alternant avec des régions faiblement conservées (**Figure 4**). La région C-terminale allant du résidu 470 au résidu 709, qui correspond vraisemblablement à la partie C-terminale des autres P, incluant le domaine de tétramérisation et le domaine C-terminal P_{XD} de fixation sur la nucléocapside, est conservée, ainsi que deux régions de la partie N-terminale qui correspondent au site potentiel d'interaction avec N⁰ (aa 1-65) et au site décrit d'interaction avec la protéine STAT1 (Signal Transducer and Activator of Transcription 1) (aa 85-140) (Ciancanelli, Volchkova et al. 2009). Les autres régions présentant des taux élevés de conservation n'ont pas été impliquées dans une fonction de la P.

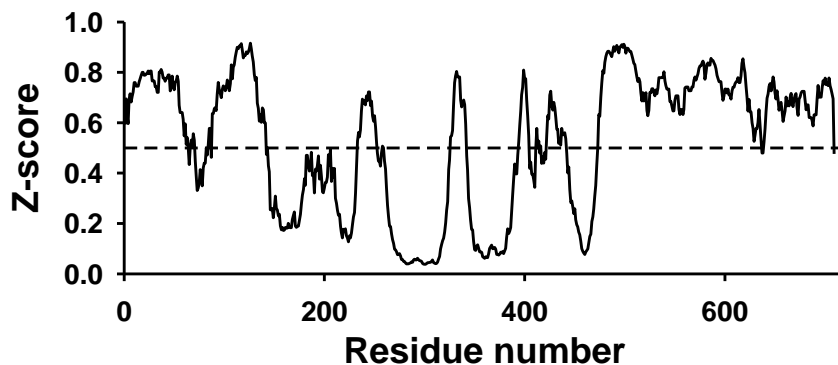


Figure 4 : Score de conservation de séquence pour la phosphoprotéine (P) des virus Nipah (NiV) et Hendra (HeV). Score de conservation (Z-score) de la séquence calculé avec le programme AL2CO (Pei and Grishin 2001) à partir d'un alignement multiple obtenu avec ClustalW (Larkin, Blackshields et al. 2007). Les résidus dont les valeurs de Z-score sont supérieures à 0.5 sont considérés comme conservée.

Le consensus des prédictions de désordre obtenu sous la forme du D-score indique la présence de quatre régions structurées, correspondant aux résidus 1 à 35, 100 à 230, 490 à 579 et 606 à 709 (**Figure 5**). Par comparaison avec la prédiction de désordre du D-score et le positionnement du domaine d'oligomérisation de la protéine P du virus de Sendai, la troisième région correspondrait au domaine d'oligomérisation (Tarbouriech, Curran et al. 2000; Gérard, Ribeiro et al. 2009). Le programme COILS (Lupas, Van Dyke et al. 1991) prédit la présence d'un faisceau d'hélices α surenroulées (coiled-coil) dans la région 520-570 (**Figure 6**), renforçant l'idée que cette région correspond au domaine d'oligomérisation. Par comparaison avec les autres P des paramyxovirus, la quatrième région comprendrait le

domaine P_{XD} qui est le site d'interaction avec la queue désordonnée de la nucléocapside, N_{TAIL} (Johansson, Bourhis et al. 2003; Blanchard, Tarbouriech et al. 2004; Kingston, Hamel et al. 2004; Houben, Marion et al. 2007; Kingston, Gay et al. 2008).

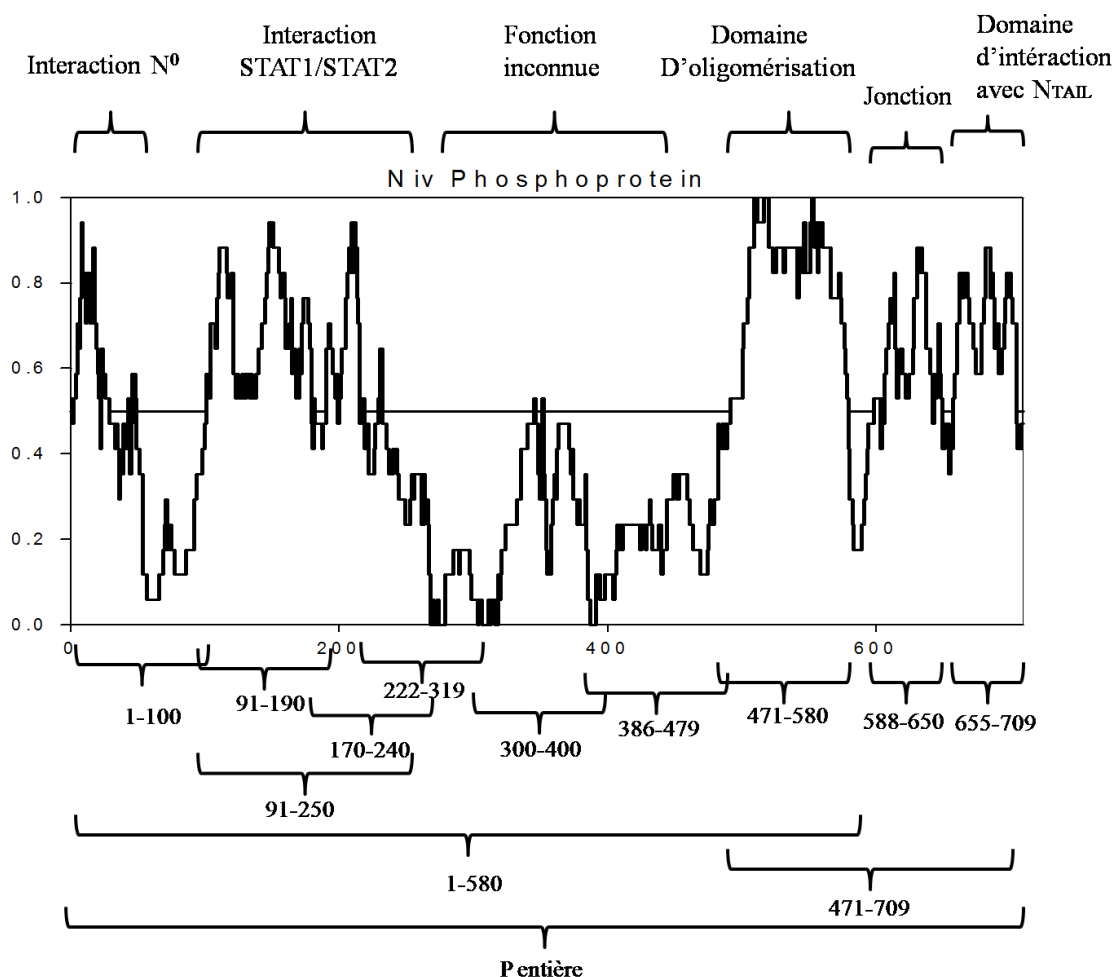


Figure 5 : Prédiction de désordre pour la phosphoprotéine (P) du virus Nipah (NiV). En haut sont représentées dans l'ordre la région d'interaction avec N⁰ (hypothétique) la région d'interaction avec STAT1/2 (Rodriguez, Cruz et al. 2004; Ciancanelli, Volchkova et al. 2009), la région correspondant au domaine d'oligomérisation (hypothétique), la région de jonction (PQ, hypothétique) et la région d'interaction avec le domaine N_{TAIL} de la nucléocapside (Habchi, Blangy et al. 2011). La fonction de la région entre les domaines d'interaction avec STAT1/2 et d'oligomérisation n'est pas connue. Au milieu, le graph D-score qui représente la prédiction de désordre en fonction du numéro des résidus. En bas sont représentées les différentes constructions de NiV P générées au cours du travail de thèse présenté ici.

Toutefois, le domaine P_{XD} des autres virus est plus court. Il correspondrait aux 50 derniers résidus. Une prédiction de structure pour la partie C-terminale par homologie avec le P_{XD} du virus de la rougeole confirme cette localisation (Habchi, Blangy et al. 2011). Dans la partie N-terminale, la première région prédite structurée correspond au site d'interaction avec N⁰ dans d'autres virus (Curran, Marq et al. 1995; Mavrakis, Mehoulas et al. 2006; Chen, Ogino et al. 2007). En contradiction avec nos prédictions, une étude récente par chromatographie d'exclusion de taille et dichroïsme circulaire montre que la partie N-terminale de la P NiV est globalement désordonnée (Habchi, Mamelli et al. 2010). Cette

observation est similaire à celle faite avec VSV et RaV, suggérant qu'il s'agit d'une région qui se structure lors de la formation d'un complexe avec son partenaire N⁰ (Gérard, Ribeiro et al. 2009). Quant à la seconde région prédite structurée, elle comprend le site identifié d'interaction avec les protéines STAT1, STAT2 (Shaw, Garcia-Sastre et al. 2004; Ciancanelli, Volchkova et al. 2009) et PLK1 (Polo-Like Kinase 1) (Ludlow, Lo et al. 2008). Cette seconde région prédite comme structurée est plus longue que la seconde région conservée entre les P de NiV et HeV. La plus faible conservation dans la région (aa 150-230) est due à une différence entre les séquences des isolats de NiV par rapport aux séquences des isolats d'HeV qui suggère une différence dans les processus de recrutement des molécules STAT entre les deux virus. Les deux régions de la partie N-terminale de NiV P, ainsi que le début de la région C-terminale, pourraient donc constituer des « éléments de reconnaissance moléculaire » ou MoRE.

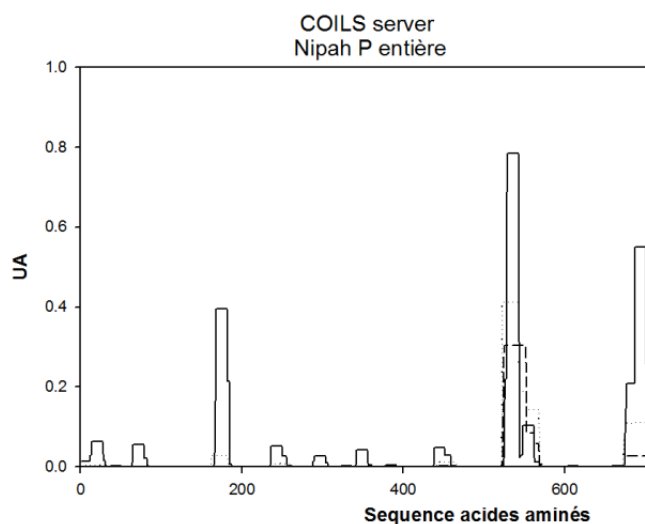


Figure 6 : Prédiction de régions formant des faisceaux d'hélices α surenroulées. Les lignes noires continue et en pointillée représentent la prédiction en utilisant une fenêtre de 14 ou 28 résidus, respectivement.

Sur la base de ces prédictions différentes constructions du gène de la phosphoprotéine de Nipah (isolat malaisien UMMC1, Genbank AY029767) ont été clonées dans le vecteur d'expression bactérien pET28 en fusion avec une étiquette poly-histidine non clivable à leur extrémité C-terminale ou dans le plasmide pETM40 en fusion avec une étiquette MBP clivable avec la protéase TEV à leur extrémité N-terminale (voir **Chapitre 2 – Matériels et Méthodes**) (**Figure 5**). La protéine entière et les différents fragments ont été exprimés en bactéries *E. coli* Rosetta et ils ont été purifiés en deux étapes, par une chromatographie d'affinité sur une colonne Ni²⁺ suivie par une chromatographie d'exclusion de taille. Les

différentes protéines présentait un degré de pureté supérieur à 95%, estimé par l'absence de contaminants visibles par SDS-PAGE en utilisant une coloration au bleu de Coomassie.

3.3.2. Le domaine P_X d'interaction avec les nucléocapsides

3.3.2.1. Structure tridimensionnelle du domaine P_X du virus Hendra

La structure tridimensionnelle du domaine P_X du virus Hendra a été déterminée par cristallographie et diffraction des rayons X. La protéine a cristallisé dans un groupe d'espace P2₁ avec deux molécules dans l'unité asymétrique. Les phases ont été déterminées par la méthode SAD (**Figure 7A**) en utilisant des données collectées à la ligne BM14-4 à ESRF avec des cristaux d'un dérivé sélénométhionine et la structure a été déterminée à une résolution de 1.65 Å (**Tableau 1**).

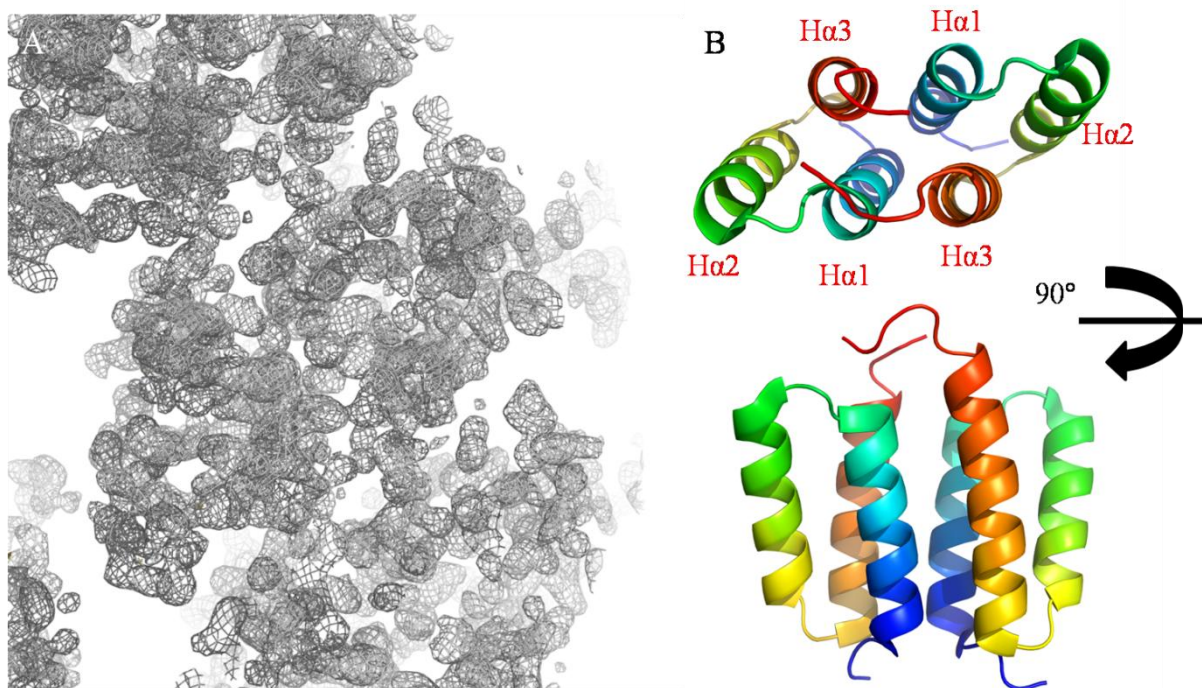


Figure 7 : Structure du domaine C-terminal (P_{XD}) du virus Hendra (HeV). **A)** Carte de densité électronique (contour à 1 σ) obtenues expérimentalement par la méthode SAD pour le domaine P_X du virus Hendra. **B)** Représentation en cartoon des deux molécules de P_{XD} composant l'unité asymétrique. Les molécules sont colorées du bleu (N-terminal) au rouge (C-terminal).

La carte de densité électronique était bien définie pour l'ensemble du domaine, à l'exception de quelques chaînes latérales à la surface de la protéine, et le modèle de la chaîne

B comprend quelques résidus supplémentaires (LEHH) appartenant à l'étiquette poly-histidine. La structure est composée de trois hélices α formant un faisceau de trois hélices antiparallèles (H α 1, H α 2 et H α 3) (**Figure 7B**). L'unité asymétrique est constituée de deux copies du domaine P_X qui sont équivalentes (RMSD = 0.65 Å) et interagissent par une interface (900 Å²) formée par les hélices H α 1 et H α 3. Cette interface, considérée comme une interface de dimérisation stable par le logiciel PISA (Krissinel and Henrick 2007), n'est pas entièrement symétrique et est composée de résidus hydrophobes (Ala 657, Val 661 et 662, Met 665, Val 692, Ala 696, Ile 703) et résidus polaires (Asp 655, Ser 658 Thr 664 et 668, His 669, Gln 693, Asn 700). La structure du domaine P_X du virus Hendra est donc semblable à celle du domaine P_X du virus des oreillons (MuV), du virus de la rougeole (MeV) et du virus Sendai (SeV) (**Figure 8**).

Tableau 1. Statistiques de collecte des données (BM14-4) et d'affinement de la structure du HeV P_{XD}

	Crystal
Data collection	
Wavelength (Å)	0.9393
Space group	P12 ₁ 1
Unit cell	
a, b, c (Å)	21.49, 43.28, 51.26
α , β , γ (°)	90.00, 79.78, 90.00,
Resolution ^a (Å)	32.85 - 1.65 (1.80 - 1.65)
I/ σ (I) ^a	14.24 (1.89)
R _{symm}	0.037 (0.546)
Completeness ^a (%)	98.6 (98.7)
Multiplicity ^a	2.7 (2.48)
Total reflections*(Friedel pairs unmerged)	58428
Refinement	
Resolution ^a (Å)	32.85 - 1.65 (1.72 - 1.65)
R-factor ^a	0.188 (0.306)
R-free ^a	0.224 (0.306)
Number of atoms	1010
macromolecules	973
ligands	2
water	35
Number of protein residues	116
R.m.s. deviations	
bond lengths (Å)	0.012
bond angles (°)	1.12
Ramachandran favored (%)	99
Ramachandran allowed (%)	1
Average B-factor	37.20
macromolecules	36.90
solvent	44.80

^aValues in parentheses are for highest-resolution shell.

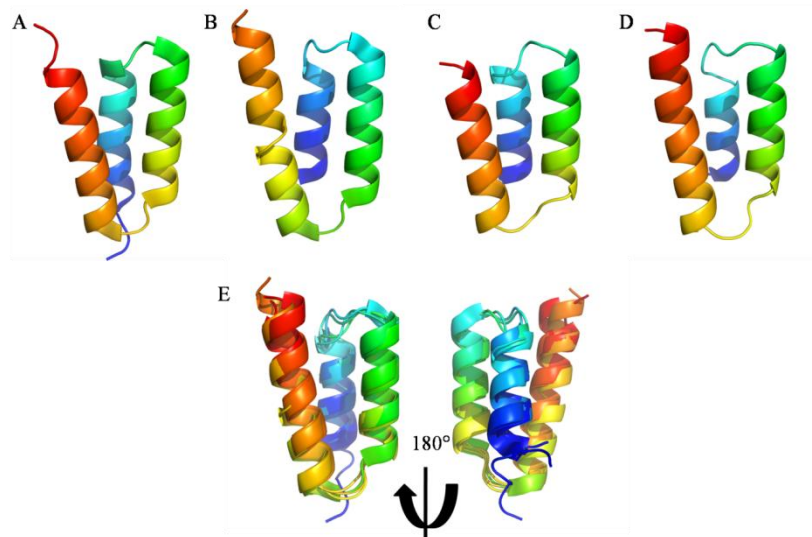


Figure 8 : Les domaines C-terminaux (P_{XD}) des *Paramyxovirinae*. Les P_{XD} des virus **A**) Hendra (HeV) ; **B**) Rougeole (MeV, code PDB : 1OKS); **C**) Oreillons (MuV, code PDB : 3BBZ); **D**) Sendai (SeV, code PDB : 1R4G). **E**) Les quatre P_{XD} superposés. Les domaines sont colorés du bleu (N-terminal) au rouge (C-terminal). Les RMSD des $C\alpha$ entre le HeV P_{XD} et Mev, Muv et SeV P_{XD} sont 1.22, 1.55 et 0.88Å, respectivement. Les P_{XD} sont superposés sur les structures secondaires en utilisant le programme CCP4MG (McNicholas, Potterton et al. 2011).

3.3.2.2. Structure tridimensionnelle du domaine P_X du virus Nipah

La structure tridimensionnelle du domaine P_X du virus Nipah a été déterminée par cristallographie et diffraction des rayons X. Le domaine P_X a cristallisé dans un groupe d'espace $P4_32_12$ avec une seule molécule dans l'unité asymétrique. Les données ont été collectées sur les lignes de lumière ID29 à l'ESRF et PROXIMA-1 à SOLEIL.

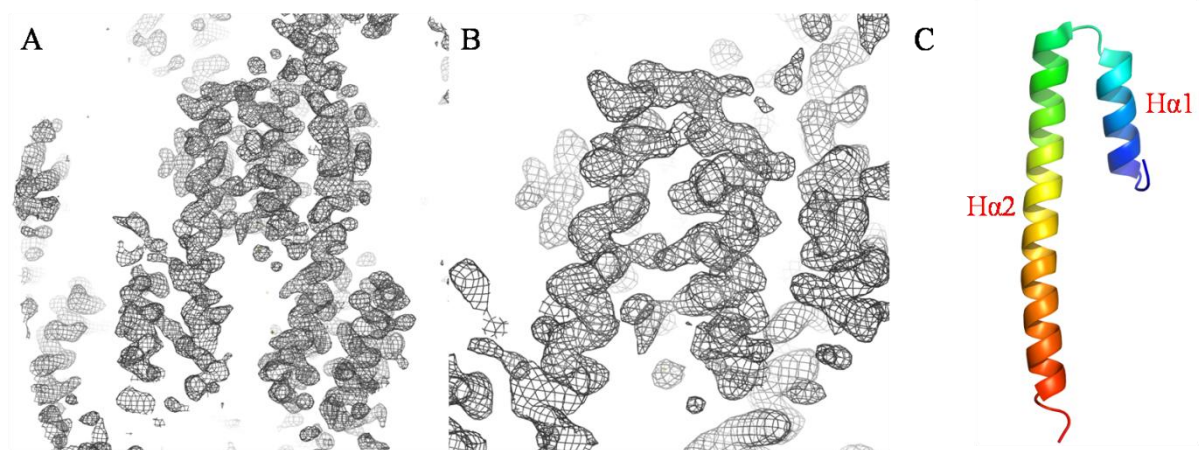


Figure 9 : Structure du domaine C-terminal (P_{XD}) du virus Nipah (NiV). **A**) et **B**) Cartes de densité électronique (contour à 1σ) obtenues expérimentalement par la méthode SIRAS pour le domaine P_X du virus Nipah. **C**) Représentation en cartoon du modèle final coloré du bleu (N-terminal) au rouge (C-terminal).

Les phases initiales ont été obtenues par la méthode SIRAS (**Figure 9A et 9B**) et la qualité de la carte initiale a permis de tracer automatiquement la quasi-totalité de la chaîne

avec le programme BUCCANEER (Cowtan 2006). La structure a été déterminée à une résolution de 2.45 Å (**Tableau 2**). L'analyse par spectrométrie de masse (plateforme IBS) a montré que la méthionine N-terminale a été clivée au cours de l'expression. Du côté N-terminal, le résidu Ala656 n'est pas visible dans la densité électronique, alors que du côté C-terminal, seule la densité électronique de la chaîne principale est visible pour les quelques derniers acides aminés et l'étiquette poly-histidine n'est pas visible.

Tableau 2. Statistiques de collecte des données (ID29) et d'affinement de la structure du NiV P_{XD}

Crystal	
Data collection	
Wavelength (Å)	1.776
Space group	P4 ₃ 2 ₁ 2
Unit cell	
a, b, c (Å)	35.5, 35.5, 112.9
α, β, γ (°)	90.00, 90.00, 90.00,
Resolution ^a (Å)	33.71 – 2.45 (2.58 – 2.45)
I/σ(I) ^a	4.3 (2.3)
R-symm ^a	0.059 (0.244)
Completeness ^a (%)	96.5 (65.95)
Multiplicity ^a	5.8 (2.0)
Unique reflections	2827
Refinement	
Resolution ^a (Å)	33.71 – 2.45 (2.51 – 2.45)
R-work ^a	0.216 (0.36)
R-free ^a	0.27.3 (0.38)
Number of atoms	455
macromolecules	444
water	11
Number of protein residues	55
R.m.s. deviations	
bond lengths (Å)	0.023
bond angles (°)	2.18
Ramachandran favored (%)	86.6
Ramachandran allowed (%)	9.4
Ramachandran outlier (%)	3.8
Average B-factor	73.4

^aValues in parentheses are for highest-resolution shell.

La protéine est constituée de deux hélices, une courte ($H\alpha 1$) et une longue ($H\alpha 2$), qui peuvent être clairement distinguées dans la carte de densité électronique initiale (**Figure 9A et 9B**). Bien qu'il n'y ait qu'une molécule dans l'unité asymétrique, l'analyse de l'empilement cristallin indique la formation d'une interface hydrophobe d'interaction (580 \AA^2) entre deux molécules de P_X dans le cristal, formées par l'hélice $H\alpha 1$ et la moitié N-terminale de l'hélice $H\alpha 2$ (**Figure 10**). L'analyse de cette interface avec le logiciel PISA (Krissinel and Henrick 2007) suggère qu'elle est spécifique et ne peut pas être expliquée uniquement par l'empilement cristallin.

Dans la structure cristallographique du complexe P_{XD} - N_{TAIL} du virus de la rougeole (Kingston, Hamel et al. 2004), et dans les cartographies réalisées par RMN des interactions entre ces deux partenaires dans les complexes des virus de Sendai (Houben, Marion et al. 2007) et d'Hendra (Communie, G. et al. dans ANNEXE), le site de fixation de l'hélice du N_{TAIL} sur le domaine P_X est constitué par la surface formée par l'assemblage côte-à-côte des hélices $H\alpha 2$ et $H\alpha 3$.

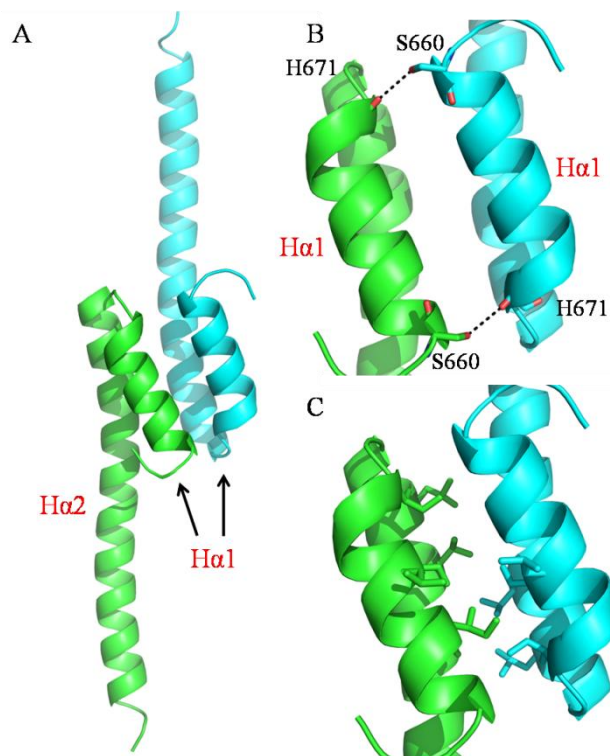


Figure 10 : Empilement cristallin de P_{XD} de NiV. **A)** Représentation en cartoon de P_{XD} de NiV (en vert) et sa molécule symétrique (en bleu). **B)** et **C)** zoom sur l'interface de contact entre les molécules symétriques de NiV P_{XD} . **B)** Le résidu Ser 660 et le groupement carbonyle de la liaison peptidique du résidu His 671 sont représentés en bâton. En pointillé noir les interactions polaires d'une distance inférieure à 3.5 \AA . **C)** Les résidus hydrophobes sont représentés en bâton.

Dans la structure cristallographique du P_{XD} de NiV, cette surface de fixation du N_{TAIL} n'existe donc pas, suscitant différentes questions. La structure du domaine P_{XD} de NiV est-elle dynamique et peut-elle s'échanger avec d'autres formes ? Comme le suggère la présence

de « dimères » dans l'empilement cristallin, existe-t-il des dimères en solution ? La forme cristallisée de la protéine est-elle la forme majoritaire de la protéine présente en solution ou est-elle au contraire une forme minoritaire piégée lors de la cristallisation ? Si elle est la forme majoritaire, la P de NiV utilise-t-elle un mode de fixation sur le N_{TAIL} différent de celui des autres paramyxovirus? Cette dernière suggestion est étonnante compte tenu du fait que le domaine P_X de NiV se fixe sur la région N_{TAIL} d'HeV et inversement (Habchi, Mamelli et al. 2010), mais afin de répondre à ces questions, nous avons entrepris d'une part une analyse détaillée des deux structures afin de comprendre le rôle joué par les acides aminés qui diffèrent entre les deux protéines, et d'autre part une étude détaillée des propriétés de ces domaines en solution.

Nous présentons les résultats des études en solutions obtenus en utilisant des P_{XD} portant aucun tag de purification. Il faut noter que des résultats équivalents ont été obtenus avec les domaines P_X en fusion avec l'étiquette hexa-histidine qui ont été utilisés pour la cristallisation (non montré).

3.3.2.3. Comparaison des structures cristallographiques de NiV et HeV P_{XD}

La simple comparaison des deux structures révèle que les hélices H α 2 et H α 3 du P_{XD} du virus Hendra correspondent à la longue hélice C-terminale (H α 2) du virus Nipah (**Figure 11B et C**).

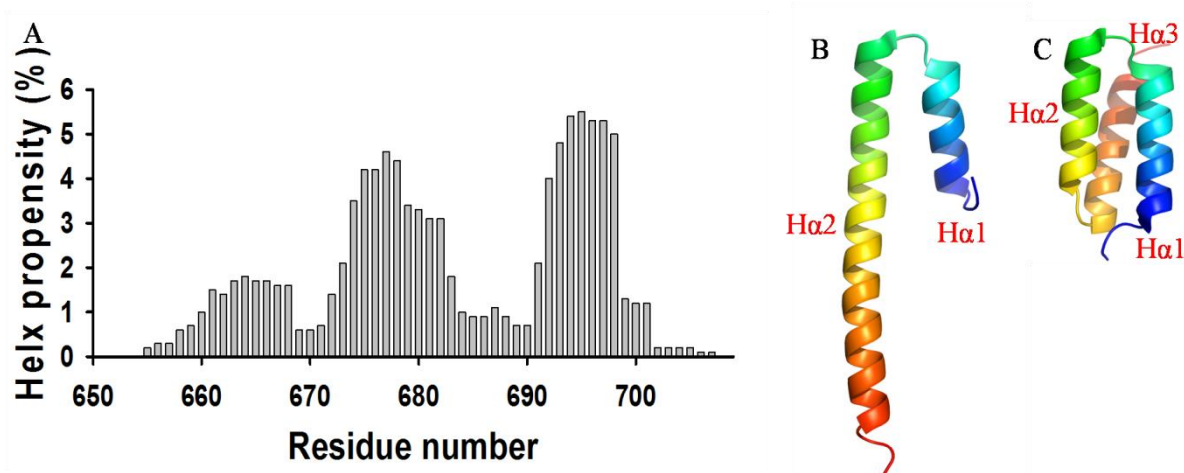


Figure 11 : Propension à former des hélices α pour le domaine P_X de NiV déterminée avec le logiciel AGADIR. A) Prédiction AGADIR. B) et C) Représentation en cartoon de NiV P_{XD} et HeV P_{XD}, respectivement, colorés du bleu (N-terminal) au rouge (C-terminal).

L'analyse de la séquence du P_{XD} de NiV avec le logiciel AGADIR (Lacroix, Viguera et al. 1998) prédit la présence de trois hélices avec un coude localisé au niveau des acides aminés

Y₆₈₄LNKAEN₆₉₀ au milieu de la longue hélice C-terminale (**Figure 11A**). Cette prédiction est basée sur l'analyse d'interaction locale, suggérant que la propagation de l'hélice dans cette région est due à des interactions tertiaires.

Afin de comprendre l'origine moléculaire de cette différence de structure, nous avons analysé l'environnement des acides aminés qui différaient entre les séquences des virus Nipah et Hendra (**Figure 12**). Il existe 10 à 12 substitutions entre les séquences des domaines P_X des différents isolats de Nipah et Hendra.

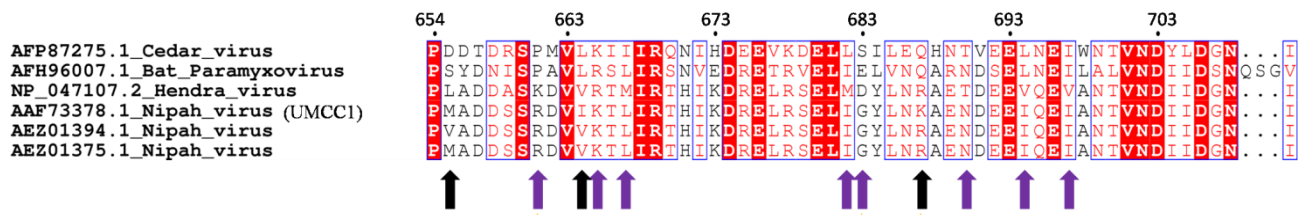


Figure 12 : Alignement de séquences de la région C-terminale (P_{XD}) des phosphoprotéines des *Henipavirus*. Représenté l'alignement de séquences des virus Cedar, paramyxovirus de chauve-souris (bat paramyxovirus), virus Hendra et trois séquences représentatives du virus Nipah. La numérotation est par rapport aux séquences de Nipah. Les résidus en blanc sur fond rouge sont identiques, les résidus en rouge sur fond blanc sont de nature similaire ou conservés >75%. Sur la figure sont marqués les résidus strictement conservés dans les différentes souches de Nipah (flèches en violet) mais non conservé chez Hendra et les résidus qui sont similaires mais pas conservés (flèches noirs) entre les souches de Nipah. Les résidus que nous avons muté sont marqués par une étoile en jaune (voir plus loin). L'alignement est fait en utilisant le programme MAFFT (Katoh and Standley 2013) et la coloration en fonction de la conservation par le programme ESPript (Gouet, Courcelle et al. 1999).

Trois résidus valine qui se situent dans l'interface entre l'hélice H α 3 et les deux autres hélices du HeV P_X (Val662, Val692 et Val695) sont remplacés par des résidus isoleucine dans la séquence de NiV P_{XD} que nous avons utilisé (Ile664, Ile694 et Ile697).

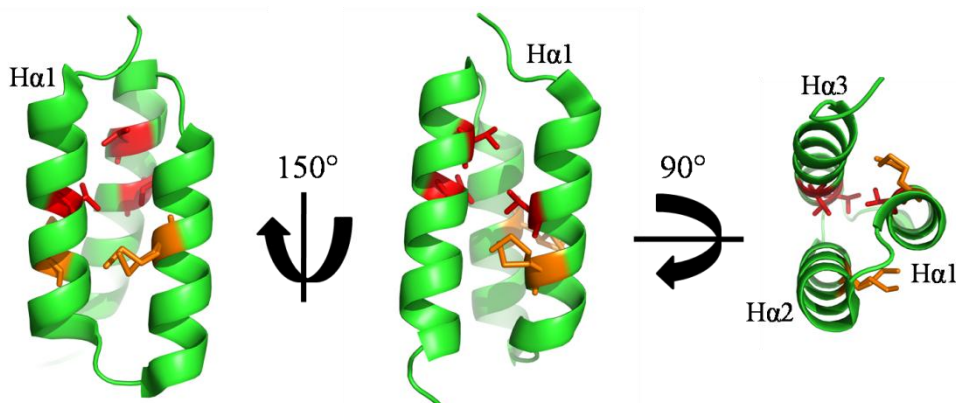


Figure 13 : Résidus hydrophobes non conservés entre NiV et HeV P_{XD}. Représentation en cartoon de HeV P_{XD}. Sont représentés en bâtons et colorés en rouge les résidus Val662, Val692 et Val695 et en orange Met665 et Met680 (double conformation) qui ne sont pas conservés entre les isolats de NiV et HeV P_{XD} étudiés.

Deux de ces trois résidus isoleucine sont strictement conservés dans les différents isolats de Nipah. Le volume plus important des résidus isoleucine et leur géométrie différente pourraient perturber l'arrimage de la troisième hélice sur la surface formée par les deux premières. De la même manière, deux résidus méthionine du HeV P_{XD} (Met665 et Met680)

qui sont localisés à la périphérie du cœur hydrophobe sont remplacés par des acides aminés à chaînes aliphatiques longues (Leu667 et Ile682) (**Figure 13**).

Dans le NiV P_{XD}, l'hélice H α 1 s'étend entre les résidus Ser660 et Ile672 et le résidu Ser659 constitue une coiffe pour l'extrémité N-terminale (N_{CAP}) de l'hélice H α 1 (**Figure 14F**). Cet arrangement permet à la Ser660 de se positionner correctement afin de former une liaison hydrogène avec le carbonyle de l'His671 (**Figure 10**). Dans HeV P_{XD} cette position est occupée par une Ala (Ala657) et non pas par une Ser ce qui résulte au gain de deux liaisons hydrogène dans la conformation observée pour P_{XD} de Nipah dans le cristal. Dans le HeV P_X l'hélice H α 1 commence au résidu Asp655 (3 résidus en plus que dans l'hélice H α 1 du NiV P_{XD}) dont la chaîne latérale constitue une coiffe pour l'extrémité N-terminale (N_{CAP}) de l'hélice pour une des molécules dans l'unité asymétrique (**Figure 14I**). Cette hélice H α 1 ne semble pas être coiffée dans la seconde molécule. Par contre, il est intéressant de noter que la boucle entre les hélices H α 1 et H α 2 ainsi que la partie C-terminale de l'hélice H α 1 et la partie N-terminal de l'hélice H α 2 sont strictement conservées. Dans les deux structures un résidu aspartate (D674 pour NiV et D672 pour HeV) forme un N_{CAP} pour l'hélice H α 2 en formant des liaisons hydrogène avec les groupements amide des résidus +2 et +3 ce qui positionne un résidu positivement chargé (Arg 678 pour NiV et Arg 676 pour Hev) en face des carbonyles du dernier tour de H α 1 et forme un C_{CAP} (**Figure 14A et 14B**). Dans le HeV P_X, un réseau de liaisons hydrogène stabilise la boucle entre les hélices H α 2 et H α 3 (**Figure 14H**). Notamment, le résidu Thr688 constitue une coiffe pour l'extrémité N-terminale (N_{CAP}) de l'hélice H α 3 en formant par l'intermédiaire du groupe OH de sa chaîne latérale une liaison hydrogène avec le groupe amide du résidu Glu691. Le remplacement du résidu Thr688 par le résidu Asn690 dans le domaine P_X de NiV pourrait fragiliser ce réseau et contribuer à la mise en place d'une structure alternative.

La substitution du résidu Gly683 de NiV P_X par un résidu Asp (Asp681) chez HeV, qui est la modification la moins conservative observée entre les domaines, ne semble pas avoir d'implication dans la stabilisation de la structure de l'un ou l'autre des domaines.

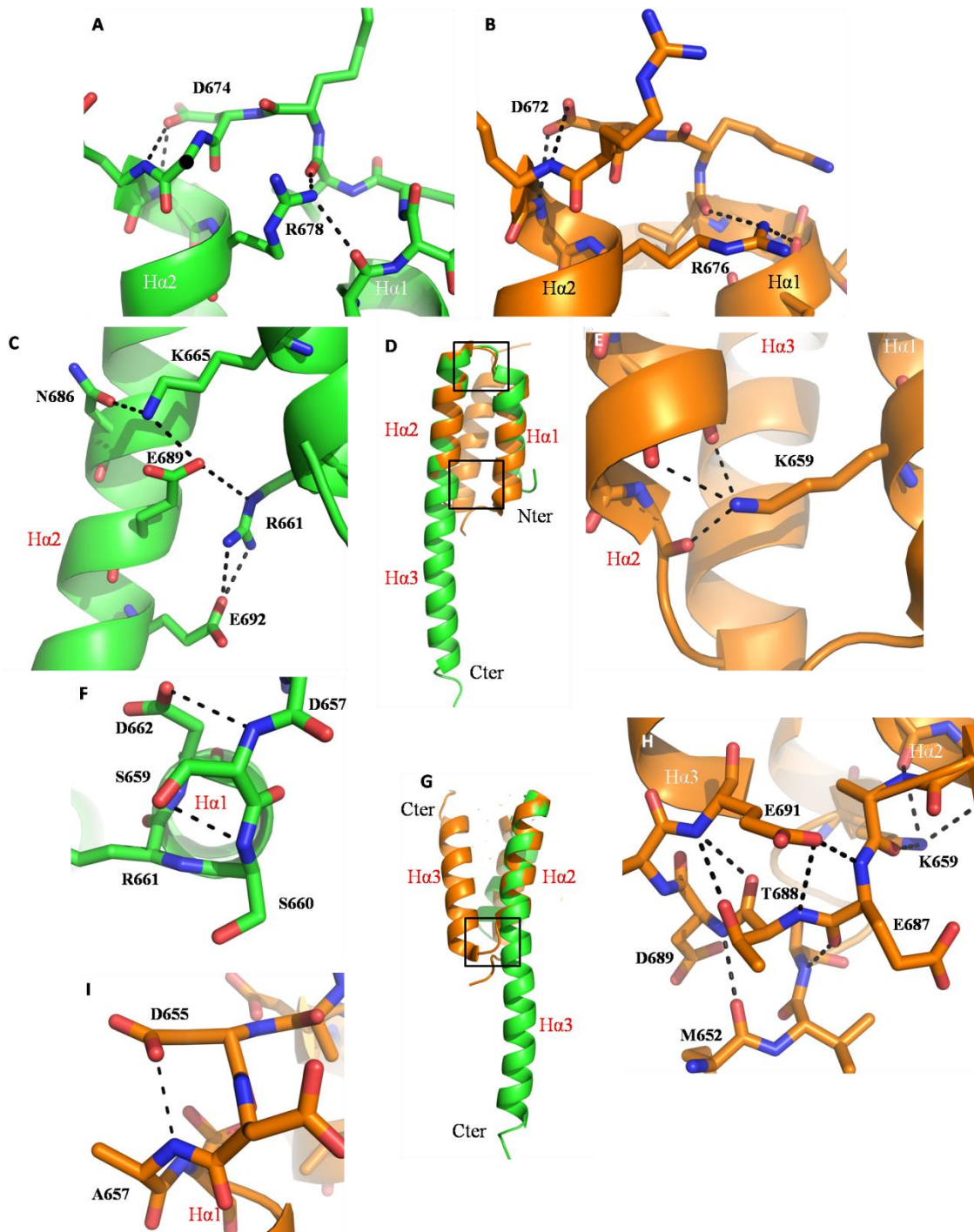


Figure 14 : Comparaison des interactions stabilisant les conformations des P_{XD} de NiV et HeV. En vert le P_{XD} de NiV et en orange le P_{XD} de HeV. Les lignes en pointillé noir montrent des interactions polaires d'une distance inférieure à 3.5 Å. **A)** et **B)** Les interactions entre les chaînes latérales et la chaîne principale stabilisant les hélices Ha1 et Ha2 de NiV et HeV P_{XD}, respectivement. Le résidu Asp 674/672 forme un N_{CAP} pour Ha2 tandis que le résidu Arg 678/676 forme un C_{CAP} pour Ha1, respectivement. **C)** et **E)** Les interactions stabilisant la forme en deux hélices de NiV et la forme en trois hélices de HeV, respectivement. Rôle central des résidus Arg 661 pour NiV et Lys 659 pour HeV. **D)** Superposition des P_{XD} de NiV et HeV sur les hélices Ha1 et Ha2. L'encadré du haut montre le positionnement des zooms de A) et B) tandis que l'encadré du bas montre le positionnement des zooms de C) et E). **F)** Stabilisation du côté N-terminale de Ha1 par les résidus Asp 662 et Ser 659 formant un N_{CAP}. **H)** Interactions stabilisant d'avantage la forme en trois hélices du P_{XD} de HeV. **G)** Même superposition que D) mais tournée de 90°. L'encadré montre le positionnement du zoom H). **I)** Stabilisation du côté N-terminale de Ha1 par le résidu Asp 655.

Deux résidus arginine et un résidu lysine d'HeV P_{XD} (Lys659, Arg663 et Arg685) sont remplacés respectivement par deux résidus lysine (Lys665 et Lys687) et un résidu arginine (Arg661) dans NiV P_{XD}. Dans certains isolats de Nipah le résidu à la position 687 est une lysine tandis que chez d'autres souches c'est une arginine. Dans la séquence de NiV P_{XD} que nous avons étudié (souche UMCC1) le résidu 687 est une lysine tandis que ce résidu est une arginine chez HeV P_{XD}.

Dans la structure de NiV P_X, le résidu Lys687 se situe au milieu de la longue hélice H α 2 et sa chaîne latérale pointe vers le solvant. Dans la structure du HeV P_X, l'Arg685 correspondante est localisée dans le dernier tour de l'hélice H α 2, et sa chaîne latérale pointe également vers le solvant. Les résidus Arg661 et Lys665 de NiV P_{XD} sont quant à eux impliqués dans un réseau de liaisons hydrogène et de ponts salins entre l'extrémité N-terminale de l'hélice H α 1 et une région située au centre de l'hélice H α 2 (**Figure 14C**). Cette région correspond précisément à la région non-prédite en hélice par le logiciel AGADIR (Y₆₈₄LNKAEN₆₉₀) (**Figure 11**), qui forme la boucle entre les hélices H α 2 et H α 3 dans le HeV P_{XD} (**Figure 14H**) et se prolongeant par un groupe de trois acides aminés chargés négativement D691EE. Le résidu Arg661 forme des contacts polaires avec les résidus Glu689 et Glu692, et le résidu Lys665 interagit avec les résidus Asn686 et Glu689 (**Figure 14C**). Dans la structure de HeV P_{XD}, le résidu Lys659 forme des liaisons hydrogène avec les carbonyles des résidus Leu683, Asn684 et Ala686 formant une coiffe du côté C-terminal de l'hélice H α 2 (**Figure 14E**). Les résidus Glu687 et Glu690, qui correspondent aux résidus Glu689 et Glu692 de NiV P_{XD} sont orientés différemment, faisant face au solvant et ne formant pas d'interactions directes avec d'autres résidus.

SEC-MALLS-RI

Les domaines P_X de NiV et d'HeV éluent de la colonne S75 sous la forme d'un seul pic. Leur rayon hydrodynamique (NiV: Rh 2.0 ± 0.1 nm et HeV: Rh 2.0 ± 0.1 nm) est supérieur à celui attendu pour des protéines compactes de même masse moléculaire. Les deux protéines éluent à des volumes inférieurs au volume d'éluion de la RNase A (124 aa, Rh 1.9 nm) dont la masse moléculaire est largement supérieure (**Figure 15A**). Ce résultats suggèrent que dans ces conditions, à pH 7.5 et en présence de 150 mM NaCl, les deux protéines sont allongées ou partiellement flexibles. Il est intéressant de noter que l'augmentation de la concentration en sel provoque une diminution du Rh et donc une compaction des deux P_{XD} tandis qu'à faible concentration en sel le Rh augmente d'avantage et donc les deux domaines

semblent être plus allongés (tableau **Figure 15**). Il faut aussi noter que P_{XD} de NiV élue de la colonne systématiquement à un volume plus petit que P_{XD} d'HeV et est donc légèrement plus étendu. La masse moléculaire déterminée par SEC-MALLS-RI est de 7.5 ±1 kDa pour le P_{XD} d'HeV et de 7.6 ±1 kDa pour le P_{XD} de NiV, en accord avec la masse moléculaire calculée à partir de la séquence en acides aminés (6.3 kDa) et ne varie pas à travers les pic.

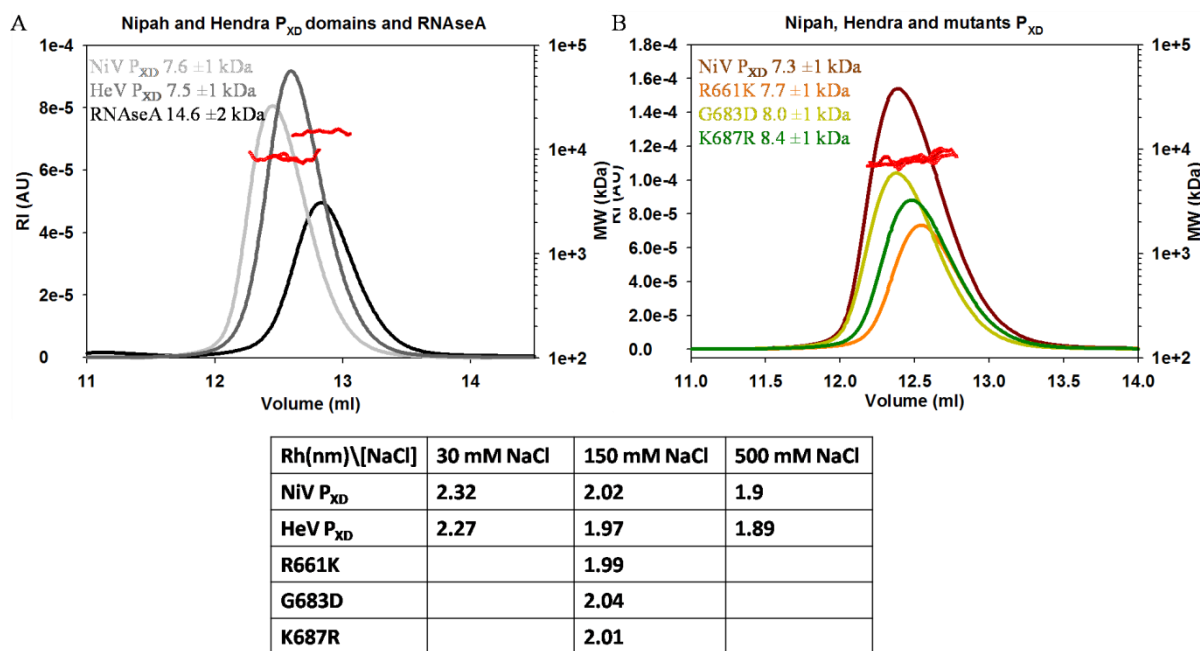


Figure 15 : Masse moléculaire et rayons hydrodynamiques (Rh) de NiV et HeV P_{XD} et les mutants R661K, G683D et K687R, mesurées par SEC-MALLS-RI. A) Profils d'élution de NiV P_{XD} (gris claire), HeV P_{XD} (gris foncé) et la RNase A (noir) de la colonne de gel filtration Superdex S75, équilibrée dans 20 mM Tris-HCl pH 7.5, 150 mM NaCl. B) Profils d'élution de NiV P_{XD} (marron), R661K (orange), G683D (vert claire) et K687R (vert foncé) dans les mêmes conditions que A). En rouge les masses moléculaires (MM) à travers les pics d'élution, calculées à partir de la diffusion statique de la lumière et la réfractométrie pour chaque protéine. Les valeurs numériques des MM ont le même code couleur que les profils d'élution. Dans le tableau sont représentés les rayons hydrodynamiques (Rh) mesurés à différentes concentrations de sel (30, 150 ou 500 mM NaCl) à pH 7.5. Les Rh ont été déterminé à partir des profils d'élution en calibrant le système avec des protéines standards (Albumine, Ovalbumine, Chymotrypsine et RNase A).

Diffusion des rayons X aux petits angles (SAXS) – effet de la concentration

Les courbes de diffusion des domaines P_{XD} de NiV et d'HeV ont été enregistrées en présence de 50 mM Arg et 50 mM Glu, puisque ces additifs semblent protéger ces protéines des dégâts causés par l'irradiation aux rayons X (voir **Chapitre 2 - Matériels et Méthodes**). Alors que la forme globale de la courbe de SAXS et le rayon de giration (R_g) du HeV P_{XD}, déterminée aux très petits angles selon l'approximation de Guinier, ne varient que très peu avec la concentration en protéine (**Figure 16, ronds rouges**), le R_g et l'intensité à angle zéro (I(0)) pour le NiV P_{XD} augmentent avec la concentration de protéine (**Figure 16, ronds noirs**). Le changement du R_g avec la concentration peut indiquer un changement de

conformation tandis que l'augmentation d' $I(0)$ suggère une oligomérisation ou une agrégation. Nous pouvons éliminer la possibilité que cette variation de masse moléculaire soit liée à des dégâts causés par l'irradiation, puisque l'analyse de la protéine récupérée à la fin de l'expérience n'a révélé aucune trace d'agrégation par SEC-MALLS-RI et que la concentration correspondait à celle mesurée avant l'expérience (non montré). Cette dépendance à la concentration est réduite à une force ionique plus élevée comme le montre la comparaison des courbes de SAXS mesurées en présence de 30 mM NaCl et de 500 mM NaCl (**Figure 16**).

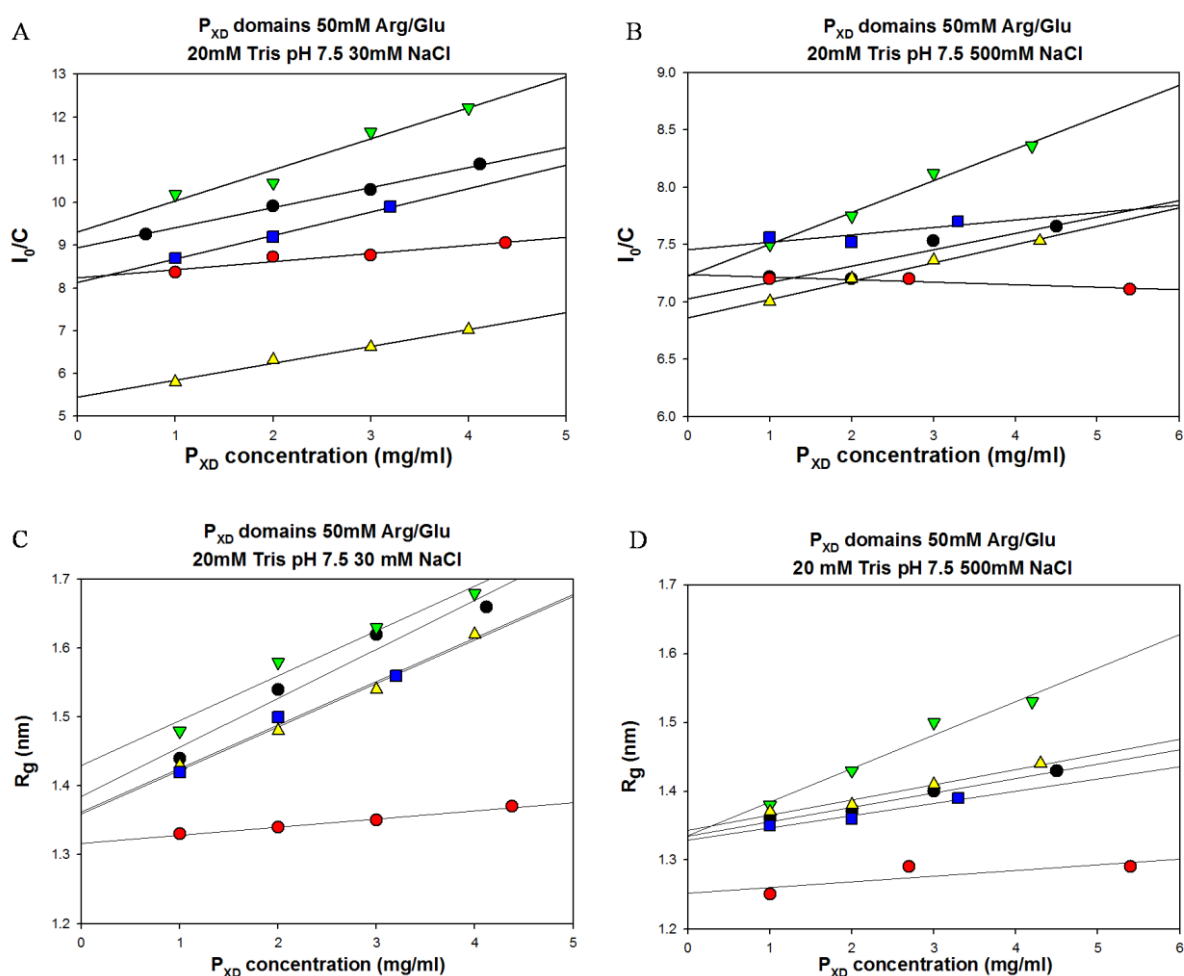


Figure 16 : Expériences de diffusion des rayons X aux petits angles (SAXS) pour NiV et HeV P_{XD} et les mutants R661K, G683D et K687R. A) et B) graphes de l'intensité à angle zéro ($I(0)$), normalisé pour la concentration en protéine, en fonction de la concentration en protéine en présence de 30mM NaCl ou 500mM NaCl, respectivement. C) et D) graphes du rayon de giration (R_g) en fonction de la concentration en protéine en présence de 30mM NaCl ou 500mM NaCl, respectivement. NiV P_{XD} (ronds noirs), HeV P_{XD} (ronds rouges), G683D (triangles verts), K687R (triangles jaunes) et R661K (carrés bleus).

Concernant l' $I(0)$ nous nous limitons à l'analyse des pentes des graphes d' $I(0)$ en fonction de la concentration en protéine car nous ne connaissons par les valeurs absolues du $I(0)$ avec une très grande précision.

Diffusion des rayons X aux petits angles (SAXS) - modélisation

La courbe de SAXS calculée à partir de la structure cristalline du NiV P_{XD} est différente des courbes de diffusion expérimentales enregistrées en présence de 150 mM et 500 mM NaCl (**Figure 17**). Un modèle de NiV P_{XD} plus compact, comprenant trois hélices en faisceau, a été construit par homologie avec le domaine HeV P_{XD} en utilisant le serveur SWISS-MODEL. La courbe de SAXS calculée pour ce modèle est également différente de la courbe expérimentale. Néanmoins, un bon ajustement est obtenu en combinant linéairement les deux courbes théoriques. La courbe en présence de 500 mM NaCl est correctement reproduite par un mélange de 80% conformation en trois hélices et de 20% conformation en deux hélices alors que la courbe en présence de 150 mM NaCl est reproduite un mélange presque équimolaire (respectivement 47% et 53%).

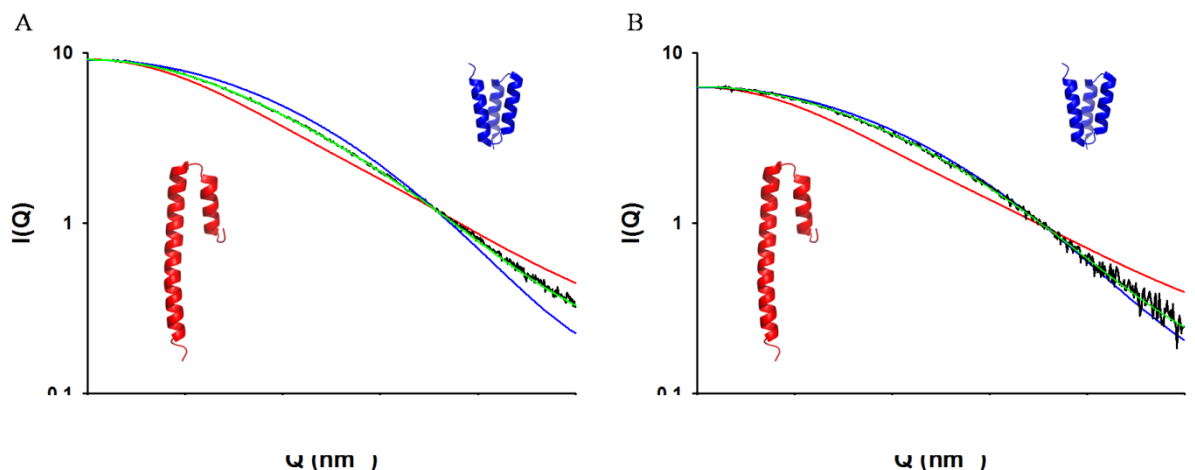


Figure 17 : Profils de SAXS pour NiV P_{XD} à différentes concentrations de sel. En noir le profil expérimental de NiV P_{XD} à **A**) 30mM NaCl et **B**) 500mM NaCl. En vert le profil de diffusion théorique obtenu en combinant les profils de diffusion calculés à partir de la forme cristalline à deux hélices (rouge) et la forme en trois hélices (bleu). Cette forme en trois hélices est modélisée par homologie en utilisant la forme cristalline de HeV P_{XD}.

De manière similaire, la courbe théorique calculée à partir de la structure du P_{XD} d'HeV ne reproduit pas parfaitement la courbe de SAXS expérimentale obtenue avec le P_{XD} d'HeV dans 150 mM de NaCl. Encore une fois, la courbe expérimentale peut être reproduite par une combinaison linéaire des deux courbes théoriques (15% de la courbe calculée avec la structure de NiV P_{XD} et 85% des la courbe calculée avec la structure d'HeV P_{XD}) (non montré). L'ajustement de la courbe expérimentale en combinant les courbes théoriques des deux formes monomériques et de la courbe obtenue pour la forme dimérique trouvée dans l'empilement cristallin reproduit moins bien les résultats expérimentaux. Ces résultats

suggèrent que les domaines P_X de NiV mais également d'HeV ne sont pas stables en solution mais existent dans au moins deux formes monomériques en échange l'une avec l'autre.

Résonance Magnétique Nucléaire (RMN)

Afin d'examiner les propriétés structurales de NiV et HeV P_{XD} en solution et à niveau atomique nous avons entrepris une caractérisation par la technique de Résonance Magnétique Nucléaire (en collaboration avec l'équipe du Dr. Martin Blackledge, IBS, Grenoble). L'attribution des résonances dans les spectres RMN de NiV et HeV P_{XD} a permis de localiser avec précision les éléments de structure secondaire en solution et en particulier les hélices α (**Figure 18**). Les valeurs de déplacements chimiques des C α et C β sont utilisées pour calculer un paramètre de propension à faire des structures secondaires (« secondary structure propensity », SSP) (Marsh, Singh et al. 2006).

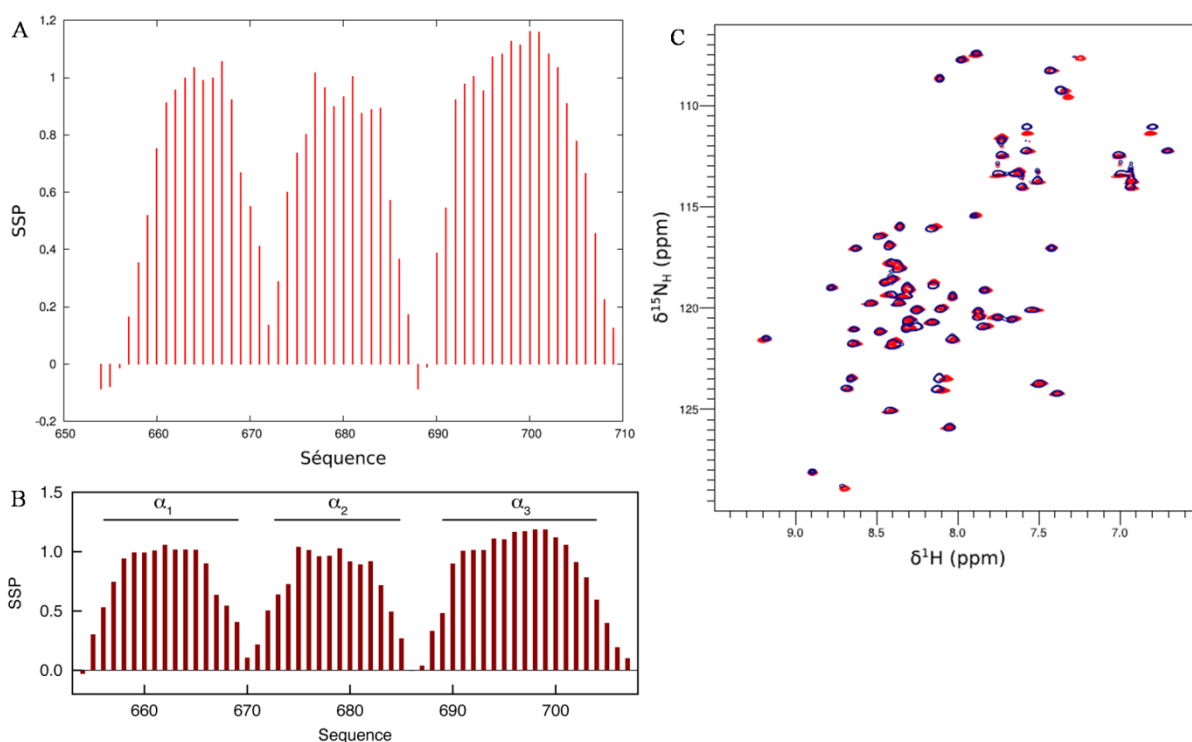


Figure 18 : Expériences de RMN sur NiV et HeV P_{XD}. A) et B) Graphique représentant la propension à former des structures secondaires (Secondary Structure Propensity, SSP) en fonction du numéro des résidus dans la séquence pour NiV et HeV P_{XD}, respectivement. Le SSP est calculé en combinant les valeurs des déplacements chimiques pour les C α et C β , obtenues expérimentalement. C) Superposition des spectres 2D [¹H-¹⁵N]-HSQC de NiV P_{XD} à 150 mM (bleu) et 500 mM (rouge) NaCl.

Les résultats concernant HeV P_{XD} démontrent qu'en solution ce domaine contient trois hélices α ce qui est en accord avec la conformation observée dans le cristal (**Figure 18B** et

Communie, G. et al. dans ANNEXES). Les résultats obtenus pour NiV P_{XD} montrent trois régions du domaine pour lesquels des résidus contigus présentent des valeurs positives indiquant la présence de trois hélices α (**Figure 18A**). Alors en solution NiV P_{XD} semble contenir trois hélices α en accord avec la prédiction d'AGADIR mais en contradiction avec la structure cristallographique. De plus l'ajout de sel n'a pas un effet notable sur le comportement de NiV P_{XD} car les spectres 2D [¹H-¹⁵N]-HSQC, enregistrés à 150 et 500 mM NaCl et à 25°C, sont presque identiques (**Figure 18C**).

Stabilité thermodynamique des domaines P_X de NiV et HeV

Le dichroïsme circulaire (CD) dans l'UV lointain (< 250 nm) renseigne sur le contenu moyen en structures secondaires des protéines et permet de suivre les réactions de dénaturation. Les spectres dans l'UV lointain enregistrés pour les domaines P_X de NiV et HeV sont caractéristiques de protéines riches en hélices α (**Figure 19**), contrairement à ce qui a été observé pour le P_{XD} du virus des oreillons (Kingston, Gay et al. 2008).

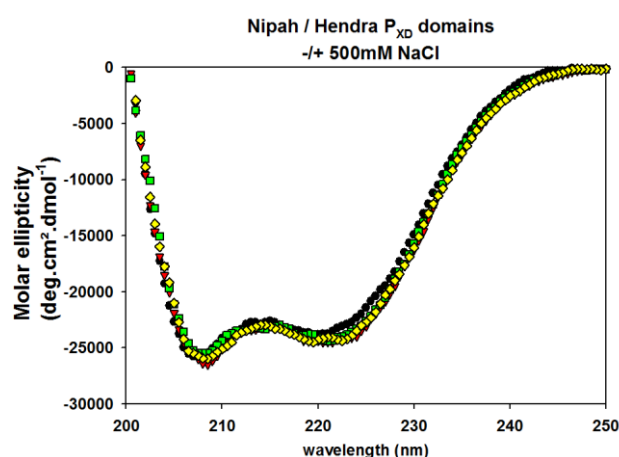


Figure 19 : Spectres de Dichroïsme Circulaire (CD) pour NiV et HeV P_{XD} Spectres dans l'UV lointain des NiV et HeV P_{XD} dans 10 mM NaP pH 7.5 et à 0 ou 500 mM NaCl. Ronds noirs et triangles rouges NiV et HeV P_{XD} à 0 mM NaCl, respectivement. Carrés verts et diamants jaunes NiV et HeV P_{XD} à 500 mM NaCl, respectivement. La concentration de protéine est de 20 μ M.

En supposant que le signal à 222 nm est uniquement dû à la présence d'hélices α , le contenu moyen en hélice est estimé à environ 66 % dans les deux protéines, en accord avec le contenu en hélices α dans les structures cristallines. L'ajout de 2 M TMAO ne modifie pas le spectre CD (non montré) tandis que l'ajout de 500 mM NaCl augmente d'environ 4% le contenu en hélice pour NiV P_{XD} (**Figure 19**).

La dénaturation thermique des domaines P_{XD} a été étudiée en enregistrant le signal CD à 222 nm tout en faisant varier continuellement la température de 4 °C à 95 °C. La réversibilité de la réaction est établie en suivant le signal en variant la température en sens

inverse. Les courbes de CD à 222 nm en fonction de la température enregistrées au cours de la dénaturation et de la renaturation sont superposées indiquant la parfaite réversibilité de la réaction (non montré).

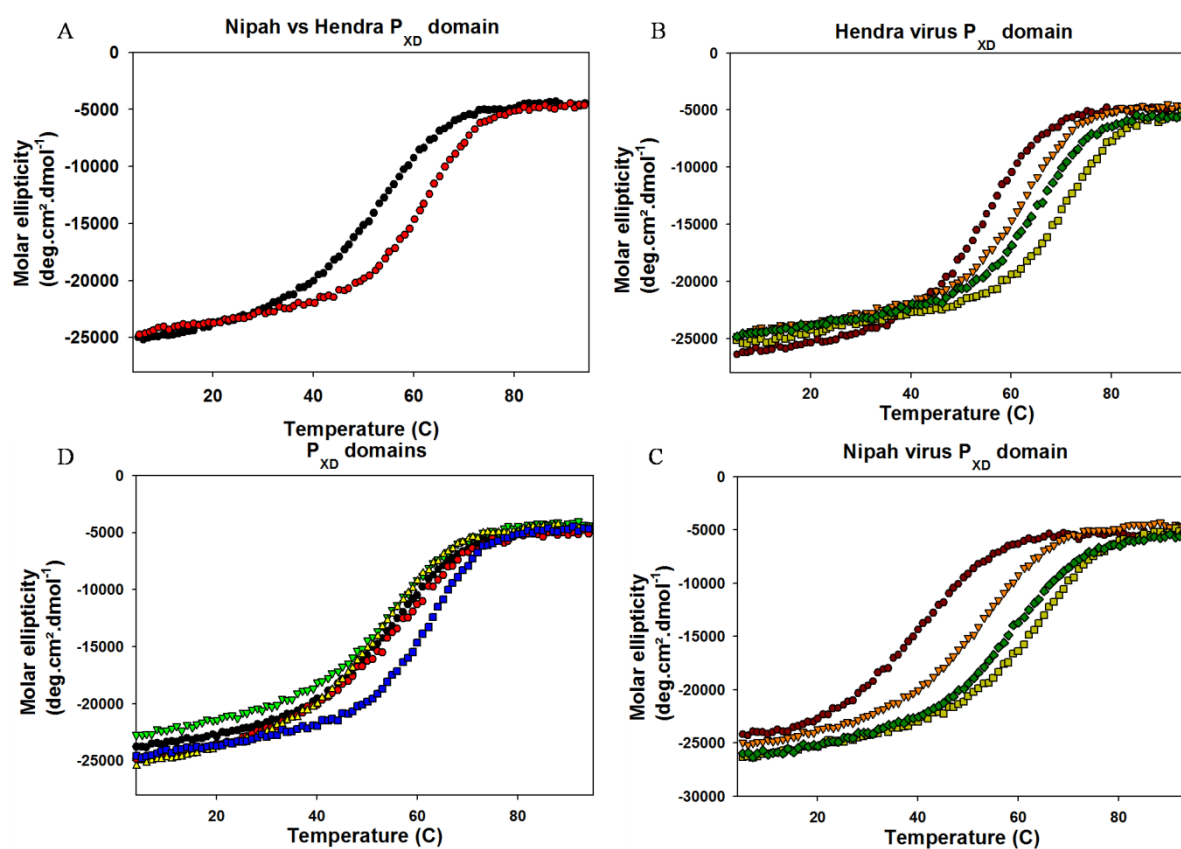


Figure 20 : Dénaturation thermique pour NiV et HeV P_{XD} et les mutants R661K, G683D et K687R. A) Superposition des courbes de dénaturation de HeV et NiV P_{XD} dans 150 mM NaCl et pH 7.5, coloré en rouge et noir, respectivement. B) et C) Dénaturations thermiques de HeV et NiV P_{XD} , respectivement. 20 mM Tris-HCl pH 7.5 et 0 mM NaCl (ronds marron), 150 mM NaCl (triangles orange), 500 mM NaCl (carrés vert clair) et 20 mM NaAc pH 2.0 et 0 mM NaCl (diamants vert foncé). D) Dénaturations thermiques à pH 7.5 et 150 mM NaCl de NiV (triangles jaunes) et HeV P_{XD} (carrés bleus) et les mutants R661K (ronds noirs), G683D (ronds rouges) et K687R (triangles inversés verts). Dénaturations réalisées à une concentration en protéine de 20 μM .

Les résultats à pH 7.5 en présence de 150 mM montrent une différence importante de T_m et entre les deux protéines (**Figure 20A**). Le NiV P_{XD} se caractérise par un T_m d'environ 52 °C, le HeV P_{XD} par un T_m d'environ 62°C. L'effet observé de la force ionique sur le comportement hydrodynamique des P_{XD} de NiV et d'HeV nous a amené à étudier également l'effet de la concentration en sel sur la stabilité thermodynamique. En absence de sel, le T_m pour le P_{XD} de NiV est réduit à 40 °C alors qu'il augmente jusqu'à 60 °C en présence de 500 mM NaCl (**Figure 20C**). Cette variation de la stabilité avec la concentration de sel suggère un effet de la force ionique par écrantage des charges. A pH 7.5 les deux domaines sont chargés négativement (NiV P_{XD} pI 4.6, HeV P_{XD} pI 4.5) et la présence de groupes de charges de mêmes signes à la surface des deux domaines, notamment des groupes de charges négatives à l'extrémité N-terminale de H α 1 et au milieu de l'hélice H α 2 de NiV P_{XD} , pourraient produire

des répulsions qui en présence de sel seraient diminuées. Cette explication est supportée par l'augmentation significative du T_m à pH 2.0 ($T_m = 57\text{ °C}$) en absence de sel par rapport à la valeur mesurée à pH 7.5 (**Figure 20C**).

Les mêmes variations de la stabilité thermique avec la concentration de sel et avec le pH sont observées pour le domaine HeV P_X , indiquant un effet similaire de la répulsion de charges sur la stabilité (**Figure 20B**).

Mutations dans le domaine P_X du virus Nipah

Afin d'identifier clairement les résidus responsables de la différence de comportement entre les deux domaines, notamment du rayon hydrodynamique et de la stabilité thermique, nous avons entrepris une étude par mutagenèse dirigée. Dans une première approche, trois variantes du domaine P_X de NiV ont été générés, exprimés, purifiés et caractérisés.

Nous avons choisi de tester le mutant Arg661→Lys, car la modélisation d'une Arg en position 659 dans la structure en trois hélices de HeV P_{XD} , en remplacement du résidu Lys, provoque une interpénétration dans l'hélice H α 2, suggérant que le résidu Arg661 défavoriserait la conformation en trois hélices dans le P_{XD} de NiV. Le second mutant, Gly683→Asp a été utilisé comme contrôle puisqu'il s'agit de la variation la moins conservative observée entre les deux domaines. Ces résidus sont exposés au solvant et ne semblent pas former d'interactions spécifiques susceptibles d'influencer la structure ou la stabilité thermodynamique. Finalement, nous avons étudié le variant Lys687→Arg car, d'une part chez Hev le résidu lysine est remplacé par une arginine et d'autre part, une partie des isolats de Nipah contient une arginine à cette position.

L'analyse des mutants par SEC-MALLS-RI (**Figure 15B et tableau**) et SAXS (**Figure 16**) indique que la mutation Lys687→Arg n'a aucun effet significatif. La mutation Arg661→Lys provoque une réduction du rayon hydrodynamique mesuré par SEC et du rayon de giration mesuré par SAXS ce qui montre, comme nous l'avions supposé, que l'Arg661 est un des résidus du P_{XD} de NiV qui est responsable de la différence de comportement entre les deux protéines. Néanmoins la stabilité thermique ne semble pas être modifiée de façon significative (**Figure 20D**). Cette mutation n'est toutefois pas suffisante pour reproduire le comportement du P_{XD} d'HeV, indiquant donc que d'autres résidus ont également un effet sur la taille du domaine et sa stabilité thermodynamique. Quant au mutant Gly683→Asp, il était inattendu d'observer des différences de comportement. En effet, à pH 7.5 en présence de 30

mM NaCl, le mutant Gly683→Asp a un rayon de giration mesuré par SAXS comparable à celui du domaine sauvage, mais en présence de 500 mM NaCl, le rayon de giration et l'intensité à angle zéro ($I(0)$) sont supérieurs aux valeurs de la protéine sauvage et la dépendance à la concentration de protéine est plus forte que pour la protéine sauvage.

3.3.2.4. Détails sur la cristallisation et résolution des structures des domaines P_{XD} des virus Nipah et Hendra

Pendant les premiers tests de cristallisation en utilisant le robot de HTXLab, NiV P_{XD} a cristallisé dans 0.2 M MgCl₂, 25% PEG 3350, 0.1 M BisTris pH 6.0 à une concentration de protéine de 15 à 30 mg/ml. Pour ce domaine les conditions du robot ont été reproduites à la main ce qui n'a pas nécessité d'optimisation importante (**Figure 21A**). Les premiers cristaux obtenus ont diffracté à une résolution d'environ 2.5Å (ESRF ID 23-2 microfocuse) mais le remplacement moléculaire avec des structures connues de P_{XD} (faisceau de trois hélices α) n'a pas donné des solutions satisfaisantes. Le fait que le remplacement moléculaire avec des structures des P_{XD} connues ne marchait pas était très étrange car, *à priori*, le P_{XD} de Nipah serait aussi en forme de faisceau de trois hélices. La séquence du domaine Px ne contient ni méthionine, qui pourrait être remplacé par sélénométhionine et utilisé pour obtenir des phases initiales, ni cystéine, dont l'atome de soufre permettrait un phasage également expérimentale. Une solution possible à ce problème consistait à introduire un résidu Cys ou Met dans la séquence de NiV P_{XD}. Nous avons changé la leucine 667 par une méthionine, par mutagenèse dirigée mais la protéine mutante n'a pas cristallisé.

Pendant longtemps la dérivation aux métaux lourds a été largement utilisé pour obtenir des phases initiales. Pour essayer d'incorporer des métaux lourds, nous avons effectué des trempages des cristaux de la protéine native dans des solutions contenant ces métaux (Pb, Pt, Au, Sm, Nd, Eu, Th) ce qui a résulté soit en la fracturation des cristaux, soit en la perte de diffraction. La stratégie suivante que nous avons abordé a consisté à utiliser non pas des métaux lourds, qui demandent des temps de trempage parfois assez longs et qui interagissent avec certains résidus de la protéine assez fortement ce qui peut casser le packaging cristallin, mais des halogénures (voir **Chapitre 2 – Matériels et Méthodes**).

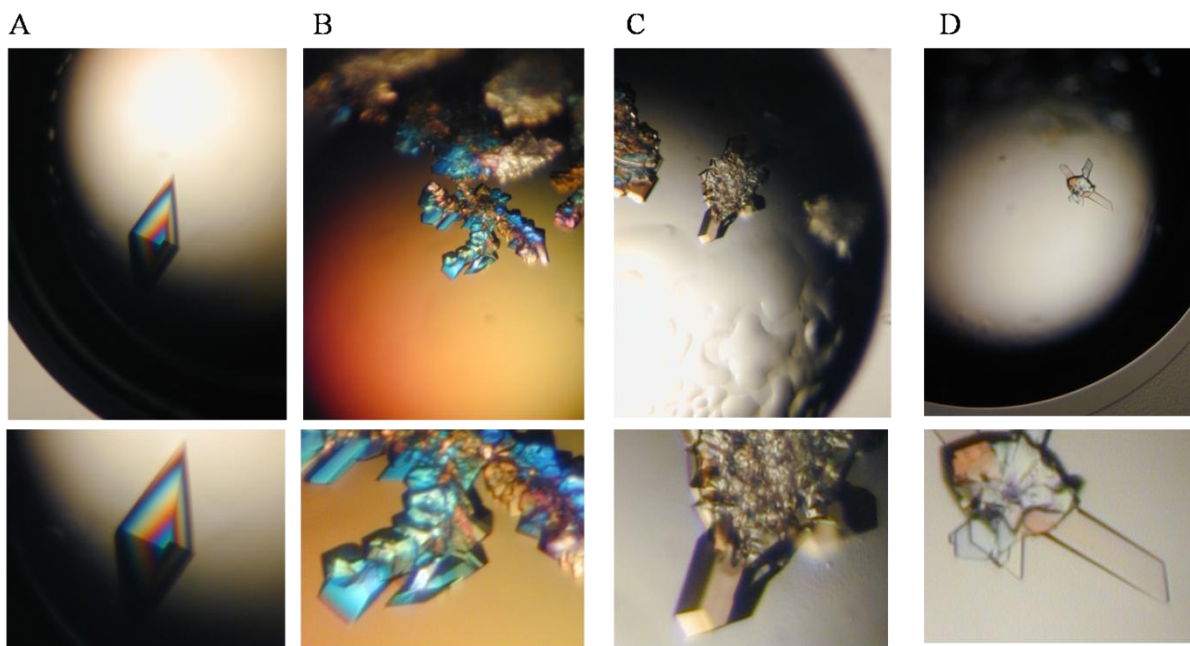


Figure 21: Crystallisation de NiV et HeV P_{XD}. **A)** Cristal typique de NiV P_{XD} obtenu dans 0.2 M MgCl₂, 25% PEG 3350, 0.1 M BisTris pH 6.0 et [P_{XD}] 15 à 30 mg/ml **B)** Cristaux de HeV P_{XD} sélénié obtenus dans 0.2 M MgCl₂, 28% PEG 3350, 0.1 M BisTris pH 6.0 et [P_{XD}] 25 mg/ml. **C)** et **D)** Micro ensemencement (microseeding) en utilisant les cristaux de B). Le précipitant contient 0.2 M MgCl₂, 28% PEG 3350, 0.1 M BisTris pH 6.0 et [P_{XD}] 12 et 6 mg/ml, respectivement pour C) et D). Le volume des gouttes est de 2 μ L.

Nous avons réalisé plusieurs trempages dans le but d'avoir des cristaux qui gardent le pouvoir de diffraction mais possèdent aussi des ions fixés sur la surface de chaque molécule de protéine dans le cristal. Les données de diffraction des cristaux dérivés, après trempages (15-30 secondes) dans des solutions contenant 0.7 M CsCl, 0.5 KBr et 0.5 M NaI, présentaient du signal anomal mais des cartes de densité électronique interprétables n'ont pas pu être obtenues. Seule la méthode SIRAS a permis d'obtenir une carte de densité électronique de bonne qualité (**Figure 9A et 9B**).

Entre temps nous nous sommes intéressés au domaine P_{XD} du virus Hendra qui a 80% d'identité avec le domaine correspondant de Nipah. Le P_{XD} de Hev contient deux méthionines dans sa séquence ce qui peut énormément faciliter la résolution de sa structure. Nous avons cristallisé le P_{XD} de Hev en utilisant des conditions très proches de ceux qui ont permis la cristallisation de NiV P_{XD}. Les premiers cristaux obtenus n'étaient pas utilisables car multiples et donc nous avons obtenu des meilleurs cristaux par la technique de microseeding en utilisant les cristaux obtenus initialement comme source de nucléations stables (**Figure 21B-D**). Brièvement les cristaux sont trempés dans une solution de stabilisation et puis cassés à l'aide de la bille « Seed Bead » de Hampton Research. Ensuite les semences cristallines ainsi obtenues sont diluées et utilisées afin d'ensemencer des gouttes fraîches. Ces cristaux améliorés ont diffracté à environ 1.5 Å de résolution. Néanmoins leurs paramètres de la maille

crystalline se sont avérés différents de ceux de P_{XD} de Niv ce qui suggérait déjà des différences dans la structure. 12 jours après l'arrivée du gène synthétique nous avons pu obtenir des phases de très bonne qualité pour le P_{XD} de Hev (**Figure 7A**) et résoudre sa structure par la méthode SAD qui, en fin de compte consiste en un faisceau de trois hélices α et est très similaire aux structures de P_{XD} d'autres virus de la famille comme MeV, MuV et SeV (**Figure 8**).

3.3.3. La phosphoprotéine de NiV forme des tétramères

3.3.3.1. Etat d'oligomérisation de la P entière en solution

L'analyse de la protéine entière et des différents fragments par chromatographie d'exclusion de taille combinée à une détection par diffusion statique de lumière (SEC-MALLS-RI) indique clairement que la protéine P de NiV forme des tétramères en solution et que la région allant du résidu 470 au résidu 580, correspond au domaine de tétramérisation (Figure 22).

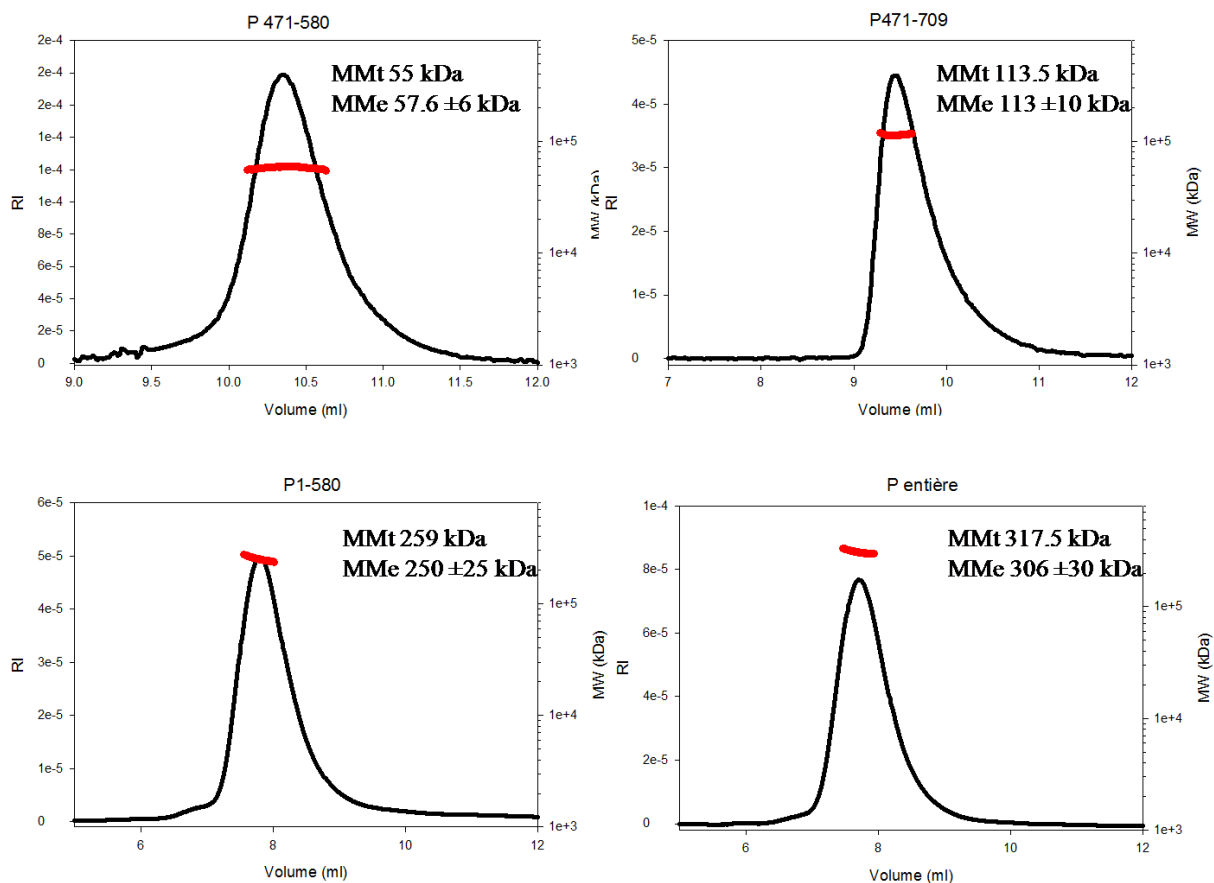


Figure 22 : Analyse par SEC-MALLS-RI des fragments P₁₋₅₈₀, P₄₇₁₋₇₀₉, P₄₇₁₋₅₈₀ ainsi que P entière. Profils d'élution des fragments et de la P entière (ligne en noir) de la colonne de gel filtration TSK G4000 (TOSOH), équilibrée dans 20 mM Tris-HCl pH 7.5, 150 mM NaCl. En rouge les masses moléculaires à travers les pics d'élution, calculées à partir de la diffusion statique de la lumière et la réfractométrie pour chaque protéine. MMt : masse moléculaire théorique correspondant à un degré d'oligomérisation tétramérique, calculée à partir de la séquence en acides aminés de la construction. MMe : masse moléculaire moyenne expérimentale.

La protéine NiV P entière élue d'une colonne TSK G4000 (TOSOH) sous la forme d'un seul pic à un volume de 7.9 mL. La masse moléculaire moyenne (MM) expérimentale de 306 ± 30 kDa est proche de la masse moléculaire calculée pour un tétramère (4 x 79.4 kDa =

317.5 kDa) et ne varie pas à travers le pic d'éluion indiquant que l'échantillon est monodisperse (**Figure 22D**). De la même manière, les différents fragments de P éluent sous la forme d'un seul pic. Le fragment correspondant au domaine central isolé (aa 471-580), ainsi que les fragments 1-580 et 471-709, qui étaient supposés de contenir également le domaine central, ont également des masses moléculaires correspondant aux tétramères respectifs (**Figure 22A-C, Tableau 3**). Il faut noter que la construction P₁₋₅₈₀ a un volume d'éluion très proche de celui de la P entière. Par contre, la masse moléculaire mesurée pour les peptides correspondant à différentes régions de la partie N-terminale, P₁₋₁₀₀, P₉₁₋₁₉₀, P₉₁₋₂₅₀, P₁₇₀₋₂₄₀, P₂₂₂₋₃₁₉, P₃₀₀₋₄₀₁, P₃₈₇₋₄₇₉, et de la région C-terminale de la protéine, P₅₈₈₋₆₅₀ et P₅₈₈₋₇₀₉, indique qu'ils sont monomériques et monodisperses (**Tableau 3**).

Le volume d'éluion de ces différents fragments indique des rayons hydrodynamiques plus grands que les rayons hydrodynamiques prédits pour des protéines globulaires de même masse moléculaire, suggérant que ces fragments de P sont allongés ou désordonnés en solution (non montré).

Tableau 3: Masses moléculaires calculées à partir des séquences ou obtenues par MALLS-SEC-RI pour les différents fragments de la phosphoprotéine du virus Nipah		
Protéine	MMt	MMe
P ₁₋₁₀₀	12.347	13.5 ±1.5
P ₉₁₋₁₉₀	11.875	12.0 ±0.5
P ₁₇₀₋₂₄₀	8.691	8.0 ±2
P ₉₁₋₂₅₀	18.583	21 ±2.0
P ₂₂₂₋₃₁₉	11.536	12.9 ±1.5
P ₃₀₀₋₄₀₁	12.421	11.5 ±1.5
P ₃₈₇₋₄₇₉	11.151	12.4 ±1.5
P ₄₇₁₋₅₈₀	13.757	57.6 ±6
P ₆₅₅₋₇₀₉	7.364	8.0 ±1
P ₅₈₈₋₆₅₀	8.289	8.0 ±2
P ₅₈₈₋₇₀₉	15.041	15.3 ±1
P ₁₋₅₈₀	64.756	250.0 ±25
P ₄₇₁₋₇₀₉	28.238	113.0 ±10
P entière	79.367	306 ±30
MMe : masse moléculaire expérimentale (en kDa) déterminée par MALLS-SEC-RI. MMt : masse moléculaire théorique (en kDa) calculé à partir de la séquence en acides aminés, la Méthionine N-terminale et le histag (LEHHHHHH) sont comprises.		

Dans une expérience de pontage chimique par du glutaraldéhyde réalisée avec le fragment P₄₇₁₋₅₈₀, la présence de quatre formes est détectée par SDS-PAGE avec des masses

moléculaires d'environ 17, 34, 53 et 68 kDa estimées par comparaison avec un marqueur de masse moléculaire (**Figure 23**), confirmant la formation de tétramères.

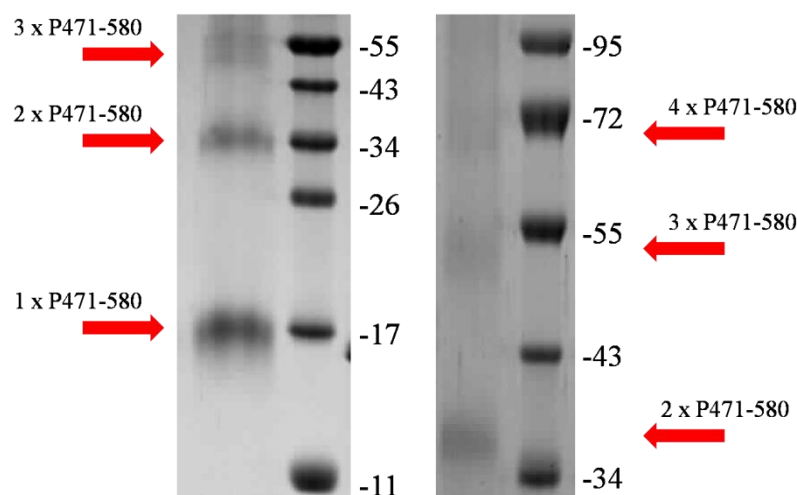


Figure 23 : Pontage chimique par glutaraldéhyde du fragment $P_{471-580}$. A gauche gel SDS-PAGE 19.5%, à droite gel SDS-PAGE 15%. Les flèches en rouge montrent les différentes formes obtenues par pontage chimique.

3.3.3.2. Structure cristalline du domaine de tétramérisation

Une première phase de recherche de conditions de cristallisation pour la construction $P_{471-580}$ englobant le domaine central (P_{CD}) de NiV P a été réalisé à l'aide du robot de cristallisation de l'EMBL (Grenoble) à 20 °C et en utilisant des concentrations de 5 et 15 $mg.mL^{-1}$ (**Figure 24A**). La protéine a cristallisé initialement dans un tampon 0.1 M citrate de sodium, pH 5.0, contenant 10 % PEG 6000 et 1 M LiCl. Ces conditions ont été optimisées manuellement et les meilleurs cristaux ont été obtenus dans une solution 0.1 M acide citrique à pH 5.0, contenant 1.1 M LiCl et 5 % PEG 6000 (**Figure 24B**). Ces premiers cristaux, constitués de « plaques » empilées en forme d'étoile, diffractaient à une résolution d'environ 3 Å (ESRF ID 23-2 microfocuse) mais de manière anisotrope, et la structure n'a pas pu être résolue par la méthode de remplacement moléculaire en utilisant soit la structure du domaine P_{CD} du virus de Sendai, soit des modèles *ab initio* générés par le Dr. Cédric Leyrat.

Le fragment 471-580 de P a ensuite été produit en incorporant de la sélénométhionine dans le but de résoudre la structure par une méthode de phasage expérimentale (voir **Chapitre 2 – Matériels et Méthodes**). La protéine sélénée n'a pas cristallisé dans les conditions établies avec la protéine native et une deuxième optimisation des conditions de cristallogénèse a dû être réalisée. Des cristaux utilisables de la protéine sélénée ont été obtenus dans 0.1 M acide citrique à pH 5.0, contenant 1 M LiCl et 17% PEG 3350 (**Figure 24D**). Ces cristaux appartenant au groupe d'espace $P4_32_12$ diffractaient à une résolution d'environ 3.5 Å, mais

ces données ne présentaient pas suffisamment de signal anomal pour le phasage. Une troisième étape d'optimisation, en utilisant l'« additive screen » d'Hampton Research, a permis d'obtenir de nouveaux cristaux en présence d'hexanediol, appartenant cette fois au groupe d'espace $P2_1$ (Figure 24F et 24G).

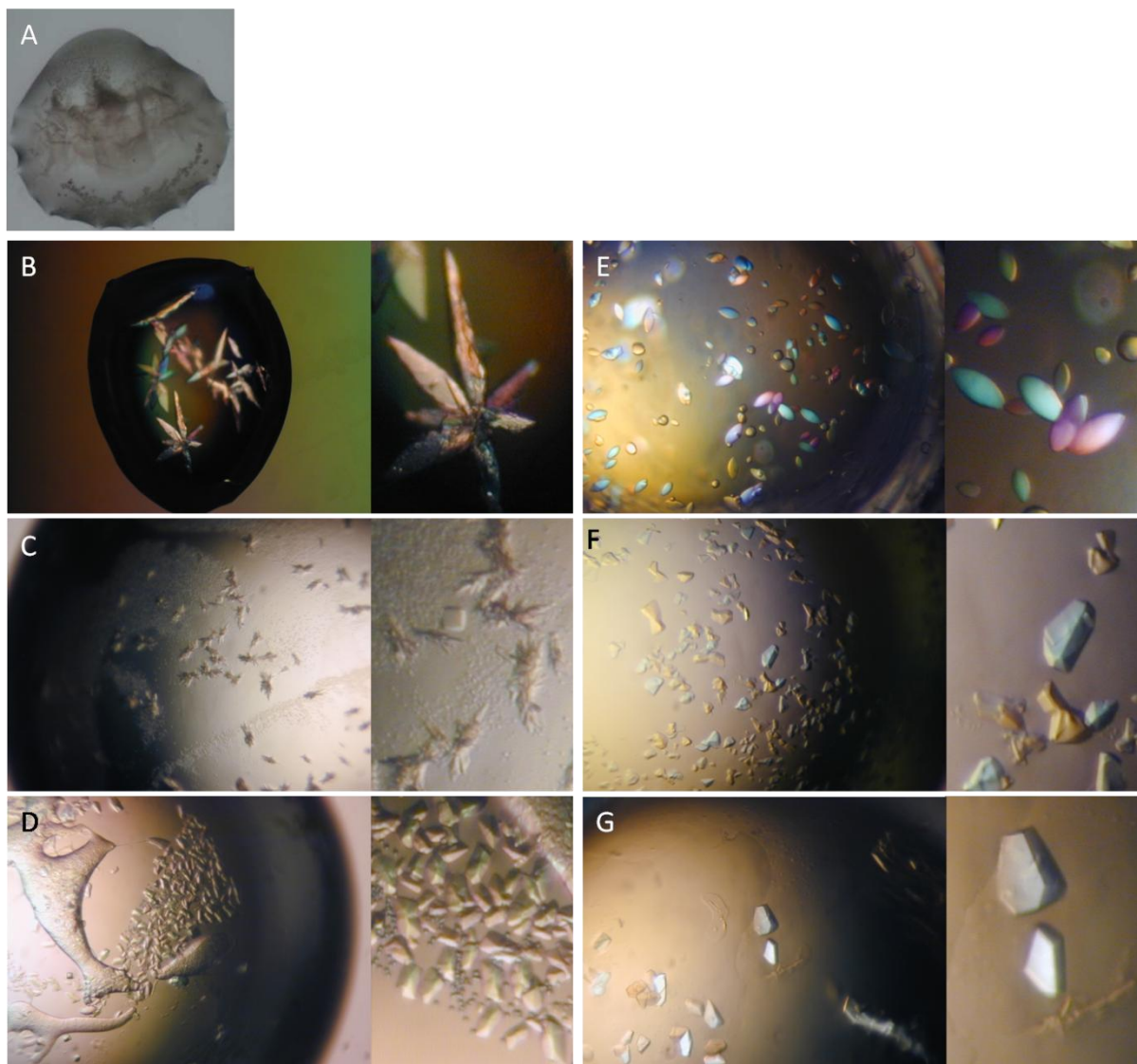


Figure 24 : Différentes formes cristallines du fragment P₄₇₁₋₅₈₀ du virus Nipah (NiV). **A)** Cristaux de la protéine native obtenus dans 1 M LiCl, 0.1 M Acide citrique pH 5.0, 10% PEG 6000 en goutte de 200 nL par le robot cartésien (EMBL, HTXlab). **B)** Cristaux de la protéine native reproduits manuellement dans 1.1 M LiCl, 0.1 M acide citrique pH 5.0, 5% PEG 6000. **C) – G)** Cristaux de la protéine sélénée. **C)** 1 M LiCl, 0.1 M acide citrique pH 5.0, 7% PEG 6000 ; **D)** 1 M LiCl, 0.1M acide citrique pH 5.0, 17% PEG 3350 ; **E)** 1.1 M LiCl, acide citrique pH 5.0, 12-13% PEG 3350 **F)** Cristaux issus d'une phase d'optimisation des conditions en D) en utilisant l' « additive creen » et en présence de 3% hexanediol. **G)** Cristaux obtenus dans les conditions optimisés de F) et notamment 1 M LiCl, 0.1 M acide citrique pH 5.0, 17% PEG 3350, 0.1 M Arginine et 4% hexanediol. Les expériences ont été menées à 20°C. B) - G) Le volume des gouttes est de 2 μ L.

Ces cristaux diffractaient à une résolution d'environ 3.0 Å mais se détruisaient rapidement lors de leur exposition aux rayons X. Un ensemble complet de données a néanmoins pu être obtenu en utilisant la technique de collecte de données de diffraction dite « en faisceau inversé » (inverse beam) par tranches de 10° et en combinant les données de

diffraction obtenues avec trois cristaux différents (ligne PROXIMA1, synchrotron SOLEIL). Des phases initiales ont été obtenues par la méthode SAD (**Figure 25**). Celles-ci permettaient de distinguer de longues hélices dans la densité électronique, mais ne permettaient pas le traçage automatique de la chaîne. Afin de faciliter ce traçage automatique, nous avons cherché à améliorer les phases initiales et donc la densité électronique expérimentale. Pour cela nous avons tiré profit de la présence de symétrie non cristallographique (NCS), qui relie les chaînes des deux tétramères dans l'unité asymétrique. Nous avons détecté la NCS au préalable en calculant une fonction d'autorotation, une procédure qui consiste à chercher des orientations de la fonction de Patterson, calculée à partir des intensités des réflexions, qui permettent de la superposer à elle-même, en utilisant le programme MOLREP (Vagin and Teplyakov 2010). Ainsi nous avons fait une moyenne par rapport à la NCS (NCS averaging) en utilisant le programme DM (Cowtan 1999). Cette technique moyenne la densité électronique entre les copies de la molécule reliées par NCS ce qui a comme effet la réduction du bruit et l'augmentation des contraintes sur les phases. Pour pouvoir faire cela les relations géométriques entre les protomères dans l'unité asymétrique (les opérateurs de NCS) doivent être connus. Nous avons déterminé ces opérateurs de NCS en exploitant les coordonnées des sites sélénium, trouvés au préalable par le programme SHELXD (Schneider and Sheldrick 2002) du script HKL2MAP (Pape and Schneider 2004), en utilisant le programme PROFESSS (Winn, Ballard et al. 2011).

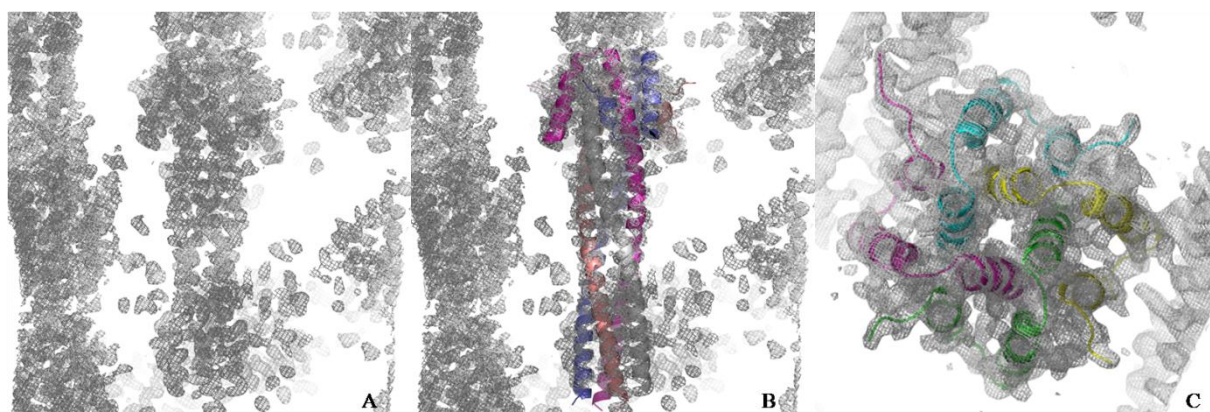


Figure 25 : Phasage expérimentale des données de diffraction du domaine d'oligomérisation P₄₇₁₋₅₈₀ (P_{CD}) du virus Nipah (NiV). A) Carte de densité électronique (contour à 1 σ) obtenue expérimentalement par la méthode SAD pour le domaine central d'oligomérisation P_{CD} de NiV. B) Le modèle affiné superposé à A). C) Vu du haut de B).

Le modèle de P_{CD} a été affiné avec le programme Phenix.refine (Afonine, Grosse-Kunstleve et al. 2012). Pendant les premiers cycles d'affinement, une méthode de recuit simulé a été utilisée qui consiste à effectuer des simulations répétées d'augmentation et de diminution de température dans le but de faire bouger les atomes et ensuite de leur permettre

de trouver leurs positions d'équilibre tout en minimisant l'énergie potentielle du système (qui est l'équivalent de la fonction cible de l'affinement (voir **Chapitre 2** - Matériels et Méthodes). Cette première étape a pour but d'éviter la modélisation d'une forme piégée dans un minimum local d'énergie (Brunger, Kuriyan et al. 1987; Brunger and Adams 2002). Un problème fréquent dans l'affinement des structures à basse résolution, est (en fonction du groupe d'espace, de la taille de la maille cristalline et de la complétude des données) le nombre limité de réflexions uniques, qui peut conduire à un ajustement excessif (« overfitting ») des données. Eviter cet ajustement excessif revient à augmenter le rapport du nombre d'observations (réflexions uniques + contraintes) par rapport au nombre de paramètres à ajuster (généralement un facteur de déplacement atomique (ADP) et trois coordonnées par atome plus le nombre d'éventuels groupes TLS). Afin d'augmenter ce rapport en cas de présence de NCS, des contraintes sur les coordonnées et les ADPs des atomes symétriques par NCS peuvent être imposées, mais une telle symétrie stricte n'autorise pas de petites différences entre les chaînes symétriques. Afin d'éviter de tels problèmes dans l'affinement du P_{CD} et de conserver des différences éventuelles entre les chaînes, tout en diminuant le nombre de paramètres à ajuster, nous avons utilisé des limitations (« restraints ») sur les distances interatomiques locales dans l'affinement (Uson, Pohl et al. 1999). Dans ce cas, l'ajout de cette procédure dans l'affinement a permis d'abaisser la valeur de R_{free} .

L'unité asymétrique du cristal contient deux copies d'un tétramère. Chaque chaîne est constituée d'une longue hélice de 64 aa, interrompue par le résidu Pro544 (H α 3 : 508-569), et de deux hélices plus courtes (H α 1 : aa 480-484, H α 2 : 490-500) (**Figure 26**). Les longues hélices forment un faisceau d'hélices α surenroulées et les deux hélices courtes forment une couronne du côté N-terminal du faisceau. Pour toutes les chaînes, les 7 premiers acides aminés du côté N-terminal et les 15 à 17 derniers acides aminés du côté C-terminal (y compris le tag hexahistidine) ne sont pas visibles dans la structure car vraisemblablement désordonnés. La chaîne D présente une faible densité pour la chaîne principale à partir de résidu 3.

Les dimensions globales de P_{CD} sont 96 Å de longueur et 45 Å de diamètre pour la couronne. Le pI, calculé à partir de la séquence en acides aminés, est d'environ 5.15 et celui de la séquence des aa visibles dans la structure est 4.9 et donc à pH neutre, la molécule a une charge nette globale négative. Le potentiel électrostatique de surface de la protéine révèle qu'une partie importante de la surface de la partie N-terminale (la couronne) est chargée négativement (**Figure 26**).

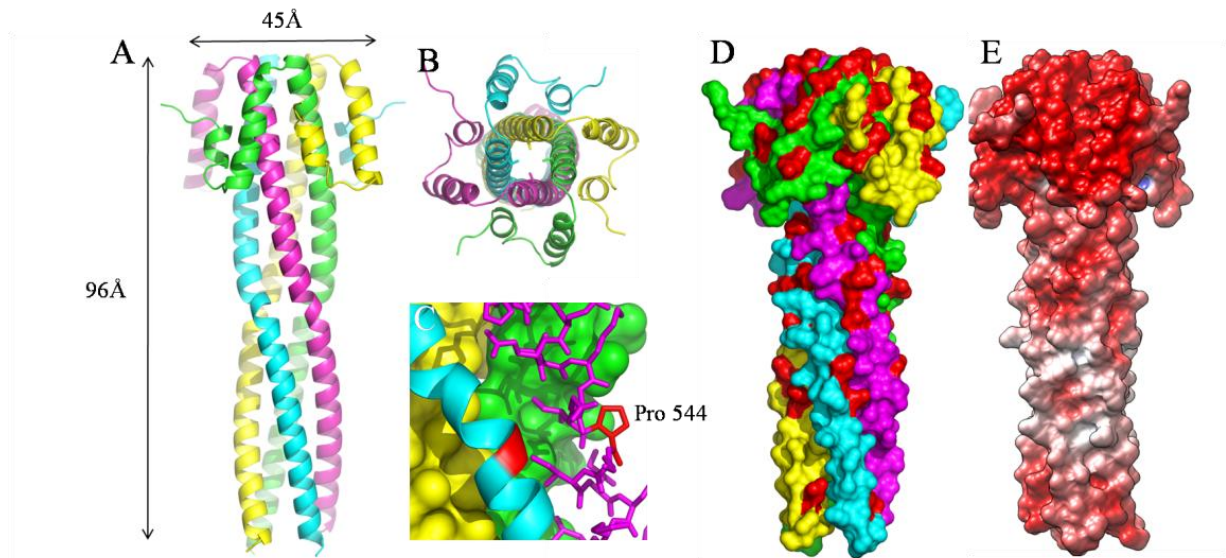


Figure 26 : La structure du domaine d'oligomérisation (P_{CD}) du virus Nipah (NiV). **A)** Représentation en cartoon et dimensions globales d'un tétramère de NiV P_{CD} . Les quatre chaînes sont colorées en violet, turquoise, jaune et vert, respectivement. **B)** Vue du haut de A). **C)** Zoom sur le résidu Pro544. **D)** Représentation en surface avec le même code couleur que A). En rouge les résidus aspartate et glutamate. **E)** Surface de potentiel électrostatique du bleu (+10K_bT) au rouge (-10K_bT). Le potentiel électrostatique a été calculé en utilisant les programmes PDB2PQR (Dolinsky, Nielsen et al. 2004) et APBS (Baker, Sept et al. 2001).

Les $C\alpha$ des 8 protomères de l'unité asymétrique se superposent avec un RMSD inférieur à 1 Å et les $C\alpha$ des résidus 489 à 539 se superposent avec un RMSD inférieur à 0.3 Å. La structure présente une symétrie d'ordre 4 dans la partie N-terminale correspondant à la couronne et à la partie du faisceau d'hélices surenroulées qui est couverte par la couronne, mais cette symétrie d'ordre 4 disparaît dans la partie C-terminale du faisceau. Les couches de résidus à l'interface de tétramérisation (à l'intérieur du faisceau) situées au delà de la cassure induite par le résidu Pro544 ont leurs $C\alpha$ situés à des distances différentes du centre du faisceau, et la symétrie est plutôt d'ordre 2 (**Figure 27**).

Les protomères du domaine P_{CD} forment un complexe très compact et la surface de contact entre deux protomères est en moyenne de 2,200 Å², une valeur supérieure de 20% à celle de l'interface formée entre les protomères du P_{CD} de SeV (PDB code : 1EZJ). En moyenne, un protomère forme 19 liaisons d'hydrogène et 11 ponts salins avec un de ces voisins ce qui est également nettement supérieur aux 12 liaisons d'hydrogène et 1 pont salin observés pour le P_{CD} de SeV. L'intérieur du faisceau d'hélices surenroulées de NiV P_{CD} , composé de 19 couches de chaînes latérales superposées, est globalement hydrophobe, avec seulement trois couches composées de résidus polaires (Ser515, Gly519 et Asn522).

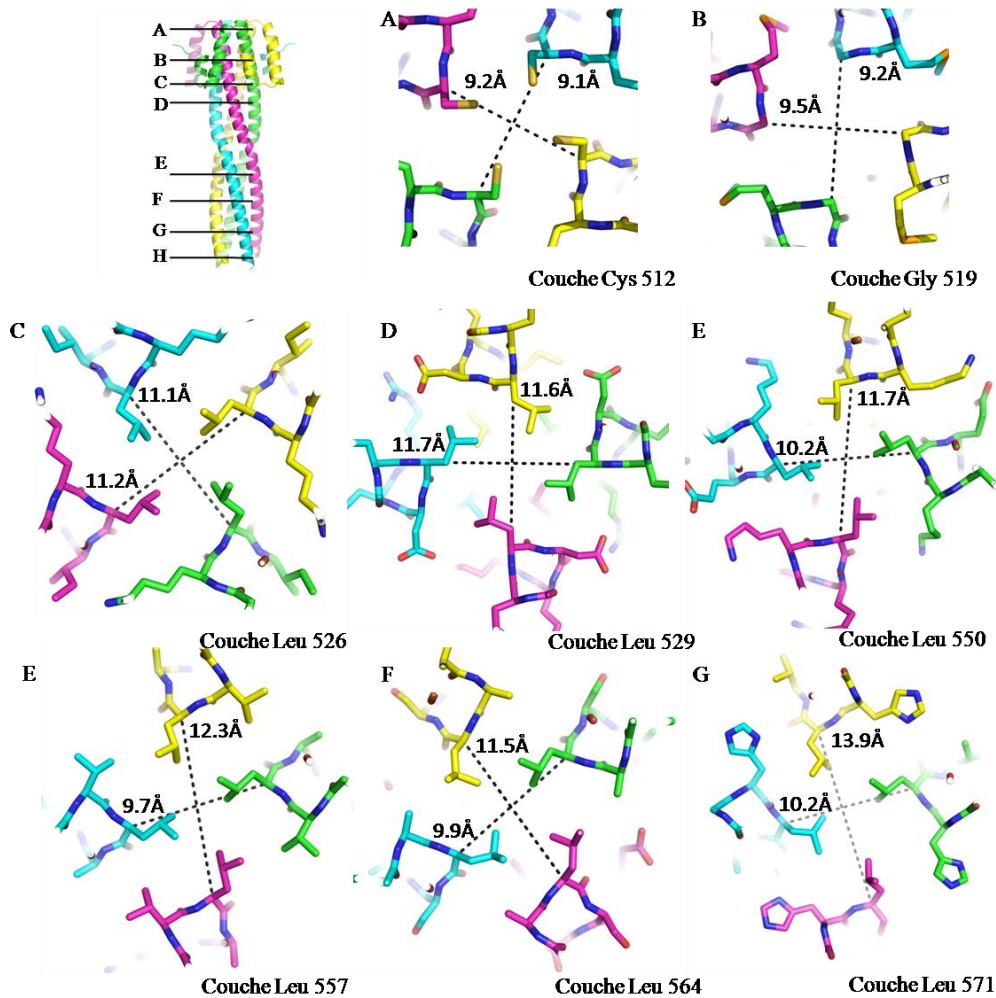


Figure 27 : La structure du domaine d'oligomérisation (P_{CD}) du virus Nipah (NiV). Sont représentées les différentes couches du faisceau central d'hélices α . Les quatre chaînes sont colorées comme dans la **Figure 25A**. Les lignes en pointillé noir représentent les distances entre les $C\alpha$ des chaînes opposées et les valeurs numériques sont données. A noter la symétrie de quatre du côté N-terminal du faisceau qui se transforme en symétrie de deux du côté C-terminal. En haut à gauche le positionnement des différentes couches dans la structure du P_{CD} .

Chaque protomère forme de nombreux contacts avec ses voisins $i+1$ et $i-1$. Ainsi, l'extrémité N-terminale d' $H\alpha 3$ est stabilisée par le résidu Asp507 qui forme deux liaisons hydrogène, l'une avec le groupement amide de la liaison peptidique entre Glu509 et Thr510 et l'autre avec le groupement hydroxyle de Thr510 et ainsi constitue un N_{CAP} (**Figure 28A**). Plusieurs réseaux de contacts polaires sont formés au niveau de la couronne et présentent une symétrie d'ordre 4.

Un premier réseau de liaisons hydrogène implique les résidus Tyr477 et Asp483 de la chaîne i et les résidus Tyr498, His499 et Arg495 de la chaîne $i-1$ (**Figure 28B**). L'hélice $H\alpha 1$ est stabilisée d'un côté par des liaisons hydrogène entre le groupement hydroxyle et le groupement amide du résidu Ser481 de la chaîne i et le résidu Glu514 de chaîne $i-1$.

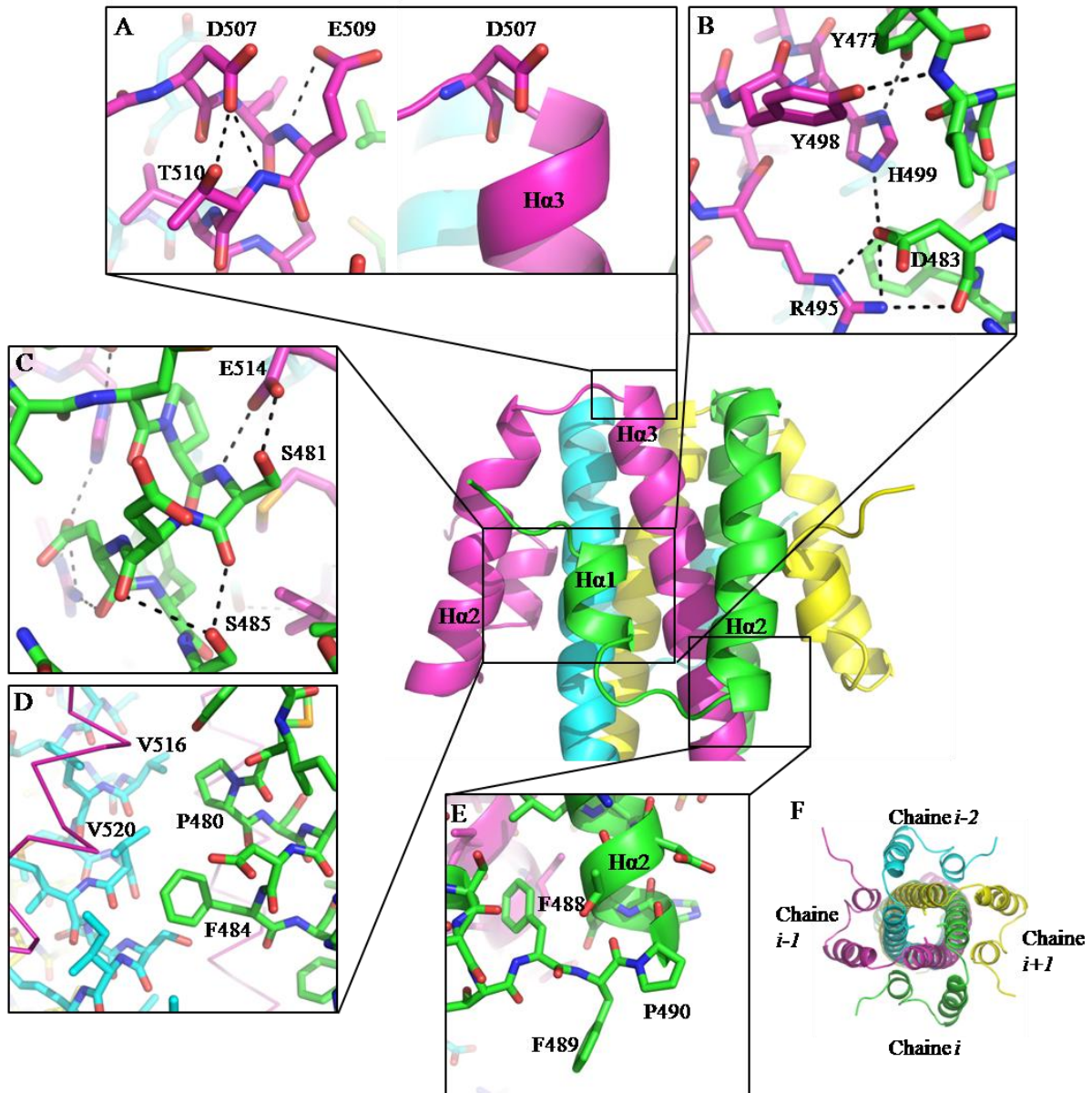


Figure 28 : La structure du domaine d'oligomérisation (P_{CD}) du virus Nipah (NiV) – la couronne. Les quatre chaînes sont colorées comme dans les **Figures 25** et **26**. Les lignes en pointillé noir représentent des interactions polaires d'une distance inférieure à 3.5Å. **A)** Les résidus Asp 507 et Glu 509 forme un N_{CAP} pour l'hélice $H\alpha 3$. **B)** Interactions entre $H\alpha 1$ de la chaîne i (en vert) et $H\alpha 2$ de la chaîne $i-1$ (violet). **C)** Stabilisation de $H\alpha 1$. Le résidu Glu 514 de la chaîne $i-1$ (violet) stabilise le N-terminale de l'hélice $H\alpha 1$ de la chaîne i (vert) ainsi formant un N_{CAP} . Le résidu Ser 485 forme un C_{CAP} . **D)** Contacts entre la chaîne i (vert) et la chaîne $i-2$ (turquoise). Les résidus Phe 484 et Pro 480 sont packés contre Val 520 et Val 516. **E)** Le N-terminal de $H\alpha 2$ est stabilisé par le résidu Pro 490.

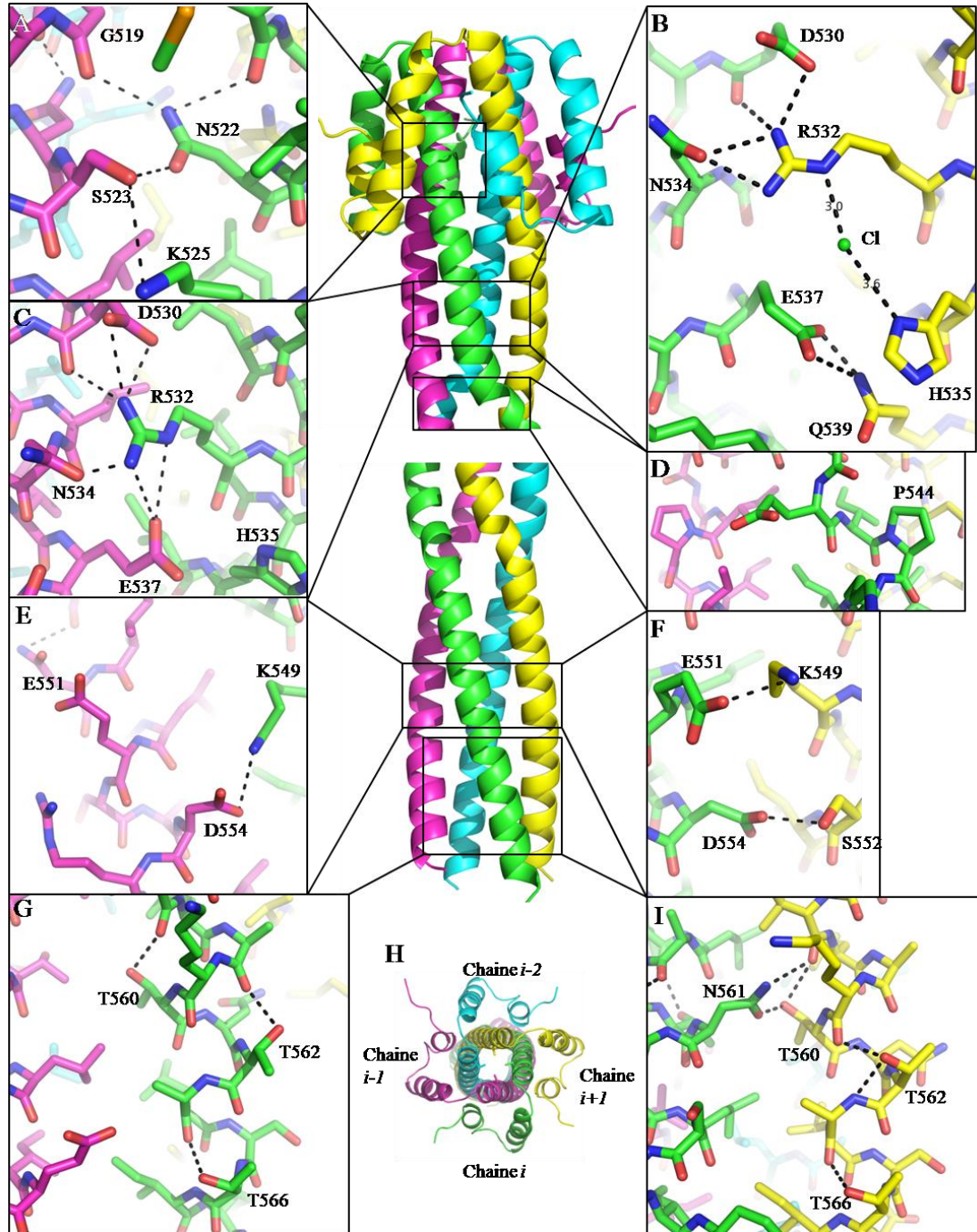


Figure 29 : La structure du domaine d'oligomérisation (P_{CD}) du virus Nipah (NiV). Les quatre chaînes sont colorées comme dans la **Figure 28**. Les lignes en pointillé noir représentent des interactions polaires d'une distance inférieure à 3.5\AA . **A)** Interactions entre les chaînes i (en vert) et $i-1$ (en violet) du faisceau central d'hélices α . Les quatre faces sont symétriques. **C), E)** et **G)** Interactions inter- et intrachaînes pour les chaînes i (en vert) et $i-1$ (en violet). **B), F)** et **D)** Interactions inter- et intrachaînes pour les chaînes $i+1$ (en jaune) et i (en vert). L'arrangement des chaînes latérales sur les faces, créés par les chaînes voisines du faisceau central d'hélices α , sont symétriques deux par deux : les faces créés par les chaînes $i/i+1$ et $i-1/i-2$ sont symétriques et les faces créés par les chaînes $i/i-1$ et $i+1/i-2$ sont également symétriques mais les faces voisines ne le sont pas.

La partie C-terminale de l'hélice H α 1 est stabilisée par le résidu Ser485 qui forme des contacts avec les carbonyles des liaisons peptidiques de Ser481 et Asp482 de la même chaîne (**Figure 28C**). L'hélice H α 1 de la chaîne i est également en contact avec la chaîne $i-2$ par

l'intermédiaire des résidus Phe484 et Pro480 qui se positionnent dans une cavité hydrophobe formée par les chaînes *i-1* et *i-2* (**Figure 28D**).

Dans la partie C-terminale des longues hélices, les interactions entre les chaînes latérales respectent plutôt une symétrie de 2. Dans la partie N-terminale des longues hélices un premier réseau de liaisons hydrogène et ponts salins stabilise le faisceau surenroulé d'hélices α (**Figure 29A**). Les résidus Arg532 appartenant à deux des quatre chaînes (chaînes *i* et *i-2*, **Figure 29C**) forment des interactions avec les résidus Asp530, Asn534 et Glu537 du protomère *i+1* tandis que pour les résidus Arg532 appartenant aux deux autres chaînes (chaînes *i+1* et *i-1*, **Figure 29B**) forment des liaisons avec les résidus Asp530 et Asn534 du protomère *i+1* et avec un atome de chlore positionné entre les résidu Arg532 et His535 du protomère *i+1*. Dans ce cas, la chaîne latérale du résidu Glu537 du protomère *i* ne forme pas de pont salin avec le résidu Arg532 mais semble former des liaisons hydrogène avec le résidu Gln539. Le résidu Glu551 de chaîne *i* forme un pont salin avec le résidu K549 de chaîne *i-1* dans deux des quatre interfaces, mais pointe vers le solvant dans les deux autres (**Figure 29F et E**). Le résidu Asn561 forme des liaisons hydrogène avec le résidu T560 et avec le carbonyle d'une liaison peptidique de la chaîne *i+1* dans deux des quatre interfaces mais pointe vers le solvant dans les deux autres (**Figure 29G et I**).

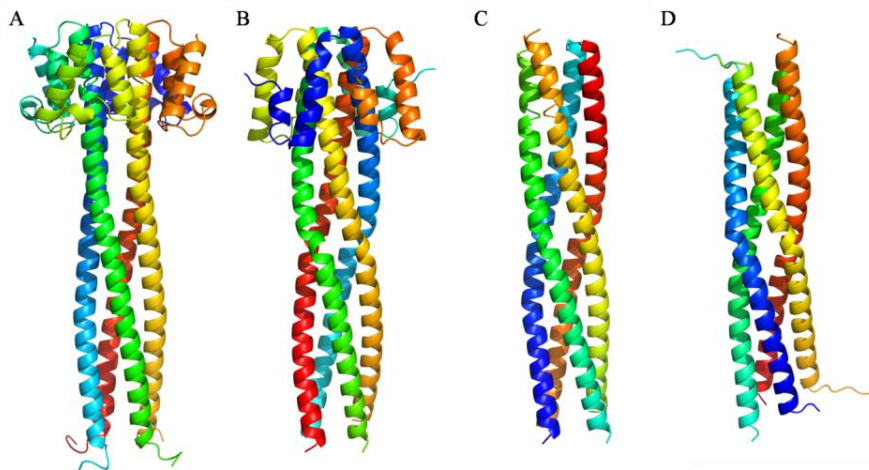


Figure 30 : Structures atomiques des domaines d'oligomérisation (P_{CD}) de différents membres des *Paramyxovirinae*. Représentation en cartoon de la structure cristalline du P_{CD} **A**) du virus Sendai (SeV) (Tarbouriech, Curran et al. 2000) ; **B**) du virus Nipah (NiV), déterminée au cours des travaux de thèse présenté ici ; **C**) du virus de la rougeole (MeV) (Communie, Crepin et al. 2013) et **D**) du virus des oreillons (MuV) (Cox, Green et al. 2013). Les différentes chaînes d'un tétramère sont colorée en rouge, bleu, vert et jaune, respectivement.

En dépit de l'absence de conservation de séquence, il est intéressant de noter la structure du domaine P_{CD} de NiV possède une structure similaire à celle du domaine P_{CD} de SeV (Tarbouriech, Curran et al. 2000) avec un faisceau de quatre hélices surenroulées et une

couronne du côté N-terminal, alors que le domaine P_{CD} de MeV ne comprend pas de couronne (Communie, Crepin et al. 2013). Néanmoins ces longues hélices sont constituées d'un nombre très similaire d'acides aminés (**Figure 30**, discuté plus loin).

Tableau 4. Statistiques de collecte des données (PROXIMA-1) et d'affinement de la structure du NiV P_{CD}.

Crystal	
Data collection	
Wavelength (Å)	0.979
Space group	P12 ₁ 1
Unit cell	
a, b, c (Å)	59.5, 58.6, 122.6
α , β , γ (°)	90.00, 9.70, 90.00,
Resolution ^a (Å)	48.75 – 3.00 (3.08 – 3.00)
1/ σ (I) ^a	5.3 (2.1)
R-merge ^{a,b}	0.14 (0.65)
Completeness ^a (%)	99.5 (99.9)
Multiplicity ^a	4.1 (1.93)*
Unique reflections*(Friedel pairs unmerged)	47924*
Refinement	
Resolution ^a (Å)	48.75 – 3.00 (3.06 – 3.00)
R-factor ^a	0.175 (0.253)
R-free ^a	0.229 (0.286)
Number of atoms	6368
macromolecules	6285
ligands	13
water	7
Number of protein residues	757
R.m.s. deviations	
bond lengths (Å)	0.013
bond angles (°)	1.159
Ramachandran favored (%)	98.04
Ramachandran allowed (%)	1.83
Ramachandran outliers (%)	0.13
Average B-factor	80

^aValues in parentheses are for highest-resolution shell.

3.3.4. La partie C-terminale de la phosphoprotéine du virus Nipah (aa 471-709) est globalement structurée mais contient des régions flexibles

La construction P₄₇₁₋₇₀₉ contient le domaine central de tétramérisation P_{CD}, les quatre domaines P_{XD} et les jonctions qui les relie (voir **Figure 5**). Des courbes de SAXS ont été enregistrées à plusieurs concentrations pour cette construction. L'allure des courbes et les valeurs du rayon de giration (R_g) sont indépendantes de la concentration, démontrant l'absence d'agrégation (**Figure 31A et B**).

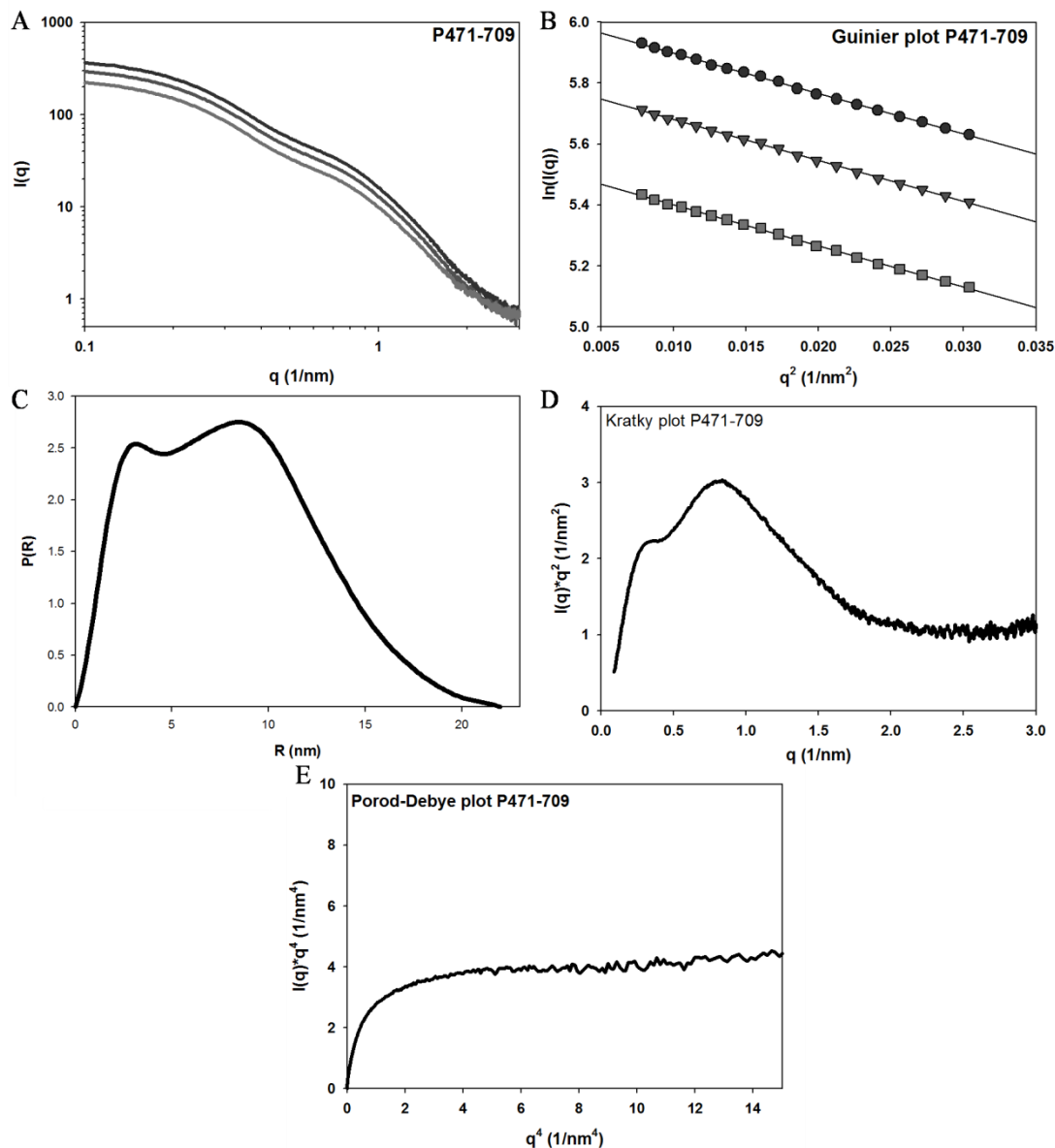


Figure 31 : Expériences de diffusion des rayons X aux petits angles (SAXS) sur le fragment P471-709 et du virus Nipah (NiV). A) Courbes de SAXS enregistrées pour des valeurs de q entre 0.05 et 4 nm^{-1} à 5.8 , 4.4 et 3.0 mg.ml^{-1} . **B)** Analyse de Guinier dans la région de $q.R_g < 1.0$. **C)** Fonction de distribution des distances $P(r)$, **D)** représentation de Kratky et **E)** représentation de Porod-Debye. Pour ces représentations la courbe enregistrée à la plus forte concentration en protéine, normalisées pour la concentration, a été utilisées.

La fonction de distribution de paires, $P(r)$, en fonction de la distance r , calculée pour cette protéine à partir des données de SAXS, présente deux maxima, caractéristique des protéines à plusieurs domaines (**Figure 31C**) (Svergun and Koch 2003). Le maximum aux environs de 3 nm pourrait correspondre aux distances intra-domaines tandis que le maximum aux environs de 9 nm pourrait correspondre aux distances inter-domaines, notamment à des distances entre le domaine P_{CD} et les quatre domaines P_X . Le graphique de Kratky pour la construction $P_{471-709}$ indique une protéine structurée contenant des régions flexibles (**Figure 31D**) mais d'une façon beaucoup moins importante comparée aux protéines P_{1-580} et à la P

entière (**Figure 34**, comparer les graphiques de $P(r)$, Kratky et Porod-Debye de $P_{471-709}$ (en vert) avec celles de la P entière (en noir) et P_{1-580} (en rouge)). De la même manière, le graphique de Porod-Debye atteint un plateau qui implique un contraste marqué entre la molécule et le solvant, typique d'une protéine structurée (**Figure 31E**) (Rambo and Tainer 2011). Ces observations sont en accord avec la présence des domaines structurés P_{CD} et P_{XD} .

La région comprise entre les résidus 588 à 650 constitue la jonction entre le domaine central et le domaine P_X et est prédite structurée par le D-score (**Figure 5**). La masse moléculaire (MM) du peptide correspondant obtenue par SEC-MALLS-RI est de 8 ± 2 kDa, ce qui démontre son état monomérique (**Figure 32A**). Le spectre RMN à deux dimensions par corrélation hétéronucléaire HSQC (*Heteronuclear Single Quantum Coherence*) présente une très faible dispersion des déplacements chimiques des groupements amides de la chaîne principale, caractéristique d'une protéine désordonnée (**Figure 32B**), et le R_g calculé à partir des données de SAXS en utilisant l'approximation de Guinier est de 2.5 ± 0.1 nm, intermédiaire entre les valeurs de 2.7 et 2.4 nm calculées en utilisant la loi de puissance de Flory et des paramètres déterminés respectivement pour une protéine dénaturée et pour une protéine intrinsèquement désordonnée du même nombre d'aa (Fitzkee and Rose 2004; Bernado and Blackledge 2009). Le peptide $P_{588-650}$ est donc globalement désordonné en solution, suggérant qu'il s'agit également d'un élément de reconnaissance qui est stabilisé lors de l'interaction avec un partenaire.

Nous avons également étudié le peptide $P_{588-709}$ qui contient la région de connexion 588-650 et le domaine structuré P_{XD} . Ce peptide élu de la colonne de gel filtration Superdex S75 sous la forme d'un seul pic, et sa MM de 15.3 ± 1.5 kDa, calculée par SEC-MASS-RI, démontre son état monomérique et monodisperse (**Figure 32C**). Le volume d'élution est nettement supérieur à celui du domaine P_{XD} en fusion avec l'étiquette hexa-histidine, confirmant que cette région n'adopte pas une forme globulaire en présence du domaine P_{XD} et ne forme pas un domaine globulaire en se repliant sur le domaine P_{XD} non plus.

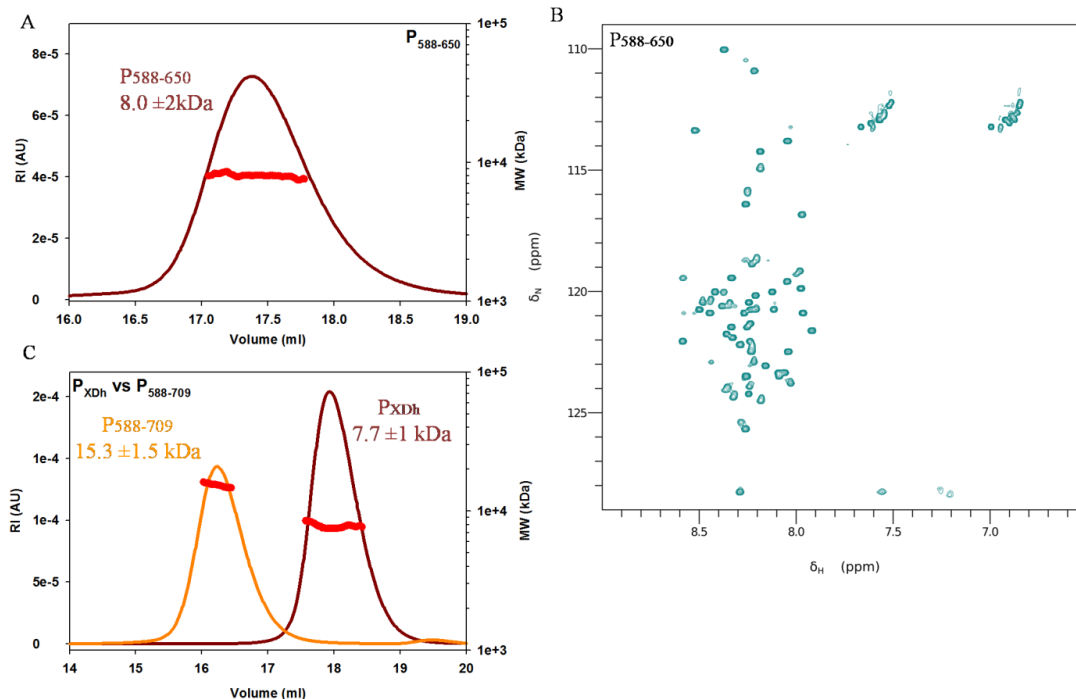


Figure 32 : Analyse par SEC-MALLS-RI et RMN des fragments $P_{588-650}$ (PQ) et $P_{588-709}$ (PQ- P_{XD}) de la phosphoprotéine (P) du virus Nipah (NiV). A) Profil d'éluion de PQ (ligne en marron) de la colonne de gel filtration Superdex S200. B) Spectre 2D [^1H - ^{15}N]-HSQC de $P_{588-650}$ enregistré à 25°C dans un tampon de 20 mM Bis-Tris pH 6.0, 150 mM NaCl et 50 mM glutamate/arginine (R/E). C) Comparaison des profils d'éluion de $P_{588-709}$ (ligne en orange) et P_{XD} (domaine P_{XD} portant une étiquette hémahistidine pour cette expérience) (ligne en marron) de la colonne de gel filtration Superdex S75. Les colonnes ont été équilibrées dans 20 mM Tris-HCl pH 7.5, 150 mM NaCl. En rouge les masses moléculaires (MM) à travers les pics d'éluion, calculées à partir de la diffusion statique de la lumière et la réfractométrie pour chaque protéine. Les valeurs numériques des masses moléculaires (MM) ont le même code couleur que les profils d'éluion.

La comparaison des spectres 2D [^1H - ^{15}N]-HSQC du peptide $P_{588-650}$ avec celui de la phosphoprotéine entière et avec celui du fragment $P_{471-709}$, qui est tronqué de la partie N-terminale désordonnée de la protéine, indique que certaines résonances se superposent alors que d'autres résonances sont déplacées, suggérant que cette région de la protéine adopte une conformation différente dans le contexte de la protéine tétramérique (**Figure 33A et B**).

D'une façon étonnante aucune résonance provenant de résidus du domaine P_{XD} n'est visible ni dans le spectre 2D [^1H - ^{15}N]-HSQC de la phosphoprotéine entière ni dans celui du fragment $P_{471-709}$. En raison de la petite taille du domaine P_{XD} et du fait que ce dernier soit lié au domaine central par un connecteur flexible (résidu 588-650), les résonances des résidus des P_{XD} devraient être visibles dans le spectre RMN. Leur absence suggère que les P_{XD} ne sont pas indépendants dans le contexte de la protéine entière tétramérique. Cette observation rappelle la formation de dimère du domaine P_{XD} dans l'empilement cristallin et la suspicion de dimérisation en solution. La présence de quatre domaines P_X connectés entre eux par l'intermédiaire du domaine de tétramérisation équivaut à une concentration effective très élevée qui pourrait expliquer la formation de dimères ou autres forme oligomériques.

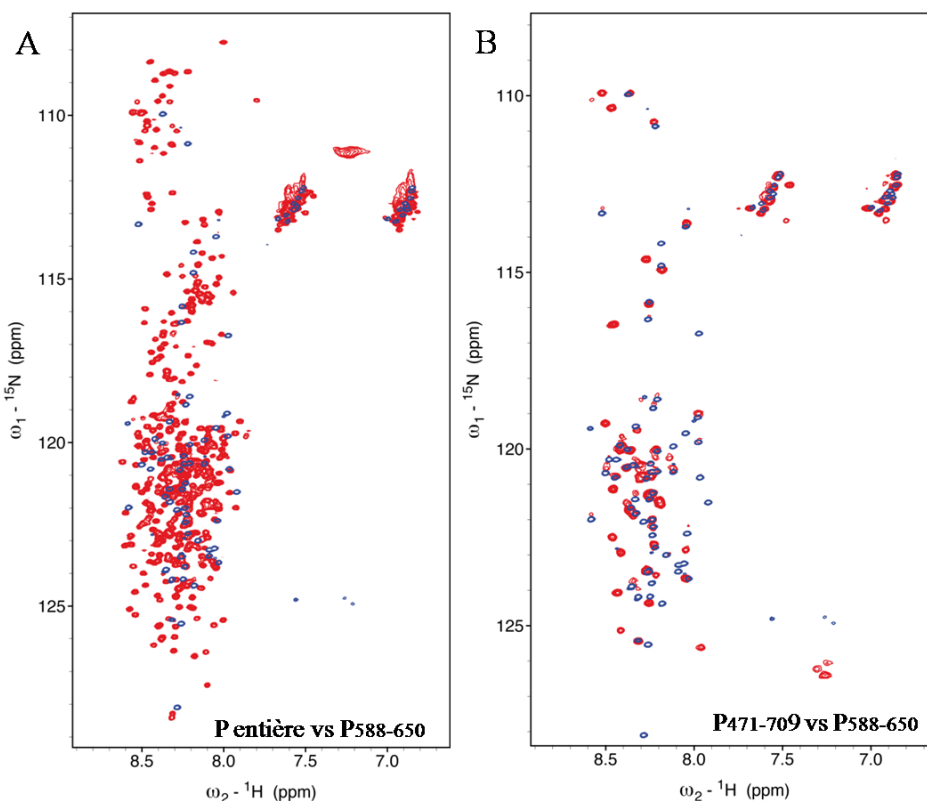


Figure 33 : Spectres RMN. **A)** Spectre 2D [^1H - ^{15}N]-HSQC de la NiV P entière (rouge) superposé au spectre 2D [^1H - ^{15}N]-HSQC du fragment P₅₈₈₋₆₅₀ (bleu). **B)** Spectre 2D [^1H - ^{15}N]-HSQC du fragment P₄₇₁₋₇₀₉ (P_{CED}) de NiV (rouge) superposé au spectre 2D [^1H - ^{15}N]-HSQC du fragment P₅₈₈₋₆₅₀ (bleu). Les spectres ont été enregistré dans 20 mM Bis-Tris pH 6.0, 150 mM NaCl et 50 mM glutamate/arginine à 25°C.

3.3.5. La partie N-terminale (P_{NTR}) de la phosphoprotéine de NiV est globalement désordonnée mais contient des structures secondaires fluctuantes

Afin de caractériser les propriétés structurales de la longue partie N-terminale (P_{NTR}) et de la phosphoprotéine entière de NiV, des mesures de SAXS et de RMN ont été réalisées avec la P entière et les fragments P₁₋₅₈₀ qui renferment le domaine de tétramérisation, ainsi qu'avec les peptides différents peptides P₁₋₁₀₀, P₉₁₋₁₉₀, P₉₁₋₂₅₀, P₂₂₂₋₃₁₉, P₃₈₇₋₄₇₉ de la région N-terminale (P_{NTR}).

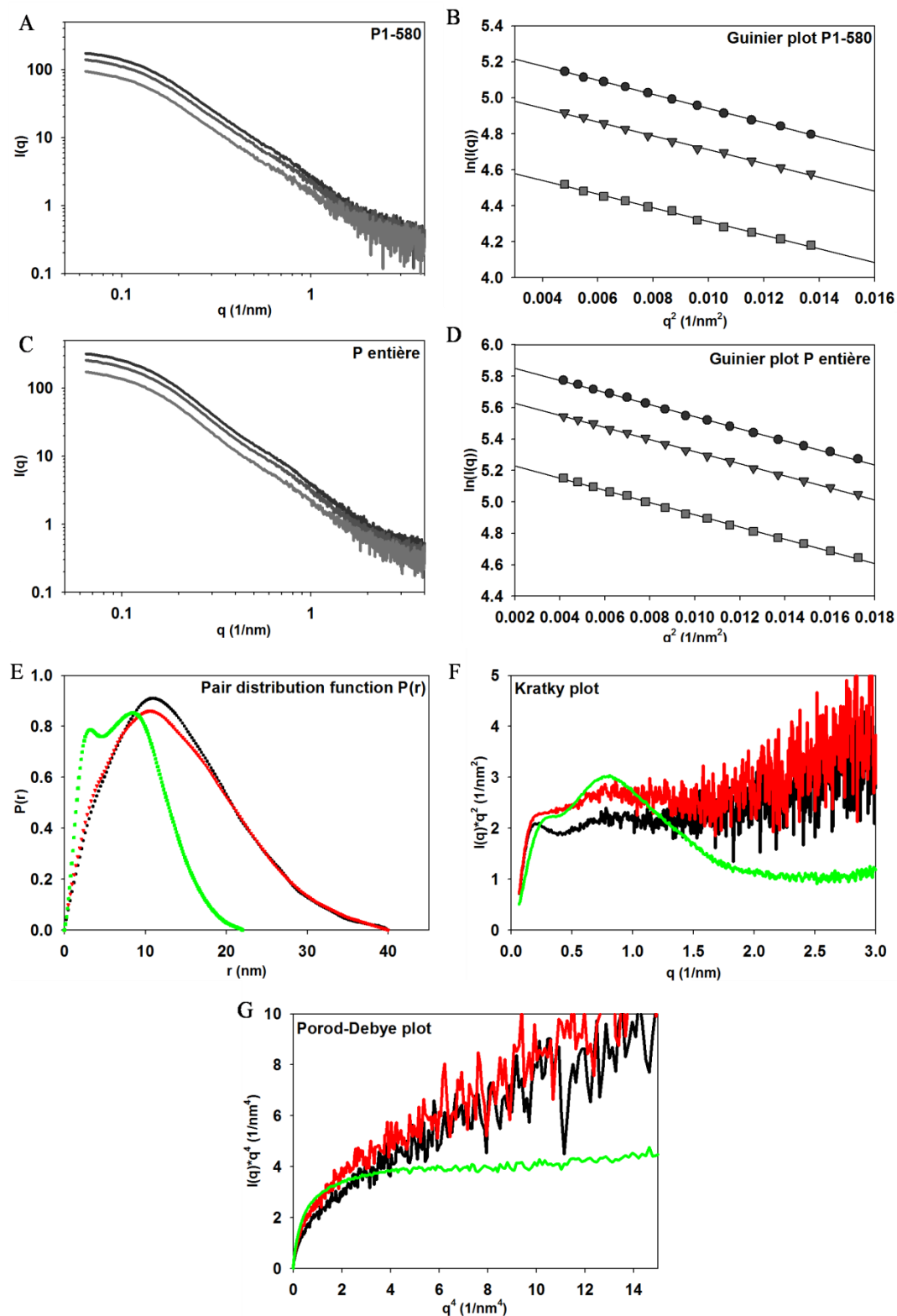


Figure 34: Expériences de diffusion des rayons X aux petits angles (SAXS) sur le fragment P1-580 et la phosphoprotéine (P) entière du virus Nipah (NiV). Courbes de SAXS enregistrées pour des valeurs de q entre 0.05 et 4 nm^{-1} pour **A)** P1-580 à 1.0 , 0.8 et $0.5 \text{ mg}\cdot\text{ml}^{-1}$ et **C)** la P entière à 1.8 , 1.4 et $1 \text{ mg}\cdot\text{ml}^{-1}$. **B)** et **D)** Analyse de Guinier dans la région de $q\cdot R_g < 1.0$ pour P1-580 et la P entière, respectivement. **E)** Fonction de distribution des distances $P(r)$, **F)** représentation de Kratky et **G)** représentation de Porod-Debye pour P471-709 (ligne en vert), P1-580 (ligne en rouge) et la P entière (ligne en noir). Pour ces représentations de **G)**, **H)** et **I)** les courbes enregistrées à la plus forte concentration en protéine, normalisées pour la concentration, ont été utilisées.

Des courbes de SAXS ont été enregistrées à plusieurs concentrations pour la P entière de NiV et pour la construction P₁₋₅₈₀. L'allure des courbes et les valeurs du rayon de giration (R_g) sont indépendantes de la concentration, démontrant l'absence d'agrégation (**Figure 34A-D**). La fonction P(r) calculée pour la P entière et pour le fragment P₁₋₅₈₀ ont une forme en cloche asymétrique lisse, avec un *Dmax* aux environs de 40 nm et un seul maximum aux environs de 10 nm (**Figure 34E**), caractéristique des protéines flexibles. Les graphiques de Kratky et de Porod-Debye confirment la présence importante de régions désordonnées dans ces deux protéines (**Figure 34F et G**). Dans le plot de Kratky, la courbe augmente aux valeurs élevées de q, et le graphique de Porod-Debye ne tend pas vers un plateau. Ces résultats sont en accord avec l'idée que la P contient de longues régions désordonnées.

Le spectre 2D [¹H-¹⁵N]-HSQC de la NiV P entière se caractérise par une faible dispersion des déplacements chimiques, caractéristique d'une protéine désordonnée (**Figure 35A**).

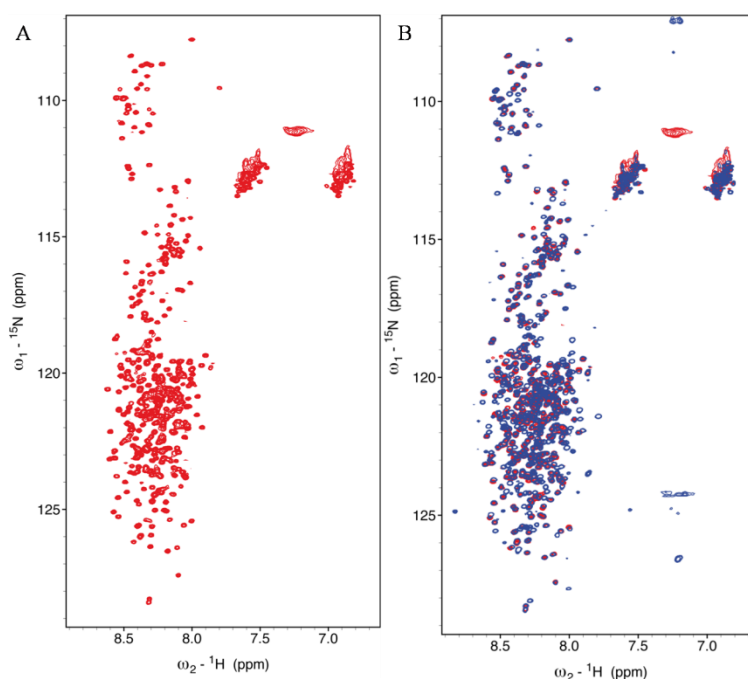


Figure 35 : Spectres RMN. A) Spectre 2D [¹H-¹⁵N]-HSQC de la NiV P entière (rouge). B) le spectre de A) superposé aux spectres 2D [¹H-¹⁵N]-HSQC des fragments P₁₋₁₀₀, P₉₁₋₁₉₀, P₉₁₋₂₅₀, P₂₂₂₋₃₁₉, P₃₀₀₋₄₀₁, P₃₈₇₋₄₇₉ et P₅₈₈₋₆₅₀ (bleu).

On dénombre environ 350 pics mais la faible dispersion des déplacements chimiques rend leur attribution difficile. Ce nombre est largement inférieur au nombre d'acides aminés de la protéine et on ne distingue aucune résonance qui pourrait correspondre aux résidus des régions structurées. Notamment, la taille du domaine P_{CD} (54 kDa) devrait conduire à un élargissement des pics au-dessous de la limite de détection. Le fait d'observer un nombre de pics inférieur à 480 (le nombre de résidus dans 1 P_{NTR}) suggère fortement que les quatre N-

terminaux se comportent comme des chaînes indépendantes dans le contexte de la protéine entière. Pour obtenir une information détaillée sur la partie N-terminale désordonnée, nous avons adopté une stratégie de « diviser pour régner » ; nous avons découpé la partie N-terminale de la protéine en peptides d'environ 100 acides aminés qui se chevauchent de 20 aa de chaque côté (voir **Tableau 3** et **Figure 5**). Dans les spectres RMN de ces différents peptides, les résonances de la chaîne principale et du C β ont été attribuées. La **Figure 35B** montre la superposition des spectres des différentes parties du Niv P_{NTR} (bleu) sur le spectre de la P entière (rouge). La grande majorité des résonances se superposent très bien, à l'exception de quelques pics. Ces différences peuvent s'expliquer par des « effets de bord » et par la présence de l'étiquette hexa-histidine en C-terminal.

Après l'attribution des résonances des différents spectres RMN pour le P_{NTR}, les déplacements chimiques pour les C α et C β ont été utilisés pour calculer la propension à former des structures secondaires (*Secondary Structure Propensity*, SSP) (Marsh, Singh et al. 2006). Le graphique du paramètre SSP en fonction de la séquence montre la présence de plusieurs îlots de structures secondaires fluctuantes, plus précisément des régions de propension positive correspondant à des hélices α résiduelles pour les régions 5 à 29, 37 à 55, 68 à 88, 120 à 126, 46 à 158, 170 à 186, 332 à 356 et 470 à 478 ainsi que des régions de propension négative correspondant à des structures étendues pour les régions allant 307 à 326 et 390 à 414 (**Figure 36**).

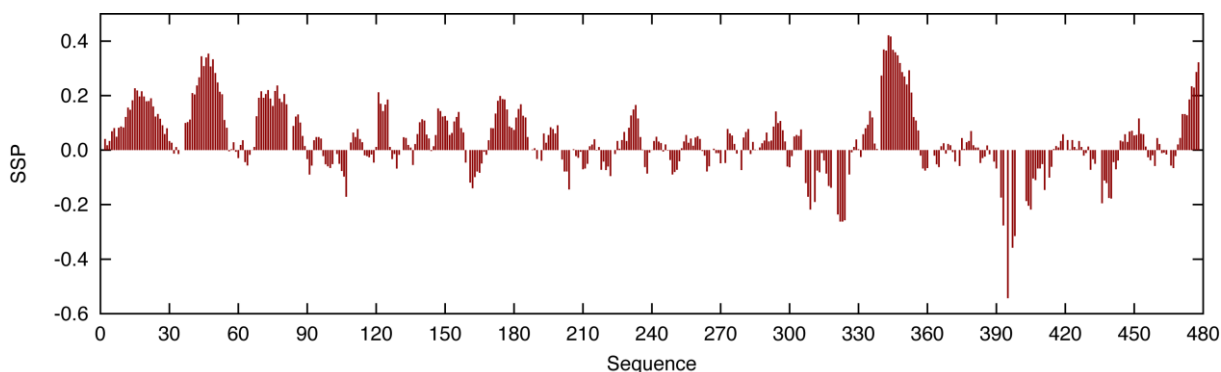


Figure 36 : Structures secondaires fluctuantes dans la région N-terminale (P_{NTR}) de la phosphoprotéine (P) du virus Nipah (NiV) détectées par la Résonance Magnétique Nucléaire (RMN). Le graphe représente la propension à former des structures secondaires (*Secondary Structure Propensity*, SSP) en fonction du numéro des résidus dans la séquence. Le SSP est calculé en combinant les valeurs des déplacements chimiques pour les C α et C β , obtenues expérimentalement pour P_{NTR}. Les valeurs positives sont caractéristiques des hélices α tandis que les valeurs négatives représentent des structures β (Wishart, Sykes et al. 1991; Marsh, Singh et al. 2006).

La première région prédite en hélice du côté N-terminal correspond à la région d'interaction avec N⁰. Les deux autres régions prédites en hélice allant des résidus 37 à 55 et 68 à 90 n'ont pas de fonction connue. Il en va de même pour la région de 332 à 356. Ces

hélices transitoires pourraient correspondre à des régions d'interaction avec d'autres partenaires et notamment avec la polymérase L.

Des données de SAXS ont été enregistrées pour les peptides P₁₋₁₀₀, P₉₁₋₁₉₀, P₉₁₋₂₅₀, P₂₂₂₋₃₁₉, P₃₈₇₋₄₇₉. Les valeurs de R_g, calculées en utilisant l'approximation de Guinier aux petits angles ($qR_g < 1.1$), ne dépendent pas de la concentration de protéine indiquant l'absence d'agrégation (non montré). Ces valeurs présentées dans le tableau sont très proches des valeurs de R_g calculées en utilisant la loi de puissance de Flory et des paramètres déterminés pour des protéines intrinsèquement désordonnées ou dénaturées contenant le même nombre de résidus (**Tableau 5**). Afin d'identifier les peptides les plus compacts et les plus étendus, nous avons calculé le rapport entre le R_g expérimental et le nombre d'acides aminés – la compaction normalisée. Ce paramètre indique que les peptides P₁₋₁₀₀ et P₉₁₋₂₅₀ sont les peptides les plus compacts alors que les peptides P₃₈₇₋₄₇₉ et P₅₈₈₋₆₅₀ sont les plus expansés (**Tableau 5**). Cette observation est confirmée par les valeurs de ν , obtenues expérimentalement à partir des courbes de SAXS. La forme plus compacte de P₁₋₁₀₀, relative aux autres peptides, peut être causée par son contenu élevé en hélices α , comme démontré par la RMN (**Figure 36**), en accord avec la proposition que les protéines intrinsèquement désordonnées semblent être plus compacte que les protéines dénaturée et cela du à la présence de structures secondaires (Bernado and Blackledge 2009). Il est intéressant de noter que le peptide P₅₈₈₋₆₅₀ semble être très allongé et le moins compacte de tous les peptides. La forme générale de la fonction P(r) en fonction de r pour les peptides P₁₋₁₀₀, P₉₁₋₁₉₀, P₉₁₋₂₅₀, P₂₂₂₋₃₁₉, P₃₈₇₋₄₇₉ est celle d'une courbe en cloche asymétrique avec un seul maximum déplacé vers les courtes distances interatomiques caractéristiques d'une flexibilité importante ou d'une molécule allongée (non montré).

Tableau 5 : Paramètres structuraux extraits des courbes de diffusion des rayons X aux petits angles (SAXS) pour les fragments de la phosphoprotéine (P) du virus Nipah (NiV) et leur comparaison avec des valeurs théoriques.

Protéine	R _g Guinier	D _{max}	R _g PID ^a	R _g protéines dénaturées ^a	Nmbre d'aa	Nmbre de charges	pI	% Prolines	Compact ion normalisée ^b	Valeurs de ν^c
P1-100	3.2 ± 0.1	12± 1	2.92	3.42	108	-8	5.11	1.9	2.96	0.526
P91-190	3.12 ± 0.1	12± 1	2.92	3.42	108	-12	4.68	3.7	2.89	0.574
P91-250	4.07 ± 0.1	15± 1	3.68	4.45	168	-18	4.64	4.8	2.42	0.552
P222-319	3.12 ± 0.1	11.5± 1	2.9	3.38	106	-16	4.51	5.7	2.94	0.562
P300-401			2.93	3.48	111	-4	5.94	10.8		
P387-479	3.22 ± 0.1	12.5± 1	2.84	3.3	102	0	7.07	8.8	3.16	0.546
P471-580					119	-12	5.15	3.4		
P655-709	1.4*		2.21	2.48	63	-7	5.31			
P588-650	2.56 ± 0.1	10± 1	2.35	2.68	72	-6	5.34	5.6	3.55	0.646
P1-580	10.8 ± 0.2	40± 1			588	-45	4.69	6		
P471-709	6.3 ± 0.2	22± 1			248	-20	4.9	3.6		
P entière	11.0 ± 0.2	40± 1			717	-53	4.69	5.6		

Le Met Nter est supposé être non clivé ; Le tag LEHHHHHHH est pris en compte dans le nombre d'aa et pour les calculs ; **a** La valeur de R_g théorique est calculée en utilisant la loi de puissance de Flory avec $\nu = 0.522$ et $R_0=2.54$ pour les PID (protéines intrinsèquement désordonnées) et $\nu = 0.598$ et $R_0=2.08$ pour les protéines dénaturées (Fitzkee et Rose ; Bernado et Blackledge). **b** La compaction normalisée est calculée en divisant le R_g, obtenu à partir de l'approximation de Guinier, par le nombre d'acides aminés dans la séquence ; la valeur obtenue est multipliée par 100 afin de se débarrasser des zéros. **c** Valeurs de ν est obtenu à partir de la représentation logI en fonction de logq de la courbe de diffusion expérimentale.* R_g de NiV P_{XD} extrapolé à concentration nulle à partir de données à 150 mM NaCl.

3.3.6. Modélisation de la phosphoprotéine de NiV

L'objectif de cette partie du travail est d'utiliser une approche similaire à celle utilisée avec la P de VSV dans le but de représenter la P de NiV sous la forme d'un ensemble de conformères. Cette approche repose, dans une première étape, sur l'utilisation des données de RMN pour générer un large ensemble de conformères qui reproduisent la propension des différents résidus à former des structures secondaires et leur combinaison avec les structures cristallines des domaines structurés. Dans une seconde étape, elle consiste à sélectionner un ensemble limité de conformères qui reproduisent collectivement les données de SAXS. Ce travail est en cours, mais nous présentons les résultats d'une modélisation *ab initio* du fragment P₄₇₁₋₇₀₉ en utilisant le programme DAMMIF. Une symétrie d'ordre 4 a été imposée dans ce processus qui a conduit à la génération d'un modèle pour lequel la courbe théorique s'ajuste au mieux avec la courbe de diffusion expérimentale. Parce que cette modélisation n'a pas une solution unique, nous avons généré de manière indépendante 40 modèles différents. Ces modèles consistent en un domaine allongé d'un côté, qui correspond vraisemblablement au domaine P_{CD} et peut facilement accommoder la structure cristalline du domaine, et en quatre « tentacules » qui correspondent aux jonctions et aux domaines P_{XD}.



Figure 37 : Modèles du fragment P₄₇₁₋₇₀₉ de la phosphoprotéine du virus Nipah (NiV) obtenus par le programme DAMMIF.

3.4. Discussion

I. La flexibilité des domaines C-terminaux de P

Les domaines P_X des virus Nipah et Hendra ont une identité de séquences d'environ 80%, néanmoins leurs structures cristallines ne sont pas identiques. Dans chaque structure, certaines interactions stabilisantes ne peuvent pas s'expliquer facilement par des contacts dus à l'empilement cristallin. La longue hélice du domaine P_X de NiV est stabilisée par un réseau de liaisons hydrogènes et électrostatiques, néanmoins des mesures en solution par RMN sont en désaccord avec une telle conformation. Les résultats de RMN indiquent que les domaines P_{XD} de NiV et HeV sont composés de trois hélices, ce qui suggère que les conformations cristallines ne sont pas figées en solution (**Figure 18**). Cette plasticité conformationnelle est également supportée par les mesures des rayons hydrodynamiques et les données de SAXS et il semble que la force ionique joue un rôle important. A faible force ionique les deux domaines semblent être très allongés tandis qu'à forte concentration en sel on observe leur compaction (**Tableau Figure 15**). Néanmoins ces observations ne peuvent pas être expliquées par un repliement de la protéine induit par le sel. Contrairement à ce qui a été observé pour le P_{XD} du virus des oreillons (MuV) (Kingston, Gay et al. 2008) les P_{XD} de NiV et HeV sont bien repliés en solution et leurs spectres CD sont caractéristiques des protéines riches en hélices α avec une faible influence du sel (**Figure 19**).

De plus les domaines P_{XD} de NiV et d'HeV se caractérisent par une stabilité thermodynamique faible. La température corporelle de la plupart des animaux infectés par ces virus est comprise entre 35 et 40 °C, et donc dans ces conditions, la concentration en sel pourrait avoir une forte influence sur le degré de repliement de NiV P_{XD} . En absence de sel, le T_m de NiV P_{XD} est de 40 °C, ce qui signifie qu'à 37°C dans ces conditions, la moitié des molécules seront dénaturées. Cette faible stabilité, associée à une grande flexibilité conformationnelle est une caractéristique commune des domaines C-terminaux d'interaction avec les nucléocapsides chez les *Rhabdoviridae* et chez les *Paramyxovirinae*. En effet, la structure du domaine P_{XD} du SeV est également très dynamique comme l'ont démontré des études par RMN (Blanchard, Tarbouriech et al. 2004). La structure du P_{XD} de MuV a été décrite comme un « molten globule » en solution, n'acquérant une structure tertiaire stable que dans le contexte de l'empilement cristallin (Kingston, Gay et al. 2008). Le domaine C-terminal de la P de VSV, dont la structure a été résolue par RMN, se caractérise aussi par une grande dynamique conformationnelle, comme en témoignent les expériences de relaxation par

RMN (Ribeiro, Favier et al. 2008). Dans le modèle où la phosphoprotéine marche sur la nucléocapside pour assurer le déplacement de la polymérase sur la matrice N-ARN (le modèle de « cartwheeling » (Curran 1998)), cette grande plasticité de structure pourrait moduler la stabilité du complexe N_{TAIL} - P_{XD} . La formation d'un complexe entre un P_{XD} flexible et la N pourrait avoir pour conséquence de réduire son temps de vie ce qui lui permettrait de s'attacher et de se détacher rapidement et continuellement, permettant le déplacement de la P sur le complexe N-ARN.

Des mesures d'affinité très différentes de 100 nM et 50 μ M ont été rapportées pour l'interaction entre P_{XD} et N_{TAIL} de MeV (Kingston, Baase et al. 2004; Bourhis, Receveur-Brechot et al. 2005; Shu, Habchi et al. 2012) alors que des mesures de l'ordre 10 μ M ont été rapportées pour NiV et HeV (Habchi, Blangy et al. 2011). En dépit de leur divergence, ces mesures sont peu représentatives de ce qui se passe dans des conditions physiologiques où ces molécules existent sous des formes oligomériques. Dans le contexte de la phosphoprotéine entière tetramérique l'augmentation de l'avidité de P pour le complexe N-ARN (les nucléocapsides) par la fixation simultanée de plusieurs P_{XD} sur plusieurs N_{TAIL} peut résulter en une interaction trop forte qui ne serait pas compatible avec le fonctionnement de la machinerie de réplication virale et donc des mécanismes diminuant l'affinité, comme par exemple la flexibilité des domaines P_{XD} , pourraient être avantageux pour le virus. Maintenant que les mécanismes d'interaction entre les protéines ou domaines isolés sont bien décrits il serait intéressant d'essayer de caractériser ces interactions en conditions oligomériques, un travail que nous avons entrepris récemment.

La séquence en acides aminés du domaine P_{XD} n'est pas conservée au sein des *Paramyxovirinae* à l'exception des résidus hydrophobes de l'intérieur du domaine (**Figure 38**). La charge globale n'est également pas conservée, les P_{XD} de Niv (pI 4.5), Hev (pI 4.6) et SeV (pI 5.7) ont une charge globale négative à pH 7.0 tandis que les P_{XD} de MeV (pI 10.1) et MuV (pI 9.4) ont une charge globale positive. La variation des séquences peut être expliquée par la faible pression évolutive exercée sur la séquence en acides aminés par la seule nécessité de préserver une interaction suffisante avec la nucléoprotéine. Chez la plupart des *Paramyxovirinae* la région C-terminale de la nucléoprotéine, le N_{TAIL} , constitue le site d'interaction avec le P_{XD} . C'est une région globalement désordonnée qui est également peu conservée parmi les différents membres des *Paramyxovirinae* (Habchi, Mamelli et al. 2010), et donc la seule contrainte sur les séquences de P_{XD} et N_{TAIL} serait de préserver leur interaction. Néanmoins le repliement du P_{XD} en faisceau de trois hélices α semble être

strictement conservé chez les *Paramyxovirinae* (**Figure 8**). La seule fonction connue de ce domaine est l'interaction avec la N et la conservation de ce repliement en faisceau de trois hélices pourrait constituer un motif minimal d'interaction.

```

[Sendai_virus]          -P-SKEKPTMHSRLRLVIESSPLRAEKAAAYVKSLSKCKTDQEVKAVMELV
[Human_parainfluenza_virus3] -P-KKVSSTMRSLVAVINNSNLSQSTKQSYINELKRCKNDEEVSELMDF
[Nipah_virus]          -P-MADDSSRDVIKTLIRTHIKDRELRSSELIGYLNKAENDEEIQEIANTV
[Hendra_virus]        -P-LADDASKDVVRTMIRTHIKDRELRSSELMDYLNRAETDEEVQEVANTV
[Cedar_virus]         -P-DDTDRSPMVLKIIIRQNIHDEEVKDELLSILEQHNTVEELNEIWNTV
[Mumps_virus]         QPVKDLAGRKVMITKMIADCVANPQMKQVFEQRLAKASTEDALNDIK---
[Parainfluenza_virus5] -PQKDLTGLKITLEQLAKDCISKPKMREEYLLKINQASSEAQLLDLK---
[Measles_virus]       -P-AALSGYKMTLLALIKECIPNQAQRQKFEMQVGGIRNEQDFKNLR---
[Measles_virus]       -P-DTGPARSVIRSIKSSRLEEDRKRYLMTLLDDIKGANDLAKFHQML
                        *      :      :      .      :      :      .      .

[Sendai_virus]          EEDIESLTN 568
[Human_parainfluenza_virus3] NEDVNNCQ - 604
[Nipah_virus]          NDIIDGNI - 709
[Hendra_virus]        NDIIDGNI - 707
[Cedar_virus]         NDYLDGNI - 737
[Mumps_virus]         RDIIRSAI - 391
[Parainfluenza_virus5] KAIIRSAI - 392
[Measles_virus]       REIIRSAQ 388
[Measles_virus]       MKIIMK - - - 507

```

Figure 38 : Alignement de séquences des domaines P_{XD} de structures connues ainsi que d'autres représentants des *Paramyxovirinae*. Les structures des domaines P_{XD} des virus Meanangle, Human parainfluenza 3 et 5 et Cedar ne sont pas connus et les séquences sont données en titre de comparaison. L'alignement a été fait en utilisant le programme ClustalW (Larkin, Blackshields et al. 2007).

II. Conservation de l'architecture du domaine central de la phosphoprotéine chez les *Paramyxovirinae*

La résolution récente des structures cristallines des domaines centraux d'oligomérisation (P_{CD}) des virus de la rougeole (Communie, Crepin et al. 2013) et des oreillons (Cox, Green et al. 2013) ainsi que la structure du P_{CD} de SeV (Tarbouriech, Curran et al. 2000) et la structure de NiV P_{CD} que nous avons déterminée pendant ces travaux permettent de faire quelques observations concernant les domaines d'oligomérisation des phosphoprotéines des *Paramyxovirinae*. La structure en faisceau surenroulé d'hélices α et le degré d'oligomérisation des phosphoprotéines semblent être conservés, malgré une faible conservation de séquences (**Figure 39**).

Il est intéressant de noter que tous ces domaines d'oligomérisation ont une longueur et un nombre de résidus formant le faisceau surenroulé de quatre hélices comparables : 9.2 nm et 63 résidus chez MeV, 9.8 nm et 64 résidus chez SeV, 9.6 nm et 64 résidus chez NiV et 8.2 nm et 58 résidus chez MuV. Une forme oligomérique semble être indispensable pour le bon fonctionnement du complexe de réplication/transcription chez le virus Sendai, modèle d'étude

des paramyxovirus, (Curran 1998; Curran and Kolakofsky 1999) ce qui peut créer une pression évolutive de conservation du degré d'oligomérisation.

```

Mumps_PCD      -----QSVISANEIMDLLRGMDARLQHLEQKVDKVLAQGSMTVQIKNELSTV-KTTLAT
Nipah_PCD      ---YDLETLC EESVLMGVIN--SIKLINLDMRLNHI EEQVKEIPKIINKLESI-DRV LAK
Measles_PCD    GDHYDDELFS DVQDIKTALAKIHEDNQKIISKLESLLLLLKG EVESIKKQINRQ-NISIST
Sendai_PCD     -ANYAEMTFNV CGLILSAEKSSARKVDENKQLLKQIQESVESFRDIYKRFSEYQKEQNSL
                .           :           .           : :           . * : : :           .           :

Mumps_PCD      IEGMMATVK--IMD- 287
Nipah_PCD      TNTALSTIEGHLVS- 573
Measles_PCD    LEGHLSSIMIAIPGL 377
Sendai_PCD     LMSNLSTLH-IITD- 433
                : : :           : .

```

Figure 39 : Alignement des domaines P_{CD} de structures connues chez les *Paramyxovirinae*. Les séquences correspondant au faisceau surenroulé central du P_{CD} ont été alignées en utilisant le programme ClustalW (Larkin, Blackshields et al. 2007).

Néanmoins, la raison pour laquelle la taille et le repliement en faisceau surenroulé du domaine P_{CD} sont conservés n'est pas claire et des délétions de différentes régions à l'intérieur du P_{CD} de SeV ne semblent pas affecter son état tétramérique (Curran, Boeck et al. 1995). Il a été également montré chez SeV que P_{CD} contient le site d'interaction avec la grande sous unité (L) de la polymérase virale (Curran, Pelet et al. 1994; Bowman, Smallwood et al. 1999). Cette observation et le fait que la protéine L est très conservée chez les *MNV* (Tordo, Poch et al. 1988; Poch, Blumberg et al. 1990) permettent d'émettre l'hypothèse que la longueur du P_{CD} chez les *Paramyxovirinae* est importante pour l'interaction avec la protéine L. Il est possible que des éléments structuraux de la polymérase nécessitent une longueur minimale du domaine P_{CD} afin d'assurer le bon fonctionnement du complexe L-P.

Les domaines d'oligomérisation de VSV et RaV sont très différents de ceux des *Paramyxovirinae* et leur structure et séquence ne sont pas conservées, contrairement à l'organisation et au degré d'oligomérisation de la P (**Figure 17 Chapitre 1**). Contrairement au P_{CD} de *Paramyxovirinae* qui constitue un site d'ancrage pour la polymérase virale, la seule fonction connue des P_{CD} des *Rhabdoviridae* est la dimérisation de la protéine. Ainsi la seule contrainte sur leur évolution serait de préserver leur état dimérique. Néanmoins chez les *Rhabdoviridae*, il a été montré que l'ablation du domaine de dimérisation P_{CD} rend la P monomérique (Gerard, Ribeiro Ede et al. 2009) et ne semble pas perturber sa fonction dans la transcription (Jacob, Real et al. 2001).

Les informations concernant l'interaction entre la P et la polymérase virale L des *Paramyxovirinae* sont plutôt rares mais il semble que la région C-terminale de la P (P_{CED}) contienne le site d'interaction avec la L (Smallwood, Ryan et al. 1994; Chen, Cortay et al. 2003). Cette interaction a été majoritairement étudiée pour le virus Sendai (SeV)

((Kolakofsky, Le Mercier et al. 2004) et références citées). Il a été proposé, dans le contexte du modèle de déplacement de type « cartwheeling » de la phosphoprotéine sur la matrice N-ARN (Curran 1998), que le domaine P_{CD} de SeV pourrait servir comme un axe moléculaire et permettre la rotation de la P tout en restant attachée sur la L (Tarbouriech, Curran et al. 2000). Cette proposition est faite sur la base de (1) l'observation que les résidus impliqués dans l'interaction avec L se trouvent à la surface du faisceau surenroulé dans la structure cristalline du SeV P_{CD} et sont très flexibles, (2) que les quatre faces du faisceau sont parfaitement symétriques et (3) que le diamètre du faisceau surenroulé est un peu plus large. Dans la structure du P_{CD} de NiV les quatre faces du faisceau surenroulé central ne sont pas identiques même si la chaîne principale du faisceau possède une symétrie d'ordre quatre (**Figure 29**). Ainsi la symétrie des chaînes latérales sur les faces est plutôt d'ordre deux ce qui est du aux propriétés géométriques du faisceau surenroulé. Dans l'hypothèse où NiV P_{CD} constitue également le site d'interaction avec L, cette non-équivalence des faces du faisceau surenroulé suggère que, lors d'une rotation du tétramère de P pendant son déplacement sur la nucléocapside, l'interface d'interaction avec la L serait continuellement modifiée. Ceci pourrait avoir un effet sur l'affinité entre les deux protéines. Dans les conditions où une des faces du NiV P_{CD} aurait plus d'affinité pour la L, la rotation hypothétique serait perturbée. Cette observation ne semble pas être en faveur d'un modèle de déplacement de type « cartwheeling » pour la NiV P. La symétrie de quatre ne semble pas non plus être parfaite dans le domaine d'oligomérisation P_{CD} du virus de la rougeole (Communie, Crepin et al. 2013).

Une différence notable entre les quatre structures de P_{CD} connues est la présence d'une couronne constituée de courtes hélices du côté N-terminal chez NiV et SeV alors qu'elle est absente des structures du P_{CD} de MeV et MuV (**Figure 30**). Chez MuV l'existence d'une telle couronne semble peu probable à cause de l'arrangement antiparallèle des hélices (Cox, Green et al. 2013). Chez MeV l'absence de cette couronne a été également confirmée par RMN (Communie, Crepin et al. 2013) et ne semble pas être importante pour la stabilité du domaine. Ainsi chez SeV et NiV, la couronne pourrait être impliquée dans l'interaction avec d'autres protéines. De ce point de vue, il est intéressant de noter la présence de deux résidus hydrophobes, Phe488 et Phe489, suivis par la Pro490, dont les chaînes latérales sont exposées au solvant (**Figure 28E**). Généralement les acides aminés hydrophobes forment le cœur hydrophobe de la protéine. Ce type d'acides aminés est exposé au solvant seulement lorsque cela est nécessaire, par exemple pour former le site d'interaction avec une protéine partenaire. Les résidus Phe488 et Phe489 pourraient ainsi constituer un site d'ancrage pour un partenaire

viral ou cellulaire, et pourquoi pas pour la protéine L. Dans l'hypothèse où ces résidus constitueraient le site d'ancrage initial pour la L, la fixation de cette dernière pourrait provoquer l'ouverture de la couronne et créer une conformation du P_{CD} spécifique du complexe L-P. L'importance de ces résidus pourrait être testée en les remplaçant par mutagenèse dirigée et en comparant la réplication du virus mutant avec le virus sauvage.

III. Structure d'une protéine tentaculaire

Il est difficile d'obtenir une description structurale à haute résolution de P car c'est une protéine modulaire flexible comprenant plusieurs domaines. Cette étude fournit pour la première fois une représentation atomique de ces protéines virales.

Pour NiV et VSV P, le spectre RMN a confirmé la présence de la région désordonnée P_{NTR} et de la jonction flexible entre les domaines structurés P_{CD} et P_{CTD}. La molécule dimérique ressemble à une pieuvre à quatre (VSV) ou huit (NiV) tentacules avec le corps constitué du domaine central dimérique ou tétramérique P_{CD}, de longs tentacules P_{NTR} et de tentacules plus courts constitués de la jonction flexible terminée par le domaine globulaire P_{CTD} ou P_{XD}. Les différentes parties flexibles de ces molécules bougent indépendamment les unes des autres sans interaction entre elles, à l'exception peut être des domaines C-terminaux (P_{XD}) de la NiV P.

Contrairement à ce qui a été observé pour les domaines P_{CTD} de la VSV P, les P_{XD} de la NiV P ne sont pas visibles sur les spectres RMN dans le contexte de la protéine entière. Cette observation, combinée à la suspicion de dimérisation des P_{XD} en solution (**Figure 16**) et à la concentration effective élevée de P_{XD} dans le contexte de la P entière tétramérique permet d'imaginer un scénario dans lequel les quatre P_{XD} dans une molécule de P forment des interactions stables. Dans un tel cas, et si ces interactions modifient ou masquent la surface de P_{XD}, impliquée dans l'interaction avec le N_{TAIL}, ce phénomène pourrait modifier l'affinité de la P pour les nucléocapsides (NC) et ainsi avoir un effet direct sur le déplacement de la P le long des NC et donc sur la réplication et la transcription virale. Un tel mécanisme constituerait une façon astucieuse de modifier l'affinité entre la P et les nucléocapsides.

D'autre part la disparition des P_{XD} des spectres RMN dans le contexte de la protéine tétramérique peut être causée par une dynamique de ces domaines à une échelle de temps de la milliseconde sans forcément former des interactions stables. Cette hypothèse est rendue plausible également par le fait que les modèles *ab initio* du fragment P₄₇₁₋₇₀₉, générés par DAMMIF, contiennent quatre tentacules bien séparés (**Figure 37**).

Une des propriétés caractéristiques des régions désordonnées est leur aptitude à établir des interactions moléculaires avec de multiples partenaires structurellement hétérogènes. L'absence de repliement confère aux régions désordonnées une plus grande surface de contact externe disponible par acide aminé et une souplesse de conformation permettant de s'adapter à la structure des partenaires. Ceci se traduit par une économie substantielle de longueur de séquence requise pour assurer une interaction spécifique entre différents partenaires par rapport à un domaine structuré dont la fraction de séquence enfouie, la rigidité et l'encombrement spatial limite le partenariat potentiel (Tompa and Fuxreiter 2008; Tompa, Fuxreiter et al. 2009).

Pour la phosphoprotéine de VSV, une méthode de sélection d'ensemble a permis de générer des ensembles de conformères compatibles à la fois avec les données de RMN et de SAXS. Dans ces ensembles, la protéine adopte une large gamme de conformations avec des dimensions et des orientations relatives des parties N- et C-terminales très différentes. Malgré le fait que nous n'ayons pas encore modélisé la NiV P (cette modélisation est en cours) nous pouvons spéculer que ses propriétés structurales et leurs implications seront similaires à celles de VSV P décrites ci-dessous. La grande flexibilité et la taille de ces protéines pourraient être déterminantes dans les fonctions de P au sein du complexe de transcription/réplication. En agissant comme une « brosse » entropique, les molécules de P pourraient auto-réguler leur espacement sur la nucléocapside. Les longs tentacules P_{NTR} flexibles, en échantillonnant un grand volume autour de la molécule favoriseraient la capture des partenaires viraux et cellulaires tels que N⁰, la protéine L, des kinases ou des karyophérines. Cela faciliterait également le positionnement et l'orientation correcte de la polymérase au site de synthèse de l'ARN et la livraison du substrat N⁰ au site d'encapsulation. Comparées à la taille de la particule virale ou de la nucléocapside, les dimensions de la molécule sont singulières. La particule en balle de fusil de VSV a une longueur moyenne de 200 nm, un rayon extérieur de 35 nm et une cavité centrale d'un rayon de 15 nm. Or, cette cavité est supposée contenir les molécules P dont la distance maximale entre les extrémités N- et C-terminale est de l'ordre de 18 nm, donc supérieure au rayon de la cavité. La distance entre les deux domaines C-terminaux d'un dimère peut atteindre 9 nm soit une valeur bien supérieure à la distance séparant deux tours successifs de la nucléocapside (environ 5 nm). Donc, un même dimère de P pourrait connecter des tours successifs de la NC et stabiliser la particule virale.

Les particules de NiV sont pléomorphes, de dimension allant de 40 à 600 nm (Hyatt, Zaki et al. 2001) et à ce jour nous ne disposons d'aucune donnée structurale détaillée concernant leur organisation interne. Des reconstructions récentes par cryo-microscopie

électronique pour un virus apparenté, le virus de la rougeole (MeV), ont révélé qu'à l'intérieur du virion, les nucléocapsides (NC) sont majoritairement recouvertes par une couche de protéine de matrice (M) (Liljeroos, Huisken et al. 2011), un phénomène qui peut s'avérer commun pour tous les *Paramyxovirinae*. L'espace entre la NC et la couche de M est très étroite pour permettre à la P d'y résider ce qui signifie que dans un virion, dont la majorité des NC sont recouvertes de M, elles ne seront pas accessibles pour la phosphoprotéine et donc la question se pose de savoir comment la P est recrutée dans de telles particules virales ? Une des possibilités serait que seulement quelques molécules de P soient recrutées dans le virion en s'attachant sur les parties de NC qui ne sont pas recouvertes de M. Dans un tel scénario, après la fusion du virion avec la cellule hôte et après le démantèlement de la couche de M, la P assurerait la transcription et la réplication virale en se déplaçant sur le complexe N-ARN (les nucléocapsides) par le mécanisme de « cartwheeling ».

Néanmoins concernant le virus Nipah plusieurs observations nous semblent en désaccord avec un tel mécanisme de déplacement de la P. Tout d'abord le volume occupé par les quatre P_{NTR} de NiV P est très grand et la dimension maximale de 40 nm de la molécule signifie que les P_{NTR} occupent le volume d'une sphère d'un rayon d'environ 20 nm (**Figure 40**).

A pH physiologique le P_{NTR} de NiV P (pI 4.56) et le N_{TAIL} de la N (pI 4.8) seraient négativement chargés et se repousseraient. Dans ces conditions, un phénomène de compaction de la P, lorsqu'elle est attachée à la surface des nucléocapsides, semble peu probable. Donc la P occuperait un volume considérable ce qui aurait comme effet de rendre son déplacement sur les nucléocapsides difficile.

Un deuxième argument défavorisant le déplacement de la P sur les nucléocapsides vient des mesures d'affinité entre la P entière et des nucléocapsides recombinantes (Yabukarski, Carriel, Fender, résultats non publiés). Ces mesures préliminaires ont montré que l'affinité entre les deux partenaires est relativement forte (K_D d'environ 50 nM) et que la dissociation de la P est très lente ce qui est incompatible avec un modèle de « cartwheeling » dans lequel l'interaction entre P et la nucléocapside doit être dynamique afin de permettre le déplacement de la P.

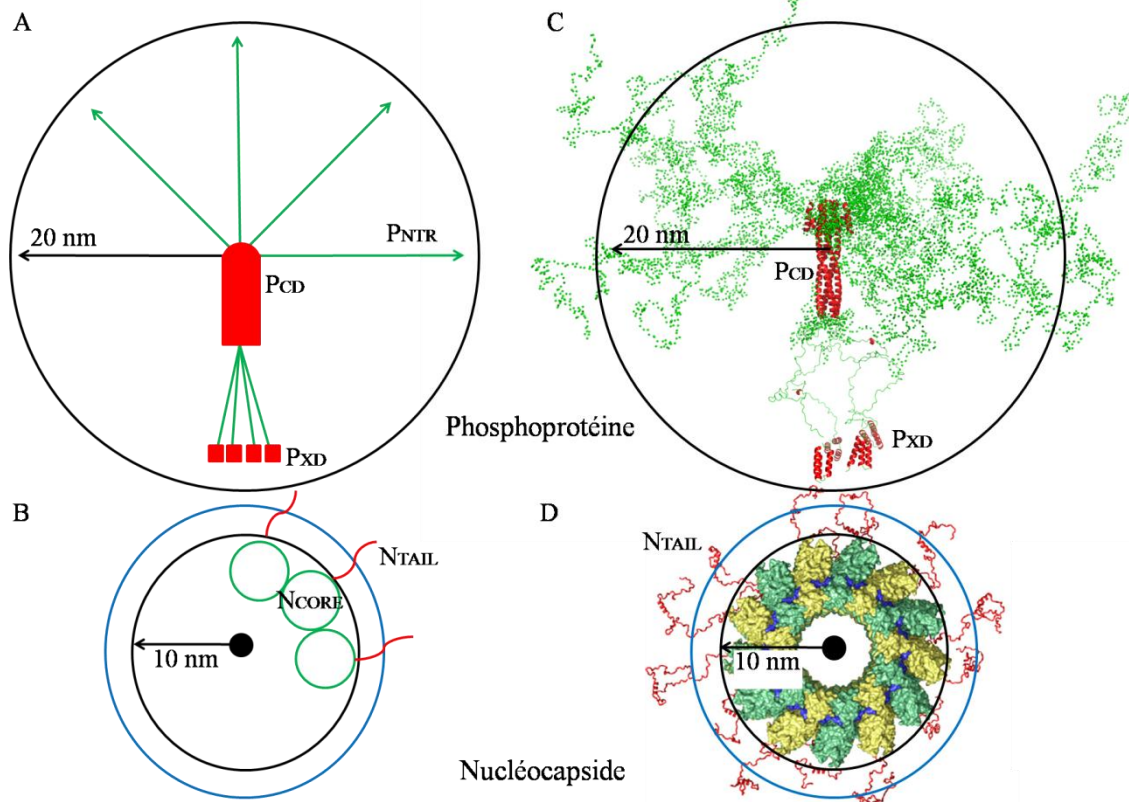


Figure 40 : La phosphoprotéine et la nucléocapside. **A)** Représentation schématique de la NiV P. Les domaines structurés (P_{CD} et P_{XD}) sont colorés en rouge, les parties désordonnées (P_{NTR} et la jonction entre P_{CD} et P_{XD}) en vert. Le cercle représente l'espace échantillonné par le P_{NTR} . Son rayon est de 20 nm et représente le $D_{max}/2$. Le centre du cercle se trouve sur le P_{CD} car la région de jonction et les P_{XD} contribuent très peu aux dimensions globales de la P. **B)** Schéma de la nucléocapside (NC) de NiV vue dans le sens de son axe. Le cercle noir représente la NC de 20 nm de diamètre, comme déterminé par microscopie électronique (Hyatt, Zaki et al. 2001). Les cercles verts représentent les N_{CORE} de chaque N et les filaments en rouge les N_{TAIL} . Le cercle bleu représente l'espace autour des N_{CORE} échantillonné par N_{TAIL} ; son diamètre n'est pas connu mais les rayons de giration de cross section (R_c) pour la NC, contenant le N_{TAIL} ou pas, sont très proches (Jensen, Communie et al. 2011). **C)** Modèles pseudo atomiques de la NiV P entière. Les résidus des régions flexibles sont représentés par des billes comme modélisé par le programme RANCH (Bernado, Mylonas et al. 2007). Quatre modèles sont superposés mais les domaines P_{XD} d'un seul modèle sont représentés pour plus de clarté. La coloration est la même qu'en A). **D)** Représentation d'un tour de NC de MeV (virus apparenté à NiV) vue dans le sens de son axe. Les N_{CORE} sont colorés en vert et jaune, les N_{TAIL} en rouge. Modèle issu de (Jensen, Communie et al. 2011).

Toutes les phosphoprotéines des *Paramyxovirinae* sont supposées être des protéines modulaires contenant une longue région N-terminale (P_{NTR}) flexible appartenant à la famille des protéines intrinsèquement désordonnées (Karlin, Longhi et al. 2002; Karlin, Ferron et al. 2003; Habchi, Mamelli et al. 2010). Néanmoins chez HeV et NiV, cette région est de 100 à 200 résidus plus longue que celles des autres virus dans la sous-famille (Wang, Michalski et al. 1998; Wang, Yu et al. 2000). Les résidus additionnels se trouvent dans la région décrite comme étant le site d'interaction avec les protéines STAT1 (résidus 100 à 150) (Shaw, Garcia-Sastre et al. 2004). Cette région semble être conservée dans toutes les souches de NiV

et HeV que nous avons analysé (**Figure 4**) et son absence chez les autres membres de la sous-famille suggère une évolution divergente pendant laquelle les Henipavirus ont « acquis » ou « inventé » cette région.

La résonance magnétique nucléaire a identifié plusieurs îlots de structures secondaires résiduels dans le P_{NTR} de NiV et notamment les régions des résidus 1 à 30, 40 à 60, 65 à 90 et 340 à 360 (**Figure 36**). Dans la partie N-terminale, les résidus 1 à 30 sont impliqués dans l'interaction avec N⁰ (voir **Chapitre 4**), néanmoins la fonction des autres régions est inconnue. Une possibilité serait que, en plus du domaine d'oligomérisation P_{CD}, les régions des résidus 40 à 60 et 65 à 90 soient impliquées dans l'interaction avec la polymérase virale, par homologie avec les résidus 41 à 106 de la P de VSV qui constituent le site d'interaction avec la polymérase de VSV (Rahmeh, Morin et al. 2012). Dans ce contexte il est très intéressant de noter que la région de la NiV P comprenant les résidus 81 à 113 semble être importante pour sa fonction de cofacteur de la polymérase virale (Ciancanelli, Volchkova et al. 2009) ; l'existence d'une hélice α préformée dans la région des résidus 65 à 90 pourrait faciliter une éventuelle interaction avec la L. Une telle interaction pourrait aussi jouer un rôle dans le transfert du N⁰ sur l'ARN néo-synthétisé.

Il est intéressant de noter que la région 340 à 360 peuple une conformation α hélicoïdale 40% du temps et ainsi représente la partie du P_{NTR} qui contient le plus de structure secondaire résiduelle. Cette région, prédite partiellement structurée par le D-score (**Figure 5**) et partiellement conservée chez les NiV et HeV (**Figure 4**), est flanquée par des séquences riches en résidus proline. Une recherche par le programme SH3-Hunter (Ferraro, Peluso et al. 2007) a identifié le fragment **R**₃₈₉**Q****R****P****G****T****P**₃₉₅ comme site potentiel d'interaction avec des domaines SRC homology-3 (SH3). D'autre part, le fragment **P**₃₆₂**P****Y****H**₃₆₅ ressemble étonnamment à la séquence consensus PPxY (ou x est n'importe quel acide aminé) reconnue par les domaines WW (Zarrinpar, Bhattacharyya et al. 2003). Ainsi cette région de la P, comprenant les résidus 360 à 400, commune aux protéines V et W, pourrait constituer un motif d'interaction avec d'autres protéines. A ce jour, absolument aucune fonction n'est attribuée à cette région de la P, et il serait très intéressant d'étudier son rôle dans le contexte de la réplication ou dans le contexte de l'inhibition du système immunitaire cellulaire.

Les résultats que nous avons obtenus décrivent la phosphoprotéine des MNV comme un acteur très dynamique dans les processus de réplication/transcription et comme une protéine charnière capable d'interagir avec de nombreuses protéines cellulaires ou virales.

Chapitre 4 : Interaction entre la phosphoprotéine du virus Nipah et la protéine STAT1

4.1. Introduction

Depuis des millions d'années les virus infectent des organismes appartenant aux différents domaines du monde vivant et ont développé des « outils » pour mieux envahir leurs hôtes tandis que les cellules hôtes ont évolués pour se défendre. Parmi les *Mononegavirales*, différentes stratégies d'interaction avec leur hôte sont adoptées par les différents virus. Certains, comme le virus de la stomatite vésiculaire, ont adopté une stratégie de réplication très rapide associée à l'inhibition de la production des protéines cellulaires dans le but de ne pas laisser du temps à la cellule hôte pour réagir. D'autres, tel que le virus de la rage, combinent un faible taux de réplication avec des moyens efficaces pour inhiber la signalisation cellulaire et rester « inaperçus » par les systèmes de défense de l'hôte. Tous ces virus ont donc en commun d'interagir d'une manière ou d'une autre avec le système immunitaire cellulaire.

Les henipavirus interfèrent avec le système immunitaire de l'hôte par l'intermédiaire des protéines V, W, C et dans un moindre degré de la protéine P (voir **Chapitre 1** et (Eaton, Broder et al. 2006) et références citées). Ces protéines agissent à plusieurs niveaux et à différentes étapes de la signalisation cellulaire. En particulier, les protéines V, W et P inhibent la voie de signalisation des interférons α/β et γ en interagissant avec la protéine STAT1 (Rodriguez, Cruz et al. 2004; Shaw, Garcia-Sastre et al. 2004). Un site essentiel pour cette interaction a été localisé dans la région commune N-terminale de ces différentes protéines allant des résidus 110 à 140 (Ciancanelli, Volchkova et al. 2009).

Aujourd'hui, aucune information structurale ou thermodynamique n'est disponible concernant cette interaction. Afin de mieux comprendre les détails moléculaires de cette interaction, nous nous sommes intéressés au complexe formé entre la partie N-terminale de la P/V/W et la protéine STAT1 (Signal Transducer and Activator of Transcription 1). Dans le but d'étudier ce complexe nous avons produit des fragments des deux partenaires et nous avons étudié leur interaction.

Architecture de STAT1

Dans ce travail, nous avons utilisé un fragment de STAT1 (aa 132-683) qui contient le domaine en faisceau d'hélices, le domaine de fixation à l'ADN, le domaine de connexion et le domaine SH2 (**Figure 1**). Selon des études récentes, la partie N-terminale de STAT1, absente de notre construction, n'est pas responsable de l'interaction avec P/V/W de Nipah (Rodriguez, Cruz et al. 2004) et nous avons arrêté notre construction au résidu 683 car (1) c'est le dernier résidu visible dans la densité électronique de la structure cristalline (Chen, Vinkemeier et al.

1998), ce qui suggère que le reste de la molécule est désordonné et (2) car la partie C-terminale manquante est responsable de la dimérisation de STAT1.

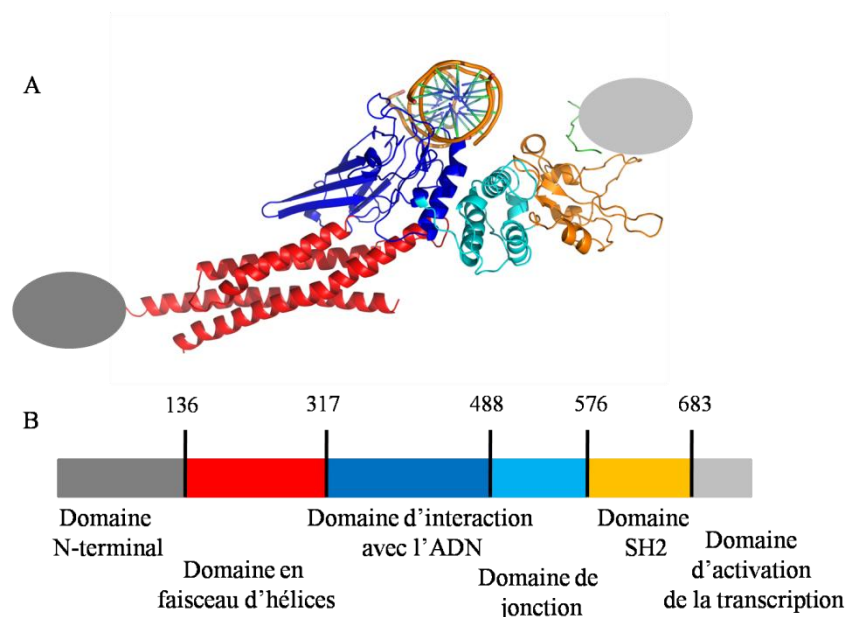


Figure 1 : Architecture de la protéine STAT1h. A) La structure cristalline de STAT1h (PDB code : 1BF5) et B) représentation schématique des différents domaines de la protéine : le domaine N-terminal n'est pas présent dans la structure (gris foncé), le domaine en faisceau d'hélices (rouge), le domaine d'interaction avec l'ARN (bleu) le domaine de jonction (turquoise), le domaine SH2 (jaune) et le domaine d'activation désordonnée dans la structure cristalline (gris clair). En orange est représenté le double brin d'ADN.

4.2. P₉₁₋₁₉₀ est intrinsèquement désordonné mais contient des structures secondaires fluctuantes

La stabilité à long terme du peptide P₉₁₋₁₉₀ a été significativement améliorée par l'ajout de 50mM glutamate et de 50 mM arginine dans la solution tampon (Golovanov, Hautbergue et al 2004). Nous avons obtenu une masse moléculaire de 11.4 ± 1 kDa par SEC-MALLS-RI, proche de la masse moléculaire de 11.9 kDa calculée à partir de la séquence, qui confirme l'état monomérique de la protéine jusqu'à des concentrations de 3 mM (**Figure 2A**).

Des expériences de diffusion des rayons X aux petits angles sur P₉₁₋₁₉₀ ont été enregistrées sur la ligne BioSAXS BM29 à l'ESRF (Grenoble). Les valeurs des paramètres structuraux sont données dans le **Tableau 5** du **Chapitre 3** mais nous avons détaillé l'explication ci-dessous. Des courbes de diffusions sont collectées pour des vecteurs de diffusion q allant de 0.05 à 3.0 nm^{-1} et des concentrations de protéine allant de 1.5 à 2.7 mg.mL^{-1} . Les profils obtenus ont la même forme, indiquant l'absence d'agrégation (**Figure 2C**). Le rayon de giration (R_g) calculé en utilisant l'approximation de Guinier pour des

valeurs de $Q.R_g < 1.3$ ne varie pas avec la concentration, indiquant également l'absence d'agrégation (**Figure 2D**).

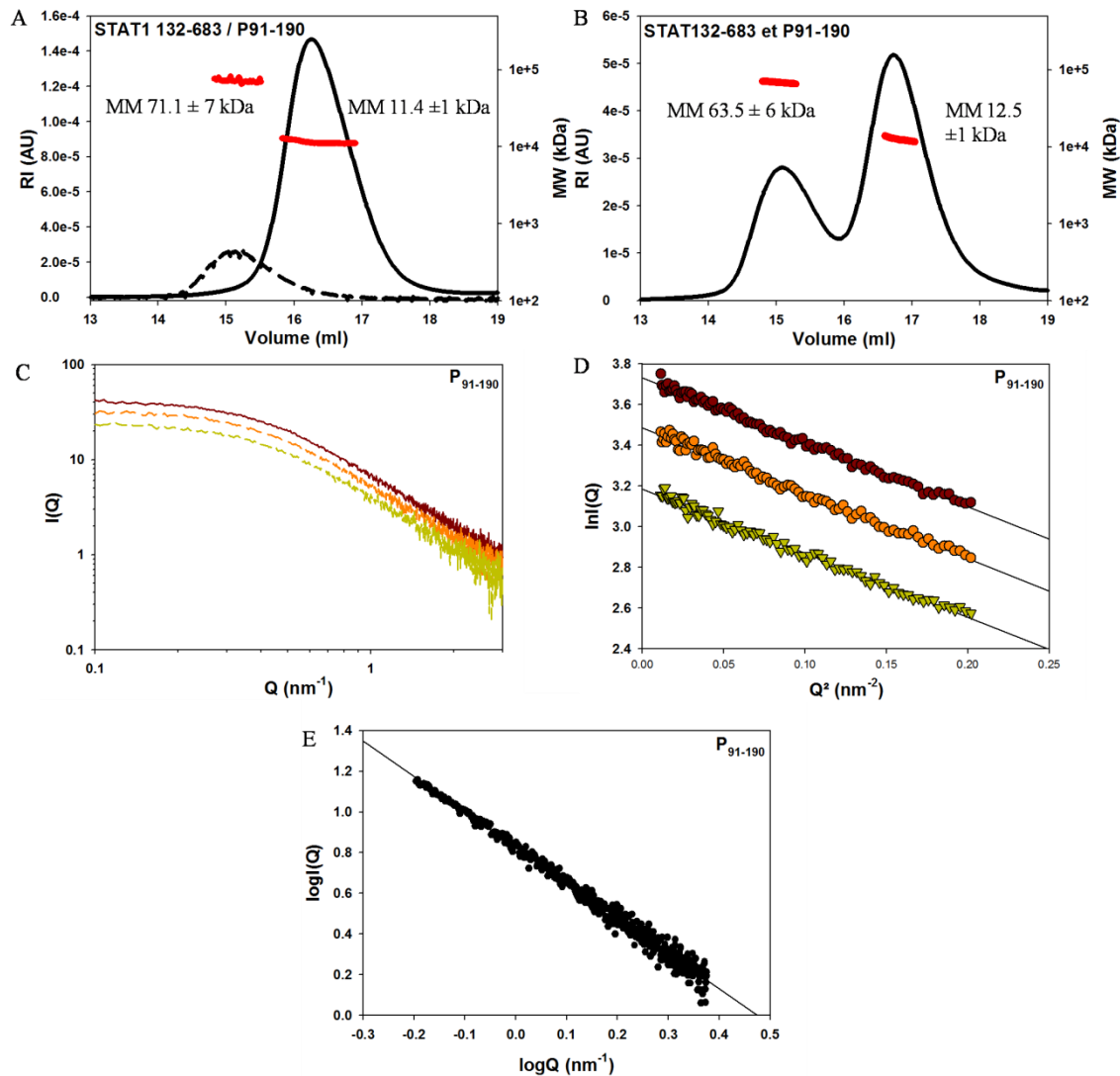


Figure 2 : Expériences de diffusion des rayons X aux petits angles (SAXS) sur le fragment P_{91-190} de la phosphoprotéine du virus Nipah (NiV) et de SEC-MALLS-RI sur ce fragment P_{91-190} et la protéine $STAT1_{132-683}$. A) Profils d'élution de P_{91-190} et $STAT1_{132-683}$ injectées séparément (P_{91-190} ligne continue et $STAT1_{132-683}$ ligne en pointillé) et B) injectées après incubation, sur colonne de gel filtration Superdex S200, équilibrée dans 20 mM Tris-HCl pH 7.5, 150 mM NaCl. En rouge les masses moléculaires (MM) à travers les pics d'élution, calculées à partir de la diffusion statique de la lumière et la réfractométrie pour chaque protéine. C) Les courbes de SAXS ont été enregistrées pour des valeurs de q entre 0.05 nm^{-1} et 3 nm^{-1} à trois concentrations de protéine (2.7 , 2.0 et 1.5 mg.ml^{-1}). D) Analyse de Guinier dans la région de $q.R_g < 1.1$. E) Représentation de $\text{Log } I(q)$ en fonction de $\text{Log } q$ aux angles moyens.

Le R_g moyen de $3.2 \pm 0.1 \text{ nm}$ est intermédiaire entre la valeur de 2.9 nm calculée pour une protéine intrinsèquement désordonnée de même taille en utilisant l'équation de Flory (**Chapitre 2**) et les paramètres déterminés par (Bernado and Blackledge 2009), et la valeur de 3.4 nm calculée pour une protéine dénaturée en utilisant la même équation et les paramètres déterminés par (Kohn, Millett et al. 2004). Le graphe de $\text{Log } I(q)$ en fonction de $\text{Log } q$ est linéaire pour les valeurs de q comprises entre 0.63 et 2.40 nm^{-1} . La pente du graphe fournit

une estimation du paramètre Dm de 1.73 (**Figure 2E**) correspondant à une valeur de l'exposant de Flory (ν) de 0.578, proche de la valeur estimée pour des protéines dénaturées (Kohn, Millett et al. 2004).

La masse moléculaire calculée à partir de l'intensité diffusée à angle zéro $I(0)$ est de 14 ± 2 kDa, en accord avec l'état monomérique de la protéine. Le spectre de dichroïsme circulaire dans l'UV lointain est caractéristique d'une protéine désordonnée (**Figure 3A**) et, en supposant que le signal à 222 nm soit uniquement attribué à la présence d'hélices α , le contenu moyen en hélice α est d'environ 7% (Luo and Baldwin 1997). L'addition de 2 M de triméthylamine N-oxide (TMAO), un composé connu pour stabiliser les formes compactes des protéines, provoque une augmentation du contenu en hélice α d'environ 2%.

Nous avons également étudié cette protéine par RMN qui est une technique sensible à la présence de structure locale et qui informe de manière quantitative sur la présence de structures secondaires transitoires. Le spectre RMN à deux dimensions par corrélation hétéronucléaire 2D [^1H - ^{15}N]-HSQC (*Heteronuclear Single Quantum Coherence*) mesuré pour la protéine P₉₁₋₁₉₀ présente une faible dispersion des déplacements chimiques caractéristique des protéines désordonnées (**Figure 3B**).

Après l'attribution des résonances des spectres RMN, les déplacements chimiques secondaires ont été calculés pour les résonances des C α et C β comme la différence entre les déplacements chimiques mesurés dans la protéine et ceux mesurés pour les chaque acides aminés dans des peptides courts dépourvus de structure. A partir des déplacements chimiques secondaires, un paramètre de propension à former des structures secondaires SSP (*Secondary Structure Propensity*) a pu être calculé (**Figure 3C**) (Wishart, Sykes et al. 1991; Marsh, Singh et al. 2006). Ces expériences démontrent que les régions comprenant les résidus 120 à 127, 136 à 158 et 169 à 190 adoptent transitoirement des conformations en hélices α tandis que les régions comprenant les résidus 95 à 106 et 158 à 169 adoptent une conformation étendue (conformations en structures β). La propension moyenne à former des hélices α est de 7%, en accord avec le contenu en hélices α estimé par CD. L'ensemble de ces résultats démontre que P₉₁₋₁₉₀ est une protéine intrinsèquement désordonnée, dont trois segments forment transitoirement des hélices α .

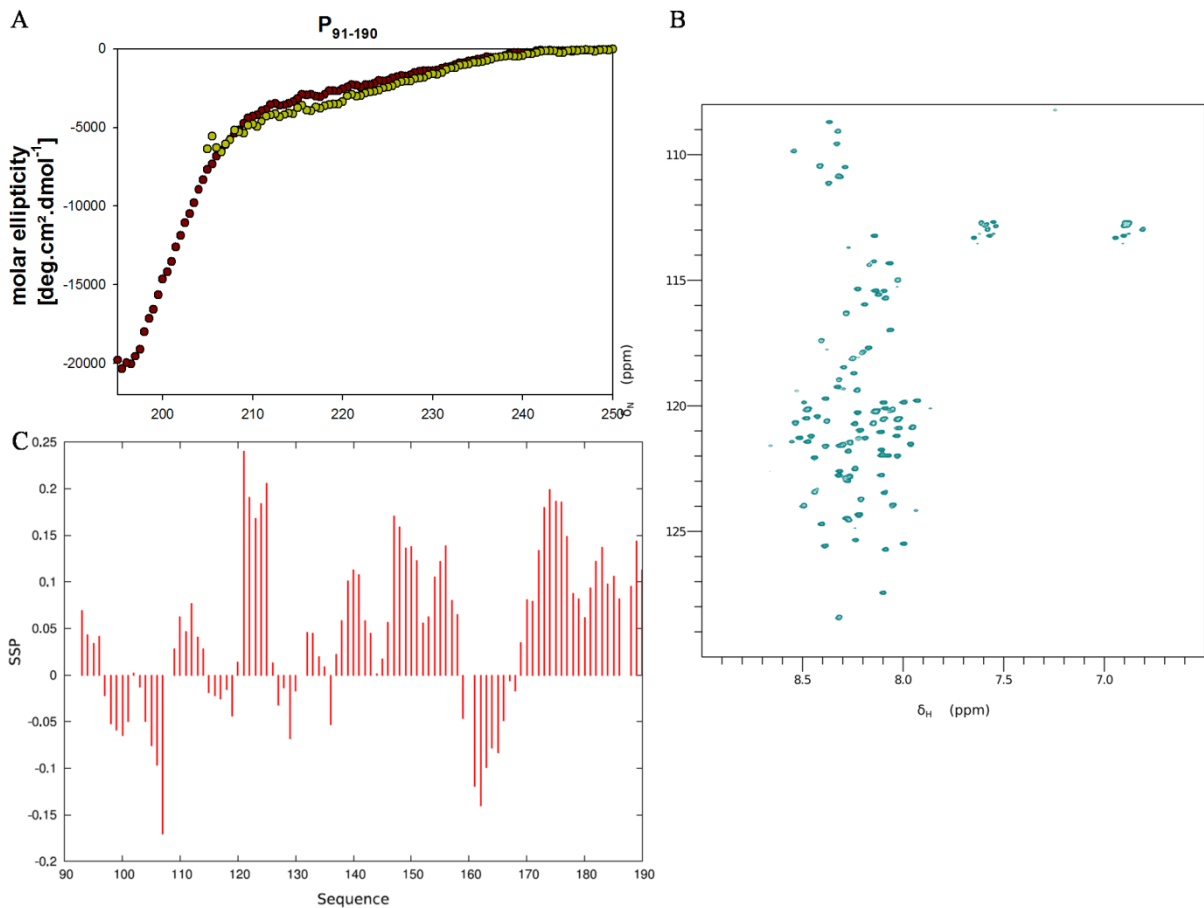


Figure 3 : Expériences de CD et RMN sur le fragment P₉₁₋₁₉₀ de la phosphoprotéine du virus Nipah (NiV). **A)** Spectre CD dans l’UV lointain dans 10 mM NaP à une concentration en protéine de 20 μ M. **B)** Spectre RMN 2D [¹H-¹⁵N]-HSQC. **C)** Le graphe représente la propension à former des structures secondaires (Secondary Structure Propensity, SSP) en fonction du numéro des résidus dans la séquence. Les valeurs positives sont caractéristiques des hélices α tandis que les valeurs négatives représentent des conformations étendues (des structures β). Les expériences sont effectuées à 20°C.

4.3. Interaction entre P₉₁₋₁₉₀ et STAT1

Le fragment STAT1₁₃₂₋₆₈₃ produit en bactéries élue de la colonne de chromatographie d’exclusion sous la forme d’un seul pic (**Figure 2A**). La masse moléculaire de 71.1 ± 7 kDa mesurée par SEC-MALLS-RI est en accord avec la masse moléculaire de 64.1 kDa calculée à partir de la séquence en acides aminés, indiquant que la protéine est monomérique en solution. La formation d’un complexe entre STAT1₁₃₂₋₆₈₃ et P₉₁₋₁₉₀ n’est pas détectée par chromatographie d’exclusion de taille, puisqu’après incubation de STAT1₁₃₂₋₆₈₃ à une concentration de 54 μ M en présence de P₉₁₋₁₉₀ à une concentration de 540 μ M, les deux protéines ont élué de la colonne aux mêmes volumes d’élution que les protéines individuelles et aucun autre pic n’a pu être détecté (**Figure 2B**). La masse moléculaire mesurée par MALLS-RI n’indique aucune modification de la masse moléculaire des composants éluant dans chacun des pics.

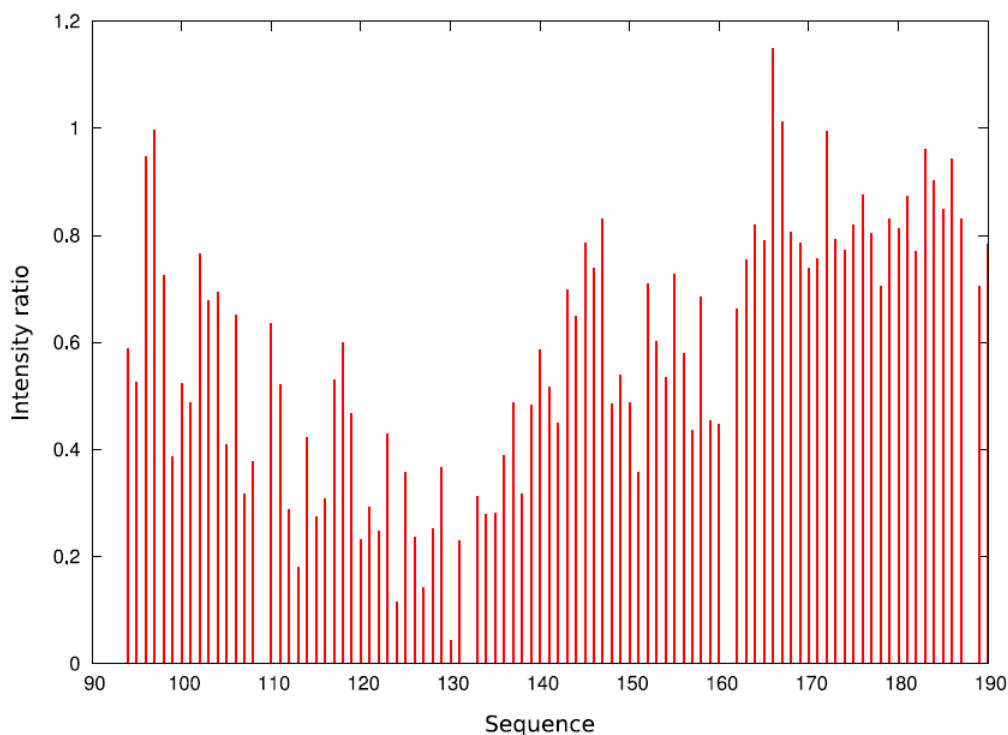


Figure 4 : Expérience de titrage de P₉₁₋₁₉₀ par STAT1₁₃₂₋₆₈₃ suivie par RMN. Le graphique représente le rapport des intensités des pics correspondant aux résonances amides des spectres RMN 2D [¹H-¹⁵N]-HSQC de P₉₁₋₁₉₀ en fonction de la séquence en acides aminés, avant (**Figure 3B**) et après l'ajout de STAT1₁₃₂₋₆₈₃ dans la cuve de mesure.

Une interaction entre P₉₁₋₁₉₀ et STAT1₁₃₂₋₆₈₃ a toutefois pu être détectée dans une expérience de titrage suivie par RMN. Le spectre 2D [¹H-¹⁵N]-HSQC obtenu pour un mélange contenant la P₉₁₋₁₉₀ stablement marquée avec ¹⁵N à une concentration de 22μM et STAT1₁₃₂₋₆₈₃ non marquée à une concentration de 44μM indique que l'intensité des résonances amides correspondant aux résidus 100 à 145 de P est réduite par comparaison avec un spectre 2D [¹H-¹⁵N]-HSQC obtenu pour P₉₁₋₁₉₀ à une concentration de 92μM en absence de STAT1₁₃₂₋₆₈₃ (**Figure 4**).

Un complexe hétérodimérique entre P₉₁₋₁₉₀ et STAT1₁₃₂₋₆₈₃ aurait une masse moléculaire d'environ 75 kDa et sa formation devrait conduire à la disparition complète des résonances amides pour les acides aminés de P₉₁₋₁₉₀ qui interagissent avec STAT1₁₃₂₋₆₈₃. La réduction partielle de l'intensité indique donc que dans ces conditions, une fraction seulement de P₉₁₋₁₉₀ fixe STAT1₁₃₂₋₆₈₃ et permet d'estimer que la constante de dissociation du complexe entre les deux protéines est comprise entre 10 μM et 100 μM. Nous observons également une faible diminution des intensités pour les résidus 145 à 165 qui ne font pas partie du site d'interaction avec STAT1 documenté dans (Ciancanelli, Volchkova et al. 2009). Nos résultats suggèrent donc que le site d'interaction avec STAT peut être plus étendu.

Afin de caractériser de manière plus précise l'interaction entre ces deux protéines, nous avons utilisé la calorimétrie de titrage isotherme (ITC). Sur la base de l'estimation de la

constante de dissociation par spectroscopie RMN entre 10 et 100 μM , nous avons évalué qu'une concentration de STAT₁₃₂₋₆₈₃ de 1 mM (64 mg.mL⁻¹) et une concentration 10 à 20 fois supérieure de P₉₁₋₁₉₀ (130 à 260 mg.mL⁻¹) seraient nécessaires pour obtenir une courbe complète de titrage. Néanmoins, des courbes de titrage ont pu être obtenues en titrant STAT₁₃₂₋₆₈₃ à une concentration de 70 μM avec une solution de P₉₁₋₁₉₀ à une concentration de 1.5 mM. Bien que la saturation ne soit pas atteinte, l'ajustement de ces courbes avec un modèle de fixation à un seul site fournit une constante de dissociation de $20 \pm 2 \mu\text{M}$ en accord avec la valeur estimée par RMN (**Figure 5**).

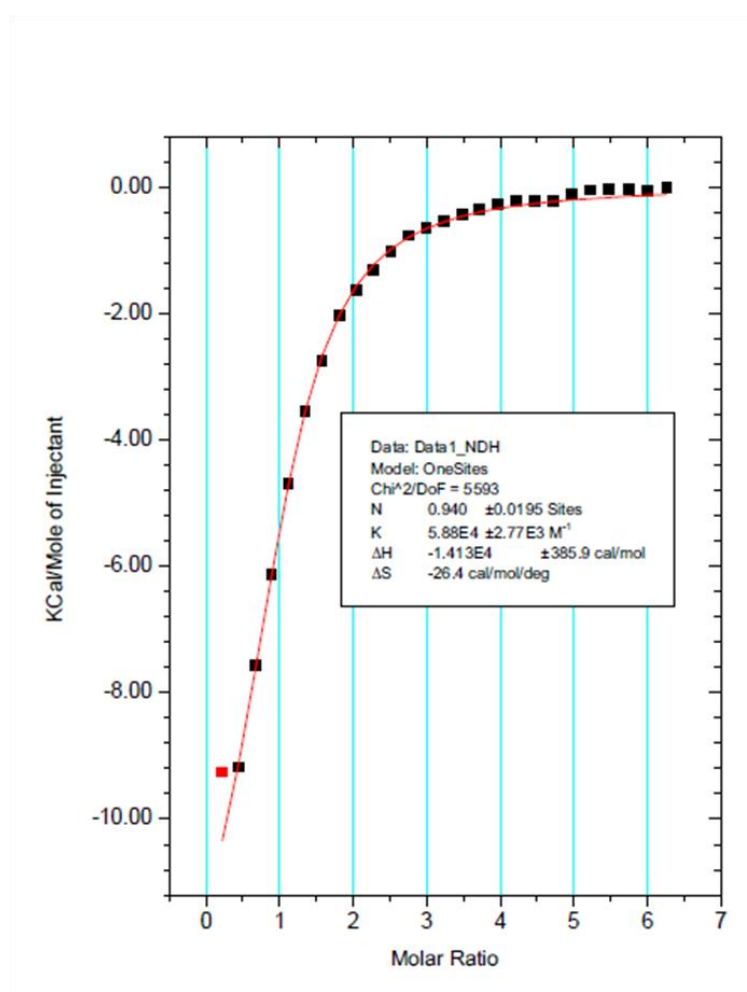


Figure 5 : Expérience de titrage de STAT₁₃₂₋₆₈₃ par P₉₁₋₁₉₀ par ITC. L'expérience est effectuée à 20°C en utilisant STAT₁₃₂₋₆₈₃ comme récepteur et P₉₁₋₁₉₀ comme ligand.

La variation d'énergie estimée à partir de cette constante de dissociation est de 6.3 kcal.mol⁻¹. L'interaction est caractérisée par un changement d'enthalpie négatif de l'ordre de 14 kcal.mol⁻¹ et donc une variation d'entropie négative de -26.4 cal.mol⁻¹.K⁻¹. La faible affinité est donc le résultat d'une grande pénalité entropique ($\Delta S \ll 0$) qui est compensée par un changement d'enthalpie également très important ($\Delta H \ll 0$). En première approximation, un $\Delta H \ll 0$ indique que la formation de liaisons hydrogène et d'interactions de van der Waals

pendant l'interaction entre les deux partenaires est favorisée par rapport aux interactions des partenaires séparés et les molécules d'eau du solvant, et un $\Delta S \ll 0$ suggère une flexibilité importante des partenaires, qui se structurent dans le complexe. Cette flexibilité pourrait expliquer que les essais de cristallisation de ce complexe ont échoué.

4.4. Discussion

Nous avons démontré que la protéine P₉₁₋₁₉₀ appartient à la famille de protéines intrinsèquement désordonnées (PIDs) avec un Rg supérieur à celui calculé pour un PID typique. La propension à former des structures secondaires (SSP) mesurée par RMN montre que le peptide P₉₁₋₁₉₀ n'est pas complètement désordonné mais possède des éléments de structure secondaire, de type hélice α et structure β , qui sont transitoirement peuplés (**Figure 3**). A noter que les régions des aa 120-127 et 136-158 peuplent des conformations hélicoïdales environ 20% et 15% du temps, respectivement. La région 120-127 se trouve au milieu de la région d'interaction avec STAT1, cartographiée expérimentalement (Ciancanelli, Volchkova et al. 2009). Ainsi la préexistence d'une faible population d'hélices pourrait faciliter la reconnaissance du site de fixation sur STAT1 et constituer un MoRE (Molecular Recognition Element). Cette hélice transitoire pourrait faciliter la formation d'un éventuel complexe de rencontre (encounter complex, (Sugase, Dyson et al. 2007)) et par la suite s'allonger des deux côtés par repliement induit. Généralement la conformation des éléments de structures secondaires fluctuants de ce type dans les PIDs isolés sont proches de leur conformation dans le complexe (Jensen, Houben et al. 2008; Lange, Lakomek et al. 2008; Leyrat, Jensen et al. 2011). Le phénomène de repliement induit par la fixation sur une protéine partenaire pour une protéine intrinsèquement désordonnée peut être utilisé afin d'assurer une bonne spécificité sans forcément avoir une très forte affinité. Pour P₉₁₋₁₉₀, ce type de comportement est en accord avec les paramètres thermodynamiques obtenus dans les expériences de calorimétrie et plus précisément avec la pénalité entropique qui accompagne l'interaction avec STAT1.

L'absence de détection du complexe sur la colonne de SEC a suggéré que l'interaction entre P₉₁₋₁₉₀ et STAT1₁₃₂₋₆₈₃ est relativement faible. Ce résultat, ensuite confirmé dans les expériences de titrage en RMN et la calorimétrie, n'est pas surprenant puisqu'une seule mutation, le remplacement du résidu Gly125 de P/V/W par un résidu Glu, est suffisante pour compromettre l'inhibition de la signalisation par l'IFN α/β dans plusieurs types cellulaires (Hagmaier, Stock et al. 2006). Il a été aussi montré que la partie N-terminale de P/V/W,

responsable de l'interaction avec STAT1, sert de point d'ancrage pour la protéine PLK (Polo like Kinase) et que cette dernière peut phosphoryler le résidu Ser131 du motif T₁₂₉SSP₁₃₂ (Ludlow, Lo et al. 2008). Le rôle de cette phosphorylation dans l'interaction avec STAT1 n'est pas clair, mais il a été montré que la formation des deux complexes V-STAT1 et V-PLK est indépendante (Ludlow, Lo et al. 2008). Dans le cas de l'interaction P/V/W-STAT1, une modification par phosphorylation pourrait déstabiliser d'avantage ce complexe. Néanmoins les expériences de RMN ont montré que la région des aa 110 à 140 de la P/V/W dans laquelle on observe une diminution des intensités des pics correspond bien à la région décrite dans la littérature, et il est intéressant de noter que la région des aa 136 à 158 est également perturbée par l'ajout de STAT lors des expériences de titrage en RMN. Etant donné sa propension à former des hélices d'environ 15%, il est possible que cette région se structure complètement lors de l'interaction.

La région comprenant les résidus 169 à 190 a également une forte propension à former des hélices α mais ne semble pas être impliquée dans la formation de complexe avec STAT1 ((Ciancanelli, Volchkova et al. 2009) et **Figure 4**). D'autre part la région de la P/V/W impliquée dans l'interaction avec STAT2 est moins bien connue et se trouve quelque part entre les résidus 100 à 300 de la P/V/W (Rodriguez, Cruz et al. 2004). Ainsi il est possible que les résidus 169 à 190 constituent un MoRE et soient impliqués dans l'interaction avec STAT2 ou d'autres partenaires.

Chapitre 5 : Le complexe N^0 -P

5.1. Introduction

Le génome des virus à ARN négatif non-segmenté est encapsidé par un homopolymère de nucléoprotéine, et ce complexe nucléoprotéique et non l'ARN nu constitue la matrice pour la polymérase (Lamb 2007). La réplication de ces virus nécessite donc la production continue de nucléoprotéine dans une forme soluble qui soit compétente pour l'encapsidation des ARN viraux génomiques ou antigénomiques néosynthétisés. Chez les *Rhabdoviridae* et les *Paramyxoviridae*, la nucléoprotéine soluble est produite sous la forme d'un complexe avec la phosphoprotéine, le complexe N⁰-P. Dans ce complexe, P joue deux rôles bien distincts, inhibant la polymérisation de N et empêchant la fixation d'ARN cellulaire. Lors de la réplication, la nucléoprotéine est transférée par un mécanisme inconnu de la phosphoprotéine vers l'ARN viral néo-synthétisé. Pour les virus de la rage et de la stomatite vésiculaire, le site d'interaction avec N⁰ a été localisé respectivement dans la région 4-40 et dans la région 11-30 (Mavrakis, Iseni et al. 2003; Mavrakis, Mehoulas et al. 2006; Chen, Ogino et al. 2007). Une stœchiométrie d'un dimère de P pour une nucléoprotéine avait été proposée sur la base d'expériences de spectrométrie de masse (Mavrakis, Iseni et al. 2003). Dans la famille des *Paramyxoviridae*, l'existence de ce complexe a été documentée pour divers virus incluant le virus Sendai (SeV) (Curran, Marq et al. 1995), le virus de la maladie de Newcastle (NDV) (Errington and Emmerson 1997), le virus parainfluenza type 2 (Nishio, Tsurudome et al. 1996), le virus Simian 5 (Precious, Young et al. 1995), le virus respiratoire syncytial (RSV) (Garcia, Garcia-Barreno et al. 1993) et le virus parainfluenza type 3 (Zhao and Banerjee 1995). Dans cette famille, la région d'interaction avec N⁰ a été également cartographiée à l'extrémité N-terminale de la P dans les 30 à 50 premiers résidus de la protéine (Curran, Marq et al. 1995).

Chez les *Rhabdoviridae* comme chez les *Paramyxoviridae*, le complexe N⁰-P a résisté pendant de nombreuses années à la détermination de sa structure à résolution atomique en raison notamment de la difficulté à co-exprimer et purifier des quantités suffisantes de complexe pour la cristallisation (Mavrakis, Iseni et al. 2003). Des essais de régénération du complexe N⁰-P par ajout de quantités importantes de P au complexe N-ARN de RaV et de VSV se sont soldés par des échecs.

Sur la base de l'analyse de la structure atomique des complexes circulaires N-ARN de VSV (Green, Zhang et al. 2006), RaV (Albertini, Wernimont et al. 2006) et RSV (Tawar, Duquerroy et al. 2009), nous avons proposé une stratégie pour reconstituer le complexe. Nous avons spéculé que la délétion d'un ou des deux sous-domaines d'échange déstabiliserait suffisamment l'homopolymère de N, de sorte qu'en présence de P ou d'un peptide dont la

séquence correspond à la séquence de la région N-terminale de P, l'homopolymère de N se dissocierait et l'ARN se séparerait de la protéine. Les premières tentatives d'expression de la nucléoprotéine de VSV tronquée des 21 résidus N-terminaux ont échoués car le complexe $N_{\Delta 21}$ -ARN était insoluble. Une étiquette « maltose binding protein » a alors été ajoutée à la $N_{\Delta 21}$ afin d'améliorer sa solubilité (**Figure 1**). Le complexe N-MBP-ARN a pu quant à lui être exprimé et purifié. L'ajout d'un peptide correspondant à la séquence des 60 premiers acides aminés de P à permis de régénérer, après clivage de l'étiquette MBP, le complexe soluble $N_{\Delta 21}^0$ -P₆₀ que nous avons par suite caractérisé structuralement. Cette régénération du complexe $N_{\Delta 21}^0$ -P₆₀ passe par la dissociation de l'homopolymère de N et le relarguage de l'ARN.

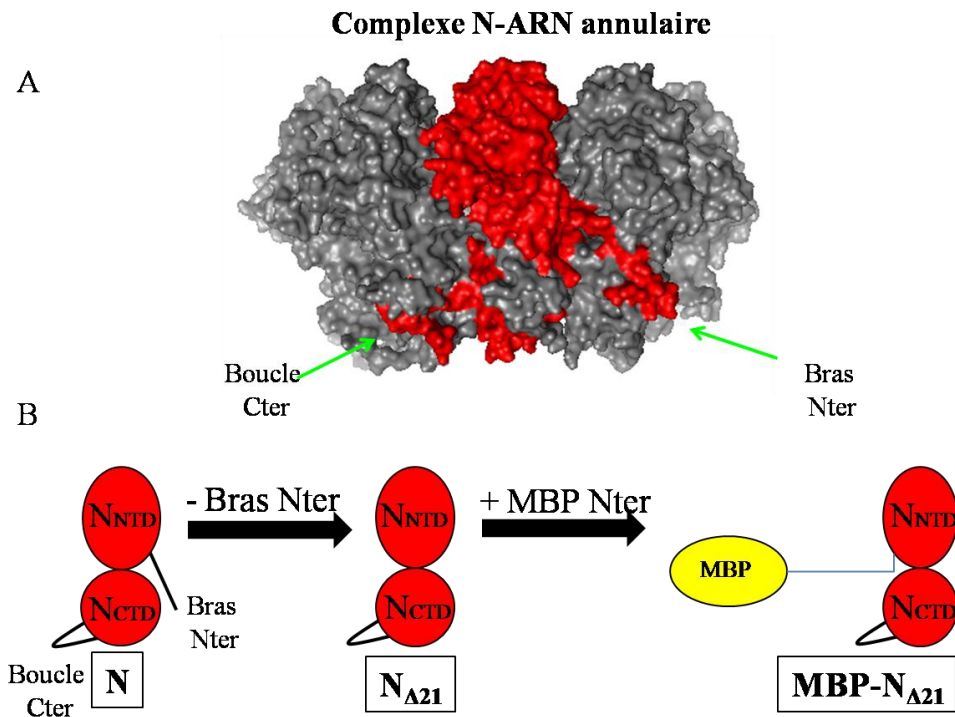


Figure 1 : Stratégie de reconstruction du complexe N^0 -P. **A)** La structure du complexe N-ARN annulaire de VSV (PDB code : 2GIC). Un protomère est représenté en rouge. **B)** Design de la construction utilisée pour la production du complexe $N_{\Delta 21}^0$ -P₆₀. D'abord le bras N-terminal a été enlevé et puis l'étiquette MBP (Maltose Binding Protein) a été ajoutée du côté N-terminal.

La suite logique du travail était d'essayer de reconstituer le complexe N^0 -P mais cette fois-ci en utilisant la phosphoprotéine entière et donc d'obtenir le complexe N^0 -P entière. Nous avons réalisé cette reconstitution en partant du complexe $N_{\Delta 21}^0$ -P₆₀ et ajoutant la P entière purifiée séparément. La P déplace le peptide et le complexe $N_{\Delta 21}^0$ -P entière a pu être purifié. Ensuite nous avons étudié ce complexe $N_{\Delta 21}^0$ -P entière par des techniques de diffusion aux petits angles des rayons X (SAXS) et des neutrons (SANS) et sa modélisation s'est transformée en projet collaboratif dans lequel les modèles atomiques du complexe $N_{\Delta 21}^0$ -P entière de VSV ont été générés par le Dr. Cédric Leyrat à partir de modèles tout atome de la P

(Leyrat, Schneider et al. 2012) et du complexe $N_{\Delta 21}^0$ -P₆₀ (Leyrat, Yabukarski et al. 2011) en utilisant des simulations de dynamique moléculaire. Les courbes de diffusion théoriques pour ces modèles ont été calculées par le Dr. Nicolas Martinez en utilisant les programmes CRY SOL et CRYSON (Svergun, Barberato et al. 1995; Svergun, Richard et al. 1998) et quelques scripts bash. La sélection d'ensemble, permettant de reproduire simultanément les données expérimentales de SAXS et SANS, a été possible grâce aux modifications apportées au programme GAJOE (Bernado, Mylonas et al. 2007) par le Dr. Pau Bérnado et son aide.

Nous avons également utilisé la construction P₁₋₁₇₇ et nous avons reconstitué le complexe $N_{\Delta 21}^0$ -P₁₋₁₇₇ que nous avons étudié par SAXS et résonance magnétique nucléaire (en collaboration avec le Dr. Malene Jensen).

Finalement, nous avons utilisé une stratégie similaire pour produire un mime du complexe N^0 -P du virus Nipah. Nous avons choisi d'étudier le complexe de ce virus pour deux raisons majeures:

(1) Des études ont montré que le peptide P₆₀ de RaV était capable d'inhiber la réaction de transcription/réplication dans un système d'expression d'un minigénome, mais aussi d'inhiber la réplication virale dans des cellules infectées (Castel, Chteoui et al. 2009). Cette inhibition pourrait résulter de la compétition entre le peptide exprimé en *trans* avec la P endogène pour la formation du complexe N^0 -P. Le complexe avec le peptide ne contenant pas le domaine C-termina de P nous pouvons spéculer qu'il ne serait pas adressé correctement au site d'encapsidation et que le complexe N^0 -P₆₀ ne permettrait donc pas une encapsidation efficace de l'ARN génomique néosynthétisé. Dans ce cas, la connaissance de la structure du complexe N^0 -P serait un atout majeur pour développer des inhibiteurs de cette interaction et donc de la réplication du virus. Toutefois, nous n'avons aucune preuve expérimentale que le peptide P₆₀ agit dans les cellules infectées en interférant avec la formation du complexe N^0 -P. Il n'est pas exclu qu'il interfère avec la polymérase. Etudier ce complexe appartenant à un membre de la famille des *Paramyxoviridae* semble très intéressant car la famille comprend de nombreux pathogènes pour l'homme, les animaux terrestres et aquatiques et notamment le virus de la rougeole (MeV), le virus des oreillons (MuV), le virus parainfluenza humain, le virus respiratoire syncytial (RSV), le virus rinderpest (RPV), le virus Sendai (SeV) ainsi que de nombreux virus émergents parmi lesquels les virus Nipah (NiV) et Hendra (HeV).

(2) A ce jour on connaît les structures des nucléoprotéines de plusieurs familles des *Monégavirales*. Pour les *Rhabdoviridae* le représentant des lyssavirus est RaV (Albertini, Wernimont et al. 2006) et des vésiculovirus – VSV (Green, Zhang et al. 2006). Pour les *Bornaviridae* le représentant est le virus de la maladie de Borna (BDV) (Rudolph, Kraus et al.

2003). Pour les *Paramyxoviridae*, sous-famille des *Pneumoviridae*, le représentant est le virus respiratoire syncytial (RSV) (Tawar, Duquerroy et al. 2009) mais on ne connaît aucune structure de nucléoprotéine de la sous-famille des *Paramyxovirinae*. L'obtention d'information atomique concernant la N pour plusieurs genres des *Paramyxovirinae* (*Henipavirus*, *Morbilivirus*, *Respirovirus*) est rendue difficile par la présence d'une longue région intrinsèquement désordonnée du côté C-terminal (le domaine N_{TAIL}) et par le fait que, contrairement aux *Rhabdoviridae* et les *Pneumoviridae*, ces virus ne forment pas de structures N-ARN en forme d'anneau mais plutôt des courtes nucléocapsides de taille variable (Bhella, Ralph et al. 2002; Ong, Yusoff et al. 2009).

Il faut noter que nous avons également reconstitué le complexe N⁰-P du virus de la rage (RaV) entre un variant de la N (N_{Δ23}) et des peptides de différentes longueurs (aa 1-68, 1-55 et 1-42) correspondant à la partie N-terminale de la P mais le complexe n'a pas cristallisé et ce travail n'es t pas présenté ici.

5.2. Article: Structure of the Vesicular Stomatitis Virus N⁰-P Complex

Authors: Leyrat C*, **Yabukarski F***, Tarbouriech N, Ribeiro E, Jensen M, Blackledge M,
Ruigrok R, Jamin M

* These authors contributed equally to the work

Publié en 2011 dans *Plos Pathogens*

Structure of the Vesicular Stomatitis Virus N⁰-P Complex

Cédric Leyrat¹*, Filip Yabukarski¹*, Nicolas Tarbouriech¹, Euripedes A. Ribeiro Jr.^{1,2}, Malene Ringkjøbing Jensen², Martin Blackledge², Rob W. H. Ruigrok¹, Marc Jamin^{1*}

1 UMI 3265 UJF-EMBL-CNRS Unit of Virus Host Cell Interactions, Grenoble, France, **2** UMR 5075 CEA-CNRS-UJF, Institut de Biologie Structurale, Grenoble, France

Abstract

Replication of non-segmented negative-strand RNA viruses requires the continuous supply of the nucleoprotein (N) in the form of a complex with the phosphoprotein (P). Here, we present the structural characterization of a soluble, heterodimeric complex between a variant of vesicular stomatitis virus N lacking its 21 N-terminal residues (N_{Δ21}) and a peptide of 60 amino acids (P₆₀) encompassing the molecular recognition element (MoRE) of P that binds RNA-free N (N⁰). The complex crystallized in a decameric circular form, which was solved at 3.0 Å resolution, reveals how the MoRE folds upon binding to N and competes with RNA binding and N polymerization. Small-angle X-ray scattering experiment and NMR spectroscopy on the soluble complex confirms the binding of the MoRE and indicates that its flanking regions remain flexible in the complex. The structure of this complex also suggests a mechanism for the initiation of viral RNA synthesis.

Citation: Leyrat C, Yabukarski F, Tarbouriech N, Ribeiro EA Jr, Jensen MR, et al. (2011) Structure of the Vesicular Stomatitis Virus N⁰-P Complex. *PLoS Pathog* 7(9): e1002248. doi:10.1371/journal.ppat.1002248

Editor: Félix A. Rey, Institut Pasteur, France

Received: May 31, 2011; **Accepted:** July 20, 2011; **Published:** September 22, 2011

Copyright: © 2011 Leyrat et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by grants from the French ANR (ANR-07-001-01 (ANRAGE)), the FINOVI foundation and Lyonbiopôle. CL and FY were supported by MENRT fellowships from the French government. EAR was supported by postdoctoral fellowships from both the ANR and the FINOVI programs. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jamin@embl.fr

† These authors contributed equally to the work.

Introduction

Negative-sense RNA viruses include numerous major human pathogens such as influenza virus, rabies virus, measles virus and respiratory syncytial virus. The (−)RNA genome of these viruses is condensed by a viral nucleoprotein (N) into a helical nucleocapsid [1], that associates with the polymerase complex and serves as the template for RNA replication and transcription [2]. Replication of the genome thus requires a continuous supply of N molecules to encapsidate both the (+)RNA intermediate copies and the newly synthesized (−)RNA genomes in single-stranded forms [3]. For non-segmented (−)RNA viruses of the *Rhabdoviridae* and *Paramyxoviridae*, N is assisted by the viral phosphoprotein (P). P binds to nascent RNA-free N, forming a N⁰-P complex (the superscript ⁰ denotes the absence of RNA) that prevents the polymerization of N and the non-specific encapsidation of host cell RNAs [4,5,6,7,8,9]. These processes are independent of each other [10] and, therefore, P has to fulfill two chaperone activities; blocking both RNA binding and self-assembly of N. P is a modular protein comprising a long N-terminal disordered region and two folded domains, a central oligomerization domain and a C-terminal nucleocapsid-binding domain, separated by a flexible linker [11,12]. The N⁰-binding region is localized in the N-terminal disordered region [7,8,9], and it has been demonstrated that in vesicular stomatitis virus (VSV), a prototypical rhabdovirus, this region of P contains transient helical elements and may thus constitute a short molecular recognition element (MoRE) that folds upon binding to its partner [13].

In rhabdovirus nucleocapsids, every N molecule binds nine nucleotides in a positively charged cavity at the interface between its N- (N_{NTD}) and C-terminal (N_{CTD}) domains [14,15]. The N-

RNA complex is stabilized by multiple salt bridges between the sugar-phosphate backbone of the RNA and basic residues of N, by contacts between neighboring N molecules involving hydrophobic side-to-side interactions, mainly between adjacent N_{CTD}, and by the exchange of N- and C-terminal sub-domains between adjacent N protomers (N_{NT}-arm, aa 1–21 and N_{CT}-loop, aa 340–375, respectively) [14,15]. Once formed, the N-RNA complex is stable and cannot be disassembled by full-length P [16]. However, on the basis of the N-RNA structure, we hypothesized that deletion of the N_{NT}-arm may sufficiently destabilize the N-RNA complex so that P or a peptide fragment of P containing the MoRE that binds N⁰, could displace the RNA molecule.

In this study, we report the reconstitution of complexes between a recombinant N of VSV lacking the 21 N-terminal residues, N_{Δ21}, and either full-length P dimer [11] or a peptide encompassing the N⁰-binding MoRE of P [13], named here P₆₀, that comprises the first 60 amino acids of P, a two-amino acid linker and a C-terminal His₆-tag. The characterization by absorbance spectroscopy and size-exclusion chromatography (SEC) combined with static light scattering (MALLS) demonstrates that both N_{Δ21}⁰-P₆₀ and N_{Δ21}⁰-P dimer complexes are free of RNA in solution, forming soluble heterodimers or heterotrimers, respectively. Therefore, P₆₀ fulfills both chaperone activities of full-length P. The heterodimer N_{Δ21}⁰-P₆₀ crystallized, but under the crystallization conditions, it assembled into a circular decamer of heterodimers very similar to the previously crystallized decameric N-RNA ring. The crystal structure of the decameric form of the N_{Δ21}⁰-P₆₀ complex reveals the molecular mechanisms by which the N⁰-binding MoRE of P attaches to N. NMR spectroscopy confirms that the MoRE of P binds to N in the N_{Δ21}⁰-P₆₀ complex in solution as in the crystal structure and shows that the regions of

Author Summary

The negative sense RNA genome of the rhabdoviruses is encapsidated by the nucleoprotein, and the replication of the genome requires a continuous supply of RNA-free, monomeric nucleoprotein (N⁰) to encapsidate the newly synthesized (+)RNA intermediate antigenomes and (−)RNA genomes. In this process, the viral phosphoprotein acts as a chaperone, forming a heterodimeric complex, named N⁰-P, which prevents nascent N molecules from self-assembling and from binding to cellular RNAs. We reconstructed the N⁰-P complex of the prototype rhabdovirus, vesicular stomatitis virus, and characterized its structure by crystal X-ray diffraction and solution experiments. Our results show how the N-terminal region of the phosphoprotein folds upon binding to the RNA-free nucleoprotein and how it prevents the non-specific encapsidation of host-cell RNA. This complex is soluble and heterodimeric, but by forcing it to polymerize into a crystal it associated into a circular decamer of heterodimers very similar to the previously crystallized decameric N-RNA ring. On the basis of our results, we propose a model that explains the role of the phosphoprotein in the encapsidation of newly synthesized RNA and in the initiation of RNA synthesis by the viral polymerase.

P flanking this MoRE remain flexible in the complex. Finally, these results suggest mechanisms for the encapsidation of newly synthesized RNA and for the initiation of RNA synthesis by the viral polymerase.

Results

Strategy for reconstituting the N_{A21}⁰-P and N_{A21}⁰-P₆₀ complexes

Production of a mutant of N deleted of its 21 first N-terminal residues (N_{A21}) in *Escherichia coli* led to the formation of inclusion bodies and of poorly soluble complexes, which could not be purified. In order to improve the solubility of the N_{A21} mutant, it was produced in *E. coli* in fusion with an N-terminal maltose binding protein (MBP) tag. The purified MBP-N_{A21} formed soluble, oligomeric N-RNA complexes, which eluted next to the exclusion volume of a Superdex S200 column (Figures S1A and S1B in Text S1). The presence of RNA was demonstrated by the absorbance ratio at 280 nm and 260 nm of 1.05 ($A_{280\text{ nm}}/A_{260\text{ nm}}$) (Figure S1C in Text S1). The MBP-N_{A21} monomer migrated as a single protein of about 100 kDa on a denaturing 4–20% gradient PAGE (Figure S1B in Text S1). Incubation of MBP-N_{A21} at 20°C overnight in the presence of P₆₀ resulted in the displacement of the bacterial RNA from N, the dissociation of the oligomeric N-RNA complexes and the formation of a new species that eluted at 14.1 mL (Figure S2A in Text S1). The analysis by SEC-MALLS indicated a weight-averaged molecular mass of 92 ± 2 kDa in agreement with the calculated molecular mass of the MBP-N_{A21}⁰-P₆₀ complex (calculated mass: 88,326 Da (MBP-N_{A21}) + 8,053 Da (P₆₀) = 96,379 Da). The co-elution of MBP-N_{A21} and P₆₀ was confirmed by denaturing 4–20% gradient PAGE (Figure S2B in Text S1). The complex contained much less RNA as shown by the absorbance spectrum ($A_{280\text{ nm}}/A_{260\text{ nm}} = 1.60$) (Figure S2C in Text S1). After cleavage of the MBP tag with the TEV protease, the resulting N_{A21}⁰-P₆₀ complex was purified by Ni²⁺ chelate affinity chromatography followed by SEC. The complex of N_{A21} with full-length P dimer [11] was then prepared by incubating the purified N_{A21}⁰-P₆₀ complex overnight with the intact P dimer.

Solution properties of the N_{A21}⁰-P and N_{A21}⁰-P₆₀ complexes

The molecular mass of the N_{A21}⁰-P dimer complex determined by SEC-MALLS was constant throughout the chromatographic peak indicating that the complex was monodisperse ($M_w/M_n = 1.00 \pm 0.01$), and the molecular mass of 104 ± 4 kDa was consistent with that of a heterotrimer composed of one N_{A21} and an intact P dimer (calculated mass: 45,377 Da (N_{A21}) + 2 × 30,976 Da = 107,329 Da) (Figure 1) in accordance with the dimeric state of P in solution [11] and with a previous determination for the rabies virus N⁰-P complex by naïve mass spectrometry [17]. The hydrodynamic radius (R_h) of 5.8 ± 0.1 nm is about 1.5 fold larger than that for a globular particle of the same molecular mass (calculated $R_s = 4.0$ nm) reflecting the elongated shape of the complex and the existence of a long N-terminal disordered region (aa 1–106) [11,12].

The N_{A21}⁰-P₆₀ complex contained no RNA (Figure 2A), and its R_s of 3.2 ± 0.1 nm and molecular mass of 53 ± 3 kDa indicated a globular 1:1 complex (Figure 2B), which agrees with the fact that P₆₀ does not contain the dimerization domain of P [12,18]. The radius of gyration (R_g) of the N_{A21}⁰-P₆₀ complex determined from SAXS data (2.7 ± 0.1 nm) (Figure S3 and Table S1 in Text S1) was similar to that of a single N protomer extracted from the N-RNA complex (2.8 nm) [15], but the calculated curve of the extracted protein poorly fitted the experimental curve of N_{A21}⁰-P₆₀, probably because of the presence of P₆₀ (Figures S4A and S4B in Text S1). *Ab initio* bead models reconstructed from SAXS data [19] (Figure 2C) could easily accommodate the structure of an isolated N protomer deleted of its N_{NT}-arm, except for the N_{CTD}-loop, which is likely to adopt a different conformation in solution (Figure 2D). Although the low resolution of the model precluded the precise localization of P₆₀, the absence of an empty groove at the interface between N_{NTD} and N_{CTD} suggests that P₆₀ could bind in this region. These results clearly show that P₆₀ fulfils both chaperone functions of P in maintaining N⁰ soluble and free of RNA, and because the size and flexibility of full-length P render the N_{A21}⁰-P complex unsuitable for X-ray crystallography and NMR studies, the N_{A21}⁰-P₆₀ complex was used for further structural characterization.

Crystal structure of a decameric N_{A21}⁰-P₆₀ complex

The N_{A21}⁰-P₆₀ complex crystallized at low pH as a decameric circular complex. The structure was solved at a resolution of 3.0 Å by molecular replacement using the structure of an N protomer derived from the N-RNA complex [15] (Table 1). The structure of N in the N_{A21}⁰-P₆₀ complex was almost identical to that of N in the N-RNA complex (rmsd = 0.96 Å) (Figure 3A and Figure S5 in Text S1) [15]. The complex contained no RNA, but instead, in each protomer, residues 6 to 35 of P₆₀ were visible in a groove formed by residues of the hinge region of N (aa 200–300) at the junction between N_{NTD} and N_{CTD} (Figure 3B). Previous observations showed that the isolated N⁰-binding region of P transiently populates α-helical conformers in the region 2–12 and 25–31 [13], and that residues 11 to 30 of P are essential for forming the N⁰-P complex [7]. Upon binding to N_{A21}, the second fluctuating helix is stabilized and extends from amino acids 17 to 31 (Figure 3B). The theoretical SAXS curve calculated from the crystal structure of one protomer of the N_{A21}⁰-P₆₀ complex fits adequately the experimental curve of the soluble complex (Figure S4A and S4C in Text S1), and the structure is perfectly accommodated within the *ab initio* bead model (Figure S4D in Text S1) showing that the structure of the N_{A21}⁰-P₆₀ complex in solution is the same as that in the crystal.

The structure of the N_{A21}⁰-P₆₀ complex shows how the N⁰-binding MoRE of P prevents both the interaction with RNA and

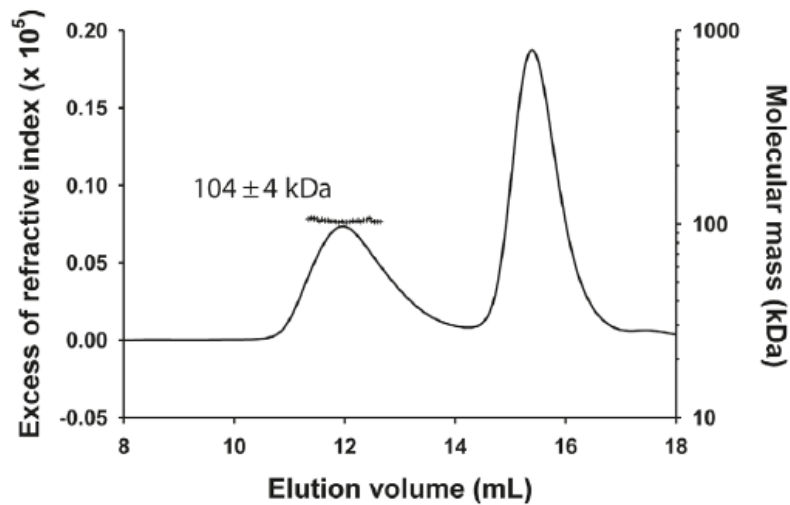


Figure 1. Heterotrimeric VSV N_{A21}⁰-P dimer complex in solution. The complex formed between N_{A21} and full-length P dimer in solution was analyzed by SEC-MALLS. The N_{A21}⁰-P complex elutes at 12.0 mL and the remaining N_{A21}⁰-P₆₀ complex elutes at 15.6 mL (line). The molecular mass of 104±4 kDa (crosses) indicates a 1:2 complex between one RNA-free N_{A21} and two P molecules in accordance with the previous observation that P forms exclusively dimers in solution and with the N⁰-P₂ complex determined for rabies virus. doi:10.1371/journal.ppat.1002248.g001

the self-assembly of soluble RNA-free N. The binding site of P is different from that of RNA but both sites do overlap (Figures 4A, 4B and Figure S5 in Text S1), and the C-terminal turn of the α -helix of P₆₀ (aa 27–31) together with the following residues (aa 32–35) block the RNA binding cavity (Figure 4A) and inhibit RNA binding. Concomitantly, the other extremity of the MoRE (aa 6–13) docks in a shallow groove on the backside of the N protomer which, in the multimeric N-RNA complex, is occupied by the N_{NT}-arm of the adjacent N_{i-1} protomer and whose bottom is made up of the N_{CT}-loop of the N_{i+1} protomer (Figure 4C and 4D). With w.t. N, the N-terminal part of the MoRE of P will compete with the N_{NT}-arm of a neighboring N molecule and therefore interferes with the polymerization of N in the absence of RNA.

The MoRE of VSV P (aa 6–35) binds to the central hinge region of N mainly through hydrophobic interactions. The amphipathic α -helix of P, together with residues 14 to 16, inserts into a hydrophobic groove of N. Tyr14 perfectly fits into a small hydrophobic pocket (Figure 5A). In addition, the complex is stabilized by three intermolecular salt bridges (Figure 5A). This region of N (aa 200–300), which also plays a central role in binding RNA is highly conserved among VSV serotypes as well as in rabies virus (RAV) [20,21] (Figure S6 in Text S1). The hinge region of VSV N exhibits 30% identity in amino acid sequence with that of RAV N, as compared with 21% and 13% for the N- and C-terminal lobes, respectively. Figure 5B shows that several hydrophobic residues lining the binding groove of P are conserved between VSV and RAV N, as well as Arg³¹², suggesting that a similar complex forms in RAV (Figure S6 in Text S1). In the N-RNA complex, the RNA molecule interacts with N through electrostatic interactions between phosphate groups of the RNA backbone and basic residues of the protein, while the bases of three nucleotides (nt. 5, 7 and 8) are docked onto an hydrophobic surface of the RNA binding groove [15], which is also part of the MoRE binding site (Figure 5B). However, none of the basic residues of VSV N directly contacting the RNA backbone in the N-RNA complex is involved in the interaction with the MoRE of

P. A similar mode of RNA binding was observed in the RAV N-RNA complex [14], but only two arginines out of the six residues involved in direct interactions with phosphate groups in the VSV complex are conserved in the RAV complex [20].

The RNA binding groove of N is rich in basic residues forming a highly positive surface area (Figure 5C), while the backside of N_{CTD} harbors a negative surface potential (Figure 5C). The MoRE of P (aa 6–35) has a bipolar distribution of charges with a positive pole at its N-terminus and a negative pole at its C-terminus. In the N_{A21}⁰-P₆₀ complex, the negative pole of P localizes in the RNA binding groove, while the positive pole docks on the backside of N_{CTD}, modifying the distribution of electrostatic potentials on these surfaces of N and suggesting that electrostatics could play a role in orientating P before binding (Figures 5C and 5D). The crystallization is at pH 4.6 and it is likely that protonation of acidic groups reduces repulsion forces that keep the N_{A21}⁰-P₆₀ complex in its isolated form at pH 7.0.

NMR spectroscopy

In the crystal structure of the N_{A21}⁰-P₆₀ complex residues 6 to 13 and 32 to 35 of P₆₀ exhibit conformational heterogeneity (Figure S7 in Text S1), while residues 1 to 5 and 36 to 68 are not visible. To further characterize the conformational dynamics of these parts of P₆₀ in the soluble N_{A21}⁰-P₆₀ complex, we used nuclear magnetic resonance (NMR) spectroscopy. Initially, spectra of ¹⁵N, ¹³C, ²H-labeled P₆₀ in complex with unlabeled N_{A21} were recorded. In a complex of this size (53 kDa) NMR signals are significantly broadened, precluding their detection, but in the HSQC spectrum of the N_{A21}⁰-P₆₀ complex, resonances corresponding to the last 28 amino acids of P₆₀ (aa 41–60+Hinker+His₆ tag) are clearly visible (Figure 6A), suggesting that this tail remains free and flexible in the complex. Comparison with the free peptide showed that most resonances superimpose. Small chemical shift differences were observed for residues Q⁴¹ to G⁴⁴, probably due to the proximity of the bulk complex, and for two aromatic residues (Y⁵³, F⁵⁴) suggesting weak interactions of these residues with N. The amide backbone ¹⁵N transverse relaxation rate constant (*R*₂) is

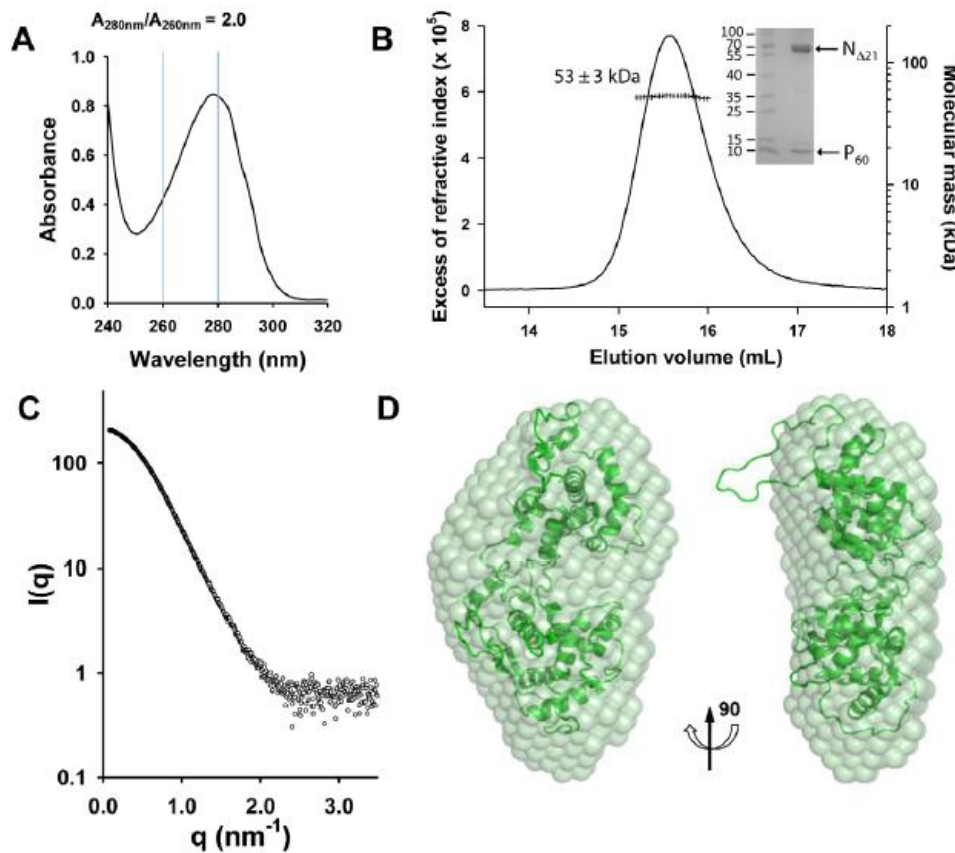


Figure 2. Heterodimeric VSV N_{Δ21}⁰-P₆₀ complex in solution. (A) The absence of RNA in the complex is revealed by the UV absorbance spectrum, which exhibits an $A_{280\text{ nm}}/A_{260\text{ nm}}$ ratio of 2.0. (B) Analysis of the N_{Δ21}⁰-P₆₀ complex by SEC-MALLS. The complex elutes as a single peak at 15.6 mL (line), and the presence of the two proteins in the complex is demonstrated by SDS-PAGE analysis of the peak fraction using Coomassie blue staining (inset). The molecular mass of 53 ± 2 kDa (crosses) indicates a 1:1 complex between N_{Δ21} and P₆₀ (calculated molecular mass = 45,377 Da (N_{Δ21}) + 8,053 Da (P₆₀) = 53,430 Da). (C) Experimental SAXS data (open circles) up to 3.5 nm^{-1} . The SAXS curve recorded at ESRF beamline ID 14-3 shows the scattering intensity $I(q)$ as a function of the scattering vector, $q = (4\pi \sin \theta / \lambda)$. (D) Average *ab initio* bead model of the N_{Δ21}⁰-P₆₀ complex. The N protomer extracted from the circular N-RNA complex (2GC chain E) fits to the SAXS-derived model, except for the N_{CT}-loop. doi:10.1371/journal.ppat.1002248.g002

sensitive to rapid fluctuations at the pico- to nanosecond time scale, as well as to chemical shift exchange on the micro- to millisecond time scale. Transverse relaxation of the visible resonances increased from the C-terminus to the region containing the bound helix, indicating a corresponding increase in rigidity of the backbone (Figure 6B). Very weak additional peaks up-field shifted in the amide proton dimension were detected in a ¹⁵N-¹H TROSY spectrum further supporting folding of the helical element upon binding (data not shown).

In a second experiment, addition of sub-stoichiometric amounts of unlabeled N_{Δ21} attached to MBP (to maintain solubility) to ¹⁵N-labeled P₆₀ resulted in an overall reduction in the intensity of the peaks in the HSQC spectrum, in proportion of the amount of added MBP-N_{Δ21} indicating that a fraction of P₆₀ formed a complex with MBP-N_{Δ21}. Spin relaxation measurements of equilibrium mixtures of free and bound peptide refine our understanding of the dynamics of the system. Systematically larger R_2 values observed for residues 1 to 17 reveal additional contributions from chemical shift exchange (R_{ex}) that are not present in the free form of the peptide and that

increase significantly upon increasing the molar ratio of MBP-N_{Δ21} to P₆₀ (Figure 5C). This indicates that the N-terminal part of P₆₀ experiences conformational exchange when in complex with N_{Δ21} with an interconversion rate on the micro to millisecond timescale. In the crystal structure, residues 6 to 13 bind in place of the N_{NT}-arm of the protomer N_{i-1} and pack onto the N_{CT}-loop of protomer N_{i+1} (Figures 4C and 4D), but in the isolated form of the complex, the N_{CT}-loop binding surface for these residues is missing. These results confirmed that in the solution, like in the crystal, residues 17 to 35 of P form a stable complex with N, while the flanking N- and C-terminal regions remain dynamic. The flanking C-terminal part (aa 40–60) behaves as a flexible tail and shows little evidence of interaction with N. The flanking N-terminal region (aa 1–16) interacts with N but undergoes conformational exchange. These flexible regions seem dispensable for the chaperone activities of P since a shorter peptide encompassing residues 7 to 40 of P is also capable of displacing bacterial RNA from N_{Δ21} and of forming a soluble heterodimeric 1:1 complex with this protein (data not shown).

Table 1. Data collection and refinement statistics (molecular replacement).

Crystal 1	
Data collection	
Space group	P2 ₁ 2 ₁ 2
Cell dimensions	
a, b, c (Å)	74.56, 171.97, 239.86
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å) ^a	59.97–3.03 (3.20–3.03)
R _{sym} ^a	11.4%(28.1%)
I/σI ^b	9(3.5)
Completeness (%) ^a	89.9 (80.7)
Redundancy	3.1 (2.5)
Refinement	
Resolution (Å)	59.97–3.03 (3.11–3.03)
No. reflections	168,372
R _{work} /R _{free}	24.4%/27.7%
No. atoms	
Protein	17,105
Water	455
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.893
Ramachandran favored	92.6%
Ramachandran allowed	98.4%

^aValues in parentheses are for highest-resolution shell.
doi:10.1371/journal.ppat.1002248.t001

Discussion

In the current study, we have determined the structure of the N⁰-binding MoRE of P bound to N and demonstrated that the regions of P flanking the MoRE conserve some flexibility in the N⁰-P complex. Because the N⁰-P complex is required for the replication of the virus [5,22], inhibition of its formation might represent an interesting target for blocking viral replication and could explain the recent observations that a homologous peptide (P₆₀) from rabies virus P inhibits viral replication [23].

The N_{A21}⁰-P₆₀ complex as a model of the N⁰-P complex

Our results demonstrate that the reconstituted N_{A21}⁰-P₆₀ complex is a suitable model for the viral N⁰-P complex in agreement with previous studies. Both the crystal structure and the NMR spectroscopy experiments show that the MoRE of P, which adopts a stable conformation upon binding to N, includes residues 6 to 35 and corresponds closely to the fragment that was previously identified as essential and sufficient for maintaining N in a soluble form (aa 11–30) [7].

Previous studies revealed that VSV P is a dimeric and modular protein in which the N-terminal part (aa 1–106) is globally disordered [11,12,13,18,24,25]. The dimerization domain of P is localized in the central region of the protein (aa 107–177) [18] and is therefore not present in P₆₀, which is monomeric [13]. The stoichiometry of the N_{A21}⁰-P₆₀ complex (1:1) shows that one MoRE of P is capable of binding one N molecule. The stoichiometry of the complex formed between N_{A21} and full-length P (1:2) in the concentration range used here suggests that a

single N_{A21} is bound to P dimer. The remaining part of the P dimer is tethered to N_{A21} through a flexible linker, in agreement with the large hydrodynamic radius measured here for the N_{A21}⁰-P dimer complex. In isolation, the N-terminal region of P contains two transient α helices (aa 2–12 and 25–38) [13]. In the crystal structure, the second helix is stabilized and extends from residue 17 to residue 35, whereas the first helix is not present. Residues 1 to 5 are not visible and residues 6 to 13 adopt different conformations in the different protomers of the circular complex. Because NMR spectroscopy reveals that this N-terminal region of P (aa 1–17) is in chemical shift exchange, it is possible that in solution it adopts different conformations bound in different orientations on the surface of N, but that only those docked into the backside groove of N allow the packing of the N_{A21}⁰-P₆₀ complex into crystals and were thus selected during the crystallization process.

The truncated form of N (N_{A21}) conserves the ability of self-association in the presence of RNA and, like w.t. N, forms oligomeric N-RNA complexes when expressed in bacteria. As assumed from the structure of the oligomeric N-RNA complex [15], the N_{NT}-arm stabilizes the multimeric N-RNA complexes by linking together adjacent N protomers. The N-RNA complex formed with w.t. N could not be dissociated by the addition of full-length P or of a fragment of P encompassing the N⁰-binding region. However, the deletion of the N-terminal sub-domain destabilized the complex and allowed P₆₀ to displace the RNA molecule and disassemble the multimeric N-RNA complex. In a previous study, the co-expression of P with a similar variant of VSV N lacking the first 22 amino acids (N_{A22}) led to the production of complexes of different sizes containing N_{A22} and P but not of N-RNA complexes, suggesting a role for the N-terminal region of N in the encapsidation of the RNA [10]. Assuming an equilibrium between the N⁰-P complex and the multimeric N-RNA complex, with w.t. N, the stabilization brought by the N_{NT}-arm to the multimeric assembly would displace the equilibrium towards the formation of the N-RNA complex. In the absence of the N_{NT}-arm, the truncated N molecules assemble onto cellular RNAs as seen in our expression system, but in the presence of co-expressed P, like upon addition of P₆₀ to our purified MBP-N_{A21}-RNA complexes, the equilibrium is displaced towards the formation of the N⁰-P complexes. The absence of N-RNA complex in cells co-expressing N_{A22} and P may not result from a default of encapsidation but rather from the displacement of the equilibrium towards N⁰-P.

Unexpectedly, the N_{A21}⁰-P₆₀ complex failed to crystallize as a heterodimer but crystallized into circular decamers of heterodimers. With the exception of the missing N_{NT}-arm, the structure of N_{A21} in the N_{A21}⁰-P₆₀ complex is very similar to that of N in the decameric N-RNA complex, with less than 1 Å r.m.s.d. between the two structures. Different explanations why multimerization occurs under crystallization conditions can be proposed. Firstly, the N_{A21}⁰-P₆₀ complex crystallized at pH 4.6 like VSV circular N-RNA complexes [15], while solution SEC-MALLS and SAXS experiments were performed at pH 7.5 and NMR experiments at pH 6.0. A modification of the electrostatic surface potential (Figures 5C and 5D) could affect the equilibrium between heterodimeric and multimeric N_{A21}⁰-P₆₀ complex. No evidence of multimerization was, however, found in solution at pH 4.6 in the concentration range used for SEC-MALLS and SAXS experiments. Secondly, the ring-like structure of ten protomers appears as a favored organization of VSV N, likely reflecting on some geometrical and/or surface properties of the protein. VSV N forms ring-like structures mostly containing ten N subunits in the presence of non-specific RNA when expressed in a recombinant

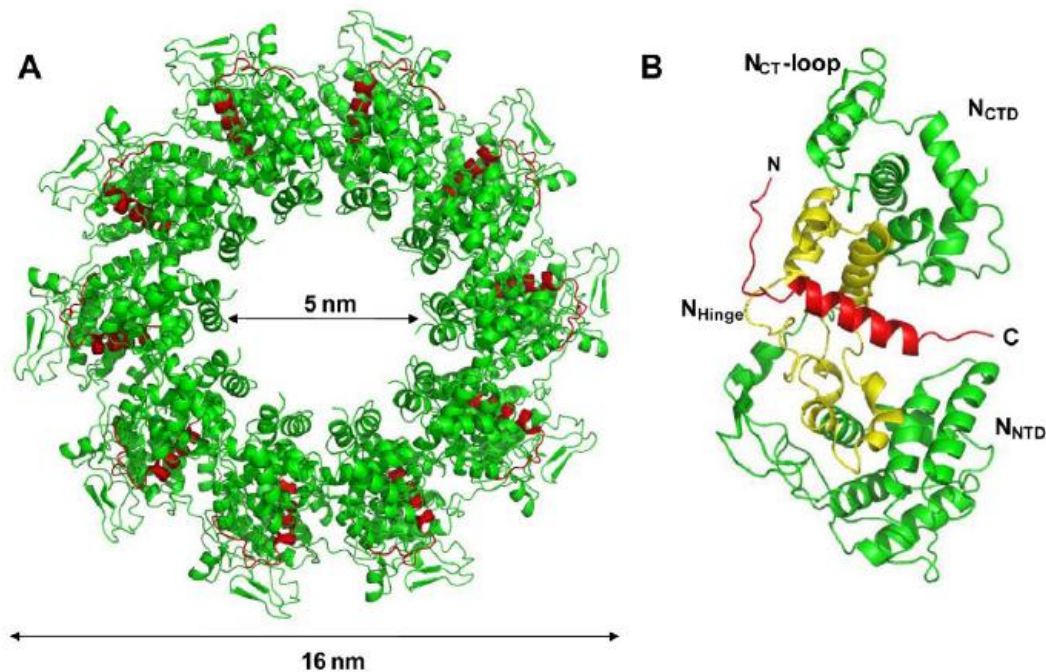


Figure 3. Crystal structure of a decameric form of the N_{A21}⁰-P₆₀ complex. (A) Overall structure of the decamer of N_{A21}⁰-P₆₀ complex. N_{A21} is shown in green and P₆₀ in red. (B) Ribbon representation of one protomer of N_{A21}⁰-P₆₀. The central hinge region of N (aa 200–300) is shown in yellow and P₆₀ (aa 6–35) is shown in red.
doi:10.1371/journal.ppat.1002248.g003

system [15,26]. The RNA can be removed from these ring-like structures without disrupting the multimeric assembly [27]. A single amino acid variant of VSV N that is no longer capable of binding RNA also crystallized into a decameric assembly of empty N molecules [10]. These circular N-RNA complexes are artifacts of the crystallization process because the actual nucleocapsid is very long and cannot form rings. However, in the virion, the nucleocapsid adopts a bullet-shaped structure composed of a trunk in which the nucleocapsid regularly spirals into superposed turns of 37.5 subunits of N and of a tip which is formed of seven turns containing varying numbers of subunits [1]. The upper turn of the bullet tip, which may represent the nucleation centre from which the particle assembles, resembles a decameric ring, suggesting that the assembly in ten members ring or spiral corresponds to an optimal side-by-side orientation between adjacent N subunits. The RNA-free N_{A21}⁰-P₆₀ complex is capable of assembling into circular multimers, and it seems likely that an increase of the concentration of the N_{A21}⁰-P₆₀ complex under the crystallization conditions together with a change in pH shift the equilibrium towards the multimers.

This raises the question of the effect of crystallization on the structure of the N_{A21}⁰-P₆₀ complex. The SAXS curve calculated for N_{A21}⁰-P₆₀ protomer extracted from the crystal structure perfectly reproduced the experimental curve of the soluble complex, while NMR spectroscopy clearly shows that the same segment of P (aa 17–35) is involved in a stable complex with N in solution and in the crystal, arguing that crystallization has no major effect on the structure of the more rigid part of the complex. In solution, the N-terminal part of the MoRE of P (aa 1–16) appears to be in conformational exchange and could thus exist in different conformers including those observed in the crystal in

which residues 6 to 16 are docked onto the backside groove of N. Crystallization of the N_{A21}⁰-P₆₀ complex may thus select the more compact conformers and therefore not reproduce the conformational diversity of this region that is found in solution. In addition, the high conservation rate of residues of N forming the binding surface for the MoRE of P, both within VSV serotypes and between the evolutionarily more distant VSV and RAV, supports the localization of the interface between the two proteins and hints at the formation of a similar complex in RAV.

Mechanisms of chaperone activities

The characterizations in solution indicate that the N_{A21}⁰-P₆₀ complex is RNA-free heterodimer and that, therefore, P₆₀ or a shorter fragment of P (aa 7–40) fulfill both chaperone activities of P. The crystal structure of the N_{A21}⁰-P₆₀ complex clearly shows how the N⁰-binding region of P inhibits RNA binding by filling the RNA-binding groove of N. In solution, this part of P also forms a stable complex with N as seen by NMR spectroscopy. The structure also suggests how P prevents the self-assembly of N in the absence of RNA. In the bound form observed in the crystal structure, the N-terminal extremity of the MoRE (aa 6–16) directly competes with the N_{NT} arm of a neighboring N molecule. Assuming this region of P fluctuates between bound and free forms in the soluble N⁰-P complex, the free form may act as an entropic bristle, thereby also preventing the oligomerization of N. With full-length P dimer, the flexibility and the bulkiness of the remainder of the protein may also contribute to this effect by masking the binding interfaces for RNAs or other N molecules. In addition, the MoRE of P exhibits a bipolar distribution of charges, with a positive pole at its N-terminal extremity and a negative pole at its C-terminal extremity. Binding of the MoRE of P modifies the

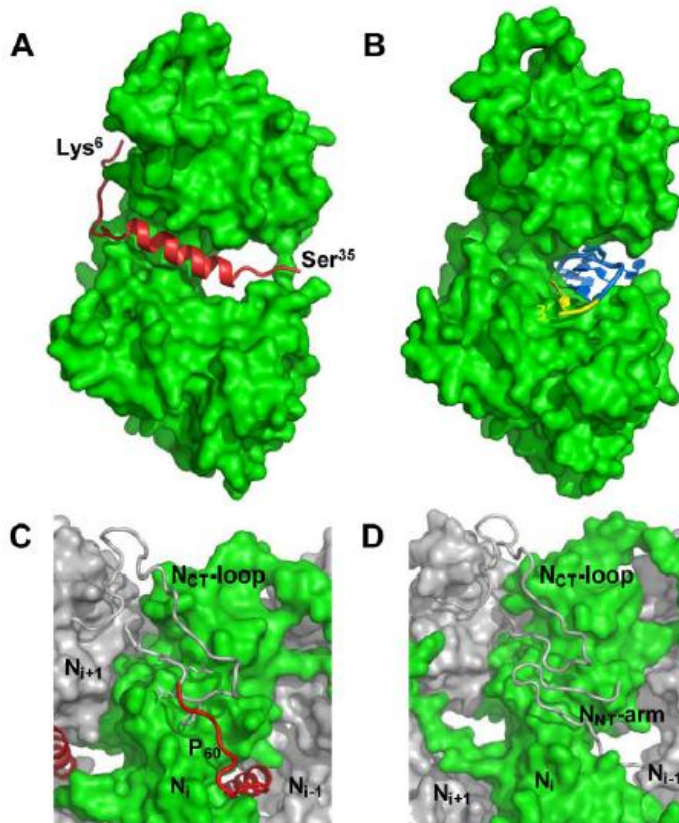


Figure 4. P binding hinders RNA binding and self-assembly of soluble N. (A) Representations of one protomer from the $N_{\Delta 21}^0$ - P_{60} complex. The N protomer is shown as space filling model in green and P_{60} is shown as a cartoon representation in red. The visible N- and C-terminal residues of P_{60} are labeled. (B) Representations of one protomer from the w.t. N-RNA complex. The 3' terminal nucleotide of the RNA molecule is shown in yellow. These representations show that P_{60} fills the RNA-binding cavity on the side of N that accommodates the 3'-end of the RNA molecule. (C, D) Close-up of the interactions between exchangeable sub-domains in the circular $N_{\Delta 21}^0$ - P_{60} and N-RNA complexes. In the w.t. N-RNA complex (D), the N_{NT} -arm of protomer N_{i-1} contacts the N_{CT} -loop of protomer N_{i+1} , while both sub-domains are docked on the back-side of protomer N_i (in green). In the $N_{\Delta 21}^0$ - P_{60} complex (C), the N-terminal extremity of P_{60} docks on protomer N_i (in green) at the position of the N_{NT} -arm of protomer N_{i-1} and contacts the N_{CT} -loop of protomer N_{i+1} . These representations suggest that P_{60} interferes with the assembly of N in the absence of RNA. doi:10.1371/journal.ppat.1002248.g004

electrostatic surface potential of N, notably reducing the positive surface potential on one side of the molecule, and may thereby affect the side-by-side interaction with another N molecule.

From the results presented here, we propose a hypothesis for the encapsidation of a newly synthesized RNA molecule during viral genome replication. By forming a complex with P, a nascent N molecule is prevented from binding to host-cell RNA and is preserved in a soluble form. During RNA replication, N is transferred to a growing RNA molecule and P is released. Little is known about the mechanism of this reaction or about the role played by the polymerase complex in this process. Our results show that the N_{NT} -arm stabilizes the multimeric N-RNA complex and therefore suggest that the multimeric N-RNA complex is more stable than the N^0 -P complex. The transfer of N from the N^0 -P complex to the growing N-RNA complex could simply be driven by a higher stability of the N-RNA complex. Upon transfer of N onto the RNA and release of P, the backside groove of N of the last added N molecule is liberated and becomes available for accepting the N_{NT} -arm of the next incoming N molecule (Figure 7A). By blocking the backside groove of N, the N-terminal part of P ensures that N molecules do not

assemble into empty N polymers but assemble only onto an RNA molecule. It is also noteworthy that, in VSV, a high affinity binding site for the L protein was localized in the second half of the N-terminal disordered region of P [7,28]. The dynamic nature of the N-terminal region of P and the proximity of the two binding sites may have significance for the mechanism of action of the transcription/replication machinery. The binding of N^0 to P may prevent the simultaneous binding of L, or conversely, the simultaneous binding of N^0 and L may modify the activity of the polymerase.

In addition to its role in RNA encapsidation, the binding of the N-terminal region of P to N may also provide a mechanism for the initiation of (+)RNA synthesis at the genome 3' end and of (-)RNA synthesis at the antigenome 3' end [29,30]. Encapsidated RNA genome and antigenome are completely covered with the nucleoprotein and are not accessible to the RNA polymerase. However, the first N molecule at the 3' extremity of nucleocapsids exposes its binding site for the N-terminal MoRE of P (Figure 7B). By binding to this surface, P may destabilize the N-RNA complex sufficiently to displace several nucleotides from the first N protomer and allow the polymerase access to the RNA.

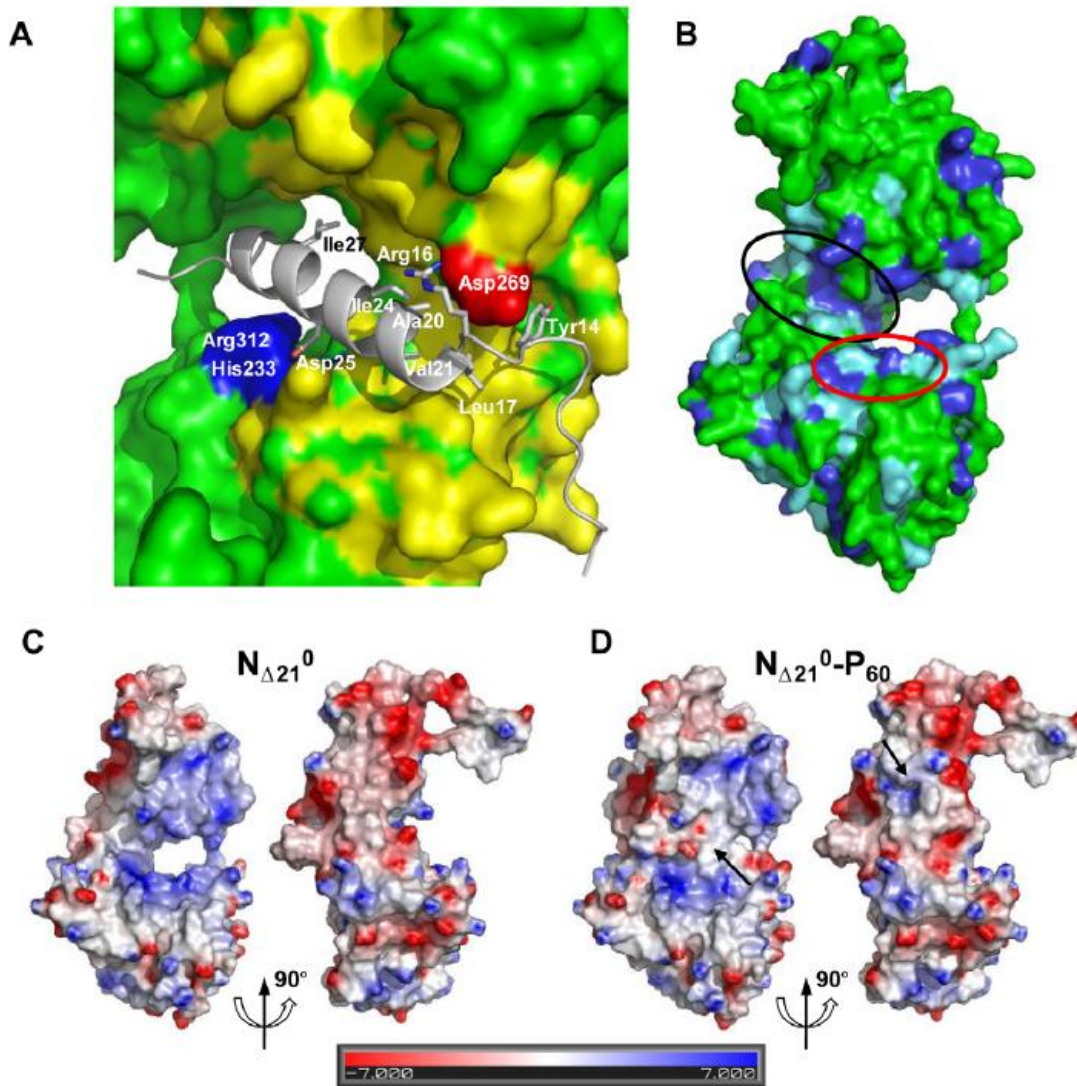


Figure 5. Surface properties and amino acid conservation in the P binding site. (A) Close up of the interface between RNA-free $N_{\Delta 21}$ and P_{60} showing the hydrophobic contacts and salt bridges. Residues 17 to 31 of P_{60} fold into an amphipathic α -helix that lies in a hydrophobic cavity formed by residues of the hinge region of N and is stabilized by hydrophobic contacts involving residues of P_{60} spaced i+3 or i+4 (Leu¹⁷, Val²¹, Ile²⁴ and Ile²⁷). Tyr¹⁴ docks into a small cavity lined with hydrophobic residues. Hydrophobic side chains in N are colored in yellow and hydrophobic side chains of P_{60} are labeled. The complex is also stabilized by salt bridges between Asp²⁵ of P_{60} and Arg³¹² and His²³³ of N (in blue) and between Arg¹⁶ of P_{60} and Asp²⁶⁹ of N (in red). (B) Amino acid sequence conservation between VSV and RAV N. Identical residues are shown in dark blue and similar residues are shown in light blue. The surface area circled in black shows the binding groove of P and the surface area circle in red shows the hydrophobic site common to both P and the bases at the 3' end of the RNA. (C, D) Electrostatic surface potential of the $N_{\Delta 21}^0$ protein (C) compared with that of the $N_{\Delta 21}^0$ - P_{60} complex (D). Both panels show in the same orientations the two sides of the $N_{\Delta 21}^0$ protein involved in binding the MoRE of P. The arrows indicate regions in which the electrostatic surface potential of N is modified by the presence of the peptide. The surface potentials were calculated with the Delphi program and are color-coded on the surface from red (negatively charged residues, -7 kcal/mol) to blue (positively charged residues, $+7$ kcal/mol).
doi:10.1371/journal.ppat.1002248.g005

Materials and Methods

Reconstitution of the $N_{\Delta 21}^0$ -P and $N_{\Delta 21}^0$ - P_{60} complexes

The cDNAs encoding vesicular stomatitis virus nucleoprotein or a fragment of this protein deleted of the 21 N-terminal residues were amplified by PCR and introduced into the pET-M40

plasmid (EMBL) using NcoI and XhoI restriction sites. The resulting constructs code for chimeric proteins that comprise an N-terminal maltose binding protein tag (MBP) and a tobacco etch virus (TEV) cleavage site. The cDNA encoding the 60 first N-terminal amino acids of VSV P (P_{60}) was amplified by PCR and cloned into the pET28a plasmid containing a C-terminal His₆-tag

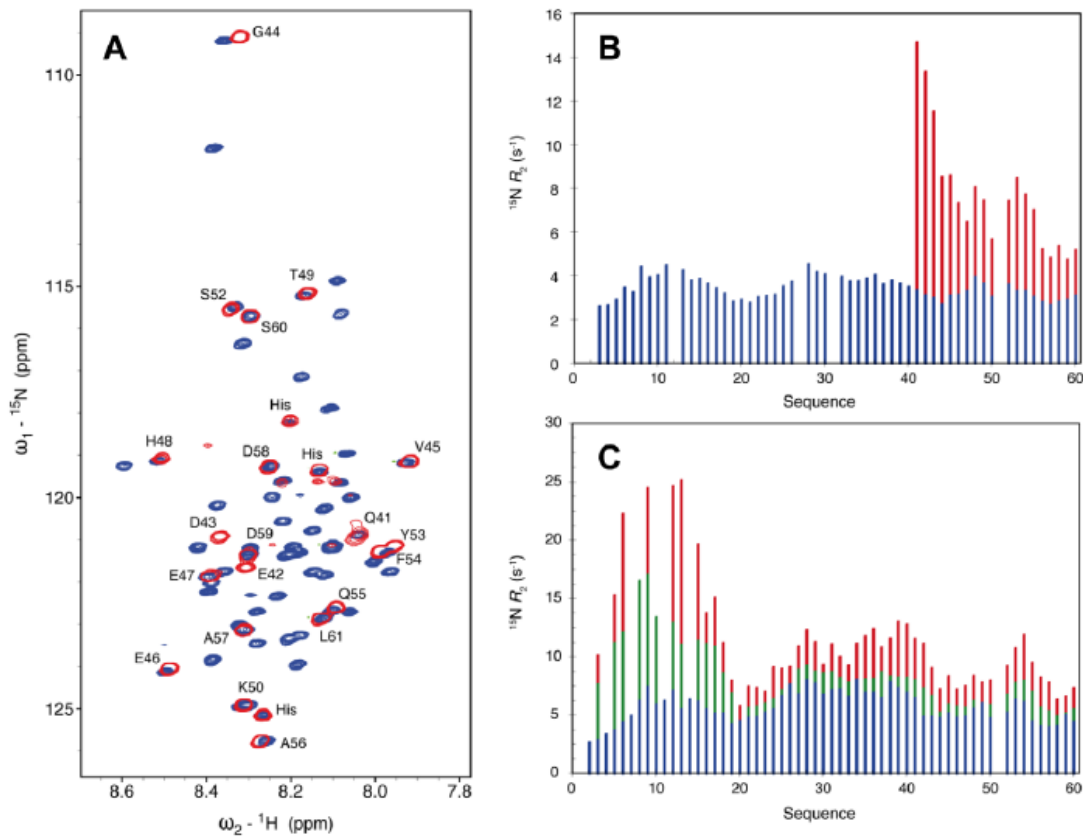


Figure 6. The N-terminal and C-terminal region of P₆₀ flanking the MoRE exhibits conformational flexibility in the soluble complex. (A) Comparison of the 2D ¹H-¹⁵N HSQC NMR spectra of free ¹⁵N, ¹³C, ²H-labeled P₆₀ (blue) and in complex with N_{A21} (red). Both spectra were recorded at 14.1 T and 25°C in 20 mM Tris-HCl, 150 mM NaCl, 50 mM Glu, 50 mM Arg with 10% D₂O adjusted to pH 6.0. The labels indicate the assignment of the resonances of the complex. (B) ¹⁵N R₂ spin relaxation rates measured under the same conditions as (A) for the free ¹⁵N-labeled P₆₀ (blue) and the N_{A21}⁰-P₆₀ complex (red). (C) ¹⁵N R₂ spin relaxation rates measured at 14.1 T and 10°C of the free ¹⁵N-labeled P₆₀ (blue), a mixture of 0.27 mM ¹⁵N-labeled P₆₀ and 0.09 mM unlabeled MBP-N_{A21} (green) and a mixture of 0.24 mM ¹⁵N-labeled P₆₀ and 0.17 mM unlabeled MBP-N_{A21} (red).

doi:10.1371/journal.ppat.1002248.g006

and a linker of two amino-acids (EL) using NcoI and XhoI restriction sites. All constructions were checked by DNA sequencing.

The plasmids were transformed into *Escherichia coli* Rosetta (DE3) cells and the expression of the recombinant proteins was induced with 1 mM isopropyl-1-thio-β-d-galactopyranoside (IPTG) for 18 h at 16°C. Cells were harvested by centrifugation, suspended in buffer A (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM DTT) containing protease inhibitors (Complete EDTA-Free, Roche Diagnostics) and disrupted by sonication. The extract was centrifuged at 20,000 g during 30 min at 4°C and the supernatant was filtered (0.45 μm). The MBP-fusion proteins were purified by affinity chromatography on amylose resin (New England Biolabs) followed by size exclusion chromatography (SEC) on a Superdex S200 column (GE Healthcare) equilibrated in buffer A. P₆₀ was purified by affinity chromatography on a Ni²⁺ resin column (Quiagen) followed by SEC on a Superdex S75 column (GE Healthcare) equilibrated in buffer A supplemented with 50 mM Glu and 50 mM Arg. Samples for NMR spectroscopy were produced in M9 minimal medium containing MEM

vitamins (Gibco). For producing ¹⁵N-labeled P₆₀, the medium was supplemented with 1.0 g.L⁻¹ of ¹⁵NH₄Cl and 2.0 g.L⁻¹ of unlabeled glucose, while for producing ¹⁵N, ¹³C, ²H-labeled P₆₀ the minimal medium was prepared in D₂O and supplemented with 1.0 g.L⁻¹ of ¹⁵NH₄Cl and 2.0 g.L⁻¹ of ¹³C glucose.

The N_{A21}⁰-P₆₀ complex was prepared by incubating overnight at 4°C an excess of P₆₀ with the MBP-N_{A21}-RNA complexes. The MBP-N_{A21}⁰-P₆₀ complex was purified by Ni²⁺ chelate affinity chromatography by using the His-tag present on P₆₀ to remove the excess of free MBP-N_{A21}-RNA complex, followed by SEC on a Superdex S200 column equilibrated in buffer A and amylose affinity chromatography to eliminate unbound P₆₀.

The MBP tag was removed by incubating the protein with the TEV protease overnight at 4°C. The N protein contains the additional N-terminal tripeptide GAM. The N_{A21}⁰-P₆₀ complex was then purified using a Ni²⁺ chelate affinity chromatography followed by SEC on a Superdex S200 column equilibrated in buffer A. This procedure yielded pure N_{A21}⁰-P₆₀ complex. The samples were checked by SDS-PAGE using denaturing 4–20% gradient PAGE (Biorad).

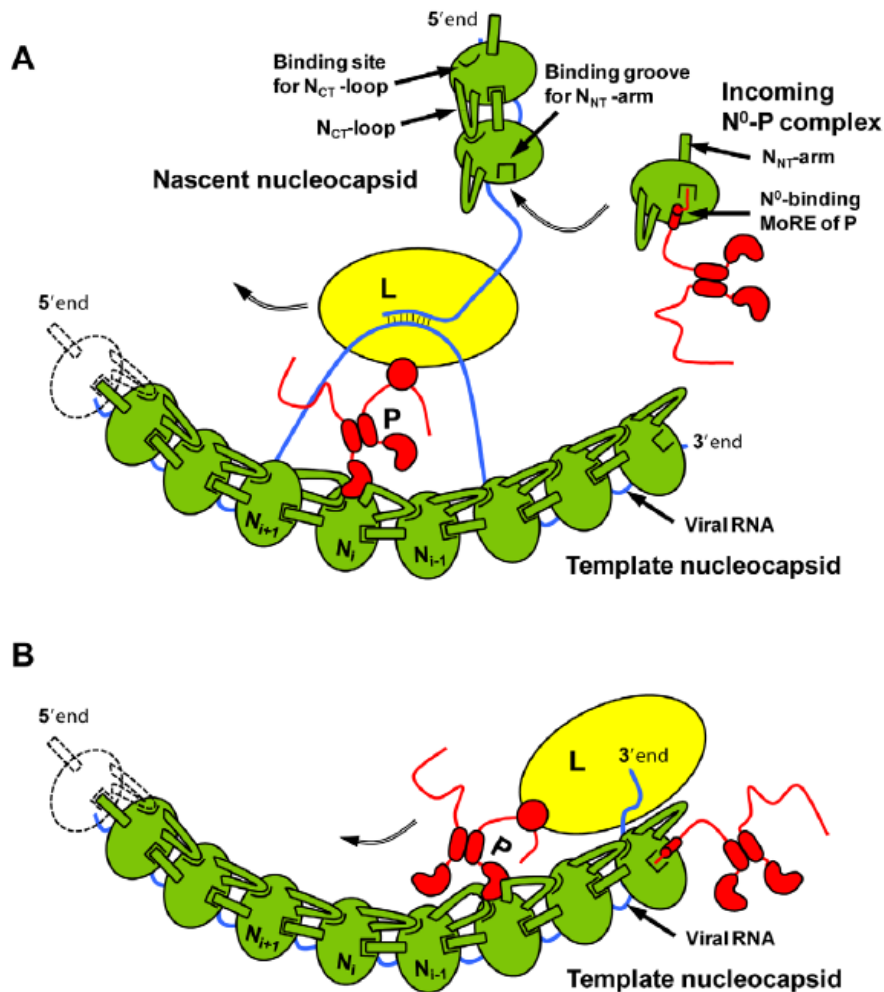


Figure 7. Schematic representations of the mechanism of RNA replication of VSV. The nucleoprotein (in green) forms with the RNA genome (blue line) the active template for the polymerase complex comprising the L (in yellow) and P (in red) proteins. (A) Encapsidation during RNA replication. During replication, the newly synthesized antigenomic or genomic RNA is encapsidated by nascent N molecules that are transferred from the soluble N⁰-P complex. In the N⁰-P complex, the N-terminal MoRE of P prevents host-cell RNA binding by obstructing the RNA binding groove and the self-assembly of N by interfering with the docking of the N_{NT}-arm of another N. Upon the transfer of N to the growing viral RNA P is released, the binding groove for the N_{NT}-arm is freed in the RNA-bound form and can accept the next incoming N molecule. (B) Initiation of RNA synthesis. By binding at the 3' extremity of the nucleocapsid, the N-terminal MoRE of P might displace nucleotides from the N molecule and allow the polymerase to initiate RNA synthesis.

doi:10.1371/journal.ppat.1002248.g007

Size exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry: SEC-MALLS

SEC was performed with a Superdex S200 column (GE Healthcare) equilibrated in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl. Separations were performed at 20°C with a flow rate of 0.5 ml.min⁻¹. 50 μL of a protein solution at a concentration ranging from 2.7 to 8.0 mg.mL⁻¹ were injected. On-line multi-angle laser light scattering (MALLS) detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting at 690 nm. Protein concentration was measured on-line by refractive index measurements using a RI2000 detector (Schambeck SFD) and a refractive

index increment $dn/dc = 0.185 \text{ mL.g}^{-1}$. Data were analyzed and weight-averaged molecular masses (M_w) were calculated using the software ASTRA V (Wyatt Technology Corp., Santa Barbara, CA) as described previously [11]. For size determination, the column was calibrated with proteins of known Stokes' radius (R_s) [31].

Small angle X-ray scattering (SAXS) and *ab initio* modeling

SAXS data were collected at the European Synchrotron Radiation Facility (E.S.R.F., Grenoble, France) on beamline ID14-3. The sample-to-detector distance was 1 m and the wavelength of the X-rays was 0.931 Å. Samples were contained

in a 1.9 mm wide quartz capillary. The exposition time was optimized for reducing radiation damage. Data acquisition was performed at 20°C. Data reduction was performed using the established procedure available at ID14-3 and buffer background runs were subtracted from sample runs.

The SAXS profile of the N_{A21}⁰-P₆₀ complex was recorded for scattering vectors, $q = \frac{4\pi \sin \theta}{\lambda}$, in the range $0.05 \text{ nm}^{-1} < q < 3.5 \text{ nm}^{-1}$. The profiles obtained at three different protein concentrations 2.7–8.0 mg.mL⁻¹ had the same shape and were flat at low q values indicating the absence of significant aggregation (Figure S4A). The radius of gyration and forward intensity at zero angle ($I(0)$) were determined with the programs PRIMUS [32] by using the Guinier approximation at low q values (Figure S4B), in a $q \cdot R_g$ range up to 1.3:

$$\ln I(q) = \ln I(0) - \frac{R_g^2 q^2}{3} \quad (1)$$

The forward scattering intensity was calibrated using bovine serum albumin and lysozyme as references. The radius of gyration, R_g , and the pair distance distribution function, $P(r)$, were calculated with the program GNOM [33] (Figure S4C). The maximum dimension (D_{max}) value was adjusted so that the R_g value obtained from GNOM agreed with that obtained from the Guinier analysis. The R_g values showed no significant dependence on protein concentration, confirming the absence of aggregation or intermolecular interactions in the concentration range used in this study. The measured R_g value of $2.7 \pm 0.1 \text{ nm}$ (Table S1 in Text S1), and the molecular mass of $65 \pm 15 \text{ kDa}$ derived from the scattering intensity at zero angle, I_0 , were in agreement with SEC-MALLS results. Additionally, the distance distribution function (Figure S4C) and the Kratky plot (Figure S4D) were typical of globular, folded proteins.

Ab initio low-resolution bead model reconstructions of the N_{A21}⁰-P₆₀ complex were performed from the scattering curves using the program DAMMIF [19]. This program restores a low-resolution shape of the protein as a volume filled with densely packed spheres (dummy atoms) that reproduces the experimental scattering curve by a simulated annealing minimization procedure. DAMMIF minimizes the interfacial area between the molecule and the solvent by imposing compactness and connectivity constraints. 20 independent models were generated with DAMMIF with no symmetry restriction and were combined with the program DAMAVER [34] yielding an average model that exhibited the common structural features of all reconstructions. The models were aligned pairwise by minimizing the normalized spatial discrepancy (NSD) score with the program SUPCOMB [35]. The mean NSD score of 0.75 (values ranging from 0.74 to 0.77) indicated an adequate convergence of the models [34]. All figures were generated with PyMOL (<http://www.pymol.org>).

Crystallization, data collection and structure determination and refinement

Crystallization conditions were screened by the hanging drop vapor diffusion method using a PixSys4200 Cartesian robot (high-throughput crystallization laboratory at EMBL Grenoble, France). The screen was performed by combining 0.1 μl of protein solution at 8 mg.mL^{-1} in buffer A with 0.1 μl of Hampton Crystal Screen solutions. Hits were reproduced manually. The N_{A21}⁰-P₆₀ complex crystallized at 20°C in 0.1 M sodium acetate buffer, pH 4.6,

containing 4% (w/v) of PEG4000. Single crystals were harvested from the drop, briefly soaked in the reservoir solution supplemented with 25% glycerol and flash frozen in liquid nitrogen at 100 K before data collection. X-ray diffraction data were collected at a wavelength of 0.933 Å on the ID14-2 beamline at the ESRF (Grenoble, France).

The data were processed using the program iMosflm [36] and scaled with the program Scala from the ccp4 suite [37]. The structure was solved by molecular replacement with the program Phaser [38] using residues 22 to 422 of a protomer of N (2GIC, chain E) extracted from the N-RNA crystal structure [15] as a search model. The visible part of P₆₀ was assigned and constructed with the program Buccaneer [39], and the overall structure was refined to 3.0 Å resolution using Coot [40] and Refmac5 [41]. The quality of the model was checked with PROCHECK [42]. Data collection and refinement statistics are summarized in Table 1.

The N_{A21}⁰-P₆₀ complex was crystallized in space group P2₁2₁2. The crystallographic asymmetric unit contained five protomers of N_{A21}⁰-P₆₀ complex. Each protomer includes residues 22 to 422 of N and residues 6 to 33 of P₆₀. In some protomers, residues 34 and 35 could also be constructed.

NMR spectroscopy

NMR experiments were performed on a Varian spectrometer operating at a ¹H frequency of 600 MHz. All samples contained 20 mM Tris-HCl, 150 mM NaCl, 50 mM Glu, 50 mM Arg with 10% D₂O adjusted to pH 6.0. The concentration of free ¹⁵N, ¹³C, ²H-labeled P₆₀ was 0.9 mM and the concentration of the N_{A21}⁰-P₆₀ complex (¹⁵N, ¹³C, ²H-labeled P₁₋₆₀ in complex with unlabeled N) was 0.28 mM. In the titration experiment, ¹⁵N-labeled P₆₀ was initially at 0.28 mM. MBP-N_{A21} was added at final concentrations of 0.09 mM or 0.17 mM. ¹⁵N R_2 (CPMG) relaxation experiments were acquired using standard pulse sequences [43]. The spectra were acquired with a sweep width of 8.0 kHz and 512 complex points in the ¹H dimension, and a sweep width of 1.2 kHz and 200 complex points in the ¹⁵N dimension. The magnetization decay was sampled at 10, 30, 50, 70, 90, 130, 170, 210 and 250 ms and the peak heights were used to extract the relaxation rates. To obtain estimates of the errors on the relaxation rates, a repeat measurement of one of the relaxation delays (70 ms) was carried out.

Supporting Information

Text S1 This file contains seven additional figures named S1 to S7 and an additional table S1. (DOC)

Acknowledgments

We thank Ivan Ivanov and Danielle Blondel for helpful discussions. We thank the Partnership for Structural Biology for the excellent structural biology environment. Crystallographic coordinates and structure factors are deposited in the Protein Data Bank with accession code 3PMK.

Author Contributions

Conceived and designed the experiments: CL FY MRJ MB RWHR MJ. Performed the experiments: CL FY NT EAR MRJ MB. Analyzed the data: CL FY NT MRJ MB RWHR MJ. Contributed reagents/materials/analysis tools: CL FY EAR MRJ MB MJ. Wrote the paper: CL MRJ MB RWHR MJ.

References

- Ge P, Tsao J, Schein S, Green TJ, Luo M, et al. (2010) Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* 327: 689–693.
- Amheiter H, Davis NL, Wertz G, Schubert M, Lazzarini RA (1985) Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell* 41: 259–267.
- Patton JT, Davis NL, Wertz GW (1984) N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus. *J Virol* 49: 303–309.
- Peluso RW, Moyer SA (1988) Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology* 162: 369–376.
- Howard M, Wertz G (1989) Vesicular stomatitis virus RNA replication: a role for the NS protein. *J Gen Virol* 70(Pt 10): 2683–2694.
- Masters PS, Banerjee AK (1988) Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J Virol* 62: 2658–2664.
- Chen M, Ogino T, Banerjee AK (2007) Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. *J Virol* 81: 13478–13485.
- Mavrakīs M, Mehoulas S, Real E, Iseni F, Blondel D, et al. (2006) Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology* 349: 422–429.
- Curran J, Marq JB, Kolakofsky D (1995) An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J Virol* 69: 849–855.
- Zhang X, Green TJ, Tsao J, Qiu S, Luo M (2008) Role of intermolecular interactions of vesicular stomatitis virus nucleoprotein in RNA encapsidation. *J Virol* 82: 674–682.
- Gérard FCA, Ribeiro EA, Albertini A, Zaccai G, Ebel C, et al. (2007) Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry* 46: 10328–10338.
- Gérard FCA, Ribeiro EA, Leyrat C, Ivanov I, Blondel D, et al. (2009) Modular organization of rabies virus phosphoprotein. *J Mol Biol* 388: 978–996.
- Leyrat C, Jensen MR, Ribeiro EA, Gérard F, Ruigrok R, et al. (2011) The N⁰-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α -helices. *Prot Sci* 20: 542–556.
- Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, et al. (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 313: 360–363.
- Green TJ, Zhang X, Wertz GW, Luo M (2006) Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 313: 357–360.
- Ribeiro EA, Leyrat C, Gérard FC, Albertini AA, Falk C, et al. (2009) Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes. *J Mol Biol* 394: 558–575.
- Mavrakīs M, Iseni F, Mazza C, Schoehn G, Ebel C, et al. (2003) Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells. *Virology* 305: 406–414.
- Ding H, Green TJ, Lu S, Luo M (2006) Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J Virol* 80: 2808–2814.
- Franko D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Cryst* 42: 342–346.
- Luo M, Green TJ, Zhang X, Tsao J, Qiu S (2007) Conserved characteristics of the rhabdovirus nucleoprotein. *Virus Res* 129: 246–251.
- Nayak D, Panda D, Das SC, Luo M, Pattnaik AK (2009) Single-amino-acid alterations in a highly conserved central region of vesicular stomatitis virus N protein differentially affect the viral nucleocapsid template functions. *J Virol* 83: 5525–5534.
- Gupta AK, Banerjee AK (1997) Expression and purification of vesicular stomatitis virus N-P complex from *Escherichia coli*: role in genome RNA transcription and replication in vitro. *J Virol* 71: 4264–4271.
- Castel G, Chtoumi M, Caignard G, Prehaud C, Mehoulas S, et al. (2009) Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. *J Virol* 83: 10808–10820.
- Ding H, Green TJ, Luo M (2004) Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). *Acta Crystallogr D Biol Crystallogr* 60: 2087–2090.
- Ribeiro EA, Jr., Favier A, Gerard FC, Leyrat C, Brutscher B, et al. (2008) Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J Mol Biol* 382: 525–538.
- Green TJ, Macpherson S, Qiu S, Lebowitz J, Wertz GW, et al. (2000) Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J Virol* 74: 9515–9524.
- Green TJ, Rowse M, Tsao J, Kang J, Ge P, et al. (2010) Access of RNA encapsidated in the nucleocapsid of vesicular stomatitis virus. *J Virol* 85: 2714–2722.
- Emerson SU, Schubert M (1987) Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc Natl Acad Sci U S A* 84: 5655–5659.
- Abraham G, Banerjee AK (1976) Sequential transcription of the genes of vesicular stomatitis virus. *Proc Natl Acad Sci U S A* 73: 1504–1508.
- Emerson SU (1982) Reconstitution studies detect a single polymerase entry site on the vesicular stomatitis virus genome. *Cell* 31: 635–642.
- Uversky VN (1993) Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* 32: 13288–13298.
- Komarev PV, Volkov VV, Sokolova A, M.H.J. K, Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* 36: 1277–1282.
- Semenyuk AV, Svergun D (1991) GNOM - a program package for small-angle scattering data processing. *J Appl Crystallogr* 24: 537–540.
- Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Cryst* 36: 860–864.
- Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Cryst* 34: 33–41.
- Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* 55: 1690–1695.
- Potterton E, Briggs P, Turkenburg M, Dodson E (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* 59: 1131–1137.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40: 658–674.
- Cowan KD (2006) The Buccaneer software for automated model building. *Acta Crystallogr D Biol Crystallogr* 62: 1002–1011.
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501.
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240–255.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26: 283–291.
- Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, et al. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* 33: 5984–6003.

SUPPORTING INFORMATION

**STRUCTURE OF THE VESICULAR STOMATITIS VIRUS N⁰-P
COMPLEX**

Cédric Leyrat, Filip Yabukarski, Nicolas Tarbouriech, Euripedes de Almeida Ribeiro Jr,
Malene Ringkjøbing Jensen, Martin Blackledge, Rob W. H. Ruigrok and Marc Jamin

Supplementary figures

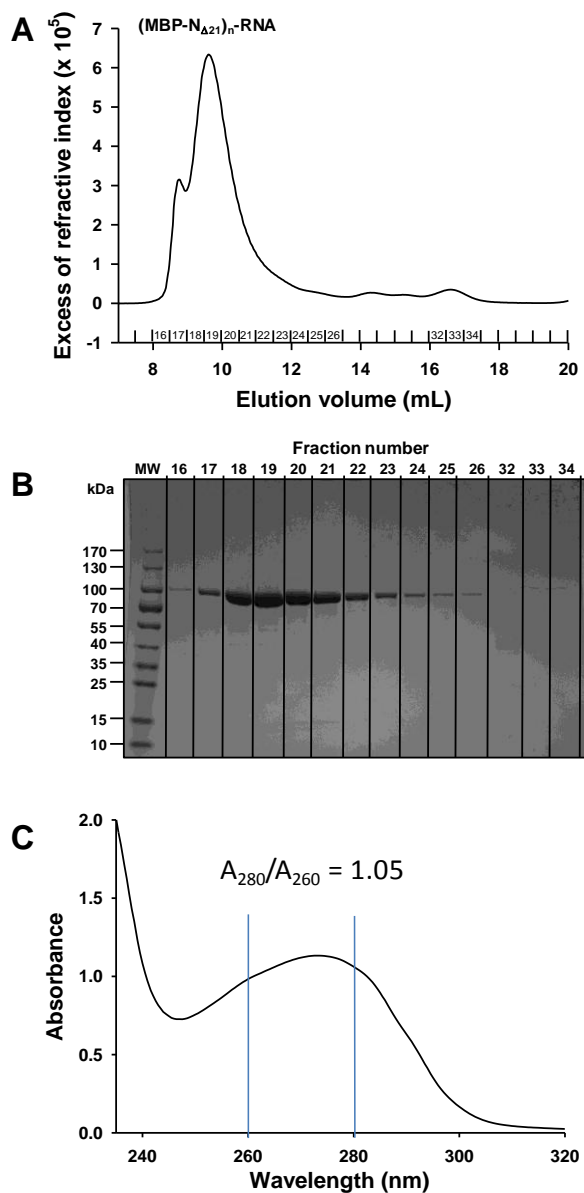


Figure S1. Production and purification of MBP-N₂₁-RNA complexes. (A) Expression of MBP-N₂₁ in *E. coli* led to the formation of soluble particles that eluted mainly in a peak at 9.4 mL close to the exclusion volume (8.6 mL) of a Superdex S200 column. The collected fractions analyzed by denaturing gel electrophoresis are numbered at the bottom of the figure. (B) Denaturing 4-20 % gradient PAGE stained by Coomassie blue. Fraction numbers correspond to those indicated in fig. S1A. The gel shows the presence of a single protein below the 100 kDa marker in agreement with the molecular mass expected for the chimeric MBP-N₂₁ (88,326 Da). (C) Absorbance spectrum of the fraction eluting in the main peak (9.4 mL) indicating the co-elution of nucleic acid with the protein.

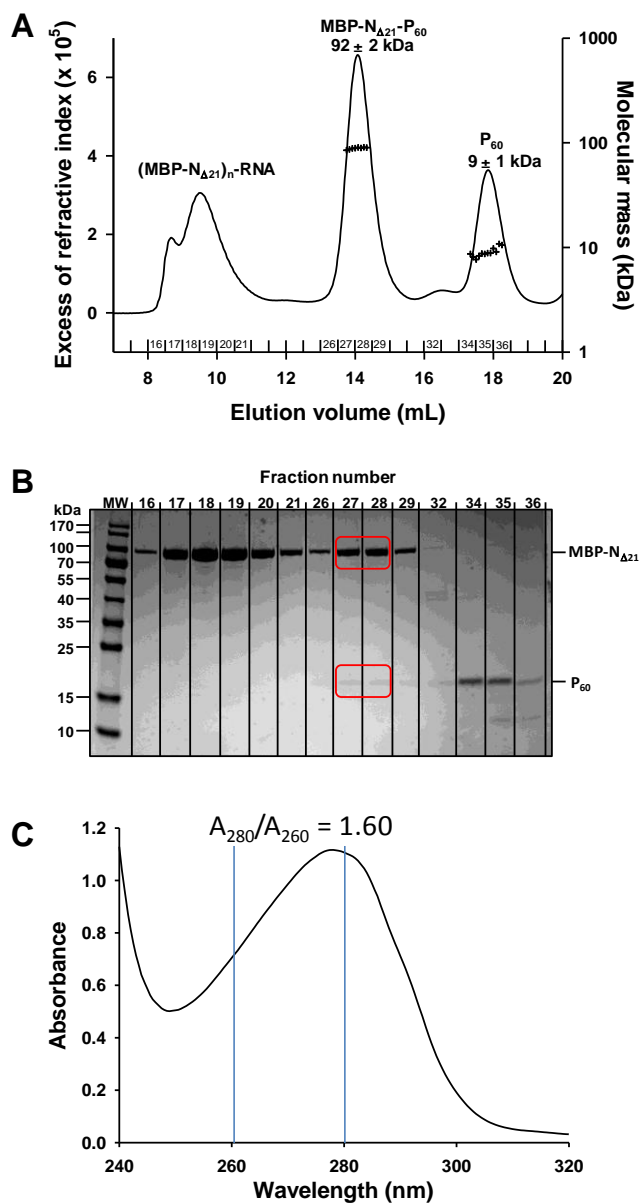


Fig. S2. Purification of the MBP-N Δ 21⁰-P₆₀ complex. (A) Overnight incubation of the MBP-N Δ 21 complex with an excess of P₆₀ led to the production of a new complex that eluted at 14.2 mL from the Superdex S200. The component eluting at 17.8 mL had a molecular mass of 9 ± 1 kDa and corresponded to the excess of P₆₀. The collected fractions analyzed by denaturing gel electrophoresis are numbered at the bottom of the figure. (B) Denaturing 4-20% gradient PAGE stained with Coomassie blue. MBP-N Δ 21 and P₆₀ co-eluted in the peak at 14.2 mL (red boxes), while the peak at 17.2 mL contains only P₆₀. Fraction numbers correspond to those indicated in Fig. S2A. (C) Absorbance spectrum of the fraction eluting in the peak at 14.2 mL indicating a lower nucleic acid content.

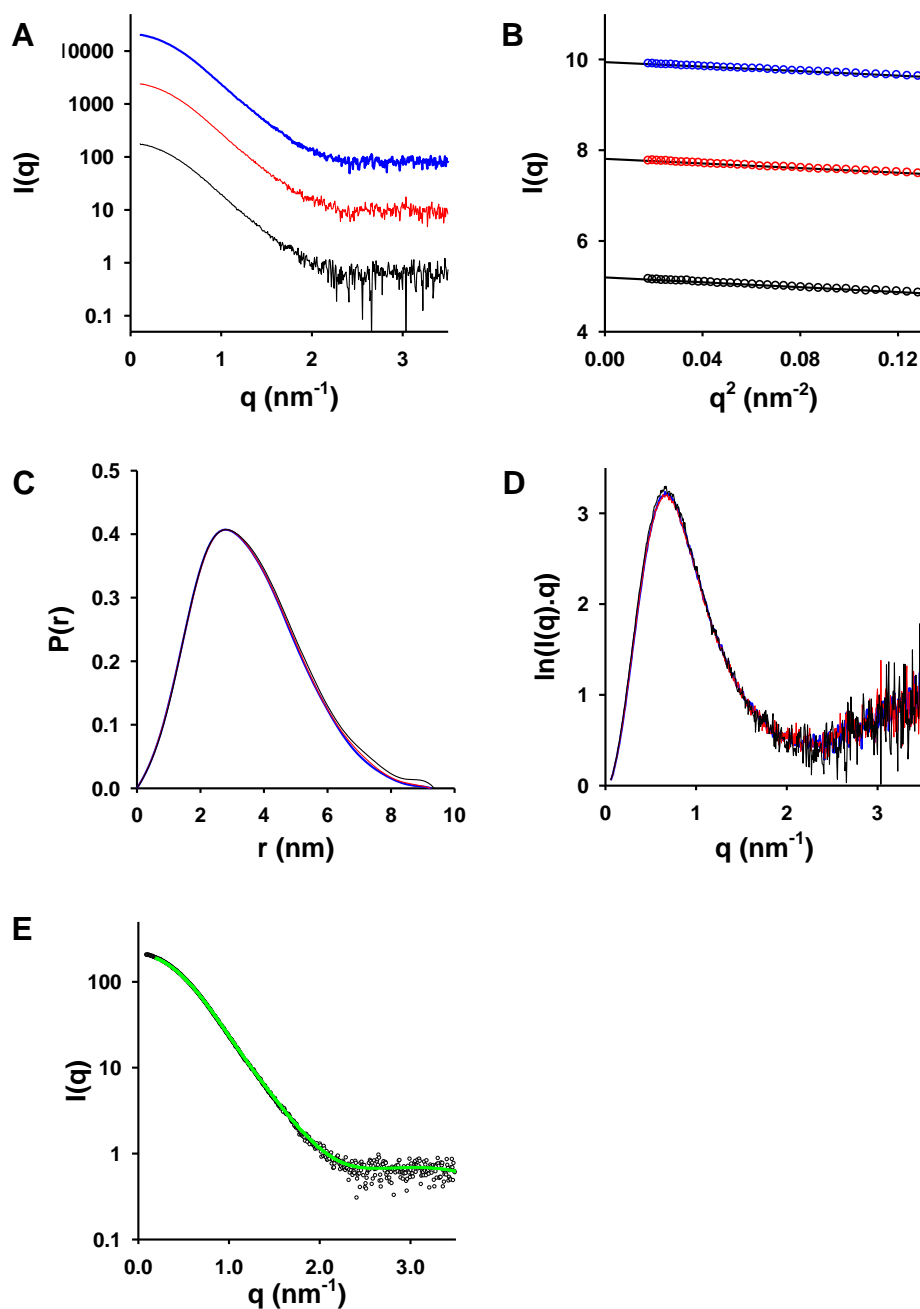


Fig. S3. Small-angle X-ray scattering experiments. (A) Scattering curve of N_{□21}⁰-P₆₀ recorded at different concentrations. Protein concentrations: 2.7 mg.mL⁻¹ (black curve), 5.3 mg.mL⁻¹ (red curve) or 8.0 mg.mL⁻¹ (blue curve). (B) Guinier plot. (C) Distance distribution function. D_{\max} values are shown in Table S1. The surface areas under the curves were normalized to account for the differences in protein concentration. (D) Kratky plot. The bell-shaped curves indicated a well-folded globular protein. (E) Modeling of SAXS data with DAMMIN. The theoretical SAXS curve derived from the bead model shown in Figure 2D (green line) reproduces the experimental SAXS data (open circles) up to 3.5 nm⁻¹ (shown in Figure 2C).

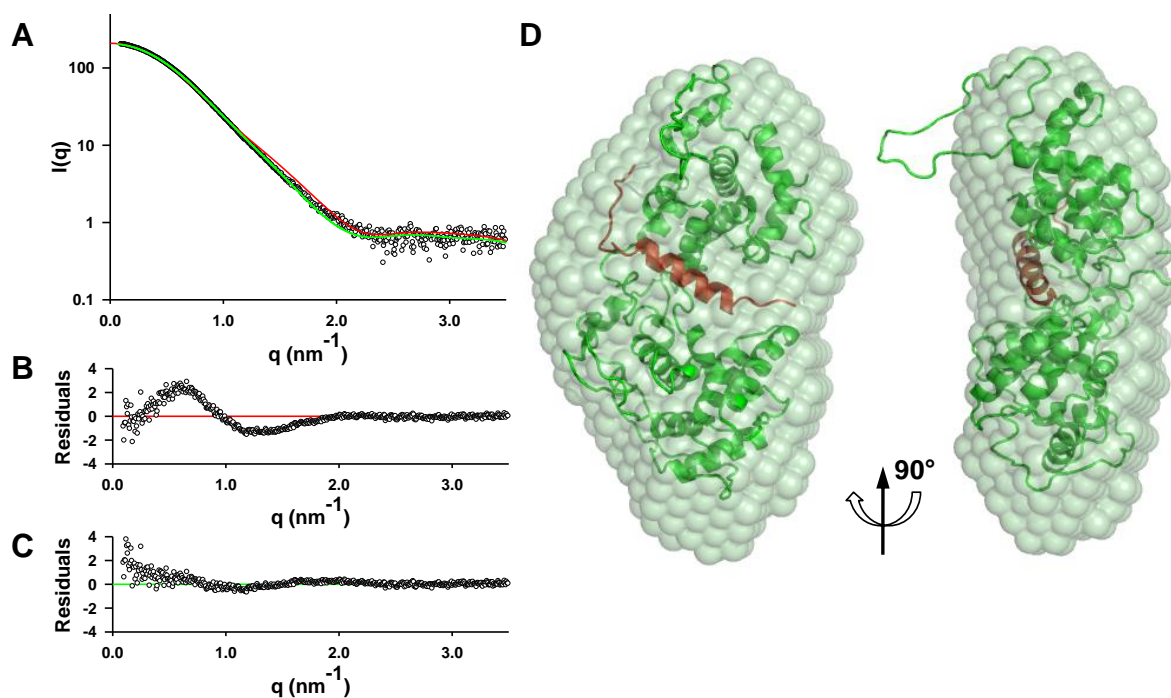


Fig. S4. *Ab initio* bead models of the $N_{21}^0-P_{60}$ complex. (A) Comparison of the theoretical SAXS curve calculated for a N protomer extracted from the circular N-RNA complex (2GIC chain E) (red line) or for a $N_{21}^0-P_{60}$ protomer extracted from the circular complex (green line) with the experimental SAXS curve recorded for the $N_{21}^0-P_{60}$ complex (dots). The SAXS curve was recorded at ESRF beamline ID 14-3. (B) Plot of the residuals between the theoretical curve calculated for a N protomer extracted from the circular N-RNA complex and the experimental curve of the soluble complex. (C) Plot of the residuals between the theoretical curve calculated for a $N_{21}^0-P_{60}$ and the experimental curve. (D) Superposition of the average *ab initio* bead model of the $N_{21}^0-P_{60}$ complex with a $N_{21}^0-P_{60}$ protomer extracted from the atomic structure of the circular complex.

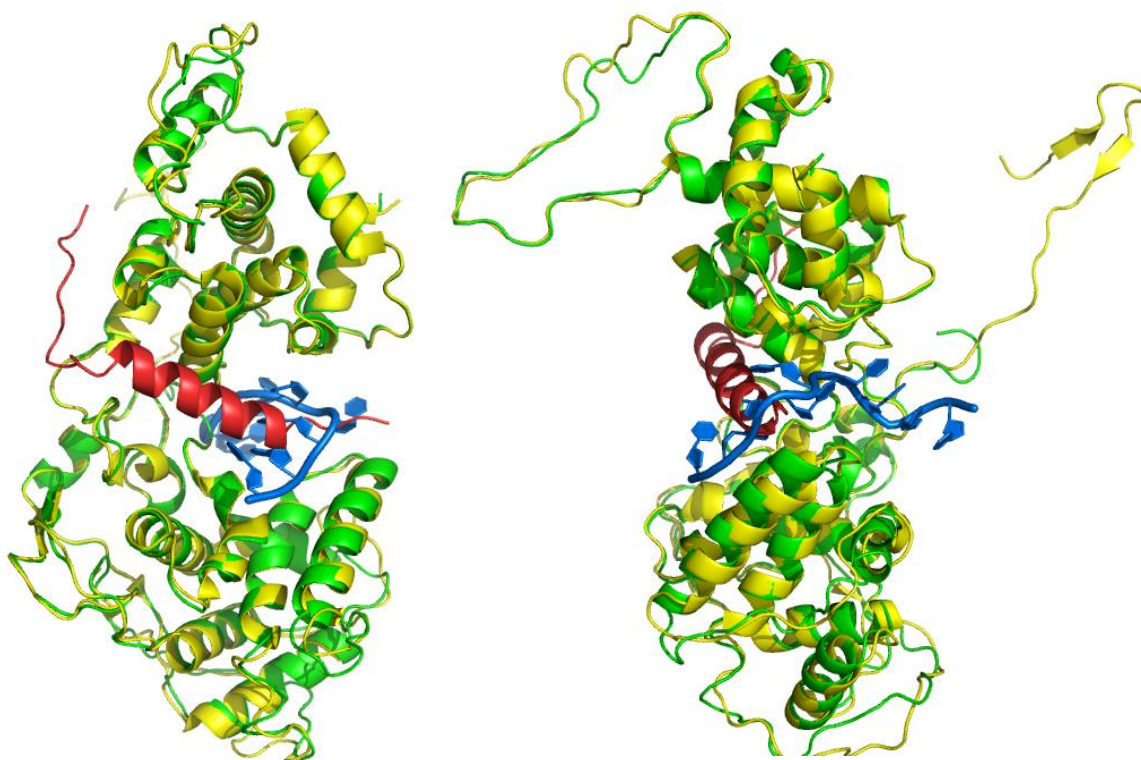


Fig. S5. Comparison of the structure of N in the N_{21} - P_{60} complex and in the N-RNA complex. The structure of N_{21} and of P_{60} (aa 6-35) extracted from the decameric complex (chains E and N) are shown in green and red, respectively. The structures of N and of the bound RNA extracted from the decameric complex (2GIC chain E) are shown in yellow and blue, respectively. The r.m.s.d. between the N molecule is 0.96 Å.



- ↑ Conserved residue – binds MoRE of P
- ↑ Similar residue – binds MoRE of P
- ↑ Conserved residue – binds MoRE of P and RNA
- ↑ Similar residue – binding MoRE of P and RNA
- ↓ Forms a salt bridge with a residue in the MoRE of P
- ↓ Interacts with RNA phosphate group in VSV – conserved in RAV
- ↓ Interacts with RNA phosphate group in VSV – not conserved in RAV

Figure S6. Sequence conservation between VSV (2GIC) and RAV N (2GTT). The alignment was performed with ClustalW2 [1] and the figure was prepared with ESPript [2]. The identities are displayed as white characters in a red box, while the similarities (using the Risler score [3]) are displayed in red. The blue box highlights the central hinge region (aa 200-300). The close arrows in dark and light blue show the identical and similar residues lining the binding groove of the MoRE of P (circled in black in figure 5B), respectively. The open arrows in dark and light blue show the identical and similar residues forming a hydrophobic surface of N that interacts with both P and the RNA. The arrows in magenta show the three residues of VSV N forming salt bridges with the MoRE of P (Figure 5A). The green arrows show the basic residues of VSV N that binds directly phosphate groups of the RNA. The close green arrows show the residues that are conserved in RAV.

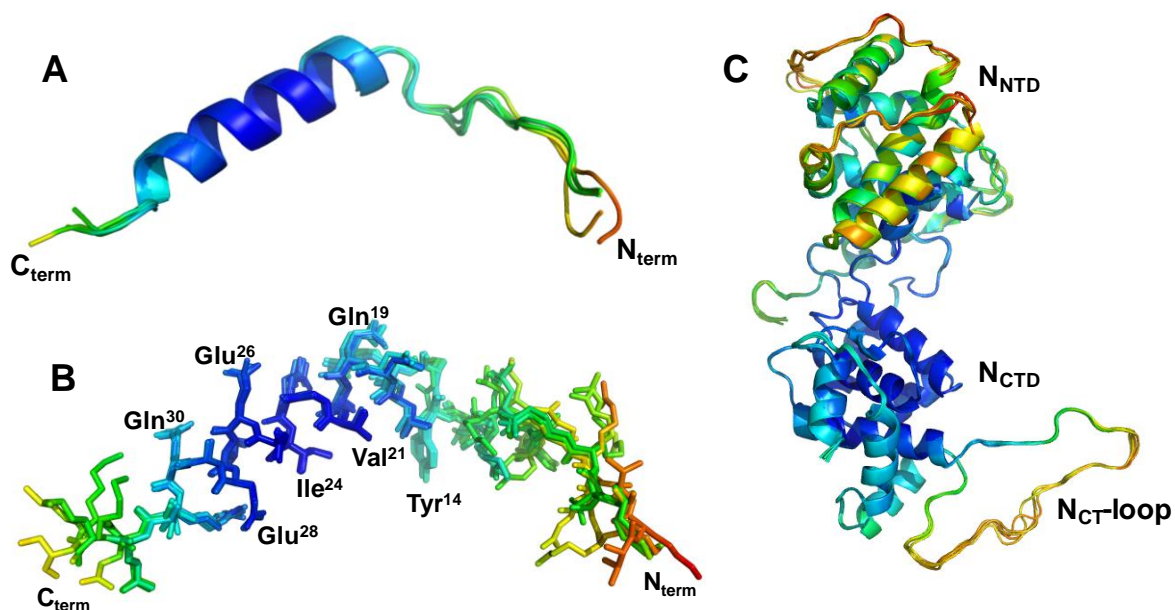


Fig. S7. Superposition of the 5 protomers of P₆₀ and of N₂₁ forming the asymmetric unit. (A, B) The overlay of the peptide moiety from the five protomers of the asymmetric unit revealed that residues 14 to 31 exhibit the lowest level of flexibility. The backbones of the different protomers take similar conformations but the side chains of residues 6 to 11 and of residues 32 to 35 take various orientations in the different protomers. The remaining residues (aa 1-5 and aa 36-68) are not visible in the crystal structure probably owing to their structural flexibility. (C) The different protomers of N₂₁ are very similar. The backbone structures are color coded according to the crystallographic B-factors ranging from blue (low B-factor values) to red (high B-factor value).

Table S1. Molecular dimensions of the N₂₁⁰-P₆₀ complex calculated from SAXS data

Protein concentration (mg.mL⁻¹)	R_g (Guinier) (nm)	D_{max} (nm)
2.7	2.7 ± 0.1	9.3
5.3	2.7 ± 0.1	9.2
8.0	2.8 ± 0.1	9.4

References

1. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31: 3497-3500.
2. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31: 3320-3323.
3. Risler JL, Delorme MO, Delacroix H, Henaut A (1988) Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. *J Mol Biol* 204: 1019-1029.

5.3. Le Complexe N⁰-P entière de VSV

5.3.1. Stratégie de production des complexes N⁰-P de VSV

La production du complexe N⁰-P entière nécessite plusieurs étapes. Tout d'abord le complexe N_{Δ21}⁰-P₆₀ est produit comme décrit dans (Leyrat, Yabukarski et al. 2011) mais en conservant l'étiquette MBP. La P entière de VSV est produite comme décrit dans (Gérard, Ribeiro et al. 2007). Ensuite, afin de déplacer le peptide P₆₀, les deux protéines sont mélangées et incubées à 4°C pendant la nuit. Le mélange de protéines est passé sur une colonne d'amylose afin de se débarrasser de la P qui n'a pas formé le complexe N_{Δ21}⁰-P et du peptide P₆₀ déplacé. Ainsi le complexe MBP-N_{Δ21}⁰-P est retenue sur la colonne tandis que la P libre et le peptide P₆₀ déplacé passent à travers la résine pendant les étapes de lavage.

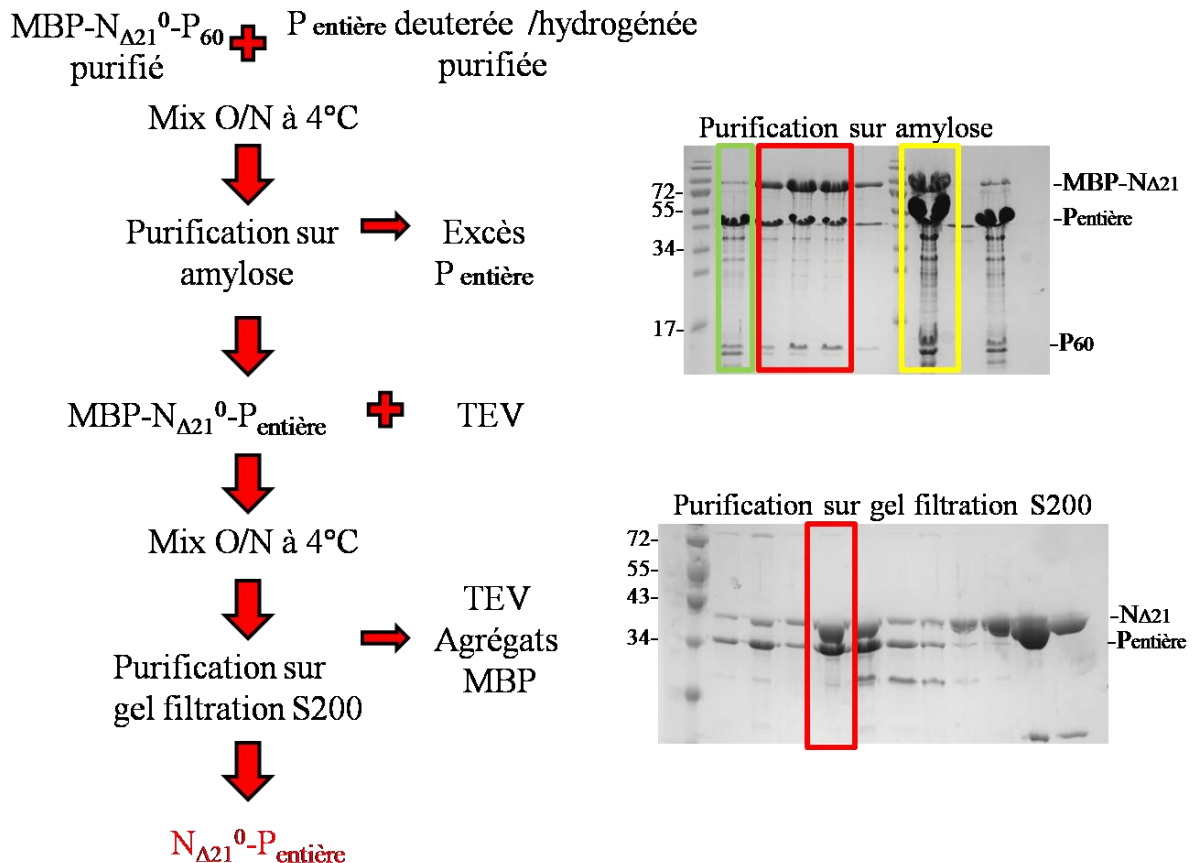


Figure 2 : Stratégie de purification du complexe N⁰-P entière du virus de la stomatite vésiculaire (VSV). Le complexe MBP-N_{Δ21}⁰-P₆₀ et la P entière sont mélangées et incubées pendant la nuit à 4°C. Le mélange est passé sur une colonne d'amylose (New England Biolabs) et ensuite la protéase TEV est ajoutée afin de cliver la protéine de fusion MBP. La dernière étape de purification est une colonne de gel filtration Superdex S200. Sur les gels : les fractions contenant le complexe d'intérêt en rouge, le matériel de départ (jaune) et la P lavée en vert.

Cette étape est importante car le complexe N_{Δ21}⁰-P entière et la P seule ont des volumes d'éluion proches et la séparation sur chromatographie d'exclusion de taille est

difficile (voir **Figure 3**). Après élution du complexe MBP-N $_{\Delta 21}^0$ -P de la résine amylose, l'étiquette MBP est clivée par la protéase TEV. Le mélange est ensuite concentré et injecté sur une colonne de chromatographie d'exclusion de taille Superdex S200 afin d'éliminer la MBP, la TEV et des agrégats éventuels. Les fractions contenant le complexe N $_{\Delta 21}^0$ -P sont regroupées et concentrées. De cette manière, nous avons produit le complexe N $_{\Delta 21}^0$ -P dans lequel la P est deutérée ou hydrogénée. La stratégie de purification est résumée dans la **Figure 2**. Le complexe contenant la P deutérée a été étudié par SANS avec variation de contraste en utilisant des solutions contenant 0, 20, 40 et 100% de D $_2$ O dans un tampon 20 mM Tris-HCl à pH 7.5, contenant 500 mM NaCl et 5 mM DTT. Le complexe contenant la P hydrogénée a été étudié par SAXS dans du tampon 20 mM Tris-HCl à pH 7.5 contenant 150 mM NaCl, 50 mM Glutamate et 50 mM Arginine afin de réduire les dégâts causés par l'irradiation.

La stratégie résumée dans **Figure 2** a été également employée afin de produire un complexe N 0 -P contenant le fragment P $_{1-177}$ - le complexe N $_{\Delta 21}^0$ -P $_{1-177}$. Ce complexe a été également étudié par SAXS (**Figure 5**).

Une deuxième stratégie de production du complexe N 0 -P consiste à mélanger directement la MBP-N $_{\Delta 21}$ avec la P entière ou avec le fragment P $_{1-177}$ purifiés séparément. Cette stratégie a l'avantage de permettre de contrôler plus facilement le rapport $[N]/[P]$ lors de la formation du complexe N 0 -P. Une partie de la P entière élue dans le volume mort de la colonne chromatographie d'exclusion de taille, probablement en raison de la fixation de ses domaines C-terminaux sur les complexes MBP-N-ARN.

Nous avons ainsi utilisé cette méthode afin de produire le complexe N $_{\Delta 21}^0$ -P $_{1-177}$ pour les études de RMN. Ainsi après le mélange de MBP-N $_{\Delta 21}$ et la P $_{1-177}$ marquée à l'isotope stable ^{15}N et leur incubation à 4 °C pendant une nuit la solution est chargée sur une résine Ni $^{2+}$. Le complexe contenant une étiquette hexahistidine sur la P $_{1-177}$ s'accroche sur la résine tandis que la MBP-N $_{\Delta 21}$ n'est pas retenue et passe à travers la résine. Après lavage, le complexe MBP-N $_{\Delta 21}^0$ -P $_{1-177}$ est élué, concentré et injecté sur une colonne de chromatographie d'exclusion de taille Superdex S200. Les fractions correspondant au complexe MBP-N $_{\Delta 21}^0$ -P $_{1-177}$ sont collectées et mises ensemble. Après le clivage du tag MBP par la protéase TEV, les protéines sont à nouveau concentrées et injectées sur la même colonne de chromatographie d'exclusion de taille afin de se débarrasser de la MBP, la TEV et des agrégats éventuels. Afin d'étudier l'importance du rapport des concentrations de départ de N et P nous avons produit le complexe N $_{\Delta 21}^0$ -P $_{1-177}$, dans lequel la P $_{1-177}$ est marquée ^{15}N , à deux rapports différents : $[N]/[P] \sim 15$ et $[N]/[P] \sim 2$

5.3.2. Le complexe N⁰-P n'a pas une stœchiométrie fixe

La **Figure 3** montre les résultats obtenus avec des préparations représentatives de N_{Δ21}⁰-P entière et N_{Δ21}⁰-P₁₋₁₇₇ et la comparaison de leurs volumes d'éluion avec ceux de la P entière ou la P₁₋₁₇₇ seules, respectivement. L'analyse par SEC-MALLS-RI de plusieurs préparations du complexe N_{Δ21}⁰-P entière et N_{Δ21}⁰-P₁₋₁₇₇ a montré que la stœchiométrie du complexe dépend de la manière dont l'échantillon est préparé, et plus précisément, dépend des concentrations relatives de N et P utilisées. Ainsi les masses moléculaires (MM) obtenues par MALLS sont souvent intermédiaires entre la MM correspondant à 1N:P et une MM correspondant à 2N-P. (Rappel : la P entière et la P₁₋₁₇₇ sont toujours des homodimères à la différence de P₁₋₁₀₅ et P₁₋₆₀ qui sont monomériques (Gérard, Ribeiro et al. 2007; Leyrat, Schneider et al. 2012))

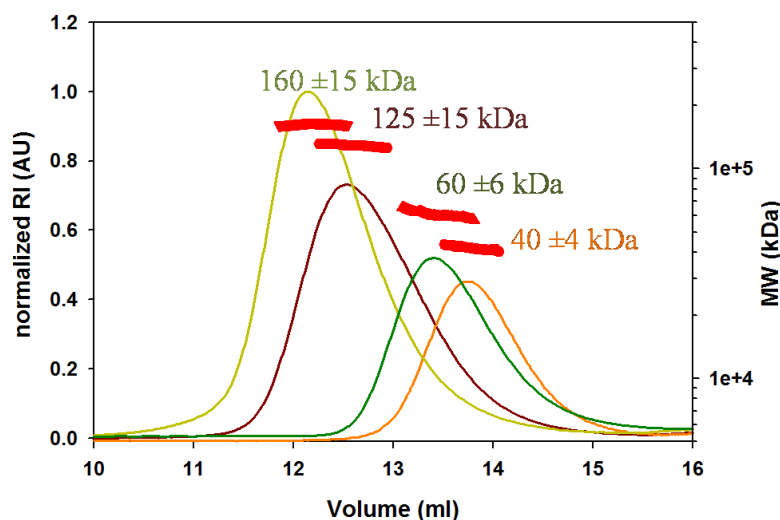
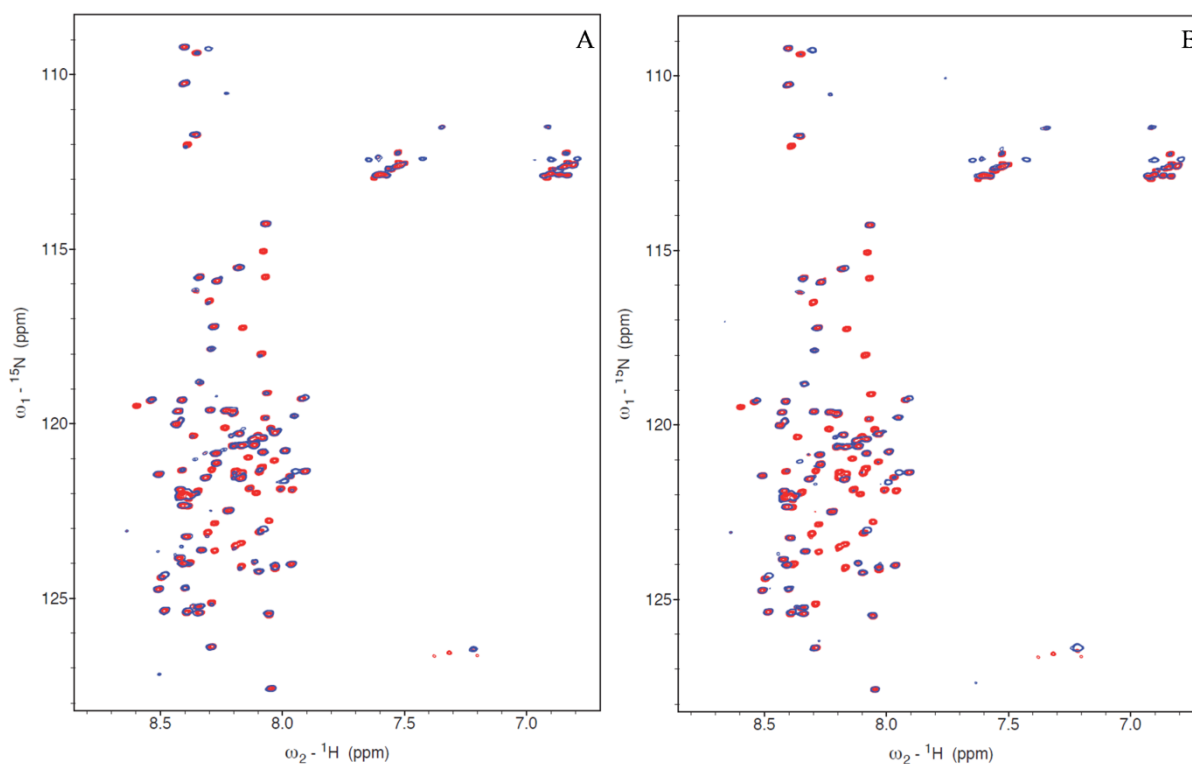


Figure 3 : Masse moléculaire et volumes d'éluion de la phosphoprotéine (P) entière et le fragment P₁₋₁₇₇ ainsi que des complexes N_{Δ21}⁰-P₁₋₁₇₇ et N_{Δ21}⁰-P entière du virus de la stomatite vésiculaire (VSV), mesurées par SEC-MALLS-RI. Profils d'éluion de P₁₋₁₇₇ (ligne en orange), P entière (ligne en vert foncé), N_{Δ21}⁰-P₁₋₁₇₇ (ligne en marron) et N_{Δ21}⁰-P entière (ligne en vert clair) de la colonne de gel filtration Superdex S200, équilibrée dans 20 mM Tris-HCl pH 7.5, 150 mM NaCl. En rouge les masses moléculaires (MM) à travers les pics d'éluion, calculées à partir de la diffusion statique de la lumière et la réfractométrie pour chaque protéine. Les valeurs numériques des MM ont le même code couleur que les profils d'éluion.

Typiquement, les MM obtenues par MALLS pour le complexe N_{Δ21}⁰-P₁₋₁₇₇ varient entre 109 ± 10 et 129 ± 10 kDa (MM théorique 88.6 kDa pour 1N-P (1x45.6 kDa + 43 kDa) et 134.2 kDa pour 2N-P (2x45.6 + 43 kDa)) tandis que pour N_{Δ21}⁰-P entière la MM expérimentale varie entre 150 ± 15 et 160 ± 15 kDa (MM théorique 107.6 kDa pour 1N-P (1x45.6 kDa + 62 kDa) et 153.2 kDa 2N-P (2x45.6 kDa + 62 kDa)). Afin de déterminer si le complexe 2N : P pouvait être formé, nous avons eu recours à la spectroscopie RMN. Les

spectres 2D [^1H - ^{15}N]-HSQC obtenus avec des complexes $\text{N}_{\Delta 21}^0\text{-P}_{1-177}$ préparés dans des différentes conditions de saturation par la N se caractérisent par des variations d'intensité des résonances. Dans le spectre RMN du P_{1-177} , seule la partie N-terminale donne des résonances visibles. Les résonances pour les résidus appartenant au domaine de dimerisation (aa 107-177) sont élargies au delà de la détection. Nous avons donc utilisé l'attribution des résonances réalisée pour la partie désordonnée N-terminale avec le peptide P_{105} . La comparaison des intensités des résonances des spectres du complexe $\text{N}_{\Delta 21}^0\text{-P}_{1-177}$ avec l'intensité des résonances dans le spectre de la protéine P_{105} seule montre que les résonances pour les résidus 1 à 40 sont réduites en présence de N^0 . Le rapport normalisé des intensités pour ces résidus (aa 1 à 40) nous donne la fraction de peptide libre et nous informe donc directement sur le degré de saturation du complexe $\text{N}_{\Delta 21}^0\text{-P}_{1-177}$. La **Figure 4C et D** montre que pour un rapport $[\text{N}]/[\text{P}] \sim 15$, nous observons la présence de 5-10% de P_{NTR} libre tandis que pour un rapport $[\text{N}]/[\text{P}] \sim 2$ nous observons la présence de 15-20% de P_{NTR} libre. Dans le cas d'un complexe contenant essentiellement une molécule de N par dimère de P (P_{1-177}), comme proposé dans (Mavrakis, Iseni et al. 2003) on s'attendrait à observer 50% de P_{NTR} libre. En conclusion, un dimère de phosphoprotéine peut donc fixer deux N, une sur chaque protomère, et le taux de saturation dépend des concentrations relatives de N et P.



Légende sur la page suivante

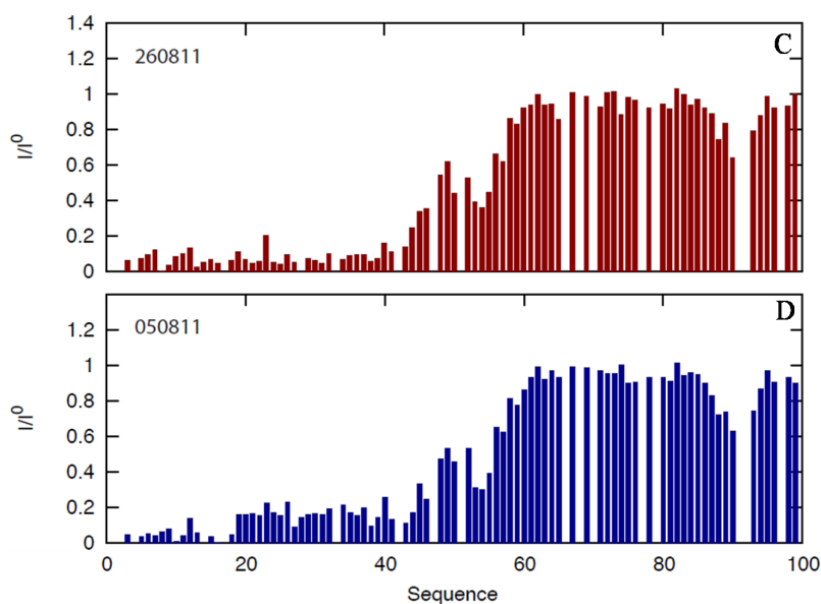


Figure 4: Etude en solution du complexe $N_{\Delta 21}^0$ -P₁₋₁₇₇ du virus de la stomatite vésiculaire (VSV) par Résonance Magnétique Nucléaire (RMN). A) et B) Superpositions des spectres 2D [¹H-¹⁵N]-HSQC du complexe $N_{\Delta 21}^0$ -P₁₋₁₇₇ (en bleu), reconstitué en utilisant des ratios $[N]/[P]$ de départ d'environ 15 et 2, respectivement, et du spectre 2D [¹H-¹⁵N]-HSQC du VSV P₁₀₅ (en rouge). Les spectres ont été enregistrés dans 20 mM Bis-Tris pH 6.0, 150 mM NaCl et 50 mM glutamate/arginine (RE). C) et D) Les graphiques représentent le rapport normalisé des intensités des pics des spectres 2D [¹H-¹⁵N]-HSQC du complexe $N_{\Delta 21}^0$ -P₁₋₁₇₇ de A) et B), respectivement, comparés aux intensités des pics du spectre 2D [¹H-¹⁵N]-HSQC du peptide VSV P₁₋₁₀₅ en fonction du numéro des résidus dans la séquence. Le rapport des intensités des résidus 1 à 40 nous renseigne sur le degré de saturation de la P₁₋₁₇₇ par la N.

5.3.3. Etude en solution des complexe $N_{\Delta 21}^0$ -P₁₋₁₇₇ et $N_{\Delta 21}^0$ -P entière

Des courbes de SAXS ont été enregistrées sur la ligne BioSAXS de l'ESRF pour chacun de ces complexes. L'allure des courbes ne change pas avec la concentration (1.0, 0.8 et 0.5 mg.mL⁻¹ pour $N_{\Delta 21}^0$ -P₁₋₁₇₇ et 1.3, 1.0 et 0.8 pour mg.mL⁻¹ pour $N_{\Delta 21}^0$ -P entière) ce qui montre l'absence d'agrégation (**Figure 5A et B**). Le R_g obtenu à partir de l'approximation de Guinier est de 6.8 ± 0.2 pour $N_{\Delta 21}^0$ -P₁₋₁₇₇ et 7.1 ± 0.2 pour $N_{\Delta 21}^0$ -P entière et ne dépend pas de la concentration ce qui confirme l'absence d'agrégation. La fonction de distribution des paires P(r) des deux complexes est très asymétrique, caractéristique d'une molécule allongée. Elle possède deux maxima, un vers 3.5 nm et un second vers 10 nm, avec une D_{max} d'environ 23 ± 0.5 nm et 25 ± 0.5 pour $N_{\Delta 21}^0$ -P₁₋₁₇₇ et $N_{\Delta 21}^0$ -P entière respectivement (**Figure 5E et F**). La présence de plusieurs maxima dans la fonction P(r) est caractéristique des protéines à plusieurs domaines (Putnam, Hammel et al. 2007; Bernado 2010).

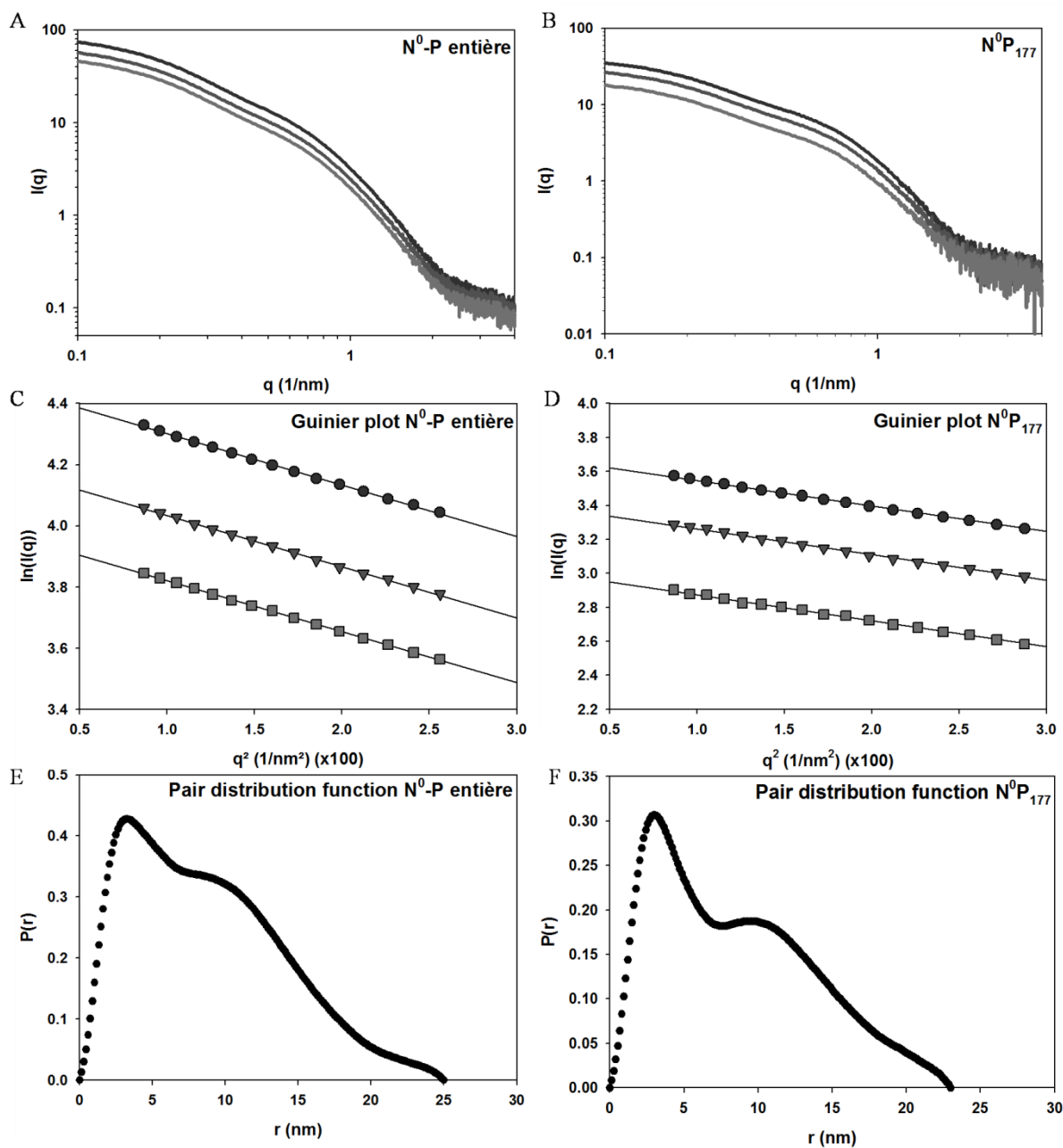


Figure 5: Expériences de diffusion des rayons X aux petits angles (SAXS) sur les complexes $N_{\Delta 21}^0$ -P₁₋₁₇₇ et $N_{\Delta 21}^0$ -P entière du virus de la stomatite vésiculaire (VSV). Les courbes de SAXS ont été enregistrées pour des valeurs de q entre 0.05 nm^{-1} et 4 nm^{-1} pour **A)** $N_{\Delta 21}^0$ -P entière à 1.3, 0.1 et 0.8 mg.ml^{-1} et **B)** $N_{\Delta 21}^0$ -P₁₋₁₇₇ à 1.0, 0.8 et 0.5 mg.ml^{-1} . **C)** et **D)** Analyse de Guinier dans la région de $q.R_g < 1.1$ pour $N_{\Delta 21}^0$ -P et $N_{\Delta 21}^0$ -P₁₋₁₇₇, respectivement. **E)** et **F)** Fonctions de distribution des distances $P(r)$ pour $N_{\Delta 21}^0$ -P et $N_{\Delta 21}^0$ -P₁₋₁₇₇, respectivement.

Le graphique de Kratky montre un maximum symétrique pour les deux protéines centré sur 0.7 nm^{-1} , et revient à une ligne de base horizontale caractéristique d'une protéine structurée contenant des régions flexibles (**Figure 6G et H**) (Putnam, Hammel et al. 2007). Les plateaux à très faible pente observés dans les graphiques de Porod-Debye suggèrent également que les complexes sont plutôt structurés (**Figure 5I et J**) (Rambo and Tainer 2011).

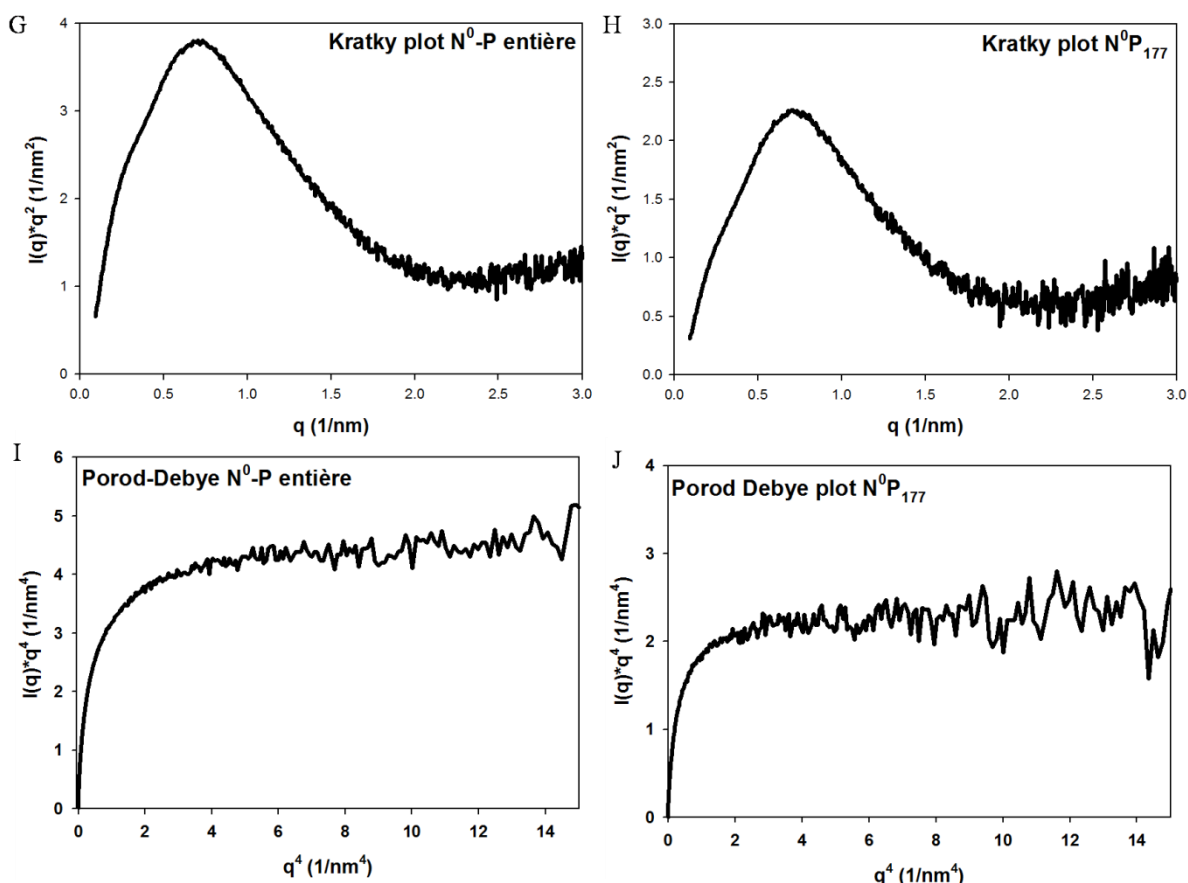


Figure 6: Expériences de diffusion des rayons X aux petits angles (SAXS) et SEC-MALLS-RI sur les complexes $N_{\Delta 21}^0$ - P_{1-177} et $N_{\Delta 21}^0$ - P entière du virus de la stomatite vésiculaire (VSV). **G** et **H**) Représentations de Kratky ; **I** et **J**) Représentations de Porod-Debye pour $N_{\Delta 21}^0$ - P et $N_{\Delta 21}^0$ - P_{1-177} , respectivement.

Dans le but d'obtenir de l'information par rapport aux propriétés structurales internes de chaque partenaire du complexe N^0 - P nous avons utilisé la technique de SANS combiné à la variation de contraste. Nous avons utilisé un complexe $N_{\Delta 21}^0$ - P entière qui est produit de la même façon que le complexe étudié par SAXS et dans lequel la P est deutérée. Ainsi nous avons enregistré des courbes de diffusion à 0, 20, 40 et 100% D_2O et à une seule concentration de complexe de 4.4 à 5.6 $mg.ml^{-1}$ par condition. Les courbes expérimentales sont représentées dans la **Figure 7B** et les R_g obtenus par l'approximation de Guinier sont présentés dans le tableau de la **Figure 7B**. A 0% D_2O (100% H_2O) les deux protéines du complexe, la N « hydrogénée » et la P « deutérée », contribuent à la courbe de diffusion. A 20% la contribution de la N est réduite, et est annulée à 40% D_2O .

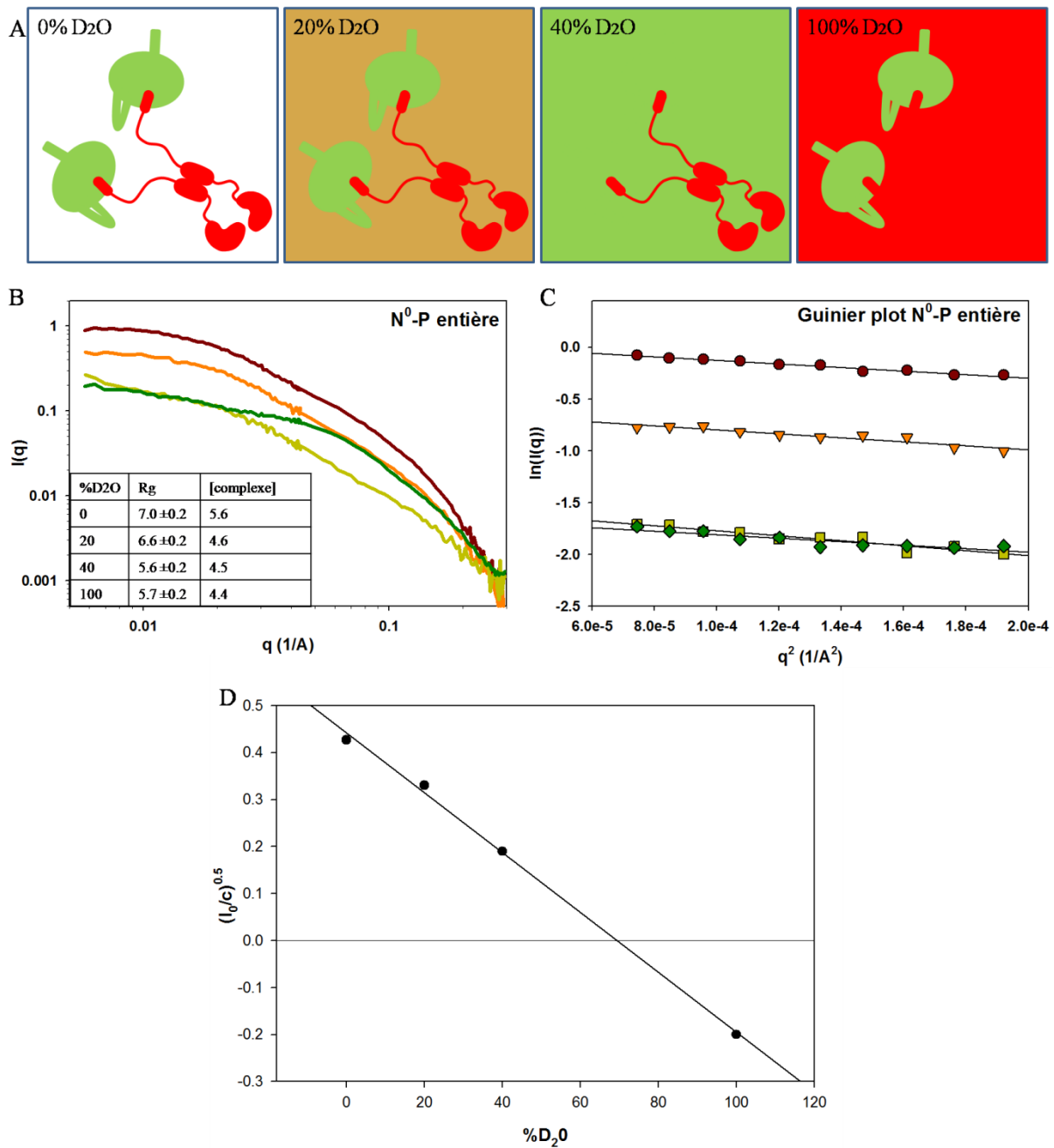


Figure 7 : Expériences de diffusion de neutrons aux petit angles (SANS) combiné à la variation de contraste sur le complexe $N_{\Delta 21}^0$ -P entière du virus de la stomatite vésiculaire (VSV). **A**) Illustration de la variation du contraste dans le cas du complexe $N_{\Delta 21}^0$ -P entière. Un modèle 2N:P est représenté. A 0% D₂O un contraste entre le complexe $N_{\Delta 21}^0$ -P entière et le solvant existe et tout le complexe contribue à la diffusion ; à 40% D₂O il n'y a pas de contraste entre N et le solvant et la diffusion provient seulement de la P ; à 100% D₂O il n'y a pas de contraste entre P et le solvant et la diffusion provient seulement de la N. **B**) Courbes de SANS enregistrées pour des valeurs de q entre 0.05 nm⁻¹ et 3 nm⁻¹ à 0% (courbe en marron), 20% (courbe en orange), 40% (courbe en vert foncé) et 100% D₂O (courbe en vert claire). Les concentrations du complexe et les valeurs de R_g, obtenues à partir de l'approximation de Guinier, sont données dans le tableau. **C**) Analyse de Guinier pour des valeurs de $q \cdot R_g < 1.0$. **D**) Graphe de Stuhrmann.

Dans ces conditions, le signal mesuré provient exclusivement de la P. A 100% D₂O, seul la protéine N hydrogénée contribue à la courbe de diffusion (**Figure 7A** pour illustration). Le graphe de Stuhrmann (**Figure 7D**) montre une dépendance linéaire du

$\sqrt{I(0)/C}$ (C étant la concentration en protéine) par rapport au pourcentage de D₂O indiquant l'absence d'agrégation pendant la collecte de données.

5.3.4. Modélisation du complexe N_{Δ21}⁰-P entière

L'étude de la P entière de VSV ((Leyrat, Schneider et al. 2012) **Chapitre 3**) a montré que la partie N-terminale de la protéine (aa 1-105) est globalement désordonnée et que les chaînes provenant des deux protomères qui forment le dimère de P sont indépendantes l'une de l'autre. Puisque la comparaison des spectres RMN de P₁₋₁₀₅ et de P₁₋₁₇₇ au sein du complexe N_{Δ21}⁰-P₁₋₁₇₇ indique que les seuls résidus qui sont affectés par la formation du complexe sont ceux de la région d'interaction (aa 1 à 40), nous avons décidé de modéliser le complexe N_{Δ21}⁰-P entière sous la forme d'un ensemble de conformères (Bernado, Mylonas et al. 2007; Bernado and Svergun 2012).

Afin de prendre en compte l'existence de complexes avec les stoechiométries 1N:P et 2N:P dans les échantillons préparés, nous avons généré un ensemble de départ de 10,000 conformères qui contient des modèles atomiques du complexe N_{Δ21}⁰-P entière avec 1N ou 2N. Dans les modèles 2N:P les deux P_{NTR} de la P portent chacun une N tandis que dans les modèles 1N:P seulement un P_{NTR} porte une N – l'autre est libre (**Figure 8A**). Nous avons ensuite utilisé le programme GAJOE (Bernado, Mylonas et al. 2007) pour sélectionner un sous-ensemble de 50 conformères qui représentent collectivement les quatre courbes de diffusion de neutrons aux petits angles (SANS) et la courbe de diffusion de rayons X aux petits angles (SAXS). La répétition de cette procédure de sélection montre que différents ensembles de conformères peuvent reproduire de manière équivalente les courbes expérimentales. En moyenne, ces différents ensembles de 50 conformères contiennent 65% de complexe 1N:2P et 35 % de complexe 2N:2P.

La **Figure 8** montre le résultat d'une telle sélection d'un ensemble de 50 modèles qui reproduit les 5 courbes de diffusion (les quatre courbes de SANS et la courbe de SAXS) et des modèles représentatifs de l'ensemble sélectionné. Les conformères sont présentés en superposant le domaine P_{CD} de chaque conformère. Cette représentation montre que les molécules de N occupent des positions extrêmement diverses autour la P en accord avec la grande flexibilité de la P.

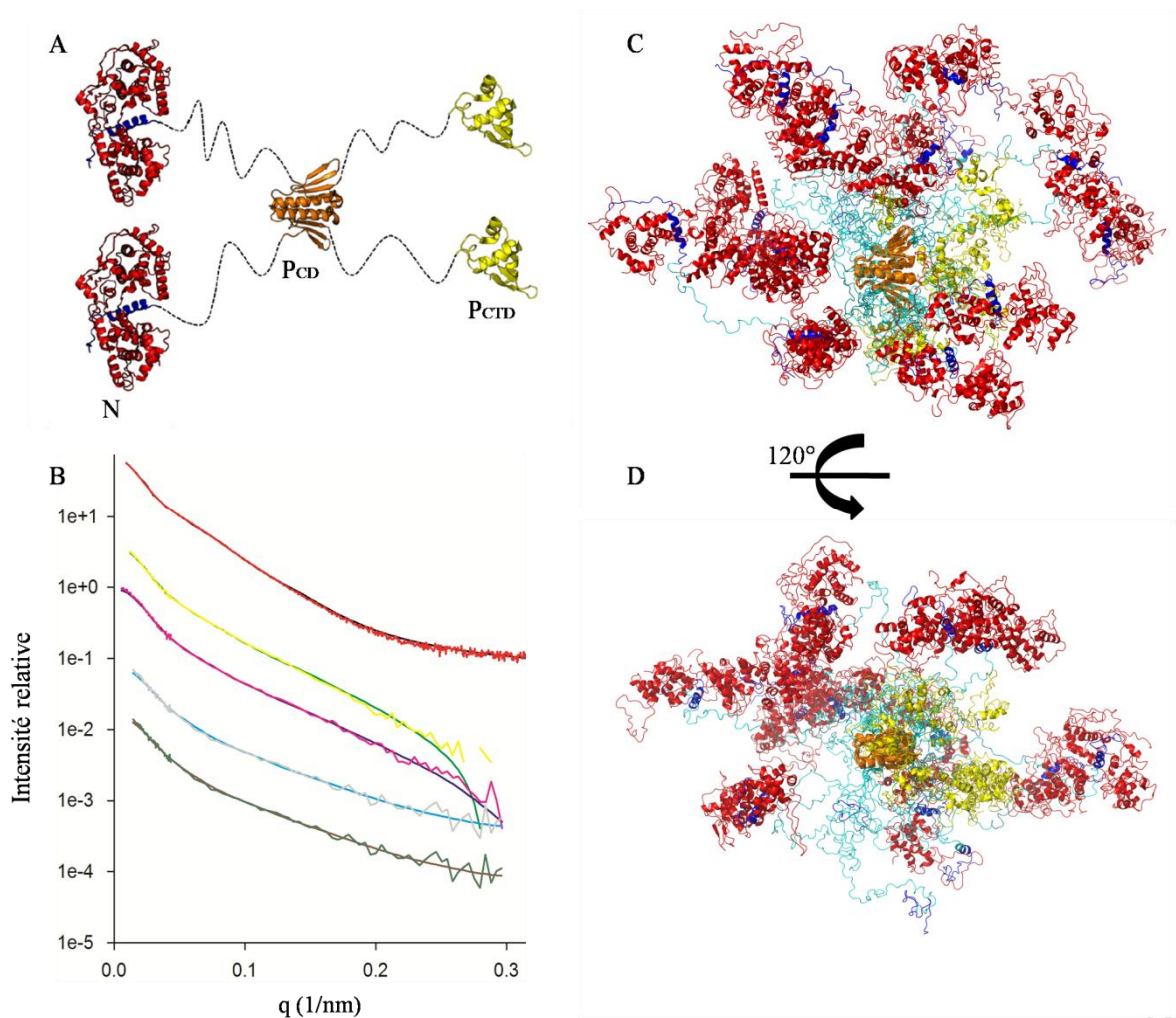


Figure 8 : Modélisation du complexe $N_{\Delta 21}^0$ -P entière. **A)** Schéma illustrant l'organisation d'un complexe N^0 -P de type 2N:P. En jaune - le domaine P_{CTD} de fixation sur N-ARN ; en orange - le domaine P_{CD} de dimérisation ; en bleu - l'hélice du P_{NTR} qui forme le complexe avec N^0 ; en rouge - N^0 . **B)** Courbes expérimentales (rouge – courbe de SAXS ; jaune, violet, gris, bleu foncé – les courbes de SANS) et courbes théoriques (ligne ajustée pour chaque courbe expérimentale) de l'ensemble de modèles sélectionné pour le complexe $N_{\Delta 21}^0$ -P entière. **C)** et **D)** Six modèles représentatifs de l'ensemble sélectionné, superposés sur le P_{CD} et colorés comme en A). Les régions désordonnées sont colorées en turquoise.

5.3.5. Discussion

Les protéines P₁₋₁₇₇ et P entière ont des volumes d'éluion proches et la même tendance est observée pour les volumes d'éluion des complexes $N_{\Delta 21}^0$ -P₁₋₁₇₇ et $N_{\Delta 21}^0$ -P entière, suggérant que la partie C-terminale de la protéine contribue peu et d'une façon similaire au comportement hydrodynamique des phosphoprotéines isolées et dans le contexte du complexe N^0 -P (**Figure 3**).

Nous avons étudié les complexes $N_{\Delta 21}^0$ -P₁₋₁₇₇ et $N_{\Delta 21}^0$ -P entier en solution par SAXS. Les R_g obtenus en utilisant l'approximation de Guinier pour les deux complexes (6.8 ± 0.2 pour $N_{\Delta 21}^0$ -P₁₋₁₇₇ et 7.1 ± 0.2 pour $N_{\Delta 21}^0$ -P entière) sont très proches ce qui confirme que la

partie C-terminale contribue peu à la dimension globale de la protéine. Les plots de Kratky pour les deux complexes sont plutôt caractéristiques des protéines bien structurées et cette observation est confirmée par la présence du plateau dans les plots de Porod-Debye contrairement à ce qui est observé pour la P₁₋₁₇₇ seule (**Figure 8**).

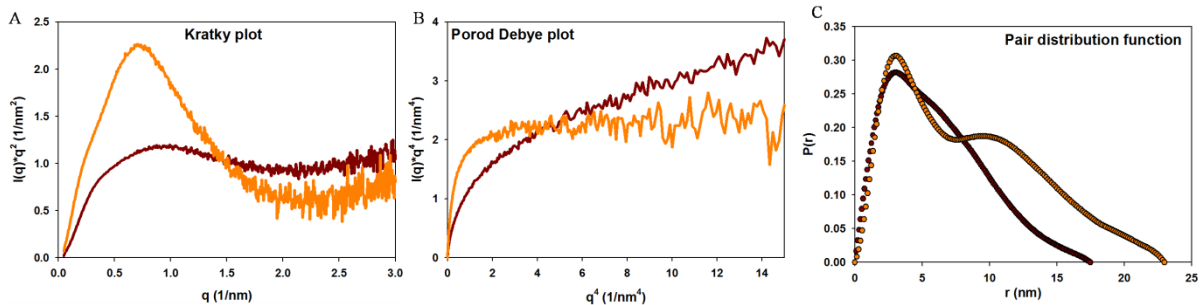


Figure 8 : Comparaison entre le fragment de la phosphoprotéine P₁₋₁₇₇ et le complexe N_{Δ21}⁰-P₁₋₁₇₇ du virus de la stomatite vésiculaire (VSV). Représentations de A) Kratky, B) Porod-Debye et C) Fonction de distribution pour P₁₋₁₇₇ (marron) et N_{Δ21}⁰-P₁₋₁₇₇ (orange).

Cela peut être dû à la présence de la N qui est une protéine globulaire d'environ 45 kDa. Une hypothèse serait que la phosphoprotéine se structure dans le complexe N⁰-P. Néanmoins dans ce dernier cas on attendrait que la protéine se compacte et donc que le R_g diminue par rapport au R_g de la P entière ou de la P₁₋₁₇₇ seules, en contradiction avec l'observation d'une augmentation du R_g. L'augmentation du R_g peut être expliquée en supposant que les nucléoprotéines fixées aux P_{NTR} échantillonnent un large volume d'espace, indépendamment du reste du complexe. Cette hypothèse est renforcée par l'observation que, à part les résidus qui interagissent directement avec la N, la majorité des résonances dans les spectres 2D [¹H-¹⁵N]-HSQC de la P₁₀₅ seule et de la P₁₋₁₇₇ du complexe N_{Δ21}⁰-P₁₋₁₇₇ sont parfaitement superposables (**Figure 4A et B**). D'ailleurs la superposition de ces résonances confirme que le comportement des P_{NTR} n'est pas influencé par le degré d'oligomérisation de la protéine.

La courbe de diffusion de SANS à 0% D₂O provient du complexe entier et son R_g de 7.0 ± 0.2 nm est en très bon accord avec la valeur de 7.1 ± 0.2 nm obtenue à partir de la courbe de SAXS. A 40% D₂O la diffusion provient essentiellement de la P et le R_g de 5.6 ± 0.2 nm est en bon accord avec la valeur de 5.3 ± 0.2 nm déterminée précédemment (Gérard, Ribeiro et al. 2007) et dans les travaux de (Leyrat, Schneider et al. 2012) ce qui exclut la compaction de la P et démontre que dans le complexe N⁰-P la P reste flexible. A 100% D₂O la diffusion provient de la N et le R_g de 4.4 nm déterminé à partir de la courbe est plus grand que le R_g de 2.8 calculé à partir de la structure cristalline ce qui peut être causé par des corrélations spatiales entre les 2N dans les molécules de 2N:2P. L'ensemble de modèles

sélectionnés sur la base des données de SANS et SAXS renforce l'hypothèse que la N du complexe N^0 -P échantillonne un grand volume de l'espace. Le complexe $N_{\Delta 21}^0$ -P entière se fixe sur les nucléocapsides virales (Yabukarski et Ivanov, résultats non publiés) et sa grande flexibilité peut faciliter son positionnement correct pour le transfert du N0 et l'encapsidation d'ARN pendant la réplication virale.

Nous n'avons pas encore analysé l'ensemble(s) sélectionné(s) en détail et ce travail est actuellement en cours. Après une première analyse, les domaines structurés de la P et les N^0 attachées ne semblent pas occuper des positions corrélées dans l'espace mais des analyses corrélant le positionnement des domaines C-terminaux ou les N^0 par rapport au domaine central P_{CD} doivent nous apporter plus d'information.

Les résultats des travaux présentés ici démontrent également que la stœchiométrie du complexe N^0 -P n'est pas fixe et dépend des proportions relatives de N et P. Cela peut avoir des conséquences pour la multiplication virale. Pendant les stades précoces de l'infection la concentration de P est plus importante que celle de la N du au fait qu'environ 450 molécules de P sont déjà présentes dans la particule virale tandis que la N n'est pas encore synthétisée (Thomas, Newcomb et al. 1985). Dans ces conditions, le complexe N^0 -P, formé juste après la synthèse de N, sera probablement de la forme 1N:P. Dans les phases plus tardives, lorsque la réplication et la transcription se passent en majorité dans les corps d'inclusion (Lahaye, Vidy et al. 2009; Heinrich, Cureton et al. 2010), caractéristiques d'une infection par des rhabdovirus et concentrés en protéines (appelés corps de Negri chez RaV) dans le cytoplasme de la cellule, la concentration de la N est probablement supérieure à celle de la P (Barr, Whelan et al. 2002). Ainsi le complexe N^0 -P contiendrait majoritairement 2N. La présence de 1N ou 2N par dimère de P peut également influencer la fixation de la grande sous unité L de la polymérase virale sur ce complexe car son site de fixation se trouve également dans la région N-terminale de la P (Chenik, Schnell et al. 1998; Rahmeh, Morin et al. 2012).

5.4. Article: Crystal structure of the N⁰-P complex of Nipah virus

Authors: **Yabukarski F**, Tarbouriech N, Bourhis JM, Jensen M, Blackledge M, Ruigrok R, Jamin M

In preparation

Crystal structure of the N⁰-P complex of Nipah virus

**Filip Yabukarksi, Nicolas Tarbouriech, Jean-Marie Bourhis, Malene Ringjobing Jensen,
Martin Blackledge, Rob W.H. Ruigrok
and Marc Jamin**

¹UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions
6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

²Institut Laue Langevin, 6 Rue Jules Horowitz 38042 Grenoble Cedex 9, France.

Running title: Structure of NiV N⁰-P

Keywords:

* **Corresponding author:** Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France

E-mail : jamin@embl.fr, Phone: + 33 4 76 20 94 62, Fax : + 33 4 76 20 94 00

Nipah virus (NiV) is an emergent zoonotic Paramyxovirus responsible of recent outbreaks of human encephalitis with high fatality rate in south-east Asia ¹. No vaccine or specific treatment is currently available, and therefore, NiV requires biosecurity level 4 (*BSL-4*) handling. NiV is a nonsegmented negative-strand RNA virus of the *Paramyxovirinae* subfamily; the *Paramyxovirinae* is grouped with the *Pneumovirinae* subfamily in the *Paramyxoviridae* family, which is classified within the order *Mononegavirales* (*MNV*), with the *Rhabdoviridae*, *Filoviridae* and *Bornaviridae* families.

As with other *MNVs*, the genomic RNA of NiV is condensed by a homopolymer of viral nucleoprotein (N) into a helical ribonucleoprotein complex, named nucleocapsid (NC). This complex associates with the viral polymerase through the phosphoprotein cofactor and serves as the template for both RNA transcription and replication ^{2,3}. Viral replication thus requires the continuous production of N in a form that is competent for encapsidating newly synthesized positive antigenomic and negative genomic RNAs ⁴. RNA-free N⁰ is thought to be delivered to nascent RNA in the form of a complex with the viral phosphoprotein (P), named the N⁰-P complex ⁵. In this essential complex, the N-terminal region of P acts as a chaperon, preventing both polymerization of N and encapsidation of cellular RNAs ⁵.

The nucleoprotein from other *MNV* comprises a homologous core (N_{CORE}) responsible for RNA encapsidation, which is made of two globular helical domains, N-terminal (N_{NTD}) and C-terminal (N_{CTD}), connected through a hinge region that define a basic groove and grab the RNA molecule ⁶⁻⁸. The N homopolymer is stabilized by lateral contacts between C-terminal domains, by an N-terminal extension or “arm” docking to the back of the neighboring N_{i+1} protomer, and by a C-terminal loop or extension docking to the back or top of the N_{i-1} protomer ^{7,9,10}. N of *Paramyxovirinae* contains an additional long flexible C-terminal extension (N_{TAIL}), which escapes from the interior of the helical NC between the helical turns ¹¹ and interacts with the phosphoprotein. Using a strategy similar to that used recently for reconstituting a mimic of the N⁰-P of VSV, we produced mimics of NiV N⁰-P and we characterize the structure of these complexes in solution and in crystal.

Reconstitution and properties of the N⁰-P complex

In order to reconstitute a structural analog of the N⁰-P complex and characterize its structure, we produced the fragment N₃₂₋₃₈₃ of the NiV protein deleted of its 31 aa N-terminal arm and C-terminal tail in fusion with the maltose binding protein (MBP). The armless N-MBP fusion expressed in bacteria was bound to RNA, but addition of peptides encompassing the N⁰-binding region of P induced the displacement of the RNA and the formation of soluble analogs of the N⁰-P complex. After removal of the MBP tag by cleavage with the TEV protease, the armless N₃₂₋₃₈₃⁰ co-eluted from the size exclusion chromatography column with P₁₋₅₀ (**Fig. 1b**). On-line multi-angle laser light scattering (MALLS) (molecular mass of 46 ± 1 kDa) and small-angle X-ray scattering (SAXS) showed that the N₃₂₋₃₈₃⁰-P₁₋₅₀ complex is a compact heterodimer (**Fig. 1c**) with an overall bean-shape typical of MNVN protein (**Fig. 1d**). In the nuclear magnetic resonance spectrum (NMR) of the N₃₂₋₃₈₃⁰-P₁₋₁₀₀ complex (**Fig. 1e**), the amide proton resonances of the 50 N-terminal residues are broadened beyond detection (**Fig. 1f**), showing that the binding site of P does not extend downstream of residue 50.

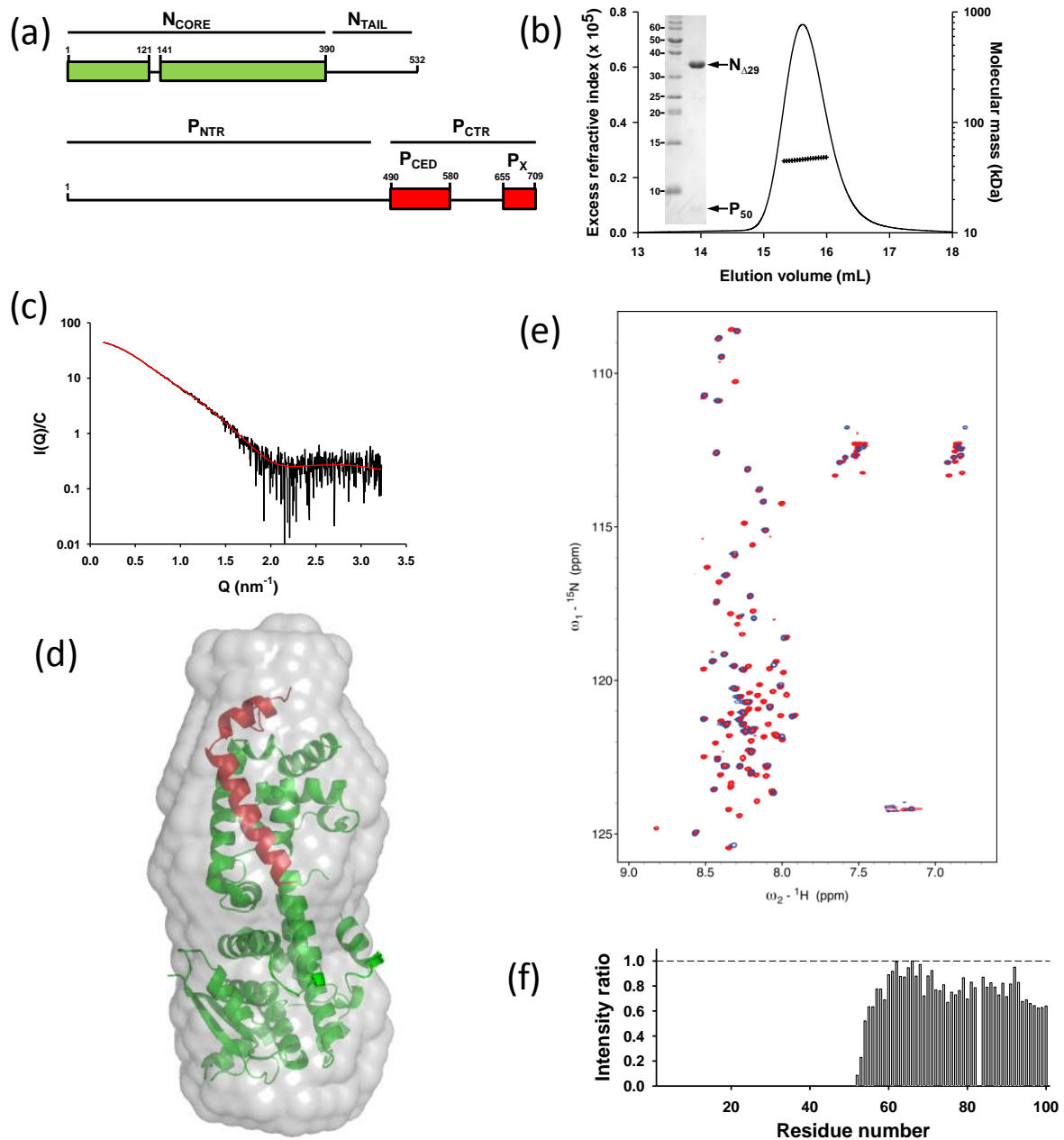


Figure 1. Heterodimeric NiV N⁰-P₁₋₅₀ complex of Nipah Virus. (a) Modular architecture of the NiV N and P. Lines show disordered regions and boxes show structured domains. (b) SEC-MALLS-RI. The N₃₂₋₃₈₃⁰-P₁₋₅₀ complex was injected on a Superdex S200 column in 20 mM Tris-HCl at pH 7.5 containing 150 mM NaCl. The line shows the chromatogram monitored by differential refractive index. The crosses show the molecular mass calculated from the intensity of scattered light. (c) Small-angle X-ray scattering curve of the complex. The red line shows the fit for the bead model shown in figure 1d. (d) Ab initio bead model calculated from the SAXS curve. The crystal structure of one protomer of the N⁰-P complex is docked in the model. (e) ¹H, ¹⁵N-HSQC NMR spectrum of the N⁰-P complex. The red spectrum is for isolated ¹⁵N-labelled P₁₋₁₀₀ and the blue spectrum is for the labeled peptide bound to N⁰. (f) Localization of the N⁰-binding site on P. The plot shows the intensity ratio of the NMR resonances of the P₁₋₁₀₀ peptide in its bound and free form.

Crystal structure

The structure of the NiV N₃₂₋₃₈₃⁰-P₁₋₅₀ complex was determined by the X-ray crystallography at 2.5 Å resolution and refined to an R_{factor} of 19.9% and an R_{free} of 25.9% (**Fig. 2, Table 1**). The asymmetric unit of the crystal contains three copies of the complex (**Fig. 3a**); two copies are well defined in the electron density map, while the N_{NTD} of the third copy is less densely packed in the crystal and ill-defined. (**Fig. 3c**) Structural overlay of the three copies of N present in the crystal asymmetric unit revealed slight differences in the position of the helices, suggesting that this region of N is dynamic at least in the absence of RNA (**Fig. 3b**). As with other *MNV*, the core of NiV N (N_{CORE}) comprises two globular helical domains, N-terminal (N_{NTD}) and C-terminal (N_{CTD}), connected through a hinge region that define a basic groove and grab the RNA molecule⁶⁻⁸ (**Fig. 4**). N_{NTD} comprises ten α -helices, and a two-strand parallel β -sheet, but part of the α_{N6} - α_{N7} loop is not visible in the crystal electron density. N_{CTD} comprises four α -helices, which pack like in RSV and rhabdovirus N^{7,9,10} (**Fig. 4a**).

Residues 1 to 38 of P₁₋₅₀ are visible in the three copies of the complex within the asymmetric crystal unit (**Fig. 2c and 4a**). They form a 2.9 nm-long helix (residues 1-19) that docks on the back of N_{CTD} with a 90° bend at Asn20 leading to a short helix (residues 21-28) and a loop that dock on the top of N_{CTD} (**Fig. 4**). Extensive hydrophobic contacts and eleven hydrogen bonds are made between P₁₋₅₀ and N_{CTD} (**Table 2**), with a total surface area buried in the interaction of 1260 Å². In N, the interactions with P₁₋₅₀ mainly involves helices α_{C1} , η_{C1} and α_{C3} (aa 265-305), which delineate a shallow hydrophobic groove on one side of N_{CTD}, and with helix α_{C4} at the top of N_{CTD}. The key N binding residues in P are Lys3, (Asp4), Asn8, Asp9, Gln21, Gln25, Ser31, Ser32 and Ile33, which form hydrogen bonds and salt bridges with N. Additionally, the hydrophobic residues Leu6, Val7, Leu11, Ile13, Ile14, Ile17, Ile24, Tyr28 and Ile33 involved in the interactions appear to be remotely conserved among most members of the *Paramyxoviridae*¹² (**Fig 5a**). Mapping N residue conservation across *Paramyxovirinae* onto the surface of the N protein revealed that the P binding site is also rather well conserved, suggesting a conserved structural architecture for the N⁰-P complex through the entire subfamily (**Fig 5b**).

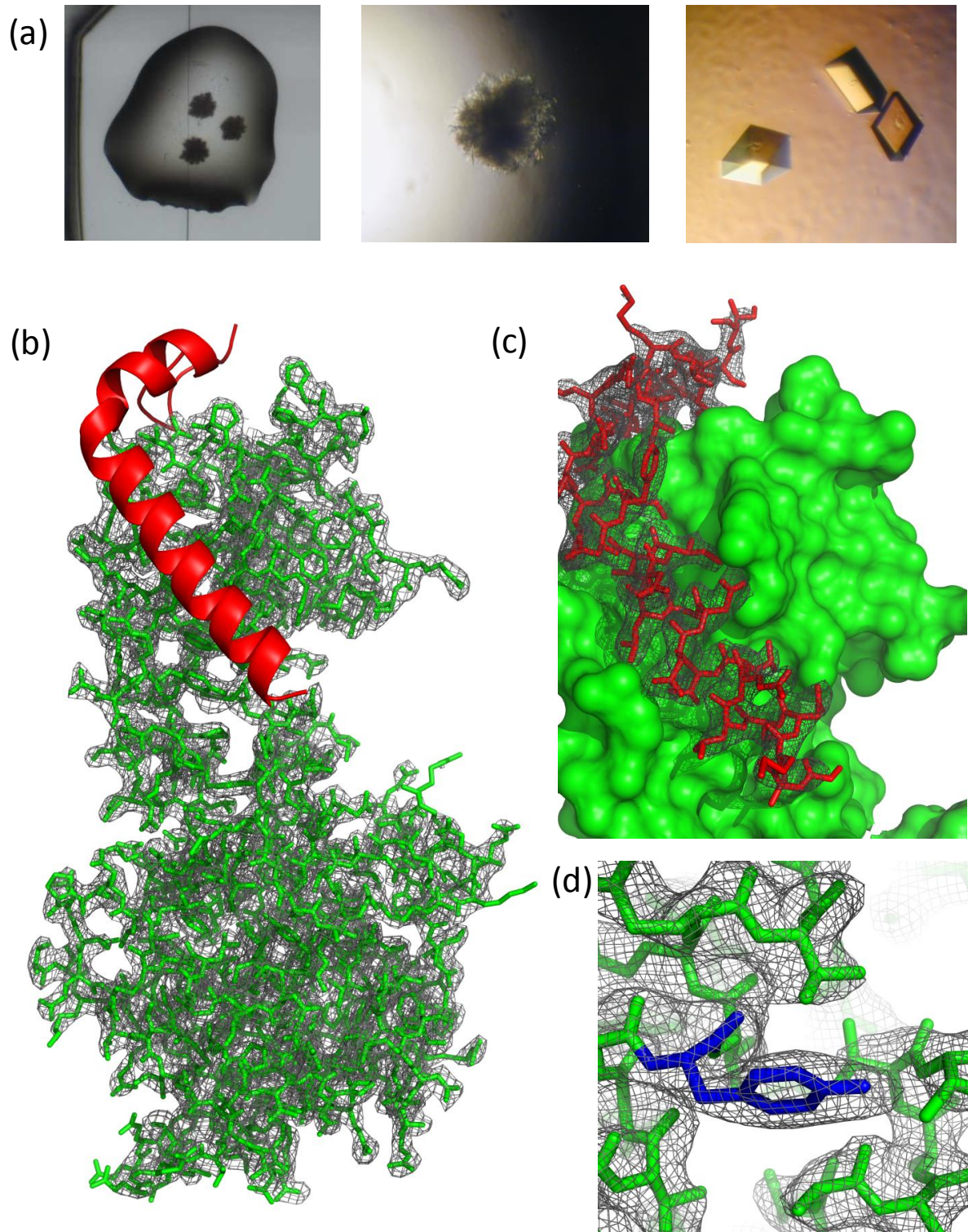


Figure 2. Crystal structure of the N⁰-P complex. (a) Crystals. Initial hit after conditions screening using a Cartesian robot from HTXLab (left), manually reproduced plate cluster using native complex (middle) and crystals obtained after microseeding (b) **The 2.5 Å resolution electron density map of N within the N⁰-P complex.** The Fo-Fc map is contoured at 1.5 σ . (c) **The 2.5 Å resolution electron density map of P within the N⁰-P complex.** (d) **Detailed view of Tyr258.**

Table 1. Data collection and refinement statistics

	Crystal 1
Data collection	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions $\square \square$	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	82.89, 98.96, 156.91
α , β , γ (°)	90.0, 90.0, 90.0
Resolution (Å) ^a	47.19 -2.498 (2.587-2.498)
<i>R</i> _{sym} ^a	%(%)
<i>I</i> / σI ^a	14.44 (2.59)
Completeness (%) ^a	99.56 (97.28)
Redundancy	3.1 (2.5)
Refinement	
Resolution (Å)	47.189-2.498 (2.553-2.498)
No. unique reflections	45321
<i>R</i> _{work} / <i>R</i> _{free}	19.87% (24.02%) / 25.94% (32.00%)
No. atoms	
Macromolecule	7622
Ligand	9
Water	105
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.21
Ramachandran favored	96 %
Ramachandran outliers	1.1 %
Average B-factor	56.50
Macromolecule	56.70
Ligand	44.40
Water	41.40

^aValues in parentheses are for highest-resolution shell.

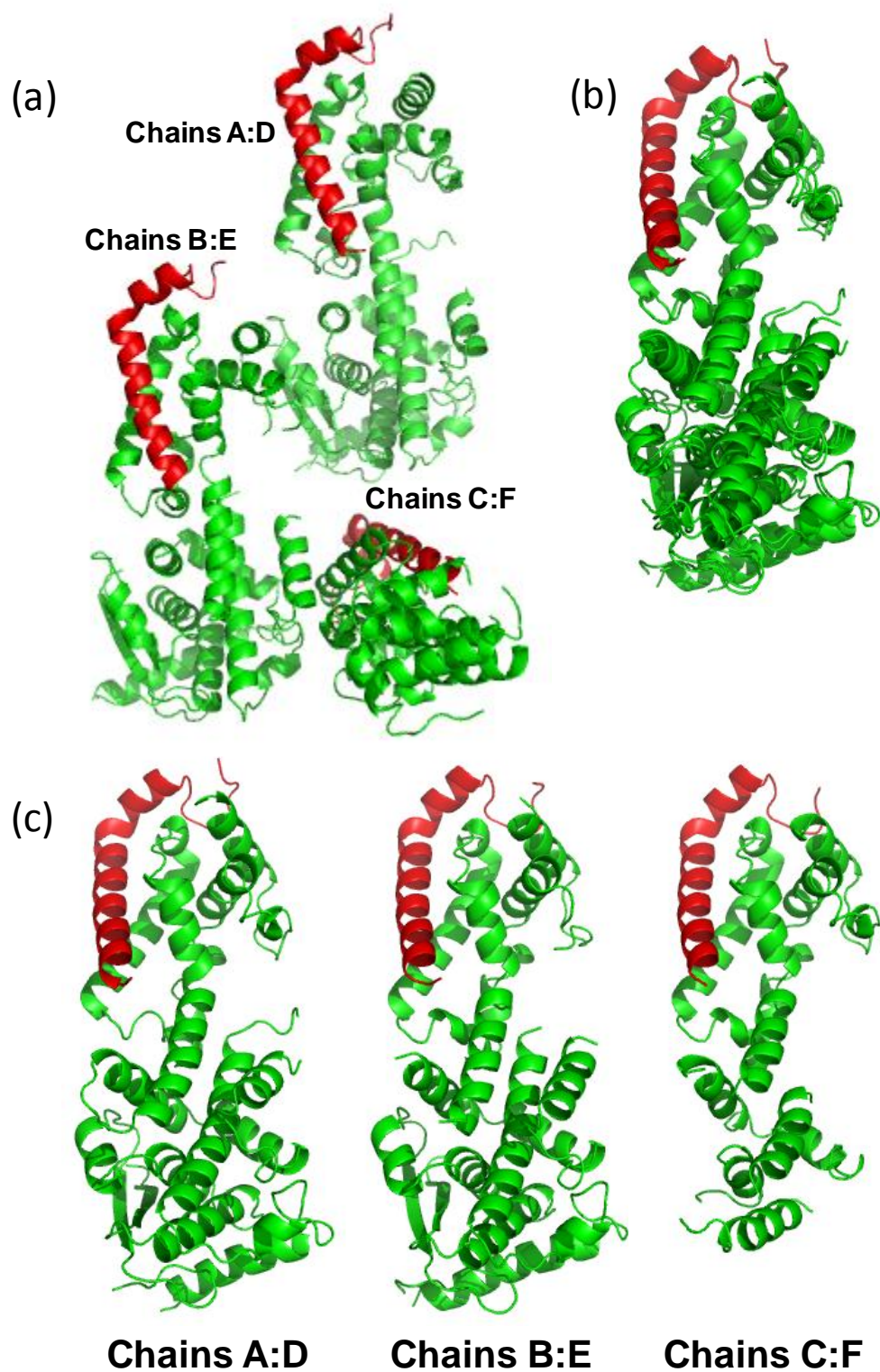


Figure 3. Asymmetric unit of the N^0 -P complex crystal. (a) Ribbon representation of the three copies of the N^0 -P complex in the asymmetric unit of the crystal. N is shown in green and the peptide of P is shown in red. (b) Superposition of the three copies. (c) Comparison of the three copies of the N^0 -P complex.

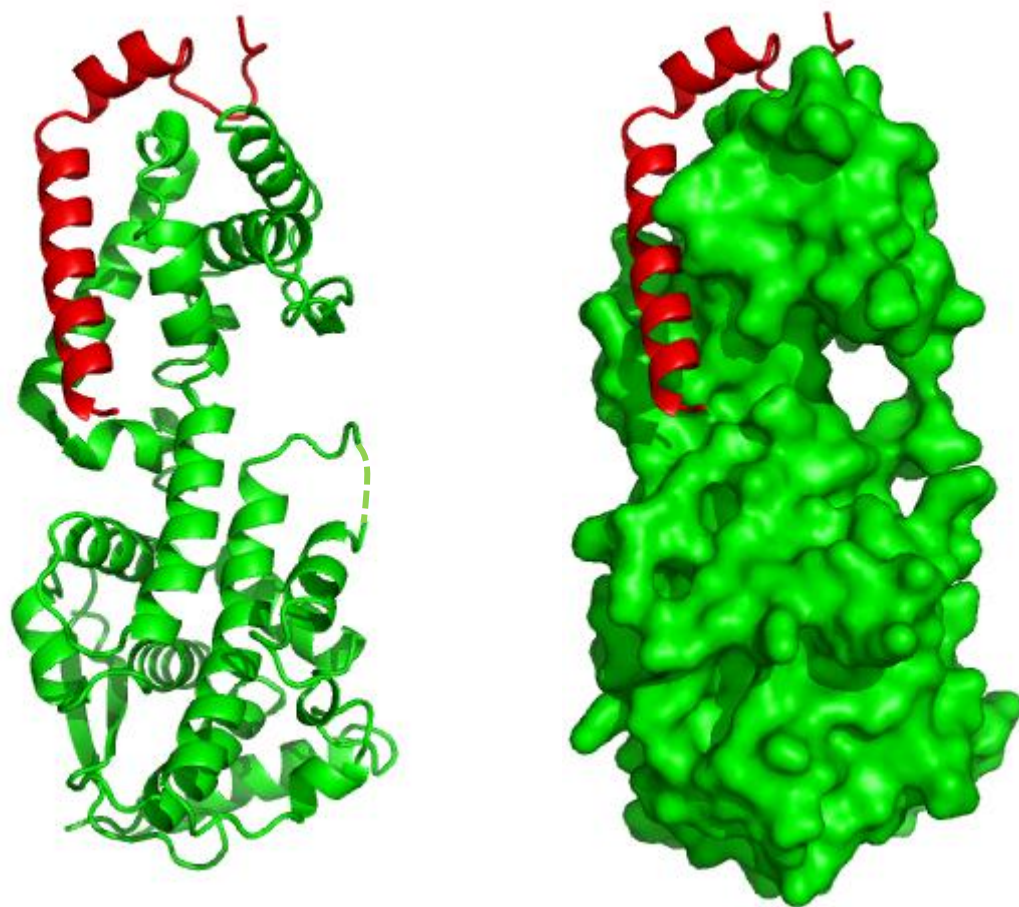


Figure 4. Structure of the N^0 -P complex. (a) Ribbon representation of the N^0 -P complex. (b) Surface representation of the N^0 -P complex.

Table 2. The binding interactions between the residues of the P₁₋₅₀ and the residues of the N₃₂₋₃₈₃ protein binding site

Interaction	P residues	Atoms	N residues	Atoms
H bond	Lys3	NZ	Glu260	OE2
H bond	Asn8	OD1	Arg302	NH1
H bond	Asp9	OD1	Arg278	NH2
H bond	Asp9	OD2	Arg278	NH2
H bond	Gln21	NE2	Asn285	OD1
H bond	Gln21	NE2	Ala283	O
H bond	Gln25	NE2	Asn285	OD1
H bond	Gln25	NE2	Ala283	O
H bond	Ser31	O	Asn285	N
H bond	Ser32	OG	Glu286	OE2
H bond	Ile33	N	Glu286	OE1

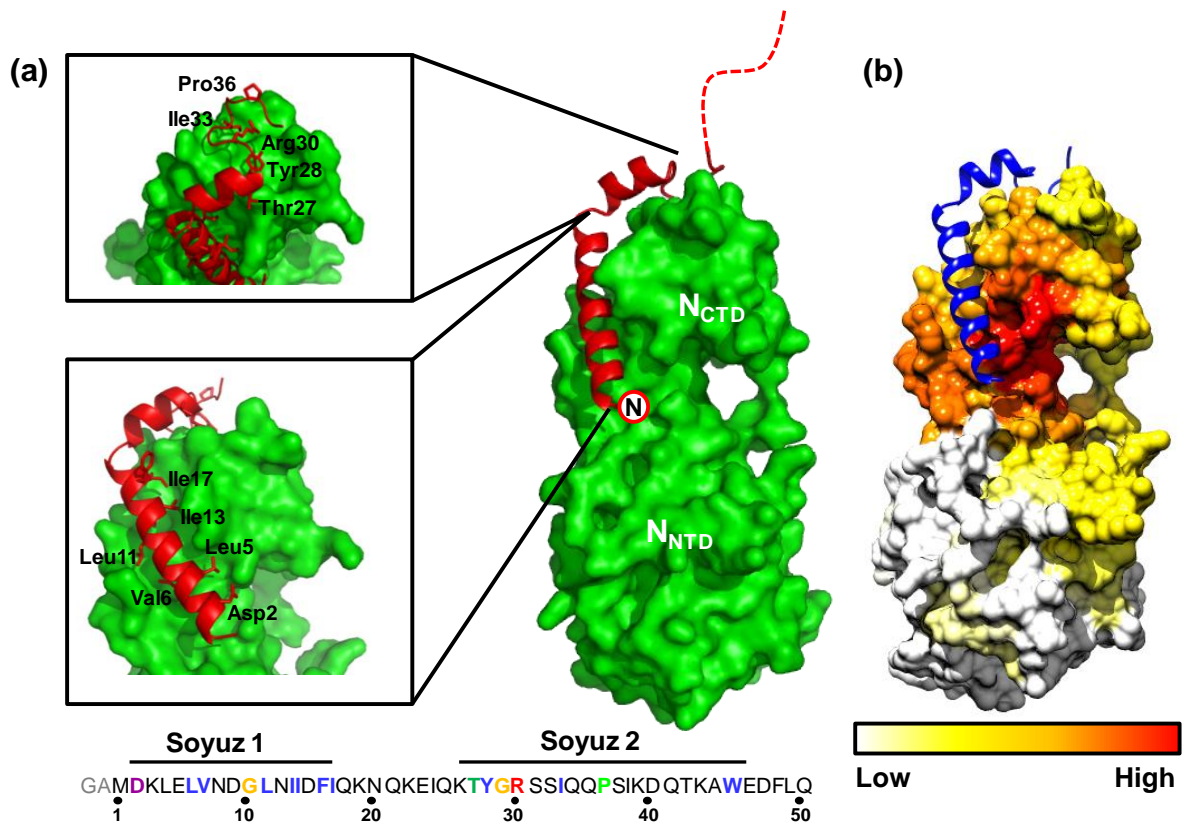


Figure 5. Amino acid sequence conservation among *Paramyxovirinae*. (a) **Conservation in P.** Several amino acids in the N-terminal region of P have been found to be conserved among Paramyxovirinae¹². These amino acids are shown in color at the bottom. They are labeled in the left side figures. (b) **Conservation of N.** The conservation is map on the surface of N.

Comparison with RSV N-RNA and RNA binding

By homology with the structure of the recombinant circular RSV N-RNA complex⁷ (**Fig. 6a**), the RNA binding site in NiV N was localized in the groove at the interface between N_{NTD} and N_{CTD} (**Fig. 6b**). The interior of the RNA binding cavity is also well-conserved among *Paramyxovirinae* (**Fig. 6c**), with large positively charged patches on both N_{NTD} and N_{CTD}. Four (Lys178, Arg192, Arg193 and Arg352) out of the five basic residues interacting with RNA in RSV N (Lys170, Arg184, Arg185, Arg338 and Arg342) are conserved among *Paramyxoviridae*. However, the 3D superposition of NiV N_{CTD} (residues 264-371) in the N⁰₃₂₋₂₈₃-P₁₋₅₀ complex onto the corresponding domain of an RSV N protomer in the recombinant circular N-RNA complex (rmsd 2.25 Å), revealed that the RNA binding groove is opened in NiV, with N_{NTD} bowing down by 30° from N_{CTD} as compared to their position in the RSV structure (**Fig 6d**). Conserved basic residues in N_{NTD} (Lys, Arg, Arg) and N_{CTD} (Arg) are too far apart for accommodating the RNA molecule, suggesting that conformational changes must occur for binding genomic RNA.

N_{NTD} can be divided into three subdomains, which are structurally similar in NiV and RSV but adopt different relative orientations in the two proteins (**Fig 6a and 6b**). Independently moving N_{NTD1} (residues 32-115), encompassing helices α_{N1} to α_{N3} , and N_{NTD3} (residues 158-263), encompassing helices α_{N6} to α_{N10} , allows superposition of NiV and RSV subdomains (with r.m.s.d. of 2.56 and XX Å, respectively). Structural overlay of N_{NTD3} brings Lys178, Arg192 and Arg193 of NiV close to their position in RSV N. RNA encapsidation by NiV N may thus involve concerted motions of these subdomains in grabbing genomic RNA. HingeProt predicts a hinge movement between N_{CTD} (residues 259-371) and N_{NTD} (residues 1-257) around hinge residue Tyr258, which results in N_{CTD} moving towards N_{NTD} (**Fig. 6d**).

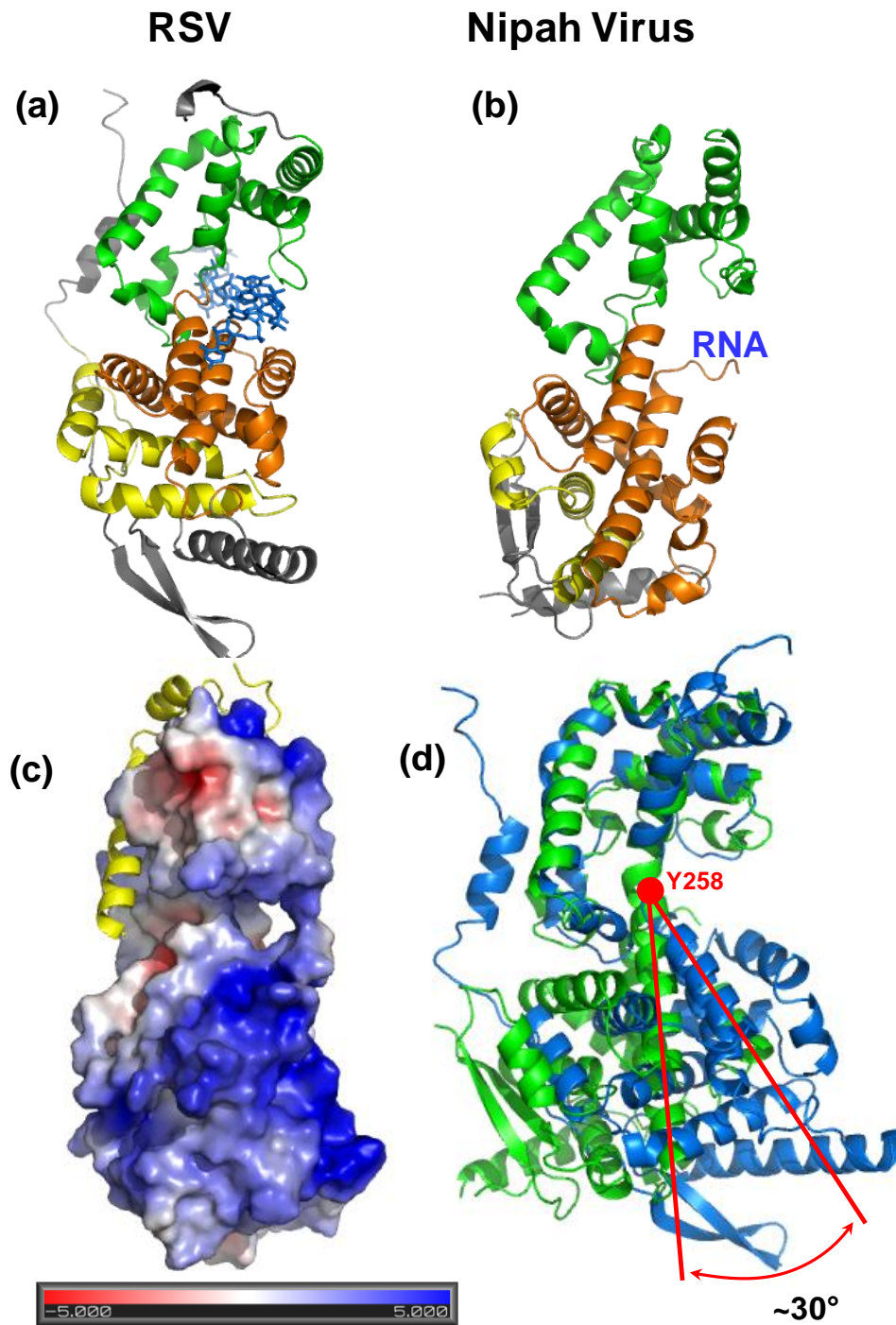


Figure 6. Comparison with respiratory syncytial virus N protein. (a) **Ribbon representation of RSV N.** The N-terminal domain is divided in three sub-domains colored in yellow (N_{NTD1}), gray (N_{NTD2}) and orange (N_{NTD3}). The C-terminal domain is shown in green and the RNA molecule in blue. (b) **Ribbon representation of NiV N.** The corresponding domains and sub-domains are shown in the same color as in figure 6a. (c) **Electrostatic surface map of NiV N.** (d) **Superposition of RSV and NiV N.** The NCTD domains of both were **superposed**. RSV N is shown in blue and NiV N is shown in green.

Chaperone activities of P

As in VSV N⁰-P complex, the N-terminal region of P interferes with the polymerization of N by competing for binding sites with the swapped sub-domains from neighboring N molecules (**Fig. 7a**). The structural overlay of NiV N⁰₃₂₋₂₈₃-P₁₋₅₀ complex onto one protomer of the RSV N-RNA complex revealed that the long N-terminal helix of P₁₋₅₀ occupies the same binding surface than the N-terminal arm of the N_{i-1} protomer on the N_i protomer of the N-RNA complex, while the short helix and loop of P₁₋₅₀ occupy the same binding surface at the top of N⁰ than the C-terminal arm of the N_{i+1} protomer on the N_i protomer (**Fig. 7b**). Binding of P to N thus weakens or prevents interaction with the swapped subdomains from neighboring N molecules.

By contrast with the position occupied by P in VSV N⁰-P complex (**Fig. 8a**), NiV P does not directly interfere with RNA binding to N (**Fig. 8b**). Several structural features of the NiV N⁰-P complex suggest that binding of P blocks the conformational changes required for accommodating the genomic RNA within the RNA binding cavity. Structural overlay of N_{NTD1} of NiV and RSV suggests that, in the closed form of N, Tyr258 at the C-terminus of helix α_{N10} , a conserved residue among *Paramyxoviridae*, points inside the RNA binding cavity (**Fig. 8d**). In RSV N-RNA complex, the side-chain of Tyr251, which occupies a similar position at the C-terminus of helix α_{N10} , docks against helix α_{C2} and does not interfere with RNA binding (**Fig. 8c**). This suggests that Tyr258 of NiV flips away from the RNA binding cavity upon closure of the molecule. Such motion of Tyr258 is prevented in the N⁰-P complex, because the N-terminal extremity of P is bound to the surface of helix α_{C2} . Closure of the RNA binding groove would require the swiveling motion of helices α_{N6} and α_{N10} around pivots at the junction between N_{NTD} and N_{CTD} (**Fig. 9a**). The connecting loop between helices α_{C2} and α_{C3} (3¹² helix η_{C1}) docks against the N-terminal end of α_{N6} and the C-terminal end of helix α_{N10} , and thus interfere with such motions (**Fig. 9a**). Motions of helices α_{N6} and α_{N10} requires motions in N_{CTD}, and binding of P to N_{CTD} bridges helices α_{C1} , α_{C2} and α_{C4} (**Fig. 9b**) rigidifying the whole domain and thus preventing closure of the molecule. In addition, two contiguous glutamate residues in helix α_{N10} (Glu260, Glu261) are H bonded to Lys3 in P providing another link between N_{NTD} and N_{CTD}.

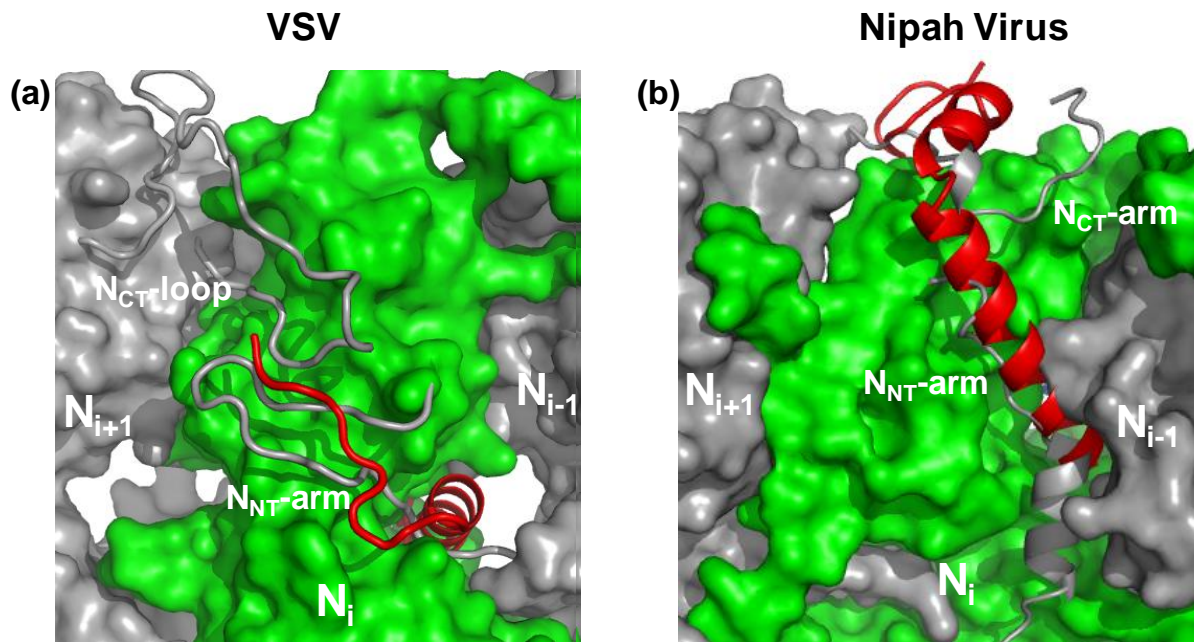


Figure 7. Chaperon activities of P – inhibition of the assembly of N. (a) Overlay of the N-RNA complex and N⁰-P complex of VSV. The N_{i+1}, N_i and N_{i-1} of VSV N-RNA complex ¹⁰ are shown in surface representation. The N_{NT}-arm of the N_{i-1} protomer and the N_{CT}-loop of the N_{i+1} protomer are shown as gray loop. The N molecule of one N⁰-P protomer has been superposed on the N_i protomer of the N-RNA complex and the peptide of P is shown in red. **(b) Overlay of the N-RNA complex of RSV and of the N⁰-P complex of NiV.** The N_{i+1}, N_i and N_{i-1} of RSV N-RNA complex ⁷ are shown in surface representation. The N_{NT}-arm of the N_{i-1} protomer and the N_{CT}-arm of the N_{i+1} protomer are shown as gray loop. The N molecule of one N⁰-P protomer has been superposed on the N_i protomer of the RSV N-RNA complex by superposing the N_{CTD} domains. The peptide of P is shown in red.

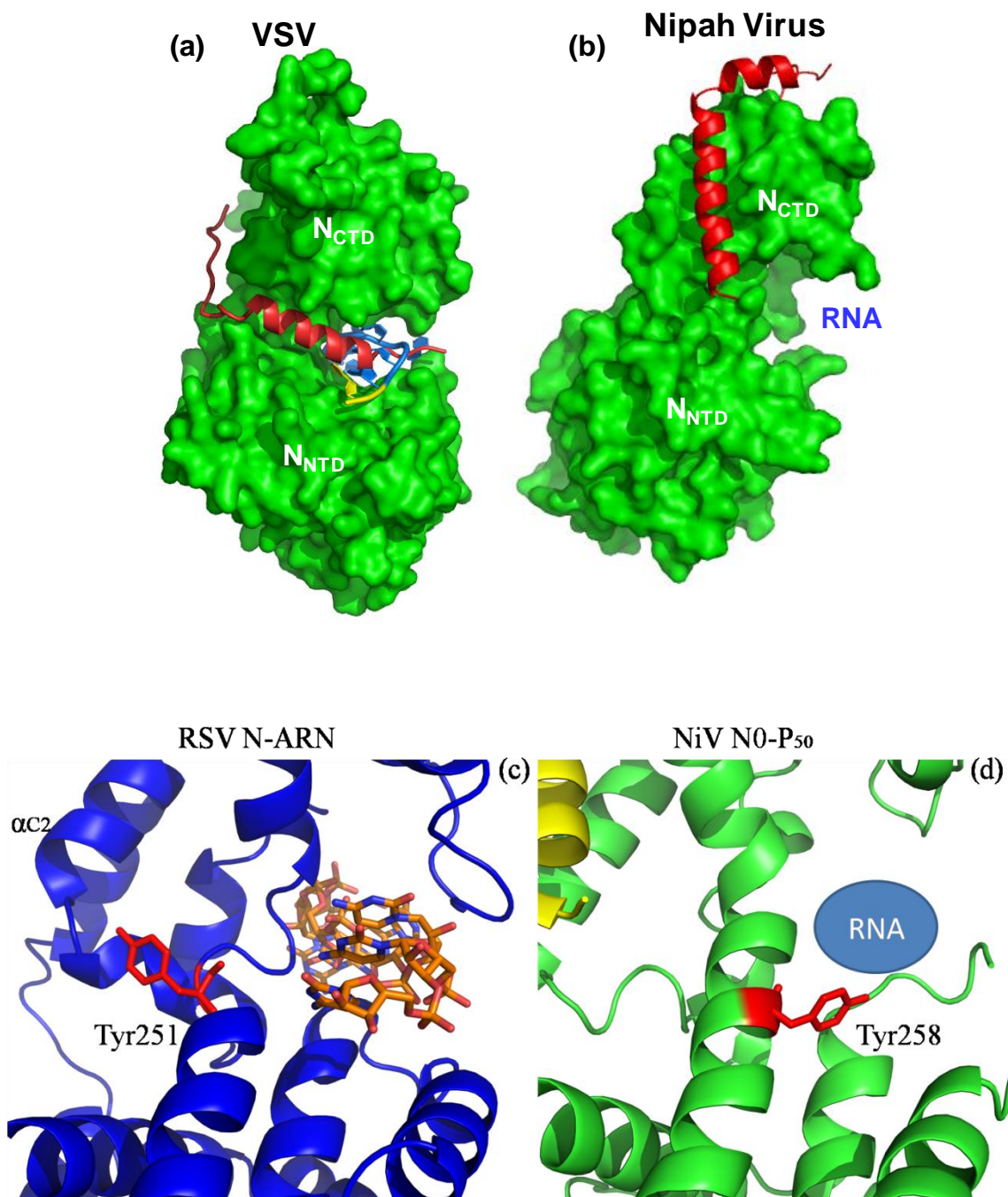


Figure 8. Chaperon activities of P – inhibition of RNA binding. **(a)** Overlay of the N-RNA and N⁰-P complexes of VSV. N is shown in green in surface representation. The peptide of P is shown in red and the RNA is shown in blue and yellow. **(b)** Structure of NiV N⁰-P complex. The N_{CTD} of NiV N has been superposed on N_{CTD} of VSV and NiV N is shown in same orientation as N VSV in figure 8a. **(c)** RSV N-ARN complex. Tyr251, conserved among all *Paramyxoviridae* points towards α C2 and away from the RNA binding cavity. **(d)** NiV N⁰-P complex. In NiV N⁰-P complex conserved Tyr258 points towards the putative RNA binding cavity.

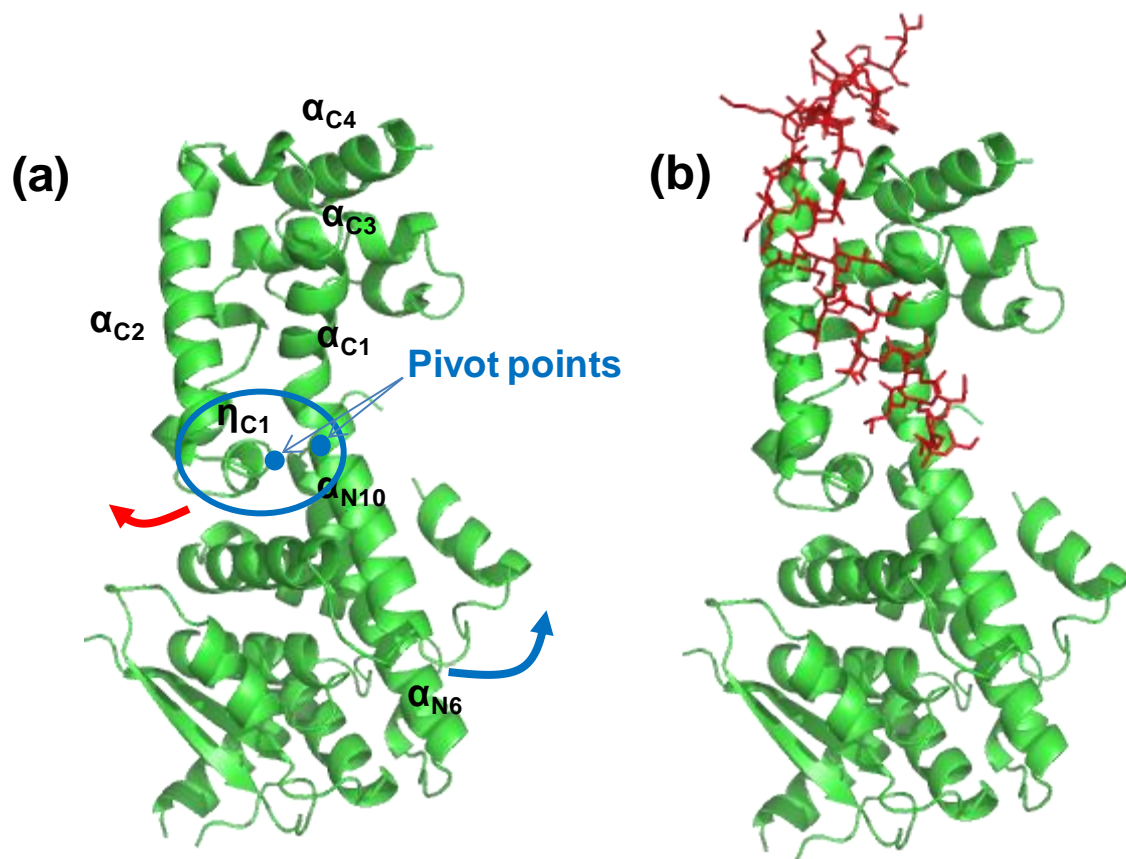


Figure 9. Model of NiV N conformational change. (a) Proposed conformational change of NiV N involved in RNA binding. The blue arrow shows the proposed motions of N_{NTD} around pivot points at the junction between N_{NTD} and N_{CTD} . The red arrows shows the proposed motion of the loop connecting helices α_{C2} and α_{C3} required for allowing motion of N_{NTD} . **(b) Bridging action of P.**

RNA binding to NiV N

In the *Rhabdoviridae* N-RNA complexes, each N binds 9 ribonucleotides (nt), and the RNA is inside the ring, whereas in the *Pneumovirinae* complex, each N binds 7 nt, and the RNA is outside^{7,9,10}. Peculiarly, the genome of the *Paramyxovirinae* obeys a rule of six, a strict requirement for the genome to consist of a number of nt that is a multiple of six^{13,14}. The RNA is also outside the helical nucleocapsid, and each N binds 6 nt⁶.

A motif conserved among all *Paramyxovirinae* provides the structural basis for encapsidation of only six nt by one N molecule that underlines the rule of six, i.e. a strict requirement for the genome to consist of a number of nt that is a multiple of six^{13,14}. In RSV N, helix α N4 which contains Arg184, Arg185 ends at Asn188 leaving the surface of helix α N8 exposed to solvent. Residues in this helix form a flat surface on which base 1 of the RNA repeat binds. In NiV N, helix α N7 propagates further than in RSV, encompassing a motif conserved among all *Paramyxovirinae*, the KYxQQxRV motif. The side chains of these residues dock onto the N-terminal part of helix α N10, hiding the surface of α N4 and therefore preventing the docking of base 1 (**Fig 10**).

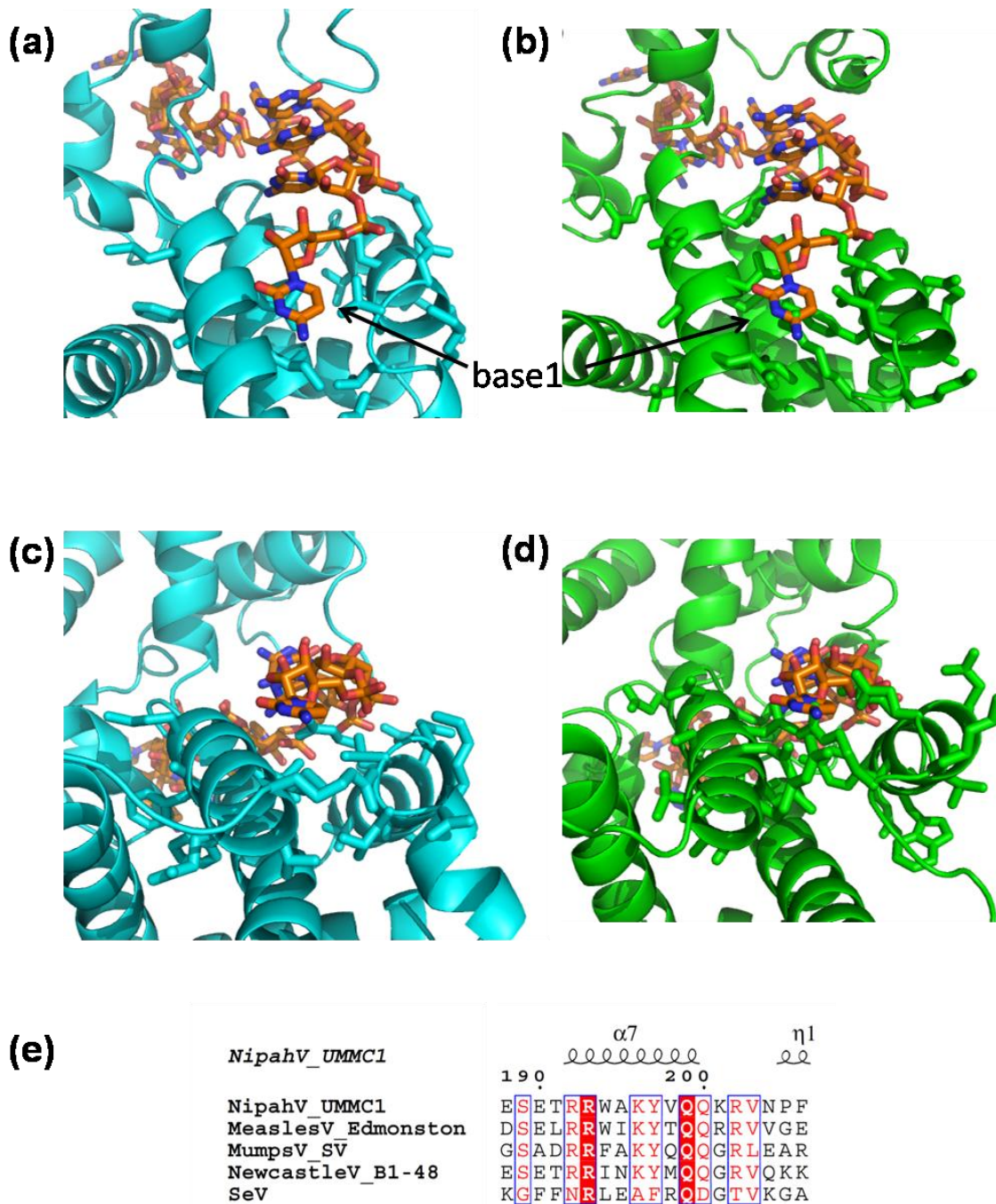


Figure 10. Structural basis for the rule of six. (a) and (c) Stacking of base 1 on α N4 of N RSV. α N4 presents a flat surface on which base 1 can stack. (b) and (d) Closed conformation of NiV N modeled by superposing N_{NTD} and N_{CTD} separately on the corresponding domains of N RSV.

Disorder-to-order transition in the N⁰-binding region of P

NMR spectroscopy, SAXS and gel filtration showed that isolated P₁₋₁₀₀ is globally disordered. The hydrodynamic radius ($R_h = 2.6$ nm) and radius of gyration ($R_g = 3.2 \pm 0.1$ nm) are larger than expected for compact domain of the same size. The poor chemical shift dispersion of amide resonances in the NMR spectra (**Fig 1e**) is typical of disordered proteins, but secondary chemical shifts revealed the presence of transient helices (aa 4-32 and aa 39-56) (**Fig. 11a**). The formation of the N⁰-P complex, thus, involves a disorder-to-order transition of the N⁰-binding region of P, leading to the stabilization of a helical conformation of the first helix. However, in the N⁰-P complex, the helix is broken in two parts (aa 1-19 and aa 21-28). Prediction with AGADIR also indicates propensity for forming a long N-terminal helix (aa 4-25) (**Fig. 11b**). Prediction and experiments show the presence of local propensity at forming a long helix, demonstrating that the kink observed in the crystal structure is induced by tertiary contacts upon docking of P to N⁰. The strain in P resulting from this opposition between local and tertiary constraints may also contribute to keeping N in an open conformation. A similar situation is found for the bridging domain of RIG-I¹⁵, in which a long stretch of amino acids predicted to form a long helix bridges two domains of the protein by forming two helices.

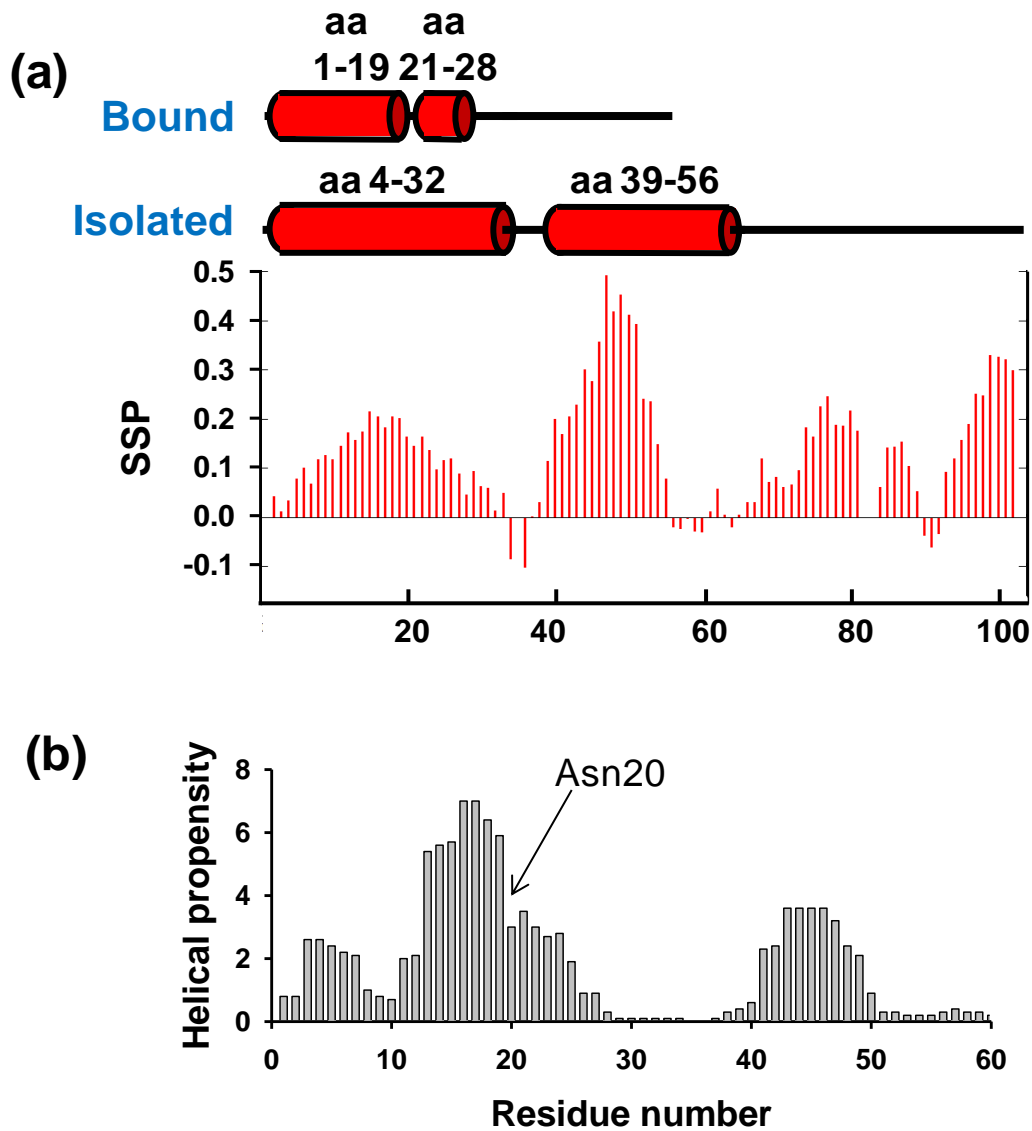


Figure 11. Disorder-to-order transition in P upon formation of the N0-P complex. (a) Secondary structure propensity in isolated P₁₋₁₀₀ calculated from the NMR spectrum. The scheme above the plot shows the position of the helices in the crystal structure (bound) and in isolation in solution calculated from the NMR spectrum. (b) Prediction using the program AGADIR.

In conclusion, the structure of a mimic of NiV N⁰-P complex, reconstituted between the armless N₃₂₋₃₈₃ and the N-terminal P₁₋₅₀, revealed the mechanisms by which P chaperons N during viral replication. P prevents the assembly of N by interfering with the binding of sub-domains from neighboring N molecules. P prevents binding to cellular RNA by blocking the N molecule in an open conformation, suggesting that encapsidation of viral RNA will involve a conformational change leading to the closure of the RNA binding groove. This structure also provides the first atomic structure of N from a member of the sub-family *Paramyxovirinae* and suggests an explanation for the rule of six.

Finally, because the continuous production of the N⁰-P complex is essential for encapsidating newly synthesized genomes, it is hoped that this structure will provide the rational for designing inhibitors of viral replication.

Methods

Cloning, expression and purification

P₁₋₅₀ and N₃₂₋₃₈₃ fused to an N-terminal maltose binding protein (MBP) tag were cloned in petM40 vector. Both proteins were expressed in *E. coli* BL21 (DE3) Rosetta cells. Cells were grown at 37 °C in LB medium until O.D. reached 0.6 and protein expression was induced overnight at 20°C by addition of isopropyl-β-D-thiogalactoside (IPTG) to the final concentration of 1 mM. Cell pellets were resuspended in 20mM Tris-HCl buffer at pH 7.5 containing 150 mM NaCl, 50 mM Arginine, 50 mM Glutamate and 0.5mM tris(2-carboxyethyl)phosphine (TCEP) for P₁₋₅₀ and in Tris-HCl buffer at pH 7.5 containing 150 mM NaCl for N₃₂₋₃₈₃ MBP and supplemented with complete protease inhibitor cocktail (Roche). Cells were disrupted by sonication, and the crude extract was centrifuged at 40000g at 4°C for 20 min. The supernatant was applied to an amylose resin column (New England Biolabs) pre-equilibrated in 20 mM Tris-HCl buffer pH 7.5 containing 150 mM NaCl. Each column is washed with 20mM Tris pH7.5, 500mM NaCl and protein is eluted in buffer CB plus 50mM Maltose (Sigma) and supplemented with complete anti-protease tablets. Protein purity is checked by SDS-PAGE.

For producing P₁₋₅₀ peptide, TEV protease is added to the protein containing fractions at a ratio of 1/100 mass TEV/mass fusion protein. Cleavage is done O/N at 4 °C and then the solution is concentrated in Sartorius concentrators of 3000 Da MWCO. Concentrated protein is injected on s75 Superdex column equilibrated in buffer CB supplemented with R/E 50mM and 0.5mM TCEP at 4°C. Fractions containing P₁₋₅₀ peptide are collected, pooled and concentrated in Sartorius concentrator 3000 Da MWCO.

Purified P₁₋₅₀ peptide was mixed with N₃₂₋₃₈₃-MBP O/N at 4 °C. Solution is concentrated in Sartorius concentrator of 3000 Da MWCO and injected on a S75 column equilibrated in buffer A. Fractions corresponding to the N₃₂₋₃₈₃-MBP⁰-P₁₋₅₀ complex are pooled and TEV protease is added at a ratio of 1/100 mass TEV/mass fusion protein and let stand O/N at 4°C. The solution is concentrated and then injected on a S200 column coupled to a packed amylose resin (Pharmacia) column all equilibrated in 20 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP. Fractions corresponding to the N₃₂₋₃₈₃⁰-P₅₀ complex are pooled and concentrated in Millipore concentrator 10000 Da MWCO at 4 °C. After each purification step the protein purity was checked by SDS-PAGE.

The P₁₋₁₀₀ peptide was clones in pet28 vector with a C-terminal His tag. The protein was expressed in in *E.coli* DE3 BL21 Rosetta. Cells were grown in LB medium at 37 °C until O.D. reached 0.6 and then induced with IPTG at a final concentration of 1 mM. Protein expression is induced O/N at 18°C. Cultures are centrifuged at 5000 rpm at 4°C for 20min and then resuspended in 20mM Tris pH 7.5, 150mM NaCl, R/E 50mM 0.5mM TCEP (buffer A) and supplemented with complete anti-protease tablets. Cells are sonicated and centrifuged at 20000 rpm at 4 °C for 20 mins. Supernatant is then applied to a column containing His Select resin (Sigma) pre-equilibrated in buffer A. The column is washed with 20 mM Tris pH7.5, 500 mM NaCl (buffer B) plus 10 mM Imidazole and protein is eluted in buffer A plus 300mM Imidazole (Sigma). Fractions containing protein are pooled and concentrated in Sartorius concentrators of 5000 Da MWCO. Concentrated protein is injected on S200 Superdex column equilibrated in buffer A at 4 °C. Fractions containing P₁₋₁₀₀ peptide are collected, pooled and concentrated in Sartorius concentrator 5000 Da MWCO.

Niv P₁₋₁₀₀ protein is ¹³C/¹⁵N labeled according to the protocol described in (1). The protein is purified as described above. The last gel filtration step is performed in Superdex 200 column equilibrated with 20 mM Bis-Tris at pH 6.0, 150 mM NaCl, supplemented with 50 mM Arg and 50 mM Glu. N₃₂₋₃₈₃⁰-P₁₋₁₀₀ complex was reconstituted as described above.

SEC–MALLS-RI experiments.

Multi-angle laser light scattering (MALLS) coupled to size exclusion chromatography was performed as follows. Size-exclusion chromatography (SEC) was carried out with an S200 Superdex column (GE Healthcare) equilibrated with 20mMTris-HCl, 150mMNaCl and 1mMb-mercaptoethanol buffer. On-line MALLS detection was performed with a DAWNEOS detector (Wyatt Technology Corp.) using a laser emitting at 690nm and by refractive index measurement using an RI2000 detector (Schambeck SFD). Weight-averaged molar masses (Mw) were calculated using the ASTRA software (Wyatt Technology Corp.).

Small angle X-ray scattering experiments.

For SAXS measurements purified and concentrated protein was injected on S200 column equilibrated in CBRE 0.5mM tcep and middle fraction from the elution peak was used. This sample was used to prepare three protein concentrations and their exact concentration was obtained from SEC-MALLS-RI measurements.

Small angle X-ray scattering (SAXS) data were collected at the BioSAXS beamline (BM29) of the ESRF(http://www.esrf.fr/UsersAndScience/Experiments/MX/About_our_beamlines/BM29). BM29 is a tunable energy beamline for SAXS experiments of biological macromolecule solutions. Samples are exposed using 30-ml volumes loaded into a 1.8 mm quartz capillary mounted in vacuum using an automated robotic system which enables the sample to be moved through the beam during exposure to minimize the effect of radiation damage. The measurement cell is cleaned (using a commercial detergent for quartz capillaries), rinsed (with distilled water) and dried using compressed air after each measurement using the automated system to minimize cross contamination and dilution of the samples. Two-dimensional scattering images were collected using the Pilatus 1M detector at a distance of 1.0m from the sample. Standard data collection time of 5 min was used for all samples split into ten 10-s time frames to assess and remove effects from radiation damage to the samples. Individual time frames are processed automatically and independently by the data collection software (BsxCUBE) developed by the BLISS team at the ESRF, yielding individual radially averaged curves of normalized intensity versus scattering angle ($Q = 4\pi \sin\theta/\lambda$). Time frames are combined, excluding any data points which are affected by aggregation induced by radiation damage, to give the average scattering curve for each measurement. The scattering from the buffer alone was measured before and after each sample measurement and was used for background subtraction using the program PRIMUS from the ATSAS package developed by EMBL Hamburg. A range of protein concentrations (1.3 to 7.1 mg.ml⁻¹) was measured to assess and remove any concentration-dependant inter-particle effects. The theoretical scattering from the different conformations was fitted to the merged scattering curve using the program CRY SOL22.

Crystallization

Initial crystallization conditions for N⁰₃₂₋₃₈₃-P₅₀ complex were identified at the High Throughput Crystallization Laboratory of the EMBL Grenoble Outstation and these conditions gave “sea urchin” type crystal clusters (Fig. 2A left) (<https://htxlab.embl.fr>). These conditions were reproduced manually by the hanging-drop vapor diffusion method at 20 °C with a protein concentration of 10-15 mg.ml⁻¹ in 0.2 M NaI at pH 6.9 in the presence of 22% PEG 3350 and resulted in the formation of plate clusters (Fig. 2A middle). Selenomethionine derivative crystals were obtained by microseeding. Crystals of native protein were manually crashed in stabilizing solution using the Seed Beads (Hampton Research) and then 0.5 µL aliquots of this solution were used to seed freshly prepared drops, containing 1µl of 10-20 mg.ml⁻¹ selenomethionine derivative of the complex and 1µl of precipitant solution (0.2M KBr, 17-20% PEG 3350). Crystal suitable for data collection appeared after 12 hours (Fig. 2A right). Best crystals diffracted at about 2.5Å resolution.

Phasing and structure refinement

Data were collected at the ID-29 beam line at ESRF. Initial phases were obtained using the anomalous signal from Selenium atoms by the SAD technique using the program HKL2MAP. Initial model was obtained by the Autobuild program from the phenix suite and subsequently refined using the phenix.refine program and after visual inspection in Coot.

References

- 1 Chua, K. B. *et al.* Nipah virus: a recently emergent deadly paramyxovirus. *Science* 288, 1432-1435, doi:8529 [pii] (2000).
- 2 Morin, B., Rahmeh, A. A. & Whelan, S. P. Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. *Embo J.* 31, 1320-1329, doi:10.1038/emboj.2011.483 emboj2011483 [pii] (2012).
- 3 Arnheiter, H., Davis, N. L., Wertz, G., Schubert, M. & Lazzarini, R. A. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell* 41, 259-267 (1985).
- 4 Patton, J. T., Davis, N. L. & Wertz, G. W. N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus. *J. Virol.* 49, 303-309 (1984).
- 5 Curran, J., Marq, J. B. & Kolakofsky, D. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J. Virol.* 69, 849-855 (1995).
- 6 Desfosses, A., Goret, G., Estrozi, L. F., Ruigrok, R. W. & Gutsche, I. Nucleoprotein-RNA orientation in the measles virus nucleocapsid by three-dimensional electron microscopy. *J. Virol.*, doi:JVI.01459-10 [pii] 10.1128/JVI.01459-10 (2010).
- 7 Tawar, R. G. *et al.* Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* 326, 1279-1283, doi:326/5957/1279 [pii] 10.1126/science.1177634 (2009).
- 8 Ge, P. *et al.* Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* 327, 689-693 (2010).
- 9 Albertini, A. A. *et al.* Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 313, 360-363 (2006).
- 10 Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 313, 357-360 (2006).
- 11 Jensen, M. R. *et al.* Intrinsic disorder in measles virus nucleocapsids. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9839-9844, doi:1103270108 [pii] 10.1073/pnas.1103270108 (2011).
- 12 Karlin, D. & Belshaw, R. Detecting remote sequence homology in disordered proteins: discovery of conserved motifs in the N-termini of Mononegavirales phosphoproteins. *PLoS One* 7, e31719, doi:10.1371/journal.pone.0031719 PONE-D-11-24019 [pii] (2012).
- 13 Calain, P. & Roux, L. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *J. Virol.* 67, 4822-4830 (1993).
- 14 Halpin, K., Bankamp, B., Harcourt, B. H., Bellini, W. J. & Rota, P. A. Nipah virus conforms to the rule of six in a minigenome replication assay. *J. Gen. Virol.* 85, 701-707 (2004).
- 15 Kowalinski, E. *et al.* Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* 147, 423-435, doi:10.1016/j.cell.2011.09.039 S0092-8674(11)01150-0 [pii] (2011).

Chapitre 6: Discussion

Discussion

I. La phosphoprotéine

L'analyse bioinformatique de la séquence en acides aminés de la phosphoprotéine (P) du virus Nipah nous a révélé la présence de domaines structurés séparés par des régions flexibles (**Chapitre 3, Figure 5**) et il semble que cette organisation modulaire soit conservée chez les *Mononegavirales* (Gérard, Ribeiro et al. 2009). En particulier le graphique du D-SCORE en fonction de la séquence nous a permis de délimiter le domaine d'oligomérisation (P_{CD}) de la P de NiV ainsi que son domaine C-terminal (P_{XD}) d'interaction avec les nucléocapsides (NC). D'autres régions sont également prédites comme ordonnées et notamment la partie N-terminale, entre les résidus 1 et 250, qui contient les sites d'interaction avec N^0 et avec la protéine STAT1 (Ciancanelli, Volchkova et al. 2009), mais également la région de connexion des domaines P_{CD} et P_{XD} .

Il a été déjà montré que la partie N-terminale de la P de NiV est globalement désordonnée (Habchi, Mamelli et al. 2010). Néanmoins, dans l'optique d'obtenir des détails structuraux et de construire des modèles atomiques de la P, nous avons caractérisé cette partie de la protéine à une résolution atomique en utilisant la spectroscopie RMN (en collaboration avec l'équipe du Dr. Martin Blackledge, IBS, Grenoble) qui nous a fourni une estimation quantitative de la présence de structures secondaires fluctuantes (**Chapitre 3, Figure 36**). Il faut noter que les régions prédites comme ordonnées par le D-score dans la partie N-terminale globalement désordonnée de la protéine ont, d'après les données de RMN, une propension à former des hélices ce qui peut suggérer que ces régions se replient lorsqu'elles interagissent avec leurs partenaires viraux ou cellulaires (Dyson and Wright 2002).

Nous avons cristallisé et résolu la structure du domaine P_{CD} correspondant aux résidus 480 à 570. Sa localisation est également confirmée par la forme tétramérique des constructions P_{1-580} et $P_{471-709}$. Ce domaine est composé d'un faisceau de quatre hélices α parallèles et contient également une couronne du côté N-terminal (**Chapitre 3, Figure 26**). Même si les conservations de séquences en acides aminés sont faibles, cette structure est homologue à la structure des domaines correspondant de la P du virus Sendai, du virus de la rougeole et du virus des oreillons (Tarbouriech, Curran et al. 2000; Communie, Crepin et al. 2013; Cox, Green et al. 2013). L'homologie la plus marquée est dans la conservation de la longueur du faisceau surenroulé central d'hélices α d'environ 10 nm ce qui suggère une conservation de fonction.

Nous avons également résolu les structures des domaines P_{XD} des virus Nipah et Hendra. Tandis que la structure de HeV P_{XD} est très similaire aux structures des domaines correspondants des virus Sendai, oreillons et rougeole et est composée d'un faisceau de trois hélices α , NiV P_{XD} n'a pas cristallisé sous la même forme. La structure cristalline de NiV P_{XD} est composée de seulement deux hélices α . Malgré le fait que cette dernière forme cristalline puisse être interprétée comme un biais de cristallisation, les deux domaines présentent des propriétés hydrodynamiques et de diffusion de rayons X aux petits angles différentes suggérant que ces domaines s'échangent entre différentes conformations et qu'en dépit d'une identité de séquence d'environ 80%, ils sont caractérisés par des différences significatives notamment de stabilité thermodynamique.

Les spectres 2D [^1H - ^{15}N]-HSQC du domaine Niv P_{XD} et du peptide $P_{588-650}$ sont caractéristiques d'une protéine repliée avec une grande dispersion des déplacements chimiques et des pics bien visibles et d'une protéine désordonnée, respectivement (**Chapitre 3, Figures 18 et 32**). Néanmoins, dans le contexte de la protéine entière ou de la construction $P_{471-709}$, la totalité des signaux de NiV P_{XD} et une partie des résonances de $P_{588-650}$ ne sont pas visibles dans les spectres 2D [^1H - ^{15}N]-HSQC, suggérant que ces domaines ne se comportent plus indépendamment. Ces expériences démontrent que le comportement du P_{XD} et $P_{588-650}$ est influencé par l'état tétramérique de la phosphoprotéine et peut être causé par une interaction entre les domaines P_{XD} . Ce phénomène pourrait moduler l'affinité de P_{XD} pour les nucléocapsides et donc réguler la formation et la dynamique du complexe de transcription/réplication.

L'ensemble des informations structurales que nous avons obtenues pendant ces travaux devraient nous permettre de construire des modèles atomiques de la phosphoprotéine entière du virus Nipah, et la procédure de sélection d'ensembles représentatifs en utilisant les données de diffusion des rayons X aux petits angles devrait nous permettre d'obtenir une représentation d'ensemble de la structure de NiV P comparable à celle que nous avons obtenue pour VSV P. Nos données expérimentales démontrent que la P de NiV peut atteindre une dimension maximale d'environ 40 nm et un rayon de giration d'environ 11 nm et que la partie N-terminale est majoritairement responsable de cette taille gigantesque.

II. Le complexe N⁰-P de VSV

Le complexe N⁰-P constitue un composant essentiel du complexe de réplication en maintenant la nucléoprotéine sous forme soluble et monomérique, compétente pour l'encapsidation des génomes viraux néo-synthétisés. Néanmoins, pendant longtemps l'information structurale, nécessaire pour mieux comprendre les mécanismes par lesquels la P accompli ses fonctions de chaperon, est restée inaccessible en raison de la difficulté à obtenir des quantités suffisantes de complexe N⁰-P pour réaliser des études structurales. Le développement de la méthode de reconstruction du complexe N⁰-P entre un variant tronqué de la N (N_{Δ21}) et le peptide N-terminale de la P (P₆₀) a permis d'obtenir les quantités nécessaires afin de mener à bien ces études. Ainsi nous avons reconstitué pour une première fois un complexe N⁰-P des *Mononegavirales* à partir de composant purifiés séparément.

La masse moléculaire de 53 ± 3 kDa du complexe N_{Δ21}⁰-P₆₀ obtenue par SEC-MALLS-RI correspond bien à un complexe monomérique de stœchiométrie 1 :1 (un P₆₀ pour une N_{Δ21}). Dans la structure cristalline du complexe N_{Δ21}⁰-P₆₀ la conformation de N_{Δ21} reste très similaire à celle observée dans le complexe N-ARN et seuls les résidus 5 à 33 du peptide P₆₀ sont visibles. L'absence de densité électronique pour les résidus 1-4 et 34-60 suggère fortement qu'ils sont désordonnés dans le cristal tandis que les études par RMN ont montré que les résidus 1-19 subissent des échanges conformationnels en régime intermédiaire (à des échelles de temps de la millioseconde) et interagissent probablement avec la surface de la N. Ce complexe a cristallisé sous la forme d'anneaux, très semblables aux anneaux de N-ARN. Néanmoins les études en solution par SAXS ont confirmé que le complexe est monomérique et que la forme cristalline est probablement un effet de la cristallisation (**Chapitre 5**, article (Leyrat, Yabukarski et al. 2011)).

L'ensemble de ces résultats et l'impossibilité de former le complexe N⁰-P à partir de la protéine N sauvage (Yabukarski et Leyrat, résultats non publiés) démontrent l'importance des sous domaines N- et C-terminaux dans la stabilisation de la forme oligomérique N-ARN. La structure cristalline du complexe N_{Δ21}⁰-P₆₀ nous a dévoilé comment la partie N-terminale de la P empêche l'oligomérisation de la N et l'interaction avec l'ARN. Dans la structure, les résidus 5 à 15 du peptide P₆₀ se positionnent dans une cavité sur la surface du protomère N_i qui est normalement occupée par le bras N_{NTA} du protomère voisin N_{i-1} dans le contexte du complexe N-ARN (**Chapitre 5**, article « Structure of the vesicular stomatitis virus N⁰-P complex », **Figure 4**). Ainsi les 15 premiers résidus du côté N-terminal de la P empêcheraient la fixation du bras N_{NTA} du protomère N_{i-1} sur la surface du protomère N_i. Le peptide se fixe d'un côté de

la N au niveau de la cavité de fixation d'ARN expliquant ainsi l'inhibition de la fixation d'ARN (Leyrat, Yabukarski et al. 2011).

Nous avons également reconstitué le complexe $N_{\Delta 21}^0$ -P contenant la P entière et nous l'avons étudié par diffusion des neutrons aux petits angles et par RMN (**Chapitre 5**). Ces expériences ont montré que la P garde bien sa flexibilité dans le complexe. Les résultats de RMN démontrent que le complexe n'a pas une stœchiométrie fixe de 1 dimère de P pour 1 N (Mavrakis, Iseni et al. 2003), mais peut avoir différents degrés de saturation en fonction des concentrations relatives de N et P, qui peuvent changer pendant les différentes étapes de réplication. Ce complexe, produit à partir des composants purifiés, se fixe bien sur les NC virales et peut probablement déplacer la P présente (Yabukarski et Ivanov, résultats non publiés).

Il est intéressant de noter que la phosphoprotéine possède des sites de phosphorylation au niveau des résidus 60, 62 et 64 pour VSV (Takacs, Barik et al. 1992) et 63 et 64 chez RaV (Gupta, Blondel et al. 2000). Ces résidus se trouvent entre les sites d'interaction avec N^0 et avec la polymérase virale (Chenik, Schnell et al. 1998; Castel, Chteoui et al. 2009) et leur phosphorylation peut éventuellement modifier l'interaction pour l'un ou l'autre. Pour VSV la phosphorylation semble être importante pour la transcription (Barik and Banerjee 1992; Takacs, Barik et al. 1992) mais pas pour la réplication (Pattnaik, Hwang et al. 1997).

III. Le complexe N^0 -P de NiV

A la différence des vésiculovirus (VSV), lyssavirus (RaV) et pneumovirus (RSV), la N des morbilivirus (MeV), respirovirus (SeV) et henipavirus (NiV et HeV) ne semble pas former des anneaux N-ARN lorsque la protéine est exprimée sous forme recombinante (Bhella, Ralph et al. 2002; Ong, Yusoff et al. 2009) et donc ne peut pas être étudiée par cristallographie de la même façon que la N de VSV (Green, Zhang et al. 2006), RaV (Albertini, Clapier et al. 2006; Albertini, Wernimont et al. 2006) ou RSV (Tawar, Duquerroy et al. 2009; El Omari, Dhaliwal et al. 2011). En plus de son domaine globulaire N_{CORE} , la N de ces paramyxovirus contient une région désordonnée, N_{TAIL} , qui rend les nucléocapsides plus flexibles et complique leur étude par microscopie électronique. Ces difficultés nous ont conduit à envisager la reconstruction du complexe N^0 -P pour le virus Nipah entre un variant tronqué de la N (N_{32-383}) et un peptide N-terminal (P_{50}) dans le but obtenir à la fois une information structurale sur la N et sur les mécanismes par lesquels cette protéine est maintenue soluble et monomérique par la P - « taper deux lapins avec la même balle ». Cette

stratégie, très semblable à celle utilisée pour la reconstruction de N⁰-P de VSV, nous a permis d'obtenir la première structure de la N à résolution atomique pour un membre des *Paramyxovirinae* et également la première structure du complexe N⁰-P d'un paramyxovirus.

Les études par RMN ont montré que les 50 premiers résidus de la partie N-terminale de P sont impliqués dans la formation du complexe N⁰-P de NiV, en accord avec ce qui est observé pour SeV et RaV (Curran, Marq et al. 1995; Mavrakakis, Mehous et al. 2006) et les données de SAXS et de SEC-MALLS-RI ont confirmé l'état hétérodimérique du complexe N₃₂₋₃₈₃⁰-P₅₀ (**Chapitre 5**, article « Crystal structure of the N⁰-P complex of Nipah virus »).

Le complexe N₃₂₋₃₈₃⁰-P₅₀ a cristallisé sous forme monomérique. La structure de la N ressemble à celle de RSV, malgré le manque de conservation de séquence, et est composée de deux domaines. Étonnamment, le domaine C-terminal est pratiquement identique à celui de N RSV et constitue le site de fixation du peptide P₅₀. Le domaine N-terminal ressemble également fortement au domaine correspondant de N RSV mais les différents composants de ce domaine ne sont pas positionnés de la même façon. Dans la structure du complexe N₃₂₋₃₈₃⁰-P₅₀, la N semble avoir une conformation ouverte, probablement incapable de fixer de l'ARN. D'ailleurs la boucle englobant l'ARN dans la structure N-ARN de RSV n'est pas bien définie dans la structure de la N₃₂₋₃₈₃⁰-P₅₀ ce qui suggère qu'elle est flexible en absence d'ARN.

Dans la structure du complexe NiV N₃₂₋₃₈₃⁰-P₅₀, seuls les résidus 1 à 35 sont visibles et les résidus 16 à 50 semblent être désordonnés dans le cristal. Étonnamment, et contrairement à ce que nous avons observé pour le complexe N_{Δ21}⁰-P₆₀ de VSV, dans la structure du complexe N₃₂₋₃₈₃⁰-P₅₀ de NiV le peptide ne semble pas interférer directement avec la fixation de l'ARN car son site de fixation ne se superpose pas au site putatif de fixation d'ARN. En supposant que l'organisation N-ARN et les contacts entre les domaines et les sous-domaines du virus Nipah sont similaires à ceux observés chez RSV, la superposition du domaine C-terminal de la N de NiV du complexe N₃₀₋₃₈₃⁰-P₅₀ sur le domaine correspondant de la N de RSV dans le contexte du complexe N-ARN démontre comment le P₅₀ empêche l'oligomérisation de la N. Les 20 premiers résidus de P₅₀ se positionnent dans une cavité sur le domaine C-terminal du protomère N_i, partiellement occupé par le bras N_{NTA} du protomère N_{i-1} dans l'anneau de N-ARN de RSV. Les résidus 20 à 35 de P₅₀ forment un « chapeau » en se fixant au dessus du domaine C-terminal du protomère N_i, qui constitue la surface de fixation pour le bras N_{CTA} du protomère N_{i+1}. Ainsi les résultats obtenus suggèrent que P₅₀ empêche l'oligomérisation de N en bloquant les interactions entre les sous-domaines et les domaines des différents protomères de la N (**Figure 1**) et (**Chapitre 5**, article « Crystal structure of the N⁰-P complex of Nipah virus », **Figure 7**). Le rôle de la tyrosine 258 n'est pas clair mais le

résidu est absolument conservé chez les *Paramyxovirinae* et sa chaîne latérale est positionnée dans la cavité de fixation d'ARN, à l'opposé de la tyrosine 251 de RSV (**Chapitre 5**, « Crystal structure of the N⁰-P complex of Nipah virus » **Figure 8c et 8d**).

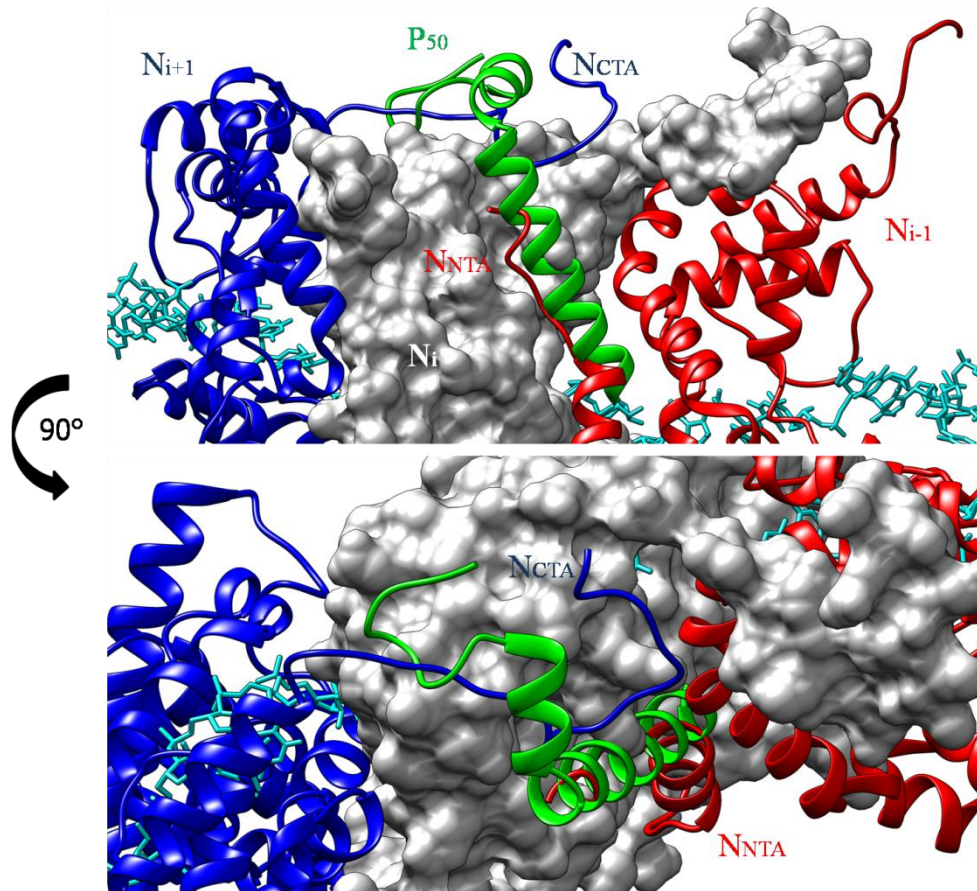


Figure 1 : Inhibition de l'assemblage de la N. Superposition du complexe N-ARN de RSV et le complexe N⁰-P₅₀ de NiV. N-ARN de RSV : le protomère Ni+1 en bleu, le protomère Ni en gris et le protomère Ni-1 en rouge. Les domaines C-terminaux de la N (N_{CTD}) du complexe N⁰-P de NiV et du protomère Ni de N-ARN de RSV sont superposés. P₅₀ (en vert) bloquerait la fixation du bras N-terminal (N_{NNTA}) du protomère Ni-1 et du bras C-terminal (N_{NCTA}) du protomère Ni+1 sur le domaine C-terminal (N_{CTD}) du protomère Ni.

Malgré le fait que la conservation de séquences en acides aminés des phosphoprotéines soit faible au sein des *Paramyxovirinae*, il a été montré qu'une conservation existe dans les 50 premiers résidus et deux motifs, Soyuz1 et Soyuz2 (Karlin and Belshaw 2012), ont été identifiés. Dans la structure du complexe N₃₂₋₃₈₃⁰-P₅₀ les motifs Soyuz1 et Soyuz2 sont impliqués dans les interactions avec la N au sein de l'hélice et du « chapeau ». Les séquences en acides aminés des domaines N_{CORE} pour les genres morbilivirus, respirovirus et henipavirus semblent également être bien conservées ce qui souligne la pertinence des résultats structuraux obtenus. De plus la cavité de fixation du P₅₀ a une conservation supérieure à la moyenne pour le N_{CORE} (**Chapitre 5**, article « Crystal structure of the N⁰-P complex of Nipah virus », **Figure 5**). Tous ces résultats suggèrent fortement que le site de fixation du P_{NTR} de la phosphoprotéine sur la nucléoprotéine dans le complexe N⁰-P

pour les virus des genres morbilivirus, respirovirus et henipavirus peut constituer une cible pour des stratégies de développement de composés antiviraux à spectre large. En accord avec cette déclaration, des résultats préliminaires montrent que la réplication du virus Nipah peut être inhibée par l'expression du peptide P₅₀ dans la cellule (Volchkov, résultats non publiés), et il sera très intéressant d'essayer d'inhiber cette réplication par des peptides P₅₀ des virus plus ou moins proches.

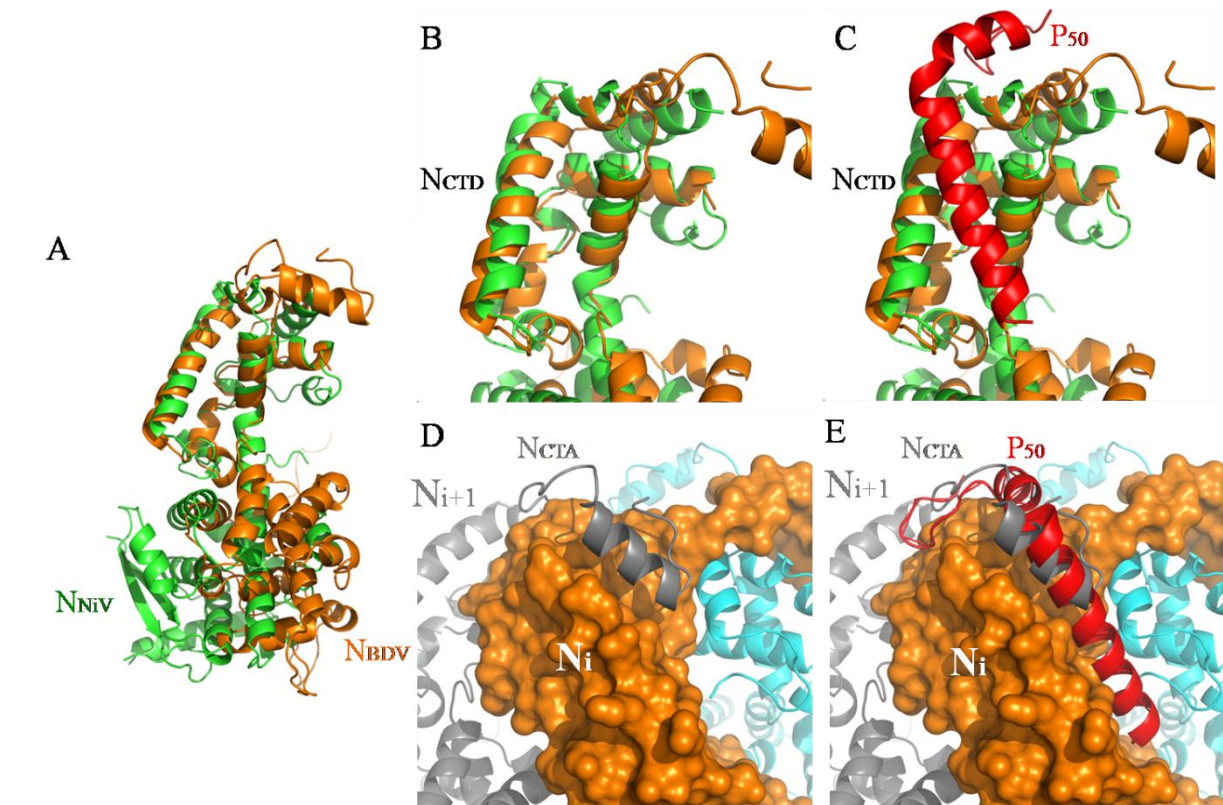


Figure 2 : Modèle proposé de l'inhibition de l'oligomérisation de la nucléoprotéine (N) chez le virus Borna (BDV). **A)** La N de BDV (en orange, code PDB : 1PP1) et la N du complexe N⁰-P₅₀ de NiV (en vert) sont superposés sur le domaine C-terminal (N_{CTD}). **B)** Zoom sur les domaines P_{CTD} superposés. **C)** Même représentation que B) mais le P₅₀ (en rouge) du complexe N⁰-P₅₀ de NiV est également visualisé. **D)** Le tétramère de N BDV est représenté. Le protomère Ni+1, Ni et Ni-1 sont colorés en gris, orange et turquoise, respectivement. **E)** La fixation d'un peptide, semblable à P₅₀ de NiV N⁰-P₅₀, bloquerait la fixation du bras C-terminal (N_{CTA}) du protomère Ni+1 (gris) sur le domaine C-terminale (N_{CTD}) du protomère Ni (orange).

Il est également intéressant de noter que le domaine C-terminal (N_{CTD}) de la N de NiV peut être facilement superposé avec le domaine correspondant de la N du virus Borna (BDV), en accord avec la conservation structurale de ce domaine (Luo, Green et al. 2007; Ruigrok, Crepin et al. 2011). La N de BDV a été cristallisée en absence d'ARN (Rudolph, Kraus et al. 2003) et forme des tétramères qui sont stabilisés par des interactions entre le corps d'un protomère N et les sous domaines des protomères voisins. La cavité de fixation du P₅₀ est bien conservée et il est probable qu'un peptide similaire chez BDV empêche l'oligomérisation de la N en la maintenant monomérique (**Figure 2**).

L'ensemble de ces résultats suggère que la stratégie généralement adoptée par les phosphoprotéines des *MNVs* afin de maintenir la nucléoprotéine soluble et monomérique est de bloquer les interactions entre les domaines et les sous domaines des protomères N_{i-1} , N_i et N_{i+1} .

IV. Implications pour les mécanismes de réplication des *Mononegavirales*

Les mécanismes de réplication chez les MNVs sont encore méconnus. Notamment nous manquons d'information structurale détaillée concernant la polymérase virale, ses interactions avec la phosphoprotéine, ou concernant la structure et la dynamique du complexe N^0 -P. Les travaux présentés ici permettent de mieux comprendre la nature des interactions entre N et P dans le complexe N^0 -P de deux virus représentatifs des *MNV*. Nous avons également démontré que la phosphoprotéine des *Rhabdoviridae* garde sa flexibilité dans le contexte du complexe N^0 -P (**Chapitre 5**) et que la P des *Paramyxovirinae* est une protéine très flexible (**Chapitre 4**).

Les connaissances que nous avons accumulées nous permettent d'imaginer le rôle de la P dans l'organisation de la machinerie de transcription/réplication. Une fois fixée sur la matrice N-ARN, l'organisation tentaculaire et la flexibilité de la protéine P permettrait de positionner correctement la protéine L sur la matrice afin d'optimiser la synthèse d'ARN. De même, à la manière d'un tentacule, P positionnerait N^0 exactement au contact du 5' de l'ARN (anti)génomique naissant, puis pour les N^0 suivantes, les orienteraient correctement pour un emboîtement.

La plasticité de la P interviendrait également au moment de la réorganisation de ce complexe multimoléculaire lors du déplacement de la polymérase ou serait directement impliquée dans son déplacement. En échantillonnant un grand volume d'espace (activité tentaculaire), l'oligomère de P ou la région N_{TAIL} augmente la probabilité de rencontre avec les différents partenaires et pourrait donc augmenter la vitesse de la réaction de fixation selon un mécanisme de « pêche à la mouche » (« fly casting ») (Shoemaker, Portman et al. 2000).

Concernant les rhabdovirus, le mécanisme d'encapsidation le plus probable d'après les données actuelles implique la fixation du complexe N^0 -P sur le complexe N-ARN matrice ou sur le complexe N-ARN en train d'être synthétisé. Pendant la réplication la N du complexe N^0 -P doit être transférée sur l'ARN viral en train d'être synthétisé. L'ARN naissant se positionnerait dans le sillon de N^0_i délogant ainsi l'extrémité C-terminale de l' α -MoRE de P_{NTRi} . Cette déstabilisation entraînerait un relâchement de la liaison de l'extrémité N-

terminale de l' α -MoRE de ce même P_{NTRi} qui encombre le sillon d'accueil du bras N_{NTA} de N^0_i . En conséquence, à l'arrivée du complexe suivant N^0_{i+1} , le bras N_{NTA} de ce dernier entrerait en compétition favorable pour déloger l'extrémité N-terminale de l' α -MoRE du P_{NTRi} en amont resté provisoirement associé à la nucléocapside en formation. Il positionnerait ainsi le protomère N^0_{i+1} de manière à accueillir le segment suivant d'ARN naissant (voir **Figure 3**).

Néanmoins il n'est pas évident d'imaginer un tel mécanisme pour les *Paramyxovirinae* car la partie N-terminale de la P ne semble pas bloquer directement la fixation de l'ARN. Une possibilité serait que dans la conformation de la N, observée dans le complexe N^0 -P₅₀ de NiV, la tyrosine 258 empêche l'ARN de se fixer et que c'est la dissociation de la partie N-terminale de la P ainsi que la réorganisation de la région charnière qui permettraient à ce résidu de changer sa position et la fixation de l'ARN. Il est difficile d'imaginer que les quelques bases d'ARN qui se fixent par protomère de N peuvent déplacer la partie N-terminale de la P et il est très tentant de spéculer que c'est la polymérase virale qui assiste cet échange.

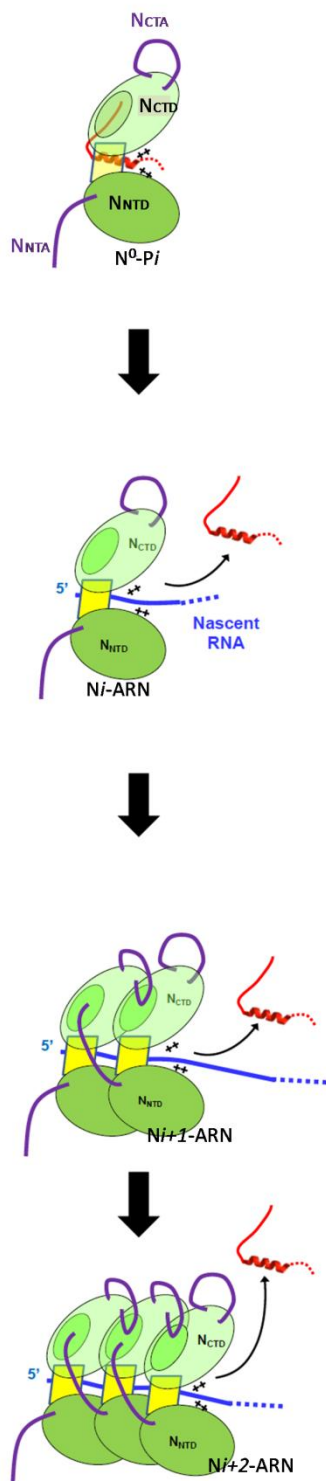
Concernant les rhabovirus les données actuellement disponibles sont en faveur d'un modèle dans lequel les dimères de P se fixent sur la nucléocapside et c'est la polymérase virale qui se déplace par saut (voir **Chapitre 1** - Introduction). Ce modèle est renforcé par la forte affinité de la P pour N-ARN et sa dissociation lente (Ribeiro, Leyrat et al. 2009) et par le fait que le domaine de dimérisation n'est pas essentiel pour la transcription (Jacob, Real et al. 2001). Concernant les paramyxovirus des mesures d'affinité préliminaires entre la phosphoprotéine entière et les nucléocapsides (NC) recombinantes du virus Nipah suggèrent une affinité de l'ordre de 50 nM et une dissociation relativement lente (Yabukarski, Carriel et Fender, résultats non publiés). Ce type d'interaction forte et le volume d'espace monstrueux occupé par la phosphoprotéine ne nous semblent pas être en faveur d'un déplacement de la P le long des NC, comme proposé par le modèle de « cartwheeling » (Curran 1998).

V. Accéder à la matrice ARN génomique enfouie dans le manchon de l'homopolymère de N

En sus de son rôle dans l'encapsidation, la fixation de P_{NTR} sur N suggère un mécanisme d'initiation de la synthèse d'ARN à l'extrémité 3' de l'ARN (Leyrat, Yabukarski et al. 2011). L'ARN encapsidé par la N dans les nucléocapsides n'est pas accessible pour la polymérase, mais la première molécule de N à l'extrémité 3' du génome expose son site

d'accueil à l' α -MoRE du P_{NTR}. Pourquoi ne pas envisager qu'en se fixant sur ce site, une molécule de P du complexe P-L déplace l'ARN de la première nucléoprotéine permettant ainsi à la polymérase de se fixer sur l'extrémité 3' du génome (**Chapitre 5**, article « Structure of the vesicular stomatitis virus N⁰-P complex », **Figure 7**) ? Ce processus expliquerait des résultats étranges obtenus avec un système de transcription et de réplication du génome de SeV reconstitué *in vitro* : l'addition de faibles concentrations du complexe N⁰-P soluble stimule la transcription, alors que l'addition de concentrations plus élevées stimule la réplication et réduit la transcription (Curran and Kolakofsky 2008). L'effet sur la transcription aux faibles concentrations de N⁰-P pourrait s'expliquer non par la présence de N mais par la présence de P_{NTR} libres. L'un d'eux, en se fixant sur N à l'extrémité 3' du génome, déplacerait N de l'ARN et stimulerait l'initiation de la synthèse d'ARN en permettant à la polymérase d'accéder plus aisément à l'ARN génomique (**Figure 7** de l'article « Structure of the Vesicular Stomatitis virus N⁰-P Complex »). Par contre, aux concentrations plus élevées de N⁰-P, ce complexe stimulerait alors la réplication par rapport à la transcription, conduisant à un basculement entre ces deux activités.

Rhabdoviridae



Paramyxovirinae

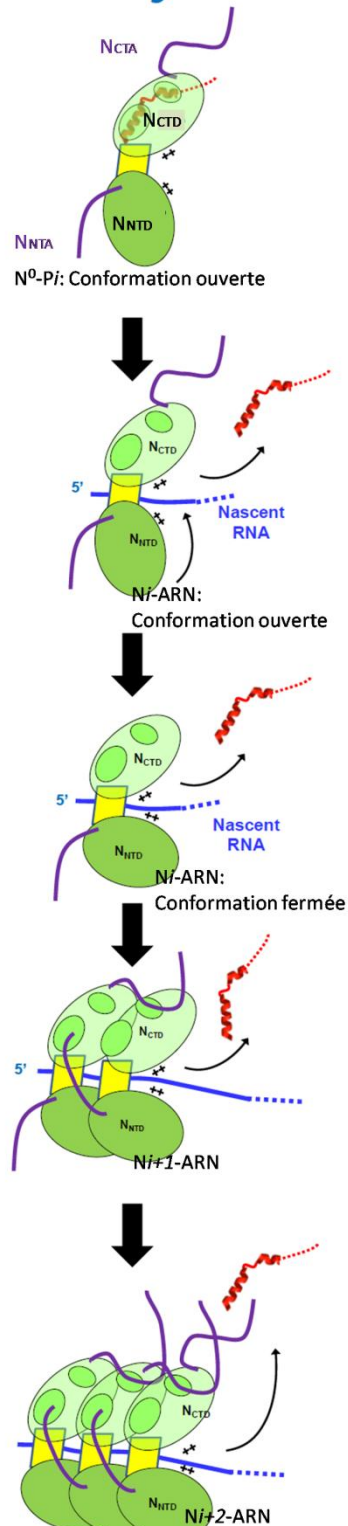


Figure 3 : Illustration du modèle d'encapsidation d'ARN proposé pour les *Rhabdoviridae* (gauche) et les *Paramyxovirinae* (droite). Pour les *Rhabdoviridae* la conformation de la N dans les structures de N-ARN et N⁰-P est identique et un changement de conformation pendant l'encapsidation d'ARN n'est pas évident. Dans le complexe N⁰-P de NiV la N semble être dans une conformation ouverte, probablement induite par la fixation du peptide N-terminal de la P, et un changement conformationnel pendant l'encapsidation d'ARN semble nécessaire pour les *Paramyxovirinae*.

ANNEXES

Review: Structural insights into the rhabdovirus transcription/replication complex

Authors: Ivan Ivanov, **Filip Yabukarski**, Rob Ruigrok, Marc Jamin

Publié en 2011 dans *Virus Research*, Volume 162, pages 126-137



Review

Structural insights into the rhabdovirus transcription/replication complex

Ivan Ivanov^{a,b}, Filip Yabukarski^a, Rob W.H. Ruigrok^a, Marc Jamin^{a,*}^a UMI 3265 UJF-EMBL-CNRS Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France^b Institut Laue-Langevin, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

ARTICLE INFO

Article history:

Available online 22 September 2011

Keywords:

Mononegavirales
 Rhabdoviridae
 Rabies virus
 Vesicular stomatitis virus
 Transcription
 Replication

ABSTRACT

The rhabdoviruses have a non-segmented single stranded negative-sense RNA genome. Their multiplication in a host cell requires three viral proteins in addition to the viral RNA genome. The nucleoprotein (N) tightly encapsidates the viral RNA, and the N–RNA complex serves as the template for both transcription and replication. The viral RNA-dependent RNA polymerase is a two subunit complex that consists of a large subunit, L, and a non-catalytic cofactor, the phosphoprotein, P. P also acts as a chaperone of nascent RNA-free N by forming a N⁰–P complex that prevents N from binding to cellular RNAs and from polymerizing in the absence of RNA. Here, we discuss the recent molecular and structural studies of individual components and multi-molecular complexes that are involved in the transcription/replication complex of these viruses with regard to their implication in viral transcription and replication.

© 2011 Elsevier B.V. All rights reserved.

Contents

1. The rhabdoviruses	126
2. The viral transcription/replication machinery	127
3. The nucleoprotein and the N–RNA complex	129
4. The phosphoprotein	130
5. Complexes between the nucleoprotein and the phosphoprotein	130
5.1. The N–RNA–P complex	130
5.2. The N ⁰ –P complex	131
6. The L subunit of the viral polymerase	134
7. The viral particle	134
8. Future work	135
Acknowledgments	135
References	135

1. The rhabdoviruses

The *Rhabdoviridae* is a family of enveloped viruses that have a non-segmented genome ((–)RNA) made of a single stranded negative-sense RNA molecule. The *Rhabdoviridae* family is grouped in the order *Mononegavirales* (MNV) (the name is a composition of three elements: *Mono* – single; *nega* – negative sense; *virales* – virus) with the *Filoviridae* (Ebola and Marburg viruses), the *Paramyxoviridae* (measles, mumps, respiratory syncytial viruses) and the *Bornaviridae* (Borna disease virus). All these viruses, share

a similar organization of their genome and virion as well as similar strategies of RNA replication and transcription (Pringle, 1997). Viruses classified as *Rhabdoviridae* are currently grouped into six genera on the basis of structural properties, antigenicity and phylogenetic analyses (Delmas et al., 2010; Tordo et al., 2004). The *Lyssaviruses*, *Vesiculoviruses* and *Ephemeroviruses*, infect a variety of animal hosts, including mammals, fishes and arthropods, the *Novirhabdoviruses* infect numerous fishes and the *Cytorhabdoviruses* and *Nucleorhabdoviruses* are arthropod-borne and infect plants (Jackson et al., 2005; Tordo et al., 2004). Numerous other rhabdoviruses remain unclassified. The rabies virus (RAV) and the vesicular stomatitis virus (VSV), respectively, a *Lyssavirus* and a *Vesiculovirus*, are prototypes of this large family of viruses.

RAV is the main causative agent of rabies. It is a neurotropic virus that causes an incurable human brain disease with a complex pathogenesis (Dietzschold et al., 2008; Hemachudha et al., 2002). It

* Corresponding author at: UMI 3265 UJF-EMBL-CNRS Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France.
 Tel.: +33 4 76 20 94 62; fax: +33 4 76 20 94 00.
 E-mail address: jamin@embl.fr (M. Jamin).

leads to death if it remains untreated before the disease is declared, but the infection can be prevented. Since its first successful use by Louis Pasteur in 1885, an anti-rabies vaccine is available as well as a post-infection treatment (Rupprecht et al., 2006). Currently, rabies is under control in Europe and North America because of efficient vaccination campaigns of dogs and wild animals (Warrell and Warrell, 2004), but, it still kills more than 50,000 people every year mainly in Africa and Asia (see WHO Fact Sheet N° 99, 2009) and causes an important economic burden in affected populations (Knobel et al., 2005). Recent reports suggest that the number of cases is even substantially underestimated in poor rural areas (Cleaveland et al., 2002). The worldwide persistence of the disease is mainly due to the cost of efficient vaccination campaigns and of post-exposure treatments. Of the 10 million people receiving post-exposure treatments each year after being exposed to rabies-suspect animals (WHO estimation in 2004), only a few benefit from local instillation of anti-rabies immunoglobulins, which has been demonstrated to be an indispensable complement to vaccination. The main natural reservoirs of rabies virus are bats or carnivores such as dogs, foxes, raccoons and skunks, but the mechanism of cross-species transmission remains poorly understood (Streicker et al., 2010). New strains, potentially unsusceptible to vaccination against classical rabies could emerge, cross the species barrier and spread in the human population. Indeed, several isolates of bat *Lyssavirus* are already less sensitive or insensitive to the current vaccine and post-infection treatments against rabies (Hanlon et al., 2005; Weyer et al., 2008). The vesicular stomatitis virus (VSV) is an arthropod-transmitted virus that causes a mild disease in horses, cattle and pigs and rarely transmits to human. It has economical impacts in the farm industry. For many years, VSV has been used as a model system for studying the replication cycle of the *Rhabdoviridae* and, more generally, the interactions between viruses and their host cells. More recently, VSV has been used as a vaccine vector and as an oncolytic virus (Lichty et al., 2004; Rose et al., 2008). Other *Vesiculoviruses* such as Chandipura virus or Piry virus also infect humans. In recent outbreaks in India, Chandipura virus produced acute encephalitis with high fatality ratio (Basak et al., 2007; Rao et al., 2004). Finally, *Vesiculoviruses* infecting fishes or plants are of major economic importance to fish farming and agricultural industries (Jackson et al., 2005).

VSV and RAV use different strategies for regulating their replication in their host cell that lead to the different pathologies caused by these viruses. VSV replicates rapidly to high levels of progenies in a minimum of time and strongly interferes with the host's cell metabolism. The virus strongly inhibits host gene expression and translation (Ahmed et al., 2003; Connor and Lyles, 2002, 2005), interferes with the host cell innate immune response (Durbin et al., 1996) and induces apoptosis of infected cells (Koyama, 1995). In contrast, RAV replicates more slowly, interferes much less with the host cell metabolism and is less cytopathic than VSV (Schnell et al., 2010). Recently, recombinant RAV and VSV have been recovered from cDNA (Lawson et al., 1995; Schnell et al., 1994; Whelan et al., 1995), but most studies analyzing the processes of transcription and replication and their regulations have been carried out for VSV, because functional transcription/replication complexes of VSV can be reconstituted from isolated components (Emerson and Yu, 1975) but not of RAV. However, the structural organization of the transcription/replication complex is similar, and comparative studies provide useful information about the mechanisms of action of this molecular machinery.

2. The viral transcription/replication machinery

The genome of the *Rhabdoviridae* is of 9–18 kb in length and comprises up to ten genes flanked by untranslated leader (le) and

trailer (tr) RNA regions. RAV and VSV genomes contain only five genes that are common to all members of the MNV and encode successively from the 3' terminus, the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the large subunit of the RNA-dependent RNA polymerase (L) (Fig. 1A).

The replication cycle of both RAV and VSV occurs entirely in the cytoplasm of their host cells (Follett et al., 1974) and leads to the formation of intracytoplasmic inclusions (Heinrich et al., 2010; Lahaye et al., 2009). The G and M proteins are involved in the entry of the virus in its host cell and in the assembly and budding of new viral particles. The viruses attach to their host cells through the G protein and are internalized by endocytosis. The ribonucleoprotein complex containing the genomic RNA together with N, P and L proteins, named hereafter the nucleocapsid (NC), is then delivered into the cytoplasm by fusion of the viral envelope with the lipid bilayer of the endocytic vesicle. The fusion is induced by a pH-dependent conformational change of the G protein (Gaudin et al., 1999; Libersou et al., 2010). The M protein maintains the structural integrity of the virus particle by connecting the lipid envelope to the NC, and in infected cells, the M protein spontaneously associates with the cytoplasmic membrane, where it recruits newly synthesized nucleocapsids (Solon et al., 2005). The sequence of the M protein contains a motif of interaction with the host cell machinery involved in multi-vesicular body (MVB) formation and it mediates the release of viral particles by redirecting this machinery to the cytoplasmic membrane (Harty et al., 1999; Irie et al., 2004).

The NC constitutes the active machinery that carries out both the transcription and replication of the genome and the actual infectious part of the virus (Szilagy and Uryvayev, 1973). For VSV, transcription could be reconstituted *in vitro* using a mixture of L, P and N–RNA complex isolated from the virus (Emerson and Yu, 1975), and both transcription and replication could be reconstituted *in vivo* using a three-plasmid (L, P and N) inverse genetic system (Pattnaik et al., 1997). For VSV, no host cell protein is necessary in stoichiometric amount relative to the viral proteins, but the involvement of low amounts of cellular proteins cannot be ruled out. Numerous results in the literature support a role of the phosphorylation of P or N in the activation and regulation of the viral transcription/replication machinery (Pattnaik et al., 1997; Toriumi and Kawai, 2004), and it is likely that cellular kinases associated with the NC are required for the proper action of this machinery (Lenard, 1999).

In the NC (Fig. 1B), the N protein entirely coats the RNA genome, and it is the N–RNA complex, rather than the naked RNA, that serves as a template for the two-subunit RNA-dependent RNA polymerase comprising the P and L proteins (Moyer et al., 1991). However, the structural organization of the polymerase complex on the N–RNA complex and the role of N and its interactions with the polymerase complex remain unknown. In addition, the encapsidation by N also protects the genome from cellular nucleases and hides it from the intracellular receptors of the innate immune response. The 5' end of the viral RNA bears a triphosphate that is recognized by RIG-I (Hornung et al., 2006) and would certainly trigger an immune response if it were freely accessible.

At first, when it is delivered into a host cell, the NC carries out the primary transcription of the genome, generating a short uncapped and untranslated leader RNA (leRNA) and five capped and polyadenylated messenger RNAs (mRNA) that encode the viral proteins. The amounts of leRNA and of the five mRNAs produced in this process decrease in the order of the genes on the genome, suggesting that the polymerase initiates at the 3' end of the genome (Abraham and Banerjee, 1976; Ball and White, 1976; Emerson, 1982) and transcribes the genes sequentially from the 3' end to the 5' end, although it can also initiate at the N gene start sequence (Chuang and Perrault, 1997; Whelan and Wertz, 2002). It is assumed that the polymerase complex proceeds by a

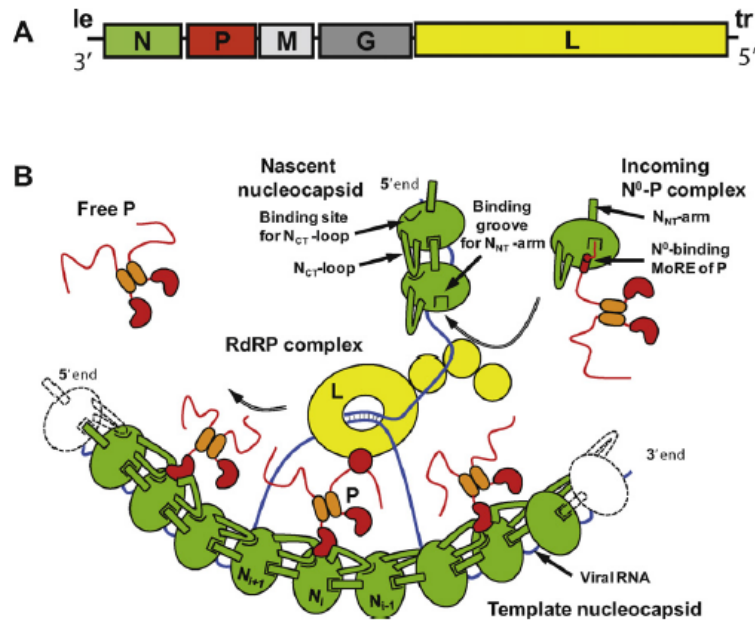


Fig. 1. The rhabdovirus transcription/replication machinery. (A) Schematic organization of the genome. The gene products are denoted as follows: N, nucleoprotein; P, phosphoprotein; M, matrix protein; G, surface glycoprotein and L, large subunit of the RNA-dependent RNA polymerase. Le and tr indicates the leader and trailer RNA sequences. (B) Schematic representations of the transcription/replication machinery. The nucleoprotein (in green) forms with the RNA genome (blue line) the active template for the RNA-dependent RNA polymerase (RdRP) complex comprising the L (in yellow) and P (in orange and red) proteins. P forms dimers and each monomer of P contains a central dimerization domain (in orange), a C-terminal N-RNA binding domain (red bean-like shape) and two intrinsically disordered regions (red lines). During the replication of the genome, the newly synthesized antigenomic or genomic RNA copies are encapsidated by nascent N molecules that are transferred from the soluble N^0 -P complex. In the N^0 -P complex, the N-terminal MoRE of P (shown as a red cylinder in the N^0 -P complex) prevents host-cell RNA binding by obstructing the RNA binding groove and the self-assembly of N by interfering with the docking of the N_{NT} -arm of another N. Upon transfer of N to the growing NC P is released, the binding groove for the N_{NT} -arm is freed in the RNA-bound form and can accept the next incoming N molecule.

stop-and-start mechanism remaining attached to the template after pausing and re-initiating at the next gene start (Iverson and Rose, 1981). In this process, the polymerase terminates the transcription at each intergenic region by iterative slippage on a poly-U tract that generates the poly-A tail (Barr and Wertz, 2001; Barr et al., 1997; Hwang et al., 1998; Stillman and Whitt, 1999). The mRNAs are capped and methylated at their 5' end and poly-adenylated at their 3' end by the L protein. The mRNAs are then used by the cellular ribosomes for synthesizing viral proteins. In a second stage, the viral polymerase complex switches from transcription to replication. During the replication process the viral polymerase initiates at the 3' end of the genome, ignores all regulatory signals at intergenic regions and produces full-length complementary (+)RNA intermediates that then serves as a template for generating new copies of the (–)RNA genome. These antigenomes and genomes, but not mRNAs, are encapsidated by N, and therefore the replication process requires a continuous supply of nascent RNA-free N molecules (Howard and Wertz, 1989; Patton et al., 1984; Peluso and Moyer, 1988). Finally, at a later stage of the viral cycle, a secondary phase of transcription occurs, thus increasing the amounts of viral proteins present in the cell, possibly to permit the assembly of viral progenies (Simonsen et al., 1979).

The phosphoprotein (P) is an essential non-catalytic cofactor of the polymerase (Emerson and Yu, 1975), but the L subunit carries out the multiple enzymatic activities that are required for the transcription and the replication of the genome (Abraham et al., 1975; Barr and Wertz, 2001; Barr et al., 1997; Grdzlishvili et al., 2005; Hercyk et al., 1988; Li et al., 2006; Moyer and Banerjee, 1975). L catalyzes the template-based RNA synthesis, the capping and methylation of the 5' extremity of the mRNAs and the polyadenylation of their 3' extremity. The role of P as a cofactor of

the polymerase is unclear. P may stabilize L (Canter and Perrault, 1996), but its main known role is in the attachment of L to the N-RNA template. By itself, the L subunit has a poor affinity for the N-RNA template (Mellon and Emerson, 1978), whereas P binds to the N-RNA template and to the polymerase, suggesting that P allows the correct positioning of the polymerase onto its template (Fig. 1B). It is possible that P acts as a processivity factor that keeps L attached to its template during the process at intergenic regions and thereby facilitates the re-initiation of RNA synthesis at the next start signal.

The second major role of P is to chaperone N and allows the assembly of new NCs from nascent N molecules. N has a strong affinity for RNA and must be prevented from encapsidating cellular RNAs. P binds to nascent N molecules, forming a soluble N/P complex, named the N^0 -P complex, with the superscript 0 indicating the absence of RNA (Fig. 1B). The continuous production of the N^0 -P complex is required for the replication of the viral genome and it has been hypothesized that the accumulation of this complex triggers the switch between transcription and replication (Howard and Wertz, 1989; Masters and Banerjee, 1988; Peluso and Moyer, 1988).

Here, we review recent structural studies of the components of the transcription/replication machinery of either VSV or RAV and the new hypotheses that these results raise about the structural organization of the NC and its dynamic remodeling during the functioning of the viral polymerase. Low- and high-resolution structures of N-RNA complexes revealed how N molecules assemble into NCs and how they interact with and protect the RNA molecule. Various biophysical and structural studies demonstrated that the phosphoproteins are highly flexible, made of concatenated disordered regions and structured domains. The structures of

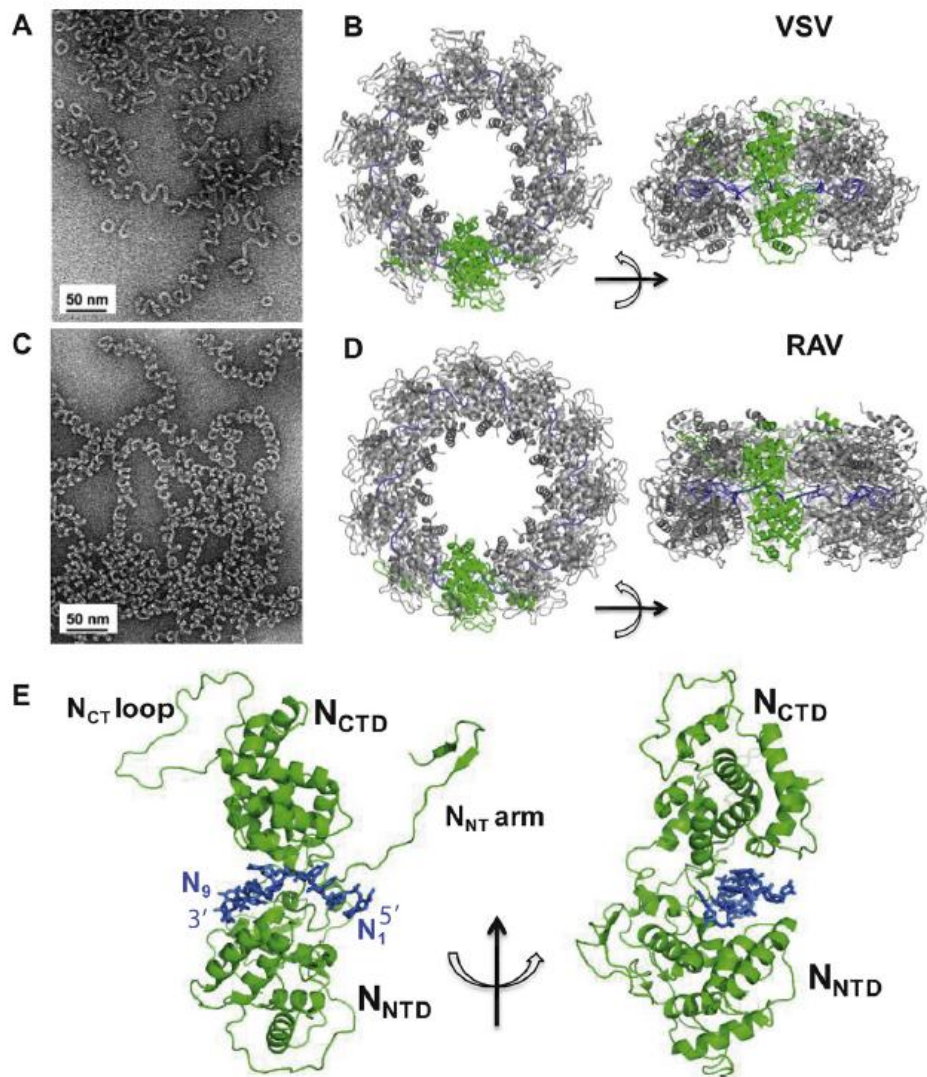


Fig. 2. Structure of the nucleoproteins and N-RNA complexes. (A) Typical electron micrograph of negatively stained VSV nucleocapsids. The viral nucleocapsids were isolated with centrifugation in CsCl gradient and negative staining was done with sodium silicotungstate. (B) Structure of the circular decameric N-RNA complex of VSV (PDB ID: 2gic). One nucleoprotein is shown in green and the RNA is shown in blue. (C) Electron micrographs of negatively stained RAV nucleocapsid. Negative staining was done with sodium silicotungstate. (D) Structure of the circular 11-mer N-RNA complex of RAV (PDB ID: 2gtt). One nucleoprotein is shown in green and the RNA is shown in blue. (E) Ribbon representation of one protomer of VSV N bound to 9 nucleotides. The N protein, shown in green, has an N-terminal domain (N_{NTD}), a C-terminal domain (N_{CTD}) and two subdomains, the N-terminal arm (N_{NT}-arm, aa 1–21) and the C-terminal loop (N_{CT}-loop, aa 340–375). The RNA, shown in blue, is completely wrapped by N_{NTD} and N_{CTD} and is not accessible for the polymerase.

two types of complexes between N and P, revealed how different regions of P specifically binds to different regions of N and carry out different functions in the transcription/replication complex. Finally, electron micrographs of purified VSV L protein provided a first glimpse at the shape of this large protein.

3. The nucleoprotein and the N-RNA complex

In rhabdovirus NCs, each N molecule binds nine nucleotides (Iseni et al., 1998; Thomas et al., 1985). The number of N molecules present in a virion of VSV (1250 N) (Ge et al., 2010; Thomas et al., 1985) corresponds closely to the theoretical number required for

covering the entire genome (VSV: 11161 nt → 1240 N), although the length of the genomic RNA molecule is not an entire multiple of nine nucleotides. By contrast, in *Paramyxoviridae* each N molecule binds six nucleotides and the length of the genome corresponds to a multiple of six nucleotides (Kolakofsky et al., 2005).

The N-RNA complex extracted from infected cells or from virus particles forms a long and irregular helix (Fig. 2A and C). The electron micrographs reveal a striated structure corresponding to the side by side arrangement of the nucleoproteins along the RNA molecule. The dimensions of the nucleoprotein in the NCs from VSV and RAV are similar as well as the diameter of individual coils (Iseni et al., 1998), but NCs of RAV appear more tightly coiled (Fig. 2A and C). In the NC, the RNA is tightly bound to N, resisting centrifugation

in CsCl density gradient (Lynch and Kolakofsky, 1978). It is highly protected against digestion by nucleases independently of the salt concentration, although the bases are accessible to small molecules (Iseni et al., 2000). The RNA can however be removed by treatment with very high concentrations of RNase A at 60°C (Green et al., 2010).

The expression of the N as a recombinant protein in bacteria or in insect cells leads to the formation of long N–RNA complexes that resemble the natural NCs and to circular N–RNA complexes that contain short unidentified RNAs (Green et al., 2000; Iseni et al., 1998; Schoehn et al., 2001). The recombinant circular RAV N–RNA complexes contain from 9 to 13 N-protomers with a larger proportion of rings with 10 protomers (Albertini et al., 2006a; Iseni et al., 1998), while recombinant VSV N–RNA complexes contain predominantly 10 protomers (Blumberg et al., 1983; Ge et al., 2010; Green et al., 2000). The N–RNA rings of RAV and VSV are less flexible than the long NCs and their structure was reconstructed from electron micrographs (Chen et al., 2004; Iseni et al., 1998; Schoehn et al., 2001) and, more recently, solved at high-resolution by X-ray crystallography (Albertini et al., 2006b; Green et al., 2006) (Fig. 2B and D).

The N protein has two domains, an N-terminal domain (N_{NTD}) and a C-terminal domain (N_{CTD}), that are connected by a hinge region (Fig. 2E). The RNA molecule is bound in a groove of the protein located at the interface between N_{NTD} and N_{CTD} and is completely sequestered from the solvent. In addition N has two subdomains, an N-terminal arm (N_{NT} -arm) and a C-terminal flexible loop (N_{CT} -loop), which exchange in opposite direction between adjacent subunits of the N–RNA complex. The polymer of N in complex with RNA is stabilized by multiple salt bridges between the sugar-phosphate backbone of the RNA and basic residues of N, by contacts between neighboring N molecules involving hydrophobic side-to-side interactions, mainly between adjacent N_{CTD} , and by the exchange of N- and C-terminal sub-domains between adjacent N protomers (Albertini et al., 2006b; Green et al., 2006).

4. The phosphoprotein

Early studies showed that P contains independent functional regions (Das et al., 1997; Takacs et al., 1993) and its hydrodynamic properties suggested that it is a non-globular molecule (Gérard et al., 2007). These characteristic features are explained by the recent findings that P is made of concatenated disordered regions and structured domains. The structure and functions of rhabdovirus phosphoproteins have been reviewed recently (Leyrat et al., 2010, 2011b), and therefore, their structural organization is only briefly discussed here (Fig. 3).

The structured domains have been localized in the amino acid sequence by limited proteolysis (Ding et al., 2004; Mavrakis et al., 2004) or by a meta-prediction of disorder (Gérard et al., 2009). The protein is dimeric (Gérard et al., 2007) and a central autonomous folding domain (P_{CED}) is responsible for the dimerization of the protein (Ding et al., 2006; Ivanov et al., 2010). In RAV and VSV P dimers, the monomers are oriented in a parallel orientation, but the structure of the dimerization domain is different, suggesting that oligomerization is conserved but not the structure (Ivanov et al., 2010). The structure of the C-terminal folded domain (P_{CTD}) of VSV and RAV P clearly adopt a similar fold clearly showing that they are homologous (Mavrakis et al., 2004; Ribeiro et al., 2008). VSV P_{CTD} lacks two α -helices of RAV P_{CTD} , but the common parts can be superimposed. In VSV, the junction between P_{CED} and P_{CTD} (Ribeiro et al., 2008) and the N-terminal region of both P proteins were shown to be globally disordered (Gérard et al., 2009; Leyrat et al., 2011a). The N-terminal intrinsically disordered region (IDR_{NT}) contains the binding sites for the RNA-free N^0 (Chen et al.,

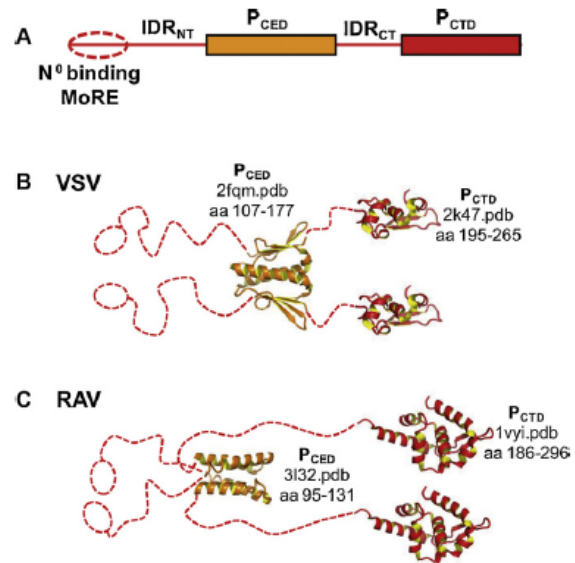


Fig. 3. Structure of the phosphoproteins. (A) Schematic organization of the P proteins. P proteins of rhabdoviruses consist of a long N-terminal intrinsically disordered region (IDR_{NT}), a central oligomerization domain (P_{CED}), a C-terminal IDR (IDR_{CT}) and a C-terminal N–RNA binding domain (P_{CTD}). In VSV, IDR_{CT} is about 16 aa long, whereas in RAV, it is about 60 aa long. The N-terminal region contains the binding site for RNA-free N. In VSV P, this region consists of a MoRE (molecular recognition element) that forms transient α -helices in the isolated protein and adopts a rigid fold in the RNA-free N^0 -P complex. (B) Structured domains of VSV and RAV P. The high resolution structures of two domains corresponding closely to the predicted central domain of P (P_{CED}) and the C-terminal domain (P_{CTD}) have been determined at high resolution. The known three-dimensional structure of P_{CED} and P_{CTD} are shown as cartoon models and the disordered regions are shown as dotted lines. The dotted circles represent the localization of N^0 binding MoRE. The PDB codes are given in figure.

2007; Mavrakis et al., 2006). A detailed NMR characterization of the N-terminal region of VSV revealed the transient presence of two α -helices in the N^0 -binding region, suggesting that this part of P is a molecular recognition element (MoRE) (Leyrat et al., 2011a). A MoRE is a small segment of a protein that is disordered in its isolated form but folds upon binding to a molecular partner (Fuxreiter et al., 2004). The binding site for L has also been localized in the N-terminal part of P but has not yet been precisely mapped (Castel et al., 2009; Emerson and Schubert, 1987).

5. Complexes between the nucleoprotein and the phosphoprotein

During the replication cycle, P forms two different complexes with N, the N^0 -P complex and the N–RNA–P complex. These interactions are independent of each other, and the recent structural characterization of these two complexes reveals the molecular origin of the dual behavior of P.

5.1. The N–RNA–P complex

The attachment of the L protein onto the N–RNA template requires the presence of P (Mellon and Emerson, 1978), suggesting that one function of P in the polymerase complex is to attach the L protein to its N–RNA template. For both VSV and RAV, the interaction between P and L was mapped to the N-terminal region of P (Castel et al., 2009; Emerson and Schubert, 1987) and the C-terminal region of L (Castel et al., 2009; Chenik et al., 1998), and

the interaction between P and the N–RNA complexes was mapped to the C-terminal domain of P (Emerson and Schubert, 1987; Jacob et al., 2001; Takacs et al., 1993) and the C-terminal domain of N (Kouznetsoff et al., 1998; Schoehn et al., 2001). For both viruses, only the C-terminal domain of P is required for the attachment onto the N–RNA template (Green et al., 2000; Mavrakakis et al., 2004), although the dimerization of P could modulate the affinity of these two proteins.

For VSV, a complex between P_{CTD} and an N–RNA rings containing 10 protomers of N was produced by co-expressing both proteins, and the structure was solved by X-ray crystallography showing one P_{CTD} bound to each N protomer (Green and Luo, 2009) (Fig. 4A). For RAV, titration experiments showed that only two P_{CTD} could bind simultaneously to an N–RNA rings containing 10 or 11 protomers of N and that the binding of these P_{CTD} is non-cooperative with a dissociation constant of 160 ± 20 nM (Ribeiro et al., 2009). A model of the complex between an N–RNA ring and P_{CTD} was obtained by a flexible cross-docking approach that used biochemical and small-angle X-ray scattering (SAXS) data as constraints (Fig. 4B). A previous analysis of mutations in the P_{CTD} of Mokola virus revealed the existence of two important patches of amino acid residues on opposite faces of the P_{CTD} that are critical for the interaction with N, a hydrophobic pocket with an exposed tryptophan residue (W-hole) and a patch of positive residues (Jacob et al., 2001; Mavrakakis et al., 2004). In accordance, the modeling procedure included the selection of models that maximize the number of contacts between these residues of P_{CTD} and N (Ribeiro et al., 2009).

In the best model of the RAV complex, like in the VSV structure, each P_{CTD} is bound at the C-terminal top of one N protomer (N_i) (Fig. 4A and B). The C-terminal helix of P_{CTD} lies along helix α_{14} of N_{CTD} but, because P_{CTD} of VSV is shorter than that of RAV, the helix is oriented in the opposite direction (in VSV, the C-terminal side of the α -helix points toward the outside of the ring, and in RAV, it points toward the centre of the ring), while in both complexes, the N-terminal extremity of P_{CTD} points toward the centre of the ring (Fig. 4C and D). Also in both complexes, the N_{CT}-loops of protomer N_i and N_{i+1} act like a pair of pincers that grasp the P_{CTD} domain (Fig. 4C and D). In the VSV complex, the N_{CT}-loop is slightly shifted upon binding to P_{CTD} and several side-chains are reoriented, while in the RAV model, a large part of the flexible N_{CT}-loop that was not visible in the N–RNA complex alone interacted with P_{CTD}.

The binding site of P_{CTD} is distributed over two adjacent N protomers and thus requires the polymerization of N, nicely explaining why P_{CTD} can bind only to polymerized N–RNA complexes but not to the isolated N⁰ (Mavrakakis et al., 2006). A main difference is that in VSV N–RNA complex, every N is capable of simultaneously accommodating a P_{CTD} whereas in the RAV complex, the N_{CT} loop on each side of P_{CTD} adopted different conformations and the N-terminal extremity of P_{CTD} protruded in the centre of the ring, thus preventing the binding of P_{CTD} on adjacent N molecules. Also, the N_{CT}-loop of RAV is longer than that of VSV, and the two flanking loops bury a larger part of P_{CTD} surface area in the RAV complex than in the VSV complex (950 \AA^2 in the VSV complex and 1700 \AA^2 in the RAV complex). In the RAV model, hydrophobic residues of the N_{CT}-loop of the N_i protomer inserts in the W-hole, while acidic residues of the N_{CT}-loop of the N_{i+1} protomer form salt-bridges with the basic patches on the opposite face of P_{CTD}. In VSV, the N_{CT} loops of protomers N_i and N_{i+1} interact with P_{CTD} mainly by forming side-chain H-bonds, and a hydrophobic cavity at the surface of P_{CTD} that was postulated to be a potential binding site is capped by the N_{CT}-loop of protomer N_{i+1} but no side-chain of N penetrates within this cavity. In the VSV complex, however, an additional stabilization of the complex may result from the interactions between P_{CTD}s bound to the adjacent protomers (Fig. 4C).

In the N–RNA complexes of RAV and VSV, the RNA bases are completely enwrapped by N_{NTD} and N_{CTD} (Albertini et al., 2006b;

Green et al., 2006). The binding of P_{CTD} to the N–RNA ring revealed no significant conformational change in N that would suggest that binding of P triggers the opening of the N jaws in order to allow the RNA polymerase access to the RNA (Green and Luo, 2009; Ribeiro et al., 2009). There is, however, no evidence that the circular N–RNA complexes are active templates for RNA synthesis, and it is possible that ring closure blocks a potential conformational change. Alternatively, the conformational change required for giving access for the polymerase to the RNA could be induced by the polymerase itself.

Two different models have been proposed to explain how P keeps L attached to the N–RNA complex and how the polymerase complex moves along its template. A first model posits that P progresses along the template by a cartwheeling mechanism in which each C-terminal domain alternatively associates with and dissociates from N (Curran, 1998; Kolakofsky et al., 2004). This model was proposed on the basis of results obtained for Sendai virus (SEV), a prototype member of the *Paramyxoviridae* (Curran, 1998). The P protein of paramyxoviruses shares a similar modular organization with that of the rhabdoviruses (Gérard et al., 2009; Karlin et al., 2003), except that SEV P forms a tetramer rather than a dimer (Tarbouriech et al., 2000). For SEV, the affinity of a single P_{CTD} for the N–RNA template is weak, in the micromolar range ($60 \pm 20 \mu\text{M}$) (Houben et al., 2007), and both the central tetramerization domain and C-terminal domain of P are required for binding to the N–RNA template in pull-down experiments performed at sub-micromolar concentrations of P and for efficient RNA synthesis (Curran, 1998). For rhabdoviruses, however, several results are in apparent contradiction with the cartwheeling model. Firstly, a mutant of RAV P deleted from its central dimerization domain and that remained functional in transcription (Jacob et al., 2001) was found to behave as a monomer in solution (Gérard et al., 2009). The oligomerization of P is not required for proper motion of the polymerase, but a monomer cannot cartwheel along the template. Secondly, a single RAV P_{CTD} is able to bind to the N–RNA rings with a good affinity (160 ± 20 nM) and the kinetics of association and dissociation of this complex on the second time range are incompatible with a mechanism in which P_{CTD}s continuously associate and dissociate to progress along the N–RNA template at a rate compatible with the rate of RNA transcription and replication (Ribeiro et al., 2009). A second model proposes that P molecules are bound at regular intervals along the N–RNA complex and that L jumps between bound Ps, helped in this process by the high flexibility of P (Ribeiro et al., 2009). According to this model, P molecules remain continuously attached to the N–RNA complex in the virion and in the cytoplasm of infected cells, and the oligomerization state of P is selected to obtain the adequate affinity for its partner. The number of P molecules bound to the N–RNA template in an active transcription/replication complex is unknown, but because of the dimensions and flexibility of the P protein, the binding of one P dimer every five N protomers is sufficient to allow one P dimer to catch the L protein from the previous P dimer and to transfer it to the next one. This model needs now to be tested.

5.2. The N⁰–P complex

The encapsidation of newly synthesized antigenomes and genomes is an essential step in the replication process that requires the continuous production of nascent N molecules in a soluble, RNA-free form (Masters and Banerjee, 1988; Peluso and Moyer, 1988). It has been established that N first forms a soluble complex with P and is then transferred from P to the newly made RNA molecules and that the growing NC assembles unidirectionally from the 5' extremity to the 3' extremity of the RNA (Blumberg and Kolakofsky, 1981; Dillon and Gupta, 1988). How the transfer occurs and is regulated are questions that need to be addressed. During

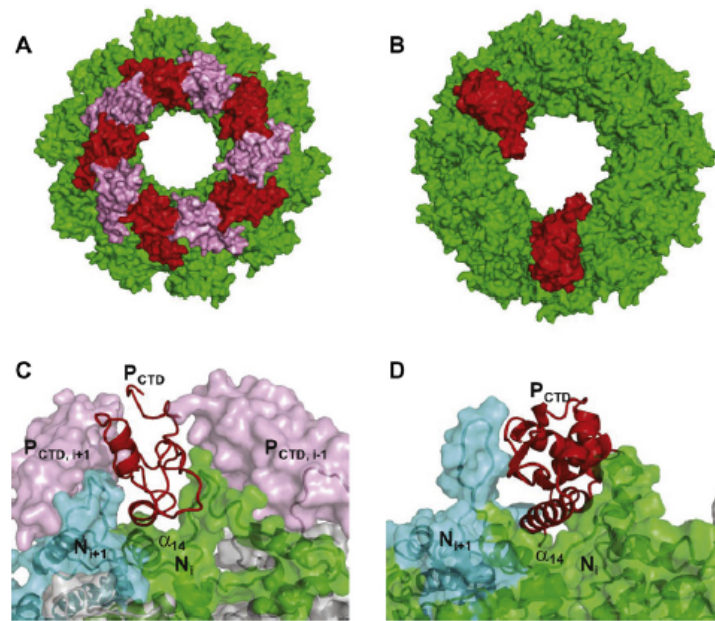


Fig. 4. Structure of the N-RNA-P_{CTD} complexes. (A) Space-filling model of VSV decameric N-RNA-P_{CTD} complex (PDB ID: 3hhz). N protomers are shown in green. The P_{CTD} are shown alternatively in red and pink. (B) Space-filling model of RAV N-RNA-P_{CTD} complex. The model, generated by molecular modeling, contains two P_{CTD} (in red) bound to a 11-mer N-RNA ring (in green). (C) Side-view of the N-RNA-P_{CTD} complex of VSV. The complex is seen from outside the ring. P_{CTD} is shown in red. The N protomer on which P_{CTD} lies is shown in green and is labeled N_i. The adjacent protomer on the 3' side of the N_i protomer is shown in gray (N_{i+1}) and that on the 5' side is shown in cyan (N_{i-1}). The adjacent P_{CTD} are shown in pink. (D) Side-view of the N-RNA-P_{CTD} complex of RAV. The N-RNA-P_{CTD} is shown in the same orientation, at the same scale and with the same color scheme than that of VSV. In RAV complex, only two P_{CTD} binds to a N-RNA ring and they are not interacting with each other.

the replication process, RNA synthesis is initiated at the 3' end of the template and the commitment to synthesize full-length RNA antigenomes and genomes rather than mRNAs must be taken by the polymerase complex before or at the moment where it reaches the end of the leader or trailer sequences, respectively. The concentration of N⁰-P plays a role in switching from transcription to replication (Arnheiter et al., 1985; Gupta and Banerjee, 1997) and the transfer of N could be controlled by its relative affinity for P and for encapsidation signals present in the terminal untranslated regions (Blumberg et al., 1981). Conversely, it could depend on the modification of the polymerase complex (Qanungo et al., 2004).

Because of the difficulty to prepare the N⁰-P complex (Mavrakis et al., 2003), this complex escaped structural characterization for a long time. The region of P involved in the formation of the N⁰-P complex was localized to a short segment of its N-terminal region (Chen et al., 2007; Mavrakis et al., 2006). For RAV, a fragment containing residues 4–40 of P was identified by limited proteolysis (Mavrakis et al., 2006), while for VSV, residues 11–30 were found to be essential for keeping N soluble (Chen et al., 2007). For VSV, this N⁰-binding region encompasses the transient α -helices that were discovered by NMR spectroscopy in the N-terminal region of P (Leyrat et al., 2011a).

Recently, a N⁰-P complex was reconstituted between an armless N, lacking the 21 N-terminal residues, N _{Δ 21}, and either a peptide corresponding to the 60 first residues of P (P₆₀) or a dimer of full length P (Leyrat et al., 2011c). These purified complexes are soluble, lack RNA and correspond to a heterodimer in the case of P₆₀ or to a mixture of hetero-trimers (1N:2P) and hetero-tetramers (2N:2P) in the case of full length P (Leyrat et al., 2011c). The structural characterization of the N _{Δ 21}-P₆₀ complex in solution by SAXS revealed a particle of similar size and shape as those of an N monomer extracted from the N-RNA crystal structure, while NMR spectroscopy showed that only residues 6–40 of P are directly

involved in the complex with N⁰, whereas the flanking regions (aa 1–5 and 41–60) remain flexible in the complex (Leyrat et al., 2011c). The N _{Δ 21}-P₆₀ complex was crystallized in the form of a multimeric circular complex, in which the N molecules are arranged as in the N-RNA complex (Leyrat et al., 2011c). In each protomer, residues 6–35 of P were clearly visible in the crystal structure. Residues 17–30 form an amphipathic α -helix that is docked on one side of N in a groove at the interface between N_{NTD} and N_{CTD}. The peptide is oriented in a roughly perpendicular direction relative to the RNA binding groove (Fig. 5A and B), and none of the basic residues of VSV N directly contacting the RNA backbone in the N-RNA complex are involved in the interaction with the MoRE of P. The second transient α -helix (aa 25–31) present in the isolated N⁰-binding region is stabilized upon binding to its N⁰ partner and extends from residue 17 to residue 31. In contrast with RNA, P binds mainly through intermolecular hydrophobic interactions although the bipolar distribution of charges on the surface of P, with a positive pole at its N-terminus that docks on the negatively charged backside of N_{CTD} and a negative pole at its C-terminus that localizes in the highly positive RNA binding groove, could play a role in orientating P before binding. The hinge region of N (aa 200–300) is more conserved among VSV isolates as well as between the more evolutionary distant VSV and RAV (Luo et al., 2007) than the rest of the protein (Fig. 5C) and participates in the regulation of the transcription and replication (Nayak et al., 2009) in accordance with its central roles in binding both the RNA and MoRE of P. The conservation of several hydrophobic residues lining the binding groove of P, as well as of Arg³¹², which forms a salt bridge with an acidic residue of P, suggests that a similar N⁰-P complex forms in RAV.

The N⁰-binding region of P must prevent two processes, the polymerization of N and the non-specific encapsidation of host cell RNAs (Chen et al., 2007; Howard and Wertz, 1989; Masters and Banerjee, 1988; Mavrakis et al., 2006; Peluso and Moyer, 1988),

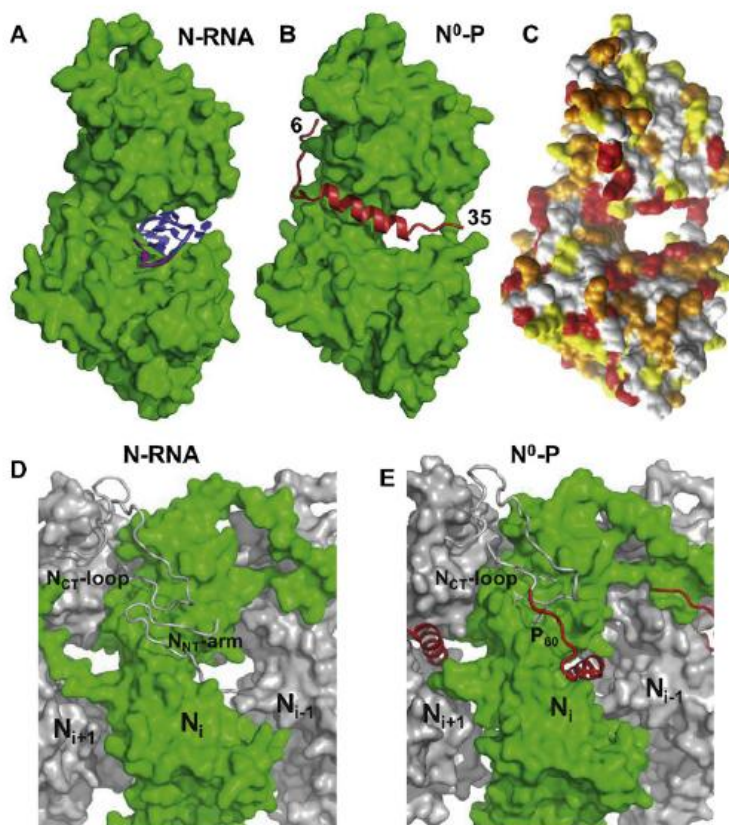


Fig. 5. Structure of the N⁰-P complex. (A) Representations of one protomer from VSV N-RNA complex (PDB ID: 2gtt). The N protomer is shown as a space filling model in green and the RNA is shown in blue. The 3' terminal nucleotide of the RNA molecule is shown in magenta. (B) Representations of one protomer from the N⁰_{Δ21}-P₆₀ complex (PDB ID: 3pmk). The N protomer is shown as a space filling model in green and P₆₀ is shown as a cartoon representation in red. The N- and C-terminal extremities of P₆₀ that are visible in the crystal are labeled. (C) Sequence conservation between VSV and RAV N mapped on a protomer of VSV N. Conserved residues are shown in red, similar residues are shown in orange and yellow, according to clustalW2 similarity scheme (Larkin et al., 2007). (D, E) Close-up of the interactions between exchangeable sub-domains in the circular N-RNA and N⁰_{Δ21}-P₆₀ complexes of VSV. In the VSV N-RNA complex (D), the N_{NT}-arm of protomer N_{i-1} contacts the N_{CT}-loop of protomer N_{i+1} while both sub-domains are docked on the back-side of protomer N_i (in green). In the N⁰_{Δ21}-P₆₀ complex (E), the N-terminal extremity of P₆₀ docks on protomer N_i (in green) at the position of the N_{NT}-arm of protomer N_{i-1} and contacts the N_{CT}-loop of protomer N_{i+1}. These representations suggest that P₆₀ interferes with the assembly of N in the absence of RNA.

which are independent of each other (Zhang et al., 2008). The structure of the N⁰_{Δ21}-P₆₀ complex unveils how P fulfills its two chaperone functions. The C-terminal extremity of the MoRE overlaps with the RNA binding groove and directly competes with RNA binding (Fig. 5A and B), whereas its N-terminal part interferes with the binding of the N_{NT} arm of the neighboring N_{i+1} molecule and thereby prevents the polymerization of N (Fig. 5C and D). Also, the binding of the MoRE of P notably reduces the positive surface potential on one side of the N molecule, and may thereby affect the side-by-side interaction with another N molecule. In addition, the flanking parts of the MoRE, by fluctuating between multiple conformations, may also contribute to the chaperone activities of P by acting as entropic bristles that repel incoming RNA or N molecules or by masking their binding interfaces. In the complex with the full-length P dimer, these effects may be exacerbated by the high flexibility and the bulkiness of the remainder of the protein.

These results suggested a hypothesis to explain how P controls the encapsidation of a newly synthesized RNA molecule during the replication of the viral genome (Fig. 1B). Little is known about the mechanism by which N is transferred from N⁰-P to a RNA molecule or about the role played by the polymerase complex in this process. The N⁰-P complex conserves its ability to bind to the N-RNA

complex (Gupta and Banerjee, 1997) more likely through the binding of P_{CTD} and it is possible that the P is transferred to the growing NC from a bound N⁰-P complex. However, the recent experiments showed that the N_{NT}-arm stabilizes the multimeric assembly of N molecules since its removal allowed reversing the assembly process (Leyrat et al., 2011c) and suggested that the multimeric N-RNA complex is more stable than the monomeric N⁰-P complex. The assembly of the growing NC could thus simply be driven by the higher stability of the multimeric N complex. In the soluble N⁰-P complex, P blocks the backside groove of N in which the N_{NT} arm of a neighboring N molecule binds and thereby prevents the assembly of soluble N molecules in the absence of RNA. Upon transfer of N onto a RNA molecule, P is released and the backside groove of the last added N molecule is liberated and becomes available for accepting the N_{NT}-arm of the next incoming N molecule (Fig. 1B).

In addition to its role in RNA encapsidation, the binding of the N-terminal region of P to N may also provide a mechanism for the initiation of (+)RNA synthesis at the genome 3' end and of (-)RNA synthesis at the antigenome 3' end (Leyrat et al., 2011c). Encapsidated RNA genome and antigenome are completely covered by the nucleoprotein and are not accessible to the RNA polymerase. However, the first N molecule at the 3' extremity of nucleocapsids

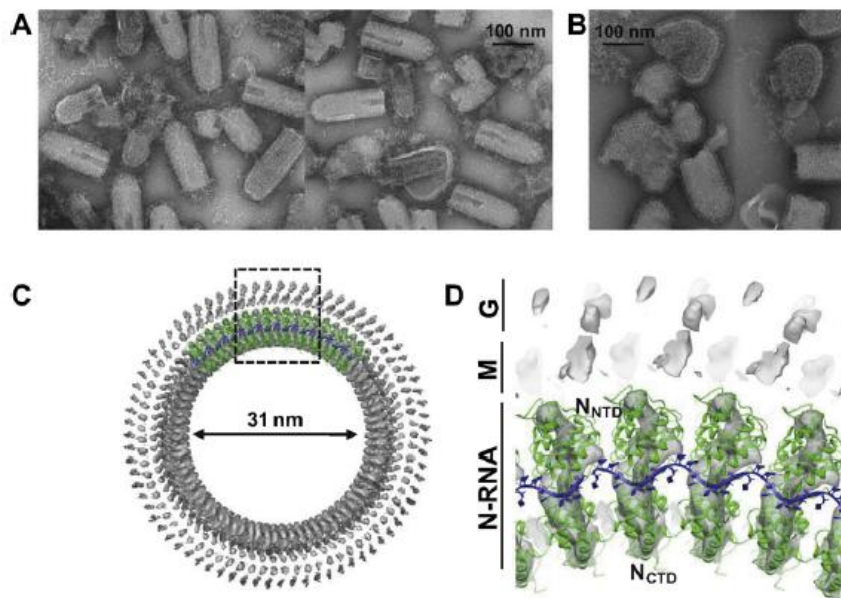


Fig. 6. Structure of the virion. (A) Typical negative staining electron microscopy micrographs of VSV virions. Negative staining was done with sodium silicotungstate. (B) Typical negative staining electron microscopy micrograph of RAV virions. (C) Transverse section in the cryo-electron microscopy density map of a reconstructed trunk of a VSV particle. The electron density map corresponding to 10 turns of 37.5 N molecules was reconstructed from the electron density map deposited in the Electron Microscopy Databank (EMD ID: EMD.1663). It is displayed at a threshold of 2.9σ above the mean. The trunk is seen from its basal face. A decamer of N adapted from the high resolution structure of the N–RNA complex (PDB ID: 2wyy) was fitted into the electron density map with the program VEDA (graphical interface of UROX, Navaza et al., 2002) and is shown as a cartoon representation. N protomers are shown in green and the RNA molecule in blue. (D) Close up of the transverse section. A slice 5 nm thick of the reconstructed electron density map shows the organization of the N molecules in the trunk of the virion. The volumes of electron density on the outside of the virion correspond to the M protein (intermediate layer) and to the cytoplasmic tail of the G protein (outer layer).

exposes its binding site for the N-terminal MoRE of P. By binding to this surface, P may destabilize the N–RNA complex sufficiently to displace several nucleotides from the first N protomer and allow the polymerase access to the 3' extremity of the viral RNA.

6. The L subunit of the viral polymerase

The L protein is a 250 kDa multi-enzymatic protein, which catalyzes RNA synthesis as well as mRNA capping, methylation and polyadenylation. Its sequence is well conserved among all non-segmented (–)RNA viruses, and sequence alignments revealed the existence of six conserved regions, numbered I–VI (Poch et al., 1990). The conserved region III contains different motives involved in the RNA-dependent RNA polymerase activity, while the conserved regions V and VI carry out mRNA capping. The capping mechanism of the non-segmented (–)RNA viruses is different from that of their eukaryotic hosts. GDP, rather than GMP, is transferred to a 5' monophosphate RNA by a polyribonucleotidyltransferase (PRNTase) (Ogino and Banerjee, 2007, 2008) that was mapped to region V (Li et al., 2008). This reaction proceeds through the formation of a covalent intermediate between the 5' monophosphate mRNA and a histidine residue of L (Ogino et al., 2010). After its formation, the cap is methylated successively at ribose 2'-O and guanine-N-7 positions by region VI (Grdzlishvili et al., 2005; Li et al., 2005, 2006; Rahmeh et al., 2009; Zhang et al., 2010). The polyadenylation also depends on region III and is performed by slippage of the L protein on a U-rich region present at the 5' end of each gene (Barr et al., 1997).

The first glimpse at the molecular structure of this large protein has been obtained recently by negative-staining electron microscopy (Rahmeh et al., 2010). Different regions of the

protein were localized in the structure by deletion studies. Conserved domains I–IV encompassing the site of RNA synthesis were localized in a ring-like domain, similar to that found in other RNA polymerases (Ferrer-Orta et al., 2006). Conserved regions V and VI involved in cap formation and methylation were mapped to three globular domains linked by flexible linkers to the ring-like core (Rahmeh et al., 2010). The binding of P was also shown to induce a conformational change in the L protein that results in the rearrangement of the globular domains at its C-terminus and to promote the formation of dimers (Rahmeh et al., 2010). The activities of L are controlled by P (Das et al., 1997; Hwang et al., 1998; Pattnaik et al., 1997) and this rearrangement of L may represent a mechanism of this regulation. In addition, the various activities catalyzed by the L protein are coordinated and regulated (Li et al., 2009; Stillman and Whitt, 1999) and it has been shown that a minimal length of the transcript is required to allow its capping, suggesting that the spatial organization of the different domains of L could control its activities (Tekes et al., 2011).

7. The viral particle

Animal rhabdoviruses have an overall bullet shape with one conical end and one flat end, while plant rhabdoviruses have a bacillus shape with two conical ends. Fig. 6A and B shows negative-staining electron micrographs of RAV and VSV particles. Recently, the cryo-electron microscopy (EM) reconstruction of the virion of VSV, revealed the molecular organization of the viral proteins in three concentric layers (Ge et al., 2010). The outer shell is a host cell-derived lipid membrane decorated with trimeric G spikes. The middle layer is formed by M proteins which are organized in a regular network onto which the trimeric G spikes attach. The NC

constitutes the inner shell and has itself a bullet shape particle (Ge et al., 2010). Its striated appearance comes from the regular arrangement of multimeric N (Schoehn et al., 2001). In the trunk of the bullet, N packs the genomic RNA into a regular helical N–RNA complex that comprises 37.5 subunits of N per turn (Fig. 6C). The N molecules are tilted upward by 27° from the horizontal plane. The tip of the bullet is formed of seven turns containing varying numbers of subunits, in which the tilt from the horizontal plane draws closer to that in the recombinant N–RNA rings. The upper turn which encapsidates the 5' extremity of the RNA molecule and may represent the nucleation centre from which the particle assembles, resembles a decameric ring, suggesting that a ten-membered ring or spiral corresponds to an optimal side-by-side orientation between adjacent N subunits. The N subunits are oriented with their N-terminal domain pointing toward the exterior of the virion and their C-terminal domain pointing toward the centre of the particle. The N_{NTD} connects the NC to the M network providing rigidity to the assembly, whereas on the inside of the NC the N_{CTD} is available for binding P molecules. A similar organization has been found for the RAV particle, although the RAV nucleocapsid forms helical turn of about 53 protomers (Albertini et al., 2008; Iseni et al., 1998; Schoehn et al., 2001).

The bullet-like particle contains a cavity of 31 nm in diameter and about 150 nm in length that was previously described as a “cigar” (Barge et al., 1993). This central cavity is large enough to accommodate the 450 molecules of P, 55 molecules of L protein and about 500 M molecules that are involved in the middle shell (Thomas et al., 1985), but its content could not be reconstructed from the cryo-electron micrographies because its components are not organized in a highly symmetrical fashion (Ge et al., 2010). The exact position of the L/P complexes in the viral particle remains, however, unknown. They may remain attached to the NC at randomly distributed localizations as they were captured by condensation of the NC during budding and as they are usually represented in drawings of rhabdovirus particles (Jackson et al., 2005; Schnell et al., 2010; Wertz, 1978). Alternatively, the M and perhaps P proteins may form a scaffold onto which the polymerase molecules await the unpacking of the NC in the host-cell cytoplasm to successively initiate RNA synthesis at the 3' end of the genome. The M protein has the ability to self-assemble reversibly into filaments that share some structural features with the observed “cigar” (Gaudin et al., 1995), and the P protein was shown to form filamentous aggregates upon incubation at 37 °C (Gérard et al., 2007).

8. Future work

The recent progress made in the preparation of highly purified components of the rhabdovirus transcription/replication complex and in the characterization of their stoichiometry, size, shape and structure opens new avenues for deciphering the molecular mechanisms of this machinery and the subtleties of their regulations. In the near future, it should be possible to reconstitute a functional system from purified components and to test hypotheses concerning the motion of the polymerase on its template, the exact role of the dimerization of P or the effect of concentration of N⁰-P and of the phosphorylation of P on the switch between the transcription and replication of the genome. The structural characterization of the L protein and of its complex with P is a long-sought grail that may find its epilogue in the near future. One can also hope that these new structural data will lead to the development of new antiviral therapies. Finally, the recent progresses made in *in vivo* imaging should also allow us to investigate the organization of the rhabdovirus transcription/replication machinery in its cellular context and to a better understanding of the connections with host cell components.

Acknowledgments

Research activities in the authors' laboratory were supported by grants from the French ANR (ANR-07-001-01 (ANRAGE)), the FINOVI foundation and Lyonbiopôle. Ivan Ivanov was supported by a PhD fellowship from the ILL and Filip Yabukarski was supported by a MENRT fellowship from the French government. We thank Guy Schoehn and Leandro Estrozi for their help in preparing Figs. 2 and 6, our colleagues for extensive discussions and the Partnership for Structural Biology for the excellent structural biology environment.

References

- Abraham, G., Banerjee, A.K., 1976. Sequential transcription of the genes of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. U.S.A.* 73, 1504–1508.
- Abraham, G., Rhodes, D.P., Banerjee, A.K., 1975. The 5' terminal structure of the methylated mRNA synthesized *in vitro* by vesicular stomatitis virus. *Cell* 5, 51–58.
- Ahmed, M., McKenzie, M.O., Puckett, S., Hojnacki, M., Poliquin, L., Lyles, D.S., 2003. Ability of the matrix protein of vesicular stomatitis virus to suppress beta interferon gene expression is genetically correlated with the inhibition of host RNA and protein synthesis. *J. Virol.* 77, 4646–4657.
- Albertini, A.A., Clapier, C.R., Wernimont, A.K., Schoehn, G., Weissenhorn, W., Ruigrok, R.W., 2006a. Isolation and crystallization of a unique size category of recombinant rabies virus nucleoprotein–RNA rings. *J. Struct. Biol.* 158, 129–133.
- Albertini, A.A., Schoehn, G., Weissenhorn, W., Ruigrok, R.W., 2008. Structural aspects of rabies virus replication. *Cell Mol. Life Sci.* 65, 282–294.
- Albertini, A.A., Wernimont, A.K., Muziol, T., Ravelli, R.B., Clapier, C.R., Schoehn, G., Weissenhorn, W., Ruigrok, R.W., 2006b. Crystal structure of the rabies virus nucleoprotein–RNA complex. *Science* 313, 360–363.
- Arnheiter, H., Davis, N.L., Wertz, G., Schubert, M., Lazzarini, R.A., 1985. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell* 41, 259–267.
- Ball, L.A., White, C.N., 1976. Order of transcription of genes of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. U.S.A.* 73, 442–446.
- Barge, A., Gaudin, Y., Coulon, P., Ruigrok, R.W., 1993. Vesicular stomatitis virus M protein may be inside the ribonucleocapsid coil. *J. Virol.* 67, 7246–7253.
- Barr, J.N., Wertz, G.W., 2001. Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination. *J. Virol.* 75, 6901–6913.
- Barr, J.N., Whelan, S.P., Wertz, G.W., 1997. cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *J. Virol.* 71, 8718–8725.
- Basak, S., Mondal, A., Polley, S., Mukhopadhyay, S., Chattopadhyay, D., 2007. Reviewing Chandipura: a vesiculovirus in human epidemics. *Biosci. Rep.* 27, 275–298.
- Blumberg, B.M., Giorgi, C., Kolakofsky, D., 1983. N protein of vesicular stomatitis virus selectively encapsidates leader RNA *in vitro*. *Cell* 32, 559–567.
- Blumberg, B.M., Kolakofsky, D., 1981. Intracellular vesicular stomatitis virus leader RNAs are found in nucleocapsid structures. *J. Virol.* 40, 568–576.
- Blumberg, B.M., Leppert, M., Kolakofsky, D., 1981. Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication. *Cell* 23, 837–845.
- Canter, D.M., Perrault, J., 1996. Stabilization of vesicular stomatitis virus L polymerase protein by P protein binding: a small deletion in the C-terminal domain of L abrogates binding. *Virology* 219, 376–386.
- Castel, G., Chteoui, M., Caignard, G., Prehaud, C., Mehous, S., Real, E., Jallet, C., Jacob, Y., Ruigrok, R.W., Tordo, N., 2009. Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. *J. Virol.* 83, 10808–10820.
- Chen, M., Ogino, T., Banerjee, A.K., 2007. Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N-terminus of P that are involved in N⁰-P complex formation and encapsidation of viral genome RNA. *J. Virol.* 81, 13478–13485.
- Chen, Z., Green, T.J., Luo, M., Li, H., 2004. Visualizing the RNA molecule in the bacterially expressed vesicular stomatitis virus nucleoprotein–RNA complex. *Structure* 12, 227–235.
- Chenik, M., Schnell, M., Conzelmann, K.K., Blondel, D., 1998. Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J. Virol.* 72, 1925–1930.
- Chuang, J.L., Perrault, J., 1997. Initiation of vesicular stomatitis virus mutant polR1 transcription internally at the N gene *in vitro*. *J. Virol.* 71, 1466–1475.
- Cleaveland, S., Fevre, E.M., Kaare, M., Coleman, P.G., 2002. Estimating human rabies mortality in the United Republic of Tanzania from dog bite injuries. *Bull. World Health Organ.* 80, 304–310.
- Connor, J.H., Lyles, D.S., 2002. Vesicular stomatitis virus infection alters the eIF4F translation initiation complex and causes dephosphorylation of the eIF4E binding protein 4E-BP1. *J. Virol.* 76, 10177–10187.
- Connor, J.H., Lyles, D.S., 2005. Inhibition of host and viral translation during vesicular stomatitis virus infection. eIF2 is responsible for the inhibition of viral but not host translation. *J. Biol. Chem.* 280, 13512–13519.
- Curran, J., 1998. A role for the Sendai virus P protein trimer in RNA synthesis. *J. Virol.* 72, 4274–4280.

- Das, T., Pattnaik, A.K., Takacs, A.M., Li, T., Hwang, L.N., Banerjee, A.K., 1997. Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. *Virology* 238, 103–114.
- Delmas, O., Assenberg, R., Grimes, J.M., Bourhy, H., 2010. The structure of the nucleoprotein binding domain of *Lyssavirus* phosphoprotein reveals a structural relationship between the N-RNA binding domains of *Rhabdoviridae* and *Paramyxoviridae*. *RNA Biol.* 7, 322–327.
- Dietzschold, B., Li, J., Faber, M., Schnell, M., 2008. Concepts in the pathogenesis of rabies. *Future Virol.* 3, 481–490.
- Dillon, P.J., Gupta, K.C., 1988. Early steps in the assembly of vesicular stomatitis virus nucleocapsids in infected cells. *J. Virol.* 62, 1582–1589.
- Ding, H., Green, T.J., Lu, S., Luo, M., 2006. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* 80, 2808–2814.
- Ding, H., Green, T.J., Luo, M., 2004. Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). *Acta Crystallogr. D: Biol. Crystallogr.* 60, 2087–2090.
- Durbin, J.E., Hackenmiller, R., Simon, M.C., Levy, D.E., 1996. Targeted disruption of the mouse *Stat1* gene results in compromised innate immunity to viral disease. *Cell* 84, 443–450.
- Emerson, S.U., 1982. Reconstitution studies detect a single polymerase entry site on the vesicular stomatitis virus genome. *Cell* 31, 635–642.
- Emerson, S.U., Schubert, M., 1987. Location of the binding domains for the RNA polymerase I and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. U.S.A.* 84, 5655–5659.
- Emerson, S.U., Yu, Y., 1975. Both NS and L proteins are required for *in vitro* RNA synthesis by vesicular stomatitis virus. *J. Virol.* 15, 1348–1356.
- Ferrer-Orta, C., Arias, A., Escarmis, C., Verdagué, N., 2006. A comparison of viral RNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* 16, 27–34.
- Follett, E.A., Pringle, C.R., Wunner, W.H., Skehel, J.J., 1974. Virus replication in enucleate cells; vesicular stomatitis virus and influenza virus. *J. Virol.* 13, 394–399.
- Fuxreiter, M., Simon, I., Friedrich, P., Tompa, P., 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* 338, 1015–1026.
- Gaudin, Y., Barge, A., Ebel, C., Ruigrok, R.W., 1995. Aggregation of VSV M protein is reversible and mediated by nucleation sites: implications for viral assembly. *Virology* 206, 28–37.
- Gaudin, Y., Tuffereau, C., Durrer, P., Brunner, J., Flamand, A., Ruigrok, R., 1999. Rabies virus-induced membrane fusion. *Mol. Membr. Biol.* 16, 21–31.
- Ge, P., Tsao, J., Schein, S., Green, T.J., Luo, M., Zhou, Z.H., 2010. Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* 327, 689–693.
- Gérard, F.C.A., Ribeiro, E., Albertini, A., Zaccari, G., Ebel, C., Ruigrok, R., Jamin, M., 2007. Unphosphorylated *Rhabdoviridae* phosphoproteins form elongated dimers in solution. *Biochemistry* 46, 10328–10338.
- Gérard, F.C.A., Ribeiro, E.A., Leyrat, C., Ivanov, I., Blondel, D., Longhi, S., Ruigrok, R.W.H., Jamin, M., 2009. Modular organization of rabies virus phosphoprotein. *J. Mol. Biol.* 388, 978–996.
- Grzelishvili, V.Z., Smallwood, S., Tower, D., Hall, R.L., Hunt, D.M., Moyer, S.A., 2005. A single amino acid change in the L-polymerase protein of vesicular stomatitis virus completely abolishes viral mRNA cap methylation. *J. Virol.* 79, 7327–7337.
- Green, T.J., Luo, M., 2009. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11721–11726.
- Green, T.J., Macpherson, S., Qiu, S., Lebowitz, J., Wertz, G.W., Luo, M., 2000. Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J. Virol.* 74, 9515–9524.
- Green, T.J., Rowse, M., Tsao, J., Kang, J., Ge, P., Zhou, Z.H., Luo, M., 2010. Access of RNA encapsidated in the nucleocapsid of vesicular stomatitis virus. *J. Virol.* 85, 2714–2722.
- Green, T.J., Zhang, X., Wertz, G.W., Luo, M., 2006. Structure of the vesicular stomatitis virus nucleoprotein–RNA complex. *Science* 313, 357–360.
- Gupta, A.K., Banerjee, A.K., 1997. Expression and purification of vesicular stomatitis virus N–P complex from *Escherichia coli*: role in genome RNA transcription and replication *in vitro*. *J. Virol.* 71, 4264–4271.
- Hanlon, C.A., Kuzmin, I.V., Blanton, J.D., Weldon, W.C., Manangan, J.S., Rupprecht, C.E., 2005. Efficacy of rabies biologics against new *Lyssaviruses* from Eurasia. *Virus Res.* 111, 44–54.
- Harty, R.N., Paragas, J., Sudol, M., Palese, P., 1999. A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding. *J. Virol.* 73, 2921–2929.
- Heinrich, B.S., Cureton, D.K., Rahmeh, A.A., Whelan, S.P., 2010. Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. *PLoS Pathog.* 6, e1000958.
- Hemachudha, T., Laothamatas, J., Rupprecht, C.E., 2002. Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges. *Lancet Neurol.* 1, 101–109.
- Hercyk, N., Horikami, S.M., Moyer, S.A., 1988. The vesicular stomatitis virus L protein possesses the mRNA methyltransferase activities. *Virology* 163, 222–225.
- Hornung, V., Elegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K.K., Schlee, M., Endres, S., Hartmann, G., 2006. 5'-Triphosphate RNA is the ligand for RIG-I. *Science* 314, 994–997.
- Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R.W., Blanchard, L., 2007. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase–nucleocapsid interactions within the paramyxovirus family. *J. Virol.* 81, 6807–6816.
- Howard, M., Wertz, G., 1989. Vesicular stomatitis virus RNA replication: a role for the NS protein. *J. Gen. Virol.* 70, 2683–2694.
- Hwang, L.N., Englund, N., Pattnaik, A.K., 1998. Polyadenylation of vesicular stomatitis virus mRNA dictates efficient transcription termination at the intercistronic gene junctions. *J. Virol.* 72, 1805–1813.
- Irie, T., Licata, J.M., McGettigan, J.P., Schnell, M.J., Harty, R.N., 2004. Budding of PPxY-containing rhabdoviruses is not dependent on host proteins TGS101 and VPS4A. *J. Virol.* 78, 2657–2665.
- Iseñi, F., Barge, A., Baudin, F., Blondel, D., Ruigrok, R.W., 1998. Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. *J. Gen. Virol.* 79, 2909–2919.
- Iseñi, F., Baudin, F., Blondel, D., Ruigrok, R.W., 2000. Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *RNA* 6, 270–281.
- Ivanov, I., Crepin, T., Jamin, M., Ruigrok, R., 2010. Structure of the dimerisation domain of the rabies virus phosphoprotein. *J. Virol.* 84, 3707–3710.
- Iverson, L.E., Rose, J.K., 1981. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. *Cell* 23, 477–484.
- Jackson, A.O., Dietzgen, R.G., Goodin, M.M., Bragg, J.N., Deng, M., 2005. Biology of plant rhabdoviruses. *Annu. Rev. Phytopathol.* 43, 623–660.
- Jacob, Y., Real, E., Tordo, N., 2001. Functional interaction map of *Lyssavirus* phosphoprotein: identification of the minimal transcription domains. *J. Virol.* 75, 9613–9622.
- Karlin, D., Ferron, F., Canard, B., Longhi, S., 2003. Structural disorder and modular organization in Paramyxoviridae N and P. *J. Gen. Virol.* 84, 3239–3252.
- Knobel, D.L., Cleveland, S., Coleman, P.G., Fevre, E.M., Meltzer, M.I., Miranda, M.E., Shaw, A., Zinsstag, J., Meslin, F.X., 2005. Re-evaluating the burden of rabies in Africa and Asia. *Bull. World Health Organ.* 83, 360–368.
- Kolakofsky, D., Le Mercier, P., Iseñi, F., Garcin, D., 2004. Viral RNA polymerase scanning and the gymnastics of sendai virus RNA synthesis. *Virology* 318, 463–473.
- Kolakofsky, D., Roux, L., Garcin, D., Ruigrok, R.W., 2005. Paramyxovirus mRNA editing, the “rule of six” and error catastrophe: a hypothesis. *J. Gen. Virol.* 86, 1869–1877.
- Kouznetsoff, A., Buckle, M., Tordo, N., 1998. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. *J. Gen. Virol.* 79, 1005–1013.
- Koyama, A.H., 1995. Induction of apoptotic DNA fragmentation by the infection of vesicular stomatitis virus. *Virus Res.* 37, 285–290.
- Lahaye, X., Vidy, A., Pomier, C., Obiang, L., Harper, F., Gaudin, Y., Blondel, D., 2009. Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: evidence that NBs are sites of viral transcription and replication. *J. Virol.* 83, 7948–7958.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lawson, N.D., Stillman, E.A., Whitt, M.A., Rose, J.K., 1995. Recombinant vesicular stomatitis viruses from DNA. *Proc. Natl. Acad. Sci. U.S.A.* 92, 4477–4481.
- Lenard, J., 1999. Host cell protein kinases in nonsegmented negative-strand virus (mononegavirales) infection. *Pharmacol. Ther.* 83, 39–48.
- Leyrat, C., Gérard, F.C., de Almeida Ribeiro Jr., E., Ivanov, I., Ruigrok, R.W., Jamin, M., 2010. Structural disorder in proteins of the *Rhabdoviridae* replication complex. *Protein Pept. Lett.* 17, 979–987.
- Leyrat, C., Jensen, M.R., Ribeiro, E.A., Gérard, F., Ruigrok, R., Blackledge, M., Jamin, M., 2011a. The N⁰-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α -helices. *Protein Sci.* 20, 542–556.
- Leyrat, C., Ribeiro, E.A., Gérard, F.C., Ivanov, I., Ruigrok, R.W.H., Jamin, M., 2011b. Structure, interactions with host-cell and functions of rhabdovirus phosphoprotein. *Future Virol.* 6, 465–481.
- Leyrat, C., Yabukarski, F., Tarbouriech, N., Ribeiro, E.A., Jensen, M.R., Blackledge, M., Ruigrok, R.W., Jamin, M., 2011c. Structure of the vesicular stomatitis virus N⁰-P complex. *PLoS Pathog.* 7, e1002248.
- Li, J., Fontaine-Rodríguez, E.C., Whelan, S.P., 2005. Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity. *J. Virol.* 79, 13373–13384.
- Li, J., Rahmeh, A., Brusci, V., Whelan, S.P., 2009. Opposing effects of inhibiting cap addition and cap methylation on polyadenylation during vesicular stomatitis virus mRNA synthesis. *J. Virol.* 83, 1930–1940.
- Li, J., Rahmeh, A., Morelli, M., Whelan, S.P., 2008. A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. *J. Virol.* 82, 775–784.
- Li, J., Wang, J.T., Whelan, S.P., 2006. A unique strategy for mRNA cap methylation used by vesicular stomatitis virus. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8493–8498.
- Libersou, S., Albertini, A.A., Ouldali, M., Maury, V., Maheu, C., Raux, H., de Haas, F., Roche, S., Gaudin, Y., Lepault, J., 2010. Distinct structural rearrangements of the VSV glycoprotein drive membrane fusion. *J. Cell Biol.* 191, 199–210.
- Lichty, B.D., Power, A.T., Stojdl, D.F., Bell, J.C., 2004. Vesicular stomatitis virus: re-inventing the bullet. *Trends Mol. Med.* 10, 210–216.
- Luo, M., Green, T.J., Zhang, X., Tsao, J., Qiu, S., 2007. Conserved characteristics of the rhabdovirus nucleoprotein. *Virus Res.* 129, 246–251.
- Lynch, S., Kolakofsky, D., 1978. Ends of the RNA within sendai virus defective interfering nucleocapsids are not free. *J. Virol.* 28, 584–589.

- Masters, P.S., Banerjee, A.K., 1988. Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J. Virol.* 62, 2658–2664.
- Mavrakis, M., Iseni, F., Mazza, C., Schoehn, G., Ebel, C., Gentzel, M., Franz, T., Ruigrok, R.W., 2003. Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells. *Virology* 305, 406–414.
- Mavrakis, M., McCarthy, A.A., Roche, S., Blondel, D., Ruigrok, R.W., 2004. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J. Mol. Biol.* 343, 819–831.
- Mavrakis, M., Mehous, S., Real, E., Iseni, F., Blondel, D., Tordo, N., Ruigrok, R.W., 2006. Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology* 349, 422–429.
- Mellon, M.G., Emerson, S.U., 1978. Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus. *J. Virol.* 27, 560–567.
- Moyer, S.A., Banerjee, A.K., 1975. Messenger RNA species synthesized *in vitro* by the virion-associated RNA polymerase of vesicular stomatitis virus. *Cell* 4, 37–43.
- Moyer, S.A., Smallwood-Kent, S., Haddad, A., Prevec, L., 1991. Assembly and transcription of synthetic vesicular stomatitis virus nucleocapsids. *J. Virol.* 65, 2170–2178.
- Navaza, J., Lepault, J., Rey, F.A., Alvarez-Rua, C., Borge, J., 2002. On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr. D: Biol. Crystallogr.* 58, 1820–1825.
- Nayak, D., Panda, D., Das, S.C., Luo, M., Pattnaik, A.K., 2009. Single-amino-acid alterations in a highly conserved central region of vesicular stomatitis virus N protein differentially affect the viral nucleocapsid template functions. *J. Virol.* 83, 5525–5534.
- Ogino, T., Banerjee, A.K., 2007. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol. Cell* 25, 85–97.
- Ogino, T., Banerjee, A.K., 2008. Formation of guanosine(5')tetraphospho(5') adenosine cap structure by an unconventional mRNA capping enzyme of vesicular stomatitis virus. *J. Virol.* 82, 7729–7734.
- Ogino, T., Yadav, S.P., Banerjee, A.K., 2010. Histidine-mediated RNA transfer to GDP for unique mRNA capping by vesicular stomatitis virus RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3463–3468.
- Pattnaik, A.K., Hwang, L., Li, T., Englund, N., Mathur, M., Das, T., Banerjee, A.K., 1997. Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication. *J. Virol.* 71, 8167–8175.
- Patton, J.T., Davis, N.L., Wertz, G.W., 1984. N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus. *J. Virol.* 49, 303–309.
- Peluso, R.W., Moyer, S.A., 1988. Viral proteins required for the *in vitro* replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology* 162, 369–376.
- Poch, O., Blumberg, B.M., Bougueleret, L., Tordo, N., 1990. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J. Gen. Virol.* 71, 1153–1162.
- Pringle, C.R., 1997. The order *Mononegavirales* – current status. *Arch. Virol.* 142, 2321–2326.
- Qanungo, K.R., Shaji, D., Mathur, M., Banerjee, A.K., 2004. Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. *Proc. Natl. Acad. Sci. U.S.A.* 101, 5952–5957.
- Rahmeh, A.A., Li, J., Kranzusch, P.J., Whelan, S.P., 2009. Ribose 2'-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N7 methylation by the large polymerase protein. *J. Virol.* 83, 11043–11050.
- Rahmeh, A.A., Schenk, A.D., Danek, E.J., Kranzusch, P.J., Liang, B., Walz, T., Whelan, S.P., 2010. Molecular architecture of the vesicular stomatitis virus RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20075–20080.
- Rao, B.L., Basu, A., Wairagkar, N.S., Gore, M.M., Arankalle, V.A., Thakare, J.P., Jadhav, R.S., Rao, K.A., Mishra, A.C., 2004. A large outbreak of acute encephalitis with high fatality rate in children in Andhra Pradesh, India, in 2003, associated with Chandipura virus. *Lancet* 364, 869–874.
- Ribeiro Jr., E.A., Favier, A., Gerard, F.C., Leyrat, C., Brutscher, B., Blondel, D., Ruigrok, R.W., Blackledge, M., Jamin, M., 2008. Solution structure of the C-terminal nucleoprotein-RNA binding domains of the vesicular stomatitis virus phosphoprotein. *J. Mol. Biol.* 382, 525–538.
- Ribeiro, E.A., Leyrat, C., Gérard, F.C., Albertini, A.A., Falk, C., Ruigrok, R.W., Jamin, M., 2009. Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes. *J. Mol. Biol.* 394, 558–575.
- Rose, N.F., Publicover, J., Chattopadhyay, A., Rose, J.K., 2008. Hybrid alphavirus-rhabdovirus propagating replicon particles are versatile and potent vaccine vectors. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5839–5843.
- Rupprecht, C.E., Willoughby, R., Slate, D., 2006. Current and future trends in the prevention, treatment and control of rabies. *Expert Rev. Anti Infect. Ther.* 4, 1021–1038.
- Schnell, M.J., McGettigan, J.P., Wirblich, C., Papaneri, A., 2010. The cell biology of rabies virus: using stealth to reach the brain. *Nat. Rev. Microbiol.* 8, 51–61.
- Schnell, M.J., Mebatsion, T., Conzelmann, K.K., 1994. Infectious rabies viruses from cloned cDNA. *EMBO J.* 13, 4195–4203.
- Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D., Ruigrok, R.W., 2001. Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J. Virol.* 75, 490–498.
- Simonsen, C.C., Batthumphries, S., Summers, D.F., 1979. RNA-synthesis of vesicular stomatitis virus-infected cells – *in vitro* regulation of replication. *J. Virol.* 31, 124–132.
- Solon, J., Gareil, O., Bassereau, P., Gaudin, Y., 2005. Membrane deformations induced by the matrix protein of vesicular stomatitis virus in a minimal system. *J. Gen. Virol.* 86, 3357–3363.
- Stillman, E.A., Whitt, M., 1999. Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription. *J. Virol.* 73, 7199–7209.
- Streicker, D.G., Turmelle, A.S., Vonnhof, M.J., Kuzmin, I.V., McCracken, G.F., Rupprecht, C.E., 2010. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science* 329, 676–679.
- Szilagyi, J.F., Uryvayev, L., 1973. Isolation of an infectious ribonucleoprotein from vesicular stomatitis virus containing an active RNA transcriptase. *J. Virol.* 11, 279–286.
- Takacs, A.M., Das, T., Banerjee, A.K., 1993. Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. *Proc. Natl. Acad. Sci. U.S.A.* 90, 10375–10379.
- Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R.W., Burmeister, W.P., 2000. On the domain structure and the polymerization state of the sendai virus P protein. *Virology* 266, 99–109.
- Tekes, G., Rahmeh, A.A., Whelan, S.P., 2011. A freeze frame view of vesicular stomatitis virus transcription defines a minimal length of RNA for 5' processing. *PLoS Pathog.* 7, e1002073.
- Thomas, D., Newcomb, W.W., Brown, J.C., Wall, J.S., Hainfeld, J.F., Trus, B.L., Steven, A.C., 1985. Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. *J. Virol.* 54, 598–607.
- Tordo, N., Benmansour, A., Calisher, C., Dietzgen, R.G., Fang, R.-X., Jackson, A.O., Kurath, G., Nadin-Davis, S., Tesh, R.B., Walker, P.J., 2004. Rhabdoviridae. In: Fauquet, M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), *Virus Taxonomy*, VIIIth Report of the ICTV. Elsevier/Academic Press, London, pp. 623–644.
- Toriumi, H., Kawai, A., 2004. Association of rabies virus nominal phosphoprotein (P) with viral nucleocapsid (NC) is enhanced by phosphorylation of the viral nucleoprotein (N). *Microbiol. Immunol.* 48, 399–409.
- Warrell, M.J., Warrell, D.A., 2004. Rabies and other *Lyssavirus* diseases. *Lancet* 363, 959–969.
- Wertz, G.W., 1978. Isolation of possible replicative intermediate structures from vesicular stomatitis virus-infected cells. *Virology* 85, 271–285.
- Weyer, J., Kuzmin, I.V., Rupprecht, C.E., Nel, L.H., 2008. Cross-protective and cross-reactive immune responses to recombinant vaccinia viruses expressing full-length *Lyssavirus* glycoprotein genes. *Epidemiol. Infect.* 136, 670–678.
- Whelan, S.P., Ball, L.A., Barr, J.N., Wertz, G.T., 1995. Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8388–8392.
- Whelan, S.P., Wertz, G.W., 2002. Transcription and replication initiate at separate sites on the vesicular stomatitis virus genome. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9178–9183.
- Zhang, X., Green, T.J., Tsao, J., Qiu, S., Luo, M., 2008. Role of intermolecular interactions of vesicular stomatitis virus nucleoprotein in RNA encapsidation. *J. Virol.* 82, 674–682.
- Zhang, X., Wei, Y., Ma, Y., Hu, S., Li, J., 2010. Identification of aromatic amino acid residues in conserved region VI of the large polymerase of vesicular stomatitis virus is essential for both guanine-N-7 and ribose 2'-O methyltransferases. *Virology* 408, 241–252.

Review: Transcription et réplication des *Mononegavirales* : une machine moléculaire originale

Authors: David Blocquel, Jean-Marie Bourhis, Jean-François Éléouët, Denis Gerlier, Johnny Habchi, Marc Jamin, Sonia Longhi, **Filip Yabukarski**

Publié en 2012 dans *Virologie*, Volume 16, Numéro 4, pages 225-257

Transcription et réplication des *Mononegavirales* : une machine moléculaire originale

David Blocquel¹
 Jean-Marie Bourhis^{1,2}
 Jean-François Éléouët³
 Denis Gerlier⁴
 Johnny Habchi¹
 Marc Jamin²
 Sonia Longhi¹
 Filip Yabukarski²

¹ Aix-Marseille Université,
 CNRS, AFMB UMR 7257,
 163, avenue de Luminy, Case 932,
 13288 Marseille, France

<sonia.longhi@afmb.univ-mrs.fr>
² Unit of Virus Host Cell Interactions,
 UMI 3265 UJF-EMBL-CNRS,
 6, rue Jules-Horowitz,
 38042 Grenoble cedex 9, France

³ Inra,
 unité de virologie et immunologie
 moléculaires (UR892),
 Jouy-en-Josas, France

⁴ Université Lyon-I,
 virologie humaine, Inserm U758,
 ENS Lyon,
 21, avenue Tony-Garnier,
 69007 Lyon, France

Résumé. Les virus à ARN de polarité négative et non segmenté ou *Mononegavirales* sont une source notable de pathogènes végétaux, animaux et humains avec des impacts socioéconomiques et en santé publique majeurs. Citons, par exemple, les virus de la rougeole, des oreillons et respiratoire syncytial chez l'homme. De plus, le réservoir animal apparaît vaste, d'où un accroissement considérable de la menace de nouvelles zoonoses aussi effrayantes que rage, Ebola, Marburg, Nipah ou Hendra. D'un point de vue fondamental, ces virus ont un mécanisme de transcription et de réplication qui en font une exception dans le monde du vivant. En effet, l'ARN génomique n'est pas infectieux car ni transcritible, ni traduisible, et la polymérase virale L est, au mieux, capable d'amorcer une synthèse d'un oligonucléotide long de quelques dizaines de bases. Pour être utilisable comme matrice, l'ARN génomique doit être encapsidé dans un homopolymère hélicoïdal fait d'un emboîtement régulier de protéines N, dans lequel le squelette ribose-phosphate est entièrement enfoui. Cet ensemble, appelé nucléocapside, est reconnu par L via son cofacteur P pour une transcription séquentielle de cinq gènes en cinq ARNm matures pour le plus simple des virus. Dans un deuxième temps, il y a commutation vers la réplication d'une copie ARN antigénomique complète encapsidée concomitamment à sa synthèse. En retour, celle-là sert de matrice pour la production d'une capsidie génomique fonctionnelle. Cette revue résume l'ensemble des données structurales, dynamiques et fonctionnelles de cette mécanique particulière et propose un modèle de son fonctionnement. Elle met en lumière les stratégies globalement similaires et les variations subtiles d'exécution selon les virus, et le rôle-clé de la dualité structurelle ordre/désordre des protéines dans la dynamique de la machinerie virale.

Mots clés : *Mononegavirales*, *Henipaviridae*, *Paramyxoviridae*, *Rhabdoviridae*, rougeole, rage, stomatite vésiculaire, virus respiratoire syncytial, transcription, réplication virale, nucléoprotéine, phosphoprotéine, polymérase, protéines intrinsèquement désordonnées, structure, substrat d'encapsidation

Abstract. Viruses with a non-segmented negative-sense RNA genome, or *Mononegavirales*, are important pathogens for plants, animals and humans with major socio-economic and health impacts. Among them are well-known human pathogens such as measles, mumps and respiratory syncytial virus. Moreover, animal reservoirs appear much larger than previously thought, hence broadening the risk of emergence of life-threatening zoonotic viruses such as Rabies, Ebola, Marburg, Nipah or Hendra related viruses. These viruses have peculiar transcription and replication machinery that make them unique in the living world. Indeed, their genomic RNA, when naked, is non-infectious because it can be neither transcribed nor translated, and the L RNA-dependent RNA-polymerase is at best able to initiate the synthesis of an RNA copy of a few of tens of nucleotides in length. To serve as a template, the genomic RNA must be encapsidated in a helicoïdal homopolymer made of a regular and continuous array of docked N protomers in which the ribose-phosphate backbone is fully embedded. This complex, or nucleocapsid, is recognized by the L polymerase thanks to its cofactor,

doi:10.1684/vir.2012.0458

Tirés à part : S. Longhi

Virologie, Vol 16, n° 4, juillet-août 2012

225

Pour citer cet article : Blocquel D, Bourhis JM, Éléouët JF, Gerlier D, Habchi J, Jamin M, Longhi S, Yabukarski F. Transcription et réplication des *Mononegavirales* : une machine moléculaire originale. *Virologie* 2012; 16(4) : 225-57 doi:10.1684/vir.2012.0458

© John Libbey Eurotext, 2012

the P protein, to sequentially transcribe the five genes into five processed mRNAs for the simplest viruses. Subsequently, a switch occurs and the polymerase replicates a full copy of antigenomic RNA that is concurrently encapsidated. This new template is then used for the production of new infectious genomic nucleocapsids. This review summarizes current structural, dynamic and functional data of this peculiar molecular machinery and provides a unified model of how it can function. It illuminates the overall common strategies and the subtle variations in the different viruses, along with the key role of the dual ordered/disordered structure of the protein components in the dynamics of the viral polymerase machinery.

Key words: *Mononegavirales*, *Henipaviridae*, *Paramyxoviridae*, *Rhabdoviridae*, measles, rabies, vesicular stomatitis virus, respiratory syncytial virus, transcription, virus replication, nucleoprotein, phosphoprotein, polymerase, intrinsically disordered proteins, structure, encapsidation substrate

Glossaire et acronymes

Mot/acronyme (anglais)	Définition/nom complet	Remarques
Box Conformère	Région conservée de N _{TAIL} , pouvant contenir un MoRE Conformation transitoire dont la durée de vie est compatible avec une détection biophysique	<i>Paramyxovirinae</i>
L	Large protéine ou polymérase, support des activités catalytiques de polymérase (RdRp)	Toujours associée à son cofacteur P
M2-1 MoRE (α - β -I-)	Cofacteur de processivité de la polymérase de RSV « <i>molecular recognition element</i> », courte région au sein d'un domaine désordonné ayant une tendance à lier un partenaire et à subir un repliement	<i>Pneumovirinae</i> De type α , β ou I selon le type de transition structurale qu'il subit suite à interaction avec le partenaire
N	Nucléoprotéine	
NC	Nucléocapside	Utilisé comme terme générique, n'existe jamais sous forme isolée
N ⁰	N sous forme monomérique soluble	Toujours en complexe avec P
N ⁰ -P	Substrat d'encapsidation	Complexe soluble
N _{CORE}	Région N-terminale de N, cœur structuré liant l'ARN	Toujours associée à de l'ARN ou à P
N _{CTA}	Bras (« arm ») C-terminal de N verrouillant la sous-unité aval sur son N _{CTD}	<i>Rhabdoviridae</i> , RSV
N _{CTD}	Domaine C-terminal de liaison à l'ARN	<i>Rhabdoviridae</i> , RSV
N _{NTA}	Bras (« arm ») N-terminal de N venant verrouiller la sous-unité sur le N _{CTD} amont	<i>Rhabdoviridae</i> , RSV
N _{NTD}	Domaine N-terminal de liaison à l'ARN	<i>Rhabdoviridae</i> , RSV
N ^{NUC}	N sous forme homopolymérique (NUCLÉOCAPSIDE)	Toujours associé à de l'ARN
N ^{NUC} -ARN	Nucléocapside, matrice de la polymérase	De forme hélicoïdale
N _{TAIL}	(demi-)région C-terminale de N (structure désordonnée)	<i>Paramyxovirinae</i>
P	Phosphoprotéine	Chaperon de N, L et anti-immunité innée
P _{10CT}	Séquence de 10 résidus en C-terminal de P	RSV
P _{CT}	Région C-terminale de P	Non commune avec V (<i>Paramyxoviridae</i>)
P _{CTD}	Domaine C-terminal de P	<i>Rhabdoviridae</i>
PID	Protéine intrinsèquement désordonnée	Dépourvue de structure secondaire et tertiaire stable en l'absence de partenaire, peut contenir des structures secondaires transitoires
P-L	Transcriptase	Embarquée dans le virion, active dès l'entrée
P _{MD}	Domaine d'oligomérisation de P	<i>Rhabdoviridae</i> = dimérisation, <i>Paramyxoviridae</i> = tétramérisation
P _{NT}	Région N-terminale de P	Commune avec V (<i>Paramyxoviridae</i>)
P _{NTD}	Domaine N-terminal de P	<i>Rhabdoviridae</i> , <i>Paramyxoviridae</i> . Contient un α -MoRE
P _{XD}	Domaine C-terminal de P des <i>Paramyxovirinae</i>	<i>Paramyxovirinae</i>
RdRp	ARN polymérase ARN dépendante (L)	Porte les fonctions catalytiques
VP30	Cofacteur de processivité de la polymérase d'EboV	<i>Filoviridae</i>

Les virus à ARN négatif

Les virus ayant un génome ARN de polarité négative sont responsables de nombreuses maladies chez l'homme, les animaux et les plantes [1]. Le caractère segmenté ou non de leur génome les classe en deux ordres, *Multinegavirales*, dont la machinerie de réplication a fait l'objet d'une autre revue [2] et les *Mononegavirales* (MNV). Ce dernier comprend quatre familles : *Bornaviridae* (virus de la maladie de Borna), *Filoviridae* (virus Ebola, EboV et Marburg), *Paramyxoviridae* et *Rhabdoviridae*. Cette dernière comprend sept genres (*Vesiculovirus*, *Lyssavirus*, *Ephemerovirus*, *Novirhabdovirus*, *Cytorhabdovirus*, *Nucleorhabdovirus* et non défini) [3] avec comme prototypes les virus de la stomatite vésiculaire (VSV) et de Chandipura (ChanV) pour les *Vesiculovirus* et les virus de la rage (RABV) et Mokola (MoV) pour les *Lyssavirus*. Deux sous-familles constituent la famille des *Paramyxoviridae* : *Paramyxovirinae* et *Pneumovirinae*. La première est organisée en sept genres (*Morbillivirus* [prototype : virus de la rougeole, MeV], *Respirovirus* [prototype : virus Sendaï, SeV], *Rubulavirus* [prototype : virus des oreillons, MuV], *Henipavirus* [prototypes : virus Nipah et Hendra (NiV et HeV)], *Aquaparamyxovirus*, *Ferlavirus* et *Avulavirus*). La seconde sous-famille comprend deux genres, *Pneumovirus* et *Metapneumovirus*, dont le virus respiratoire syncytial (RSV) en est le représentant le mieux connu.

Certains virus sont prévalents chez l'homme, comme les virus de la rougeole, des oreillons, para-influenza, respiratoire syncytial ou le *Metapneumovirus* dont l'impact en santé publique est majeur. Ainsi, la rougeole est encore une cause majeure de mortalité chez les jeunes enfants [4] et sujette à une ré-émergence massive à l'instar de l'épidémie survenue en France de 2009 à 2011 consécutive à la négligence accrue de couverture vaccinale ; le RSV est le principal responsable des bronchiolites et pneumonies sévères du nourrisson et de pneumonies et pathologies respiratoires chez les personnes âgées et les receveurs d'une transplantation de moelle osseuse (<http://www.cdc.gov/rsv/>). D'autres MNV émergent ou ré-émergent à partir de réservoirs animaux, en particulier la chauve-souris, provoquant des épidémies de maladies graves ou mortelles comme les virus RABV, ChanV, EboV, HeV et NiV [5, 6]. De même peut-on évoquer les pathologies vétérinaires telles la maladie de Carré chez les carnivores, la maladie des bovins provoquée par le virus *Rinderpest* dont l'éradication vient d'être annoncée, diverses maladies des poissons provoquées par des *Rhabdovirus* redoutés dans les piscicultures, etc.

Si pour lutter contre certains MNV, nous disposons de vaccins très efficaces, la couverture vaccinale reste incomplète [7]. De manière plus inquiétante, de nombreux MNV, auxquels la population humaine n'a vraisemblablement pas

encore été exposée, sont aujourd'hui découverts chez certains animaux, notamment les chauves-souris. Ainsi, de nouvelles souches de rage pour lesquelles les vaccins contre la rage classique semblent moins efficaces ont été découvertes chez ces dernières [8], d'où un risque d'émergence dans la population humaine. Plus encore, un recensement systématique des MNV chez les chauves-souris et les rongeurs terrestres à travers le monde (Europe incluse) a mis à jour d'autres virus apparentés aux *Henipaviridae* pour ne citer que les plus effrayants [9]. Dans ce contexte, l'originalité et la conservation de la machinerie de synthèse d'ARN chez les MNV justifient son étude approfondie avec la perspective de développement de composés antiviraux ciblant sélectivement leur mécanisme de réplication.

Homogénéité structurale et fonctionnelle des Mononegavirales

Très variables dans leur spécificité d'hôtes, leur mode de transmission et les maladies qu'ils provoquent, les MNV partagent une organisation similaire de leur génome et de leur virion, et des stratégies de transcription et de réplication de l'ARN viral très proches et très originales dans le monde du vivant (*figure 1*) [3, 10, 11]. Ce sont des virus enveloppés contenant une nucléocapside (NC) hélicoïdale en un ou plusieurs exemplaires. Leur génome, dont la taille est comprise entre 9 et 20 kb, comprend entre cinq et 12 gènes encadrés par des régions non codantes complémentaires au mieux sur une douzaine de nucléotides et dénommées « *leader* » et « *trailer* » respectivement aux extrémités 3' et 5' du génome. Chaque gène est encadré par des séquences conservées d'initiation (*gene start* [GS]) et de terminaison (*gene end* [GE]) qui délimitent les gènes transcrits. Le génome des MNV code pour une ARN polymérase ARN-dépendante (RdRp) présente dans le virion qui, lors de l'infection, catalyse la transcription et la réplication du génome viral. La transcription débute par la synthèse d'un court ARN « *leader* » et se poursuit par la production d'ARNs (ARNm) messagers pour chacun des gènes. L'ajout immédiat de la coiffe est réalisé par un mécanisme spécifique [12]. La réplication de l'ARN génomique requiert la synthèse continue des protéines virales. Ces similarités suggèrent une évolution à partir d'un ancêtre commun avec une divergence expliquant la faible similarité des séquences.

Organisation du génome et protéines virales

Le génome des MNV code pour des protéines structurales requises pour former la particule virale et des protéines

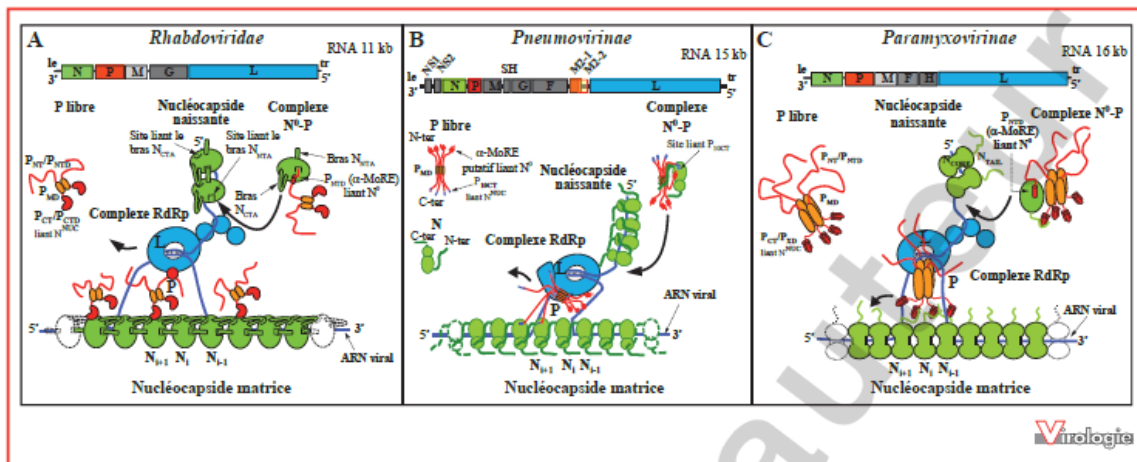


Figure 1. Représentation schématique des complexes réplcatifs des *Rhabdoviridae* (A), *Pneumovirinae* (B) et *Paramyxovirinae* (C). En haut de chaque panneau est présentée l'organisation schématique du génome avec les gènes N : nucléoprotéine (vert) ; P : phosphoprotéine (rouge) ; M : protéine matrice ; G : glycoprotéine ; L : grande sous-unité de la polymérase (bleue) ; F : protéine de fusion. Le (*leader*) et Tr (*trailer*) sont des régions non codantes supports des promoteurs transcriptionnel (Le) et réplcatif (Le & Tr) à partir des extrémités 3' des (anti)génomés. Dans la partie inférieure de chaque panneau sont schématisés les complexes de réplication, avec l'ARN symbolisé par une ligne bleue. L'ARN néo-synthétisé est montré déjà partiellement encapsidé. Dans P et N, les régions intrinsèquement désordonnées sont représentées par des lignes. La conformation étendue des régions désordonnées au sein de N et de P permettrait la formation d'un complexe tripartite entre N⁰, P, et L, prélude à l'assemblage de N⁰ avec la nucléocapside (NC) en néoformation. **A) *Rhabdoviridae*.** P est un dimère. Chaque monomère contient un domaine central de dimérisation (orange) et un domaine C-terminal de fixation sur le complexe N^{NUC}-ARN (rouge). Le domaine C-terminal de P se fixe sur le sommet du domaine C-terminal d'un protomère de N de la matrice N-RNA et les bras N_{CTA} de ce protomère et du protomère voisin se ferment autour du P_{CTD} à la manière d'une pince. Le domaine d'interaction avec L est montré sous la forme d'un cercle rouge symbolisant un repliement putatif induit lors de l'interaction. L progresserait le long de la matrice N^{NUC}-ARN en sautant d'une P à l'autre comme indiqué par la flèche. La flexibilité de la P assurerait la capture de L en aval du site de fixation de P, l'accompagnerait dans sa progression dans la synthèse d'ARN, jusqu'à la transmettre à la P immédiatement en aval. Dans le complexe N⁰-P, l' α -MoRE de la région N-terminale de P (P_{NTD}) qui subit une transition désordre/ordre est symbolisée par un cylindre rouge. Elle se fixe dans la gorge qui, dans la forme polymérique de N, accueille le bras N-terminale d'un autre protomère, bloquant ainsi la polymérisation de N. **B) *Pneumovirinae*.** P de RSV est un tétramère. Son domaine d'interaction avec la NC est localisé dans les dix derniers résidus C-terminaux. Il se fixe à une région située dans le domaine globulaire N_{NTD} situé à la périphérie de la NC hélicoïdale. Le domaine d'interaction avec N⁰ et le complexe N⁰-P sont spéculatifs. Seul son domaine d'oligomérisation P_{MD} (résidus 120-160) est structuré et prédit en coiled-coil. Le domaine d'interaction de P avec L est soit localisé dans le domaine d'oligomérisation de P soit dépendant de l'état oligomérique de P. N est bilobée, l'ARN se loge dans un sillon situé entre les deux lobes N_{NTD} et N_{CTD} et les bras N- et C-terminaux permettent les interactions N-N au sein de la NC. **C) *Paramyxovirinae*.** P de MeV est supposée être un tétramère par analogie avec SeV. Le segment qui connecte la région de multimérisation de P (P_{MD}) et P_{XD} est désordonné. Dans le complexe N^{NUC}-P, P est représentée liée à N^{NUC} à travers trois de ses quatre « bras » terminaux P_{XD} selon le modèle d'une progression « en faisant la roue » (flèche). Le complexe N⁰-P est représenté avec une stoechiométrie de 1:4 par analogie avec le SeV. La région N_{CORE} au sein du complexe N⁰-P a été représentée avec une forme légèrement différente par rapport à celle du complexe N^{NUC}-P pour refléter son changement de conformation [180]. Au sein de la NC, la protéine N est bilobée par analogie avec les N de RSV, VSV et RABV. La région N_{TAIL} est exposée à la surface de la NC. La morphologie de la protéine L est calquée sur l'image en microscopie électronique de la L de VSV. L interagit avec P via P_{MD} et P_{NT}, par analogie respectivement avec les interactions décrites chez SeV et le virus *Rinderpest*.

non structurales essentiellement garantes d'une propagation chez l'hôte. Cinq du premier groupe sont communes à tous les MNV avec un agencement de leurs gènes identique le long du génome comprenant de l'extrémité 3' à l'extrémité 5' : ceux des nucléoprotéine (N), phosphoprotéine (P), protéine de matrice (M), glycoprotéine de surface et protéine L (Large) ou RdRp *stricto sensu* (figure 1). N encapside l'ARN formant un complexe N^{NUC}-ARN qui sert de matrice pour la transcription et la réplication. P et L

constituent la polymérase RdRp à deux sous-unités [10, 11]. M joue un rôle structural dans la particule virale, participe au processus d'assemblage des nouveaux virions dans le cytoplasme de la cellule hôte ([13] et références citées) et est un régulateur négatif de la transcription [14]. Chez les *Rhabdoviridae*, la glycoprotéine de surface G assure reconnaissance et fixation du virus sur la cellule hôte et induit la fusion de l'enveloppe virale avec la membrane cellulaire, préalablement à la pénétration du virus dans le cytoplasme

de la cellule infectée [15]. Chez les *Paramyxoviridae*, ces fonctions sont dissociées et assurées par deux glycoprotéines transmembranaires, la glycoprotéine d'attachement (G ou H) et la glycoprotéine de fusion (F) [11].

À ce socle de gènes communs, viennent s'ajouter pour certaines familles virales, d'autres gènes insérés à différents endroits du génome et codant pour des protéines structurales et/ou non structurales. Ainsi, le complexe RdRp du RSV comprend deux autres cofacteurs, M2-1 et M2-2. Le premier augmente la processivité de la polymérase lors de la transcription et le second favorise la réplication de l'ARN génomique en fin de cycle [16]. Par ailleurs, chez les *Rhabdoviridae* et les *Paramyxovirinae*, le gène *P* code pour d'autres protéines, des isoformes de *P* tronquées, des isoformes issues d'un mécanisme d'édition de l'ARNm (protéines V et W) et des protéines distinctes par l'initiation de la traduction à partir d'un codon d'initiation alternatif (protéines C, Y) [3, 11, 17]. Ces protéines sont notamment impliquées dans la régulation de la réplication virale et dans l'inhibition de la réponse immunitaire innée intrinsèque de la cellule hôte.

Architecture de la particule virale

Les virions des *Paramyxoviridae* sont pléiomorphes, de forme grossièrement sphérique (150-350 nm) ou filamenteuse [11]. La forme des *Rhabdoviridae* est plus régulière, en balle de fusil ou en bacille [3]. De façon générale, la particule des MNV est constituée de trois couches superposées. La couche externe ou enveloppe est une bicouche lipidique dérivée de la membrane plasmique de la cellule hôte riche en glycoprotéines qui forment des spicules visibles en surface. La couche intermédiaire, constituée de l'assemblage en un réseau tridimensionnel d'une partie des protéines *M*, assure la cohésion entre les couches interne et externe. La couche interne est la NC composée de protéines *N* polymérisées sur l'ARN génomique (N^{NUC} -ARN) auxquelles sont associées les protéines *L* et *P* et, dans le cas du RSV, la protéine *M2-1*. Cette couche n'est visible en microscopie électronique que pour les virus ayant une forme géométrique bien régulière comme les *Rhabdoviridae*. La reconstitution de la partie cylindrique du virion de VSV par cryo-microscopie électronique a révélé une géométrie hélicoïdale des couches intermédiaire et interne [18]. La localisation des protéines *P* et *L* reste indéterminée, et il est proposé que ces protéines, associées avec le complexe N^{NUC} -ARN, soient localisées dans la cavité centrale de la particule en forme de « cigare » définie par l'enroulement hélicoïdal de la NC [19]. En cryomicroscopie électronique, la structure hélicoïdale de N^{NUC} -ARN de MeV est visible à l'intérieur d'un virion, et elle est recouverte, au moins partiellement, d'un manchon de protéine *M* [20].

Machinerie de transcription/réplication et son fonctionnement

La réplication se déroule exclusivement dans le cytoplasme [3, 11] et conduit à la formation d'inclusions cytoplasmiques [21-23]. Cependant, certains *Rhabdoviridae* de plantes se répliqueraient en partie dans le noyau [24] et les *Bornaviridae*, dont les particularités structurales ne seront pas traitées, ont un cycle nucléaire [25].

Machinerie moléculaire

La machinerie responsable de la transcription et de la réplication est un complexe ribonucléoprotéique (NC) qui comprend l'ARN viral et les protéines *L*, *P*, *N* auxquelles s'ajoutent les protéines *M2-1* et *M2-2* des *Pneumovirinae* ou la protéine *VP30* des *Filoviridae*. Ce complexe constitue le module infectieux minimal capable de produire de nouveaux virions à l'intérieur d'une cellule hôte (figure 1) [26]. Au sein de ce complexe, l'ARN est encapsidé par un homopolymère non covalent de *N* parfaitement régulier couvrant le génome sur son intégralité. À l'entrée dans le cytoplasme, le complexe N^{NUC} -ARN, associé à des protéines *P* isolées indispensables à l'activité polymérase, sert de matrice pour le complexe polymérasique *L/P* ayant seulement une activité transcriptase [3, 10, 11]. En effet, l'ARN génomique ou antigénomique nu et isolé n'est pas infectieux car ni transcriptible, ni traduisible. *In vitro*, sur un ARN nu, la protéine *L* purifiée est capable d'amorcer la synthèse d'un ARN complémentaire sur une longueur de quelques dizaines de nucléotides, alors que le cofacteur *P* et l'ARN sous forme encapsidé par de la *N* polymérique sont requis pour transcrire et répliquer le génome entier [12, 27].

Transcriptase et transcription

Le cycle de réplication débute par la transcription séquentielle du génome avec production d'ARN « leader » non mûré et des ARNm de surcroît coiffés et polyadénylés et donc traduisibles par la machinerie cellulaire [28, 29]. Il n'y a qu'un seul promoteur transcriptionnel pour tous les gènes, situé à l'extrémité 3' du génome et chevauchant le début du premier gène [3, 10, 11]. De ce fait, la transcription est initiée exclusivement à cette extrémité et les gènes sont transcrits séquentiellement dans l'ordre *N*, *P*, *M*, *G* et *L* par un mécanisme d'arrêt/démarrage. *In vitro*, en utilisant des NC virales comme matrice, la polymérase des *Rhabdoviridae* et des *Paramyxoviridae* n'engage la synthèse d'ARN qu'à l'extrémité 3' du génome encapsidé avec production d'ARN leader. *In vivo*, certains arguments expérimentaux suggèrent qu'elle puisse initier la synthèse d'ARN au début du gène *N*. Dans ce cas, la polymérase éviterait la synthèse de l'ARN « leader » et balayerait (« scan ») l'ARN jusqu'à trouver le

signal d'initiation du premier gène. En atteignant une région intergénique, la polymérase fait une pause. Puis, suite à un balayage aval – ou même amont pour certains gènes de *Filoviridae* et de *Pneumoviridae* ayant leur GS en amont du GE du gène qui le précède – sur au plus une centaine de nucléotides, elle reconnaît le signal d'initiation GS du gène suivant et initie sa transcription [10, 30]. La ré-initiation échoue dans une proportion variable selon les séquences intergéniques et la transcriptase devient inactive. Des arguments cinétiques en cellules infectées indiquent que cette inactivité est transitoire jusqu'à ce que la transcriptase se réamorce en retrouvant l'unique promoteur transcriptionnel à l'extrémité 3' du génome. Y-a-t-il migration non productive de la polymérase le long de la NC ou bien détachement puis ré-attachement à distance à l'extrémité 3' du génome ? Aucune donnée expérimentale ne permet de trancher [31]. Ce mécanisme de régulation, bien caractérisé chez plusieurs virus, en particulier VSV et SeV, conduit à un gradient d'abondance des ARNm de chacun des gènes décroissant avec l'éloignement de l'extrémité 3' du génome et ajuste ainsi la production des protéines virales selon leur quantité requise pour l'assemblage de nouvelles particules. Les co-facteurs additionnels, M2-1 chez les Pneumovirus et VP30 chez les Filovirus, assurent la processivité de la polymérase nécessaire à la complétion des ARNm et à la transcription des gènes les plus avaux.

Réplique et réplication

La complétion de la phase de transcription primaire conditionne le démarrage de la réplication. Dans cette seconde phase, la polymérase reconnaît le promoteur de réplication, sis dans la région *leader* en amont du promoteur de transcription et le chevauchant, et elle synthétise une copie complémentaire complète du génome, appelée antigénome en ignorant les signaux de début et de fin de gène [10]. Puis, l'antigénome sert à son tour de matrice pour la synthèse de nouvelles copies du génome à partir du promoteur antigénomique sis dans la région *trailer*. Dans une cellule infectée, et malgré leur complémentarité parfaite, les ARNs génomique et antigénomique ne forment pas d'ARN double brin, car ils sont toujours étroitement encapsidés sous forme d'ARN simple brin par un homopolymère ininterrompu de N (N^{NUC}) formant la NC. En effet, la réplication se distingue de la transcription par l'encapsidation concomitante de l'ARN génomique (ou antigénomique) naissant grâce à l'assemblage de monomères de N reconstituant ainsi le complexe N^{NUC} -ARN qui sert de matrice pour la polymérase. Ainsi, la réplication du génome nécessite la production continue de nouvelles molécules de N [10]. En vertu de son rôle cardinal dans l'encapsidation du génome, N est la protéine virale la plus abondante. La disponibilité de quantités suffisantes de molécules de N contrôlerait le basculement

entre la transcription et la réplication [10, 31]. Ainsi, dans les cellules infectées par les MNV, la protéine N existe soit sous forme monomérique soluble (appelée N^0) soit sous forme polymérique complexée à l'ARN viral (N^{NUC}). Comme la protéine N a une forte affinité pour l'ARN et une forte propension à s'auto-assembler, la production de N monomérique nécessite, dès sa biosynthèse, l'intervention d'un chaperon, la protéine P. En s'associant à N^0 sous la forme du complexe soluble N^0 -P, elle empêche l'interaction de N avec un ARN cellulaire et son auto-assemblage illégitime. Le complexe N^0 -P sert de substrat d'encapsidation des ARN génomiques et antigénomiques durant la réplication (voir pour revue [3, 11] et références citées). La décision de produire des ARNm ou des génomes entiers doit donc être prise au moment de la synthèse de l'ARN complémentaire du « leader » et implique *a priori* la participation de la protéine N selon des mécanismes non élucidés. Chez RSV, le basculement entre les activités de transcription et de réplication impliquerait également la protéine M2-2. Chez les *Rhabdoviridae*, la phosphorylation de P ou de N jouerait ce rôle [32, 33].

Comment ça marche ?

À cause de l'interdépendance entre ARN génomique, support nucléocapsidique (fait d'un homopolymère de N), et polymérase L associée à son cofacteur/chaperon P, la mécanique moléculaire de transcription/réplication des MNV est unique dans le monde du vivant et une source de perplexité. Une première question, non traitée ici, est de comprendre comment ces quatre partenaires ont pu co-évoluer de manière interdépendante. Une autre est de comprendre comment fonctionne cette machinerie. Comment l'ARN génomique est-il extrait de N^{NUC} pour être lu par la polymérase, notamment au moment de l'initiation à l'extrémité 3' du génome ? Comment la polymérase progresse-t-elle le long de la matrice N^{NUC} -ARN ? Quel est le rôle de la P ou d'autres facteurs viraux et cellulaires dans ce mouvement ? Par quels mécanismes la machinerie bascule-t-elle entre la transcription et la réplication ? Comment l'encapsidation des nouveaux génomes se déroule-t-elle et comment est-elle couplée à la synthèse d'ARN ? Quels sont les rôles de P et L dans ce processus ?

Les progrès récents

Deux difficultés, au moins, ont entravé pendant longtemps notre marche vers une compréhension des mécanismes moléculaires mis en jeu. La première fût de reconstituer l'intégralité des étapes de transcription et de réplication à partir des composants purifiés. La seconde fût de déterminer la structure des composants de cette machinerie. À cet égard, la prédiction puis la démonstration de l'existence au sein des protéines N et P de longues régions

intrinsèquement désordonnées (jusqu'à 400 résidus) (PIDs) grâce aux travaux pionniers de l'une de nous [34-40] a représenté l'un des progrès-clé. Les régions désordonnées sont dépourvues de structures secondaire et tertiaire stables dans des conditions physiologiques et en absence de partenaires, et existent sous forme d'un ensemble de conformères (pour revue voir [41] et références citées). Elles diffèrent donc radicalement des protéines structurées caractérisées par une structure secondaire (hélices α , brins β et boucles) et tertiaire (agencement tridimensionnel des éléments de structure secondaire) stable.

L'abondance du désordre au sein des protéines du complexe répliatif de ces virus implique une pertinence fonctionnelle [35, 36, 38, 42, 43]. Si la structure de N de BDV a été résolue dans une forme libre sans ARN [44], il a fallu de la persévérance doublée d'élégantes astuces biochimiques pour résoudre les structures atomiques d'anneaux réguliers de NC de RABV, VSV et RSV [45-47] puis des complexes $N^{\text{NUC}}\text{-ARN-P}$ [48] et $N^0\text{-P}$ [49]. Trois des cinq structures sont à l'actif d'équipes françaises dont deux des auteurs de cette revue.

Aujourd'hui, notre compréhension des mécanismes de transcription et de répliation des MNV provient essentiellement d'études fonctionnelles menées avec VSV (*Rhabdoviridae*) et SeV (*Paramyxoviridae*), et d'études structurales récentes de divers composants de VSV, RABV, MoV, SeV, MeV, NiV, HeV et RSV. Ces travaux fournissent une vue de plus en plus détaillée de l'architecture de la machinerie virale de synthèse d'ARN, étrangement similaire dans son principe entre les *Rhabdoviridae* et les *Paramyxoviridae*, mais bâtie et fonctionnant de manière subtilement distincte, y compris au sein d'une même famille virale.

Dans cette revue, nous discuterons des apports récents de la caractérisation structurale à notre compréhension de la machinerie de transcription et répliation des *Rhabdoviridae* et des *Paramyxoviridae*, notamment la structure des protéines ou complexes $N^{\text{NUC}}\text{-ARN}$, $N^{\text{NUC}}\text{-ARN-P}$, $N^0\text{-P}$ et M2-1. Leur intégration dans des modèles nourrira notre compréhension des mécanismes moléculaires de la transcription et de la répliation. Un absent de marque, cependant, est la protéine L ou l'ARN polymérase ARN dépendante dont le support structural et les activités font l'objet de la revue jumelle de Morin et Whelan [12].

Organisation structurale des nucléocapsides

Structure du cœur nucléoprotéique $N^{\text{NUC}}\text{-ARN}$

L'ARN génomique des MNV est encapsidé par la nucléoprotéine sous la forme d'une NC hélicoïdale et le même

complexe $N^{\text{NUC}}\text{-ARN}$ sert de matrice à la fois pour la transcription et pour la répliation. Dans ces complexes, l'ARN est fortement lié à la N, résistant à la centrifugation en gradient de CsCl à froid [50]. La fixation de N permet de maintenir l'ARN viral dans une forme simple brin, évitant la formation d'ARN double brin qui constituerait un *microbe associated molecular pattern* (MAMP) reconnu par les récepteurs de l'immunité innée [51], et protège l'ARN viral contre les nucléases et les ARN interférants [52, 53]. En microscopie électronique, les NC de MNV, extraites de virus ou de cellules infectées ou exprimées sous forme recombinante, apparaissent comme de longues hélices irrégulières. Les NC de *Rhabdoviridae* sont très souples et ressemblent au câble de combiné téléphonique surenroulé. Celles des *Paramyxoviridae* forment des tubes creux avec une morphologie plus régulière en forme d'épi (figure 2) [54]. L'apparence striée de ces structures vues en microscopie électronique correspond à l'arrangement côte-à-côte des nucléoprotéines le long de l'ARN.

La NC peut adopter différentes conformations plus ou moins compactes caractérisées par différents pas d'hélice [55-59]. Dans la conformation la plus étendue de la NC du SeV (pas hélicale de 37,5 nm contre 5,3 nm pour la forme la plus compacte), l'ARN viral est probablement plus accessible à la polymérase. Il est tentant d'imaginer qu'un tel phénomène de « déenroulement » pourrait intervenir au moment de la synthèse d'ARN pour permettre à la polymérase d'accéder à l'ARN génomique enfoui entre les mâchoires des protomères de N.

Chez les MNV, chaque monomère de N interagit avec un nombre précis de nucléotides, caractéristique de chaque famille : six pour les *Paramyxovirinae*, sept pour les *Pneumovirinae* [47], neuf pour les *Rhabdoviridae* [60, 61] et entre 12 et 15 pour les *Filoviridae* (voir [54] et références citées). Ainsi, la longueur du génome des *Paramyxovirinae* obéit à une « règle de six » stricte selon laquelle le génome doit être constitué d'un multiple entier de six nucléotides pour assurer une répliation efficace [62]. Cependant, même si chez les *Rhabdoviridae*, le nombre de molécules de N dans le virion correspond approximativement aussi au nombre théorique nécessaire pour couvrir le génome entier [61], les génomes des *Pneumovirinae* et des *Rhabdoviridae* n'ont pas nécessairement une longueur multiple exacte du nombre de nucléotides couvert par un protomère de leur N. Dans les cellules infectées, N se fixe exclusivement sur l'ARN génomique viral. Cependant, exprimées isolément en bactéries ou en cellules eucaryotes, les N des *Rhabdoviridae* et des *Paramyxoviridae* s'auto-assemblent sur des ARNs cellulaires formant des structures nucléocapsidiques de taille hétérogène mais de forme identique à celles contenant l'ARN viral [57, 58, 60, 63-68] (Papageorgiou, Habchi et Longhi, données non publiées).

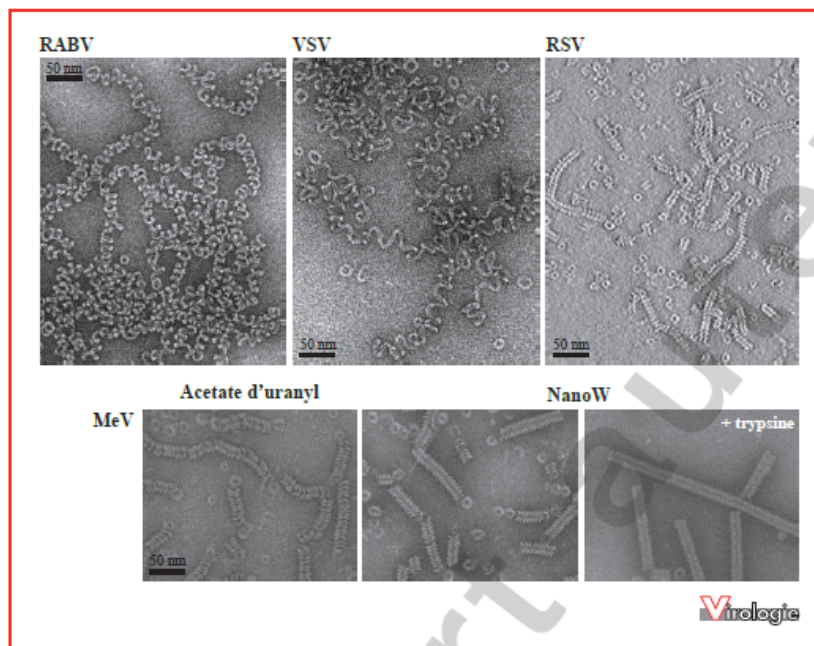


Figure 2. Clichés de microscopie électronique des nucléocapsides N^{NUC} -ARN de RABV, VSV, RSV et MeV. **A)** *Rhabdoviridae*. Images en coloration négative au silicotungstate de sodium de nucléocapsides de RABV et VSV isolées par centrifugation en gradient de CsCl (clichés R.W. Ruigrok, UVHCI Grenoble). **B)** *Pneumovirinae*. Nucléocapsides de RSV purifiées à partir de cellules d'insecte infectées par un baculovirus recombinant exprimant la nucléoprotéine et observées par cryomicroscopie après coloration négative (cliché David Bhella, université de Glasgow). **C)** *Paramyxovirinae*. Nucléocapsides de MeV non traitées ou traitées à la trypsine, après coloration négative avec de l'acétate d'uranyl ou du NanoW (avec une couche carbonique simple ou double, respectivement). Noter la rigidité des nucléocapsides obtenues après digestion tryptique de la région N_{TAIL} (panel du bas et à droite). Clichés modifiés à partir de [74]. Les nucléocapsides ont une longueur d'environ $1 \mu m$ pour un diamètre de 18 nm, avec un vide central de 4 à 5 nm de diamètre.

Fort opportunément, la N de certains virus forme également des anneaux composés d'un petit nombre de protomères de N, de neuf à 13 chez RABV avec majorité d'anneaux à dix sous-unités [45, 60, 69], essentiellement des anneaux de dix protomères pour VSV [63, 70], et dix ou 11 protomères pour RSV [47]. Chez les *Rhabdoviridae*, ces anneaux ont un diamètre et un nombre de sous-unités comparables au premier tour de l'extrémité conique des particules virales [18, 45, 71] ; et chez RSV, ils ont le même diamètre et le même nombre de sous-unités que les NC hélicoïdales, suggérant qu'un anneau corresponde à un tour d'hélice refermé sur lui-même par des interactions N-N. En revanche, la N des *Paramyxovirinae* ne semble pas former d'anneaux.

Des anneaux homogènes en taille de RABV, VSV et RSV ont été cristallisés, révélant ainsi les détails de la structure polymérique de N et des interactions entre N et l'ARN (figures 3 et 4) [45-47]. La protéine N des *Rhabdoviridae* et des *Pneumovirinae* possède deux domaines globulaires, un domaine N-terminal (N_{NTD}) et un domaine C-terminal (N_{CTD}) qui sont connectés entre eux par une région

chamnière. La molécule d'ARN est fixée dans un sillon délimité par l'interface entre les deux domaines qui semblent agir comme les mâchoires d'une pince emprisonnant entièrement la molécule d'ARN (figures 3 et 4). La région de N en contact avec l'ARN est riche en résidus basiques dont certains sont conservés au sein des différentes familles [71-73], et la fixation de l'ARN implique la formation de ponts salins entre les résidus chargés positivement de la protéine et les groupes phosphates chargé négativement du squelette de l'ARN [45-47]. Ces structures ont confirmé la fixation de neuf nucléotides par les protéines N de RABV et VSV et de sept nucléotides par la protéine N de RSV.

La fixation par N protège essentiellement le squelette sucre-phosphate de l'ARN, expliquant la résistance aux nucléases, et laisse certaines bases accessibles au solvant [53]. Chez les *Rhabdoviridae*, sur les neuf nucléotides fixés à chaque protomère de N, les bases des nucléotides (nt) 3, 4 et 6 sont enfouies dans la protéine (figure 3). Les autres sont exposées au solvant [45, 46], avec celles des nt 7, 8 et 9 empilées avec les bases 1 et 2 des nt fixés par le protomère

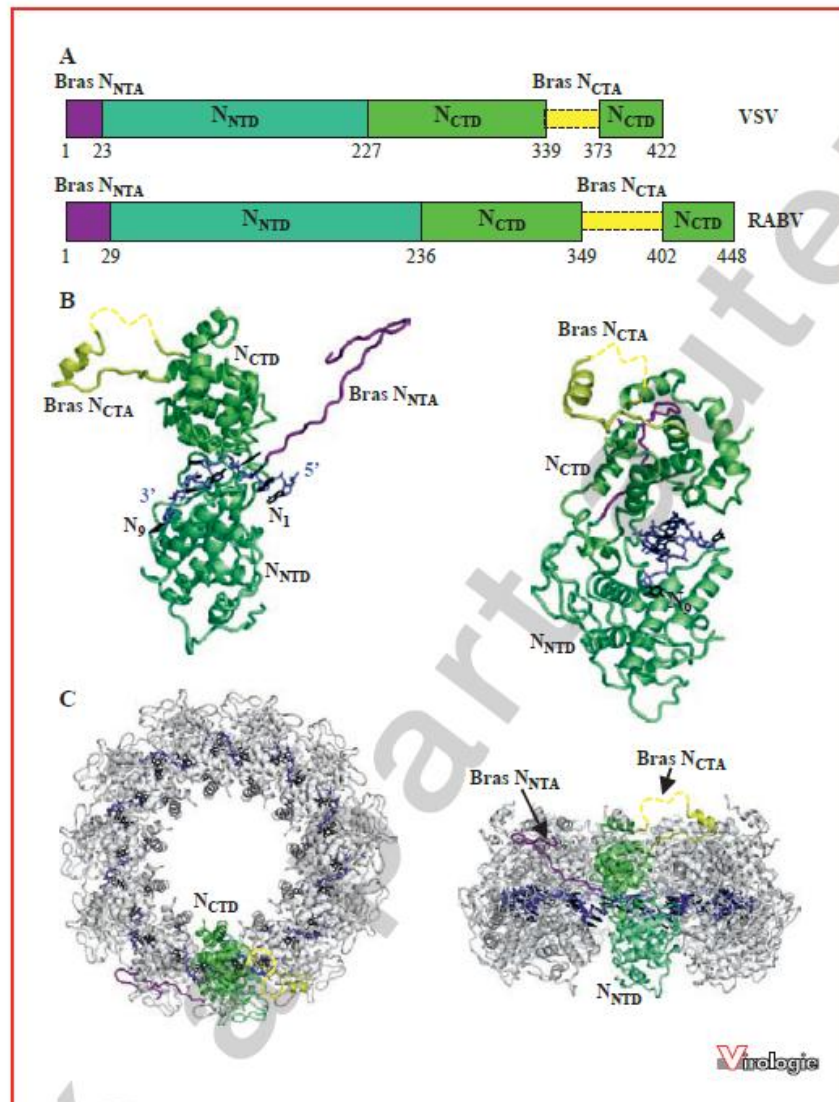


Figure 3. Structure de la nucléoprotéine de *Rhabdoviridae*. **A)** Schéma de l'organisation de N de VSV et de RABV. Localisation des deux domaines N- et C-terminaux (en vert) et des deux sous-domaines qui s'échangent avec les protomères voisins lors de la polymérisation de N, le bras N-terminal (N_{NTA}) (en magenta) et le bras C-terminal (N_{CTA}) (en jaune). **B)** Représentation en ruban de la structure d'un protomère du complexe N^{NUC}-ARN de RABV (code pdb 2GTT) avec l'ossature de l'ARN en bleu et les bases en noir prise en étai entre les deux lobes de N. La couleur des domaines de N respecte le code du panneau A. **C)** Structure de N polymérique sous forme d'un anneau de [N^{NUC}]₁₀-ARN de VSV (code pdb 2GIC). L'ARN et un protomère de N sont colorés comme en B. Les bras N_{NTA} et N_{CTA} se fixent sur les protomères adjacents respectivement à gauche et à droite.

aval. La base du nt 5 est libre. Chez RSV, l'organisation est similaire avec les bases des nt 2, 3 et 4 enfouies dans la protéine et celles des nt 5, 6 et 7 empilées avec la base 1 fixée sur le protomère aval (figure 4) [47]. La modélisation de la structure atomique de N de RSV dans la carte de densité de cryo-microscopie électronique de MeV (figure 5)

prédit comme attendu une organisation similaire avec toutefois six nucléotides par protomère de N^{NUC} [74]. Les bases orientées vers l'extérieur de la structure pourraient constituer des signaux, notamment pour l'encapsulation ou la polymérase virale, même si les analyses de séquences ne sont pas en faveur d'un tel schéma. Ces structures

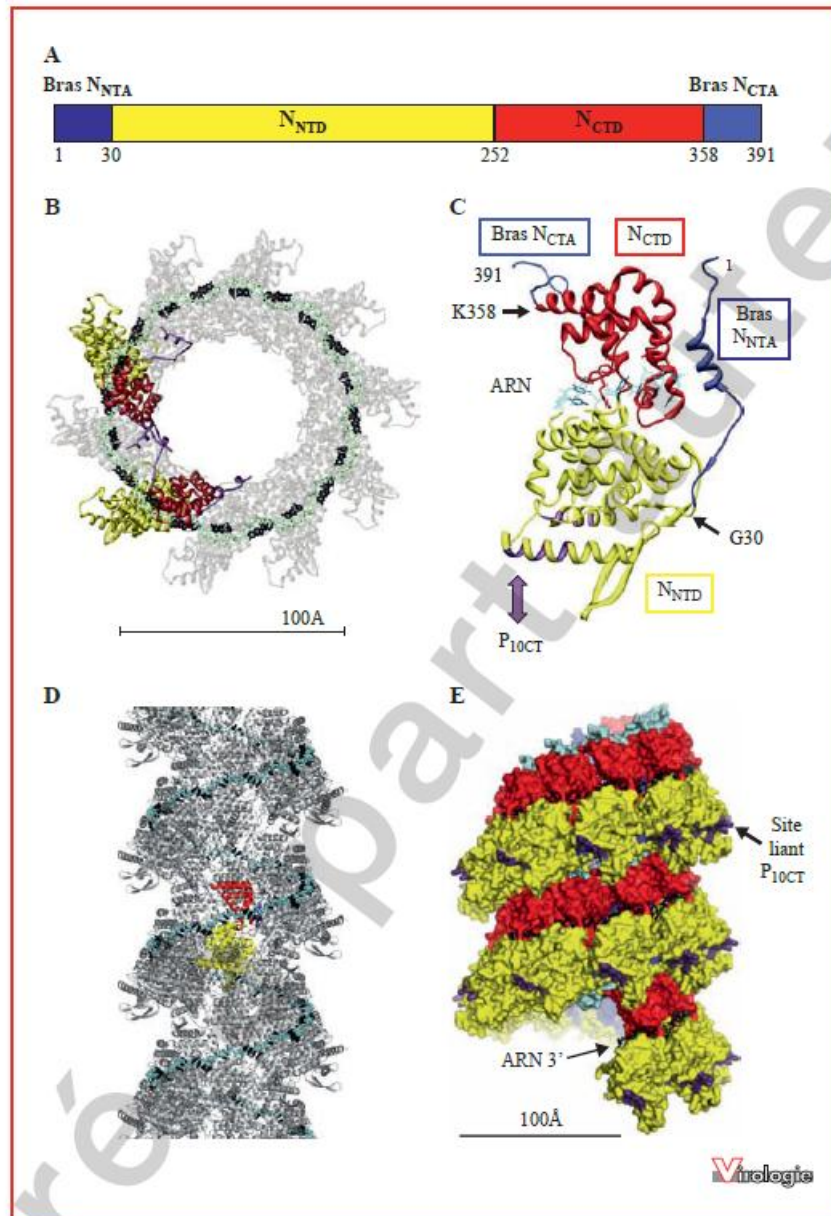


Figure 4. Structure atomique de la nucléocapside N^{NUC} -ARN du RSV résolue par cristallographie (code pdb 2WJ8). **A)** Schéma de l'organisation de la N de RSV avec les deux domaines N- et C-terminaux, N_{NTD} en jaune et N_{CTD} en rouge, et les bras N- et C-terminaux, N_{NTA} et N_{CTA} en bleu. **B)** Représentation en rubans de la structure de la protéine N assemblée en anneaux contenant dix monomères de N reliés entre eux par les bras N_{NTA} et N_{CTA} liant les N_{CTD} aval et amont respectivement. Même code couleur qu'en (A) L'ARN est en cyan (squelette) et noir (bases). Chaque monomère de N^{NUC} contacte sept nucléotides. **C)** Structure d'un monomère de N^{NUC} . Le site d'interaction avec la partie C-terminale de P est représenté en mauve. **D)** Modèle d'hélice droite (représentation en rubans) dérivé de la structure en anneau. **E)** Modèle d'hélice droite (avec vue de la surface) montrant la localisation du site d'interaction avec la partie C-terminale de P (en mauve) à la périphérie. Même code couleur qu'en (A) sauf pour le bras C-ter de N (en cyan) et l'ARN (entièrement noir).

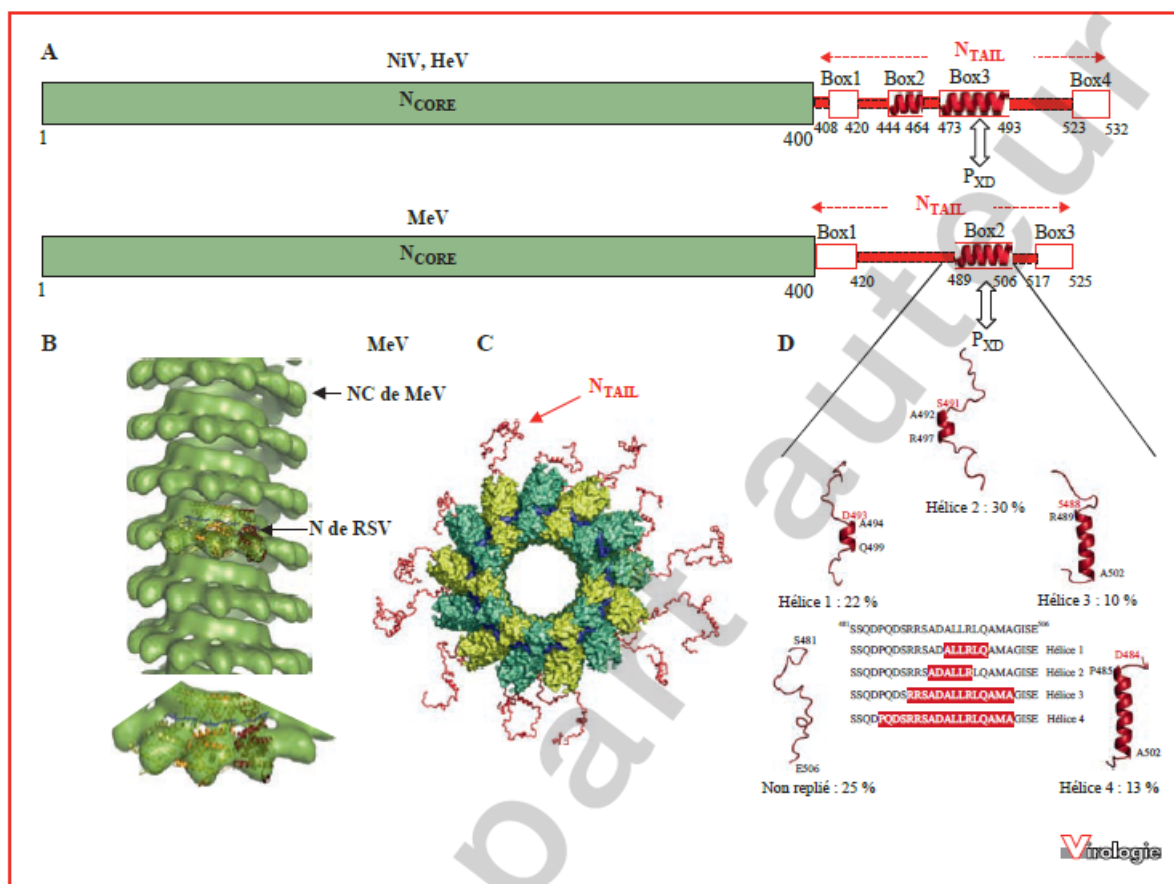


Figure 5. **A)** Représentation schématique de l'organisation modulaire de la protéine N chez les *Henipavirus* et chez MeV avec les domaines N_{TAIL} désordonné (ligne rouge) et structuré N_{CORE} (en vert), responsable de l'interaction N^{NUC} - N^{NUC} et N^{NUC} -ARN. Au sein des domaines N_{TAIL} sont représentés les I-MoREs (rectangles blancs) et les α -MoREs (hélices α rouges). **B)** Reconstitution tridimensionnelle, obtenue par cryo-microscopie électronique, de la nucléocapside du MeV à l'intérieur de laquelle a été modélisée la structure cristalline de N du RSV (modifiée à partir de [74] et préparée par Ambroise Desfosses). **C)** Vue supérieure de la nucléocapside de MeV avec les régions N_{TAIL} , exposées à la surface de la nucléocapside représentées en rouge, et l'ARN en bleu (modifiée à partir de [92]). **D)** Les multiples conformères de l' α -MoRE au sein de la Box2 du domaine N_{TAIL} du MeV et leur équilibre dynamique (modifiée à partir de [92]).

suggèrent que N doit s'ouvrir au passage de la polymérase de manière à rendre l'ARN accessible [45]. L'ARN peut être déplacé ou échangé de N^{NUC} sans affecter la structure polymérique à 37°C et au-delà [75]. Cette stabilité thermodynamique de l'homopolymère N^{NUC} indépendamment d'une âme ARN est confirmée par une structure N^{NUC} quasiment inchangée d'un mutant $S^{290}W$ dépourvu d'ARN [76]. Donc, cela est compatible avec une persistance de la forme polymérique N^{NUC} lors du passage de la polymérase et à une ré-encapsidation consécutive du brin d'ARN ayant servi de matrice.

Outre les deux domaines globulaires, les protéines N de MNV possèdent un sous-domaine N-terminal (bras N_{NTA}) et un sous-domaine C-terminal (bras N_{CTA}) qui ne se struc-

turent partiellement ou entièrement que dans le contexte du polymère de N (figures 3 et 4). N_{NTA} et N_{CTA} assurent l'assemblage et la cohésion du polymère de N via des contacts avec les protomères amont et aval adjacents. Chez les *Rhabdoviridae*, N_{CTA} consiste en une boucle plus ou moins longue et plus ou moins flexible située à l'intérieur du domaine C-terminal (figure 3). Chez la N de RABV, une grande partie de cette boucle n'est pas visible dans la structure cristallographique du complexe N^{NUC} -ARN indiquant une grande flexibilité. Chez VSV, la boucle N_{CTA} est plus courte et apparaît entièrement structurée dans certains protomères. Des mutations perturbant l'agencement de cette boucle accroissent la réplication sans modifier la transcription [77]. Le bras N_{CTA} du protomère N_i se fixe

sur le domaine C-terminal du protomère aval N_{i+1} et son bras N_{NTA} se fixe sur le domaine C-terminal du protomère amont N_{i-1} et contacte le bras N_{CTA} du protomère précédent N_{i-2} (figure 3C). La mutation F348A prédite pour rompre ce contact entre bras N_{NTA} et N_{CTA} contigus abolit sélectivement la transcription sans affecter la réplication [77]. Ainsi donc, les contacts à distance entre protomères N^{NUC} semblent mis en jeu de manière différentielle lors des étapes de transcription et de réplication. Dans les anneaux de RSV (*Pneumovirinae*), le bras N_{NTA} se fixe d'une manière comparable à celle observée chez les *Rhabdoviridae*, mais le bras N_{CTA} reste localisé à proximité du N_{CTD} , bien que les 12 à 20 derniers résidus C-terminaux soient désordonnés dans la structure cristallographique (figure 4B). À côté du rôle joué par les bras N_{NTA} et N_{CTA} , les N_{CTD} adjacents participent à la cohésion du polymère N^{NUC} -ARN grâce à une surface d'interaction de type mixte associant des contacts hydrophobes et polaires. En revanche, les N_{NTD} n'interagissent pas, d'où une certaine souplesse dans le positionnement des sous-unités N les unes par rapport aux autres [45, 46]. Chez RSV, les interfaces entre N_{NTD} et N_{CTD} adjacentes sont très hydratées et n'impliquent que des contacts de van der Waals intermittents [47]. Les interactions limitées aux interfaces entre N_{CTD} chez RABV et VSV ou très lâches entre les protomères de N^{NUC} chez RSV autorisent, par conséquent, des grandes variations des pas d'hélice et des tors des NC comme observé en microscopie électronique [58, 59, 69].

Une différence majeure entre la structure de la NC des *Rhabdoviridae* et celle des *Paramyxoviridae* (figures 3-5) réside dans le positionnement de la molécule d'ARN. Dans la structure cristallographique des anneaux de RABV [45] et de VSV [46], l'ARN se situe sur la face concave du complexe à l'intérieur de l'anneau (figure 3C), alors que dans celle des anneaux du RSV (figure 4B, D, E) [47] et dans la reconstruction par cryo-microscopie électronique de la NC du MeV (figure 5) [74], l'ARN est situé sur la face convexe à l'extérieur de l'anneau. Ces structures suggèrent donc *a priori* une organisation différente de la NC, bien que l'on ne puisse exclure chez l'un des groupes une possible inversion de conformation au moment du passage du complexe polymérisique. Un tel scénario est cependant peu plausible compte tenu des contraintes structurales.

Une autre différence affecte la taille et la structure de la partie C-terminale de N qui participe à la fixation de la P sur la forme polymérique N^{NUC} -ARN. Chez les *Rhabdoviridae* et les *Paramyxovirinae*, cette région de N participe à la fixation de la P sur la forme polymérique N^{NUC} -ARN (cf. plus loin). Chez les *Rhabdoviridae*, hormis le bras N_{CTA} , elle est parfaitement structurée [45, 58]. Chez les *Paramyxovirinae*, et d'après des études convergentes de microscopie

électronique, d'analyses bio-informatiques et d'analyses de mutants de délétion, la nucléoprotéine est constituée de deux parties. Alors que la partie N-terminale, dénommée N_{CORE} (approximativement résidus 1 à 400 chez MeV, SeV, et chez les *Henipavirus*), est prédite structurée en deux domaines N_{NTD} et N_{CTD} responsables de l'auto-assemblage et de la liaison à l'ARN [78-85], la partie C-terminale, dénommée N_{TAIL} (résidus 401 à 525 pour MeV, 401 à 524 pour SeV et 400 à 532 pour les *Henipavirus*), est désordonnée et fixe P (figure 5A). Chez les *Pneumovirinae*, l'appendice C-terminal est court (16 résidus) et n'intervient pas dans la fixation de P.

Structure désordonnée de l'appendice N_{TAIL} chez les *Paramyxovirinae*

N_{TAIL} des *Paramyxovirinae* présente les caractéristiques typiques de régions désordonnées, à savoir :

- une hypersensibilité à la protéolyse [85, 86] ;
- une variabilité de séquence au sein des membres phylogénétiquement proches [39] ;
- un biais compositionnel en acides aminés [34, 39, 40, 87] avec notamment une richesse en résidus promoteurs de désordre et une pauvreté en résidus promoteurs d'ordre ;
- l'impossibilité à être visualisé en microscopie électronique [59].

Les N_{TAIL} de *Paramyxovirinae* sont prédits être majoritairement désordonnés par le métaserveur MeDor [88]. D'après des approches hydrodynamiques et spectroscopiques, la nature désordonnée des domaines N_{TAIL} isolés est confirmée, bien qu'ils ne soient pas complètement étendus en solution et qu'ils conservent une certaine compacité résiduelle associée à un certain degré de structure secondaire fluctuante [34, 39, 87, 89].

D'après des reconstructions par microscopie électronique [56, 59], l'élimination de la région N_{TAIL} partiellement exposée à la surface de la NC [85, 90, 91], conduit à une augmentation de la rigidité de la NC (figure 2) avec des changements significatifs dans le pas de l'hélice et dans les tors [56, 59, 74, 87]. Bien que les domaines N_{TAIL} désordonnés ne soient pas résolus dans la reconstruction de la NC N^{NUC} -ARN du MeV, l'origine de l'appendice N_{TAIL} désordonné est positionné à l'intérieur de la NC (figure 5B) [74]. L'analyse par RMN combinée à la diffusion de rayons X aux petits angles (SAXS) a confirmé que N_{TAIL} du MeV reste désordonné *in situ* dans le contexte des NC intactes, d'où la proposition d'un modèle de l'organisation de N_{TAIL} dans la NC [92]. La région N_{TAIL} conserve sa flexibilité au sein des NC intactes, à l'exception des 50 premiers acides aminés de N_{TAIL} qui sont conformationnellement restreints (figure 5C). Cette rigidité pourrait résulter de la prise en sandwich de ces 50 acides aminés dans l'espace interstitiel

situé entre les tours d'hélice successifs permettant ainsi à la région N_{TAIL} d'émerger à l'extérieur de la NC hélicoïdale [92]. Cette architecture de la NC permet de comprendre l'augmentation de sa rigidité lors du clivage protéolytique de N_{TAIL} [59] et de faire des hypothèses sur le rôle de N_{TAIL} dans le positionnement de la polymérase sur sa matrice. La sensibilité au clivage protéolytique de N_{TAIL} suppose cependant une bonne accessibilité au clivage par la trypsine en position 392 située sur la face interne de la NC [85], probablement grâce à la flexibilité inhérente des NC intactes.

Au sein de la région N_{TAIL} de HeV, NiV et MeV, des analyses bio-informatiques et biochimiques révèlent la présence d'éléments de reconnaissance moléculaires ou *molecular recognition elements* (MoREs) (figure 5A). Les MoREs sont de courtes régions au sein d'une région désordonnée ayant une propension à subir un repliement induit lors de la liaison à un partenaire [93]. Dans le cas du MeV, un MoRE constitué d'hélices α transitoires (α -MoRE) et impliqué dans l'interaction avec le domaine C-terminal X (P_{XD}, résidus 459-507) de P et dans le repliement α -hélicale de N_{TAIL} a été localisé [34] dans l'une (Box2, résidus 489 à 506 de N) des trois régions de N_{TAIL} (dénommées Box1, Box2 et Box3) qui sont conservées au sein des Morbillivirus [94]. Cet α -MoRE est partiellement pré-configuré en solution en l'absence de P_{XD} [95-97]. D'après des approches quantitatives combinant des mesures de couplages résiduels dipolaires et des méthodes d'optimisation d'ensembles [92], cette région existe en tant qu'ensemble conformationnel comprenant une forme étendue et quatre conformères hélicaux de différentes longueurs (figure 5D). En plus de cet α -MoRE, deux autres régions avec un potentiel de repliement ont été identifiées au sein du domaine N_{TAIL} du MeV : une localisée dans la région 400-420 et une autre dans la région C-terminale (résidus 516-525) (figure 5A) dénommées Box1 et Box3, et capables de recruter des facteurs cellulaires décrits plus loin.

Au sein des domaines N_{TAIL} des Henipavirus, quatre MoREs putatifs ont été identifiés (figure 5A). L'un d'eux (Box3, résidus 473-493) est la contrepartie fonctionnelle de l' α -MoRE localisée dans la Box2 de MeV [34, 78, 98] dans la mesure où il constitue le site de liaison à P_{XD} [86, 99]. Comme la Box2 de N_{TAIL} de MeV, la Box3 des Henipavirus est aussi partiellement pré-configurée en hélice α [100].

La pré-structuration partielle de ces régions permettrait de réduire la pénalité entropique associée à la transition désordre-ordre lors de la liaison au partenaire P_{XD} [101].

La polymérase virale

La polymérase virale est constituée de deux sous-unités, une grande sous-unité appelée protéine L (large) qui porte

différentes activités catalytiques et une petite sous-unité non catalytique, la phosphoprotéine (P). La protéine L est une grande protéine d'environ 2 200 acides aminés, dans laquelle six régions conservées ont été identifiées. Chez les *Paramyxovirinae*, elle est oligomérique. La protéine L possède les différentes activités catalytiques nécessaires pour la synthèse de nouvelles copies du génome et pour la synthèse d'ARNm et leur maturation avec synthèse et méthylation de la coiffe à leur extrémité 5' et polyadénylation de leur extrémité 3'. Les activités méthyltransférases du ribose en position 2'-OH et de la guanine en position N-7 ont été prédites en région C-terminale, puis validées. La polyadénylation résulte d'un mécanisme itératif dans lequel la polymérase « bégaie » sur une courte séquence poly-U à l'extrémité 5'-terminale de chaque gène [12].

Quelles sont les bases moléculaires de la dualité fonctionnelle transcriptase/réplacase du complexe polymérasique L/P ? N pourrait en être un des régulateurs [10, 11]. Les activités transcriptase et réplacase feraient intervenir des complexes de composition différente avec, chez VSV [102], un complexe composé de N, P et L pour répliquer l'ARN génomique en présence de complexe N⁰-P et un autre, composé de P et de L pour la production des ARNm. Chez RSV, le facteur de transcription M2-1 est nécessaire à une activité transcriptase optimale mais n'intervient pas dans la réplication. Nous invitons les lecteurs à se référer à la revue jumelle de B. Morin qui décrit plus en détail la protéine L et ses activités [12].

P, une protéine modulaire oligomérique

P est une protéine multifonctionnelle essentielle de la machinerie virale. L'existence de régions fonctionnelles indépendantes, d'abord proposée d'après le phénotype de mutants de délétion [103-105], a été confirmée par la démonstration d'une organisation structurale modulaire associant des régions intrinsèquement désordonnées et deux domaines structurés [40, 43, 106-108] (figures 6-8). P agit comme un cofacteur non catalytique de la polymérase [109] et comme un chaperon de N⁰ monomérique [110-112] et de L [113, 114]. De plus, comme les autres protéines dérivées du gène P (V, W, C...), elle participe à la régulation de la réplication virale et à l'inhibition de la réponse immunitaire innée chez certains MNV. Les P des *Rhabdoviridae* et des *Paramyxoviridae* ne sont pas conservées dans leur séquence, mais elles sont oligomériques, partagent une organisation modulaire similaire avec une distribution des fonctions répartie de manière identique entre les régions désordonnées et structurées. En conséquence, nous proposons de diviser la P en deux régions distinctes (figures 6-8) : une région N-terminale (PNT), commune chez les *Paramyxovirinae* à P, V et W et une

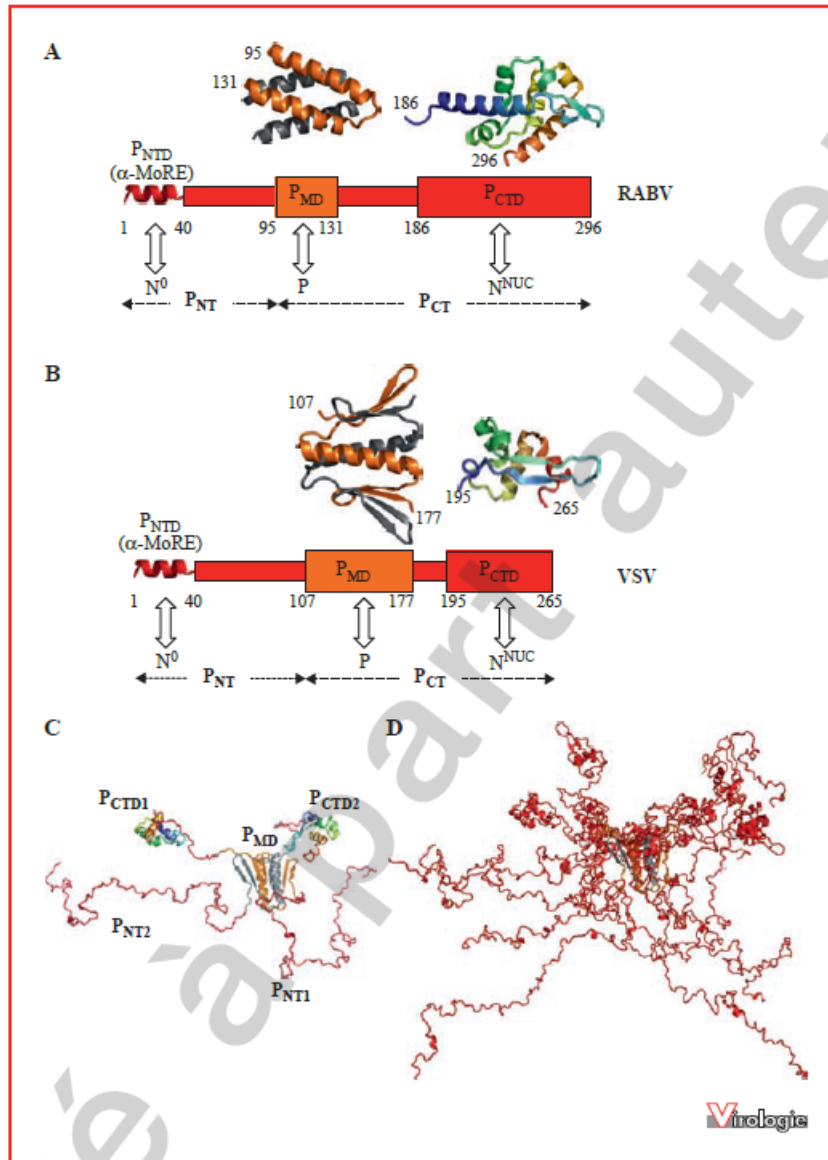


Figure 6. Structure de P des *Rhabdoviridae*, avec sa partie N-terminale (P_{NT}) et sa partie C-terminale (P_{CT}), contenant deux domaines structurés (boîtes larges) et deux régions intrinsèquement désordonnées (boîtes étroites). Organisation schématique de P de RABV (A) et de VSV (B). L' α -MoRE de P_{NTD} , site de fixation à N^0 , est partiellement désordonné lorsqu'il est isolé, mais il est replié en hélice α dans sa forme liée au partenaire. En position centrale, P_{MD} est le domaine de dimérisation. Le domaine C-terminal (P_{CTD}) se fixe sur la forme polymérique N^{NUC} -ARN. Chaque domaine est surmonté de sa structure atomique quand elle est connue. Un des protomères du P_{MD} dimérique est montré en orange et l'autre en gris. L'arc-en-ciel de la structure du P_{CTD} va de l'extrémité N-terminale (bleue) à C-terminale (rouge). La région désordonnée du P_{CT} est plus longue chez RABV (60 résidus) que chez VSV (16 résidus). C) Structure atomique d'un dimère de P de VSV obtenue par reconstruction avec incorporation des structures atomiques de P_{MD} et P_{CTD} . Les parties désordonnées reproduisent les déplacements chimiques mesurés par spectroscopie RMN de la P_{NT} . D) Représentation d'un ensemble de cinq conformères de P de VSV choisis parmi un ensemble de 8000 conformères, dont la courbe moyenne de diffusion de rayons X aux petits angles (SAXS) reproduit la courbe expérimentale. L'ajustement obtenu avec des ensembles de cinq modèles ou plus est meilleur qu'avec un modèle unique démontrant la diversité de conformations de la protéine en solution. Les cinq modèles sont centrés sur leur P_{MD} (orange et gris). Les régions désordonnées et P_{CTD} sont montrées en rouge [138].

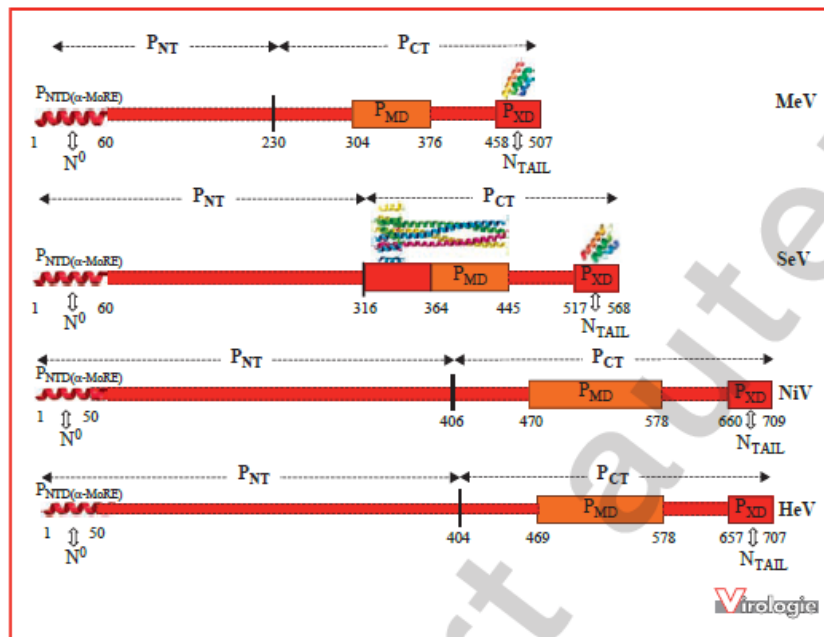


Figure 7. Organisation schématique de P de MeV et SeV et des Henipavirus avec les régions P_{NT} (commune à P et à V) et P_{CT} délimitées au centre par la barre noir verticale. Les régions globulaires et désordonnées sont respectivement représentées par des encadrés larges et étroits. La région chevauchant le cadre de lecture de V est désordonnée chez MeV et les Henipavirus, mais ordonnée chez SeV. L'α-MoRE à l'extrémité N-terminale et les sites de liaison à N_{TAIL} et à N⁰ sont indiqués. Les domaines sont surmontés par leurs structures atomiques connues : P_{XD} de MeV (code pdb 1OKS) et de SeV (code pdb 1R4G) et P_{MD} de SeV (pdb code 1EZJ).

région C-terminale (P_{CT}), propre à la protéine P. P_{NT} est globalement désordonnée, alors que P_{CT} contient un domaine d'oligomérisation et un domaine C-terminal structurés et reliés entre eux par un lien flexible plus ou moins long selon les virus.

Présence de régions intrinsèquement désordonnées

Des prédictions de désordre ont révélé l'organisation modulaire et la présence de longues régions intrinsèquement désordonnées chez la P des *Paramyxovirinae* [40, 107], *Henipavirus* [39], *Pneumovirinae* [115] et *Rhabdoviridae* [106, 116]. Les domaines P_{NT} des paramyxovirus sont tous prédits par le métaserveur MeDor qui regroupe les prédicteurs de désordre [88]. Cependant, individuellement, ceux-là divergent quant à la localisation précise des régions désordonnées dans la séquence. Afin de faciliter l'interprétation de ces données, nous avons proposé un calcul de score simple que nous avons appelé D-score permettant d'intégrer les résultats de multiples prédictions de désordre en une méta-prédiction consensus. Ce D-score a permis la localisation correcte des régions désordonnées en accord avec les limites des domaines structurés de P antérieurement définies par la résolution de leur

structure chez RABV et SeV (figures 6A et 7) [106, 116]. De même, la prédiction correcte des domaines P_{MD} de RABV et P_{CTD} de VSV a permis de définir les bornes de constructions compatibles avec une purification de qualité et avec la détermination de leur structure atomique (figure 6A, B) [116, 117]. De plus, des analyses biophysiques, dont notamment des expériences hydrodynamiques de SAXS et de RMN, ont confirmé l'état majoritairement désordonné des régions prédites comme telles [106, 108].

Présence de MoRE dans P_{NT}

La partie N-terminale désordonnée de longueur variable selon le virus comprend des régions de reconnaissance pour différentes protéines virales et cellulaires. Elle comprend notamment un MoRE conservé responsable de l'activité chaperon sur N⁰ qui contient des hélices α transitoires dont l'une est stabilisée lors de la formation du complexe N⁰-P (figures 6 et 7) [49]. Chez certains virus, cette région désordonnée comprend également des sites de phosphorylation qui participeraient à la régulation du cycle réplcatif viral [33]. Enfin, chez les *Rhabdoviridae*, cette région comprendrait également un site de fixation à la protéine L [118, 119]. La caractérisation par RMN et SAXS du peptide correspondant aux 60 premiers résidus de P liant N⁰ de VSV

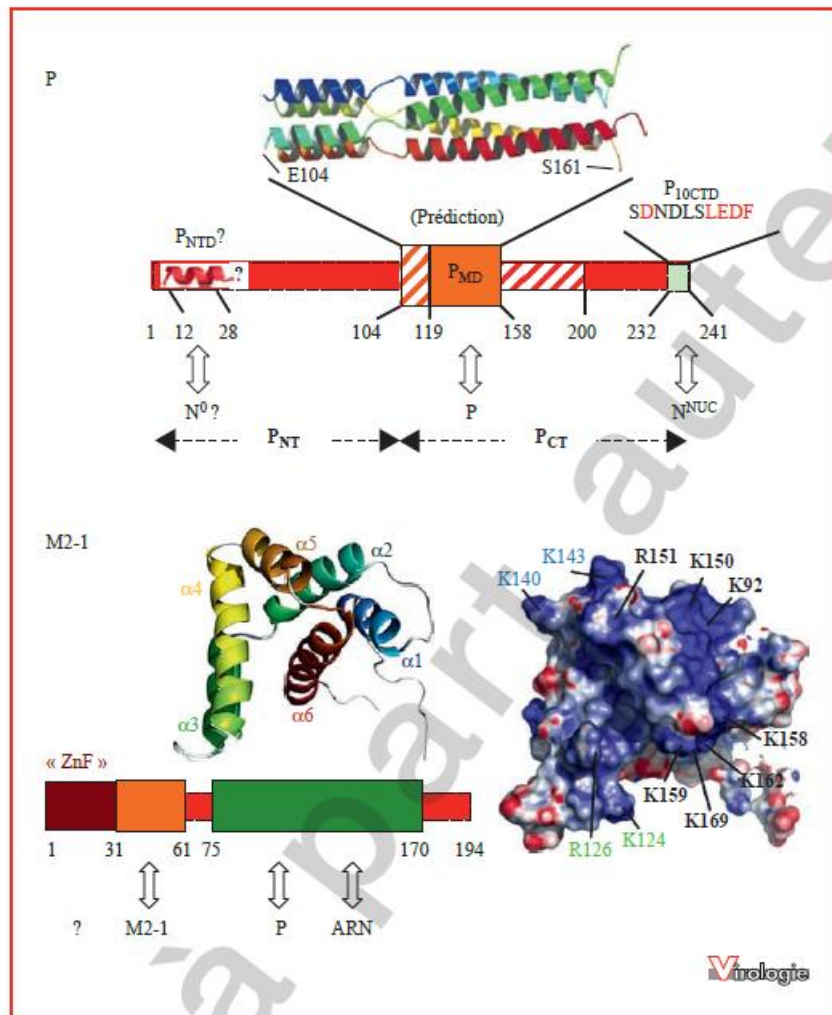


Figure 8. Organisation des protéines P (haut) et M2-1 (bas) de RSV : P (haut) avec son domaine de tétramérisation central (P_{MD}) surmonté de sa structure prédite, les régions N-terminale et C-terminale prédites non structurées avec des hélices α possibles entre les résidus 12-28 et 158-200. Les dix derniers résidus en C-terminal P_{10CT} constituent un domaine d'interaction avec N^{NUC} . M2-1 (bas) et ses quatre domaines : un domaine en doigt de zinc (ZnF) contenant trois cystéines et une histidine conservées, un domaine d'oligomérisation [32-61], un domaine cœur « core » responsable de l'interaction avec la protéine P et l'ARN [75-170] surmonté de sa structure déterminée en RMN selon le code couleur arc-en-ciel du N-terminus (bleu) au C-terminus (rouge) (code pdb 2L9J) et une région C-terminale non structurée. Le cœur présente une face très chargée positivement représentée en bas à droite. Les résidus basiques de différents clusters sur la séquence primaire de la protéine sont indiqués selon le code couleur arc-en-ciel de la structure de gauche en ruban.

confirme son caractère globalement désordonné. Cependant, elle montre la présence de deux hélices α dont la structuration alterne continuellement à partir de, et vers, des conformations désordonnées de la chaîne peptidique [112]. D'où l'adoption d'un modèle de cette région sous la forme d'un ensemble de conformères qui s'échangent sous l'influence de l'agitation thermique plutôt que sous la forme d'une structure unique. Cette représentation structurale

d'ensemble rend bien compte des propriétés hydrodynamiques du peptide et met en évidence la diversité des conformations de la protéine et le volume d'espace qui peut être échantillonné par cette dernière. Cette modélisation révèle l'existence d'une distribution bimodale des valeurs du rayon de giration (R_g) suggérant la coexistence de deux sous-populations de conformères, les uns compacts, les autres étendus, comme observé avec d'autres

PIDs [120, 121]. Lors de la formation du complexe N^0 -P, une des deux hélices α s'allonge et subit une transition ordre-désordre, et c'est elle qui joue un rôle prépondérant dans l'activité chaperon de P [49, 112].

Dans le cas des Henipavirus et de MeV (figure 7), les domaines P_{NT} sont en grande partie pauvres en paquets hydrophobes et possèdent très peu d'éléments de structure secondaire prédits, ce qui est une caractéristique des régions désordonnées. L'état désordonné des domaines P_{NT} de MeV, NiV et HeV a été confirmé expérimentalement par de multiples approches biochimiques et biophysiques [39, 107].

Les analyses bio-informatiques prédisent également une courte région ordonnée (40-50 résidus) au sein de la région N-terminale de la P de MeV, NiV et HeV (figure 7). Ce domaine N-terminal, qui possède une propension de repliement en hélice α , est aussi conservé chez les Avulavirus et Rubulavirus [40]. Chez ces derniers, cette région interagit avec N^0 [122]. Une région de 11 à 16 résidus (appelée Soyuz1) au sein des premiers 40 résidus des protéines P de tous les membres de la sous-famille des *Paramyxovirinae* peut être définie [123]. Cette région, probablement conservée par évolution divergente, serait impliquée dans la liaison à N^0 chez tous les MNV [123]. En effet, expérimentalement, les domaines P_{NT} de MeV, NiV et HeV ne sont pas totalement désordonnés et conservent une certaine compacité résiduelle observable en chromatographie d'exclusion de taille associée à l'analyse de diffusion dynamique de la lumière (DLS) [39, 107]. De plus, dans un milieu mimant l'environnement hydrophobe d'interactions protéine-protéine, les domaines P_{NT} de ces trois virus ont un spectre de dichroïsme circulaire dans l'UV lointain compatible avec un repliement en hélices α [39, 107].

Le domaine d'oligomérisation

P possède un domaine d'oligomérisation appelé P_{MD} , dimérique chez les *Rhabdoviridae* (figure 6A, B) [117, 124] et tétramérique chez les *Paramyxovirinae* (figure 7) et les *Pneumovirinae* (figure 8) [115, 125-127]. Le P_{MD} dimérique des *Rhabdoviridae* est assez court comme l'indiquent les structures pour RABV [117] et VSV (figure 6A, B) [128]. Le P_{MD} tétramérique de SeV comprend un faisceau de quatre hélices α parallèles surenroulées (*coiled-coil*) complété par une collerette d'autres hélices α qui stabilise la partie N-terminale du faisceau (figure 7) [125, 126]. Cette structure en faisceau de quatre hélices α parallèles surenroulées semble être conservée au sein des *Paramyxovirinae*, car elle est prédite pour la P d'autres membres de cette sous-famille. À l'exception de l'organisation parallèle des chaînes α au sein de l'oligomère, le degré d'oligomérisation et les différences structurales ne permettent pas de postuler une origine commune pour les P_{MD} des *Rhabdoviridae* et

des *Paramyxovirinae*. De plus, chez les *Paramyxovirinae*, ce domaine est impliqué dans la liaison avec la polymérase L [104, 105, 114].

Le domaine C-terminal de fixation à N^{NUC} -ARN

Chez tous les MNV, la région de fixation au complexe polymérique N^{NUC} -ARN est localisée en C-terminal de P. Elle est constituée soit d'un domaine structuré connecté au domaine central d'oligomérisation par un lien flexible plus ou moins long selon les virus (*Paramyxovirinae*, *Rhabdoviridae*), soit d'une région désordonnée (*Pneumovirinae*). Chez les *Paramyxovirinae*, la structure conservée du domaine C-terminal est un faisceau de trois hélices α (figure 7). Chez les *Rhabdoviridae*, le domaine C-terminal, plus long, comprend quatre (VSV, figure 6B) ou six hélices α (RABV, figure 6A) et un court feuillet β à deux brins (figure 6). Au sein d'une même famille, la superposition des structures fait apparaître une parenté évidente en dépit de l'absence de conservation de la séquence et de la présence d'insertions et de délétions (comparer les P_{CTD} , figure 6A, B) [117].

En accord avec les analyses bio-informatiques [39, 40], les domaines X C-terminaux (P_{XD}) des protéines P de MeV et de SeV (résidus 459-507 et 516-568, respectivement) sont organisés en trois hélices α (figure 7) [95, 98, 129]. De même, d'après des études par CD et RMN, les domaines X des Henipavirus (résidus 660-709 pour NiV et 657-707 pour HeV), très similaires avec $\sim 90\%$ de similarité et $\sim 80\%$ d'identité, adoptent également une structuration α -hélicale prédominante [130]. Chez ces virus, le domaine X est le domaine de liaison à la NC via l'interaction avec N_{TAIL} [78, 89, 98, 130].

Bien que le domaine X (résidus 343-391 de P) du MuV adopte en solution une structure dite en globule fondu (« molten globule ») et soit donc dépourvu de structure tertiaire stable [78], il a cristallisé et sa structure a été déterminée par cristallographie aux rayons X [131]. Il s'agit d'un bel exemple où l'empilement cristallin a permis de « capturer » une conformation stable. Corrélativement, ce domaine possède des propriétés distinctes par rapport aux domaines X des Morbillivirus, Respirivirus et Henipavirus car il n'interagit pas avec N_{TAIL} , mais avec la région structurée, N_{CORE} [132].

La comparaison de la structure des domaines C-terminaux des protéines P de RABV, VSV et MoV, [116, 133, 134] avec celle des domaines X de MeV [98] et SeV [89] suggère une topologie conservée entre *Paramyxovirinae* et *Rhabdoviridae* en dépit d'une faible conservation de séquence [135]. En fait, en dépit d'une absence de conservation identifiable dans la séquence en acides aminés, une similarité structurale semble exister dans le domaine C-terminal de la P de tous les MNV [123, 135].

La connexion flexible entre P_{MD} et P_{CTD}/P_{XD}

Chez tous les *Paramyxovirinae*, P_{MD} et P_{XD} sont séparés par une région flexible prédite comme faiblement ordonnée [40]. D'après des études par RMN de la région 474-568 de P de SeV, la région en amont de P_{XD} (474-515) est désordonnée [136, 137]. Chez les Morbilivirus et les Henipavirus, une région flexible supplémentaire est prédite en amont de P_{MD} (figure 7) [39, 40]. De la même manière, des données de RMN et de SAXS indiquent que la région entre P_{MD} et P_{CTD} est désordonnée dans la P de VSV [116, 138].

La P entière a une structure tentaculaire

Il est difficile d'obtenir une description structurale à haute résolution de P car c'est une protéine modulaire flexible comprenant plusieurs domaines. Une représentation atomique de la P entière de VSV a été obtenue récemment sur la base de données de RMN et de SAXS sous la forme d'un ensemble de conformères qui s'échangent continuellement entre eux sous l'effet de l'agitation thermique [138]. Le spectre RMN de P a confirmé la présence de la région désordonnée P_{NT} , de l'entretoise flexible et des domaines structurés P_{MD} et P_{CTD} . La molécule dimérique ressemble à une pieuvre à quatre tentacules avec le corps constitué du domaine central dimérique P_{MD} , deux longs tentacules P_{NT} et deux autres bras plus courts associant l'entretoise flexible terminée par le domaine globulaire P_{CTD} (figure 6C). Les différentes parties flexibles de la molécule bougent indépendamment les unes des autres sans interaction entre elles. Une méthode de sélection d'ensemble [139] a permis de générer des ensembles de conformères compatibles à la fois avec les données de RMN et de SAXS (figure 6D). Dans ces ensembles, la protéine adopte une large gamme de conformations avec des dimensions et des orientations relatives des parties N- et C-terminales très différentes. La grande flexibilité et la taille de la molécule pourraient être déterminantes dans les fonctions de P au sein du complexe de transcription/réplication. En agissant comme une « brosse » entropique, les molécules de P pourraient auto-réguler leur espacement sur la NC. Les longs tentacules P_{NT} flexibles, en échantillonnant un grand volume autour de la molécule, favoriseraient la capture des partenaires viraux et cellulaires telles que N^0 , la protéine L, des kinases ou des karyophérines. Cela faciliterait également le positionnement et l'orientation correcte de la polymérase au site de synthèse de l'ARN et la livraison du substrat N^0 au site d'encapsidation. Comparées à la taille de la particule virale ou de la NC, les dimensions de la molécule sont singulières. La particule en balle de fusil de VSV a une longueur moyenne de 200 nm, un rayon extérieur de 35 nm et une cavité centrale d'un rayon de 15 nm. Or, cette cavité est supposée contenir les molécules P dont la distance maximale

entre les extrémités N- et C-terminale de la protéine est de l'ordre de 18 nm, donc supérieure au rayon de la cavité. La distance entre les deux domaines C-terminaux d'un dimère peut atteindre 9 nm soit une valeur bien supérieure à la distance séparant deux tours successifs de N^{NUC} -ARN (environ 5 nm). Donc, un même dimère de P pourrait connecter des tours successifs de la NC et stabiliser la particule virale.

Une P plus courte chez les *Pneumovirinae*

Comparée aux P des autres familles de MNV, la protéine P des *Pneumovirinae* est beaucoup plus courte et ne partage aucune similarité de séquence. Aussi indispensable et assurant les mêmes fonctions clés dans le complexe de transcription/réplication à travers un partenariat moléculaire identique, elle a une organisation modulaire et fonctionnelle canonique (figure 8). La P de RSV est un homotétramère de forme allongée d'environ $12,5 \pm 1$ nm de long [115, 127] avec un domaine central (résidus ~100-160) modélisé en hélices α surenroulées (« coiled-coil ») résistant à la protéolyse ménagée [115, 140] et responsable de la tétramérisation [115, 127, 140-142], une région N-terminale (résidus 1-99) non structurée et enfin une région C-terminale peu ou pas structurée. Les dix derniers résidus C-terminaux sont responsables de l'interaction avec la NC (voir ci-dessous) [143]. La région située entre les résidus 161 et 180 est la plus conservée entre les P des Pneumovirus et Métopneumovirus. De fonction inconnue, elle pourrait présenter un certain degré de structure secondaire [140].

Contrairement aux P des *Rhabdoviridae* et *Paramyxovirinae*, chez RSV le domaine d'interaction entre P et N^{NUC} n'est pas structuré et beaucoup plus petit [143]. En effet, dix résidus sont suffisants à P pour fixer des NC en anneaux, d'où son acronyme P_{10CT} . Les résidus impliqués dans cette interaction sont hydrophobes ou acides et comportent une phénylalanine indispensable comme dernier résidu (F241) (figure 8).

Cette région terminale de P se fixe au niveau du domaine N-terminal globulaire de N (N_{NTD} représentée en mauve en figure 4C, E) au niveau d'une poche hydrophobe entourée de résidus plutôt basiques. L'introduction de mutations dans cette poche est délétère pour l'interaction P-N et la synthèse d'ARN viral [144]. Donc, la topologie structurale des interactions P_{CT} et N_{NTD} impliquée dans la synthèse d'ARN est inversée entre *Paramyxovirinae* et *Pneumovirinae*, les domaines d'interaction entre N et P étant respectivement structurés et non structurés chez les *Pneumovirinae*, et vice et versa chez les *Paramyxovirinae*.

Structure de M2-1

La protéine M2-1 est codée par la première ORF du gène bicistronique M2. Elle est spécifique des *Pneumovirinae*.

Elle ne partage aucune homologie de séquence avec d'autres protéines de MNV, mais, chez les *Filoviridae*, elle a un homologue fonctionnel, VP30. M2-1 est un facteur de transcription pour le RSV. Elle accroît la processivité de la transcriptase virale, en augmentant l'efficacité de transcription. Elle favorise la traversée non-stop de la région intergénique, et donc la synthèse de transcrits bicistroniques. Cela pourrait constituer un mécanisme pour permettre à la polymérase d'accéder aux gènes avals avec un meilleur rendement. Cette protéine très basique de 22 kDa est un tétramère très stable avec une organisation modulaire comprenant une région dite en « doigt de zinc » (résidus 1-31), qui fixerait un atome de zinc par monomère [145], un domaine d'oligomérisation (résidus 32-61), un domaine globulaire responsable des interactions avec P et l'ARN (domaine cœur ou « core », résidus 75-170), et une région C-terminale non structurée (figure 8) [143]. La structure atomique de la région cœur a pu être résolue par RMN [146]. Ce domaine est structuré en six hélices α avec une face riche en résidus basiques et donc très chargée positivement (figure 8). Ce domaine interagit de manière compétitive avec P et l'ARN. Une analyse par mutagenèse systématique a permis de définir précisément les surfaces d'interaction avec P et l'ARN comme se recouvrant partiellement avec certains résidus engagés exclusivement dans l'interaction avec P ou l'ARN [146]. Cela laisse présager une compétition entre la fixation de la P et de l'ARN. La mutagenèse de résidus impliqués dans l'interaction avec P ou l'ARN est fatale à l'activité transcriptionnelle. L'interaction avec P permet de recruter M2-1 au niveau des centres de réplication viraux appelés corps d'inclusion cytoplasmiques. Malgré l'absence de similarité de séquence, la similarité structurale est frappante entre le domaine cœur et la région C-terminale de la protéine VP30 du virus Ebola, ce qui suggère une homologie [146, 147]. Comme M2-1, VP30 forme des oligomères, possède un domaine prédit en doigt de zinc, et assure la complétion des ARNm naissants. Cette homologie structurale et fonctionnelle plaide en faveur d'une origine commune pour VP30 et M2-1, sans exclure pour autant une convergence.

En résumé, d'un point de vue structural, les complexes polymérasiques des *Pneumovirinae* et des *Paramyxovirinae* présentent des différences notoires avec comme éléments distinctifs des premiers, une taille réduite de la P, l'absence de N_{TAIL}, le site d'interaction avec P sis au sein du domaine globulaire N_{NTD}, l'inversion de la combinaison structuré/non structuré dans l'interaction entre P_{CTD} et N_{CORE}, et la présence de cofacteurs supplémentaires M2-1 et M2-2. En fait, l'homologie structurale et fonctionnelle entre M2-1 (RSV) et VP30 (virus Ebola), et la phylogénie basée sur la séquence du gène L [148], indiquent que le RSV et les *Pneumovirinae* seraient, d'un point de vue évolutif, des cousins plus proches des *Filoviridae* que des *Paramyxovirinae*.

Rôle de P dans l'activité de la polymérase virale à deux sous-unités

Structure du complexe N^{NUC}-ARN-P chez les *Rhabdoviridae*

P est un cofacteur essentiel de la polymérase, car elle assure l'attachement de la protéine L sur la matrice N^{NUC}-ARN. En effet, la polymérase seule se fixe avec une faible affinité sur le complexe N^{NUC}-ARN, et son activité en absence de P est peu processive. En possédant un site de fixation pour le complexe N^{NUC}-ARN dans sa région C-terminale et un site de fixation pour la polymérase dans sa partie N-terminale ou dans le domaine central d'oligomérisation, P assure le recrutement de L sur le complexe homopolymérique N^{NUC}-ARN. De plus, P interviendrait dans le déplacement de la polymérase le long de sa matrice au cours des phases de transcription et de réplication.

Chez les *Rhabdoviridae* comme chez les *Paramyxovirinae*, l'attachement de P sur la matrice N^{NUC}-ARN implique l'interaction entre le domaine C-terminal de P et l'extrémité C-terminale de N, mais le mode d'interaction est différent. Chez les *Rhabdoviridae*, la région C-terminale de P forme un domaine globulaire P_{CTD} (figure 6A, B). La région C-terminale de N, N_{CTD}, est constituée d'une boucle suivie de trois hélices qui forment la surface d'interaction au sommet de la molécule de N. La structure d'un complexe entre un anneau N^{NUC}-ARN de VSV contenant dix protomères de N et dix domaines P_{CTD} a été résolue par cristallographie [48]. Chez RABV, d'après des données de titration, seulement deux P_{CTD} peuvent se fixer simultanément sur un anneau de dix ou 11 protomères de N (figure 9A). Cette fixation est non coopérative avec une constante de dissociation de 160 ± 20 nM [149]. Un modèle structural de ce complexe a été construit par une méthode d'arrimage moléculaire flexible sur la base des données de SAXS et en utilisant des contraintes de contact entre P_{CTD} et certains résidus de N identifiés par mutagenèse chez MoK (figure 9A) [150].

Dans le modèle de RABV et dans la structure cristallographique de VSV, le domaine P_{CTD} est fixé sur le sommet du N_{CTD} d'un protomère N et à l'extrémité C-terminale du bras N_{CTA}. L'hélice C-terminale de P_{CTD} se positionne sur l'hélice α_{14} de N_{CTD} et l'extrémité N-terminale de P_{CTD} pointe vers l'intérieur de l'anneau. Dans les deux complexes, les boucles C-terminales du bras N_{CTA} de deux protomères adjacents (N_i et N_{i+1}) se fixent sur les faces latérales opposées du P_{CTD} à la manière des mâchoires d'une pince (figure 9B). Cette géométrie N_{iCTA}/P_{CTD}/N_{i+1CTA} en pince explique que P_{CTD} se fixe à l'homopolymère N^{NUC}-ARN (ou NC) mais non au monomère N⁰. La cohérence architecturale de l'hélice α_{14} de N_{CTD} avec son

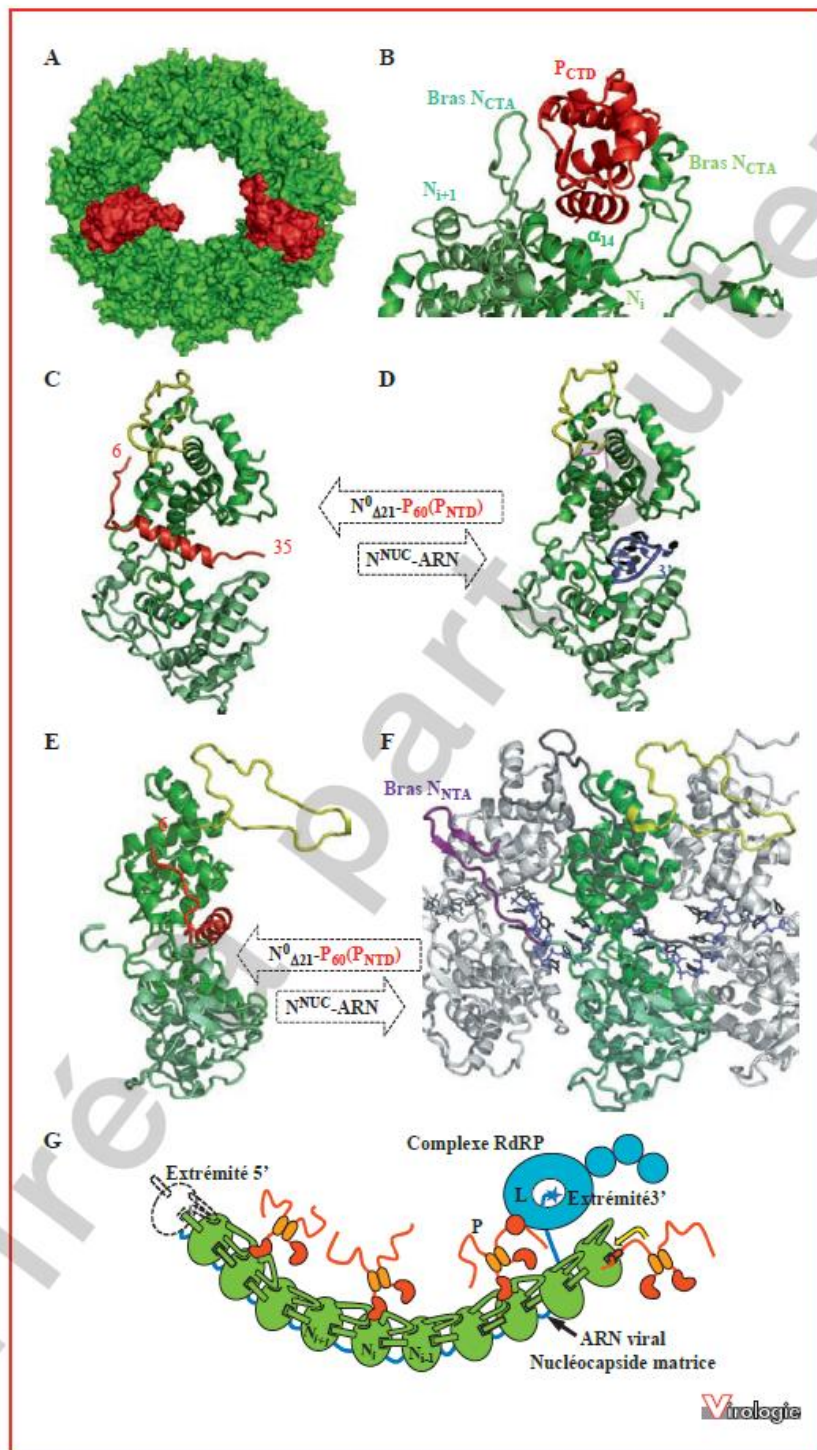


Figure 9.

_ 244

environnement est critique avec des mutations ponctuelles ayant un effet délétère soit sur l'encapsidation de l'ARN soit sur la transcription, et ce indépendamment de la liaison à P_{CTD} [151].

Dans les structures du complexe N^{NUC}-ARN du RABV et du VSV, l'ARN est totalement séquestré par N. Comment donc la polymérase peut-elle avoir accès à l'ARN ? Selon une première hypothèse, la fixation de P sur le complexe N^{NUC}-ARN induirait un changement de conformation de N avec libération de l'ARN qui deviendrait donc accessible pour la polymérase. Cependant, la structure du complexe N^{NUC}-ARN-P_{CTD} du VSV résolue par cristallographie et la modélisation du complexe correspondant du RABV à partir de données de SAXS ne montrent aucun changement de conformation de N après fixation de P_{CTD} sur des anneaux nucléocapsidiques de N^{NUC}-ARN [48, 149]. Les données expérimentales sont-elles biaisées à cause de la circularisation artificielle du complexe N^{NUC}-ARN étudié qui piègerait N dans sa conformation « fermée », et/ou du fait de l'absence des autres régions de P ? Ou bien la protéine L est-elle nécessaire pour ce processus ? Quoiqu'il en soit, la juxtaposition immédiate au sein de N_{CTD} du bras N_{NTA} liant le lobe N_{CTD} du protomère N^{NUC} amont et du site de fixation à P_{NTD} suggère l'intervention de P dans l'ouverture des protomères N^{NUC} pour dégager l'ARN génomique et son utilisation comme matrice par la polymérase virale. Une seconde hypothèse, formulée d'après la structure du complexe N⁰-P, sera discutée ci-dessous.

Complexes N_{TAIL}-P_{XD} des Paramyxovirinae

La fixation de la P sur le complexe N^{NUC}-ARN des Paramyxovirinae repose sur une stratégie moléculaire très différente. En effet, la partie de N fixant la P est désordonnée en l'absence de cette dernière.

D'après des analyses par CD et RMN, les domaines N_{TAIL} de MeV, NiV et HeV subissent un repliement α -hélicale induit lors de la liaison avec le domaine X de la protéine P homologue [34, 37, 86, 98]. Comme prédit [98], et d'après

la cristallographie aux rayons X [78], l' α -MoRE localisé au sein de la Box2 de N_{TAIL} du MeV adopte une conformation α -hélicale et vient s'insérer dans un large sillon hydrophobe délimité par les hélices α 2 et α 3 de P_{XD} avec une constante de dissociation (K_D) faible de 100 nM [37, 152] (figure 10). Comme chez MeV [98, 132], l'interaction entre P_{XD} et N_{TAIL} chez HeV et NiV a une stœchiométrie 1:1 [130]. En revanche, les complexes N_{TAIL}-P_{XD} d'Henipavirus sont caractérisés par un K_D de l'ordre du μ M (2 μ M pour NiV et 8 μ M pour HeV) [86]. D'après des études par titration calorimétrique isotherme (ITC) des complexes N_{TAIL}-P_{XD} des Henipavirus, ces complexes sont stables même en présence de fortes concentrations en NaCl (1 M) [86], ce qui suggère une interaction non polaire comme chez MeV [98, 132]. D'après l'analyse de formes de N_{TAIL} des Henipavirus tronquées d'un ou plusieurs des quatre α -MoREs prédits, c'est l' α -MoRE localisé dans la Box3 (résidus 473 à 493) qui est impliqué dans la liaison à P_{XD} [99]. De plus, les domaines N_{TAIL} et P_{XD} des Henipavirus forment des complexes hétérologues, ce qui témoigne de l'interchangeabilité fonctionnelle de ces α -MoREs entre les deux virus [99].

La structure des complexes N_{TAIL}-P_{XD} des Henipavirus a été modélisée d'après la structure du complexe du MeV [86]. Dans ces modèles, le côté hydrophobe de l' α -MoRE localisé au sein de la Box3 de N_{TAIL} viendrait se loger à la surface hydrophobe de P_{XD} délimitée par les hélices α 2 et α 3 (figure 10). Ces modèles sont validés par des approches de spectroscopie de résonance paramagnétique électronique (RPE) en marquage de spin [100]. En accord avec les différentiels de K_D, l'interface mise en jeu est plus petite chez NiV et HeV (respectivement, 439 Å² et 337 Å²) que chez MeV (634 Å²) (figure 10).

Chez SeV, le complexe N_{TAIL}-P_{XD} du SeV partage une organisation structurale remarquablement similaire, mais l'interface est dominée par des résidus chargés contrairement aux interfaces hydrophobes chez MeV, HeV et NiV, et l'affinité est plus faible avec un K_D d'environ 60 μ M [89]. C'est une bonne illustration de la façon dont la pression de sélection a permis aux domaines C-terminaux de N et de

Figure 9. (Suite) Structures des complexes N^{NUC}-P et N⁰-P des *Rhabdoviridae*. **A**) Structure du complexe entre un anneau de N^{NUC}-ARN de RABV contenant 11 protomères de N (vert) et deux domaines P_{CTD} (rouge). Le modèle a été obtenu par une procédure de modélisation moléculaire impliquant de l'arrimage flexible et utilisant des données de SAXS et de mutagenèse comme contraintes. **B**) Vue externe détaillée de l'interaction entre un P_{CTD} (rouge) et deux protomères adjacents de N^{NUC} (vert foncé et clair). P_{CTD} repose sur l'hélice α 14 du protomère N_i et est pris en étau entre les bras N_{CTA} des protomères N_i et N_{i+1}. **C, D**) Vue antérieure des complexes N⁰ _{Δ 21}-P₆₀ et N^{NUC}-ARN, avec P₆₀ correspondant aux premiers 60 résidus de P et couvrant donc P_{NTD}/ α -MoRE : C : complexe soluble N⁰ _{Δ 21}-P₆₀ (code pdb 3PMK) avec l' α -MoRE de P_{NTD} visible dans la structure cristallographique (rouge) qui encombre la moitié gauche du site de fixation de l'ARN et bloque la fixation de ce dernier. **D**) Protomère du complexe N^{NUC}-ARN (code pdb 2GIC) avec les domaines N_{CTD} (vert foncé) N_{NTD} (vert clair), le bras N_{CTA} (jaune), le bras N_{NTA} (magenta) et l'ossature sucre-phosphate (bleue) et les bases nucléotidiques de l'ARN (noir). **E**) Vue arrière du complexe soluble N⁰ _{Δ 21}-P₆₀ avec l'extrémité N-terminale de P₆₀ (rouge) fixée sur N_i en lieu et place du bras N_{NTA} du protomère N_{i+1} dans sa gorge d'accueil de la forme polymérique N^{NUC}-ARN. **F**) Vue arrière du complexe N^{NUC}-ARN avec l'échange entre protomères adjacents des bras N_{NTA} et N_{CTA}. Le bras N_{NTA} du protomère N_i (protomère vert) vient se fixer dans une gorge à l'arrière du protomère N_{i+1} (protomère en gris à droite). **G**) Mécanisme proposé de stimulation de l'initiation de la synthèse d'ARN par P [49]. En fixant son P_{NTD}/ α -MoRE sur la première N^{NUC} à l'extrémité 3' du génome à droite (flèche jaune), comme dans le complexe N⁰-P, P initierait l'ouverture de ce premier protomère N^{NUC} permettant ainsi à la polymérase L d'accéder à l'ARN.

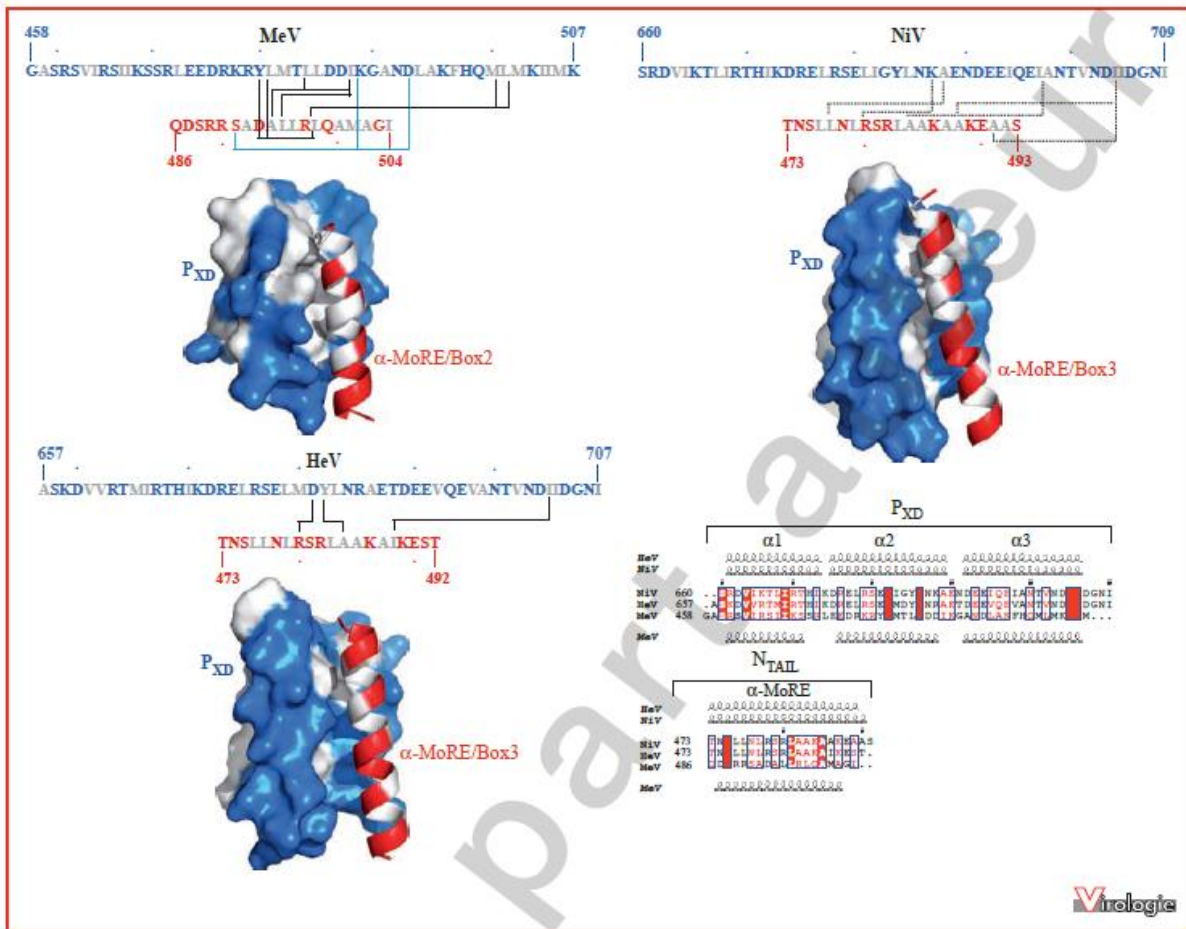


Figure 10. Structure (MeV) et modèles structuraux (NiV et HeV [86]) des complexes entre P_{XD} (bleu) et l'α-MoRE (ruban rouge) au sein de la Box2 (MeV) ou de la Box3 (Henipavirus) avec les résidus hydrophobes (gris-blanc), surmontés des séquences en acides aminés de P_{XD} et de l'α-MoRE selon le même code couleur. Les contacts hydrophobes établis à l'interface dans les différents complexes sont indiqués par des lignes noires en pointillé, alors que les liaisons hydrogènes sont indiquées par des lignes bleues. En bas à droite : alignement multiple de séquences obtenu avec ClustalW et ESPript avec les résidus ayant une similarité > 60 % montrés en rouge sur fond blanc, et les identiques en blanc sur fond rouge. Les nombres indiqués devant les séquences correspondent à la position au sein de P et de N. Les points indiquent des intervalles de dix résidus. Les éléments de structure secondaire des domaines P_{XD} et des α-MoREs de NiV et de HeV prédite par le serveur PSIPRED, et dans le complexe cristallin P_{XD}-N_{TAIL} de MeV (code pdb 1T6O) sont montrés respectivement au-dessus et en-dessous de l'alignement.

P de co-évoluer au sein de la famille des *Paramyxovirinae* avec l'adoption d'une configuration 3D, une architecture de complexe et une fonction identique mais avec une très faible identité de séquence [89].

Bien que les analyses par CD indiquent une transition de type α-hélicale au sein des domaines N_{TAIL} à la fois chez HeV et NiV lors de l'interaction avec le partenaire P_{XD}, des différences subtiles existent entre ces deux virus. En effet, d'après des expériences de titration par RMN, réalisées avec des formes de N_{TAIL} du HeV marquées à l'azote 15, l'ajout de P_{XD} du même virus entraîne la disparition d'une

quinzaine de pics qui ne réapparaissent pas dans le spectre, même une fois la saturation atteinte [86]. Ce comportement, souvent observé chez les PIDs [153], indique un régime d'échange intermédiaire et suggère que l'α-MoRE de N_{TAIL} reste très dynamique à la surface de P_{XD}. Le complexe N_{TAIL}-P_{XD} du HeV serait donc caractérisé par une hétérogénéité conformationnelle beaucoup plus prononcée que celui du NiV.

Dans le cas du MeV, la disponibilité de l'attribution complète des spectres des formes libres et liées de N_{TAIL} [95], et les études de RPE en marquage de spin [96, 97]

ont permis d'élucider les mécanismes de formation du complexe N_{TAIL} - P_{XD} . La région 488-502 subit une transition α -hélicale lors de l'interaction avec P_{XD} , ce qui entraîne une réduction de mobilité dans la région en aval [96, 97], qui conserve cependant une liberté conformationnelle élevée [154] au sein du complexe. La mobilité réduite de la région 488-502, même en l'absence du partenaire, suggère que la Box2 est au moins partiellement pré-configurée en α -hélice avant d'interagir avec P_{XD} . Cette pré-configuration est confirmée par des études de RMN [92, 95]. Grâce à des approches originales permettant d'obtenir une description quantitative d'ensembles conformationnels des PIDs sur la base de données RMN expérimentales, une description à résolution atomique de la forme isolée de MeV N_{TAIL} a ainsi été obtenue. Les déplacements chimiques et les couplages dipolaires résiduels (RDCs), ont été combinés pour sonder directement le niveau et la nature des structures résiduelles du domaine N_{TAIL} du MeV. Si la majorité de N_{TAIL} se comporte comme une chaîne désordonnée, l' α -MoRE adopte un équilibre dynamique entre une forme non repliée et des conformères α -hélicaux (figure 5D). Un comportement très similaire est également observé dans le domaine N_{TAIL} de SeV [155]. La présence d'une hélice α transitoire en l'absence du partenaire suggère que le mécanisme moléculaire qui gouverne le repliement de N_{TAIL} induit par P_{XD} reposerait sur une sélection conformationnelle (à savoir une sélection par le partenaire d'une conformation préexistante). Cependant, d'après une analyse quantitative de titrations RMN, l'interaction pourrait impliquer également un intermédiaire de liaison sous la forme d'un complexe faible et non spécifique [156]. Ainsi, le repliement couplé à la liaison du domaine N_{TAIL} du MeV résulterait plutôt d'un mécanisme impliquant à la fois un repliement induit et une sélection conformationnelle, un comportement qui a déjà été observé pour d'autres complexes impliquant des PIDs [157].

Des études par SAXS ont fourni un modèle à basse résolution du complexe N_{TAIL} - P_{XD} du MeV où une grande partie de N_{TAIL} reste désordonnée au sein du complexe [37]. De plus, d'après des études combinant la spectroscopie RPE en marquage de spin et la modélisation, la région 505-525 de N_{TAIL} conserve une liberté conformationnelle élevée dans sa forme liée en dépit du gain de rigidité induit par la liaison avec P_{XD} [154]. De ce fait, le complexe N_{TAIL} - P_{XD} constitue un exemple illustratif de liaison floue (« fuzziness »), un terme récemment introduit pour désigner la persistance de régions désordonnées au sein de complexes dont l'un des partenaires est une PID [158]. Les complexes N_{TAIL} - P_{XD} des Henipavirus fournissent un autre exemple de liaison floue. En effet, d'après le comportement des résonances du domaine N_{TAIL} du HeV lors des titrations RMN, ce dernier demeure dynamique après liaison à P_{XD} . En outre,

le rayon de Stokes (R_S) du complexe N_{TAIL} - P_{XD} du NiV ($35,4 \pm 3,1 \text{ \AA}$) est compatible avec une liaison à P_{XD} associée à une compaction seulement partielle (R_S attendu de $22,3 \text{ \AA}$ si complexe compact) et la conservation par N_{TAIL} d'une grande flexibilité [86]. De plus, chez les trois virus, la majorité des pics de N_{TAIL} n'est pas affectée par l'ajout du partenaire [37, 86, 95], ce qui souligne une fois de plus la persistance du désordre au sein de la forme liée de N_{TAIL} . D'ailleurs, l'introduction d'une mutation dans la Box2, entraînant la perte de la capacité à subir un repliement α -hélicale lors de l'interaction avec P_{XD} ne perturbe notablement ni l'affinité pour P_{XD} , ni l'activité polymérase dans un MeV recombinant [152].

Interaction P - N^{NUC} -ARN chez les *Pneumovirinae*

Rappelons que la situation est encore différente chez RSV. Le site d'interaction de P sur N polymérique complexée à l'ARN est situé dans le domaine globulaire N-terminal de N (N_{NTD}) (figure 4C) [144]. Il est constitué d'un sillon hydrophobe entouré de résidus plutôt basiques (chargés positivement). La partie C-terminale de P (P_{10CT}), hydrophobe et acide, vient se loger dans ce sillon situé à la périphérie de la NC hélicoïdale avec un K_D d'environ $50 \mu\text{M}$ (figure 4E) [143]. La structure du complexe n'est pas établie, ni sa dynamique.

Déplacements de la polymérase sur la matrice N^{NUC} -ARN

Deux modèles ont été proposés pour expliquer comment la polymérase virale se déplace le long de sa matrice lorsqu'elle synthétise de l'ARN.

Modèle d'une progression de P-L « faisant la roue » ou « cart-wheeling »

Un premier modèle propose que P (associée à L) avance le long du complexe N^{NUC} -ARN « en faisant la roue » (« cart-wheeling ») par un mécanisme dans lequel les domaines P_{XD} s'attachent et se détachent alternativement de la matrice nucléocapsidique hélicoïdale à la manière d'un tétrapode « faisant la roue » en descendant chaque marche (c'est-à-dire chaque N_{TAIL}) d'un escalier en colimaçon (figure 1C et animation en fichier supplémentaire) [10, 159]. Chez SeV, ce modèle est inféré :

- de la faible affinité du P_{XD} pour le complexe N^{NUC} -ARN ($60 \pm 20 \mu\text{M}$) ;
- du caractère indispensable du domaine central de tétramérisation pour permettre la fixation de P_{XD} dans des expériences de « pull-down » ;

– pour une synthèse efficace de l'ARN à partir de minigénomes [160].

À travers la fixation simultanée de plusieurs P_{XD}, c'est-à-dire en augmentant l'avidité de P pour le complexe N^{NUC}-ARN, le domaine tétramérique permettrait la synthèse d'ARN.

Ce modèle prédit qu'une stabilité accrue du complexe N_{TAIL}-P_{XD} ralentit ou même bloque le cheminement du complexe polymérase le long de la matrice nucléocapsidique [10, 159]. Cependant, une forte affinité entre les protéines N et P n'est probablement pas incompatible avec une bonne processivité de la polymérase d'après les faits expérimentaux suivants. Malgré une différence d'un facteur 600 entre le K_D du couple N_{TAIL}-P_{XD} chez SeV (60 μM) [89] et chez MeV (100 nM) [152], la vitesse d'élongation apparente de leur polymérase est du même ordre de grandeur avec 1,7 nt par seconde chez SeV [161] et 3,3 nt par seconde chez MeV [31]. De plus, cette vitesse reste relativement constante quelle que soit l'affinité entre N_{TAIL} et P_{XD} : en effet des mutations au sein de la Box2 chez ce dernier virus entraînant une baisse de l'affinité envers P_{XD} pouvant aller jusqu'à un facteur 30 ne semblent pas modifier la vitesse d'élongation [152].

Comment réconcilier le modèle d'une progression de P-L « faisant la roue » avec une forte affinité du couple N_{TAIL}-P_{XD} du MeV, si tant est qu'il soit correct ? Peut-être faut-il faire intervenir un troisième acteur tel qu'un facteur cellulaire capable de moduler la force de l'interaction N_{TAIL}-P_{XD}. Hsp70 est l'un des candidats possibles [162-164]. Des études fonctionnelles visant à élucider le rôle de la multimérisation de la protéine P du MeV et/ou l'importance de l'expression d'Hsp70 par l'utilisation de cellules infectées appauvries en Hsp70 (et/ou en hsc70, une protéine apparentée) apporteront sans doute des éléments de compréhension.

Modèle d'une progression par saltation de L sur plots fixes de P

Chez les *Rhabdoviridae*, les données expérimentales ne sont pas en faveur du modèle de progression « en faisant la roue » :

- l'affinité du domaine P_{CTD} monomérique de RABV pour le complexe N^{NUC}-ARN est forte (K_D de 160 ± 20 nM) [149] ;
- la lente cinétique de dissociation mesurée par résonance plasmonique de surface semble incompatible avec un mécanisme dans lequel la P progresserait en dissociant continuellement sa paire de P_{CTD} de la N^{NUC} ;
- l'ablation du domaine de dimérisation P_{MD} rend la P monomérique [106] et n'empêche pas une synthèse effi-

cace d'ARN dans un essai fonctionnel de transcription d'un minigénome *in cellula*¹ [150].

Un mécanisme alternatif a donc été proposé selon lequel la polymérase L progresse par saltation entre les P fixées à intervalles réguliers tout le long de la matrice N^{NUC}-ARN (figure 1A) [71, 149]. La grande flexibilité de la P permettrait de capturer et relayer la L à longue distance le long de cette matrice. Comme les particules de VSV contiennent ~1 dimère de P pour 5 N [61], on imagine une L séjournant temporairement sur un même dimère de P le temps de copier l'ARN encapsidé par les cinq monomères de N contigus entre chaque saut de P_i à P_{i+1}. Selon ce modèle, et pour reprendre notre langage imagé, les marches de l'escalier N^{NUC}-ARN en colimaçon ne seraient plus les N_{TAIL}, mais des dimères de P fixés tous les cinq protomères de N sur lesquels L, au lieu de P-L, se déplacerait.

Le complexe N⁰-P et les mécanismes d'encapsidation et d'initiation

Le processus d'encapsidation

L'encapsidation des antigénomes et des génomes néo-synthétisés est une étape essentielle dans le processus de la multiplication du virus dans la cellule infectée car la polymérase utilise exclusivement de l'ARN encapsidé comme matrice pour une transcription et une réplication efficace.

Le processus d'encapsidation se déroule de manière concomitante à la synthèse d'ARN. Il requiert donc la production continue de N monomérique et soluble dépourvue d'ARN cellulaire [111, 165]. Lors de la synthèse de N en présence de P se forme un complexe soluble N⁰-P. Il sert de substrat pour la polymérase en assurant le transfert de N⁰ sur le brin naissant d'ARN pour former la NC par homopolymérisation dirigée, unidirectionnelle et régulière depuis l'extrémité 5' de l'ARN jusqu'à son extrémité 3' [166, 167]. Cette exigence expliquerait qu'un peptide comprenant l'α-MoRE N-terminal de P et le site de fixation de N⁰ puisse inhiber la réplication d'un mini-génome ou du virus [118]. Comment s'effectue le transfert de N⁰-P vers la NC naissante et comment est-il régulé ? À ce jour, trop peu de données expérimentales viennent nourrir notre réflexion. Lors des processus de transcription et de réplication, la synthèse d'ARN débute à l'extrémité 3' de la matrice d'ARN. Donc, la décision de synthétiser des ARNm en cascade ou de produire une copie complète de l'ARN génomique doit être prise avant, ou au moment, où la polymérase atteint la fin de la séquence « leader ». D'après les rares observations expérimentales disponibles, et intuitivement,

¹ En latin, *cellula* est féminin. L'expression « *in cellulo* », malheureusement très répandue, est incorrecte.

la concentration du complexe N^0 -P devrait jouer un rôle dans le basculement entre les activités de transcription et de réplication [168], ne serait-ce que pour éviter l'initiation d'une réplication finissant de manière abortive suite à une pénurie de substrat d'encapsidation. La réplication est toujours différée d'un laps de temps (en heures) suffisant pour une production détectable de N et P, ce qui est compatible avec un effet seuil d'une concentration minimale de substrat N^0 -P. Corrélativement, dans une cellule infectée, l'arrêt de la synthèse protéique abroge la réplication sans affecter la transcription primaire. Lorsque cet arrêt intervient en cours de réplication, celle-ci s'interrompt rapidement et les polymérases en mode réplisase semblent basculer en mode transcriptase, comme si le complexe N^0 -P, ou un cofacteur cellulaire clef, avait une demi-vie courte [31]. Chez VSV, la phosphorylation de P serait impliquée dans cette régulation [33]. Le transfert de N^0 vers l'ARN pourrait être contrôlé par une affinité plus élevée de N^0 pour des signaux d'encapsidation présents dans les régions *leader* (et *trailer*) en 3' de l'ARN génomique (et antigénomique) comparativement à celle vis-à-vis de P [166]. Alternativement, en s'associant à P-L, le complexe N^0 -P pourrait commuter le complexe polymérasique de transcriptase à réplisase [102].

Le substrat d'encapsidation N^0 -P

L'interaction entre N^0 et P fait intervenir un court α -MoRE de la partie N-terminale de P couvrant les résidus 4 à 40 chez RABV [169] et 11 à 30 chez VSV, ce dernier peptide étant capable de maintenir N dans une forme soluble [170]. Chez VSV, cette région contient deux hélices α instables détectables par RMN [49]. Le complexe N^0 -P a pu être reconstitué entre une N « manchote », dont le bras N-terminal de 21 acides aminés a été tronqué ($N_{\Delta 21}$), et soit un peptide correspondant aux 60 résidus N-terminaux de la P (P_{60}), soit la P entière dimérique [43, 49]. Ces deux complexes sont solubles, ne contiennent pas d'ARN et correspondent respectivement à un hétérodimère ($N_{\Delta 21}:P_{60}$) et à un mélange d'hétérotrimères ($1N_{\Delta 21}:2P$) et d'hétérotétramères ($2N_{\Delta 21}:2P$). Selon une analyse par SAXS, le complexe $N_{\Delta 21}:P_{60}$ a des dimensions et une forme similaire à celle d'un monomère de N. En spectroscopie RMN, seuls les résidus 6-40 de P sont directement impliqués dans l'interaction et les régions en amont (résidus 1-5) et en aval (résidus 41-60) restent flexibles dans le complexe. Le complexe $N_{\Delta 21}:P_{60}$ a cristallisé sous la forme d'un complexe multimérique circulaire, dans lequel les molécules de N sont organisées de manière similaire à celle du complexe circulaire N^{NUC} -ARN, autrement dit selon un agencement identique à la forme d'une NC malgré l'absence d'une âme ARN ! Dans chaque protomère de la structure cristallographique, les résidus 6 à 35 de

P_{60} sont parfaitement visibles (figure 9C, E). Les résidus 17 à 30 forment une hélice amphipathique arrimée sur une face latérale de N dans une gorge à l'interface entre les domaines N_{NTA} et N_{CTA} , pratiquement perpendiculaire au sillon d'accueil de l'ARN (figure 9D). Principalement de nature hydrophobe, le site de fixation de P sur N comprend également une distribution de charge bipolaire à sa surface avec son pôle négatif orienté vers le sillon d'accueil de l'ARN riche en résidus basiques, et son pôle positif orienté vers une région négativement chargée sur la face opposée de N. Cette bipolarisation pourrait orienter les partenaires au moment de l'interaction. La région charnière de N doublement impliquée dans la fixation de l'extrémité N-terminale de P et de l'ARN est la région la plus conservée chez les *Rhabdoviridae* [171]. Elle est également impliquée dans la régulation de la transcription et de la réplication [172].

Un scénario pour l'encapsidation de l'ARN naissant

La P est donc un double chaperon, car la fixation de P sur N^0 doit à la fois prévenir la fixation d'ARN non spécifique et la polymérisation de N, deux processus intervenant sur deux sites indépendants l'un de l'autre [76]. La comparaison de la structure cristalline du complexe $N_{\Delta 21}:P_{60}$ avec celle d'un anneau de NC N^{NUC} -ARN illumine notre compréhension de cette dualité fonctionnelle. D'une part, en occupant une partie de la cavité de fixation de l'ARN, la partie C-terminale de l' α -MoRE de P_{NTD} bloque la fixation de l'ARN. D'autre part, l'extrémité N-terminale de l' α -MoRE P_{NTD} se fixe sur la face arrière de N dans le sillon d'accueil du bras N_{NTA} du protomère aval N_{i+1} dans une situation de compétition (comparer figure 9E et F) prévenant ainsi la polymérisation de N^0 .

Le mécanisme d'encapsidation de l'ARN en cours de synthèse est inconnu, mais, de ces données structurales et fonctionnelles se déduit un scénario plausible par lequel le transfert de N^0 sur l'ARN serait couplé à la polymérisation de N^{NUC} (figure 11) :

- le complexe N_i^0 - P_i se fixerait sur le complexe N^{NUC}_{i-1} -ARN par l'intermédiaire du PCTD de P_i [173] afin de positionner N_i^0 près du site de synthèse de l'ARN ;
- grâce à sa grande flexibilité et à sa longueur, P_i orienterait correctement N_i^0 pour qu'elle accueille l'ARN viral ;
- l'ARN naissant se positionnerait dans le sillon de N_i^0 délogeant ainsi l'extrémité C-terminale de l' α -MoRE de P_{NTD}_i ;
- cette déstabilisation entraînerait un relâchement de la liaison de l'extrémité N-terminale de l' α -MoRE de ce même P_{NTD}_i qui encombre le sillon d'accueil du bras N_{NTA} de N_i^0 ;
- en conséquence, à l'arrivée du complexe suivant N^0 - P_{i+1} , le bras N_{NTA} de ce dernier entre en compétition favorable pour déloger l'extrémité N-terminale de l' α -MoRE du

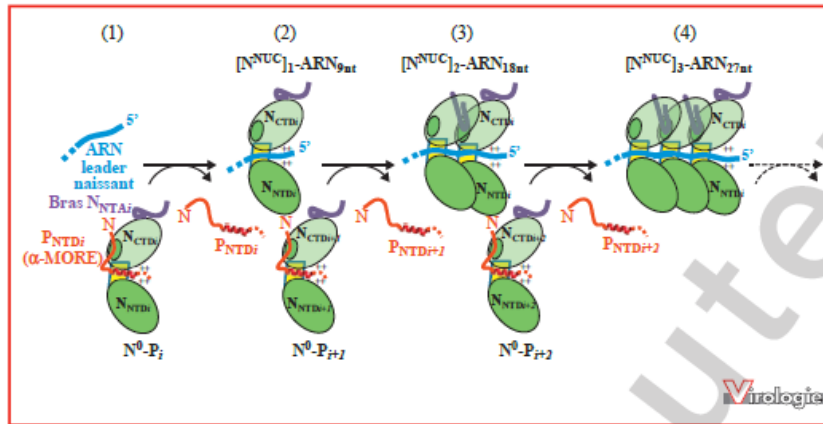


Figure 11. Modèle de l'encapsidation de l'extrémité « leader » de l'ARN génomique naissant par le substrat N^0 -P chez les *Rhabdoviridae* [49]. 1. L'extrémité libre du sillon chargée positivement (ici en face antérieure de la charnière représentée en jaune) d'un premier protomère N^0 -P_i vient au contact de l'extrémité 5' de l'ARN leader chargée négativement en cours de synthèse. La fixation des neuf premiers nucléotides résulte en la formation du premier protomère de la NC $[N^{NUC}]_1$ -ARN_{9nt} et occasionne le déplacement par compétition de l'extrémité P_{NTDi}/α-MORE localisée en écharpe latérale chevauchant la moitié gauche du sillon antérieur et la gorge postérieure (représentée par l'ovale vert en arrière de N_{CTD}). 2. Le départ de P_{NTDi}/α-MORE libre, sur la face arrière de la charnière, la gorge pour permettre. 3. La fixation du bras N_{NNTAi+1} du protomère suivant N^0 -P_{i+1}. Cela permet l'adossement de ce dernier au protomère N^{NUC} _i-ARN et l'accueil des neuf nucléotides suivants de l'ARN dans son sillon. Ce nouveau segment d'ARN vient déloger le P_{NTDi}/α-MORE du complexe N^0 -P_{i+1} pour former $[N^{NUC}]_2$ -ARN_{18nt}. 4. La répétition de ce cycle permet l'enchaînement continu et ordonné de l'encapsidation au fur et à mesure de l'élongation du génome. Par simplification, P n'est représentée que par son domaine P_{NTD}. De même, ne sont représentés ni le complexe polymérisique, ni la matrice NC. Glossaire : gorge : site de fixation du bras N_{NNTA} du monomère de N^{NUC} suivant ; sillon : emplacement délimité par les deux domaines N_{CTD} et N_{NTD} dans lequel vient se loger la chaîne phospho-glycosidique de l'ARN génomique viral.

P_{NTDi} amont resté provisoirement associé à la NC en formation. Il positionne ainsi le protomère N^0 _{i+1} de manière à accueillir le segment suivant d'ARN naissant ;
 – devenue libre, P_i peut alors être recyclée (ou détruite).
 Il est remarquable que les structures de N^0 et de N^{NUC} soient parfaitement similaires, exception faite de la présence de partenaires et du positionnement du bras N_{NNTA} et du bras N_{NCTA}. Sinon, comment expliquer que l'empilement cristallin du complexe $N^0_{\Delta 21}$ -P₆₀ monomérique soluble reproduise à l'identique la structure d'un anneau de NC N^{NUC} -ARN ? Cet élégant scénario serait particulièrement économique en énergie tout en étant d'une grande fiabilité, la structure de chaque protomère de N étant « prête à poser » et son emboîtement guidé. De plus, il est défavorable à une recombinaison, recombinaison non documentée à ce jour chez les MNV.

Accéder à la matrice ARN génomique enfouie dans le manchon de l'homopolymère de N

En sus de son rôle dans l'encapsidation, la fixation de P_{NTD} sur N suggère un mécanisme d'initiation de la synthèse d'ARN à l'extrémité 3' de l'ARN [49]. L'ARN encapsidé

par N^{NUC} n'est pas accessible pour la polymérase, mais la première molécule N^{NUC} à l'extrémité 3' du génome expose son site d'accueil à l'α-MoRE du P_{NTD} [49]. Pourquoi ne pas envisager qu'en se fixant sur ce site, une molécule de P du complexe P-L déplace l'ARN de la première nucléoprotéine permettant ainsi à la polymérase de se fixer sur l'extrémité 3' du génome (figure 9G) ? Ce processus expliquerait des résultats étranges obtenus avec un système de transcription et de répllication du génome de SeV reconstitué *in vitro* [168] : l'addition de faibles concentrations du complexe N^0 -P soluble stimule la transcription, alors que l'addition de concentrations plus élevées stimule la répllication et réduit la transcription. L'effet sur la transcription aux faibles concentrations de N^0 -P pourrait s'expliquer non par la présence de N mais par la présence de P_{NTD} libres. L'un d'eux, en se fixant sur N à l'extrémité 3' du génome, déplacerait N de l'ARN et stimulerait l'initiation de la synthèse d'ARN en permettant à la polymérase d'accéder plus aisément à l'ARN génomique (figure 9G). En revanche, aux concentrations plus élevées de N^0 -P, ce complexe stimulerait alors la répllication par rapport à la transcription, conduisant à un basculement entre ces deux activités comme discuté plus haut.

Désordre structural de N et P : un pléomorphisme dynamique gouvernant transcription et réplication ?

N et P des *Rhabdoviridae* et *Paramyxoviridae* sont caractérisées par un important contenu en régions désordonnées. Quel avantage cela représente-il pour la machinerie virale ? Passons en revue les propriétés des régions intrinsèquement désordonnées, leur implication plausible dans différents processus moléculaires, et le bénéfice fonctionnel potentiel pour la transcription et/ou la réplication des MNV.

Propriétés du désordre structural

Une des propriétés caractéristiques des régions désordonnées, qui découle de leur malléabilité, est leur « promiscuité », autrement dit leur aptitude à établir un partenariat moléculaire multiple vis-à-vis de partenaires structurellement hétérogènes. L'absence de repliement confère aux régions désordonnées une plus grande surface de contact externe disponible par acide aminé et une souplesse de conformation permissive à se lover en complexe avec d'autres macromolécules selon une géométrie variable. Cela permet une économie substantielle de longueur de séquence requise pour une interaction spécifique avec des partenaires par rapport à un domaine structuré dont la fraction de séquence enfouie, la rigidité et l'encombrement spatial limite le partenariat potentiel. Ainsi, le domaine N_{TAIL} de MeV interagit avec plusieurs partenaires, parmi lesquels on dénombre les protéines virales P [87], M [14] et N (interaction de la Box1 entre deux N^{NUC} contiguës, positionnées sur deux tours successifs de la NC [92]) et les protéines cellulaires Hsp70 [162, 163], IRF-3 [174], un récepteur membranaire putatif NR [175], une protéine impliquée dans le transport nucléocytoplasmique de N [176] et la peroxirédoxine 1 [177]. De même, la P_{NT} des MNV, en plus de la présence d'un site commun de fixation à N⁰, possède, selon les virus, des sites de fixation pour la protéine L, des kinases et des protéines intervenant dans les voies de signalisation de l'immunité innée et/ou des karyophérines [43, 130].

Une autre propriété clé des régions désordonnées est leur très longue portée qui leur confère un rôle chaudière et tentaculaire permettant le recrutement de partenaires sur de grands assemblages macromoléculaires. Ainsi, d'après des mesures de SAXS, la longueur de P de VSV est de l'ordre de 18 nm entre ses extrémités N- et C-terminales (figure 6D) [138]. L'extension maximale de P_{NT} et de N_{TAIL} mesurée par SAXS chez MeV atteint respectivement 40 nm (Longhi et Receveur-Bréchet, données non publiées) et 13 nm

[87]. Par comparaison, un tour de la NC de MeV représente 18 nm de diamètre et 6 nm de hauteur [57]. Dans les particules de *Rhabdoviridae*, les tours successifs d'hélice sont séparés d'environ 5 nm. Ainsi, chaque protomère de P peut s'étendre facilement sur plusieurs sous-unités de N le long du complexe N^{NUC}-ARN et/ou connecter entre eux des tours successifs de la NC.

Rôle du désordre structural dans la transcription et la réplication

Une fonction de P est d'ancrer la polymérase sur la matrice N^{NUC}-ARN tout au long du déplacement de la polymérase, notamment lorsqu'elle interrompt (avant de la reprendre) la synthèse d'ARN au niveau des régions intergéniques. Une fois fixée sur la matrice N^{NUC}-ARN, l'organisation tentaculaire de la protéine P positionnerait correctement la protéine L sur la matrice afin d'optimiser la synthèse d'ARN. De même, à la manière d'un tentacule, P positionnerait N⁰ exactement au contact du 5' de l'ARN (anti-)génomique naissant, puis pour les N⁰ suivantes les orienterait correctement pour un emboîtement au nm près. Sa plasticité interviendrait également au moment de la réorganisation de ce complexe multimoléculaire lors du déplacement de la polymérase ou serait directement impliquée dans le déplacement comme discuté ci-dessus.

En échantillonnant un grand volume d'espace (activité tentaculaire), l'oligomère de P (figure 6D) ou la région N_{TAIL} (figure 5D) augmente la probabilité de rencontre avec les différents partenaires et pourrait donc augmenter la vitesse de la réaction de fixation selon un mécanisme de « pêche à la mouche » (« fly casting »).

La plasticité des régions désordonnées et la diversité conformationnelle associée aux phénomènes de repliement induit fournissent un mécanisme additionnel permettant de moduler la conformation de la NC. Chez les *Paramyxovirinae*, le repliement induit de N_{TAIL} résultant de l'interaction avec P (et/ou avec d'autres partenaires physiologiques) est postulé avoir un impact sur la conformation de la NC, favorisant ainsi la réplication et/ou la transcription [35]. D'après des observations en microscopie électronique, la présence de N_{TAIL} est associée à un grand degré de fragilité et donc de flexibilité de la NC et modifie sa conformation (figure 2) [56, 74, 87, 92]. Dans le même ordre d'idée, l'incubation des NC du MeV avec P_{XD} induit un déroulement de ces dernières, d'où une amélioration possible de l'accessibilité de l'ARN génomique au complexe polymérase (Bhella et Longhi, données non publiées).

Finalement, lors de la formation de complexes impliquant le repliement induit de régions intrinsèquement désordonnées sur un partenaire, les régions voisines restent désordonnées dans le complexe. Cette liaison floue (« fuzziness ») pourrait également contribuer à l'assemblage

et à la dynamique de complexes multi-moléculaires et interviendrait dans la réplication et la transcription des MNV.

En effet, la nature désordonnée de N_{TAIL} des *Paramyxovirinae* après formation d'un premier complexe pourrait servir d'amorce pour la fixation d'autres partenaires. En accord avec cette hypothèse, la Box3 interagit avec la protéine de choc thermique Hsp70 [163, 178], ce qui stimule la transcription et la réplication de MeV [162-164, 179]. N_{TAIL} se lie à Hsp70 via deux sites de liaison : l'un de haute affinité comprenant l' α -MoRE (K_D de 10 nM), l'autre de faible affinité associé à la Box3 [162, 163, 179]. D'après des expériences de miniréplicon de MeV, l'extrémité C-terminale de N_{TAIL} régule négativement la transcription et la réplication du génome viral [162], reflétant probablement le rôle de la Box3 dans la stabilisation du complexe N_{TAIL}-P_{XD} via l'établissement de contacts faibles et non spécifiques avec P_{XD}. Par compétition avec P_{XD} pour la liaison à N_{TAIL} [163], Hsp70 stimulerait la transcription et la réplication virale en déstabilisant le complexe N_{TAIL}-P_{XD}. Cela entraînerait ainsi des cycles successifs de liaison/dissociation entre la polymérase P-L et sa matrice nucléocapsidique, cycles qui seraient essentiels pour la progression de la polymérase le long de cette dernière dans le contexte d'un déplacement « en faisant la roue » [37, 38, 42, 163].

Dans les complexes N⁰-P et N-RNA-P des *Rhabdoviridae*, la grande taille et la grande mobilité de la P agirait par exclusion entropique et régulerait l'accessibilité à la N. Ainsi, la fixation de dimères de P sur le complexe N-RNA pourrait restreindre l'accès à l'espace autour du complexe, notamment pour d'autres molécules de P. De cette manière, la P régulerait l'espacement entre les molécules de P le long de la matrice N^{NUC}-ARN. Dans le complexe N⁰-P, l'occupation de l'espace autour de la molécule de N par les parties du dimère de P non directement impliquées dans l'interaction avec N⁰ pourrait contribuer, par encombrement stérique dynamique, à l'activité chaperon de P : cela préviendrait notamment le recrutement d'une molécule d'ARN illégitime ou le contact avec d'autres copies de complexe N⁰-P.

Perspectives

Les principes de la transcription et de la réplication des MNV sont connus depuis les travaux biochimiques et fonctionnels pionniers effectués dans les années 1970-2000 essentiellement sur VSV, SeV et RSV. Depuis les années 2000, les données structurales s'accumulent, et sans remettre en cause les principes, nous apportent des éléments moléculaires, fonctionnels et cinétiques qui nous

permettent aujourd'hui de proposer des *scenarii* mécanistiques plausibles et, même dans certains cas, des pistes pour la conception d'antiviraux. Outre les questions posées dans cette revue, d'autres attendent de nouveaux progrès dans l'analyse de cette mécanique. Sans vouloir les citer toutes, une des plus excitantes est de comprendre comment s'effectue l'ouverture du sillon d'accueil de l'ARN au sein de la NC, non seulement à l'entrée en 3' du génome, mais tout du long. Combien de protomères de N^{NUC} sont-ils ouverts simultanément ? Par qui et par quel mécanisme ? Comment se referment-ils et bénéficient-ils d'un effet chaperon ? Comment s'effectue leur ouverture/fermeture bidirectionnelle lors du balayage à la recherche d'un démarrage de gène ? Quelle est la force motrice du déplacement de la polymérase, l'énergie libérée par la condensation des nucléotides et/ou un gradient d'affinité inhérent à l'asymétrie des protomères N^{NUC} ? Le complexe polymérase accède-t-il à la matrice nucléocapsidique par l'intérieur ou par l'extérieur ? Le partenariat N-P dans les complexes N⁰-P et N^{NUC}-ARN-P n'utilise-t-il pas des variantes dynamiques de mécanismes communs ? *In fine*, quel est le rôle primaire de N ? Empêcher toute structure secondaire de l'ARN génomique évitant ainsi une coûteuse activité hélicase soumise de plus aux contraintes variables selon la richesse en GC du segment d'ARN traversé ? Comment est assurée la processivité de la polymérase est une autre question clé. Comment fonctionnent les co-facteurs M2-1 et VP30 ? Ont-ils des équivalents fonctionnels chez les virus dépourvus de ces co-facteurs ? Comment les autres protéines virales de type V et C ou M modulent-elles les activités de la polymérase ? Affaire à suivre.

Note : la cartographie de la région PNT du VSV s'enrichit de trois propriétés, liaison à la L, accroissement de la processivité de la transcriptase sur une courte matrice d'ARN, et induction d'un changement conformationnel de la L [181].

Remerciements. Jean-François Élouët remercie Marie Galloux, Bogdan Tarus, Ilfad Blazevic, Jenna Fix, Marie-Lise Blondot, Virginie Dubosclard, Thi-Lan Tran, Nathalie Castagné, Sylvie Noinville, Mohammed Moudjou, Bruno Da Costa, Bernard Delmas, Stefan Chilmonczyk, Julie Bernard (unité VIM, Inra, Jouy-en-Josas), Marie-Anne Rameix-Welti (Université Versailles-Saint-Quentin), Safa Lassoued, François Bontems et Christina Sizun (Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Gif-sur-Yvette, France), Magali Aumont-Nicaise (IBBMC, Université Paris-XI, Orsay, France), Robert Paul Yeo (Durham University), Jean Lepault et Paloma F. Varela (Laboratoire de Virologie Moléculaire et Structurale, CNRS UPR 3296, Gif sur Yvette), David Bhella (Université de Glasgow), Felix Rey, Rajiv Tawar, Laurence

Damier-Piolle et Stéphane Duquerroy (Unité de Virologie structurale, Institut Pasteur-CNRS URA 3015, Paris) pour leur collaboration aux travaux sur le RSV.

Denis Gerlier remercie S. Plumet (CNRS-UMR5537), P. Duprex (Queen's University of Belfast) et J. Brunel (Inserm U758) pour leur collaboration aux travaux cités.

Marc Jamin remercie Rob Ruigrok, Nicolas Tarbouriech, Irina Gutsche, Guy Schoehn, Thibaut Crépin, Julien Pérard (UVHCI, Grenoble), Euripèdes de Almeida Ribeiro Jr. (MFPL, Vienna), Francine Gérard (IRCM, Montréal), Cédric Leyrat (STRUBI, université d'Oxford), Ivan Ivanov (EMBL, Grenoble), Nicolas Martinez (ILL, Grenoble), Robert Schneider, Malene Ringkjøbing Jensen et Martin Blackledge (IBS, Grenoble), Aurélie Albertini et Danielle Blondel (IVMS, Gif-sur-Yvette), Ambroise Desfosses (EMBL, Heidelberg), Jo Zaccai (ILL, Grenoble) et Noël Tordo (Institut Pasteur, Paris) pour leur collaborations aux travaux sur les Rhabdoviridae. FG, CL et FY ont été financé par une bourse thèse MENRT du gouvernement Français, II par une bourse de thèse de l'Institut Laue Langevin et NM par une bourse de thèse de la Région Rhône-Alpes. EAR a été financé par une bourse post-doctorale de l'université Joseph-Fourier de Grenoble. MJ remercie la fédération TGIR-RMN-THC pour l'accès aux spectromètres, l'ILL et l'ESRF pour l'accès aux lignes et le Partnership for Structural Biology pour l'excellent environnement scientifique et l'accès à ses différentes plateformes.

Sonia Longhi remercie David Karlin (université d'Oxford), François Ferron, Bruno Canard, Philippe Lieutaud, Stéphane Gely, Hervé Darbon (AFMB, CNRS et université d'Aix-Marseille), Cédric Bernard (université de New York), Valérie Belle, Bruno Guigliarelli (BIP, CNRS et université d'Aix-Marseille), Kenth Johansson (AstraZeneca), Michael Oglesbee (université de Ohio), Marie Couturier (BCF, Inra Marseille), Laurent Mamelli (Bioxtal), Martin Blackledge et Malene Ringkjøbing-Jensen (IBS, Grenoble), Gary Daughdrill (université de la Floride du Sud), Matteo Colombo (IBS, Grenoble), Hélène Valentin, Chantal Rabourdin-Combe (Inserm, Lyon) et David Bhella (université de Glasgow) pour leur collaboration aux travaux sur le virus de la rougeole et les Henipavirus. DB est financé par une bourse de thèse CNRS-DGA. LM a été financé par une allocation post-doctorale du CNRS et JMB par une bourse de thèse MENRT du gouvernement français.

Les travaux cités ont été rendus possibles grâce aux soutiens suivants. ANR : ANR-11-BSV8-024 01 (JFE) ; ANR-07-001-01 (ANRAGE, MJ.), ANR-05-MIIM-035 (DG, SL), ANR-08-PCVI-0020-01 (SL, DG), ANR-09-BLAN-0100-03 (SL). CNRS (Programme Interdisciplinaire CNRS « Maladies infectieuses émergentes », 2006 et 2007, MJ),

Lyonbiopôle (MJ). EU RTD programme 'Quality of Life and Management of Living Resources', QLK2-CT2001-01225 (DG, SL) et NIH (R01 NS031693-11A, SL).

Conflits d'intérêts : aucun.

Références

1. Ball LA. Virus replication strategies. In : Knipe D, Howley P, eds. *Fields virology*, vol. 1, 5th ed. Philadelphia : Lippincott Williams & Wilkins, 2007, p. 120-40.
2. Chenavas S, Guilligay D, Cusack S, Hart DJ, Ruigrok RWH, Crépin T. Avancées dans l'étude structurale et la compréhension moléculaire du mécanisme de transcription du virus de la grippe. *Virologie* 2010 ; 14 : 435-45.
3. Lyles DS, Rupprecht CE. Rhabdoviridae. In : Knipe DM, Howley PM, eds. *Fields virology*, vol. 1, 5th ed. Philadelphia : Lippincott Williams & Wilkins, 2007, p. 1364-408.
4. Anonymous. WHO Fact Sheet N° 286 – Measles. 2011.
5. Wong S, Lau S, Woo P, Yuen KY. Bats as a continuing source of emerging infections in humans. *Rev Med Virol* 2007 ; 17 : 67-91.
6. Menghani S, Chikhale R, Raval A, Wadibhasme P, Khedekar P. Chandipura Virus: an emerging tropical pathogen. *Acta Trop* 2012 ; 124 : 1-14.
7. Anonymous. WHO Fact Sheet N° 99 – Rabies. 2010.
8. Hanlon CA, Kuzmin IV, Blanton JD, Weldon WC, Manangan JS, Rupprecht CE. Efficacy of rabies biologics against new lyssaviruses from Eurasia. *Virus Res* 2005 ; 111 : 44-54.
9. Drexler JF, Corman VM, Muller MA, et al. Bats host major mammalian paramyxoviruses. *Nat Commun* 2012 ; 3 : 796.
10. Kolakofsky D, Le Mercier P, Iseni F, Garcin D. Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology* 2004 ; 318 : 463-73.
11. Lamb RA, Parks GD. *Paramyxoviridae: the viruses and their replication*. In : Knipe DM, Howley PM, eds. *Fields virology*, Vol. 1, 5th ed. Philadelphia : Lippincott Williams & Wilkins, 2007, p. 1449-96.
12. Morin B, Whelan SP. La protéine L des *Mononégavirales*. *Virologie* 2012 (soumis).
13. Solon J, Gareil O, Bassereau P, Gaudin Y. Membrane deformations induced by the matrix protein of vesicular stomatitis virus in a minimal system. *J Gen Virol* 2005 ; 86 : 3357-63.
14. Iwasaki M, Takeda M, Shirogane Y, Nakatsu Y, Nakamura T, Yanagi Y. The matrix protein of measles virus regulates viral RNA synthesis and assembly by interacting with the nucleocapsid protein. *J Virol* 2009 ; 83 : 10374-83.
15. Albertini AA, Baquero E, Ferlin A, Gaudin Y. Molecular and cellular aspects of Rhabdovirus entry. *Viruses* 2012 ; 4 : 117-39.
16. Collins PL, Crowe JEJ. Respiratory syncytial virus and *Metapneumovirus*. In : Knipe D, Howley P, eds. *Fields virology*, vol. 2, 5th ed. Philadelphia : Lippincott Williams & Wilkins, 2007, p. 1601-46.
17. Eaton BT, Mackenzie JS, Wang LF. *Henipaviruses*. In : Knipe DM, Howley PM, eds. *Fields virology*, vol. 2, 5th ed. Philadelphia : Lippincott Williams & Wilkins, 2007, p. 1587-600.
18. Ge P, Tsao J, Schein S, Green TJ, Luo M, Zhou ZH. Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* 2010 ; 327 : 689-93.
19. Barge A, Gaudin Y, Coulon P, Ruigrok RW. Vesicular stomatitis virus M protein may be inside the ribonucleocapsid coil. *J Virol* 1993 ; 67 : 7246-53.
20. Liljeroos L, Huiskonen JT, Ora A, Susi P, Butcher SJ. Electron cryotomography of measles virus reveals how matrix protein coats the ribonucleocapsid within intact virions. *Proc Natl Acad Sci U S A* 2011 ; 108 : 18085-90.

21. Carlos TS, Young DF, Schneider M, Simas JP, Randall RE. Parainfluenza virus 5 genomes are located in viral cytoplasmic bodies whilst the virus dismantles the interferon-induced antiviral state of cells. *J Gen Virol* 2009; 90: 2147-56.
22. Heinrich BS, Cureton DK, Rahmeh AA, Whelan SP. Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. *PLoS Pathog* 2010; 6: e1000958.
23. Lahaye X, Vidy A, Pomier C, et al. Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: evidence that NBs are sites of viral transcription and replication. *J Virol* 2009; 83: 7948-58.
24. Goodin MM, Chakrabarty R, Yelton S, Martin K, Clark A, Brooks R. Membrane and protein dynamics in live plant nuclei infected with *Sonchus yellow net virus*, a plant-adapted rhabdovirus. *J Gen Virol* 2007; 88: 1810-20.
25. Lipkin WI, Briese T. *Bornaviridae*. In : Knipe DM, Howley PM, eds. *Fields virology*, vol. 2. Philadelphia : Lippincott Williams & Wilkins, 2007, p. 1829-51.
26. Szilagyi JF, Uryvayev L. Isolation of an infectious ribonucleoprotein from vesicular stomatitis virus containing an active RNA transcriptase. *J Virol* 1973; 11: 279-86.
27. Morin B, Rahmeh AA, Whelan SP. Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. *EMBO J* 2012; 31: 1320-9.
28. Ogino T, Banerjee AK. An unconventional pathway of mRNA cap formation by vesiculoviruses. *Virus Res* 2011; 162: 100-9.
29. Barr JN, Wertz GW. Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination. *J Virol* 2001; 75: 6901-13.
30. Plattet P, Strahle L, le Mercier P, Hausmann S, Garcin D, Kolakofsky D. Sendai virus RNA polymerase scanning for mRNA start sites at gene junctions. *Virology* 2007; 362: 411-20.
31. Plumet S, Duprex WP, Gerlier D. Dynamics of viral RNA synthesis during measles virus infection. *J Virol* 2005; 79: 6900-8.
32. Toriumi H, Eriguchi Y, Takamatsu F, Kawai A. Further studies on the hyperphosphorylated form (p40) of the rabies virus nominal phosphoprotein (P). *Microbiol Immunol* 2004; 48: 865-74.
33. Pattnaik AK, Hwang L, Li T, et al. Phosphorylation within the amino-terminal acidic domain of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication. *J Virol* 1997; 71: 8167-75.
34. Bourhis J, Johansson K, Receveur-Bréchet V, et al. The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Res* 2004; 99: 157-67.
35. Bourhis JM, Canard B, Longhi S. Désordre structural au sein du complexe réplcatif du virus de la rougeole : implications fonctionnelles. *Virologie* 2005; 9: 367-83.
36. Bourhis JM, Canard B, Longhi S. Structural disorder within the replicative complex of measles virus: functional implications. *Virology* 2006; 344: 94-110.
37. Bourhis JM, Receveur-Bréchet V, Oglesbee M, et al. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Sci* 2005; 14: 1975-92.
38. Longhi S. Nucleocapsid structure and function. *Curr Top Microbiol Immunol* 2009; 329: 103-28.
39. Habchi J, Mamelli L, Darbon H, Longhi S. Structural disorder within henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLoS ONE* 2010; 5: e11684.
40. Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in *Paramyxovirinae* N and P. *J Gen Virol* 2003; 84: 3239-52.
41. Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim Biophys Acta* 2010; 1804: 1231-64.
42. Longhi S, Oglesbee M. Structural disorder within the measles virus nucleoprotein and phosphoprotein. *Protein Pept Lett* 2010; 17: 961-78.
43. Leyrat C, Ribeiro EA, Gérard FC, Ivanov I, Ruigrok RWH, Jamin M. Structure, interactions with host-cell and functions of Rhabdovirus phosphoprotein. *Future Virol* 2011; 6: 465-81.
44. Rudolph MG, Kraus I, Dickmanns A, Eickmann M, Garten W, Ficner R. Crystal structure of the borna disease virus nucleoprotein. *Structure (Camb)* 2003; 11: 1219-26.
45. Albertini AA, Wernimont AK, Muziol T, et al. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 2006; 313: 360-3.
46. Green TJ, Zhang X, Wertz GW, Luo M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 2006; 313: 357-60.
47. Tawar RG, Duquerry S, Vourheine C, et al. Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* 2009; 326: 1279-83.
48. Green TJ, Luo M. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor P. *Proc Natl Acad Sci U S A* 2009; 106: 11713-8.
49. Leyrat C, Yabukarski F, Tarbouriech N, et al. Structure of the vesicular stomatitis virus N(0)-P complex. *PLoS Pathog* 2011; 7: e1002248.
50. Lynch S, Kolakofsky D. Ends of the RNA within Sendai virus defective interfering nucleocapsids are not free. *J Virol* 1978; 28: 584-9.
51. Chen S, Short JA, Young DF, et al. Heterocellular induction of interferon by negative-sense RNA viruses. *Virology* 2010; 407: 247-55.
52. Bifko V, Musiyenko A, Shulyayeva O, Barik S. Inhibition of respiratory viruses by nasally administered siRNA. *Nat Med* 2005; 11: 50-5.
53. Iseni F, Baudin F, Blondel D, Ruigrok RW. Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *RNA* 2000; 6: 270-81.
54. Albertini AAV, Schoehn G, Ruigrok RW. Structures impliquées dans la réplication et la transcription des virus à ARN non segmentés de sens négatif. *Virologie* 2005; 9: 83-92.
55. Egelman EH, Wu SS, Amrein M, Portner A, Murti G. The Sendai virus nucleocapsid exists in at least four different helical states. *J Virol* 1989; 63: 2233-43.
56. Schoehn G, Mavrakis M, Albertini A, Wade R, Hoenger A, Ruigrok RW. The 12 Å structure of trypsin-treated measles virus N-RNA. *J Mol Biol* 2004; 339: 301-12.
57. Bhella D, Ralph A, Murphy LB, Yeo RP. Significant differences in nucleocapsid morphology within the *Paramyxoviridae*. *J Gen Virol* 2002; 83: 1831-9.
58. Schoehn G, Iseni F, Mavrakis M, Blondel D, Ruigrok RW. Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J Virol* 2001; 75: 490-8.
59. Bhella D, Ralph A, Yeo RP. Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction. *J Mol Biol* 2004; 340: 319-31.
60. Iseni F, Barge A, Baudin F, Blondel D, Ruigrok RW. Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. *J Gen Virol* 1998; 79 (Pt 12): 2909-19.
61. Thomas D, Newcomb WW, Brown JC, et al. Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. *J Virol* 1985; 54: 598-607.
62. Roux L. Dans le génome des *Paramyxovirinae*, les promoteurs et leurs activités sont façonnés par la « règle de six ». *Virologie* 2005; 9: 19-34.
63. Green TJ, Macpherson S, Qiu S, Lebowitz J, Wertz GW, Luo M. Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J Virol* 2000; 74: 9515-24.
64. Albertini AA, Schoehn G, Weissenhorn W, Ruigrok RW. Structural aspects of rabies virus replication. *Cell Mol Life Sci* 2008; 65: 282-94.

65. Spehner D, Kim A, Drillien R. Assembly of nucleocapsid-like structures in animal cells infected with a vaccinia virus recombinant encoding the measles virus nucleoprotein. *J Virol* 1991; 65: 6296-300.
66. Warnes A, Fooks AR, Dowsett AB, Wilkinson GW, Stephenson JR. Expression of the measles virus nucleoprotein gene in *Escherichia coli* and assembly of nucleocapsid-like structures. *Gene* 1995; 160: 173-8.
67. Tan WS, Ong ST, Eshaghi M, Foo SS, Yusoff K. Solubility, immunogenicity and physical properties of the nucleocapsid protein of Nipah virus produced in *Escherichia coli*. *J Med Virol* 2004; 73: 105-12.
68. Kerdiles YM, Cherif B, Marie JC, et al. Immunomodulatory properties of *Morbillivirus* nucleoproteins. *Viral Immunol* 2006; 19: 324-34.
69. Albertini AA, Clapier CR, Wermimont AK, Schoehn G, Weissenhorn W, Ruijgrok RW. Isolation and crystallization of a unique size category of recombinant rabies virus nucleoprotein-RNA rings. *J Struct Biol* 2007; 158: 129-33.
70. Blumberg BM, Giorgi C, Kolakofsky D. N protein of vesicular stomatitis virus selectively encapsidates leader RNA *in vitro*. *Cell* 1983; 32: 559-67.
71. Ivanov I, Yabukarski F, Ruijgrok RW, Jamin M. Structural insights into the rhabdovirus transcription/replication complex. *Virus Res* 2011; 162: 126-37.
72. Ruijgrok RW, Crepin T, Kolakofsky D. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* 2011; 14: 504-10.
73. Luo M, Green TJ, Zhang X, Tsao J, Qiu S. Structural comparisons of the nucleoprotein from three negative strand RNA virus families. *Virol J* 2007; 4: 72.
74. Desfosses A, Goret G, Farias Estrozi L, Ruijgrok RW, Gutsche I. Nucleoprotein-RNA orientation in the measles virus nucleocapsid by three-dimensional electron microscopy. *J Virol* 2011; 85: 1391-5.
75. Green TJ, Rowse M, Tsao J, et al. Access to RNA encapsidated in the nucleocapsid of vesicular stomatitis virus. *J Virol* 2011; 85: 2714-22.
76. Zhang X, Green TJ, Tsao J, Qiu S, Luo M. Role of intermolecular interactions of vesicular stomatitis virus nucleoprotein in RNA encapsidation. *J Virol* 2008; 82: 674-82.
77. Harouaka D, Wertz GW. Mutations in the C-terminal loop of the nucleocapsid protein affect vesicular stomatitis virus RNA replication and transcription differentially. *J Virol* 2009; 83: 11429-39.
78. Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc Natl Acad Sci U S A* 2004; 101: 8301-6.
79. Curran J, Homann H, Buchholz C, Rochat S, Neubert W, Kolakofsky D. The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. *J Virol* 1993; 67: 4358-64.
80. Buchholz CJ, Spehner D, Drillien R, Neubert WJ, Homann HE. The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly. *J Virol* 1993; 67: 5803-12.
81. Bankamp B, Horikami SM, Thompson PD, Huber M, Billeter M, Moyer SA. Domains of the measles virus N protein required for binding to P protein and self-assembly. *Virology* 1996; 216: 272-7.
82. Liston P, Batal R, DiFlumeri C, Briedis DJ. Protein interaction domains of the measles virus nucleocapsid protein (NP). *Arch Virol* 1997; 142: 305-21.
83. Myers TM, Pieters A, Moyer SA. A highly conserved region of the Sendai virus nucleocapsid protein contributes to the NP-NP binding domain. *Virology* 1997; 229: 322-35.
84. Myers TM, Smallwood S, Moyer SA. Identification of nucleocapsid protein residues required for Sendai virus nucleocapsid formation and genome replication. *J Gen Virol* 1999; 80: 1383-91.
85. Karlin D, Longhi S, Canard B. Substitution of two residues in the measles virus nucleoprotein results in an impaired self-association. *Virology* 2002; 302: 420-32.
86. Habchi J, Blangy S, Mamelli L, et al. Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipavirus. *J Biol Chem* 2011; 286: 13583-602.
87. Longhi S, Receveur-Brechot V, Karlin D, et al. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J Biol Chem* 2003; 278: 18638-48.
88. Lieutaud P, Canard B, Longhi S. MeDor: a metaserver for predicting protein disorder. *BMC Genomics* 2008; 9: S25.
89. Houben K, Marion D, Tarbouriech N, Ruijgrok RW, Blanchard L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the *Paramyxovirus* family. *J Virol* 2007; 81: 6807-16.
90. Heggenes MH, Scheid A, Choppin PW. Conformation of the helical nucleocapsids of *Paramyxoviruses* and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc Natl Acad Sci U S A* 1980; 77: 2631-5.
91. Heggenes MH, Scheid A, Choppin PW. The relationship of conformational changes in the Sendai virus nucleocapsid to proteolytic cleavage of the NP polypeptide. *Virology* 1981; 114: 555-62.
92. Ringkjøbing Jensen M, Communie G, Ribeiro Jr. ED, et al. Intrinsic disorder in measles virus nucleocapsids. *Proc Natl Acad Sci U S A* 2011; 108: 9839-44.
93. Garner E, Romero P, Dunker AK, Brown C, Obradovic Z. Predicting binding regions within disordered proteins. *Genome Inform Ser Workshop Genome Inform* 1999; 10: 41-50.
94. Diallo A, Barrett T, Barbron M, Meyer G, Lefevre PC. Cloning of the nucleocapsid protein gene of peste-des-petits-ruminants virus: relationship to other *Morbilliviruses*. *J Gen Virol* 1994; 75 (Pt 1): 233-7.
95. Gely S, Lowry DF, Bernard C, et al. Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. *J Mol Recognit* 2010; 23: 435-47.
96. Belle V, Rouger S, Costanzo S, et al. Mapping alpha-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. *Protein Struct Funct Bioinformatics* 2008; 73: 973-88.
97. Morin B, Bourhis JM, Belle V, et al. Assessing induced folding of an intrinsically disordered protein by site-directed spin-labeling EPR spectroscopy. *J Phys Chem B* 2006; 110: 20596-608.
98. Johansson K, Bourhis JM, Campanacci V, Cambillau C, Canard B, Longhi S. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem* 2003; 278: 44567-73.
99. Blocquel D, Habchi J, Gruet A, Blangy S, Longhi S. Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies. *Mol Biosyst* 2012; 8: 392-410.
100. Martinho M, Habchi J, El Habre Z, et al. Assessing induced folding within the intrinsically disordered C-terminal domain of the Henipavirus nucleoproteins by site directed spin labeling EPR spectroscopy. *J Biomol Struct Dyn* 2012 (in press).
101. Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* 2004; 338: 1015-26.
102. Qanungo KR, Shaji D, Mathur M, Banerjee AK. Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. *Proc Natl Acad Sci U S A* 2004; 101: 5952-7.
103. Das T, Patmaik AK, Takacs AM, Li T, Hwang LN, Banerjee AK. Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. *Virology* 1997; 238: 103-14.

104. Curran J, Pelet T, Kolakofsky D. An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation. *Virology* 1994; 202: 875-84.
105. Bowman MC, Smallwood S, Moyer SA. Dissection of individual functions of the Sendai virus phosphoprotein in transcription. *J Virol* 1999; 73: 6474-83.
106. Gerard FC, Ribeiro Ede Jr. A, Leyrat C, et al. Modular organization of rabies virus phosphoprotein. *J Mol Biol* 2009; 388: 978-96.
107. Karlin D, Longhi S, Receveur V, Canard B. The N-terminal domain of the phosphoprotein of *Morbilliviruses* belongs to the natively unfolded class of proteins. *Virology* 2002; 296: 251-62.
108. Leyrat C, Gerard FC, de Almeida Ribeiro Jr. E, Ivanov I, Ruigrok RW, Jamin M. Structural disorder in proteins of the rhabdoviridae replication complex. *Protein Pept Lett* 2010; 17: 979-87.
109. Emerson SU, Yu Y. Both NS and L proteins are required for *in vitro* RNA synthesis by vesicular stomatitis virus. *J Virol* 1975; 15: 1348-56.
110. Arnharter H, Davis NL, Wertz G, Schubert M, Lazzarini RA. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell* 1985; 41: 259-67.
111. Peluso RW, Moyer SA. Viral proteins required for the *in vitro* replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology* 1988; 162: 369-76.
112. Leyrat C, Jensen MR, Ribeiro Jr. EA, et al. The N(0)-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient alpha-helices. *Protein Sci* 2011; 20: 542-56.
113. Canter DM, Perrault J. Stabilization of vesicular stomatitis virus L polymerase protein by P protein binding: a small deletion in the C-terminal domain of L abrogates binding. *Virology* 1996; 219: 376-86.
114. Cevik B, Holmes DE, Vrotsos E, Feller JA, Smallwood S, Moyer SA. The phosphoprotein (P) and L binding sites reside in the N-terminus of the L subunit of the measles virus RNA polymerase. *Virology* 2004; 327: 297-306.
115. Llorente MT, Garcia-Barreno B, Calero M, et al. Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. *J Gen Virol* 2006; 87: 159-69.
116. Ribeiro Jr. EA, Favier A, Gerard FC, et al. Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J Mol Biol* 2008; 382: 525-38.
117. Ivanov I, Crepin T, Jamin M, Ruigrok R. Structure of the dimerisation domain of the rabies virus phosphoprotein. *J Virol* 2010; 84: 3707-10.
118. Castel G, Chteoui M, Caignard G, et al. Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. *J Virol* 2009; 83: 10808-20.
119. Emerson SU, Schubert M. Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc Natl Acad Sci U S A* 1987; 84: 5655-9.
120. Boze H, Marlin T, Durand D, et al. Proline-rich salivary proteins have extended conformations. *Biophys J* 2010; 99: 656-65.
121. Paz A, Zeev-Ben-Mordehai T, Lundqvist M, et al. Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3. *Biophys J* 2008; 95: 1928-44.
122. Watanabe N, Kawano M, Tsurudome M, et al. Identification of the sequences responsible for nuclear targeting of the V protein of human parainfluenza virus type 2. *J Gen Virol* 1996; 77: 327-38.
123. Karlin D, Belshaw R. Detecting remote sequence homology in disordered proteins: discovery of conserved motifs in the N-termini of *Mononegavirales* phosphoproteins. *PLoS ONE* 2012; 7: e31719.
124. Gérard FCA, Ribeiro E, Albertini A, et al. Unphosphorylated *Rhabdoviridae* phosphoproteins form elongated dimers in solution. *Biochemistry* 2007; 46: 10328-38.
125. Tarbouriech N, Curran J, Ebel C, Ruigrok RW, Burnmeister WP. On the domain structure and the polymerization state of the sendai virus P protein. *Virology* 2000; 266: 99-109.
126. Tarbouriech N, Curran J, Ruigrok RW, Burnmeister WP. Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat Struct Biol* 2000; 7: 777-81.
127. Castagne N, Barbier A, Bernard J, et al. Biochemical characterization of the respiratory syncytial virus P-P and P-N protein complexes and localization of the P protein oligomerization domain. *J Gen Virol* 2004; 85: 1643-53.
128. Ding H, Green TJ, Lu S, Luo M. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J Virol* 2006; 80: 2808-14.
129. Blanchard L, Tarbouriech N, Blackledge M, et al. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology* 2004; 319: 201-11.
130. Habchi J, Mamelli L, Longhi S. Structural disorder within the nucleoprotein and phosphoprotein from measles, Nipah and Hendra viruses. In: Uversky VN, Longhi S, eds. *Flexible viruses: structural disorder in viral proteins*. Hoboken, New Jersey: John Wiley, Sons, 2012.
131. Kingston RL, Gay LS, Baase WS, Matthews BW. Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. *J Mol Biol* 2008; 379: 719-31.
132. Kingston RL, Walter AB, Gay LS. Characterization of nucleocapsid binding by the measles and the mumps virus phosphoprotein. *J Virol* 2004; 78: 8630-40.
133. Mavrakis M, McCarthy AA, Roche S, Blondel D, Ruigrok RW. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J Mol Biol* 2004; 343: 819-31.
134. Assenberg R, Delmas O, Ren J, et al. Structure of the nucleoprotein binding domain of *Mokola* virus phosphoprotein. *J Virol* 2010; 84: 1089-96.
135. Delmas O, Assenberg R, Grimes JM, Bourhy H. The structure of the nucleoprotein binding domain of *Lyssavirus* phosphoprotein reveals a structural relationship between the N-RNA binding domains of *Rhabdoviridae* and *Paramyxoviridae*. *RNA Biol* 2010; 7: 322-7.
136. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci U S A* 2005; 102: 17002-7.
137. Houben K, Blanchard L, Blackledge M, Marion D. Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P. *Biophys J* 2007; 93: 2830-44.
138. Leyrat C, Schneider R, Ribeiro Jr. EA, et al. Ensemble structure of the modular and flexible full-length vesicular stomatitis virus phosphoprotein. *J Mol Biol* 2012.
139. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 2007; 129: 5656-64.
140. Llorente MT, Taylor IA, Lopez-Vinas E, et al. Structural properties of the human respiratory syncytial virus P protein: evidence for an elongated homotetrameric molecule that is the smallest orthologue within the family of paramyxovirus polymerase cofactors. *Proteins* 2008; 72: 946-58.
141. Asenjo A, Gonzalez-Armas JC, Villanueva N. Phosphorylation of human respiratory syncytial virus P protein at serine 54 regulates viral uncoating. *Virology* 2008; 380: 26-33.
142. Asenjo A, Mendieta J, Gomez-Puertas P, Villanueva N. Residues in human respiratory syncytial virus P protein that are essential for its activity on RNA viral synthesis. *Virus Res* 2008; 132: 160-73.
143. Tran TL, Castagne N, Bhella D, et al. The nine C-terminal amino acids of the respiratory syncytial virus protein P are necessary and sufficient for binding to ribonucleoprotein complexes in which six ribonucleotides are contacted per N protein protomer. *J Gen Virol* 2007; 88: 196-206.

144. Galloux M, Tarus B, Blazevic I, Fix J, Duquerroy S, Éléouët JF. Characterization of a viral phosphoprotein binding site on the surface of the respiratory syncytial nucleoprotein. *J Virol* 2012; 86 : 8375-87.
145. Esperante SA, Chemes LB, Sanchez IE, de Prat-Gay G. The respiratory syncytial virus transcription antiterminator M(2-1) is a highly stable, zinc binding tetramer with strong pH-dependent dissociation and a monomeric unfolding intermediate. *Biochemistry* 2011; 50 : 8529-39.
146. Blondot ML, Dubosclard V, Fix J, et al. Structure and functional analysis of the ma- and viral phosphoprotein-binding domain of respiratory syncytial virus M2-1 protein. *PLoS Pathog* 2012; 8 : e1002734.
147. Hartlieb B, Muziol T, Weissenhorn W, Becker S. Crystal structure of the C-terminal domain of Ebola virus VP30 reveals a role in transcription and nucleocapsid association. *Proc Natl Acad Sci U S A* 2007; 104 : 624-9.
148. King AMQ, Lefkowitz E, Adams MJ, Carstens EB. *Virus taxonomy, 10th report of the International Committee on Taxonomy of Viruses*. London : Elsevier/Academic Press, 2011.
149. Ribeiro Ede Jr. A, Leyrat C, Gerard FC, et al. Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes. *J Mol Biol* 2009; 394 : 558-75.
150. Jacob Y, Real E, Tordo N. Functional interaction map of *Lyssavirus* phosphoprotein: identification of the minimal transcription domains. *J Virol* 2001; 75 : 9613-22.
151. Heinrich BS, Morin B, Rahmeh AA, Whelan SP. Structural properties of the C-terminus of vesicular stomatitis virus n protein dictate N-RNA complex assembly: Encapsidation and RNA synthesis. *J Virol* 2012; 86 : 8720-9.
152. Shu Y, Habchi J, Costanzo S, et al. Plasticity in structural and functional interactions between the phosphoprotein and nucleoprotein of measles virus. *J Biol Chem* 2012; 287 : 11951-67.
153. Mittag T, Orlicky S, Choy WY, et al. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci U S A* 2008; 105 : 17772-7.
154. Kavalenka A, Urbancic I, Belle V, et al. Conformational analysis of the partially disordered measles virus N-TAIL-XD complex by SDSL EPR spectroscopy. *Biophys J* 2010; 98 : 1055-64.
155. Jensen MR, Houben K, Lescop E, Blanchard L, Ruijgrok RW, Blackledge M. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J Am Chem Soc* 2008; 130 : 8055-61.
156. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* 2000; 97 : 8868-73.
157. Espinoza-Fonseca LM. Leucine-rich hydrophobic clusters promote folding of the N-terminus of the intrinsically disordered transactivation domain of p53. *FEBS Lett* 2009; 583 : 556-60.
158. Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* 2008; 33 : 2-8.
159. Curran J, Kolakofsky D. Replication of paramyxoviruses. *Adv Virus Res* 1999; 54 : 403-22.
160. Curran J. A role for the Sendai virus P protein trimer in RNA synthesis. *J Virol* 1998; 72 : 4274-80.
161. Gubbay O, Curran J, Kolakofsky D. Sendai virus genome synthesis and assembly are coupled: a possible mechanism to promote viral RNA polymerase processivity. *J Gen Virol* 2001; 82 : 2895-903.
162. Zhang X, Glendening C, Linke H, et al. Identification and characterization of a regulatory domain on the carboxyl terminus of the measles virus nucleocapsid protein. *J Virol* 2002; 76 : 8737-46.
163. Zhang X, Bourhis JM, Longhi S, et al. Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus. *Virology* 2005; 337 : 162-74.
164. Oglesbee M. Nucleocapsid protein interactions with the major inducible 70kDa heat shock protein. In : Longhi S, ed. *Measles virus nucleoprotein*. Hauppauge, NY : Nova Publishers Inc, 2007, p. 53-98.
165. Masters PS, Banerjee AK. Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J Virol* 1988; 62 : 2658-64.
166. Blumberg BM, Leppert M, Kolakofsky D. Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication. *Cell* 1981; 23 : 837-45.
167. Dillon PJ, Gupta KC. Early steps in the assembly of vesicular stomatitis virus nucleocapsids in infected cells. *J Virol* 1988; 62 : 1582-9.
168. Curran J, Kolakofsky D. Nonsegmented negative-strand RNA virus RNA synthesis *in vivo*. *Virology* 2008; 371 : 227-30.
169. Mavrikis M, Mehous S, Real E, et al. Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology* 2006; 349 : 422-9.
170. Chen M, Ogino T, Banerjee AK. Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. *J Virol* 2007; 81 : 13478-85.
171. Luo M, Green TJ, Zhang X, Tsao J, Qiu S. Conserved characteristics of the Rhabdovirus nucleoprotein. *Virus Res* 2007; 129 : 246-51.
172. Nayak D, Panda D, Das SC, Luo M, Pattnaik AK. Single-amino-acid alterations in a highly conserved central region of vesicular stomatitis virus N protein differentially affect the viral nucleocapsid template functions. *J Virol* 2009; 83 : 5525-34.
173. Gupta AK, Banerjee AK. Expression and purification of vesicular stomatitis virus N-P complex from *Escherichia coli*: role in genome RNA transcription and replication *in vitro*. *J Virol* 1997; 71 : 4264-71.
174. Colombo M, Bourhis JM, Chamontin C, et al. The interaction between the measles virus nucleoprotein and the Interferon Regulator Factor 3 relies on a specific cellular environment. *Virol J* 2009; 6 : 59.
175. Laine D, Trescol-Biémont M, Longhi S, et al. Measles virus nucleoprotein binds to a novel cell surface receptor distinct from FcγRII via its C-terminal domain: role in MV-induced immunosuppression. *J Virol* 2003; 77 : 11332-46.
176. Sato H, Masuda M, Miura R, Yoneda M, Kai C. Morbillivirus nucleoprotein possesses a novel nuclear localization signal and a CRM1-independent nuclear export signal. *Virology* 2006; 352 : 121-30.
177. Watanabe A, Yoneda M, Ikeda F, Sugai A, Sato H, Kai C. Peroxiredoxin 1 is required for efficient transcription and replication of measles virus. *J Virol* 2011; 85 : 2247-53.
178. Couturier M, Buccellato M, Costanzo S, et al. High Affinity Binding between Hsp70 and the C-terminal domain of the measles virus nucleoprotein requires an Hsp40 Co-Chaperone. *J Mol Recognit* 2010; 23 : 301-15.
179. Carsillo T, Traylor Z, Choi C, Niewiesk S, Oglesbee M. Hsp72, a host determinant of measles virus neurovirulence. *J Virol* 2006; 80 : 11031-9.
180. Gombart AF, Hirano A, Wong TC. Conformational maturation of measles virus nucleocapsid protein. *J Virol* 1993; 67 : 4133-41.
181. Rahmen AA, Morin B, Schenk AD, et al. Critical phosphoprotein elements that regulate polymerase architecture and function in vesicular stomatitis virus. *Proc Natl Acad Sci U S A* 2012; 109 : 14628-33.

Article: Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus

Authors: Guillaume Communie*, Johnny Habchi*, **Filip Yabukarski**, David Blocquel, Robert Schneider, Nicolas Tarbouriech, Nicolas Papageorgiou, Rob W.H. Ruigrok, Marc Jamin, Malene Ringkjøbing Jensen, Sonia Longhi and Martin Blackledge

Publié en 2013 dans *Plos Pathogens*

Atomic Resolution Description of the Interaction between the Nucleoprotein and Phosphoprotein of Hendra Virus

Guillaume Communie^{1,2,3,4,5,6*}, Johnny Habchi^{7*}, Filip Yabukarski^{4,5,6}, David Blocquel⁷, Robert Schneider^{1,2,3}, Nicolas Tarbouriech^{4,5,6}, Nicolas Papageorgiou⁷, Rob W. H. Ruigrok^{4,5,6}, Marc Jamin^{4,5,6}, Malene Ringkjøbing Jensen^{1,2,3*}, Sonia Longhi^{7*}, Martin Blackledge^{1,2,3}

1 Université Grenoble Alpes, Institut de Biologie Structurale (IBS), Grenoble, France, **2** CEA, DSV, IBS, Grenoble, France, **3** CNRS, IBS, Grenoble, France, **4** Université Grenoble Alpes, UVHCI, Grenoble, France, **5** CNRS, UVHCI, Grenoble, France, **6** Unit for Virus Host Cell Interactions, Université Grenoble Alpes-EMBL-CNRS, Grenoble, France, **7** CNRS and Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques, UMR 7257, Marseille, France

Abstract

Hendra virus (HeV) is a recently emerged severe human pathogen that belongs to the *Henipavirus* genus within the *Paramyxoviridae* family. The HeV genome is encapsidated by the nucleoprotein (N) within a helical nucleocapsid. Recruitment of the viral polymerase onto the nucleocapsid template relies on the interaction between the C-terminal domain, N_{TAIL}, of N and the C-terminal X domain, XD, of the polymerase co-factor phosphoprotein (P). Here, we provide an atomic resolution description of the intrinsically disordered N_{TAIL} domain in its isolated state and in intact nucleocapsids using nuclear magnetic resonance (NMR) spectroscopy. Using electron microscopy, we show that HeV nucleocapsids form herringbone-like structures typical of paramyxoviruses. We also report the crystal structure of XD of P that consists of a three-helix bundle. We study the interaction between N_{TAIL} and XD using NMR titration experiments and provide a detailed mapping of the reciprocal binding sites. We show that the interaction is accompanied by α -helical folding of the molecular recognition element of N_{TAIL} upon binding to a hydrophobic patch on the surface of XD. Finally, using solution NMR, we investigate the interaction between intact nucleocapsids and XD. Our results indicate that monomeric XD binds to N_{TAIL} without triggering an additional unwinding of the nucleocapsid template. The present results provide a structural description at the atomic level of the protein-protein interactions required for transcription and replication of HeV, and the first direct observation of the interaction between the X domain of P and intact nucleocapsids in *Paramyxoviridae*.

Citation: Communie G, Habchi J, Yabukarski F, Blocquel D, Schneider R, et al. (2013) Atomic Resolution Description of the Interaction between the Nucleoprotein and Phosphoprotein of Hendra Virus. *PLoS Pathog* 9(9): e1003631. doi:10.1371/journal.ppat.1003631

Editor: Félix A. Rey, Institut Pasteur, France

Received: April 18, 2013; **Accepted:** August 1, 2013; **Published:** September 26, 2013

Copyright: © 2013 Communie et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was carried out with financial support from the French Agence Nationale de la Recherche (ANR) through the specific programs "Physico-Chimie du Vivant" ANR-08-PCVI-0020-01, "Blanc" ANR-09-BLAN-0100 and "ASTRID" ANR-11-ASTR-003-01 (to SL), ANR JCJC ProteinDisorder (to MRJ), ANR MALZ TAUSTRUCT (to MB), the Human Frontier Science Program (long-term fellowship to RS), the Rhône-Alpes region (Ph.D. fellowship to FY), the CNRS and the Direction Générale de l'Armement (Ph.D. fellowship to DB) and the Finovi foundation (to MB and RWHR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: malene.ringkjobering-jensen@ibs.fr (MRJ); Sonia.Longhi@afmb.univ-mrs.fr (SL)

☞ These authors contributed equally to this work.

Introduction

The Hendra (HeV) and Nipah (NiV) viruses are recently emerged, severe human pathogens within the *Paramyxoviridae* family [1]. A few distinctive properties set them aside from other paramyxoviruses and led to their classification within the *Henipavirus* genus of the *Paramyxoviridae* family that also contains the newly identified Cedar virus [2,3]. The genome of Henipaviruses is encapsidated by the nucleoprotein (N) within a helical nucleocapsid. This helical N:RNA complex, rather than the naked RNA, serves as substrate for both transcription and replication [4]. By analogy with other paramyxoviruses, the viral polymerase complex is thought to consist of the L protein and the phosphoprotein P. The P protein is an essential polymerase cofactor as it tethers L onto the nucleocapsid template [5,6]. *Henipavirus* N and P proteins were shown to interact with each other [7–9], being able to form both homologous and heterologous

N-P complexes [9–11]. The functional significance of the N-P interaction for genome transcription and replication makes it a promising target for antiviral drug design [12].

So far, high-resolution structural data are limited to *Henipavirus* surface proteins, where crystallographic studies led to the determination of the three-dimensional structure of the fusion (F) and the attachment (G) proteins [13,14]. A more complete picture of the molecular mechanisms governing transcription and replication of the virus awaits an atomic resolution characterization of the three key players N, P and L. The only available molecular data on *Henipavirus* N and P proteins come from our previous studies [7–9,15] showing that the N and P proteins possess a modular organization consisting of both ordered and intrinsically disordered regions [15,16]. Intrinsically disordered proteins (IDPs) are ubiquitous, functional proteins that lack significant amounts of secondary and tertiary structure under physiological conditions and are, therefore, not amenable to

Author Summary

The polymerase of negative strand RNA viruses reads the viral RNA that is associated with the nucleoprotein N forming a helical nucleocapsid. The interaction between N and the cofactor of the polymerase, the phosphoprotein P, is essential for transcription and replication of the viral genome. The mechanism by which the polymerase dislodges the RNA from the nucleoprotein for its polymerising activity remains unknown, although it has been proposed that binding to P causes a conformational change in the nucleocapsid. Here, we use nuclear magnetic resonance (NMR) spectroscopy to develop an atomic resolution description of the intrinsically disordered C-terminal domain, N_{TAIL} , of N of Hendra virus, an emerging paramyxovirus, and X-ray crystallography to determine the structure of the X domain (XD) of P. Characterization of the interaction between XD and N_{TAIL} provides evidence for folding of N_{TAIL} upon binding to P. Crucially, we were also able to study, for the first time, the interaction between XD and recombinant paramyxoviral nucleocapsids. NMR spectra of N_{TAIL} in its isolated form and in the context of nucleocapsids demonstrate that binding of XD does not change the dynamics of N_{TAIL} and that the nucleocapsid does not undergo any major rearrangements or unwinding upon interaction with P.

characterization by X-ray crystallography [17–21]. IDPs populate a vast conformational space and ensemble descriptions have emerged as the preferred tool for representing their highly dynamic nature. In this context, nuclear magnetic resonance (NMR) spectroscopy, in combination with ensemble selection techniques, has contributed immensely to our understanding of the conformational behaviour of IDPs providing a detailed mapping of their inherent potential energy landscape [22–29].

The HeV nucleoprotein (N) consists of two domains: a folded domain (referred to as N_{CORE} , residues 1–400) that is responsible for the interaction with the viral RNA and for maintaining the nucleocapsid structure, and an intrinsically disordered domain (referred to as N_{TAIL} , residues 401–532) that is responsible for the interaction with the P protein (Figure 1) [7–10]. The atomic resolution structure of N_{CORE} remains unknown, although, the crystal structure of the decameric ring-like N:RNA complex of a related paramyxovirus, the respiratory syncytial virus (RSV), has been solved recently [30]. The structure reveals two domains of the nucleoprotein, where the RNA is accommodated in a positively charged pocket at the hinge between the two domains [31].

The HeV phosphoprotein (P) consists of a large intrinsically disordered N-terminal domain (residues 1–469), a central domain (PMD, residues 470–578) that is thought to be responsible for the oligomerization of P through a coiled-coil arrangement by analogy with the related Sendai (SeV) [32] and Measles viruses (MeV) [33], and a flexible linker of 80 amino acids connecting the

globular C-terminal X domain (XD, residues 657–707) to PMD (Figure 1) [15].

In this study, we provide an atomic resolution description of the structure and dynamics of both N_{TAIL} and XD using a combination of nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography. The complex between N_{TAIL} and XD is characterised using NMR revealing a characteristic folding of the molecular recognition element (MoRE) of N_{TAIL} upon binding to XD. In addition, we determine the conformational behaviour of N_{TAIL} in the context of intact nucleocapsids revealing that N_{TAIL} remains flexible, as in the case of MeV nucleocapsids, and exhibits differential flexibility along its primary sequence. While previous solution NMR studies investigated the residual structure and dynamics of N_{TAIL} in the context of intact MeV nucleocapsids, we here describe the first study of the interaction between a paramyxoviral nucleocapsid and the X domain of the phosphoprotein. The present results reveal that HeV XD can be accommodated on N_{TAIL} without triggering additional unwinding of the helical nucleocapsid template. Our results provide an atomic resolution description of the protein-protein interactions that play an essential role in transcription and replication of HeV.

Results

Residual structure and dynamics of HeV N_{TAIL}

In order to characterize the residual structure and dynamics of N_{TAIL} , we carried out the spectral assignment of the backbone and $C\beta$ nuclei using standard triple resonance experiments (Figure S1, Table S1 in Text S1). Carbon chemical shifts are exquisitely sensitive to the presence of secondary structure allowing us to directly probe the nature and level of residual secondary structure within N_{TAIL} . The secondary structure propensity (SSP) of N_{TAIL} was calculated on the basis of the experimental $C\alpha$ and $C\beta$ chemical shifts (Figure 2A) [34]. Positive and negative SSP scores indicate propensity to form α -helical and β -extended conformations, respectively. The SSP shows that, although N_{TAIL} is devoid of significant secondary structure, the protein adopts an α -helical conformation in the region encompassing residues 470–490. The helical propensity reaches approximately 50% for the central residues and decreases progressively towards the extremities. This region of the protein has previously been shown to constitute the binding site of the P protein [7–9].

^{15}N R_2 relaxation rates and $\{^1H\}$ - ^{15}N heteronuclear Overhauser effects (nOe) report on dynamics occurring on the pico- to nanosecond time scale and were used to probe the backbone dynamics of N_{TAIL} . They display a profile characteristic of an IDP revealing, however, more elevated values in the region 470–490 due to a reduced level of chain flexibility arising from the residual helical structure present in this region (Figure 2B,C) [35]. In addition, the glycine and serine-rich regions 417–425, 455–470 and the C-terminal part display on average higher flexibility on the pico- to nanosecond time scale compared to the remainder of the chain.

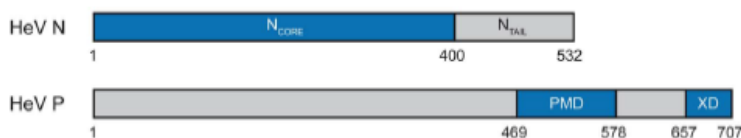


Figure 1. Domain organization of HeV N and P proteins. Gray areas indicate regions of the two proteins that are predicted to be intrinsically disordered.

doi:10.1371/journal.ppat.1003631.g001

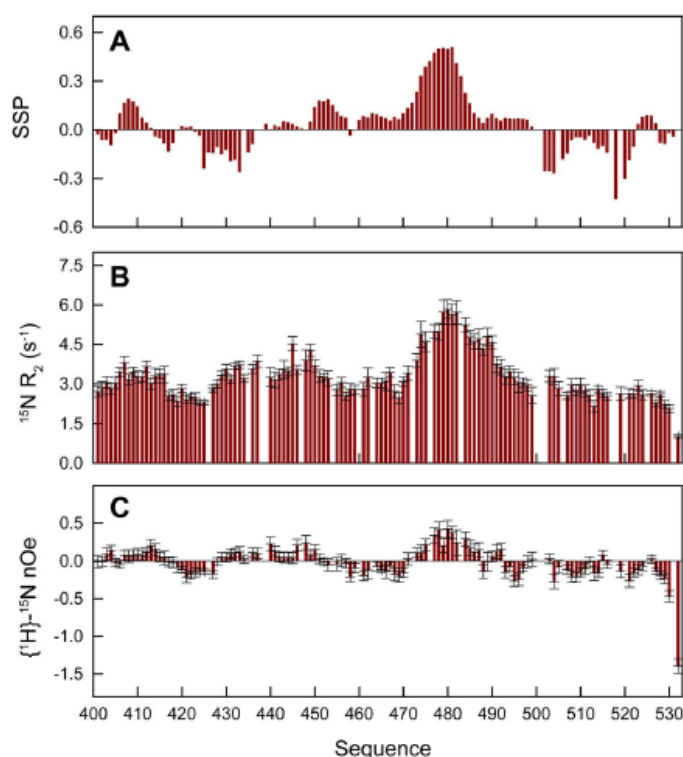


Figure 2. Characterization of the structure and dynamics of HeV N_{TAIL} . (A) Secondary structure propensity (SSP) of N_{TAIL} obtained from experimental $C\alpha$ and $C\beta$ chemical shifts. (B) ^{15}N R_2 relaxation rates of N_{TAIL} at 293 K and a 1H frequency of 600 MHz. (C) 1H - ^{15}N steady state heteronuclear Overhauser effects of N_{TAIL} at 293 K and a 1H frequency of 600 MHz. doi:10.1371/journal.ppat.1003631.g002

Three-dimensional structure of HeV XD

The three-dimensional crystal structure of HeV XD was determined from X-ray diffraction data. The protein crystallized in a $P2_1$ space group with two molecules within the asymmetric unit of the crystal. Crystallographic phases were determined by single wavelength anomalous diffraction (SAD) using synchrotron data collected from crystals of the selenomethionine-substituted protein, and the structure was refined at 1.65 Å resolution (Table S2 in Text S1). The electron density was well defined throughout the structure, except for a few side chains on the surface, and the model of chain B includes four additional residues encoded by the expression vector (LEHH). The structure is composed of three α -helices, forming an anti-parallel three-helix bundle (Figure 3). The two copies in the asymmetric unit are equivalent (RMSD = 0.65 Å) and interact through a molecular surface (900 Å²) created by helices $\alpha 1$ and $\alpha 3$.

To investigate whether the structure of XD in solution is identical to the one determined by crystallography, as well as to study the interaction of XD with N_{TAIL} , we carried out the NMR backbone assignment of XD on the basis of a set of standard triple resonance experiments (Figure S2, Table S3 in Text S1). The location of secondary structure elements obtained from the experimental $C\alpha$ and $C\beta$ chemical shifts are consistent with a three-helix bundle arrangement (Figure 4A). We also measured NMR residual dipolar couplings (RDCs) by weakly aligning the protein molecule in the magnetic field [36–38]. RDCs report on bond vector orientations with respect to a global alignment tensor that describes the preferential orientation of the molecule in the

magnetic field. Four types of RDCs measured in a concentrated suspension of filamentous phages agree well with those back-calculated from the structure determined by X-ray crystallography (Table S4 in Text S1, Figure 4B). The RDCs, therefore, confirm that the structure adopted by XD in solution is very similar to that observed in the crystal.

Atomic level structural insights into the interaction between HeV N_{TAIL} and XD

To map the reciprocal N_{TAIL} and XD interaction sites, as well as to gain insight into the molecular mechanisms controlling the binding reaction, NMR chemical shift titration experiments were performed. The addition of an increasing amount of unlabeled XD to ^{15}N labeled N_{TAIL} results in severe line broadening of the NMR resonances for residues 470–490 even at sub-stoichiometric amounts of XD (Figure 5A). This behaviour is a manifestation of contributions to the transverse relaxation rates arising from exchange occurring on the micro- to millisecond time scale between free and bound conformations of N_{TAIL} . This confirms that the residues 470–490 are the primary interaction site of N_{TAIL} with XD. The largest chemical shift perturbations in N_{TAIL} are also observed in the region of residues 470–490 (Figure 5B, Figure S3 in Text S1). The affinity of the interaction between N_{TAIL} and XD could not be derived from the chemical shift titrations as the resonances of the residues of N_{TAIL} that are involved in the interaction disappear when more than 30% (molar fraction) of XD is added to N_{TAIL} . The effective dissociation constant was previously measured by isothermal titration calorimetry to be 8.7 μM [7].

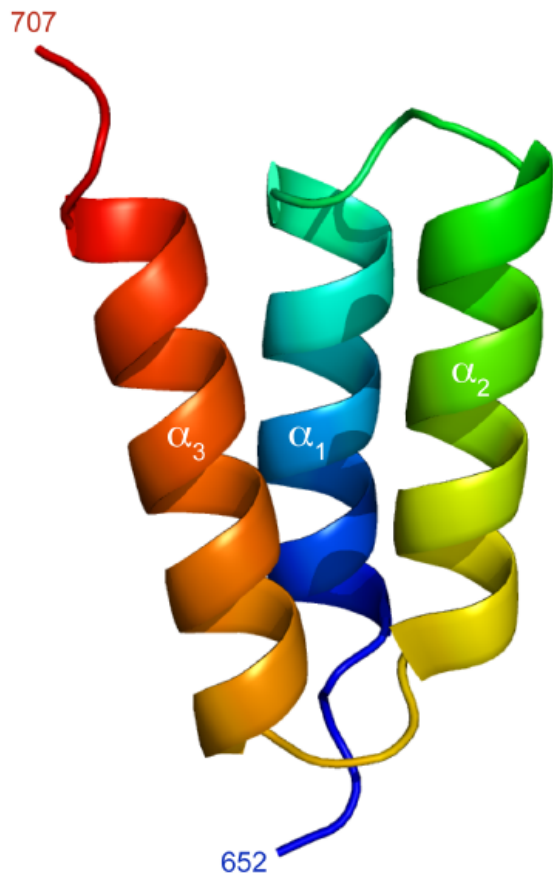


Figure 3. Crystal structure of HeV XD. Cartoon representation of the structure of XD in the crystal. The model is coloured from blue at the N-terminus to red at the C-terminus. doi:10.1371/journal.ppat.1003631.g003

To obtain further mechanistic details of the interaction between N_{TAIL} and XD, we monitored the decrease in signal intensities in the 1H - ^{15}N HSQC spectra of N_{TAIL} at each XD titration point (Figure 6B, C, D). Interestingly, the signal intensity decreases faster for the residues located at the extremities of the MoRE and for which a smaller amount of residual helical structure is observed experimentally in the isolated state of N_{TAIL} (Figure 6A). This differential broadening suggests that N_{TAIL} binds to XD via a short, central helix, which is subsequently extended via helical folding of the adjacent residues. Our data therefore indicate that N_{TAIL} interacts with XD via a folding-upon-binding mechanism [39–41]. The line broadening observed in the interaction site implies that the folding event occurs on the micro- to millisecond time scale.

In order to map the interaction site of N_{TAIL} on XD, we carried out the complementary titration, monitoring chemical shift changes in ^{15}N , ^{13}C labeled XD upon addition of unlabeled N_{TAIL} . The resonances of XD shift and gradually lose intensity with increasing N_{TAIL} concentration, and then regain intensity when approaching saturation (Figure 7A). This suggests that the exchange between the free and bound form of XD occurs on a fast to intermediate timescale (faster than the millisecond) in agreement with the complementary titration of N_{TAIL} . The largest

chemical shift and intensity changes are observed for residues of helix α_2 and α_3 identifying these two helices as the site of interaction for N_{TAIL} (Figure 7B, C). Only small chemical shift changes are observed in helix α_1 .

The crystal structure of XD provides additional information about the mechanism of interaction between N_{TAIL} and XD. The distribution of hydrophobic and charged residues on the surface of XD reveals a hydrophobic groove between helix α_2 and α_3 surrounded by numerous negatively charged residues (Figure 7D). We propose that the negatively charged residues on the surface of XD play a role in correctly orienting N_{TAIL} before adopting its final conformation in the hydrophobic pocket [42]. This is supported by the distribution of hydrophobic and charged residues within the MoRE of N_{TAIL} . Assuming a helical conformation of the bound MoRE [7,8], a helical wheel representation of this region shows four leucines (L472, L476, L479 and L483) and an alanine (A487) clustered on one side of the N_{TAIL} molecular recognition helix most likely constituting the direct interaction interface with XD (Figure 7E, F). In addition, four positively charged residues (R480, R482, K486 and K489) flank the four leucines and the alanine providing charge complementarity with respect to the negatively charged residues on the surface of XD. Note that this orientation of the MoRE at the XD surface is consistent with previous electron paramagnetic resonance (EPR) data showing that the extent to which the N_{TAIL} side chains are constrained within the N_{TAIL} -XD complex follows the order A487>S481>S475 [8].

Characterization of N_{TAIL} in the context of intact HeV nucleocapsids

In order to study the conformational behaviour of N_{TAIL} in the context of the full-length nucleoprotein, we expressed and purified ^{15}N labeled, C-terminally His-tagged N (see Materials and Methods). Expression of full-length N results in the formation of nucleocapsid-like particles containing bacterial RNA (the nucleocapsid-like particles are simply denoted nucleocapsids in the following). In order to image the structures formed by HeV N, we performed negative staining transmission electron microscopy studies. As shown in Figure 8A, the sample contains flexible nucleocapsids with the typical herringbone appearance previously reported for other paramyxoviruses [43–50]. Figure 8B shows the nucleocapsid length statistics as inferred from analysis of 170 different particles. The analysis shows that the nucleocapsids have lengths between 20 and 200 nm with more than 75% of the nucleocapsids in the range between 30 and 90 nm. Assuming a mean pitch of 5.7 nm as determined previously for MeV [51], the majority of the nucleocapsids contain 5 to 15 turns. The large size of the nucleocapsids (where 5–15 turns with 12 protomers per turn yields a mass in the order of 2.5 to 24 MDa) would normally preclude the detection of solution NMR signals due to the absence of fast rotational tumbling. However, the spectrum of the intact HeV nucleocapsids displays a large number of resonances and superimposes well on that of the isolated N_{TAIL} domain with only minor chemical shift changes (Figure 9A). This shows that N_{TAIL} retains significant conformational flexibility *in situ* and allows the resonance assignment of the isolated N_{TAIL} domain to be transferred to the nucleocapsid spectrum. Only the five C-terminal residues (528–532) of N_{TAIL} could not be easily identified in the nucleocapsid spectrum due to the different location of the His-tag in the two proteins.

The intensity profile of the nucleocapsid spectrum (normalized with respect to the spectrum of the isolated N_{TAIL} domain) reveals that N_{TAIL} has differential flexibility along its primary sequence (Figure 9B). The resonances of residues 401–454 of N_{TAIL} are

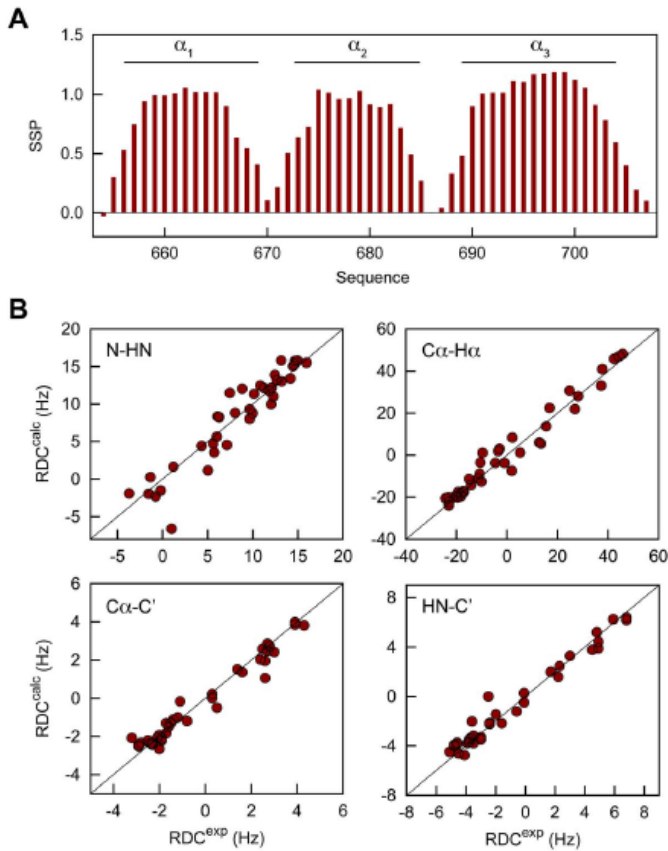


Figure 4. Structure of HeV XD in solution. (A) Secondary structure propensity (SSP) of XD obtained from experimental C α and C β chemical shifts. (B) Agreement between experimental residual dipolar couplings obtained of XD in a suspension of filamentous phages and those back-calculated on the basis of the crystal structure of XD. doi:10.1371/journal.ppat.1003631.g004

absent or have very low signal intensities indicating that N_{TAIL} is too rigid in this part of the sequence to give rise to solution NMR signals. This observation closely mimics measurements made previously on recombinant nucleocapsids of MeV [52] and indicates that the N_{TAIL} domain of HeV also requires approximately 50 residues to exfiltrate from the spindle to the exterior of the nucleocapsid helix. By assuming that each turn of the helical nucleocapsid contains 12 nucleoproteins and a pitch of 5.7 nm as determined previously for MeV [51], we estimate on the basis of our length distribution (Figure 8B) that 9% of the N_{TAIL} domains detected by NMR are located at the edge of the nucleocapsids. This percentage is very close to the experimental intensity ratio observed for residues 420–450, suggesting that the residual intensities in this region are nucleocapsid-edge effects that arise from N_{TAIL} domains that are not positioned between two helical turns of the nucleocapsid (Figure 9B). These intensities can therefore be considered as an artefact of the purification of recombinant nucleocapsids.

The NMR signal intensities of the region encompassing residues 455–490 of N_{TAIL} increase to an approximately constant level, while the C-terminal residues 491–527 gain even further in signal intensity (Figure 9B). The plateau observed in the signal intensities in the region 455–490 indicates some restriction of motion compared to residues in the C-terminal part of the protein. This

restriction of motion could be induced by a transient contact between the MoRE of N_{TAIL} and the surface of the nucleocapsids or by an increased local correlation time of the partially folded MoRE due to a “drag” effect induced by the presence of the nucleocapsid core domain [53].

Interaction of intact HeV nucleocapsids with XD

The observed NMR signals of N_{TAIL} within the nucleocapsids provide a unique probe of nucleocapsid morphology. Our next goal was thus to directly monitor the interaction between the nucleocapsids and XD in order to observe whether any changes in the local conformational behaviour of N_{TAIL}, or in the overall morphology of the nucleocapsids occur upon XD binding. The interaction between XD and full-length nucleocapsids of paramyxoviruses had previously not been amenable to structural studies, in particular, MeV nucleocapsids were found to precipitate upon addition of XD *in vitro* (data not shown). In the case of HeV, however, no such behaviour was observed, and we were able to record well-resolved spectra of nucleocapsids complexed with XD (Figure 10A). HSQC spectra of intact nucleocapsids in the absence and presence of a saturating amount of XD superimpose well and show that XD interacts only locally with the MoRE of N_{TAIL} (Figure 10A). The intensity profile of the nucleocapsids in the

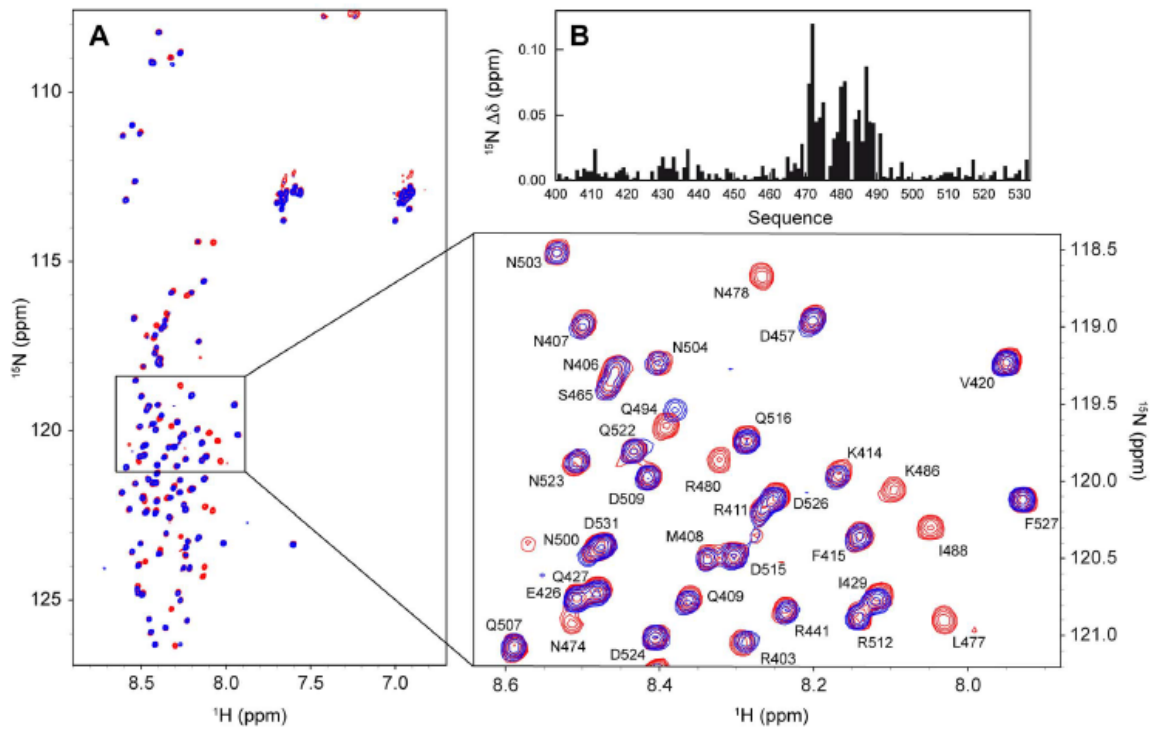


Figure 5. Interaction between ^{15}N labeled N_{TAIL} and unlabeled XD. (A) Superposition of the ^1H - ^{15}N HSQC spectra of isolated N_{TAIL} (red) and N_{TAIL} containing 4.4-fold molar excess of XD (blue). The spectra were obtained at a ^1H frequency of 800 MHz and 293 K. The expanded region shows the spectral assignment of a number of resonances in the spectrum of the isolated N_{TAIL} domain. (B) ^{15}N chemical shift changes (absolute values) between isolated N_{TAIL} and N_{TAIL} with 30% of XD (molar ratio) added. Note that at this molar fraction all the peaks of N_{TAIL} are still visible. doi:10.1371/journal.ppat.1003631.g005

presence of XD (Figure 10B) closely resembles that of the nucleocapsids alone except that the resonances of the MoRE have disappeared, as is the case in the titration of the isolated N_{TAIL} domain with XD (Figure 6D). Importantly, the intensities outside the MoRE (residues 420–468 and residues 500–527) are not modulated by the presence of XD and no additional peaks are observed (Figure 10B). The fact that we see the same intensities and chemical shifts, and therefore the same conformational signature, for these residues upon interaction with the X domain of the phosphoprotein provides evidence that the environment of N_{TAIL} is effectively identical in the free and XD-bound form of the nucleocapsids. Any major rearrangement or unwinding of the nucleocapsids upon interaction with XD would be expected to change the conformational signature of N_{TAIL} . In particular, those N_{TAIL} residues whose motion is evidently strongly restricted within intact nucleocapsids (i.e. residues 401–450) should gain in NMR signal intensity if the nucleocapsids were unwound, which is not observed. The fact that the small residual signals from residues 420–450, interpreted as nucleocapsid edge effects, are preserved indicates additionally that the length distribution of the nucleocapsids remains unchanged. Taken together, our results thus strongly suggest that the helical nucleocapsid remains intact upon interaction with XD.

Discussion

Differential flexibility and transient order in HeV N_{TAIL}

In this study, we have characterized the X domain of the polymerase-associated phosphoprotein and the N_{TAIL} domain of

the nucleoprotein whose interaction is essential for the replication and transcription process. We have provided an atomic resolution structural and dynamic characterization of HeV N_{TAIL} using NMR spectroscopy. From experimental chemical shifts and nuclear relaxation data, we have shown that N_{TAIL} adopts conformations close to a random coil, except for the interaction site with the phosphoprotein that populates up to 50% α -helical conformations in broad agreement with previous global spectroscopic analyses [9]. A similar structural architecture was also observed for N_{TAIL} of two closely related viruses, namely SeV and MeV, whose MoREs adopt a dynamic equilibrium between a completely unfolded state and several specific helical conformations [52,54–57].

To further extend the structural characterization of N_{TAIL} to a physiologically relevant environment, we studied the conformational behaviour and flexibility of N_{TAIL} within intact nucleocapsids. To this end, we expressed and purified intact nucleoprotein that assembles on cellular RNA in *E. coli*, to form helical nucleocapsids. Using solution NMR, N_{TAIL} was found to remain flexible *in situ* supporting a functional role for structural disorder in orchestrating the replicative machinery of *Paramyxoviridae*. The flexibility of N_{TAIL} in the context of the nucleocapsids was found to deviate from that observed for the isolated domain, with a gain of rigidity in the first 50 residues of N_{TAIL} . The NMR intensity profile of N_{TAIL} (Figure 9B) bears some resemblance to the profile measured from the homologous N_{TAIL} domain within intact MeV nucleocapsids [52]. In the previous study of MeV, a combination of NMR, electron microscopy and small angle X-ray scattering

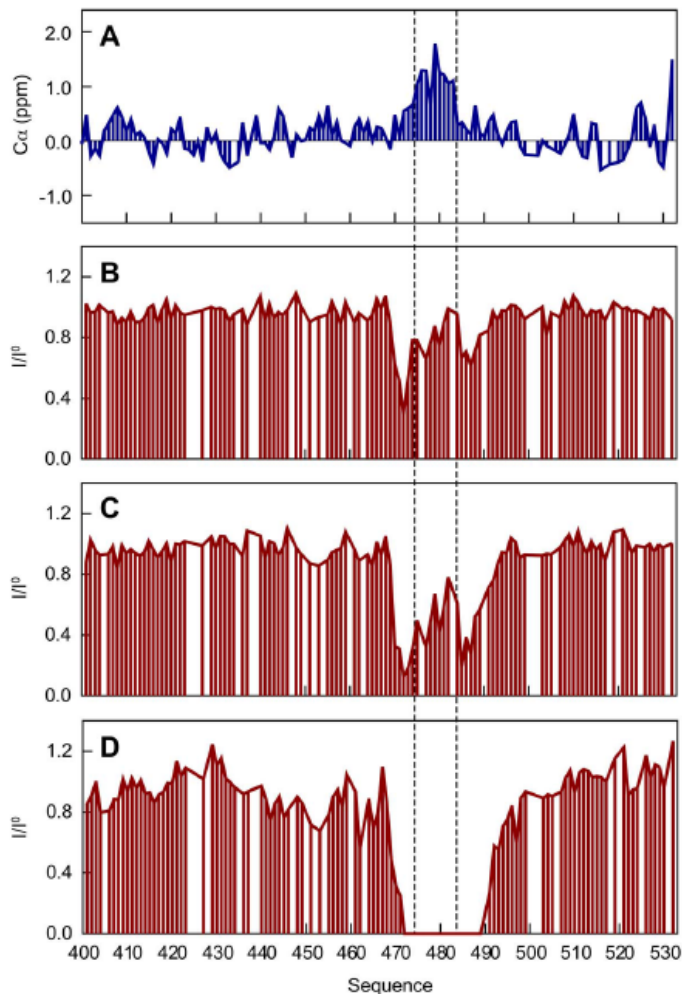


Figure 6. Conformational exchange contributions to the NMR line widths of N_{TAIL} upon interaction with XD. (A) Secondary $C\alpha$ chemical shifts of N_{TAIL} . (B)–(D) Intensity profiles of the 1H - ^{15}N HSQC spectra of N_{TAIL} with different amounts of XD: 7.5% (B), 30% (C), 440% (D). The profiles are calculated as I/I^0 , where I are the intensities in the spectrum of N_{TAIL} with XD, and I^0 are the intensities in the spectrum of the isolated N_{TAIL} domain. All intensity profiles were normalized to one outside the MoRE. In all panels the dashed lines indicate the region for which most helical propensity is present in the isolated state of N_{TAIL} according to the secondary $C\alpha$ chemical shifts. doi:10.1371/journal.ppat.1003631.g006

was used to propose a model whereby MeV N_{TAIL} exfiltrates from the inside to the outside of the helical nucleocapsids via the interstitial space between folded N_{CORE} domains. Indeed, the first 50 amino acids of MeV N_{TAIL} could not be detected by solution NMR due to restricted conformational flexibility as the disordered chain escapes from the inside to the outside of the nucleocapsids. The results obtained here are in agreement with a similar model for the location of N_{TAIL} in HeV nucleocapsids.

Despite the overall similarity in the conformational behaviour of N_{TAIL} within the nucleocapsids of HeV and MeV, the intensity profiles are however strikingly different within their MoREs. In MeV, the MoRE of N_{TAIL} appears to be co-localized on the surface of the nucleocapsids up to 95% of the time, resulting in heavily attenuated signals in the MoRE. We have previously proposed that this co-localization may have functional implications, providing a mechanism by which N_{TAIL} can position the

viral polymerase complex close to the viral RNA. In HeV, this co-localization appears weaker, as evidenced by appreciable intensities of the solution NMR signals in the MoRE (Figure 9B). This different location of the MoRE might provide a basis for a different mechanism of recognition by the polymerase complex in the two viruses.

Insight into the molecular mechanisms underlying the HeV N_{TAIL} -XD interaction

Chemical shift titrations suggest that upon interaction with the C-terminal X domain of the phosphoprotein, the MoRE of N_{TAIL} undergoes α -helical folding. This observation is supported by previous studies of this interaction using circular dichroism (CD) and EPR spectroscopy that showed that XD triggers α -helical folding of N_{TAIL} upon binding [7,8]. On the basis of the changes in the NMR signal intensities of N_{TAIL} upon addition of an

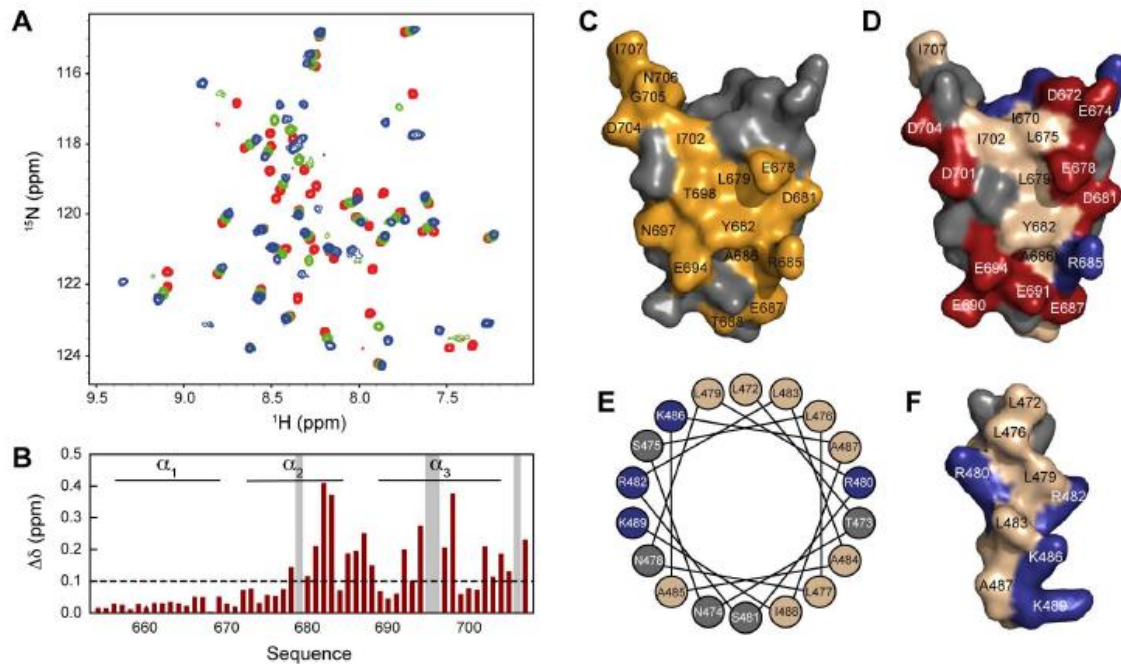


Figure 7. Interaction between ^{15}N , ^{13}C labeled XD and unlabeled N_{TAIL} . (A) Superposition of ^1H - ^{15}N HSQC spectra of XD containing increasing amounts of N_{TAIL} : 0% (red), 71% (green) and 416% (blue). The spectra were obtained at a ^1H frequency of 600 MHz and 298 K. (B) Combined ^1H and ^{15}N chemical shift difference ($\Delta\delta = \sqrt{(\Delta\delta_{\text{H}})^2 + (\Delta\delta_{\text{N}}/8)^2}$) between the resonances in the ^1H - ^{15}N HSQC spectrum of isolated XD and the spectrum with 416% N_{TAIL} . Gray vertical shading indicates residues for which the NMR resonances disappear upon addition of N_{TAIL} . Chemical shift changes above the horizontal dashed line are considered significant. (C) Surface representation of the crystal structure of XD displaying residues experiencing large chemical shift changes upon addition of N_{TAIL} (orange). The orientation of XD is the same as shown in Figure 3. (D) Surface representation of the crystal structure of XD displaying the location of hydrophobic (beige), negatively charged (red) and positively charged (blue) residues. (E) Helical wheel representation of the MoRE of N_{TAIL} encompassing residues 472–489 with the same color-coding as in (D). (F) Surface representation of the MoRE of N_{TAIL} assuming that the MoRE adopts an α -helical conformation. doi:10.1371/journal.ppat.1003631.g007

increasing amount of XD (Figure 6B, C, D), we suggest that the formation of the N_{TAIL} -XD complex relies on a combination of conformational selection and induced folding, where complex formation proceeds via an initial binding of a short helix in the centre of the MoRE that is subsequently extended by α -helical folding of the adjacent residues. The complex between N_{TAIL} and XD appears to be dynamic, as the resonances of the MoRE of N_{TAIL}

do not reappear even for large saturating amounts of XD, suggesting helical fraying of the MoRE of N_{TAIL} on the surface of XD.

To obtain further insight into the structural basis of transcription and replication of HeV, we determined the crystal structure of XD. The structure consists of a three-helix bundle arrangement in agreement with previous studies showing that the *Henipavirus* X domains form autonomously folding units adopting an α -helical

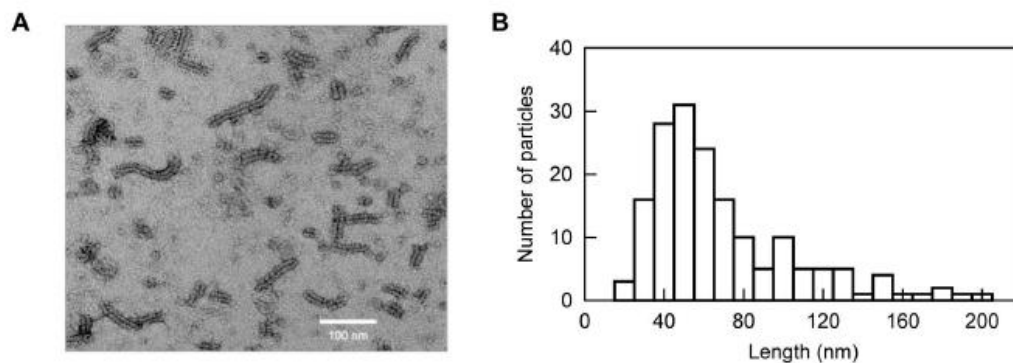


Figure 8. Electron microscopy of intact HeV nucleocapsids. (A) Negative stain electron micrograph of HeV N showing the presence of nucleocapsid-like structures. (B) Histogram showing the distribution of lengths of 170 randomly chosen nucleocapsids. doi:10.1371/journal.ppat.1003631.g008

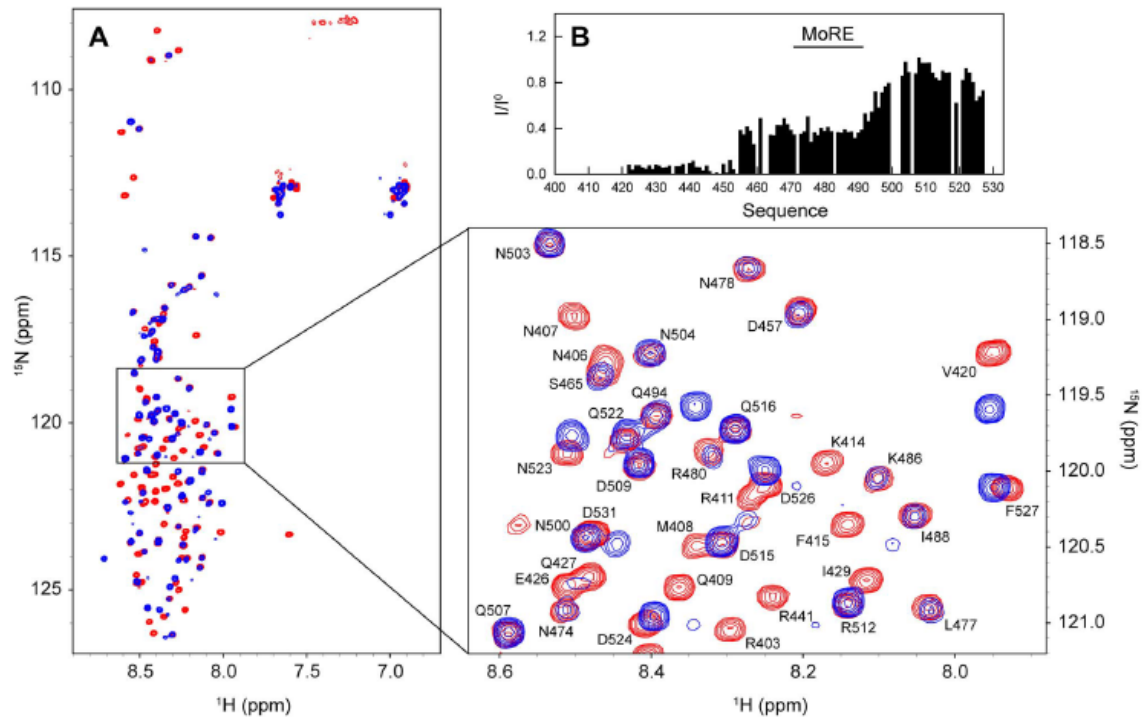


Figure 9. Characterization of N_{TAIL} in the context of intact nucleocapsids. (A) Superposition of the 1H - ^{15}N HSQC spectra of isolated N_{TAIL} (red) and intact nucleocapsids (blue). The spectra were obtained at a 1H frequency of 600 MHz and 293 K. The expanded region shows the spectral assignment of a number of resonances in the spectrum of the isolated N_{TAIL} domain. (B) Normalized intensity profile of the HeV nucleocapsid 1H - ^{15}N HSQC spectrum. The profile is calculated as I/I^0 , where I are the intensities in the spectrum of the nucleocapsids and I^0 are the intensities in the spectrum of the isolated N_{TAIL} domain. doi:10.1371/journal.ppat.1003631.g009

conformation [7]. NMR chemical shifts and RDCs of XD measured in solution confirmed the triple-helix structure. The structure of HeV XD is very similar to the structure of the XD domains of MeV [58], SeV [59] and mumps virus [60] P proteins with backbone RMSD values of 1.22, 1.34 and 0.88 Å, respectively. Although a comparison of the primary sequences of the X domains of SeV, MeV and HeV shows significant divergence, the hydrophobic residues buried within the XD structure are conserved among the three viruses (Figure S4 in Text S1). The reproduction of an essentially identical hydrophobic core in all three viruses implies that this particular triple-helix arrangement has a strong functional relevance in paramyxoviruses.

Our results show that the two helices $\alpha 2$ and $\alpha 3$ of XD constitute the binding site for the MoRE of N_{TAIL} , as also observed in the corresponding N_{TAIL} -XD pairs in both MeV and SeV [54,61]. The interaction between XD and N_{TAIL} in HeV appears to be controlled by a combination of long-range electrostatic forces that correctly orient N_{TAIL} prior to accommodation in the narrow hydrophobic pocket on the surface of XD. The extent to which these long-range electrostatic interactions play a role in complex formation remains however to be established, since previous studies showed that salt concentrations as high as 1 M do not affect the N_{TAIL} -XD binding affinity [7]. A comparison of the distribution of hydrophobic and charged residues on the surface of the X domains of SeV, MeV and HeV shows that they all display a hydrophobic pocket of varying size at the interface between helices $\alpha 2$ and $\alpha 3$ (Figure S5 in Text S1). In contrast to the hydrophobic core that is reproduced in all three viruses and provides the stabilizing scaffold for XD, the charge distribution on

and around the interaction surface with N_{TAIL} , varies significantly over X domains. These differences presumably play a role in the selectivity of the N_{TAIL} -XD interactions in SeV, MeV and HeV.

Although high resolution NMR spectroscopy, as well as previous biochemical data [7], suggest that HeV N_{TAIL} -XD forms a so-called 'fuzzy' complex [62], which is characterized by mobility within the complex, we have nevertheless attempted to model the structure of the complex in terms of a single conformation using chemical shift mapping of the two interacting interfaces. The results (Figure S6 in Text S1) reveal that neither chemical shifts, nor electrostatic interactions, are able to distinguish between the directionality of the N_{TAIL} helix on the surface of XD. The experimental data do not distinguish rotational symmetry about the axis of the N_{TAIL} helix, although two conformations are most probable, both having the hydrophobic face of the MoRE of N_{TAIL} identified in Figure 7 in contact with the hydrophobic interface of XD. In both conformations the two arginines flanking the hydrophobic face on N_{TAIL} interact with acidic patches on the surface of XD. Notably, irrespective of the direction of the MoRE, its optimal position with respect to the observed chemical shift perturbations on the surface of XD is very similar to that observed in the chimeric crystal structure of MeV N_{TAIL} -XD (Figure S6 in Text S1).

Binding of XD to intact nucleocapsids and functional implications for transcription and replication

In paramyxoviruses, genome transcription and replication rely on the N-P interaction that is critical for the recruitment of the

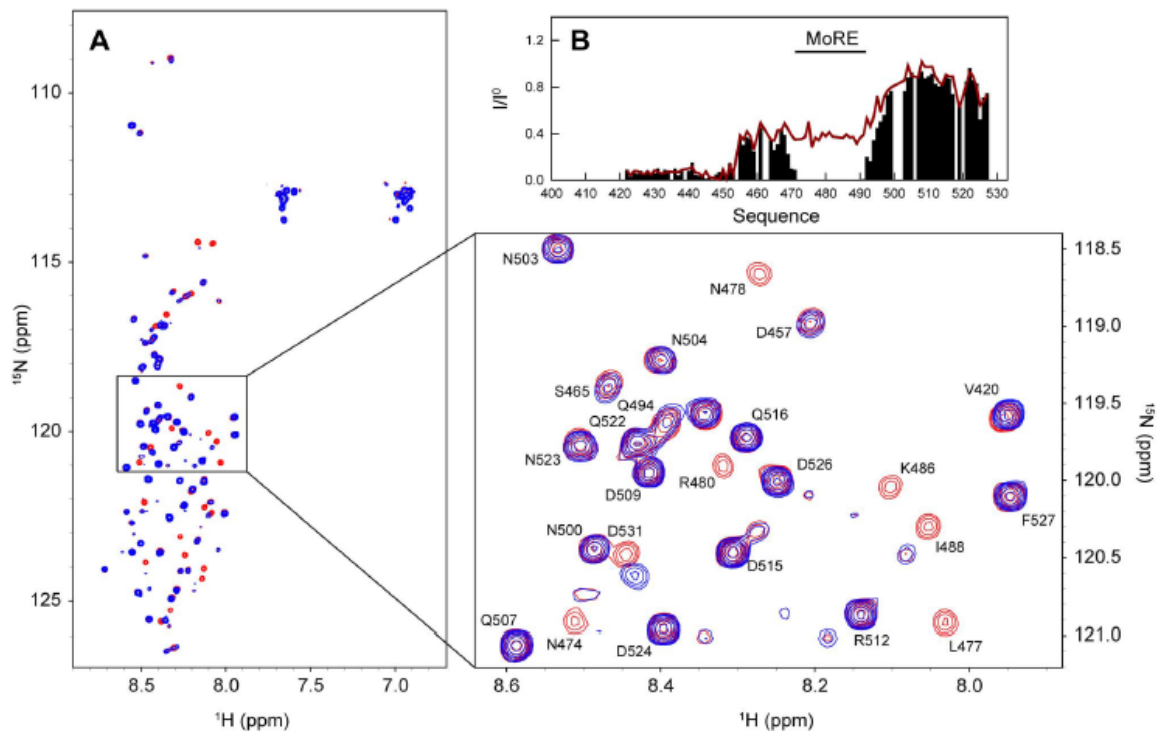


Figure 10. Characterization of the interaction between ^{15}N labeled nucleocapsids and unlabeled XD. (A) Superposition of the ^1H - ^{15}N HSQC spectra of intact nucleocapsids (red) and intact nucleocapsids with 1.5-fold molar excess of XD (blue). The spectra were obtained at a ^1H frequency of 600 MHz and 293 K. The expanded region shows the spectral assignment of a number of resonances in the spectrum of the nucleocapsids. The assignment was transferred from the spectrum of the isolated N_{TAIL} domain. (B) Intensity profile (black) of the HSQC spectrum of the nucleocapsids with XD. The profile is calculated as I/I^0 , where I are the intensities in the spectrum of the nucleocapsids with XD, and I^0 are the intensities in the spectrum of the isolated N_{TAIL} domain. The red line corresponds to the intensity profile of the nucleocapsid spectrum obtained in the absence of XD (see figure 9).

doi:10.1371/journal.ppat.1003631.g010

polymerase. This interaction is thought to make the viral RNA accessible to the polymerase L through a structural reorganization of the nucleocapsid. The molecular mechanisms of this critical step remain, however, poorly understood and to the best of our knowledge no such nucleocapsid conformational change induced by a polymerase complex has been observed so far in any *Paramyxoviridae* member. As a step towards a better understanding of the structural and molecular basis of transcription and replication, we have performed the first investigation of XD binding to N_{TAIL} in the context of intact nucleocapsids. Our results show that XD interacts only locally with the MoRE of N_{TAIL} , and that no additional residues in N_{TAIL} are involved in the interaction. Importantly, the NMR signal intensities outside the MoRE are not perturbed by the presence of XD (Figure 10B). The similarity of the high-resolution NMR spectroscopic parameters measured in the presence and absence of the partner, which together represent the signature that the N-terminal part of N_{TAIL} remains sterically hindered between turns of the nucleocapsid, indicates that there is no major reorganization of the nucleocapsid upon interaction. Any further opening of the interstitial space between turns of the nucleocapsid helix would be expected to give rise to enhanced flexibility and therefore increased signal intensity, both within the first 50 amino acids and in the 10–15 amino acids immediately following this region. In addition, RNase digestion experiments did not reveal any hyperchromic effect at 260 nm

upon addition of XD (data not shown), indicating that XD does not render the encapsidated RNA more accessible to the solvent.

In combination, these observations provide the first experimental evidence that XD can be accommodated on N_{TAIL} without triggering a nucleocapsid rearrangement that could increase the accessibility of the polymerase complex to the viral genome. We can speculate that the presence of the full-length, tetrameric P protein, the polymerase L and/or additional host factors [63], are necessary to trigger any conformational transition that may occur prior to replication of the viral RNA by the polymerase machinery. Notably, the major inducible heat shock protein 70 (hsp70) was shown to enhance both transcription and replication in MeV and in the closely related canine distemper virus (CDV), with this effect relying on an interaction between hsp70 and N [64–70]. Taking into account the fact that hsp70 has been suggested to trigger conformational changes within CDV nucleocapsids [71,72], it is tempting to speculate that unwinding of the nucleocapsid may require hsp70 also in the case of HeV.

Materials and Methods

Expression and purification of N_{TAIL}

The HeV N_{TAIL} construct, which encodes residues 401–532 of N, with an N-terminal hexahistidine tag has been described previously [15]. Isotopically labeled (^{15}N , ^{13}C) HeV N_{TAIL} was

prepared by growing transformed *E. coli* T7 pLysS (New England Biolabs) cells at 37°C in minimal M9 medium [73] containing 100 mg/L ampicillin and 34 mg/L chloramphenicol supplemented with $^{15}\text{NH}_4\text{Cl}$ (1 g/L) and ^{13}C -glucose (2 g/L). A 150 mL preculture grown overnight to saturation in 2YT medium containing 100 mg/L ampicillin and 34 mg/L chloramphenicol (2YT-AC) was harvested, washed in minimal M9 medium, and inoculated into 1 L of minimal M9 medium supplemented with ampicillin and chloramphenicol (M9-AC). The culture was grown at 37°C. When the optical density (OD) at 600 nm reached 0.6, protein expression was induced by the addition of 0.5 mM isopropyl β -D-thiogalactopyranoside (IPTG), and the cells were grown overnight at 28°C.

Isotopically ^{13}C , ^{15}N and ^{15}N labeled N_{TAIL} samples were purified as previously described [7] except that NMR buffer was used during size exclusion chromatography (SEC), and that an additional purification step was performed: the eluent from SEC was loaded onto an anion exchange MonoQ column (GE, Healthcare) and elution was carried out using a gradient of NaCl (10–1000 mM) in 10 mM Tris/HCl pH 8 buffer. The sample was concentrated using Centricon Plus-20 (molecular cutoff of 5,000 Da) (Millipore) and then dialyzed against NMR buffer. A protease inhibitor cocktail was added to the samples before storage at -20°C . The expression and purification of unlabeled N_{TAIL} has already been described [15]. All purification steps, except for gel filtrations, were carried out at 4°C. The protein concentration was estimated using the previously experimentally determined absorption coefficients [8]. Because N_{TAIL} is devoid of both Trp and Tyr residues, the absorbance was measured at 254 nm using the absorption coefficient experimentally determined at this wavelength [8].

Expression and purification of nucleocapsids

The HeV N construct, encoding full-length N (residues 1–532), with a hexahistidine tag fused to the C-terminus, has already been described [7]. Isotopically labeled (^{15}N) HeV nucleocapsids were prepared by growing transformed *E. coli* T7 pLysS (New England Biolabs) cells according to the protocol described by Marley et al [74]. Briefly, an 80 mL overnight preculture grown to saturation in 2YT-AC was used to inoculate 2 L of the same medium. When the OD at 600 nm was approximately 0.6, the culture was harvested and the pellet was resuspended in 500 mL of M9-AC medium containing $^{15}\text{NH}_4\text{Cl}$ (1 g/L) and glucose (2 g/L). After an additional growth step at 37°C for 2 hours, 1 mM IPTG was added and the culture was further incubated at 37°C for 3.5 hours. In all cases, the induced cells were harvested, washed and collected by centrifugation (5,000 g, 10 min). The resulting pellets were frozen at -20°C .

The nucleocapsids were purified by resuspending cellular pellets in 5 volumes (v/w) buffer A (50 mM Tris/HCl pH 8, 300 mM NaCl, 10 mM imidazole, 1 mM phenyl-methyl-sulphonyl-fluoride (PMSF) supplemented with 0.1 mg/mL lysozyme, 10 $\mu\text{g}/\text{mL}$ DNase I, 20 mM MgSO_4 and protease inhibitor cocktail (Sigma, 1 mL per 25 mL of bacterial lysate). After 30 min incubation with gentle agitation, the cells were disrupted by sonication (using a 750 W sonicator and 4 cycles of 30 s each at 45% power output). The lysate was clarified by centrifugation at 30,000 g for 30 min. The clarified supernatant was loaded onto a 5 mL HisTrap FF column (GE, Healthcare), previously equilibrated in buffer A. Elution was carried out using a gradient of imidazole (10–500 mM) in buffer A. Eluates were analyzed by SDS-PAGE for the presence of N. The fractions of interest were pooled and then injected onto a HiPrep 26/10 desalting column (GE, Healthcare). Elution was carried out in NMR buffer (50 mM BisTris, pH 6,

500 mM NaCl). The sample was concentrated using Centricon Plus-20 (molecular cutoff of 10,000 Da) (Millipore) and protease inhibitor cocktail (1/40) was added. The sample was immediately subjected to NMR and electron microscopy analyses. The concentration of nucleocapsid samples was roughly estimated using the theoretical absorption coefficient at 280 nm as provided by the ProtParam program at the ExpASY server [75].

Expression and purification of XD

A synthetic gene coding for a slightly enlarged XD domain (residues 654–707) of HeV P was cloned in the pET28a vector and was expressed at 20°C in *E. coli* BL21 Rosetta [DE3] with two N-terminal non-native amino acids (Met and Val) and two C-terminal non-native residues (Leu and Glu) followed by a hexahistidine tag. Selenomethionine-substituted protein was produced by growing the bacterial cells in minimal medium and by adding SeMet before induction (1 mM IPTG) of the heterologous protein expression. The protein fragment was purified by Ni^{2+} affinity chromatography on a His-Select (Sigma) and size exclusion chromatography on a Superdex S75 (GE Healthcare) column in 20 mM Tris buffer at pH 7.5 supplemented with 150 mM NaCl, 50 mM Glu/Arg, 0.5 mM TCEP and an anti-protease cocktail (Roche) (Buffer A).

For producing isotopically (^{13}C , ^{15}N) labeled XD without a hexahistidine tag for NMR spectroscopic measurements, the gene was subcloned into the pETM40 vector (EMBL) with an N-terminal MBP tag followed by a TEV protease cleavage site. The protein was expressed as a MBP fusion as described [74] and purified by affinity chromatography on an Amylose resin column (NEB) and size exclusion chromatography on a Superdex S200 (GE Healthcare) column in buffer A. After TEV protease cleavage, the XD domain was purified by size exclusion chromatography on a Superdex S75 column in 20 mM Bis-Tris at pH 6, supplemented with 150 mM NaCl and 0.5 mM TCEP. The expression and purification of unlabeled XD were previously described [7].

NMR spectral assignment of N_{TAIL}

The spectral assignment of N_{TAIL} was carried out using a 200 μM ^{13}C , ^{15}N labeled sample in 20 mM Bis-Tris, pH 6.0 and 500 mM NaCl. A set of six triple resonance BEST-type spectra were acquired at 293 K and a ^1H frequency of 800 MHz: HNC0, intra-residue HN(CA)CO, HN(CO)CA, intra-residue HNCA, HN(COCA)CB, intra-residue HN(CA)CB [76]. All spectra were acquired with sweep widths of 8.0 (^1H), 1.9 (^{15}N), 1.3 ($^{13}\text{C}'$), 2.7 ($^{13}\text{C}\alpha$) and 12.4 kHz ($^{13}\text{C}\beta$) where the number of complex points in each dimension was 512 (^1H), 38 (^{15}N), 60 ($^{13}\text{C}'$), 105 ($^{13}\text{C}\alpha$) and 122 ($^{13}\text{C}\beta$). Spectra were processed using NMRPipe [77] and analyzed using CcpNmr [78] and Sparky [79]. ^1H - ^{15}N HSQC resonances were not visible for residues S400, S439, S454 and S463. Secondary chemical shifts and secondary structure propensities were calculated using the random coil chemical shifts from RefDB [80].

^{15}N relaxation measurements of N_{TAIL}

^{15}N CPMG R_2 relaxation rates and $\{^1\text{H}\}$ - ^{15}N steady state heteronuclear Overhauser effects (nOes) of N_{TAIL} were measured at a ^1H frequency of 600 MHz and 293 K using standard pulse sequences on a 150 μM ^{15}N labeled sample [81]. The ^{15}N R_2 relaxation rates were obtained by sampling the decay of the magnetization at the following time points: 0.01, 0.03, 0.05, 0.07, 0.09, 0.13, 0.17, 0.21 and 0.25 s. A repeat of the time point at 0.07 s was measured for the purpose of error estimation. For the heteronuclear nOes, the amide protons were saturated using a 3 s WALTZ16 decoupling scheme that in the reference experiment

was replaced by a 3 s delay. The recycle delay in both experiments was set to 2 s. The nOe values were calculated from the ratio between signal intensities in the saturated and the reference experiment, where the standard deviation in the noise was taken as a measure for the error in the signal intensity.

Titration of ^{15}N labeled N_{TAIL} with unlabeled XD

The interaction between N_{TAIL} and XD was studied using a ^{15}N labeled sample of N_{TAIL} by addition of an increasing amount of unlabeled XD. The same buffer conditions as described above for N_{TAIL} was used for both proteins. 1H - ^{15}N HSQC spectra were measured at 293 K and a 1H frequency of 800 MHz for the following concentrations of $[XD]:[N_{TAIL}]$: 0.0:40.0, 3.0:39.8, 5.9:39.6, 8.9:39.4, 11.8:39.2, 14.6:39.0, 17.5:38.8, 23.2:38.6, 28.8:38.5, 34.4:38.3, 45.7:38.1, 56.9:37.9, 162.2:37.2 μM .

Crystal structure of XD

Crystals of XD were grown at 293 K by the hanging drop vapor diffusion method in 100 mM Bis-Tris buffer, pH 6 containing 50 mM $MgCl_2$ and 26–28% PEG 3350. The protein concentration was between 10 and 15 mg/mL. Crystals were optimized by micro-seeding in the same solution. X-ray diffraction data were collected on ESRF beamline ID14-4 [82] using an inverse beam strategy. Data were integrated, scaled and merged with XDS and XSCALE [83]. The anomalously scattering Se atoms were located and initial phases were obtained using SHELX C/D/E [84] as implemented in HKL2MAP [85]. An initial model obtained with ARPwARP [86] from the CCP4 suite [87] was subsequently refined at 1.65 Å resolution with phenix.refine using unmerged Friedel pairs [88] and manually rebuilt after visual inspection in Coot [89]. After multiple cycles of refinement, the final crystallographic R-factor and R_{free} values were 18.8% and 22.4%, respectively. Other statistics are reported in Table S2 in Text S1.

NMR spectral assignment of XD

The spectral assignment of XD was carried out using a 2.5 mM ^{13}C , ^{15}N labeled sample in 20 mM Bis-Tris, 50 mM Glu/Arg, 150 mM NaCl at pH 6.0. A set of six triple resonance BEST-type spectra were acquired at 298 K and a 1H frequency of 800 MHz: HNCO, intra-residue HN(CA)CO, HN(CO)CA, intra-residue HNCA, HN(COCA)CB, intra-residue HN(CA)CB [76]. All spectra were acquired with sweep widths of 10.0 (1H), 1.9 (^{15}N), 1.9 ($^{13}C'$), 3.8 ($^{13}C\alpha$) and 14.0 kHz ($^{13}C\beta$) where the number of complex points in each dimension was 704 (1H), 39 (^{15}N), 70 ($^{13}C'$), 125 ($^{13}C\alpha$) and 135 ($^{13}C\beta$). Spectra were processed using NMRPipe [77] and analyzed using CcpNmr [78] and Sparky [79]. The 1H - ^{15}N HSQC spectra of XD with and without 50 mM Glu/Arg were almost identical allowing the transfer of assignments to the spectrum without Glu/Arg buffer.

Measurement of residual dipolar couplings in XD

Four types of residual dipolar couplings ($^1D_{HN-N}$, $^1D_{C\alpha-H\alpha}$, $^1D_{C\alpha-C'}$, $^2D_{HN-C'}$) were measured in XD using BEST-type experiments at a 1H frequency of 600 MHz and 298 K [90]. The protein was aligned in a suspension of filamentous phages (ASLA Biotech) at 11.5 mg/mL giving rise to a residual deuterium splitting of 7 Hz. All spectra were acquired with sweep widths of 7.5 (1H), 1.4 (^{15}N), 1.2 ($^{13}C'$) and 2.9 ($^{13}C\alpha$) where the number of complex points in each dimension was 512 (1H), 40 (^{15}N), 80 ($^{13}C'$) and 80 ($^{13}C\alpha$). Spectra were processed using NMRPipe [77] and analyzed using CcpNmr [78]. The alignment tensor was determined by fitting the experimental RDCs to the crystal

structure of XD using the program Module [91]. Only the RDCs for residues within regular secondary structure elements were used in the analysis.

Titration of ^{15}N labeled XD with unlabeled N_{TAIL}

In order to map the interaction of N_{TAIL} on XD, a titration was carried out using a ^{15}N , ^{13}C labeled sample of XD by addition of an increasing amount of unlabeled N_{TAIL} . The titration was carried out in 20 mM Bis-Tris buffer, 150 mM NaCl at pH 6.0. 1H - ^{15}N HSQC spectra were measured at 298 K and a 1H frequency of 800 MHz for the following concentrations of $[N_{TAIL}]:[XD]$: 0.0:80.6, 9.5:80.2, 19.0:79.9, 28.4:79.5, 37.7:79.2, 56.0:78.5, 74.0:77.8, 91.7:77.1, 126.2:75.8, 159.6:74.5, 215.3:72.4, 289.6:69.5, 358.2:66.9, 659.3:55.4 μM . Concentrations were determined by amino acid analysis.

NMR experiments on recombinant HeV nucleocapsids

A 1H - ^{15}N HSQC spectrum of a 37 μM ^{15}N labeled sample of intact nucleocapsids was obtained at 293 K and a 1H frequency of 600 MHz. The interaction between the nucleocapsids and XD was studied by the addition of an excess amount of unlabeled XD (55 μM). The 1H - ^{15}N HSQC experiment was repeated for this sample with the same parameters as for the isolated nucleocapsids.

Negative staining transmission electron microscopy studies

Drops of 2 μL of freshly purified HeV nucleocapsids at 0.05 mg/mL were deposited on evaporated carbon membranes on standard copper grids (300 Meshes). Grids were exposed to plasma glow discharge for 20 seconds, prior to protein deposition in order to increase protein adhesion. Grids were then negatively stained with a 2% uranyl acetate solution and imaged using a TECNAI 120 keV transmission electron microscope. Images were collected on a 3 by 3 cm CCD camera with a 15 μm by 15 μm pixel size.

Supporting Information

Text S1 This file contains six additional figures showing the assigned HSQC spectra of N_{TAIL} and XD (Figures S1, S2 and S3), comparison of the X domains of Sendai, Measles and Hendra viruses (Figures S4 and S5) and a model of the Hendra virus N_{TAIL} -XD complex from NMR chemical shift perturbations (Figure S6). In addition, the file contains four additional tables with chemical shift values of N_{TAIL} and XD (Tables S1 and S3), data collection and refinement statistics of XD (Tables S2) and residual dipolar couplings of XD (Table S4). Crystallographic coordinates and structure factors of XD have been deposited in the Protein Data Bank with accession code 4HEO. (DOC)

Acknowledgments

We thank Dr. Jean-Marie Bourhis for helpful discussions and Jean-Pierre Andrieu from the Institut de Biologie Structurale platform of the Partnership for Structural Biology (Grenoble) for the assistance and access to the amino acid analysis facility.

Author Contributions

Conceived and designed the experiments: RWHR MJ MRJ SL MB. Performed the experiments: GC JH FY DB RS NT NP MRJ. Analyzed the data: GC JH FY NT NP MJ MRJ SL MB. Contributed reagents/materials/analysis tools: GC MB. Wrote the paper: RWHR MJ MRJ SL MB.

References

- Eaton BT, Broder CC, Middleton D, Wang L-F (2006) Hendra and Nipah viruses: different and dangerous. *Nat Rev Microbiol* 4: 23–35.
- Wang L-F, Yu M, Hansson E, Pritchard LI, Shiell B, et al. (2000) The exceptionally large genome of Hendra virus: support for creation of a new genus within the family Paramyxoviridae. *J Virol* 74: 9972–9979.
- Marsh GA, de Jong C, Barr JA, Tachedjian M, Smith C, et al. (2012) Cedar Virus: A Novel Henipavirus Isolated from Australian Bats. *PLoS Pathogens* 8: e1002836.
- Rockx B, Winegar R, Freiberg AN (2012) Recent progress in henipavirus research: molecular biology, genetic diversity, animal models. *Antiviral Res* 95: 135–149.
- Curran J, Kolakofsky D (1999) Replication of paramyxoviruses. *Adv Virus Res* 54: 403–422.
- Halpin K, Bankamp B, Harcourt BH, Bellini WJ, Rota PA (2004) Nipah virus conforms to the rule of six in a minigenome replication assay. *J Gen Virol* 85: 701–707.
- Habchi J, Blangy S, Mamelli L, Jensen MR, Blackledge M, et al. (2011) Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipavirus. *J Biol Chem* 286: 13583–13602.
- Martinho M, Habchi J, El Habre Z, Nesme I, Guigliarelli B, et al. (2013) Assessing induced folding within the intrinsically disordered C-terminal domain of the Henipavirus nucleoproteins by site-directed spin labeling EPR spectroscopy. *J Biomol Struct Dyn* 31: 453–471.
- Bloquel D, Habchi J, Gruet A, Blangy S, Longhi S (2012) Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies. *Mol Biosyst* 8: 392–410.
- Chan YP, Koh CL, Lam SK, Wang L-F (2004) Mapping of domains responsible for nucleocapsid protein-phosphoprotein interaction of Henipaviruses. *J Gen Virol* 85: 1675–1684.
- Omi-Furutani M, Yoneda M, Fujita K, Ikeda F, Kai C (2010) Novel phosphoprotein-interacting region in Nipah virus nucleocapsid protein and its involvement in viral replication. *J Virol* 84: 9793–9799.
- Messling V von, Cattaneo R (2012) A Henipavirus Vaccine in Sight. *Science* 337: 651–652.
- Lou Z, Xu Y, Xiang K, Su N, Qin L, et al. (2006) Crystal structures of Nipah and Hendra virus fusion core proteins. *FEBS J* 273: 4538–4547.
- Bowden TA, Crispin M, Harvey DJ, Aricescu AR, Grimes JM, et al. (2008) Crystal structure and carbohydrate analysis of Nipah virus attachment glycoprotein: a template for antiviral and vaccine design. *J Virol* 82: 11628–11636.
- Habchi J, Mamelli L, Darbon H, Longhi S (2010) Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLoS ONE* 5: e11684.
- Habchi J, Longhi S (2012) Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. *Mol Biosyst* 8: 69–81.
- Tomba P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27: 527–533.
- Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6: 197–208.
- Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 18: 756–764.
- Tomba P (2011) Unstructural biology coming of age. *Curr Opin Struct Biol* 21: 419–425.
- Chouard T (2011) Structural biology: Breaking the protein rules. *Nature* 471: 151–153.
- Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, et al. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci USA* 102: 17002–17007.
- Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc Natl Acad Sci USA* 102: 13099–13104.
- Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17: 3–14.
- Jensen MR, Marlswick PRI, Meier S, Griesinger C, Zweckstetter M, et al. (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure* 17: 1169–1185.
- Fisher CK, Stultz CM (2011) Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struct Biol* 21: 426–431.
- Schneider R, Huang J, Yao M, Communie G, Ozenne V, et al. (2012) Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol Biosyst* 8: 58–68.
- Ozenne V, Schneider R, Yao M, Huang J, Salmon L, et al. (2012) Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *J Am Chem Soc* 134: 15138–15148.
- Jensen MR, Ruigrok RW, Blackledge M (2013) Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol* 23: 426–435.
- Tawar RG, Duquerroy S, Vornhein C, Varela PF, Damier-Piolle L, et al. (2009) Crystal Structure of a Nucleocapsid-Like Nucleoprotein-RNA Complex of Respiratory Syncytial Virus. *Science* 326: 1279–1283.
- Ruigrok RW, Crépin T, Kolakofsky D (2011) Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Current Opinion in Microbiology* 14: 504–510.
- Tarbouriech N, Curran J, Ruigrok RW, Burmeister WP (2000) Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat Struct Biol* 7: 777–781.
- Communie G, Crépin T, Maurin D, Jensen MR, Blackledge M, et al. (2013) Structure of the tetramerization domain of measles virus phosphoprotein. *J Virol* 87: 7166–7169.
- Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-nuclein: implications for fibrillation. *Protein Sci* 15: 2795–2804.
- Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wimmer J, Duchardt E, et al. (2002) Long-Range Interactions Within a Nonnative Protein. *Science* 295: 1719–1722.
- Tolman JR, Hanagan JM, Kennedy MA, Prestegard JH (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci USA* 92: 9279–9283.
- Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278: 1111–1114.
- Blackledge M (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Prog Nucl Magn Reson Spectrosc* 46: 23–61.
- Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12: 54–60.
- Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447: 1021–1025.
- Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19: 31–38.
- Tang C, Iwahara J, Clore GM (2006) Visualization of transient encounter complexes in protein-protein association. *Nature* 444: 383–386.
- Heggness MH, Scheid A, Chopin PW (1980) Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc Natl Acad Sci USA* 77: 2631–2635.
- Finch JT, Gibbs AJ (1970) Observations on the structure of the nucleocapsids of some paramyxoviruses. *J Gen Virol* 6: 141–150.
- Fooks AR, Stephenson JR, Warnes A, Dowsett AB, Rima BK, et al. (1993) Measles virus nucleocapsid protein expressed in insect cells assembles into nucleocapsid-like structures. *J Gen Virol* 74 (Pt 7): 1439–1444.
- Bhella D, Ralph A, Murphy LB, Yeo RP (2002) Significant differences in nucleocapsid morphology within the Paramyxoviridae. *J Gen Virol* 83: 1831–1839.
- Longhi S, Receveur-Bréchet V, Karlin D, Johansson K, Darbon H, et al. (2003) The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J Biol Chem* 278: 18638–18648.
- Goldsmith CS, Whistler T, Rollin PE, Ksiazek TG, Rota PA, et al. (2003) Elucidation of Nipah virus morphogenesis and replication using ultrastructural and molecular approaches. *Virus Research* 92: 89–98.
- Bhella D, Ralph A, Yeo RP (2004) Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction. *J Mol Biol* 340: 319–331.
- Schoehn G, Mavrakīs M, Albertini A, Wade R, Hoenger A, et al. (2004) The 12 Å structure of trypsin-treated measles virus N-RNA. *J Mol Biol* 339: 301–312.
- Desfosses A, Goret G, Fariás Estrozi L, Ruigrok RWH, Gutsche I (2011) Nucleoprotein-RNA orientation in the measles virus nucleocapsid by three-dimensional electron microscopy. *J Virol* 85: 1391–1395.
- Jensen MR, Communie G, Ribeiro EA Jr, Martínez N, Desfosses A, et al. (2011) Intrinsic disorder in measles virus nucleocapsids. *Proc Natl Acad Sci USA* 108: 9839–9844.
- Bae S-H, Dyson HJ, Wright PE (2009) Prediction of the rotational tumbling time for proteins with disordered segments. *J Am Chem Soc* 131: 6814–6821.
- Houben K, Marion D, Tarbouriech N, Ruigrok RWH, Blanchard L (2007) Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J Virol* 81: 6807–6816.
- Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RWH, et al. (2008) Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J Am Chem Soc* 130: 8055–8061.
- Gely S, Lowry DF, Bernard C, Jensen MR, Blackledge M, et al. (2010) Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. *J Mol Recognit* 23: 435–447.
- Jensen MR, Bernadó P, Houben K, Blanchard L, Marion D, et al. (2010) Structural disorder within sendai virus nucleoprotein and phosphoprotein: insight into the structural basis of molecular recognition. *Protein Pept Lett* 17: 952–960.

58. Johansson K, Bourhis J-M, Campanacci V, Cambillau C, Canard B, et al. (2003) Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem* 278: 44567–44573.
59. Blanchard I, Tarbouriech N, Blackledge M, Timmins P, Burmeister WP, et al. (2004) Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology* 319: 201–211.
60. Kingston RL, Gay LS, Baase WS, Matthews BW (2008) Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. *J Mol Biol* 379: 719–731.
61. Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW (2004) Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc Natl Acad Sci USA* 101: 8301–8306.
62. Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* 33: 2–8.
63. Delpeut S, Noyce RS, Siu RWC, Richardson CD (2012) Host factors and measles virus replication. *Curr Opin Virol* 2: 773–783.
64. Oglesbee MJ, Kenney H, Kenney T, Krakowka S (1993) Enhanced production of morbillivirus gene-specific RNAs following induction of the cellular stress response in stable persistent infection. *Virology* 192: 556–567.
65. Oglesbee MJ, Liu Z, Kenney H, Brooks CL (1996) The highly inducible member of the 70 kDa family of heat shock proteins increases canine distemper virus polymerase activity. *J Gen Virol* 77: 2125–2135.
66. Vasconcelos D, Norrby E, Oglesbee M (1998) The cellular stress response increases measles virus-induced cytopathic effect. *J Gen Virol* 79: 1769–1773.
67. Vasconcelos DY, Cai XH, Oglesbee MJ (1998) Constitutive overexpression of the major inducible 70 kDa heat shock protein mediates large plaque formation by measles virus. *J Gen Virol* 79: 2239–2247.
68. Zhang X, Glendenning C, Linke H, Parks CL, Brooks C, et al. (2002) Identification and characterization of a regulatory domain on the carboxyl terminus of the measles virus nucleocapsid protein. *J Virol* 76: 8737–8746.
69. Zhang X, Bourhis J-M, Longhi S, Carsillo T, Buccellato M, et al. (2005) Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus. *Virology* 337: 162–174.
70. Carsillo T, Zhang X, Vasconcelos D, Niewiesk S, Oglesbee M (2006) A single codon in the nucleocapsid protein C terminus contributes to in vitro and in vivo fitness of Edmonston measles virus. *J Virol* 80: 2904–2912.
71. Oglesbee M, Tatalick L, Rice J, Krakowka S (1989) Isolation and characterization of canine distemper virus nucleocapsid variants. *J Gen Virol* 70: 2409–2419.
72. Oglesbee M, Ringle S, Krakowka S (1990) Interaction of canine distemper virus nucleocapsid variants with 70K heat-shock proteins. *J Gen Virol* 71: 1585–1590.
73. Xu B, Jahic M, Enfors SO (1999) Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*. *Biotechnol Prog* 15: 81–90.
74. Marley J, Lu M, Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* 20: 71–75.
75. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, et al. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research* 40: W597–W603.
76. Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* 187: 163–169.
77. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6: 277–293.
78. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59: 687–696.
79. Goddard TD, Kneller DG (n.d.) SPARKY 3. University of California, San Francisco.
80. Zhang H, Neal S, Wishart DS (2003) ReDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR* 25: 173–195.
81. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay GM, et al. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* 33: 5984–6003.
82. McCarthy AA, Brockhauser S, Nurizzo D, Theveneau P, Mairs T, et al. (2009) A decade of user operation on the macromolecular crystallography MAD beamline ID14-4 at the ESRF. *J Synchrotron Radiat* 16: 803–812.
83. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 66: 133–144.
84. Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* 66: 479–485.
85. Pape T, Schneider TR (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Cryst* 37: 843–844.
86. Langer G, Cohen SX, Lanzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3: 1171–1179.
87. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67: 235–242.
88. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, et al. (2012) Towards automated crystallographic structure refinement with phenix-refine. *Acta Crystallogr D Biol Crystallogr* 68: 352–367.
89. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501.
90. Rasia RM, Lescop E, Palatnik JF, Boisbouvier J, Brutscher B (2011) Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. *J Biomol NMR* 51: 369–378.
91. Dosset P, Hus JC, Marion D, Blackledge M (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J Biomol NMR* 20: 223–231.

Supporting Information

Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus

Guillaume Communie^{1,2,3,4,5,6†}, Johnny Habchi^{7†}, Filip Yabukarski^{4,5,6}, David Blocquel⁷, Robert Schneider^{1,2,3}, Nicolas Tarbouriech^{4,5,6}, Nicolas Papageorgiou⁷, Rob W.H. Ruigrok^{4,5,6}, Marc Jamin^{4,5,6}, Malene Ringkjøbing Jensen^{1,2,3*}, Sonia Longhi^{7*} and Martin Blackledge^{1,2,3}

¹Université Grenoble Alpes, Institut de Biologie Structurale (IBS), F-38027 Grenoble, France

²CEA, DSV, IBS, F-38027 Grenoble, France

³CNRS, IBS, F-38027 Grenoble, France

⁴Université Grenoble Alpes, UVHCI, Grenoble, France

⁵CNRS, UVHCI, Grenoble, France

⁶Unit for Virus Host Cell Interactions, Université Grenoble Alpes-EMBL-CNRS, Grenoble, France

⁷CNRS and Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques, UMR 7257, 13288 Marseille, France

† These authors contributed equally to this work

* To whom correspondence should be addressed

Malene Ringkjøbing Jensen

E-mail: malene.ringkjobering-jensen@ibs.fr

Tel: (33) 4 38 78 47 80, Fax: (33) 4 38 78 54 94

Sonia Longhi

E-mail: Sonia.Longhi@afmb.univ-mrs.fr

Tel: (33) 4 91 82 55 80, Fax: (33) 4 91 26 67 20

Figure S1: Assignment of the ^1H - ^{15}N HSQC spectrum of HeV N_{TAIL}. The spectrum was acquired at 293 K in 20 mM Bis-Tris and 500 mM NaCl at pH 6.0.

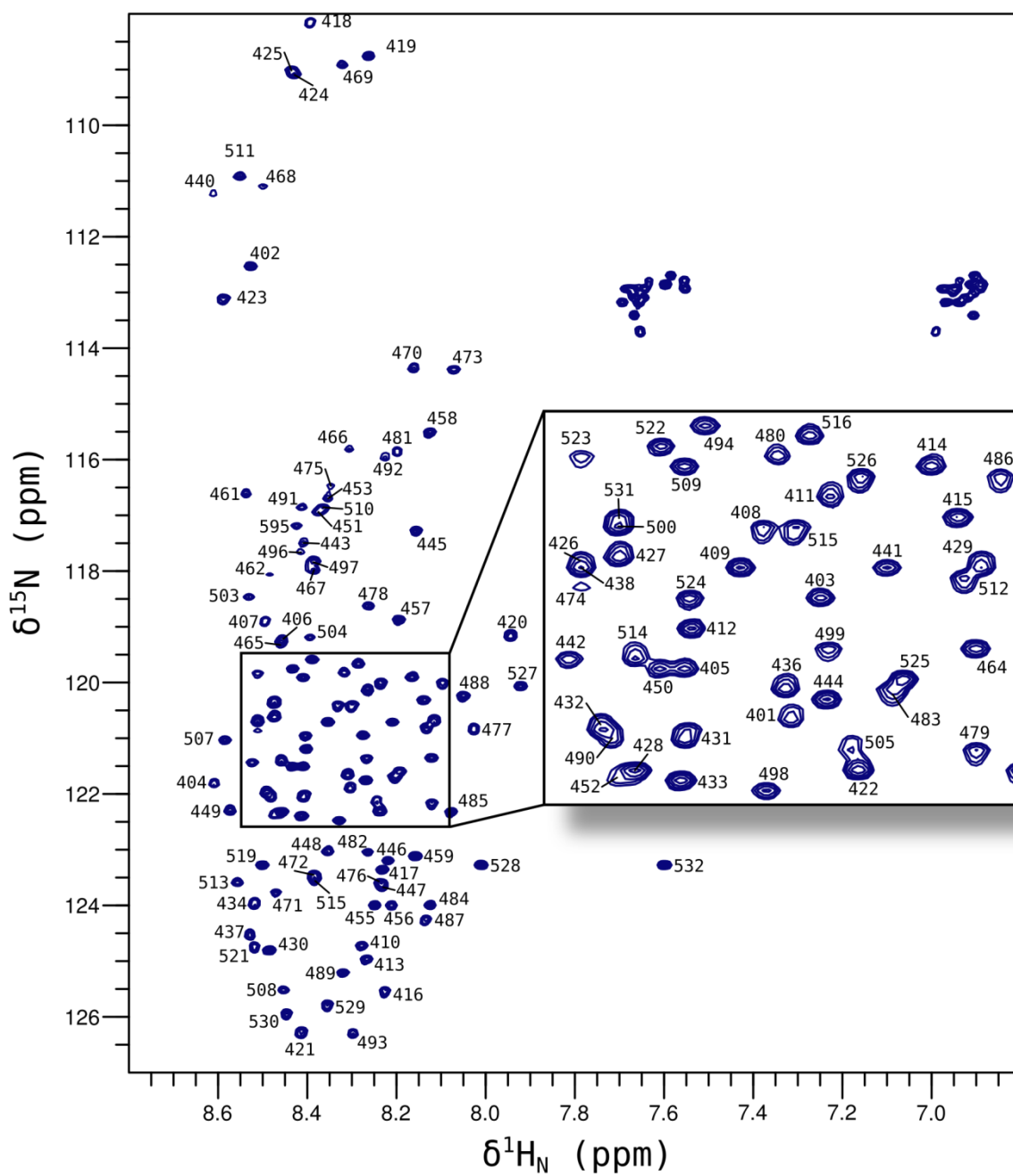


Figure S2: Assignment of the ^1H - ^{15}N HSQC spectrum of HeV XD. The spectrum was acquired at 298 K in 20 mM Bis-Tris, 50 mM Arg/Glu, 150 mM NaCl at pH 6.0.

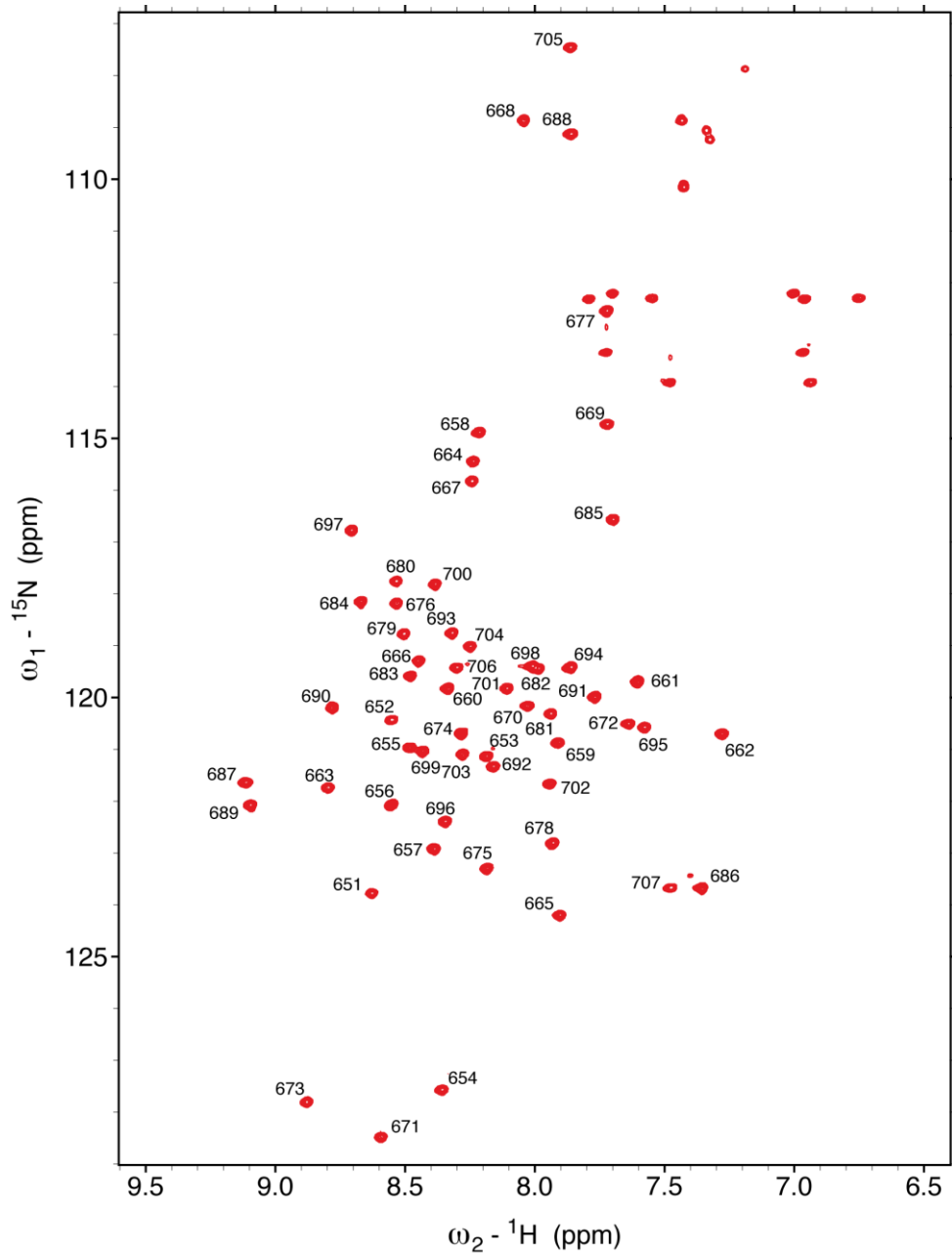


Figure S3: Expanded region of the ^1H - ^{15}N HSQC spectrum of HeV N_{TAIL}. The spectrum was obtained in the absence of XD (red) and in the presence of 30% (molar fraction) of XD (blue). The observed chemical shift changes correspond to those reported in Figure 5B.

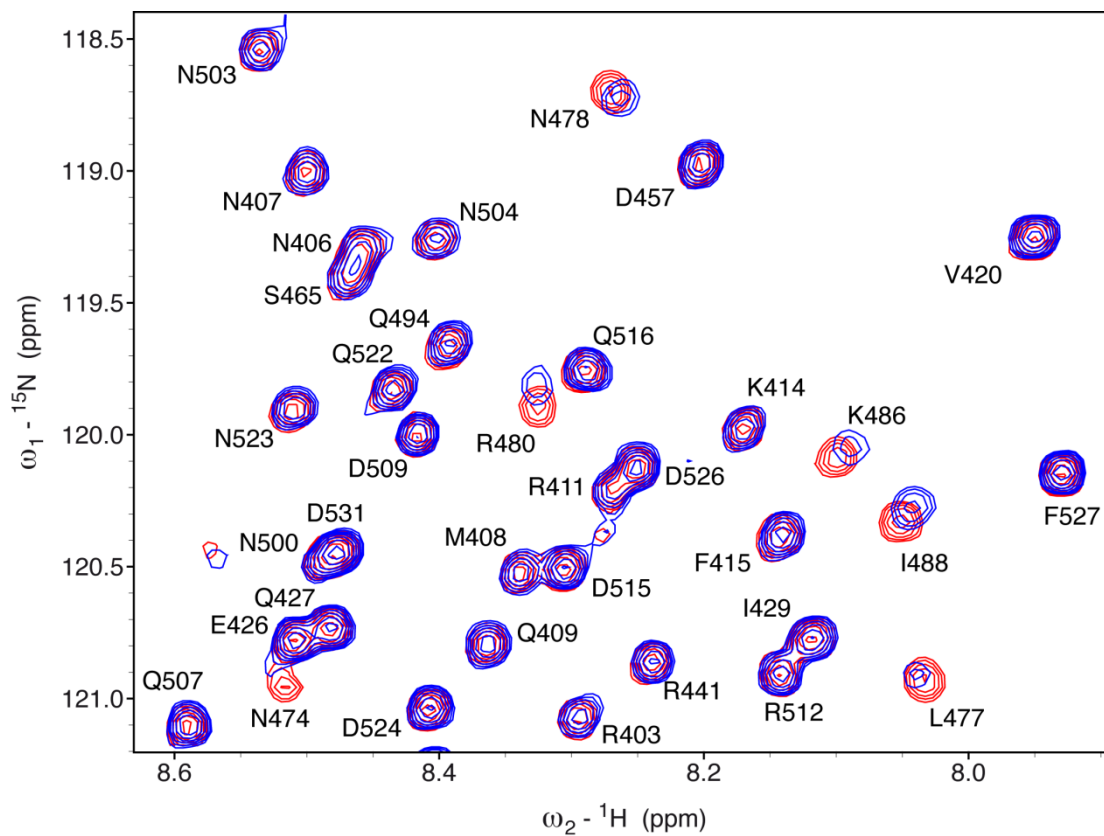


Figure S4: The three-helix arrangement in the C-terminal domains of three paramyxoviruses. (A) Sequence comparison of Sendai, Measles and Hendra virus X domains with conserved hydrophobic, positively charged, negatively charged and other residues indicated in beige, blue, red and gray colour respectively. (B) Mapping of conserved residues on the crystal structure of Hendra virus XD using the colour coding from (A). Left: front view of XD in cartoon representation. Middle: front view of XD in surface representation. Right: top view of XD in cartoon representation. The majority of conserved residues reside in the hydrophobic core of XD, while almost no conserved residues are located on the surface of XD.

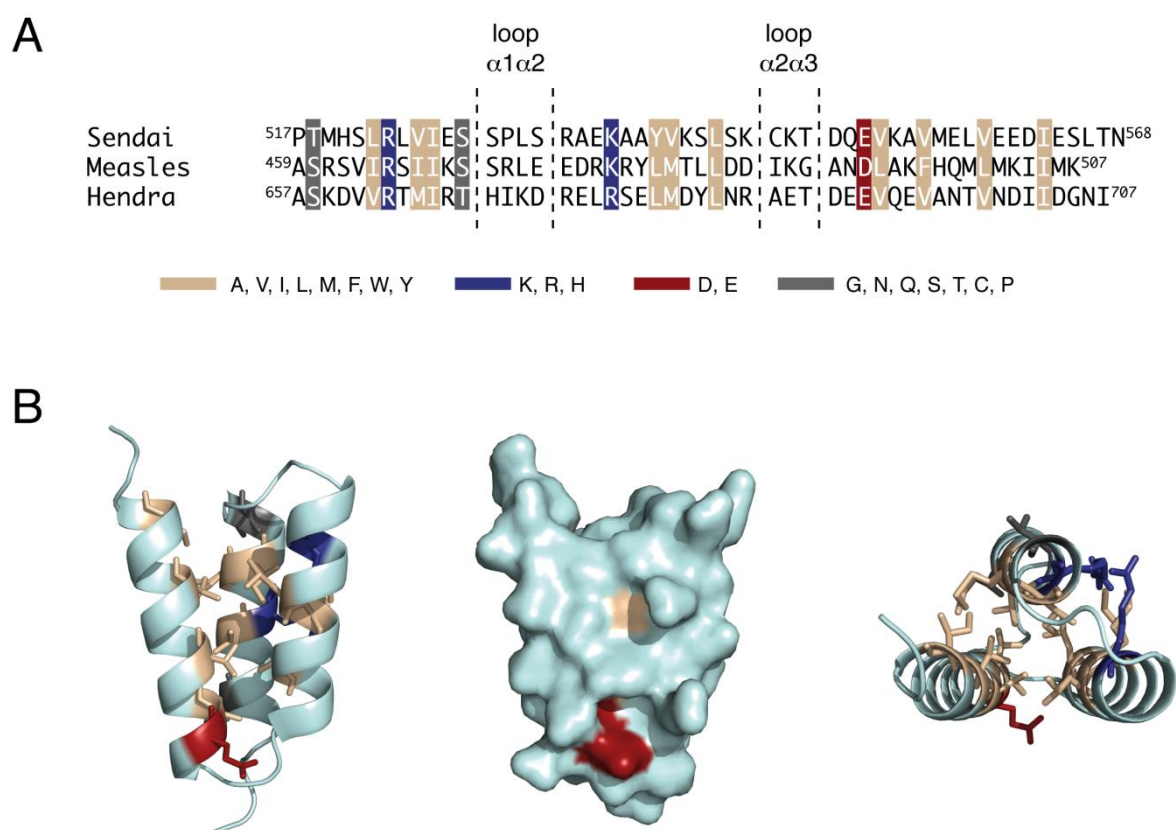


Figure S5: Comparison of the N_{TAIL}-XD interaction surfaces of Sendai, Measles and Hendra viruses. (A) Surface representation of the X domains displaying the location of hydrophobic (beige), negatively charged (red) and positively charged (blue) residues. The orientation of the X domains is the same as shown in Figures 3 and 7. (B) Helical wheel representations of the MoRE of the N_{TAIL} domains with the same color-coding as in (A). The orientation of the wheel of the MoRE of HeV N_{TAIL} is the same as shown in Figure 7. For the sake of clarity, the wheels of SeV and MeV are shown in the same orientation as that of HeV.

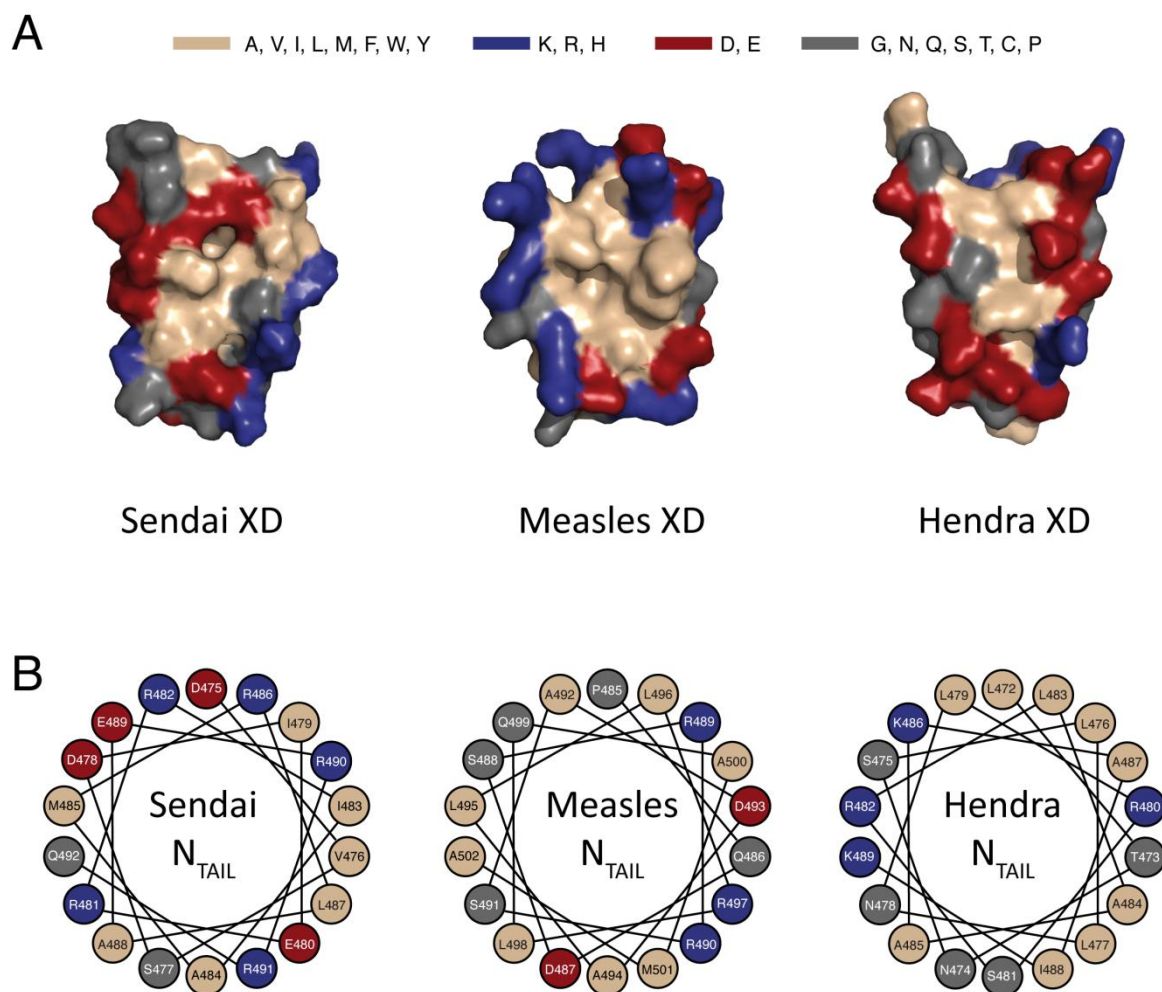


Figure S6: Docking of N_{TAIL} on the surface of XD using experimental chemical shift perturbations as restraints. (A) Observed chemical shift perturbations in XD used as ambiguous restraints in the docking calculations (purple). All residues between 471 and 489 were used as restraints in N_{TAIL}. All restraints were combined as ambiguous distance restraints using standard procedures and used in a torsion-angle molecular dynamics refinement (initial sampling and annealing at 4000 K, followed by slow cooling to 100 K, and refined using Cartesian dynamics and minimization – all calculations were performed using the program CNS-Sculptor). 100 initial structures were created with N_{TAIL} placed randomly on the surface of a sphere of radius 40 Å with the centre of mass of XD at the centre. Typical solutions are shown in (B) and (C), with the hydrophobic surfaces shown in Figure 7 in direct contact, and the basic (N_{TAIL}) and acidic (XD) residues stabilizing the interaction. All constraints and biophysical considerations are in agreement with two directions of N_{TAIL} on the surface of XD (up-down and down-up). (D) Comparison of the best-fitting solution with the X-ray crystallographic coordinates of MeV N_{TAIL}-XD (orange and blue). Note that in this particular conformation the orientation of HeV N_{TAIL} is inverted with respect to the orientation in MeV.

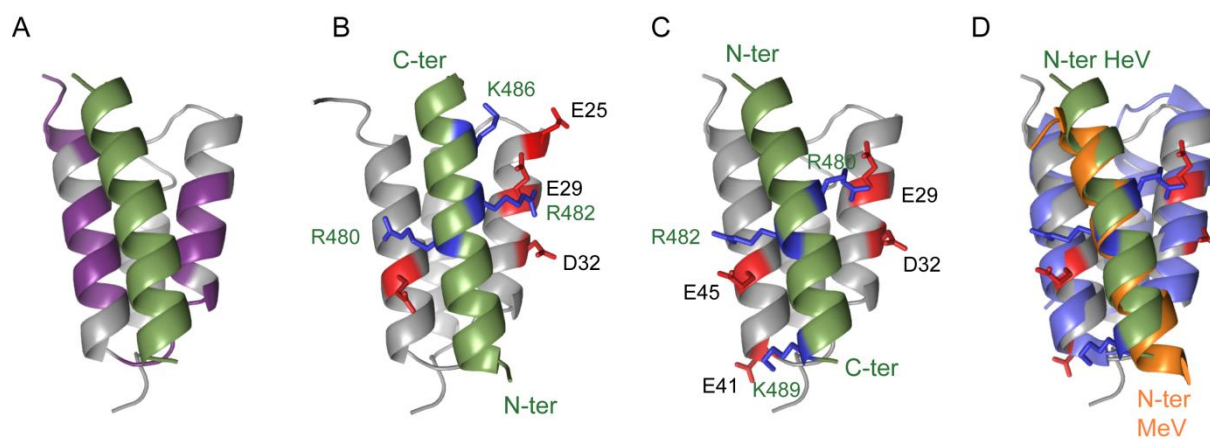


Table S1: Chemical shifts (ppm) of Hendra virus N_{TAIL}. The chemical shifts were obtained at 293 K in 50 mM Bis-Tris, 500 mM NaCl at pH 6.0.

		HN	N	C'	C α	C β
400	Ser	-	-	174.47	58.32	63.88
401	Val	8.30	121.85	176.68	62.54	32.72
402	Gly	8.53	112.59	174.05	45.23	-
403	Arg	8.29	121.03	176.55	56.29	30.92
404	Gln	8.61	121.82	175.67	55.85	29.44
405	Asp	8.42	121.54	176.10	54.37	41.24
406	Asn	8.46	119.28	175.31	53.55	38.78
407	Asn	8.50	118.94	175.67	53.71	38.68
408	Met	8.33	120.49	176.72	56.28	32.39
409	Gln	8.36	120.74	176.27	56.55	29.11
410	Ala	8.28	124.75	178.21	53.03	19.05
411	Arg	8.27	120.15	176.73	56.82	30.79
412	Glu	8.41	121.21	176.66	56.97	30.15
413	Ala	8.27	124.99	178.00	53.01	19.14
414	Lys	8.17	119.94	176.59	56.75	32.91
415	Phe	8.14	120.32	175.76	57.78	39.66
416	Ala	8.23	125.59	177.33	52.42	19.41
417	Ala	8.23	123.39	178.40	52.87	19.11
418	Gly	8.40	108.19	174.80	45.47	-
419	Gly	8.27	108.80	173.99	45.28	-
420	Val	7.95	119.23	176.01	62.28	32.85
421	Leu	8.42	126.30	177.24	55.06	42.28
422	Val	8.24	122.30	176.70	62.50	32.76
423	Gly	8.59	113.15	174.72	45.37	-
424	Gly	8.43	109.12	174.82	45.35	-
425	Gly	8.44	109.11	174.40	45.25	-
426	Glu	8.51	120.73	176.64	56.86	30.14
427	Gln	8.48	120.64	175.58	55.74	29.80
428	Asp	8.47	122.37	175.93	54.43	41.10
429	Ile	8.12	120.74	175.83	61.02	39.26
430	Asp	8.49	124.81	176.18	54.33	41.34
431	Glu	8.41	122.05	176.42	56.63	30.56
432	Glu	8.49	121.96	176.33	56.49	30.40
433	Glu	8.41	122.39	176.24	56.38	30.66
434	Glu	8.52	123.95	174.51	54.38	29.82
435	Pro	-	-	177.02	63.09	32.02
436	Ile	8.31	121.66	176.37	61.37	38.92
437	Glu	8.52	124.59	176.30	56.60	30.15
438	His	8.51	120.74	-	-	-
439	Ser	-	-	175.23	58.85	63.85
440	Gly	8.61	111.28	174.37	45.47	-
441	Arg	8.23	120.83	176.53	56.38	30.77
442	Gln	8.53	121.51	176.09	55.96	29.47
443	Ser	8.41	117.50	174.73	58.46	63.84
444	Val	8.28	121.78	176.37	62.65	32.70
445	Thr	8.16	117.33	174.20	62.10	69.82
446	Phe	8.23	123.30	175.44	58.00	39.76
447	Lys	8.24	123.63	176.23	56.38	33.13
448	Arg	8.36	123.05	176.46	56.54	30.74
449	Glu	8.57	122.33	176.76	56.87	30.02
450	Met	8.43	121.55	176.47	55.69	32.82
451	Ser	8.38	116.94	174.84	58.62	63.79
452	Met	8.47	122.39	176.65	55.85	32.73
453	Ser	8.36	116.73	174.80	58.84	63.89
454	Ser	-	-	174.80	58.58	63.84

455	Leu	8.25	124.01	177.63	55.57	42.19
456	Ala	8.22	124.02	177.68	52.95	19.23
457	Asp	8.20	118.92	176.27	54.52	41.18
458	Ser	8.13	115.57	174.17	58.39	64.04
459	Val	8.16	123.15	174.60	60.04	32.55
460	Pro	-	-	177.21	63.38	32.19
461	Ser	8.54	116.62	174.93	58.64	63.84
462	Ser	8.49	118.10	-	58.78	63.85
463	Ser	-	-	174.54	58.53	63.86
464	Val	8.12	121.37	176.38	62.42	32.77
465	Ser	8.46	119.32	175.18	58.39	64.00
466	Thr	8.31	115.85	174.77	61.90	69.93
467	Ser	8.39	117.94	175.19	58.61	63.89
468	Gly	8.50	111.17	174.80	45.57	-
469	Gly	8.33	108.97	174.47	45.36	-
470	Thr	8.17	114.41	174.67	62.12	70.01
471	Arg	8.48	123.85	176.48	56.50	30.59
472	Leu	8.39	123.52	177.75	55.47	42.20
473	Thr	8.07	114.42	174.65	62.24	69.81
474	Asn	8.51	120.82	175.82	53.90	38.61
475	Ser	8.35	116.53	175.29	59.42	63.62
476	Leu	8.24	123.59	178.10	56.21	41.92
477	Leu	8.03	120.87	177.88	56.20	42.13
478	Asn	8.27	118.66	175.86	53.90	38.57
479	Leu	8.12	122.24	178.14	56.70	42.16
480	Arg	8.32	119.85	177.68	57.66	30.24
481	Ser	8.21	115.92	175.55	59.59	63.41
482	Arg	8.26	123.14	177.45	57.47	30.47
483	Leu	8.20	121.71	178.12	56.03	42.05
484	Ala	8.13	123.98	178.21	53.13	18.93
485	Ala	8.08	122.37	178.46	53.18	18.93
486	Lys	8.10	120.06	176.66	56.90	32.99
487	Ala	8.14	124.32	178.27	52.95	19.04
488	Ile	8.05	120.29	176.87	61.68	38.65
489	Lys	8.32	125.25	176.90	56.83	32.93
490	Glu	8.49	122.01	176.87	56.95	30.25
491	Ser	8.41	116.88	175.17	58.74	63.79
492	Thr	8.23	116.00	174.68	62.12	69.85
493	Ala	8.30	126.34	177.98	52.96	19.18
494	Gln	8.39	119.61	176.36	56.07	29.45
495	Ser	8.43	117.26	174.92	58.56	63.95
496	Ser	8.41	117.60	174.95	58.72	63.83
497	Ser	8.39	117.89	174.70	58.74	63.83
498	Glu	8.33	122.55	176.41	56.78	30.18
499	Arg	8.27	121.45	175.86	56.17	30.85
500	Asn	8.48	120.40	172.37	51.52	38.85
502	Pro	-	-	176.78	63.20	32.04
503	Asn	8.53	118.46	174.90	53.23	38.80
504	Asn	8.40	119.23	174.82	53.14	38.89
505	Arg	8.24	122.23	174.17	54.24	30.07
506	Pro	-	-	177.04	63.29	32.10
507	Gln	8.59	121.08	176.03	55.89	29.53
508	Ala	8.45	125.58	177.54	52.76	19.36
509	Asp	8.41	119.95	176.59	54.32	41.31
510	Ser	8.37	116.92	175.31	58.88	63.87
511	Gly	8.55	110.96	174.08	45.42	-
512	Arg	8.13	120.80	176.52	56.13	30.95
513	Lys	8.55	123.66	176.49	56.38	32.99
514	Asp	8.46	121.45	176.07	54.51	41.16
515	Asp	8.31	120.48	176.27	54.48	41.02
516	Gln	8.29	119.71	175.91	55.59	29.62
517	Glu	8.39	123.59	174.47	54.51	29.62
518	Pro	-	-	176.68	63.04	32.16
519	Lys	8.50	123.27	174.64	54.05	32.43

520	Pro	-	-	176.69	63.08	32.15
521	Ala	8.52	124.79	177.85	52.49	19.29
522	Gln	8.43	119.81	175.71	55.95	29.52
523	Asn	8.51	119.86	174.98	53.28	39.13
524	Asp	8.41	120.99	176.55	54.79	41.02
525	Leu	8.20	121.67	177.40	55.62	42.16
526	Asp	8.24	120.05	175.90	54.60	41.00
527	Phe	7.93	120.08	175.36	57.87	39.68
528	Val	8.01	123.29	175.71	62.36	32.78
529	Arg	8.36	125.82	175.78	56.04	30.94
530	Ala	8.44	125.98	177.24	52.36	19.64
531	Asp	8.48	120.40	175.23	54.37	40.90
532	Val	7.60	123.31	174.65	63.55	33.39

Table S2: Data collection and refinement statistics of HeV XD.

	Crystal
Data collection	
Wavelength (Å)	0.9393
Space group	P12 ₁ 1
Unit cell	
a, b, c (Å)	21.49, 43.28, 51.26
α , β , γ (°)	90.00, 79.78, 90.00
Resolution ^a (Å)	32.85 - 1.65 (1.80 - 1.65)
I/ σ (I) ^a	14.24 (1.89)
Rsym ^a	0.037 (0.546)
Completeness ^a (%)	98.6 (98.7)
Multiplicity ^a	2.7 (2.48)
Total reflections ^b	58428
Refinement	
Resolution ^a (Å)	32.85 - 1.65 (1.72 - 1.65)
R-factor ^a	0.188 (0.306)
R-free ^a	0.224 (0.306)
Number of atoms	1010
macromolecules	973
ligands	2
water	35
Number of protein residues	116
R.m.s. deviations	
bond lengths (Å)	0.012
bond angles (°)	1.12
Ramachandran favored (%)	99
Ramachandran allowed (%)	1
Average B-factor	37.20
macromolecules	36.90
solvent	44.80

^aValues in parentheses are for highest-resolution shell.^bFriedel pairs unmerged

Table S3: Chemical shifts (ppm) of Hendra virus XD. The chemical shifts were obtained at 298 K in 20 mM Bis-Tris, 50 mM Arg/Glu, 150 mM NaCl at pH 6.0.

		HN	N	C'	C α	C β
XXX	Gly	–	–	178.69	43.33	–
XXX	Ala	8.62	123.77	177.47	52.50	19.56
XXX	Met	8.55	120.43	175.84	55.25	33.27
XXX	Val	8.18	121.13	175.81	62.15	33.03
654	Ala	8.35	127.57	177.05	52.23	19.27
655	Asp	8.48	120.97	176.44	53.39	41.73
656	Asp	8.55	122.07	177.72	56.77	41.12
657	Ala	8.38	122.92	180.26	54.85	18.18
658	Ser	8.21	114.88	177.08	60.98	62.87
659	Lys	7.91	120.87	178.36	60.53	32.84
660	Asp	8.33	119.82	178.71	57.96	40.65
661	Val	7.60	119.69	178.78	66.44	32.00
662	Val	7.27	120.70	177.80	65.86	30.97
663	Arg	8.79	121.74	178.65	61.05	30.14
664	Thr	8.23	115.44	176.87	66.85	68.59
665	Met	7.90	124.20	178.47	59.64	33.38
666	Ile	8.44	119.29	177.19	65.84	38.71
667	Arg	8.24	115.82	178.47	59.75	30.62
668	Thr	8.04	108.86	175.73	64.63	70.33
669	His	7.72	114.73	174.75	57.50	31.80
670	Ile	8.02	120.16	175.43	58.98	34.54
671	Lys	8.59	128.48	176.67	57.43	32.76
672	Asp	7.63	120.51	176.22	54.34	42.68
673	Arg	8.87	127.80	178.49	59.93	30.31
674	Glu	8.28	120.69	179.07	59.53	28.82
675	Leu	8.18	123.29	178.59	57.29	41.63
676	Arg	8.53	118.18	177.78	60.92	30.83
677	Ser	7.72	112.54	176.69	62.22	62.88
678	Glu	7.92	122.80	178.92	59.35	29.61
679	Leu	8.50	118.77	179.12	57.89	42.26
680	Met	8.53	117.75	178.59	57.84	31.44
681	Asp	7.93	120.31	178.58	57.70	40.76
682	Tyr	7.98	119.44	179.05	59.41	37.76
683	Leu	8.47	119.58	178.82	57.49	41.98
684	Asn	8.66	118.15	177.24	55.81	37.85
685	Arg	7.69	116.56	176.27	56.96	31.01
686	Ala	7.35	123.68	177.16	53.07	19.19
687	Glu	9.11	121.64	177.11	57.02	32.49
688	Thr	7.85	109.12	175.08	59.42	72.38
689	Asp	9.09	122.08	178.75	57.67	40.25
690	Glu	8.77	120.20	179.27	60.08	29.23
691	Glu	7.76	119.99	179.75	59.17	30.57
692	Val	8.15	121.33	177.63	67.61	31.55
693	Gln	8.31	118.75	178.01	58.00	27.90
694	Glu	7.85	119.41	179.72	59.56	29.10
695	Val	7.57	120.57	177.90	66.66	30.94
696	Ala	8.34	122.39	179.59	55.79	17.83
697	Asn	8.70	116.77	177.54	55.88	37.66
698	Thr	8.00	119.39	176.15	67.16	68.74
699	Val	8.43	121.04	177.36	67.28	31.42
700	Asn	8.38	117.81	177.16	56.59	38.29

701	Asp	8.10	119.82	178.82	57.50	40.20
702	Ile	7.94	121.66	179.46	64.97	38.38
703	Ile	8.27	121.10	178.16	64.95	38.22
704	Asp	8.24	119.01	177.04	55.48	40.79
705	Gly	7.86	107.45	174.60	45.86	-
706	Asn	8.29	119.42	174.19	53.35	38.97
707	Ile	7.47	123.67	171.81	62.76	39.91

Table S4: Residual dipolar couplings (Hz) of Hendra virus XD. The couplings were obtained in filamentous phages (11.5 mg/mL) at 298 K in 20 mM Bis-Tris, 50 mM Arg/Glu, 150 mM NaCl at pH 6.0.

				RDC	Error
654	N	654	HN	2.16	1.00
655	N	655	HN	1.32	1.00
656	N	656	HN	9.71	1.00
657	N	657	HN	6.43	1.00
658	N	658	HN	7.13	1.00
659	N	659	HN	13.14	1.00
660	N	660	HN	12.30	1.00
661	N	661	HN	-1.33	1.00
662	N	662	HN	9.99	1.00
663	N	663	HN	12.65	1.00
664	N	664	HN	5.03	1.00
665	N	665	HN	1.18	1.00
666	N	666	HN	12.09	1.00
667	N	667	HN	11.81	1.00
668	N	668	HN	-3.71	1.00
669	N	669	HN	5.24	1.00
670	N	670	HN	11.88	1.00
671	N	671	HN	-2.31	1.00
672	N	672	HN	-13.99	1.00
673	N	673	HN	-1.75	1.00
674	N	674	HN	13.21	1.00
675	N	675	HN	5.73	1.00
676	N	676	HN	-1.54	1.00
677	N	677	HN	8.04	1.00
678	N	678	HN	12.02	1.00
679	N	679	HN	0.97	1.00
680	N	680	HN	-0.21	1.00
681	N	681	HN	11.25	1.00
682	N	682	HN	9.64	1.00
683	N	683	HN	-0.77	1.00
684	N	684	HN	6.01	1.00
685	N	685	HN	14.47	1.00
686	N	686	HN	-6.58	1.00
687	N	687	HN	-7.27	1.00
688	N	688	HN	10.76	1.00
689	N	689	HN	6.99	1.00
690	N	690	HN	6.29	1.00
691	N	691	HN	14.96	1.00
692	N	692	HN	8.80	1.00
693	N	693	HN	4.33	1.00
694	N	694	HN	12.44	1.00
695	N	695	HN	14.68	1.00
696	N	696	HN	9.64	1.00
697	N	697	HN	6.11	1.00
698	N	698	HN	14.61	1.00
699	N	699	HN	14.19	1.00
700	N	700	HN	7.13	1.00
701	N	701	HN	10.83	1.00
702	N	702	HN	15.94	1.00
703	N	703	HN	10.13	1.00
704	N	704	HN	5.59	1.00
705	N	705	HN	7.48	1.00

706	N	706	HN	6.92	1.00
707	N	707	HN	-3.78	1.00

654	CA	654	HA	-1.53	2.40
655	CA	655	HA	-17.14	2.40
656	CA	656	HA	-10.97	2.40
657	CA	657	HA	32.27	2.40
658	CA	658	HA	-15.08	2.40
659	CA	659	HA	-0.99	2.40
660	CA	660	HA	2.06	2.40
661	CA	661	HA	26.88	2.40
662	CA	662	HA	-24.39	2.40
663	CA	663	HA	-19.37	2.40
664	CA	664	HA	37.68	2.40
665	CA	665	HA	-10.06	2.40
666	CA	666	HA	-10.87	2.40
667	CA	667	HA	-15.06	2.40
668	CA	668	HA	45.81	2.40
669	CA	669	HA	-5.07	2.40
670	CA	670	HA	-12.63	2.40
671	CA	671	HA	2.65	2.40
672	CA	672	HA	18.28	2.40
673	CA	673	HA	-21.31	2.40
674	CA	674	HA	-19.59	2.40
675	CA	675	HA	16.84	2.40
676	CA	676	HA	12.81	2.40
677	CA	677	HA	-17.29	2.40
678	CA	678	HA	-22.92	2.40
679	CA	679	HA	44.36	2.40
680	CA	680	HA	-11.49	2.40
681	CA	681	HA	-14.23	2.40
682	CA	682	HA	-2.99	2.40
683	CA	683	HA	37.34	2.40
684	CA	684	HA	-18.20	2.40
685	CA	685	HA	-23.15	2.40
686	CA	686	HA	27.01	2.40
687	CA	687	HA	32.32	2.40
688	CA	688	HA	-26.48	2.40
689	CA	689	HA	43.30	2.40
690	CA	690	HA	-20.24	2.40
691	CA	691	HA	-10.52	2.40
692	CA	692	HA	13.41	2.40
693	CA	693	HA	15.48	2.40
694	CA	694	HA	-3.47	2.40
695	CA	695	HA	-23.10	2.40
696	CA	696	HA	42.49	2.40
697	CA	697	HA	-17.00	2.40
698	CA	698	HA	5.19	2.40
699	CA	699	HA	-4.68	2.40
700	CA	700	HA	28.16	2.40
701	CA	701	HA	-9.62	2.40
702	CA	702	HA	-18.16	2.40
703	CA	703	HA	24.72	2.40
704	CA	704	HA	1.88	2.40
706	CA	706	HA	-13.49	2.40

654	C	654	CA	1.20	0.25
655	C	655	CA	1.10	0.25

656	C	656	CA	-1.60	0.25
657	C	657	CA	-1.20	0.25
658	C	658	CA	-0.80	0.25
659	C	659	CA	2.80	0.25
660	C	660	CA	-2.90	0.25
661	C	661	CA	-1.90	0.25
662	C	662	CA	2.50	0.25
663	C	663	CA	0.30	0.25
664	C	664	CA	-2.50	0.25
665	C	665	CA	-1.70	0.25
666	C	666	CA	3.90	0.25
667	C	667	CA	-1.90	0.25
668	C	668	CA	-3.20	0.25
669	C	669	CA	-1.80	0.25
670	C	670	CA	3.50	0.25
671	C	671	CA	-1.50	0.25
672	C	672	CA	-2.40	0.25
673	C	673	CA	1.10	0.25
674	C	674	CA	3.00	0.25
675	C	675	CA	-2.80	0.25
676	C	676	CA	-2.30	0.25
677	C	677	CA	3.90	0.25
678	C	678	CA	-1.10	0.25
679	C	679	CA	-2.40	0.25
680	C	680	CA	-1.20	0.25
681	C	681	CA	4.30	0.25
682	C	682	CA	-2.10	0.25
683	C	683	CA	-2.90	0.25
684	C	684	CA	1.60	0.25
685	C	685	CA	2.40	0.25
686	C	686	CA	-3.20	0.25
687	C	687	CA	0.50	0.25
688	C	688	CA	-2.30	0.25
689	C	689	CA	-1.70	0.25
690	C	690	CA	0.50	0.25
691	C	691	CA	2.60	0.25
692	C	692	CA	-1.70	0.25
693	C	693	CA	-2.30	0.25
694	C	694	CA	2.70	0.25
695	C	695	CA	-1.50	0.25
696	C	696	CA	-1.40	0.25
697	C	697	CA	-1.60	0.25
698	C	698	CA	2.70	0.25
699	C	699	CA	-2.00	0.25
700	C	700	CA	-1.90	0.25
701	C	701	CA	1.40	0.25
702	C	702	CA	0.30	0.25
703	C	703	CA	-0.80	0.25
704	C	704	CA	-2.00	0.25
705	C	705	CA	2.60	0.25
706	C	706	CA	-0.90	0.25

654	C	655	HN	-1.60	0.40
655	C	656	HN	-0.30	0.40
656	C	657	HN	-2.10	0.40
657	C	658	HN	3.60	0.40
658	C	659	HN	-3.30	0.40
659	C	660	HN	-3.80	0.40

660	C	661	HN	4.90	0.40
661	C	662	HN	-0.10	0.40
662	C	663	HN	-4.80	0.40
663	C	664	HN	-3.60	0.40
664	C	665	HN	6.80	0.40
665	C	666	HN	-4.60	0.40
666	C	667	HN	-3.90	0.40
667	C	668	HN	2.20	0.40
668	C	669	HN	4.80	0.40
669	C	670	HN	-4.50	0.40
670	C	671	HN	-5.00	0.40
671	C	672	HN	6.60	0.40
672	C	673	HN	5.10	0.40
673	C	674	HN	-4.20	0.40
674	C	675	HN	-4.80	0.40
675	C	676	HN	5.90	0.40
676	C	677	HN	-0.60	0.40
677	C	678	HN	-4.50	0.40
678	C	679	HN	-2.50	0.40
679	C	680	HN	6.80	0.40
680	C	681	HN	-3.80	0.40
681	C	682	HN	-4.10	0.40
682	C	683	HN	3.00	0.40
683	C	684	HN	2.30	0.40
684	C	685	HN	-3.00	0.40
685	C	686	HN	-2.30	0.40
686	C	687	HN	6.60	0.40
687	C	688	HN	-1.70	0.40
688	C	689	HN	-1.50	0.40
689	C	690	HN	3.70	0.40
690	C	691	HN	-3.70	0.40
691	C	692	HN	-4.60	0.40
692	C	693	HN	4.50	0.40
693	C	694	HN	-2.40	0.40
694	C	695	HN	-2.40	0.40
695	C	696	HN	-3.50	0.40
696	C	697	HN	4.90	0.40
697	C	698	HN	-3.00	0.40
698	C	699	HN	-3.50	0.40
699	C	700	HN	1.70	0.40
700	C	701	HN	-0.10	0.40
701	C	702	HN	-1.60	0.40
702	C	703	HN	-5.10	0.40
703	C	704	HN	4.80	0.40
704	C	705	HN	-2.00	0.40
705	C	706	HN	2.90	0.40
706	C	707	HN	1.30	0.40

Références Bibliographiques

- (2011). "WHO Fact Sheet N° 286 - Measles."
- Afonine, P. V., R. W. Grosse-Kunstleve, et al. (2012). "Towards automated crystallographic structure refinement with phenix.refine." Acta Crystallogr D Biol Crystallogr **68**: 352-367.
- Ahmed, M. and D. S. Lyles (1998). "Effect of vesicular stomatitis virus matrix protein on transcription directed by host RNA polymerases I, II, and III." J Virol **72**(10): 8413-8419.
- Albertini, A. A., E. Baquero, et al. (2012). "Molecular and cellular aspects of rhabdovirus entry." Viruses **4**(1): 117-139.
- Albertini, A. A., C. R. Clapier, et al. (2006). "Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings." J Struct Biol **158**: 129-133.
- Albertini, A. A., C. R. Clapier, et al. (2007). "Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings." J Struct Biol **158**(1): 129-133.
- Albertini, A. A., C. Merigoux, et al. (2012). "Characterization of monomeric intermediates during VSV glycoprotein structural transition." PLoS Pathog **8**(2): e1002556.
- Albertini, A. A., G. Schoehn, et al. (2008). "Structural aspects of rabies virus replication." Cell Mol Life Sci **65**(2): 282-294.
- Albertini, A. A., A. K. Wernimont, et al. (2006). "Crystal structure of the rabies virus nucleoprotein-RNA complex." Science **313**(5785): 360-363.
- Albertini, A. A. V., G. Schoehn, et al. (2005). "Structures impliquées dans la réplication et la transcription des virus à ARN non segmentés de sens négatif." Virologie **9**(2): 83-92.
- Amal, N. M., M. S. Lye, et al. (2000). "Risk factors for Nipah virus transmission, Port Dickson, Negeri Sembilan, Malaysia: results from a hospital-based case-control study." Southeast Asian J Trop Med Public Health **31**(2): 301-306.
- Andrejeva, J., K. S. Childs, et al. (2004). "The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-beta promoter." Proc Natl Acad Sci U S A **101**(49): 17264-17269.
- Arnheiter, H., N. L. Davis, et al. (1985). "Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis." Cell **41**(1): 259-267.
- Assenberg, R., O. Delmas, et al. (2010). "Genomics and structure/function studies of Rhabdoviridae proteins involved in replication and transcription." Antiviral Res **87**(2): 149-161.
- Assenberg, R., O. Delmas, et al. (2010). "Structure of the nucleoprotein binding domain of Mokola virus phosphoprotein." Journal of virology **84**(2): 1089-1096.
- Baker, N. A., D. Sept, et al. (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proc Natl Acad Sci U S A **98**(18): 10037-10041.
- Ball, L. A. (2007). Virus Replication Strategies. Fields Virology, 5th ed. D. Knipe and P. Howley. Philadelphia, Lippincott Williams & Wilkins. **1**: 120-140.
- Bankamp, B., S. M. Horikami, et al. (1996). "Domains of the measles virus N protein required for binding to P protein and self-assembly." Virology **216**(1): 272-277.
- Barge, A., Y. Gaudin, et al. (1993). "Vesicular stomatitis virus M protein may be inside the ribonucleocapsid coil." Journal of virology **67**(12): 7246-7253.
- Barik, S. and A. K. Banerjee (1992). "Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P." Proc Natl Acad Sci U S A **89**(14): 6570-6574.

- Barr, J. N. and G. W. Wertz (2001). "Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination." Journal of virology **75**(15): 6901-6913.
- Barr, J. N., S. P. Whelan, et al. (2002). "Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus." Biochim Biophys Acta **1577**(2): 337-353.
- Battye, T. G., L. Kontogiannis, et al. (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 271-281.
- Belle, V., S. Rouger, et al. (2008). "Mapping alpha-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy." Proteins: Structure, Function and Bioinformatics **73**(4): 973-988.
- Bernado, P. (2010). "Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering." Eur Biophys J **39**(5): 769-780.
- Bernado, P. and M. Blackledge (2009). "A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering." Biophys J **97**(10): 2839-2845.
- Bernado, P., L. Blanchard, et al. (2005). "A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering." Proc Natl Acad Sci U S A **102**(47): 17002-17007.
- Bernado, P., E. Mylonas, et al. (2007). "Structural characterization of flexible proteins using small-angle X-ray scattering." J Am Chem Soc **129**(17): 5656-5664.
- Bernado, P. and D. I. Svergun (2012). "Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering." Mol Biosyst **8**(1): 151-167.
- Beutler, B. (2004). "Inferences, questions and possibilities in Toll-like receptor signalling." Nature **430**(6996): 257-263.
- Bhatt, P. N. and F. M. Rodrigues (1967). "Chandipura: a new Arbovirus isolated in India from patients with febrile illness." Indian J Med Res **55**(12): 1295-1305.
- Bhella, D., A. Ralph, et al. (2002). "Significant differences in nucleocapsid morphology within the Paramyxoviridae." J Gen Virol **83**(Pt 8): 1831-1839.
- Bhella, D., A. Ralph, et al. (2004). "Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction." Journal of molecular biology **340**(2): 319-331.
- Bitko, V., A. Musiyenko, et al. (2005). "Inhibition of respiratory viruses by nasally administered siRNA." Nat Med **11**(1): 50-55.
- Black, B. L. and D. S. Lyles (1992). "Vesicular stomatitis virus matrix protein inhibits host cell-directed transcription of target genes in vivo." J Virol **66**(7): 4058-4064.
- Black, B. L., R. B. Rhodes, et al. (1993). "The role of vesicular stomatitis virus matrix protein in inhibition of host-directed gene expression is genetically separable from its function in virus assembly." J Virol **67**(8): 4814-4821.
- Blanchard, L., N. Tarbouriech, et al. (2004). "Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution." Virology **319**(2): 201-211.
- Blocquel, D., J. Habchi, et al. (2012). "Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies." Mol Biosyst **8**(1): 392-410.

- Bloor, S., J. Maelfait, et al. (2010). "Endoplasmic reticulum chaperone gp96 is essential for infection with vesicular stomatitis virus." Proc Natl Acad Sci U S A **107**(15): 6970-6975.
- Blumberg, B. M., M. Leppert, et al. (1981). "Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication." Cell **23**(3): 837-845.
- Bonaparte, M. I., A. S. Dimitrov, et al. (2005). "Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus." Proc Natl Acad Sci U S A **102**(30): 10652-10657.
- Bos, L. (1999). "Beijerinck's work on tobacco mosaic virus: historical context and legacy." Philos Trans R Soc Lond B Biol Sci **354**(1383): 675-685.
- Bossart, K. N., B. Rockx, et al. (2012). "A Hendra virus G glycoprotein subunit vaccine protects African green monkeys from Nipah virus challenge." Sci Transl Med **4**(146): 146ra107.
- Bourhis, J., K. Johansson, et al. (2004). "The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner." Virus research **99**: 157-167.
- Bourhis, J. M., V. Receveur-Brechot, et al. (2005). "The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded." Protein Sci **14**(8): 1975-1992.
- Bowden, T. A., A. R. Aricescu, et al. (2008). "Structural basis of Nipah and Hendra virus attachment to their cell-surface receptor ephrin-B2." Nat Struct Mol Biol **15**(6): 567-572.
- Bowman, M. C., S. Smallwood, et al. (1999). "Dissection of individual functions of the Sendai virus phosphoprotein in transcription." J. Virol. **73**(8): 6474-6483.
- Boze, H., T. Marlin, et al. (2010). "Proline-rich salivary proteins have extended conformations." Biophys J **99**(2): 656-665.
- Broughan, J. H. and W. H. Wunner (1995). "Characterization of protein involvement in rabies virus binding to BHK-21 cells." Arch Virol **140**(1): 75-93.
- Brunger, A. T. (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." Nature **355**(6359): 472-475.
- Brunger, A. T. and P. D. Adams (2002). "Molecular dynamics applied to X-ray structure refinement." Acc Chem Res **35**(6): 404-412.
- Brunger, A. T., J. Kuriyan, et al. (1987). "Crystallographic R factor refinement by molecular dynamics." Science **235**(4787): 458-460.
- Brzozka, K., S. Finke, et al. (2005). "Identification of the rabies virus alpha/beta interferon antagonist: phosphoprotein P interferes with phosphorylation of interferon regulatory factor 3." J Virol **79**(12): 7673-7681.
- Buchholz, C. J., D. Spehner, et al. (1993). "The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly." Journal of virology **67**(10): 5803-5812.
- Cadd, T., D. Garcin, et al. (1996). "The Sendai paramyxovirus accessory C proteins inhibit viral genome amplification in a promoter-specific fashion." J Virol **70**(8): 5067-5074.
- Calain, P. and L. Roux (1993). "The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA." J Virol **67**(8): 4822-4830.
- Canter, D. M. and J. Perrault (1996). "Stabilization of vesicular stomatitis virus L polymerase protein by P protein binding: a small deletion in the C-terminal domain of L abrogates binding." Virology **219**(2): 376-386.

- Carlos, T. S., D. F. Young, et al. (2009). "Parainfluenza virus 5 genomes are located in viral cytoplasmic bodies whilst the virus dismantles the interferon-induced antiviral state of cells." The Journal of general virology **90**(Pt 9): 2147-2156.
- Castagne, N., A. Barbier, et al. (2004). "Biochemical characterization of the respiratory syncytial virus P-P and P-N protein complexes and localization of the P protein oligomerization domain." The Journal of general virology **85**(Pt 6): 1643-1653.
- Castel, G., M. Chtouei, et al. (2009). "Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity." J Virol **83**(20): 10808-10820.
- Cevik, B., D. E. Holmes, et al. (2004). "The phosphoprotein (P) and L binding sites reside in the N-terminus of the L subunit of the measles virus RNA polymerase." Virology **327**(2): 297-306.
- Chattopadhyay, D. and A. K. Banerjee (1987). "Two separate domains within vesicular stomatitis virus phosphoprotein support transcription when added in trans." Proc Natl Acad Sci U S A **84**(24): 8932-8936.
- Chen, L., S. Zhang, et al. (2013). "N-terminal phosphorylation of phosphoprotein of vesicular stomatitis virus is required for preventing nucleoprotein from binding to cellular RNAs and for functional template formation." J Virol **87**(6): 3177-3186.
- Chen, M., J. C. Cortay, et al. (2003). "Measles virus protein interactions in yeast: new findings and caveats." Virus Res **98**(2): 123-129.
- Chen, M., T. Ogino, et al. (2007). "Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA." J Virol **81**(24): 13478-13485.
- Chen, S., J. A. Short, et al. (2010). "Heterocellular induction of interferon by negative-sense RNA viruses." Virology **407**(2): 247-255.
- Chen, X., U. Vinkemeier, et al. (1998). "Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA." Cell **93**(5): 827-839.
- Chenik, M., K. Chebli, et al. (1995). "Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism." J Virol **69**(2): 707-712.
- Chenik, M., M. Schnell, et al. (1998). "Mapping the interacting domains between the rabies virus polymerase and phosphoprotein." J Virol **72**(3): 1925-1930.
- Childs, K., R. Randall, et al. (2012). "Paramyxovirus V proteins interact with the RNA Helicase LGP2 to inhibit RIG-I-dependent interferon induction." J Virol **86**(7): 3411-3421.
- Childs, K. S., J. Andrejeva, et al. (2009). "Mechanism of mda-5 Inhibition by paramyxovirus V proteins." J Virol **83**(3): 1465-1473.
- Chua, K. B., K. J. Goh, et al. (1999). "Fatal encephalitis due to Nipah virus among pig-farmers in Malaysia." Lancet **354**(9186): 1257-1259.
- Ciancanelli, M. J. and C. F. Basler (2006). "Mutation of YMYL in the Nipah virus matrix protein abrogates budding and alters subcellular localization." J Virol **80**(24): 12070-12078.
- Ciancanelli, M. J., V. A. Volchkova, et al. (2009). "Nipah virus sequesters inactive STAT1 in the nucleus via a P gene-encoded mechanism." J Virol **83**(16): 7828-7841.
- Clore, G. M. (2011). "Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation." Protein Sci **20**(2): 229-246.
- Collins, P. L. and J. E. J. Crowe (2007). Respiratory syncytial virus and metapneumovirus. Fields Virology, 5th ed. D. Knipe and P. Howley. Philadelphia, Lippincott Williams & Wilkins. **2**: 1601-1646.

- Communie, G., T. Crepin, et al. (2013). "Structure of the tetramerization domain of measles virus phosphoprotein." *J Virol* **87**(12): 7166-7169.
- Conzelmann, K. K. (2005). "Transcriptional activation of alpha/beta interferon genes: interference by nonsegmented negative-strand RNA viruses." *J Virol* **79**(9): 5241-5248.
- Cowtan, K. (1999). "Error estimation and bias correction in phase-improvement calculations." *Acta Crystallogr D Biol Crystallogr* **55**(Pt 9): 1555-1567.
- Cowtan, K. (2006). "The Buccaneer software for automated model building. 1. Tracing protein chains." *Acta Crystallogr D Biol Crystallogr* **62**(Pt 9): 1002-1011.
- Cox, R., T. J. Green, et al. (2013). "Structural and functional characterization of the mumps virus phosphoprotein." *J Virol* **87**(13): 7558-7568.
- Curran, J. (1998). "A role for the Sendai virus P protein trimer in RNA synthesis." *J Virol* **72**(5): 4274-4280.
- Curran, J., R. Boeck, et al. (1995). "Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils." *Virology* **214**(1): 139-149.
- Curran, J., H. Homann, et al. (1993). "The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation." *J. Virol.* **67**(7): 4358-4364.
- Curran, J. and D. Kolakofsky (1999). "Replication of paramyxoviruses." *Adv. Virus Res.* **54**: 403-422.
- Curran, J. and D. Kolakofsky (2008). "Nonsegmented negative-strand RNA virus RNA synthesis in vivo." *Virology* **371**(2): 227-230.
- Curran, J., J. B. Marq, et al. (1995). "An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication." *J Virol* **69**(2): 849-855.
- Curran, J., T. Pelet, et al. (1994). "An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation." *Virology* **202**(2): 875-884.
- Das, S. C. and A. K. Pattnaik (2004). "Phosphorylation of vesicular stomatitis virus phosphoprotein P is indispensable for virus growth." *J Virol* **78**(12): 6420-6430.
- Das, T., A. K. Pattnaik, et al. (1997). "Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA." *Virology* **238**(1): 103-114.
- Dauter, Z., M. Dauter, et al. (2000). "Novel approach to phasing proteins: derivatization by short cryo-soaking with halides." *Acta Crystallogr D Biol Crystallogr* **56**(Pt 2): 232-237.
- de Weerd, N. A., S. A. Samarajiwa, et al. (2007). "Type I interferon receptors: biochemistry and biological functions." *J Biol Chem* **282**(28): 20053-20057.
- Delmas, O., R. Assenberg, et al. (2010). "The structure of the nucleoprotein binding domain of lyssavirus phosphoprotein reveals a structural relationship between the N-RNA binding domains of Rhabdoviridae and Paramyxoviridae." *RNA Biol* **7**(3): 322-327.
- Desfosses, A., G. Goret, et al. (2011). "Nucleoprotein-RNA orientation in the measles virus nucleocapsid by three-dimensional electron microscopy." *Journal of virology* **85**(3): 1391-1395.
- Desfosses, A., E. A. Ribeiro, Jr., et al. (2013). "Self-organization of the vesicular stomatitis virus nucleocapsid into a bullet shape." *Nat Commun* **4**: 1429.
- Dillon, P. J. and K. C. Gupta (1988). "Early steps in the assembly of vesicular stomatitis virus nucleocapsids in infected cells." *Journal of virology* **62**(5): 1582-1589.

- Ding, H., T. J. Green, et al. (2006). "Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus." *J Virol* **80**(6): 2808-2814.
- Dodson, E. (2003). "Is it jolly SAD?" *Acta Crystallogr D Biol Crystallogr* **59**(Pt 11): 1958-1965.
- Dolinsky, T. J., J. E. Nielsen, et al. (2004). "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations." *Nucleic Acids Res* **32**(Web Server issue): W665-667.
- Dyson, H. J. and P. E. Wright (2002). "Coupling of folding and binding for unstructured proteins." *Curr Opin Struct Biol* **12**(1): 54-60.
- Eaton, B. T., C. C. Broder, et al. (2006). "Hendra and Nipah viruses: different and dangerous." *Nat Rev Microbiol* **4**(1): 23-35.
- Eaton, B. T., J. S. Mackenzie, et al. (2007). Henipaviruses. *Fields Virology 5th Edition*. D. M. Knipe and P. M. Howley. Philadelphia, Lippincott Williams & Wilkins. **2**: 1587-1600.
- Egelman, E. H., S. S. Wu, et al. (1989). "The Sendai virus nucleocapsid exists in at least four different helical states." *Journal of virology* **63**(5): 2233-2243.
- El Omari, K., B. Dhaliwal, et al. (2011). "Structures of respiratory syncytial virus nucleocapsid protein from two crystal forms: details of potential packing interactions in the native helical form." *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**(Pt 10): 1179-1183.
- Emerson, S. U. and M. Schubert (1987). "Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus." *Proc Natl Acad Sci U S A* **84**(16): 5655-5659.
- Emerson, S. U. and Y. Yu (1975). "Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus." *J Virol* **15**(6): 1348-1356.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132.
- Epstein, J. H., H. E. Field, et al. (2006). "Nipah virus: impact, origins, and causes of emergence." *Curr Infect Dis Rep* **8**(1): 59-65.
- Errington, W. and P. T. Emmerson (1997). "Assembly of recombinant Newcastle disease virus nucleocapsid protein into nucleocapsid-like structures is inhibited by the phosphoprotein." *J Gen Virol* **78** (Pt 9): 2335-2339.
- Evans, P. R. (2011). "An introduction to data reduction: space-group determination, scaling and intensity statistics." *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 282-292.
- Faria, P. A., P. Chakraborty, et al. (2005). "VSV disrupts the Rae1/mrnp41 mRNA nuclear export pathway." *Mol Cell* **17**(1): 93-102.
- Fernandes, M. V., T. J. Wiktor, et al. (1963). "Mechanism of the Cytopathic Effect of Rabies Virus in Tissue Culture." *Virology* **21**: 128-131.
- Ferraro, E., D. Peluso, et al. (2007). "SH3-Hunter: discovery of SH3 domain interaction sites in proteins." *Nucleic Acids Res* **35**(Web Server issue): W451-454.
- Finke, S. and K. K. Conzelmann (2005). "Recombinant rhabdoviruses: vectors for vaccine development and gene therapy." *Curr Top Microbiol Immunol* **292**: 165-200.
- Finke, S., R. Mueller-Waldeck, et al. (2003). "Rabies virus matrix protein regulates the balance of virus transcription and replication." *J Gen Virol* **84**(Pt 6): 1613-1621.
- Fitzkee, N. C. and G. D. Rose (2004). "Reassessing random-coil statistics in unfolded proteins." *Proc Natl Acad Sci U S A* **101**(34): 12497-12502.
- Flory, P. J. (1960). "Conformations of macromolecules." *Brookhaven Symp Biol* **13**: 89-103.
- Galloux, M., B. Tarus, et al. (2012). "Characterization of a Viral Phosphoprotein Binding Site on the Surface of the Respiratory Syncytial Nucleoprotein." *J Virol* **86**: 8375-8387.

- Galloux, M., B. Tarus, et al. (2012). "Characterization of a viral phosphoprotein binding site on the surface of the respiratory syncytial nucleoprotein." J Virol **86**(16): 8375-8387.
- Gans, P., O. Hamelin, et al. (2010). "Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins." Angew Chem Int Ed Engl **49**(11): 1958-1962.
- Gao, Y. and J. Lenard (1995). "Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II." Embo J **14**(6): 1240-1247.
- Garcia, J., B. Garcia-Barreno, et al. (1993). "Cytoplasmic inclusions of respiratory syncytial virus-infected cells: formation of inclusion bodies in transfected cells that coexpress the nucleoprotein, the phosphoprotein, and the 22K protein." Virology **195**(1): 243-247.
- Garner, E., P. Romero, et al. (1999). "Predicting Binding Regions within Disordered Proteins." Genome Inform. Ser. Workshop Genome Inform. **10**: 41-50.
- Ge, P., J. Tsao, et al. (2010). "Cryo-EM model of the bullet-shaped vesicular stomatitis virus." Science **327**(5966): 689-693.
- Gely, S., D. F. Lowry, et al. (2010). "Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein." J Mol Recognit **23**(5): 435-447.
- Gérard, F. C. A., E. Ribeiro, et al. (2007). "Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution." Biochemistry **46**: 10328-10338.
- Gérard, F. C. A., E. A. Ribeiro, et al. (2009). "Modular organization of rabies virus phosphoprotein." J. Mol. Biol. **388**: 978-996.
- Glatter, O. (1982). Data treatment. Small Angle X-ray Scattering. O. Glatter and O. Kratky. London, UK, Academic Press: 119-165.
- Glatter, O. and O. Kratky (1982). Small Angle X-ray Scattering. London, UK, Academic Press.
- Goh, K. J., C. T. Tan, et al. (2000). "Clinical features of Nipah virus encephalitis among pig farmers in Malaysia." N Engl J Med **342**(17): 1229-1235.
- Goldsmith, C. S., T. Whistler, et al. (2003). "Elucidation of Nipah virus morphogenesis and replication using ultrastructural and molecular approaches." Virus Res **92**(1): 89-98.
- Golovanov, A. P., G. M. Hautbergue, et al. (2004). "A simple method for improving protein solubility and long-term stability." J Am Chem Soc **126**(29): 8933-8939.
- Goodbourn, S., L. Didcock, et al. (2000). "Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures." J Gen Virol **81**(Pt 10): 2341-2364.
- Goodin, M. M., R. Chakrabarty, et al. (2007). "Membrane and protein dynamics in live plant nuclei infected with Sonchus yellow net virus, a plant-adapted rhabdovirus." The Journal of general virology **88**(Pt 6): 1810-1820.
- Gotoh, B., T. Komatsu, et al. (2002). "Paramyxovirus strategies for evading the interferon response." Rev Med Virol **12**(6): 337-357.
- Gouet, P., E. Courcelle, et al. (1999). "ESPrInt: analysis of multiple sequence alignments in PostScript." Bioinformatics **15**(4): 305-308.
- Green, T. J. and M. Luo (2006). "Resolution improvement of X-ray diffraction data of crystals of a vesicular stomatitis virus nucleocapsid protein oligomer complexed with RNA." Acta Crystallogr D Biol Crystallogr **62**(Pt 5): 498-504.
- Green, T. J. and M. Luo (2009). "Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P." Proc Natl Acad Sci U S A **106**(28): 11713-11718.

- Green, T. J., S. Macpherson, et al. (2000). "Study of the assembly of vesicular stomatitis virus N protein: role of the P protein." Journal of virology **74**(20): 9515-9524.
- Green, T. J., M. Rowse, et al. (2011). "Access to RNA encapsidated in the nucleocapsid of vesicular stomatitis virus." Journal of virology **85**(6): 2714-2722.
- Green, T. J., X. Zhang, et al. (2006). "Structure of the vesicular stomatitis virus nucleoprotein-RNA complex." Science **313**(5785): 357-360.
- Gubbay, O., J. Curran, et al. (2001). "Sendai virus genome synthesis and assembly are coupled: a possible mechanism to promote viral RNA polymerase processivity." The Journal of general virology **82**(Pt 12): 2895-2903.
- Gupta, A. K. and A. K. Banerjee (1997). "Expression and purification of vesicular stomatitis virus N-P complex from Escherichia coli: role in genome RNA transcription and replication in vitro." J Virol **71**(6): 4264-4271.
- Gupta, A. K., D. Blondel, et al. (2000). "The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C." J Virol **74**(1): 91-98.
- Habchi, J., S. Blangy, et al. (2011). "Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipavirus." J Biol Chem **286**(15): 13583-13602.
- Habchi, J., L. Mamelli, et al. (2010). "Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment." PLoS One **5**(7): e11684.
- Habchi, J., L. Mamelli, et al. (2012). Structural disorder within the nucleoprotein and phosphoprotein from measles, Nipah and Hendra viruses. Flexible viruses: structural disorder in viral proteins. V. N. Uversky and S. Longhi. Hoboken, New Jersey, John Wiley and Sons.
- Hagmaier, K., N. Stock, et al. (2006). "A single amino acid substitution in the V protein of Nipah virus alters its ability to block interferon signalling in cells from different species." J Gen Virol **87**(Pt 12): 3649-3653.
- Halpin, K., P. L. Young, et al. (2000). "Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus." J Gen Virol **81**(Pt 8): 1927-1932.
- Hanlon, C. A., I. V. Kuzmin, et al. (2005). "Efficacy of rabies biologics against new lyssaviruses from Eurasia." Virus research **111**(1): 44-54.
- Hao, Q. (2006). "Macromolecular envelope determination and envelope-based phasing." Acta Crystallogr D Biol Crystallogr **62**(Pt 8): 909-914.
- Harcourt, B. H., A. Tamin, et al. (2001). "Molecular characterization of the polymerase gene and genomic termini of Nipah virus." Virology **287**(1): 192-201.
- Harouaka, D. and G. W. Wertz (2009). "Mutations in the C-terminal loop of the nucleocapsid protein affect vesicular stomatitis virus RNA replication and transcription differentially." Journal of virology **83**(22): 11429-11439.
- Harty, R. N., M. E. Brown, et al. (2001). "Rhabdoviruses and the cellular ubiquitin-proteasome system: a budding interaction." J Virol **75**(22): 10623-10629.
- Harty, R. N., J. Paragas, et al. (1999). "A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding." J Virol **73**(4): 2921-2929.
- Hayman, D. T., R. Suu-Ire, et al. (2008). "Evidence of henipavirus infection in West African fruit bats." PLoS One **3**(7): e2739.

- Heggeness, M. H., A. Scheid, et al. (1980). "Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration." Proc Natl Acad Sci U S A **77**(5): 2631-2635.
- Heggeness, M. H., A. Scheid, et al. (1981). "The relationship of conformational changes in the Sendai virus nucleocapsid to proteolytic cleavage of the NP polypeptide." Virology **114**(2): 555-562.
- Heinrich, B. S., D. K. Cureton, et al. (2010). "Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions." PLoS Pathog **6**(6): e1000958.
- Heinrich, B. S., B. Morin, et al. (2012). "Structural Properties of the C Terminus of Vesicular Stomatitis Virus N Protein Dictate N-RNA Complex Assembly, Encapsulation, and RNA Synthesis." Journal of virology **86**(16): 8720-8729.
- Herman, R. C. (1986). "Internal initiation of translation on the vesicular stomatitis virus phosphoprotein mRNA yields a second protein." J Virol **58**(3): 797-804.
- Hooper, P. T., H. A. Westbury, et al. (1997). "The lesions of experimental equine morbillivirus disease in cats and guinea pigs." Vet Pathol **34**(4): 323-329.
- Houben, K., L. Blanchard, et al. (2007). "Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P." Biophys J **93**(8): 2830-2844.
- Houben, K., D. Marion, et al. (2007). "Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family." J Virol **81**(13): 6807-6816.
- Huang, A. S. and R. R. Wagner (1966). "Comparative sedimentation coefficients of RNA extracted from plaque-forming and defective particles of vesicular stomatitis virus." J Mol Biol **22**(2): 381-384.
- Hyatt, A. D., S. R. Zaki, et al. (2001). "Ultrastructure of Hendra virus and Nipah virus within cultured cells and host animals." Microbes Infect **3**(4): 297-306.
- Izeni, F., A. Barge, et al. (1998). "Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures." The Journal of general virology **79** (Pt **12**): 2909-2919.
- Izeni, F., F. Baudin, et al. (2000). "Structure of the RNA inside the vesicular stomatitis virus nucleocapsid." Rna **6**(2): 270-281.
- Ivanov, I., T. Crepin, et al. (2010). "Structure of the dimerisation domain of the rabies virus phosphoprotein." J. Virol. **84**: 3707-3710.
- Ivanov, I., F. Yabukarski, et al. (2011). "Structural insights into the rhabdovirus transcription/replication complex." Virus research **162**(1-2): 126-137.
- Iverson, L. E. and J. K. Rose (1981). "Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription." Cell **23**(2): 477-484.
- Iwasaki, M., M. Takeda, et al. (2009). "The matrix protein of measles virus regulates viral RNA synthesis and assembly by interacting with the nucleocapsid protein." J Virol **83**(20): 10374-10383.
- Jacob, Y., E. Real, et al. (2001). "Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains." J Virol **75**(20): 9613-9622.
- Jacobs, B. L. and J. O. Langland (1996). "When two strands are better than one: the mediators and modulators of the cellular responses to double-stranded RNA." Virology **219**(2): 339-349.
- Janeway, C. A., Jr. (2001). "How the immune system protects the host from infection." Microbes Infect **3**(13): 1167-1171.

- Jelesarov, I. and H. R. Bosshard (1999). "Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition." J Mol Recognit **12**(1): 3-18.
- Jensen, M. R., G. Communie, et al. (2011). "Intrinsic disorder in measles virus nucleocapsids." Proc Natl Acad Sci U S A **108**(24): 9839-9844.
- Jensen, M. R., K. Houben, et al. (2008). "Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein." J Am Chem Soc **130**(25): 8055-8061.
- Johansen, D., C. M. Jeffries, et al. (2011). "Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching." Biophys J **100**(4): 1120-1128.
- Johansen, D., J. Trehwella, et al. (2011). "Fractal dimension of an intrinsically disordered protein: small-angle X-ray scattering and computational study of the bacteriophage lambda N protein." Protein Sci **20**(12): 1955-1970.
- Johansson, K., J. M. Bourhis, et al. (2003). "Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein." J Biol Chem **278**(45): 44567-44573.
- Kabsch, W. (2010). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.
- Karlin, D. and R. Belshaw (2012). "Detecting remote sequence homology in disordered proteins: discovery of conserved motifs in the N-termini of Mononegavirales phosphoproteins." PLoS One **7**(3): e31719.
- Karlin, D., F. Ferron, et al. (2003). "Structural disorder and modular organization in Paramyxovirinae N and P." J Gen Virol **84**(Pt 12): 3239-3252.
- Karlin, D., S. Longhi, et al. (2002). "Substitution of two residues in the measles virus nucleoprotein results in an impaired self-association." Virology **302**(2): 420-432.
- Karlin, D., S. Longhi, et al. (2002). "The N-terminal domain of the phosphoprotein of Morbilliviruses belongs to the natively unfolded class of proteins." Virology **296**(2): 251-262.
- Katoh, K. and D. M. Standley (2013). "MAFFT multiple sequence alignment software version 7: improvements in performance and usability." Mol Biol Evol **30**(4): 772-780.
- Kerdiles, Y. M., B. Cherif, et al. (2006). "Immunomodulatory properties of morbillivirus nucleoproteins." Viral Immunol **19**(2): 324-334.
- Kingston, R. L., W. A. Baase, et al. (2004). "Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins." J Virol **78**(16): 8630-8640.
- Kingston, R. L., L. S. Gay, et al. (2008). "Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization." J Mol Biol **379**(4): 719-731.
- Kingston, R. L., D. J. Hamel, et al. (2004). "Structural basis for the attachment of a paramyxoviral polymerase to its template." Proc Natl Acad Sci U S A **101**(22): 8301-8306.
- Kingston, R. L., A. B. Walter, et al. (2004). "Characterization of nucleocapsid binding by the measles and the mumps virus phosphoprotein." Journal of virology **78**(16): 8630-8640.
- Kleywegt, G. J. and A. T. Brunger (1996). "Checking your imagination: applications of the free R value." Structure **4**(8): 897-904.
- Kohn, J. E., I. S. Millett, et al. (2004). "Random-coil behavior and the dimensions of chemically unfolded proteins." Proc Natl Acad Sci U S A **101**(34): 12491-12496.

- Kolakofsky, D., P. Le Mercier, et al. (2004). "Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis." *Virology* **318**(2): 463-473.
- Komarova, A. V., E. Real, et al. (2007). "Rabies virus matrix protein interplay with eIF3, new insights into rabies virus pathogenesis." *Nucleic Acids Res* **35**(5): 1522-1532.
- Konarev, P. V., V. V. Volkov, et al. (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." *J. Appl. Cryst.* **36**: 1277-1282.
- Kopecky, S. A. and D. S. Lyles (2003). "The cell-rounding activity of the vesicular stomatitis virus matrix protein is due to the induction of cell death." *J Virol* **77**(9): 5524-5528.
- Kopecky, S. A., M. C. Willingham, et al. (2001). "Matrix protein and another viral component contribute to induction of apoptosis in cells infected with vesicular stomatitis virus." *J Virol* **75**(24): 12169-12181.
- Koyama, A. H., H. Irie, et al. (2003). "Virus multiplication and induction of apoptosis by Sendai virus: role of the C proteins." *Microbes Infect* **5**(5): 373-378.
- Kretzschmar, E., R. Peluso, et al. (1996). "Normal replication of vesicular stomatitis virus without C proteins." *Virology* **216**(2): 309-316.
- Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." *J Mol Biol* **372**(3): 774-797.
- Ksiazek, T. G., P. A. Rota, et al. (2011). "A review of Nipah and Hendra viruses with an historical aside." *Virus Res* **162**(1-2): 173-183.
- Kulkarni, S., V. Volchkova, et al. (2009). "Nipah virus edits its P gene at high frequency to express the V and W proteins." *J Virol* **83**(8): 3982-3987.
- Lacroix, E., A. R. Viguera, et al. (1998). "Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters." *J Mol Biol* **284**(1): 173-191.
- Lahaye, X., A. Vidy, et al. (2009). "Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication." *J Virol* **83**(16): 7948-7958.
- Lamb, R. A. (2007). Mononegavirales. *Fields Virology 5th edition*. D. M. Knipe and P. M. Howley. Philadelphia, Lippincott Williams & Wilkins. **1**: 1357-1361.
- Lamb, R. A. and D. Kolakofsky (2001). "Paramyxoviridae: the viruses and their replication." *Fields Virology 4th edn*: 1305-1340.
- Lamb, R. A. and G. D. Parks (2007). Paramyxoviridae: The viruses and their replication. *Fields Virology 5th edition*. D. M. Knipe and P. M. Howley. Philadelphia, Lippincott Williams & Wilkins. **1**: 1449-1496.
- Lange, O. F., N. A. Lakomek, et al. (2008). "Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution." *Science* **320**(5882): 1471-1475.
- Langer, G., S. X. Cohen, et al. (2008). "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7." *Nat Protoc* **3**(7): 1171-1179.
- Larkin, M. A., G. Blackshields, et al. (2007). "Clustal W and Clustal X version 2.0." *Bioinformatics* **23**(21): 2947-2948.
- Lee, B. and Z. A. Ataman (2011). "Modes of paramyxovirus fusion: a Henipavirus perspective." *Trends Microbiol* **19**(8): 389-399.
- Letchworth, G. J., L. L. Rodriguez, et al. (1999). "Vesicular stomatitis." *Vet J* **157**(3): 239-260.
- Leyrat, C., F. C. Gerard, et al. (2010). "Structural disorder in proteins of the rhabdoviridae replication complex." *Protein Pept Lett* **17**(8): 979-987.

- Leyrat, C., M. R. Jensen, et al. (2011). "The N(0)-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient alpha-helices." Protein Sci **20**(3): 542-556.
- Leyrat, C., E. A. Ribeiro, et al. (2011). "Structure, interactions with host-cell and functions of rhabdovirus phosphoprotein." Future Virol. **6**(4): 465-481.
- Leyrat, C., R. Schneider, et al. (2012). "Ensemble structure of the modular and flexible full-length vesicular stomatitis virus phosphoprotein." J Mol Biol **423**(2): 182-197.
- Leyrat, C., F. Yabukarski, et al. (2011). "Structure of the vesicular stomatitis virus N(0)-P complex." PLoS Pathog **7**(9): e1002248.
- Liljeroos, L., J. T. Huiskonen, et al. (2011). "Electron cryotomography of measles virus reveals how matrix protein coats the ribonucleocapsid within intact virions." Proc Natl Acad Sci U S A **108**(44): 18085-18090.
- Lipkin, W. I. and T. Briese (2007). Bornaviridae. Fields Virology. D. M. Knipe and P. M. Howley. Philadelphia, Lippincott Williams & Wilkins. **2**: 1829-1851.
- Liston, P., R. Batal, et al. (1997). "Protein interaction domains of the measles virus nucleocapsid protein (NP)." Arch Virol **142**(2): 305-321.
- Llorente, M. T., B. Garcia-Barreno, et al. (2006). "Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization." The Journal of general virology **87**(Pt 1): 159-169.
- Lodish, H. F. and M. Porter (1980). "Translational control of protein synthesis after infection by vesicular stomatitis virus." J Virol **36**(3): 719-733.
- Longhi, S., V. Receveur-Brechot, et al. (2003). "The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein." J Biol Chem **278**(20): 18638-18648.
- Luby, S. P., E. S. Gurley, et al. (2009). "Transmission of human infection with Nipah virus." Clin Infect Dis **49**(11): 1743-1748.
- Ludlow, L. E., M. K. Lo, et al. (2008). "Henipavirus V protein association with Polo-like kinase reveals functional overlap with STAT1 binding and interferon evasion." J Virol **82**(13): 6259-6271.
- Luo, M., T. J. Green, et al. (2007). "Conserved characteristics of the rhabdovirus nucleoprotein." Virus Res **129**(2): 246-251.
- Luo, M., T. J. Green, et al. (2007). "Structural comparisons of the nucleoprotein from three negative strand RNA virus families." Virology **4**: 72.
- Luo, P. and R. L. Baldwin (1997). "Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water." Biochemistry **36**: 8413-8421.
- Lupas, A., M. Van Dyke, et al. (1991). "Predicting coiled coils from protein sequences." Science **252**(5010): 1162-1164.
- Lyles, D. S. and C. E. Rupprecht (2007). Rhabdoviridae. Fields Virology 5th edition. D. M. Knipe and P. M. Howley. Philadelphia, Lippincott Williams & Wilkins. **1**: 1364-1408.
- Lynch, S. and D. Kolakofsky (1978). "Ends of the RNA within Sendai virus defective interfering nucleocapsids are not free." Journal of virology **28**(2): 584-589.
- Mao, B., R. Guan, et al. (2011). "Improved technologies now routinely provide protein NMR structures useful for molecular replacement." Structure **19**(6): 757-766.
- Marley, J., M. Lu, et al. (2001). "A method for efficient isotopic labeling of recombinant proteins." J Biomol NMR **20**(1): 71-75.
- Marsh, G. A., C. de Jong, et al. (2012). "Cedar virus: a novel Henipavirus isolated from Australian bats." PLoS Pathog **8**(8): e1002836.

- Marsh, J. A., V. K. Singh, et al. (2006). "Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation." *Protein Sci* **15**(12): 2795-2804.
- Martinez, N., E. A. Ribeiro, Jr., et al. (2013). "Structure of the C-terminal domain of Lettuce Necrotic Yellow Virus phosphoprotein." *J Virol*.
- Martinho, M., J. Habchi, et al. (2012). "Assessing induced folding within the intrinsically disordered C-terminal domain of the Henipavirus nucleoproteins by site directed spin labeling EPR spectroscopy." *J Biomol Struct Dyn* **in press**.
- Masters, P. S. and A. K. Banerjee (1988). "Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA." *J Virol* **62**(8): 2658-2664.
- Mathieu, C., V. Guillaume, et al. (2012). "Lethal Nipah virus infection induces rapid overexpression of CXCL10." *PLoS One* **7**(2): e32157.
- Mathieu, C., V. Guillaume, et al. (2012). "Nonstructural Nipah virus C protein regulates both the early host proinflammatory response and viral virulence." *J Virol* **86**(19): 10766-10775.
- Mavrakis, M., F. Ise, et al. (2003). "Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells." *Virology* **305**(2): 406-414.
- Mavrakis, M., A. A. McCarthy, et al. (2004). "Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus." *J Mol Biol* **343**(4): 819-831.
- Mavrakis, M., S. Mehoulas, et al. (2006). "Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA." *Virology* **349**(2): 422-429.
- McColl, K. A., N. Tordo, et al. (2000). "Bat lyssavirus infections." *Rev Sci Tech* **19**(1): 177-196.
- McCoy, A. J. (2007). "Solving structures of protein complexes by molecular replacement with Phaser." *Acta Crystallogr D Biol Crystallogr* **63**(Pt 1): 32-41.
- McNicholas, S., E. Potterton, et al. (2011). "Presenting your structures: the CCP4mg molecular-graphics software." *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 386-394.
- Mebatsion, T., F. Weiland, et al. (1999). "Matrix protein of rabies virus is responsible for the assembly and budding of bullet-shaped particles and interacts with the transmembrane spike glycoprotein G." *J Virol* **73**(1): 242-250.
- Mellon, M. G. and S. U. Emerson (1978). "Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus." *J Virol* **27**(3): 560-567.
- Menghani, S., R. Chikhale, et al. (2012). "Chandipura Virus: An emerging tropical pathogen." *Acta tropica* **124**(1): 1-14.
- Middleton, D. J., H. A. Westbury, et al. (2002). "Experimental Nipah virus infection in pigs and cats." *J Comp Pathol* **126**(2-3): 124-136.
- Moll, M., H. D. Klenk, et al. (2002). "Importance of the cytoplasmic tails of the measles virus glycoproteins for fusogenic activity and the generation of recombinant measles viruses." *J Virol* **76**(14): 7174-7186.
- Morin, B., J. M. Bourhis, et al. (2006). "Assessing induced folding of an intrinsically disordered protein by site-directed spin-labeling EPR spectroscopy." *J. Phys. Chem. B* **110**(41): 20596-20608.
- Morin, B., A. A. Rahmeh, et al. (2012). "Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase." *Embo J* **31**: 1320-1329.
- Morin, B. and S. P. Whelan (2012). "La protéine L des Mononegavirales." *Virologie soumise*.

- Moyer, S. A., S. Smallwood-Kentro, et al. (1991). "Assembly and transcription of synthetic vesicular stomatitis virus nucleocapsids." *J Virol* **65**(5): 2170-2178.
- Murray, K., R. Rogers, et al. (1995). "A novel morbillivirus pneumonia of horses and its transmission to humans." *Emerg Infect Dis* **1**(1): 31-33.
- Murray, K., P. Selleck, et al. (1995). "A morbillivirus that caused fatal disease in horses and humans." *Science* **268**(5207): 94-97.
- Murshudov, G. N., P. Skubak, et al. (2011). "REFMAC5 for the refinement of macromolecular crystal structures." *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 355-367.
- Myers, T. M., A. Pieters, et al. (1997). "A highly conserved region of the Sendai virus nucleocapsid protein contributes to the NP-NP binding domain." *Virology* **229**(2): 322-335.
- Myers, T. M., S. Smallwood, et al. (1999). "Identification of nucleocapsid protein residues required for Sendai virus nucleocapsid formation and genome replication." *The Journal of general virology* **80**(Pt 6): 1383-1391.
- Nagem, R. A., Z. Dauter, et al. (2001). "Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers." *Acta Crystallogr D Biol Crystallogr* **57**(Pt 7): 996-1002.
- Negrete, O. A., E. L. Levroney, et al. (2005). "EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus." *Nature* **436**(7049): 401-405.
- Newcomb, W. W. and J. C. Brown (1981). "Role of the vesicular stomatitis virus matrix protein in maintaining the viral nucleocapsid in the condensed form found in native virions." *J Virol* **39**(1): 295-299.
- Nishio, M., M. Tsurudome, et al. (1996). "Interaction between nucleocapsid protein (NP) and phosphoprotein (P) of human parainfluenza virus type 2: one of the two NP binding sites on P is essential for granule formation." *J Gen Virol* **77** (Pt 10): 2457-2463.
- Ogino, T. and A. K. Banerjee (2007). "Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus." *Mol Cell* **25**(1): 85-97.
- Ogino, T. and A. K. Banerjee (2011). "An unconventional pathway of mRNA cap formation by vesiculoviruses." *Virus research* **162**(1-2): 100-109.
- Ogino, T., S. P. Yadav, et al. (2010). "Histidine-mediated RNA transfer to GDP for unique mRNA capping by vesicular stomatitis virus RNA polymerase." *Proc Natl Acad Sci U S A* **107**(8): 3463-3468.
- Oglesbee, M. (2007). Nucleocapsid protein interactions with the major inducible 70 kDa heat shock protein. *Measles virus nucleoprotein*. S. Longhi. Hauppauge, NY, Nova Publishers Inc.: 53-98.
- Ohgimoto, S., H. Bando, et al. (1990). "Sequence analysis of P gene of human parainfluenza type 2 virus: P and cysteine-rich proteins are translated by two mRNAs that differ by two nontemplated G residues." *Virology* **177**(1): 116-123.
- Oksayan, S., L. Wiltzer, et al. (2012). "A novel nuclear trafficking module regulates the nucleocytoplasmic localization of the rabies virus interferon antagonist, P protein." *J Biol Chem* **287**(33): 28112-28121.
- Ong, S. T., K. Yusoff, et al. (2009). "Mutagenesis of the nucleocapsid protein of Nipah virus involved in capsid assembly." *J Gen Virol* **90**(Pt 2): 392-397.
- Pager, C. T. and R. E. Dutch (2005). "Cathepsin L is involved in proteolytic processing of the Hendra virus fusion protein." *J Virol* **79**(20): 12714-12720.
- Pannu, N. S. R., R.J. (1996). "Improved Structure refinement through maximum likelihood." *Acta. Cryst.* **A52**: 659-668.

- Pape, T. and T. R. Schneider (2004). "HKL2MAP: a graphical user interface for phasing with SHELX programs" *J. Appl. Cryst.* **37**: 843-844.
- Pasdeloup, D., N. Poisson, et al. (2005). "Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1-dependent nuclear export signal." *Virology* **334**(2): 284-293.
- Pattnaik, A. K., L. Hwang, et al. (1997). "Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication." *J Virol* **71**(11): 8167-8175.
- Patton, J. T., N. L. Davis, et al. (1984). "N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus." *J Virol* **49**(2): 303-309.
- Paul, P. R., D. Chattopadhyay, et al. (1988). "The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype)." *Virology* **166**(2): 350-357.
- Paz, A., T. Zeev-Ben-Mordehai, et al. (2008). "Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3." *Biophys J* **95**(4): 1928-1944.
- Pei, J. and N. V. Grishin (2001). "AL2CO: calculation of positional conservation in a protein sequence alignment." *Bioinformatics* **17**(8): 700-712.
- Peluso, R. W. and S. A. Moyer (1988). "Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA." *Virology* **162**(2): 369-376.
- Plattet, P., L. Strahle, et al. (2007). "Sendai virus RNA polymerase scanning for mRNA start sites at gene junctions." *Virology* **362**(2): 411-420.
- Plumet, S., W. P. Duprex, et al. (2005). "Dynamics of viral RNA synthesis during measles virus infection." *Journal of virology* **79**(11): 6900-6908.
- Poch, O., B. M. Blumberg, et al. (1990). "Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains." *J Gen Virol* **71 (Pt 5)**: 1153-1162.
- Pohl, C., W. P. Duprex, et al. (2007). "Measles virus M and F proteins associate with detergent-resistant membrane fractions and promote formation of virus-like particles." *J Gen Virol* **88**(Pt 4): 1243-1250.
- Poisson, N., E. Real, et al. (2001). "Molecular basis for the interaction between rabies virus phosphoprotein P and the dynein light chain LC8: dissociation of dynein-binding properties and transcriptional functionality of P." *J Gen Virol* **82**(Pt 11): 2691-2696.
- Precious, B., D. F. Young, et al. (1995). "Inducible expression of the P, V, and NP genes of the paramyxovirus simian virus 5 in cell lines and an examination of NP-P and NP-V interactions." *J Virol* **69**(12): 8001-8010.
- Pringle, C. R. (1997). "The order Mononegavirales--current status." *Arch Virol* **142**(11): 2321-2326.
- Putnam, C. D., M. Hammel, et al. (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." *Q Rev Biophys* **40**(3): 191-285.
- Qanungo, K. R., D. Shaji, et al. (2004). "Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA." *Proc Natl Acad Sci U S A* **101**(16): 5952-5957.
- Qian, B., S. Raman, et al. (2007). "High-resolution structure prediction and the crystallographic phase problem." *Nature* **450**(7167): 259-264.

- Rahmeh, A. A., J. Li, et al. (2009). "Ribose 2'-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein." J Virol **83**(21): 11043-11050.
- Rahmeh, A. A., B. Morin, et al. (2012). "Critical phosphoprotein elements that regulate polymerase architecture and function in vesicular stomatitis virus." Proc Natl Acad Sci U S A **109**(36): 14628-14633.
- Rahmeh, A. A., A. D. Schenk, et al. (2010). "Molecular architecture of the vesicular stomatitis virus RNA polymerase." Proc Natl Acad Sci U S A.
- Rambo, R. P. and J. A. Tainer (2011). "Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law." Biopolymers **95**(8): 559-571.
- Rambo, R. P. and J. A. Tainer (2013). "Accurate assessment of mass, models and resolution by small-angle scattering." Nature **496**(7446): 477-481.
- Rao, B. L., A. Basu, et al. (2004). "A large outbreak of acute encephalitis with high fatality rate in children in Andhra Pradesh, India, in 2003, associated with Chandipura virus." Lancet **364**(9437): 869-874.
- Rao, T. R., K. R. Singh, et al. (1967). "Experimental transmission of Chandipura virus by mosquitoes." Indian J Med Res **55**(12): 1306-1310.
- Raux, H., A. Flamand, et al. (2000). "Interaction of the rabies virus P protein with the LC8 dynein light chain." J Virol **74**(21): 10212-10216.
- Rhodes, D. P., S. A. Moyer, et al. (1974). "In vitro synthesis of methylated messenger RNA by the virion-associated RNA polymerase of vesicular stomatitis virus." Cell **3**(4): 327-333.
- Ribeiro, E. A., Jr., A. Favier, et al. (2008). "Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein." J Mol Biol **382**(2): 525-538.
- Ribeiro, E. A., C. Leyrat, et al. (2009). "Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes." J Mol Biol **394**(3): 558-575.
- Ringkjøbing Jensen, M., G. Communie, et al. (2011). "Intrinsic disorder in measles virus nucleocapsids." Proc Natl Acad Sci U S A **108**(24): 9839-9844.
- Roche, S., A. A. Albertini, et al. (2008). "Structures of vesicular stomatitis virus glycoprotein: membrane fusion revisited." Cell Mol Life Sci **65**(11): 1716-1728.
- Roche, S., S. Bressanelli, et al. (2006). "Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G." Science **313**(5784): 187-191.
- Roche, S., F. A. Rey, et al. (2007). "Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G." Science **315**(5813): 843-848.
- Rodriguez, J. J., C. D. Cruz, et al. (2004). "Identification of the nuclear export signal and STAT-binding domains of the Nipah virus V protein reveals mechanisms underlying interferon evasion." J Virol **78**(10): 5358-5367.
- Rodriguez, J. J., J. P. Parisien, et al. (2002). "Nipah virus V protein evades alpha and gamma interferons by preventing STAT1 and STAT2 activation and nuclear accumulation." J Virol **76**(22): 11476-11483.
- Roossinck, M. J. (2011). "The good viruses: viral mutualistic symbioses." Nat Rev Microbiol **9**(2): 99-108.
- Roux, L. (2005). "Dans le génome des Paramyxovirinae, les promoteurs et leurs activités sont façonnés par la « règle de six." Virologie **9**(1): 19-34.
- Rudolph, M. G., I. Kraus, et al. (2003). "Crystal structure of the borna disease virus nucleoprotein." Structure **11**(10): 1219-1226.

- Ruigrok, R. W., T. Crepin, et al. (2011). "Nucleoproteins and nucleocapsids of negative-strand RNA viruses." Curr Opin Microbiol **14**(4): 504-510.
- Sadler, A. J. and B. R. Williams (2008). "Interferon-inducible antiviral effectors." Nat Rev Immunol **8**(7): 559-568.
- Sakaguchi, T., A. Kato, et al. (2005). "AIP1/Alix is a binding partner of Sendai virus C protein and facilitates virus budding." J Virol **79**(14): 8933-8941.
- Sazzad, H. M., M. J. Hossain, et al. (2013). "Nipah virus infection outbreak with nosocomial and corpse-to-human transmission, Bangladesh." Emerg Infect Dis **19**(2): 210-217.
- Schmitt, A. P., B. He, et al. (1999). "Involvement of the cytoplasmic domain of the hemagglutinin-neuraminidase protein in assembly of the paramyxovirus simian virus 5." J Virol **73**(10): 8703-8712.
- Schneider, R., J. R. Huang, et al. (2012). "Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy." Mol Biosyst **8**(1): 58-68.
- Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**(Pt 10 Pt 2): 1772-1779.
- Schoehn, G., F. Iseni, et al. (2001). "Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site." Journal of virology **75**(1): 490-498.
- Schoehn, G., M. Mavrikakis, et al. (2004). "The 12 A structure of trypsin-treated measles virus N-RNA." Journal of molecular biology **339**(2): 301-312.
- Schubert, M., J. D. Keene, et al. (1980). "Site on the vesicular stomatitis virus genome specifying polyadenylation and the end of the L gene mRNA." J Virol **34**(2): 550-559.
- Shaw, M. L., W. B. Cardenas, et al. (2005). "Nuclear localization of the Nipah virus W protein allows for inhibition of both virus- and toll-like receptor 3-triggered signaling pathways." J Virol **79**(10): 6078-6088.
- Shaw, M. L., A. Garcia-Sastre, et al. (2004). "Nipah virus V and W proteins have a common STAT1-binding domain yet inhibit STAT1 activation from the cytoplasmic and nuclear compartments, respectively." J Virol **78**(11): 5633-5641.
- Shiell, B. J., D. R. Gardner, et al. (2003). "Sites of phosphorylation of P and V proteins from Hendra and Nipah viruses: newly emerged members of Paramyxoviridae." Virus Res **92**(1): 55-65.
- Shoemaker, B. A., J. J. Portman, et al. (2000). "Speeding molecular recognition by using the folding funnel: the fly-casting mechanism." Proc Natl Acad Sci U S A **97**(16): 8868-8873.
- Shu, Y., J. Habchi, et al. (2012). "Plasticity in structural and functional interactions between the phosphoprotein and nucleoprotein of measles virus." J Biol Chem **287**(15): 11951-11967.
- Sleat, D. E. and A. K. Banerjee (1993). "Transcriptional activity and mutational analysis of recombinant vesicular stomatitis virus RNA polymerase." J Virol **67**(3): 1334-1339.
- Smallwood, S., B. Cevik, et al. (2002). "Intragenic complementation and oligomerization of the L subunit of the sendai virus RNA polymerase." Virology **304**(2): 235-245.
- Smallwood, S. and S. A. Moyer (2004). "The L polymerase protein of parainfluenza virus 3 forms an oligomer and can interact with the heterologous Sendai virus L, P and C proteins." Virology **318**(1): 439-450.
- Smallwood, S., K. W. Ryan, et al. (1994). "Deletion analysis defines a carboxyl-proximal region of Sendai virus P protein that binds to the polymerase L protein." Virology **202**(1): 154-163.

- Solon, J., O. Gareil, et al. (2005). "Membrane deformations induced by the matrix protein of vesicular stomatitis virus in a minimal system." The Journal of general virology **86**(Pt 12): 3357-3363.
- Spehner, D., A. Kirn, et al. (1991). "Assembly of nucleocapsidlike structures in animal cells infected with a vaccinia virus recombinant encoding the measles virus nucleoprotein." Journal of virology **65**(11): 6296-6300.
- Spiropoulou, C. F. and S. T. Nichol (1993). "A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus." J Virol **67**(6): 3103-3110.
- Stetson, D. B. and R. Medzhitov (2006). "Type I interferons in host defense." Immunity **25**(3): 373-381.
- Stone, R. (2011). "Breaking the Chain in Bangladesh." Science **331**(6021): 1128-1131.
- Sugase, K., H. J. Dyson, et al. (2007). "Mechanism of coupled folding and binding of an intrinsically disordered protein." Nature **447**: 920-921.
- Svergun, D., C. Barberato, et al. (1995). "CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates." J. Appl. Cryst. **28**: 768-773.
- Svergun, D. I. (1992). "Determination of the regularization parameter in indirect-transform methods using perceptual criteria." J. Appl. Cryst. **25**: 495-503.
- Svergun, D. I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." Biophys J **76**(6): 2879-2886.
- Svergun, D. I. and M. H. J. Koch (2003). "Small-angle scattering studies of biological macromolecules in solution." Reports on Progress in Physics **66**: 1735-1782.
- Svergun, D. I., S. Richard, et al. (1998). "Protein hydration in solution: experimental observation by x-ray and neutron scattering." Proc Natl Acad Sci U S A **95**(5): 2267-2272.
- Szilagyi, J. F. and L. Uryvayev (1973). "Isolation of an infectious ribonucleoprotein from vesicular stomatitis virus containing an active RNA transcriptase." J Virol **11**(2): 279-286.
- Takacs, A. M., S. Barik, et al. (1992). "Phosphorylation of specific serine residues within the acidic domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro." J Virol **66**(10): 5842-5848.
- Takacs, A. M., T. Das, et al. (1993). "Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system." Proc Natl Acad Sci U S A **90**(21): 10375-10379.
- Tan, C. T. and K. B. Chua (2008). "Nipah virus encephalitis." Curr Infect Dis Rep **10**(4): 315-320.
- Tan, W. S., S. T. Ong, et al. (2004). "Solubility, immunogenicity and physical properties of the nucleocapsid protein of Nipah virus produced in Escherichia coli." J Med Virol **73**(1): 105-112.
- Tarbouriech, N., J. Curran, et al. (2000). "On the domain structure and the polymerization state of the sendai virus P protein." Virology **266**(1): 99-109.
- Tarbouriech, N., J. Curran, et al. (2000). "Tetrameric coiled coil domain of Sendai virus phosphoprotein." Nat Struct Biol **7**(9): 777-781.
- Tawar, R. G., S. Duquerroy, et al. (2009). "Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus." Science **326**(5957): 1279-1283.

- Taylor, G. (2003). "The phase problem." Acta Crystallogr D Biol Crystallogr **59**(Pt 11): 1881-1890.
- Terwilliger, T. C. (2000). "Maximum-likelihood density modification." Acta Crystallogr D Biol Crystallogr **56**(Pt 8): 965-972.
- Terwilliger, T. C. and J. Berendzen (1999). "Automated MAD and MIR structure solution." Acta Crystallogr D Biol Crystallogr **55**(Pt 4): 849-861.
- Terwilliger, T. C., R. W. Grosse-Kunstleve, et al. (2008). "Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard." Acta Crystallogr D Biol Crystallogr **64**(Pt 1): 61-69.
- Thomas, D., W. W. Newcomb, et al. (1985). "Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis." J Virol **54**(2): 598-607.
- Thoulouze, M. I., M. Lafage, et al. (1998). "The neural cell adhesion molecule is a receptor for rabies virus." J Virol **72**(9): 7181-7190.
- Tomba, P. and M. Fuxreiter (2008). "Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions." Trends Biochem Sci **33**(1): 2-8.
- Tomba, P., M. Fuxreiter, et al. (2009). "Close encounters of the third kind: disordered domains and the interactions of proteins." Bioessays **31**(3): 328-335.
- Tordo, N., O. Poch, et al. (1988). "Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses." Virology **165**(2): 565-576.
- Toriumi, H., Y. Eriguchi, et al. (2004). "Further studies on the hyperphosphorylated form (p40) of the rabies virus nominal phosphoprotein (P)." Microbiol Immunol **48**(11): 865-874.
- Tran, T. L., N. Castagne, et al. (2007). "The nine C-terminal amino acids of the respiratory syncytial virus protein P are necessary and sufficient for binding to ribonucleoprotein complexes in which six ribonucleotides are contacted per N protein protomer." The Journal of general virology **88**(Pt 1): 196-206.
- Tsumoto, K., M. Umetsu, et al. (2004). "Role of arginine in protein refolding, solubilization, and purification." Biotechnol Prog **20**(5): 1301-1308.
- Tuffereau, C., J. Benejean, et al. (1998). "Low-affinity nerve-growth factor receptor (P75NTR) can serve as a receptor for rabies virus." Embo J **17**(24): 7250-7259.
- Uson, I., E. Pohl, et al. (1999). "1.7 Å structure of the stabilized REIv mutant T39K. Application of local NCS restraints." Acta Crystallogr D Biol Crystallogr **55**(Pt 6): 1158-1167.
- Uson, I., C. E. Stevenson, et al. (2007). "Structure determination of the O-methyltransferase NovP using the 'free lunch algorithm' as implemented in SHELXE." Acta Crystallogr D Biol Crystallogr **63**(Pt 10): 1069-1074.
- Uversky, V. N. (1993). "Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule." Biochemistry **32**(48): 13288-13298.
- Vagin, A. and A. Teplyakov (2010). "Molecular replacement with MOLREP." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 22-25.
- Vidy, A., M. Chelbi-Alix, et al. (2005). "Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways." J Virol **79**(22): 14411-14420.
- Vinkemeier, U., S. L. Cohen, et al. (1996). "DNA binding of in vitro activated Stat1 alpha, Stat1 beta and truncated Stat1: interaction between NH2-terminal domains stabilizes binding of two dimers to tandem DNA sites." Embo J **15**(20): 5616-5626.

- Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." J. Appl. Cryst. **36**: 860-864.
- von Kobbe, C., J. M. van Deursen, et al. (2000). "Vesicular stomatitis virus matrix protein inhibits host cell gene expression by targeting the nucleoporin Nup98." Mol Cell **6**(5): 1243-1252.
- Wang, L. F., W. P. Michalski, et al. (1998). "A novel P/V/C gene in a new member of the Paramyxoviridae family, which causes lethal infection in humans, horses, and other animals." J Virol **72**(2): 1482-1490.
- Wang, L. F., M. Yu, et al. (2000). "The exceptionally large genome of Hendra virus: support for creation of a new genus within the family Paramyxoviridae." J Virol **74**(21): 9972-9979.
- Warnes, A., A. R. Fooks, et al. (1995). "Expression of the measles virus nucleoprotein gene in Escherichia coli and assembly of nucleocapsid-like structures." Gene **160**(2): 173-178.
- Watanabe, N., M. Kawano, et al. (1996). "Identification of the sequences responsible for nuclear targeting of the V protein of human parainfluenza virus type 2." The Journal of general virology **77**(Pt 2): 327-338.
- Westbury, H. A., P. T. Hooper, et al. (1996). "Susceptibility of cats to equine morbillivirus." Aust Vet J **74**(2): 132-134.
- Westbury, H. A., P. T. Hooper, et al. (1995). "Equine morbillivirus pneumonia: susceptibility of laboratory animals to the virus." Aust Vet J **72**(7): 278-279.
- Whelan, S. P. and G. W. Wertz (2002). "Transcription and replication initiate at separate sites on the vesicular stomatitis virus genome." Proc Natl Acad Sci U S A **99**(14): 9178-9183.
- Wild, T. F. (2009). "Henipaviruses: a new family of emerging Paramyxoviruses." Pathol Biol (Paris) **57**(2): 188-196.
- Williamson, M. M., P. T. Hooper, et al. (1998). "Transmission studies of Hendra virus (equine morbillivirus) in fruit bats, horses and cats." Aust Vet J **76**(12): 813-818.
- Winn, M. D., C. C. Ballard, et al. (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.
- Wiseman, T., S. Williston, et al. (1989). "Rapid measurement of binding constants and heats of binding using a new titration calorimeter." Anal Biochem **179**(1): 131-137.
- Wishart, D. S. and B. D. Sykes (1994). "The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data." J Biomol NMR **4**(2): 171-180.
- Wishart, D. S., B. D. Sykes, et al. (1991). "Relationship between nuclear magnetic resonance chemical shift and protein secondary structure." J Mol Biol **222**(2): 311-333.
- Wong, K. T., W. J. Shieh, et al. (2002). "Nipah virus infection: pathology and pathogenesis of an emerging paramyxoviral zoonosis." Am J Pathol **161**(6): 2153-2167.
- Wong, S., S. Lau, et al. (2007). "Bats as a continuing source of emerging infections in humans." Reviews in medical virology **17**(2): 67-91.
- Yoneyama, M., M. Kikuchi, et al. (2004). "The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses." Nat Immunol **5**(7): 730-737.
- Yu, M., E. Hansson, et al. (1998). "The attachment protein of Hendra virus has high structural similarity but limited primary sequence homology compared with viruses in the genus Paramyxovirus." Virology **251**(2): 227-233.
- Zarrinpar, A., R. P. Bhattacharyya, et al. (2003). "The structure and function of proline recognition domains." Sci STKE **2003**(179): RE8.

- Zhang, X., J. M. Bourhis, et al. (2005). "Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus." Virology **337**(1): 162-174.
- Zhang, X., C. Glendening, et al. (2002). "Identification and characterization of a regulatory domain on the carboxyl terminus of the measles virus nucleocapsid protein." Journal of virology **76**(17): 8737-8746.
- Zhang, X., T. J. Green, et al. (2008). "Role of intermolecular interactions of vesicular stomatitis virus nucleoprotein in RNA encapsidation." Journal of virology **82**(2): 674-682.
- Zhao, H. and A. K. Banerjee (1995). "Interaction between the nucleocapsid protein and the phosphoprotein of human parainfluenza virus 3. Mapping of the interacting domains using a two-hybrid system." J Biol Chem **270**(21): 12485-12490.

Résumé :

Le virus de la stomatite vésiculaire (VSV) et le virus Nipah (NiV) appartiennent respectivement aux familles des *Rhabdoviridae* et des *Paramyxoviridae*. VSV est un modèle du virus de la rage tandis que NiV est un virus émergent, appartenant à la sous-famille des *Paramyxovirinae*, pour lequel les données moléculaires et structurales sont limitées. Ce sont des virus enveloppés dont le génome code pour cinq à neuf protéines. Le complexe de réplication de ces virus est constitué de trois protéines : la phosphoprotéine (P), la nucléoprotéine (N) et la polymérase virale (L). La N encapside le génome viral et l'ensemble N-ARN sert de matrice pour la transcription et la réplication. La P joue deux rôles : elle sert de cofacteur pour la polymérase et forme le complexe N⁰-P qui maintient la N sous une forme soluble, compétente pour l'encapsidation des génomes néo-synthétisés. Un premier objectif de mon travail de thèse consistait à étudier la structure et la dynamique des protéines P de VSV et de NiV. Ce sont des protéines modulaires qui contiennent des domaines structurés, séparés par des régions flexibles. A mon arrivée au laboratoire un travail important avait été déjà réalisé sur la P de VSV et j'ai participé à l'achèvement de cette étude. Je me suis ensuite intéressé à la protéine P de NiV. J'ai cristallisé et résolu par diffraction des rayons X les structures du domaine C-terminal et du domaine central (codes PDB : 4F9X et 4GJW). La combinaison de ces modèles cristallographiques avec des données de SAXS sur la P entière et des données de résonance magnétique nucléaire (RMN, collaboration IBS) va permettre d'obtenir un modèle atomique de la P entière sous la forme d'un ensemble de conformères. Un deuxième objectif était d'étudier les complexes N⁰-P. J'ai activement participé au développement de la méthode de reconstitution et à la caractérisation structurale du complexe N⁰-P de VSV, entre un mutant de la N (NΔ21) et un peptide N-terminal de la P (code PDB : 3PMK). J'ai ensuite reconstitué, cristallisé et résolu la structure du complexe N⁰-P de NiV entre la N (tronquée de son domaine C-terminal) et la partie N-terminale de la P. Ces structures montrent par quels mécanismes moléculaires la P maintient la N sous forme monomérique, en empêchant sa polymérisation et son interaction avec l'ARN. Les résultats présentés ici ont permis de générer de nouvelles hypothèses pour expliquer les mécanismes d'encapsidation et d'initiation de la synthèse d'ARN chez ces virus. Le complexe N⁰-P étant essentiel pour la réplication du virus, l'information structurale obtenue au cours de ce travail devrait permettre d'envisager l'utilisation de ce complexe comme cible pour le développement de composés antiviraux.

Summary :

Vesicular stomatitis virus (VSV) and Nipah virus (NiV) belong to the *Rhabdoviridae* and *Paramyxoviridae* families, respectively. VSV serves as model system for rabies virus while NiV is an emerging pathogen of the *Paramyxovirinae* subfamily, for which molecular and structural data are scarce. Both viruses are enveloped and their genomes encode five to nine proteins. Three proteins form their replication complex: the phosphoprotein (P), the nucleoprotein (N) and the viral polymerase (L). N encapsidates the viral genome and this N-RNA complex serves as template for transcription and replication. P has two functions: it serves as a polymerase cofactor and forms an N⁰-P complex, which keeps the N protein in a soluble and monomeric state, competent for the encapsidation of the newly synthesized genomes. The first goal during the PhD work was to study the structure and dynamics of the VSV and NiV P proteins. These proteins are modular, containing structured domains separated by flexible regions. Before my arrival, a large amount of work was already done on the VSV P protein in the lab and I was involved in the final stages of this work. Then I studied the NiV P protein, crystallizing and solving the structures of its central and C-terminal domains by X-ray crystallography (PDB codes: 4F9X and 4GJW). Combining these structures with small angle X-ray scattering (SAXS) and nuclear magnetic resonance (NMR, collaboration with IBS group) data obtained for the entire protein will allow the construction of an atomic model of the phosphoprotein in the form of a conformational ensemble. The second goal was to study the N⁰-P complex. I actively participated in the development of the method which permitted the reconstruction of the VSV N⁰-P complex, using a truncation mutant of the N protein (NΔ21) and an N-terminal peptide from P, and to its structural determination (PDB code: 3PMK). Then I reconstructed, crystallized and solved the structure of the NiV N⁰-P complex using a C-terminally truncated N protein and the N-terminal region of the P protein. Both structures yielded insights into the molecular mechanisms used by the phosphoproteins in order to maintain the corresponding nucleoproteins in their monomeric state, thus inhibiting their polymerization and interaction with RNA. The results presented here also offered new hypothesis about mechanisms of encapsidation and of RNA synthesis initiation. Given that the N⁰-P complex is an essential component of the replication complex, the structural information gained from this work allow us to consider this complex as a potential antiviral target.