



HAL
open science

Etude structurale de la Nucléoprotéine du virus de la rage

Aurélie Albertini

► **To cite this version:**

Aurélie Albertini. Etude structurale de la Nucléoprotéine du virus de la rage. Biochimie [q-bio.BM]. Université Joseph-Fourier - Grenoble I, 2006. Français. NNT : . tel-00151191

HAL Id: tel-00151191

<https://theses.hal.science/tel-00151191>

Submitted on 1 Jun 2007

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.

THESE

Pour l'obtention du diplôme de Docteur de l'Université Joseph Fourier - Grenoble I

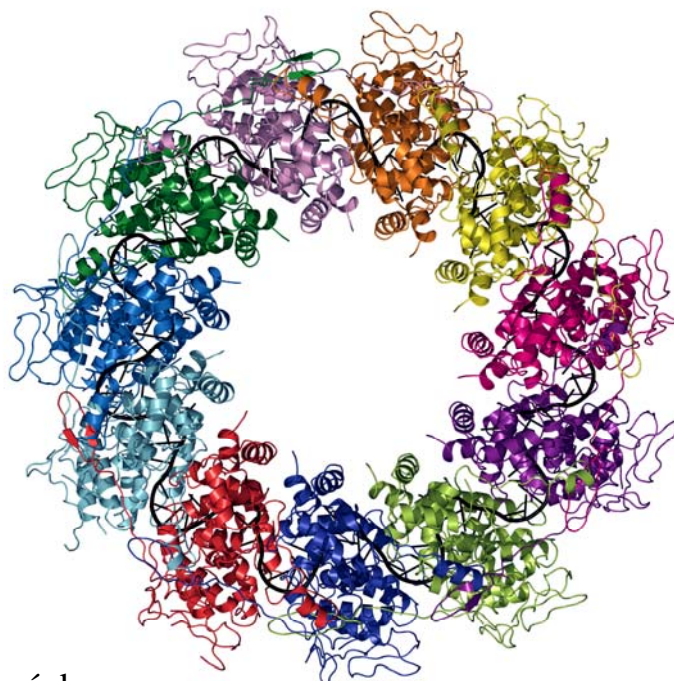
Spécialité : Biologie structurale et Nanobiologie

Présentée et soutenue publiquement par :

Aurélie ALBERTINI

Le 15 Décembre 2006

ETUDE STRUCTURALE DE LA NUCLEOPROTEINE DU VIRUS DE LA RAGE



Devant le jury composé de :

Dr Noël Tordo	Rapporteur
Pr Daniel Kolakofsky	Rapporteur
Pr Eva Pebay-Peyroula	Présidente du Jury
Dr Yves Gaudin	Examineur
Dr Felix Rey	Examineur
Pr Rob Ruigrok	Directeur de thèse

Travaux effectués à l'Institut de Virologie Moléculaire et Structurale

Résumé

Le virus de la rage appartient à l'ordre des *Mononegavirales*. Ces virus sont enveloppés, et leur génome est composé d'une seule molécule d'ARN de polarité négative. Dans le cas du virus de la rage, cet ARN encode cinq protéines virales dont la nucléoprotéine (N) qui est fortement impliquée dans la réplication et la transcription du virus. L'objectif de mon travail de thèse consistait à obtenir des informations structurales à la meilleure résolution possible sur la nucléoprotéine (N) du virus de la rage, ceci afin de mieux comprendre les mécanismes impliqués dans la multiplication virale. La nucléoprotéine existe sous deux formes : en complexe « soluble » avec la phosphoprotéine (P), ou associée à l'ARN viral. Le complexe N-ARN viral est la matrice utilisée par l'ARN-polymérase pour effectuer la transcription et la réplication. Il est possible de produire de manière recombinante des complexes N-ARN en forme d'anneaux composés d'un nombre de nucléoprotéines allant de 9 à 15 sous-unités par complexe. La mise au point d'une technique biochimique pour purifier ces différentes espèces suivie d'études en microscopie électronique et en cristallographie aux rayons X ont conduit à l'obtention d'un modèle atomique à une résolution de 3.5 Å. L'ultrastructure du complexe dont la structure a été résolue est formée par l'association entre 11 nucléoprotéines et un brin d'ARN de 99 bases. Il s'agit de la première structure de nucléoprotéine d'un virus à ARN négatif associée à son substrat, l'ARN. Cette structure permet de constater que la nucléoprotéine séquestre étroitement l'ARN limitant ainsi la formation d'ARN double brin et le protégeant de la réponse immunitaire innée. Pour devenir accessible à la polymérase virale, l'ARN génomique du virus de la rage doit être dissocié localement de quelques protomères de nucléoprotéine. Ce mode de fonctionnement spécifique existe probablement pour d'autres virus comme celui de la rougeole, Ebola ou la grippe, responsables eux aussi de pathologies humaines graves. La nucléoprotéine est une cible antivirale intéressante puisqu'elle n'existe qu'au sein des virus.

Mots clefs : Virus de la rage, virus à ARN négatif, nucléoprotéine, phosphoprotéine, protection de l'ARN, structure cristallographique, microscopie électronique, électrophorèse préparative.

Abstract

Rabies virus, a member of the *Mononegavirales* family, constitutes a serious health problem in developing countries lacking effective vaccination programs. The genome of negative-strand RNA viruses is compacted by the nucleoprotein (N) resulting in an N-RNA complex forming a tight helical nucleocapsid that is packaged into enveloped virions. Transcription and replication in the cytoplasm upon virus infection requires the N-RNA as a template for the RNA dependent RNA-polymerase complex. When N is expressed in insect cells in the absence of the other viral proteins, the protein binds to cellular RNA molecules and ring-shaped N-RNA complexes are produced. These rings are made up of 9 to 15 nucleoproteins subunits. We have been able to separate these distinct size classes and crystallised the complex containing 11 N-protomers and an RNA molecule of 99 ribonucleotides. The crystals diffracted to a resolution of 3.5 Å. The nucleoprotein subunit is composed of two domains that clamp around the RNA at their interface and shield it off from the environment, representing the replication inactive N-RNA conformation of the nucleocapsid inside the virus particle. Polymerisation of the nucleoprotein protomers is achieved by domain exchange between monomers acting as two hinges. The flexibility allowed by these hinges permits the formation of long N-RNA complexes adopting the larger helical radius required for nucleocapsid formation. The sequestering of the RNA within N indicates that the RNA has to dissociate from a monomer to become a template for the polymerase complex. Other negative-strand RNA viruses such as filoviruses, paramyxoviruses, borna viruses and influenza viruses may use the same structural principles to assemble a helical nucleocapsid. The nucleoprotein is an interesting target for drug design as it only exists in viruses. This thesis also contains the first experiments describing the characterization of the binding of the phosphoprotein, the polymerase co-factor, to the N-RNA complex.

Keywords: Rabies virus, negative strand RNA viruses, nucleoprotein, phosphoprotein, RNA protection, crystal structure, electron microscopy, preparative electrophoresis.

Remerciements...

Je tiens à remercier particulièrement mon responsable de thèse le Pr Rob Ruigrok pour avoir dirigé mes recherches, avec patience et compétence. Je lui exprime toute ma gratitude pour son soutien, ses encouragements et surtout la transmission de son savoir en virologie qui a suscité mon envie de poursuivre dans ce domaine.

Je remercie le Dr Noël Tordo et le Pr Dan Kolakofsky d'avoir accepté d'être les rapporteurs de ce travail. Je suis aussi extrêmement reconnaissante envers les Pr Eva Pebay-Peyroula, Dr Félix Rey et Dr Yves Gaudin pour leur participation à mon jury de thèse.

Je remercie le directeur de l'EMBL Grenoble, Dr Stephen Cusack, de m'avoir accueilli dans son institut durant les deux premières années de ma thèse où j'ai trouvé un environnement scientifique nécessaire pour mener à bien mes travaux.

Je tiens à témoigner ma gratitude au Dr Winfried Weissenhorn pour son intérêt et son implication dans mon travail, en particulier, pour le projet concernant les anneaux de la nucléoprotéine de la rage.

Je remercie également le Dr Guy Schoehn pour l'aide constante qu'il m'a apportée dès mon arrivée il y a déjà 4 ans, pour les nombreux essais en microscopie électronique qu'il a réalisés et pour m'avoir beaucoup aidé pour mon manuscrit.

Merci aussi au Dr Amy Wernimont pour tous les efforts qu'elle a fourni pour la partie cristallographique de ma thèse.

Ce travail doit beaucoup au Dr Cédric Clapier qui m'a apporté une aide et des conseils précieux particulièrement en biochimie.

Mes remerciements vont aussi vers le Dr Nicolas Tarbouriech pour sa patience, sa gentillesse et pour son aide en cristallographie.

Un grand merci au Dr Thibaut Crépin qui a eu l'amabilité de relire ce manuscrit et de me faire partager ses opinions.

Je remercie aussi le Dr Manos Mavrakis qui m'a enseigné les bases du travail en laboratoire à mon arrivée dans l'équipe.

Je tiens aussi à adresser toute ma gratitude au Pr Marc Jamin pour son œil de biophysicien sur mes travaux et pour les nombreuses discussions enrichissantes que nous avons eues ensemble.

Un grand merci à Francine Gérard pour sa bonne humeur constante et son amitié précieuse.

Merci aussi à Céline Fabry pour ses astuces en informatique et pour les nombreux moments de rigolade que nous avons partagés. Merci du fond du cœur à Hatice Akarsu pour son soutien amical et moral (même à distance). D'ailleurs, dans la même lignée je tiens à remercier Majida El Bakkouri, Jeanne Morinière, Lucy Freeman, Thibault Géoui, le Dr Marlyse Buisson, Isabelle Petit, le Dr Irina Gutsche, Monique Perrissin, Alex Dias, Sébastien Boulo, Julien Pérard, Joceline Visconti, Sophie Torres, Pierre Metais.

Je remercie aussi du fond du cœur Joffrey Heyraud et Aude Ta-Trung pour leur soutien sans limites ainsi que pour leur amitié, leurs encouragements et leur aide constante durant ma thèse.

Merci à mes parents sans qui je n'aurais pas pu faire de si longues études et me soucier pendant si longtemps de la structure d'une protéine et de son interaction avec de l'ARN.

*A mes parents Mariane et Jean-Yves
A mon frère Laurent et ma sœur Audrey
A Julien
A mes grands-parents*

TABLE DES MATIERES

PARTIE I : INTRODUCTION.....	5
I. INTRODUCTION GENERALE	6
A. Généralités sur les virus à ARN Négatif.....	6
B. Les Rhabdovirus	7
C. Le virus de la rage.....	10
D. Le cycle viral des Rhabdovirus.....	13
II. LA TRANSCRIPTION ET LA REPLICATION CHEZ LES MONONEGAVIRALES	15
A. Le génome des Rhabdovirus	15
B. La transcription virale.....	16
C. La réplication virale	18
III. LES PROTEINES VIRALES.....	19
A. Structures impliquées dans l'entrée et dans le bourgeonnement du virion	19
B. Structures impliquées dans la multiplication du virus et la synthèse d'ARN.....	24
OBJECTIFS DU TRAVAIL DE THESE.....	38
PARTIE II : PROTOCOLES POUR L'EXPRESSION EN CELLULES D'INSECTES ET LA PURIFICATION DES COMPLEXES N^o-P ET N-ARN.....	41
I. EXPRESSION ET PURIFICATION DU COMPLEXE N ^o -P NATIF DU VIRUS DE LA RAGE.....	42
A. Expression du complexe N ^o -P natif du virus de la rage en cellules d'insecte	42
B. Purification du complexe N ^o -P du virus de la rage.....	43
C. Purification du complexe N ^o -P _{dig}	43
II. EXPRESSION ET PURIFICATION DES COMPLEXES N-ARN DU VIRUS DE LA RAGE.....	44
A. Expression des complexes N-ARN du virus de la rage en cellules d'insectes	44
B. Purification des complexes N-ARN : gradients de chlorure de Césium	44
C. Purification des anneaux N-ARN : gradients de glycérol.....	44
PARTIE III : PURIFICATION ET CRISTALLISATION DES COMPLEXES CONTENANT LA NUCLEOPROTEINE.....	47
I. INTRODUCTION	48
II. CRISTALLISATION DU COMPLEXE ENTRE LA NUCLEOPROTEINE ET LA PHOSPHOPROTEINE	49
III. CRISTALLISATION DES COMPLEXES NUCLEOPROTEINE-ARN	51
A. Production d'un mélange d'anneaux nucléoprotéine-ARN.....	51
B. <i>ARTICLE 1: Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings.</i>	55
<i>Précisions sur la méthode de purification des différentes espèces d'anneaux N-ARN et l'obtention des cristaux (complément à l'article I)</i>	56
IV. CONCLUSION	64

**PARTIE IV : STRUCTURE CRISTALLOGRAPHIQUE DU COMPLEXE ARN-NUCLEOPROTEINE
DU VIRUS DE LA RAGE..... 66**

I.	INTRODUCTION	67
II.	ARTICLE II : CRYSTAL STRUCTURE OF THE RABIES VIRUS NUCLEOPROTEIN-RNA COMPLEX.....	70
III.	FIGURES DE LA PUBLICATION (ALBERTINI ET AL., 2006) A MEILLEURE RESOLUTION	72
IV.	PRECISIONS SUR LA METHODE DE RESOLUTION DE LA STRUCTURE EN COMPLEMENT A L'ARTICLE II ...	76
A.	<i>Détermination du groupe d'espace et collecte des jeux de données.....</i>	76
B.	<i>Résolution de la structure des anneaux N_{11}.....</i>	77
C.	<i>Construction du modèle.....</i>	81
V.	CONCLUSION	83

**PARTIE V : CARACTERISATION BIOCHIMIQUE DES COMPLEXES NUCLEOPROTEINE-ARN-
PHOSPHOPROTEINE 84**

I.	INTRODUCTION	85
II.	EXPRESSION ET PURIFICATION DE LA PHOSPHOPROTEINE DU VIRUS DE LA RAGE	85
A.	<i>Expression de la phosphoprotéine du virus de la rage en cellules d'insecte.....</i>	85
B.	<i>Purification de la Phosphoprotéine P du virus de la rage</i>	85
III.	INTERACTION ENTRE LES ANNEAUX N-ARN (N_{10}) ET LA PHOSPHOPROTEINE P.....	86
A.	<i>Etude de l'interaction entre les complexes N-ARN et la phosphoprotéine par MALLS</i>	86
B.	<i>Etude de l'interaction entre les anneaux N-ARN à 10 sous-unités (N_{10}) et la phosphoprotéine sur gel natif 88</i>	
IV.	CONCLUSION	89

PARTIE VI : DISCUSSION DES RESULTATS SCIENTIFIQUES 92

I.	COMPARAISON DE LA STRUCTURE DE LA NUCLEOPROTEINE DU VIRUS DE LA RAGE AVEC LES STRUCTURES DE NUCLEOPROTEINES DES AUTRES <i>MONONEGAVIRALES</i>	93
A.	<i>Comparaison entre la structure de la nucléoprotéine du virus de la rage et celle du VSV.....</i>	93
B.	<i>Comparaison des structures de la nucléoprotéine du virus de la rage et du Bornavirus.....</i>	98
II.	HYPOTHESE POUR LE MECANISME DE LA REPLICATION CHEZ LE VIRUS DE LA RAGE	99

RÉFÉRENCES BIBLIOGRAPHIQUES..... 103

ARTICLES SUPPLÉMENTAIRES..... 110

THE 12 Å STRUCTURE OF TRYPSIN-TREATED MEASLES VIRUS N-RNA	112
REVUE : STRUCTURES IMPLIQUEES DANS LA REPLICATION ET LA TRANSCRIPTION DES VIRUS A ARN NON SEGMENTE DE SENS NEGATIF	114

Abréviations :

2D : Deux dimensions
3D : Trois dimensions
Å : Angström
aa : acide aminé
ADN : Acide Désoxyribonucléique
ARN : Acide Ribonucléique
c/o : cut off
CCP4 : Collaborative Computational Project Number 4 in Protein Crystallography
DEAE : DiEthylAminoEthyl
EDTA : Ethylene diamine tetraacetic acid
EMTS : ethyl mercury thiosalicylate
ESRF : European Synchrotron Radiation Facility
kb : kilobase
kDa : kiloDalton
m.o.i. : Multiplicity of Infection.
MM : Masse moléculaire
MPD : 2 methyl-1,3 pentanediol
nt : Nucléotide
OMS : Organisation Mondiale de la Santé
p.i. : post infection
p/p : poids sur poids
PAGE : PolyAcrylamide Gel Electrophoresis
PBS : Phosphate Buffered Saline
PEG : Polyéthylène Glycol
rmsd : root mean square deviation
RNP : Particule Ribonucléique
rpm : tours par minute
SAD : Single Anomalous Dispersion
SDS : Sodium Dodecyle Sulfate
SNC : Système Nerveux Central
sw : swing
TBE : Tris Borate EDTA
UV : Ultraviolet
v/v : volume sur volume
VRS : Virus Respiratoire Syncytial
VSV : Virus de la Stomatite Vésiculaire

Code à une et trois lettres pour les acides aminés :

A	Ala	alanine	M	Met	méthionine
C	Cys	cystéine	N	Asn	asparagine
D	Asp	acide aspartique	P	Pro	proline
E	Glu	acide glutamique	Q	Gln	glutamine
F	Phe	phénylalanine	R	Arg	arginine
G	Gly	glycine	S	Ser	sérine
H	His	histidine	T	Thr	thréonine
I	Ile	isoleucine	V	Val	valine
K	Lys	lysine	W	Trp	tryptophane
L	Leu	leucine	Y	Tyr	tyrosine

Partie I : Introduction

I. Introduction générale

A. Généralités sur les virus à ARN Négatif

Les virus à ARN simple brin de polarité négative sont des virus enveloppés dont l'ARN viral est complémentaire de l'ARN messager. Ces virus possèdent soit un unique brin d'ARN viral (ordre des *Mononegavirales*) soit un génome segmenté. Les familles des *Rhabdoviridae* (virus de la rage), *Paramyxoviridae* (virus de la rougeole, virus respiratoire syncytial, virus Hendra et Nipah), *Filoviridae* (virus Ebola) et *Bornaviridae* appartiennent à l'ordre des *Mononegavirales*. Le génome des virus des familles des *Arenaviridae* (2 segments d'ARN, virus Lassa), *Bunyaviridae* (3 segments, virus de la fièvre de la vallée du rift) et *Orthomyxoviridae* (7-8 segments, virus de la grippe) est constitué de plusieurs molécules d'ARN simple brin de polarité négative.

Les virus à ARN négatif ont des morphologies variables, infectent un large spectre d'hôtes (plantes, invertébrés, mammifères), et sont à l'origine de nombreuses pathologies humaines (Tableau 1 et Figure 1).

Tableau 1 : Les différentes familles de l'ordre des *Mononegavirales*.

Famille	Génome	Morphologie	Hôte
Filoviridae	19 kb (7 protéines)	Filamenteuse	Réservoir inconnu Peut infecter les primates
<ul style="list-style-type: none"> • Genre : Marburg (virus Marburg) • Genre : Ebola (virus Ebola – Zaïre) 			
Paramyxoviridae	15-18 kb (10-11 protéines)	Pléomorphique	Vertébrés
Sous famille : <i>Paramyxovirinae</i> <ul style="list-style-type: none"> • Genre : Morbillivirus (virus de la rougeole). • Genre : Respirovirus (virus Sendai). • Genre : Rubulavirus (virus des oreillons). • Genre : Henipavirus (virus Hendra, virus Nipah) <i>en attente de classification définitive.</i> Sous famille : <i>Pneumovirinae</i> <ul style="list-style-type: none"> • Genre : Pneumovirus (virus respiratoire syncytial). 			
Rhabdoviridae	11-15 kb (5 protéines)	Balle de fusil	Animaux, plantes
<ul style="list-style-type: none"> • Genre : Lyssavirus (virus de la rage). • Genre : Vesiculovirus (virus de la stomatite vésiculaire). • Genre : Ephemerovirus (fièvre éphémère bovine). • Genre : Novirhabdovirus (nécrose infectieuse hématopoïétique). • Genre : Cytorhabdovirus (nécrose jaune de la laitue). • Genre : Nucleorhabdovirus (mosaïque du maïs). 			
Bornaviridae	9 kb (5 protéines)	?	Chevaux
<ul style="list-style-type: none"> • Genre : Bornavirus. 			

Le virus de la grippe et le virus respiratoire syncytial provoquent des affections aiguës des voies respiratoires. Les virus de la rage, Nipah et certains *Bunyaviridae* sont quant à eux responsables d'encéphalites graves. D'autres virus au sein de cette famille (virus Ebola, virus de la fièvre de Crimée Congo, virus Lassa) sont à l'origine de fièvres hémorragiques.

Certains de ces virus, tel que celui de la rage, sont connus depuis des siècles; d'autres sont des virus émergents qui, en s'adaptant rapidement à de nouveaux hôtes, causent des épidémies à forte mortalité comme dans le cas du virus Nipah en 1998 (Chua et al., 2000).

Lors de l'infection, après leur entrée dans la cellule hôte, la première activité de ces virus est la transcription de l'ARN viral de polarité négative en ARN messager. D'un point de vue structural, l'ARN des virus à ARN négatifs n'est jamais nu, ni dans les particules virales ni dans les cellules infectées mais toujours sous forme de complexe ribonucléoprotéique. La protéine majoritaire de ce complexe est la nucléoprotéine (N) qui recouvre entièrement l'ARN viral et forme une nucléocapside N-ARN hélicoïdale. Chez tout les *Mononegavirales*, deux autres protéines sont également associées à la nucléocapside : la polymérase virale (ARN-polymérase-ARN-dépendante, L) et son cofacteur, la phosphoprotéine (P) (Curran et al., 1994).

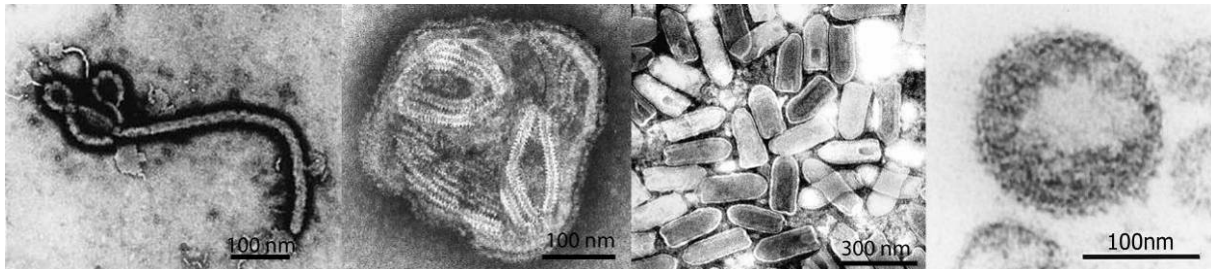


Figure 1 : Images de microscopie électronique de particules virales de *Mononegavirales*.

A : Virus Ebola (*Filoviridae*)

B : Virus Parainfluenza (*Paramyxoviridae*)

C : Virus de la stomatite vésiculaire (*Rhabdoviridae*)

D : Virus de la maladie de Borna (*Bornaviridae*)

B. Les Rhabdovirus

1) Classification

Les Rhabdovirus regroupent 5 genres différents, les genres Vesiculovirus, Lyssavirus et Ephemerovirus qui sont pathogènes pour les animaux, et les genres Cytorhabdovirus et Nucleorhabdovirus qui infectent les végétaux. Il semblerait que de nombreux Rhabdovirus soient véhiculés par des insectes (Hogenhout et al., 2003). Différentes espèces de mouches

sont impliquées dans la transmission horizontale de certains Vesiculovirus. Les moustiques peuvent transmettre le virus de la fièvre éphémère bovine (Ephemorovirus). Les sauterelles et les pucerons sont quant à eux souvent responsables de la transmission des Rhabdovirus qui infectent les végétaux.

Deux virus font l'objet de nombreuses études au sein de la famille des Rhabdovirus. Il s'agit du virus de la stomatite vésiculaire (VSV) qui fait partie du genre Vesiculovirus et de celui de la rage appartenant au genre Lyssavirus. Le VSV touche principalement le bétail (chevaux, bovin, porcs) et cause une maladie bénigne, qui se caractérise par l'apparition de lésions vésiculaires sur la bouche, la langue, le pis et les sabots des animaux. Contrairement au virus de la rage, le VSV n'est pas un virus dangereux pour l'homme; il peut donc être facilement étudié en laboratoire. Il constitue un excellent modèle pour la réplication et la transcription car c'est le virus le moins complexe de l'ordre des *Mononegavirales* ; il ne comporte que 5 gènes dont les signaux de régulation de leur expression sont plus simples que pour les autres virus de cet ordre.

Le genre Lyssavirus est composé de 7 génotypes, cette classification étant basée sur une analyse de séquence de nucléoprotéines (Badrane et al., 2001; Bourhy et al., 1993). Le génotype 1 englobe toutes les souches de virus de la rage. Les autres génotypes concernent des virus dits apparentés à la rage. Les Lyssavirus ont une distribution mondiale et infectent essentiellement des mammifères.

2) Morphologie des Rhabdovirus :

Les Rhabdovirus ont une forme caractéristique d'ogive avec une extrémité ronde et l'autre aplatie (Figure 2 A). Le diamètre des particules virales est d'environ 70 nm pour une longueur de 150 nm à 180 nm (Figure 2). Ces virus sont constitués de 74 % de protéines, 20 % de lipides, 3 % de carbohydrates et de 3 % d'ARN (Pal et al., 1987).

Les Rhabdovirus n'ont pas de capsid protéique, la membrane lipidique de ces virus provient des cellules hôtes. La glycoprotéine (G) est enchâssée dans cette membrane lipidique et forme des spicules (Gaudin et al., 1992). L'intérieur du virion est constitué d'un complexe ribonucléoprotéique appelé nucléocapside. Cette nucléocapside, qui est l'unité infectieuse du virus, présente une forme hélicoïdale caractéristique des virus de l'ordre des *Mononegavirales*. Dans le cas du virus de la rage, l'hélice est composée de 30 tours espacés de 45 Å (pas de l'hélice).

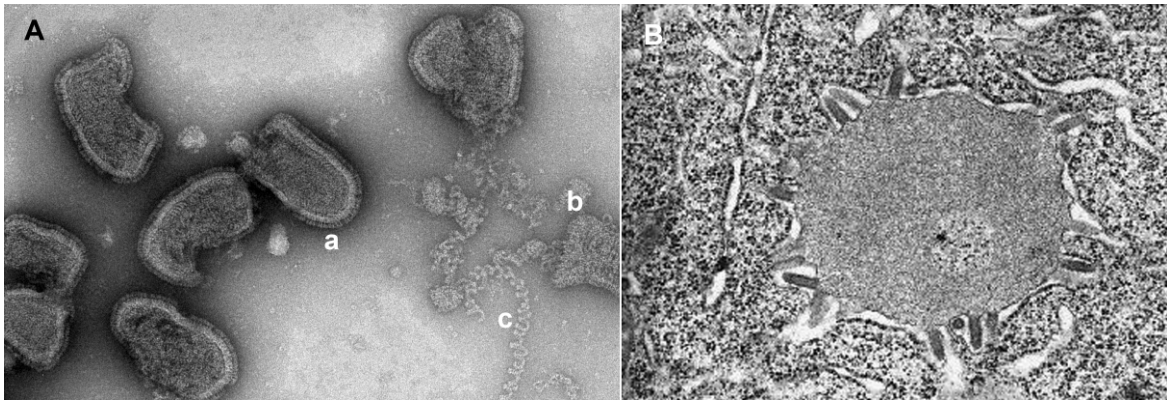


Figure 2 : Visualisations par microscopie électronique du virus de la rage.

A : Image de microscopie électronique de particules virales de rage (souche Pasteur). Photo prise par le Pr. R.W. Ruigrok. a : Particule virale. b : Nucléocapside virale se déroulant. c : Nucléocapside libre.
B : Image de microscopie électronique de particules virales de rage dans des cellules de cerveau. Photo prise par le Dr. F.A. Murphy.

D'après <http://www.tulane.edu/%7Edmsander/WWW/335/Rhabdoviruses.html>

Cette nucléocapside est formée par l'association entre l'ARN viral de polarité négative (12 kb), la nucléoprotéine (N), l'ARN polymérase virale (L) et son cofacteur la phosphoprotéine (P). La protéine matrice (M) condense la nucléocapside; pour le VSV, il a été proposé que les protéines matrice formeraient une structure en forme de cigare autour de laquelle la nucléocapside s'enroulerait. La protéine matrice ne serait alors en contact avec la membrane virale qu'au niveau des extrémités de cette structure en cigare (Barge et al., 1993) (Figure 3).

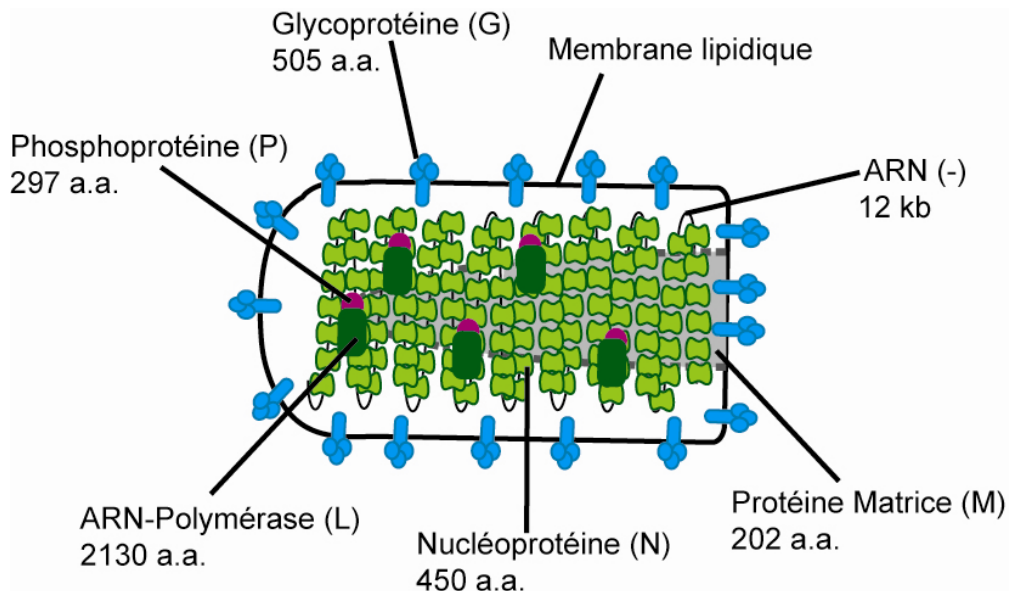


Figure 3 : Représentation schématique du virus de la rage.

Les protéines matrice sont représentées sous une forme de «cigare» autour duquel la nucléocapside s'enroulerait (Barge et al., 1993).

C. Le virus de la rage

1) Généralités

La rage est une zoonose virale très largement répandue dans le monde (Figure 4), tous les mammifères et plus particulièrement les carnivores y sont sujets. Tous les cas de rage humaine sont d'origine animale mis à part les cas exceptionnels dus à des greffes d'organes (Dietzschold and Koprowski, 2004). La majorité des cas mortels de rage chez l'homme se produisent en Asie (30000 décès chaque année au Pakistan, en Inde, et en Birmanie). L'Afrique est aussi touchée de manière conséquente par le virus de la rage. En Europe de l'est et en Amérique, quelques cas de rage humaine sont également signalés chaque année.

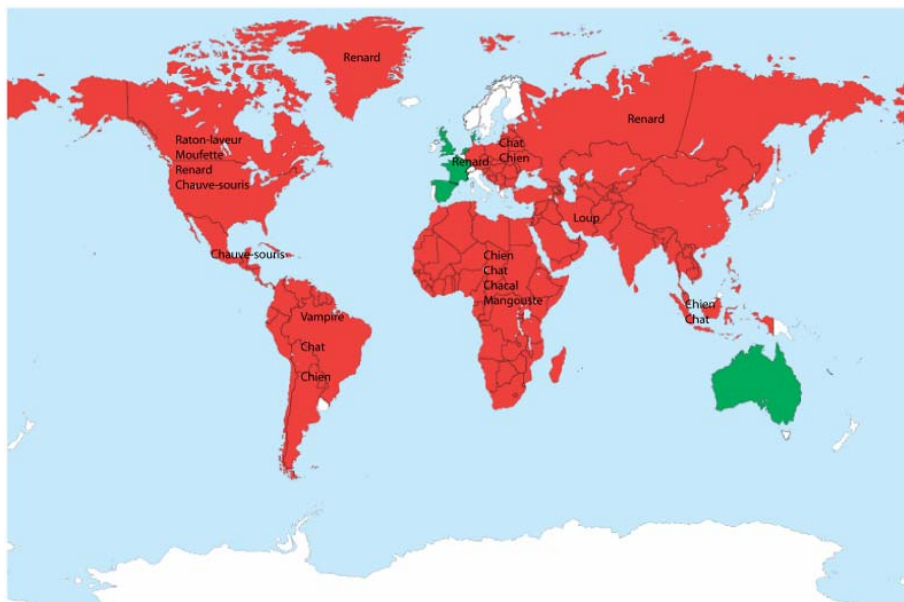


Figure 4 : Distribution mondiale du virus de la rage et des Lyssavirus.

En Rouge : Zones où le virus de la rage est présent (rouge), En vert : Zones déclarées «indemnes de rage », le virus de la rage terrestre est absent mais des virus apparentés à la rage y sont présents chez les chauves-souris. En Blanc : Zones prétendues indemnes de rage terrestre. (Warrell and Warrell, 2004).

Cette maladie est accidentellement transmissible à l'homme par inoculation transcutanée de virus. La vaccination contre la rage existe depuis plus de 100 ans et demeure de nos jours encore très efficace. Ce vaccin est utilisé à titre préventif chez l'homme ou chez l'animal mais peut aussi être utilisé à titre thérapeutique en particulier chez l'homme après morsure par un animal supposé infecté.

Le virus de la rage, après inoculation, se dirige rapidement vers les cellules neuronales. En remontant le long des neurones, le virus gagne le système nerveux central en un temps

variable dépendant non seulement de la dose virale inoculée mais aussi de la distance entre la morsure et le cerveau. Lorsque la maladie se déclare, elle provoque des encéphalomyélites (inflammation de la substance blanche du cerveau, du cervelet et de la moelle épinière) létales. Quelques cas de pseudo-guérisons «miraculeuses» ont été rapportés mais ils correspondent à des situations exceptionnelles (Madhusudana et al., 2002; Porras et al., 1976). Il n'existe pas d'antiviral efficace après l'apparition des premiers symptômes et finalement, depuis la vaccination de Joseph Meister en 1885 par Louis Pasteur, aucun progrès majeur n'a été fait en matière de thérapeutique *post infection* par le virus de la rage. Néanmoins, le vaccin a été amélioré et est désormais mieux toléré mais il n'existe toujours pas d'antiviral efficace après l'apparition des premiers symptômes. Selon l'OMS, la rage humaine est responsable d'au moins 50000 décès chaque année et ceci plus particulièrement dans les pays où il n'y a pas de vaccination systématique des animaux. Il est donc vraiment nécessaire, afin de traiter la rage en tant que maladie déclarée, de cibler le virus avant qu'il n'atteigne le système nerveux ou de bloquer sa multiplication massive au niveau du cerveau.

2) Epidémiologie

En France la rage humaine a disparu depuis 1924. Ceci a fait suite à des campagnes de vaccination massives qui ont visé principalement les chiens puis les renards. La France a été déclarée indemne de rage en 2001. Cependant, des cas de rage sont encore importés en France. Récemment, en 2004, trois cas de rage canine ont été diagnostiqués et en octobre 2003, un jeune enfant est décédé après avoir contracté le virus de la rage lors d'un voyage en Afrique du nord.

La rage canine est une enzootie touchant les régions où des mesures telles que l'élimination des chiens errants et la vaccination des animaux domestiques ne sont pas prises. Les chiens constituent un intermédiaire simple entre la rage sauvage et la rage domestique, et sont le principal vecteur du virus vers l'homme.

Au nord de l'Europe, la rage est essentiellement transmise par le renard, alors qu'en Europe de l'Est et dans les régions arctiques, le loup peut également être porteur de rage. En Amérique du nord, le raton laveur et la mouffette sont des hôtes potentiels de la rage. Certaines espèces de chiroptères hébergent le virus dans leur organisme pendant une très longue durée et peuvent l'excréter par l'intermédiaire de leur salive. Ces animaux sont alors de redoutables agents potentiels de contamination. Il est possible que ces chauves-souris soient des porteurs

sains du virus de la rage. Ces chauves-souris enragées sont présentes dans la plupart des régions du monde, y compris dans des pays déclarés indemnes de rage tels que l'Angleterre et l'Australie (Figure 5).

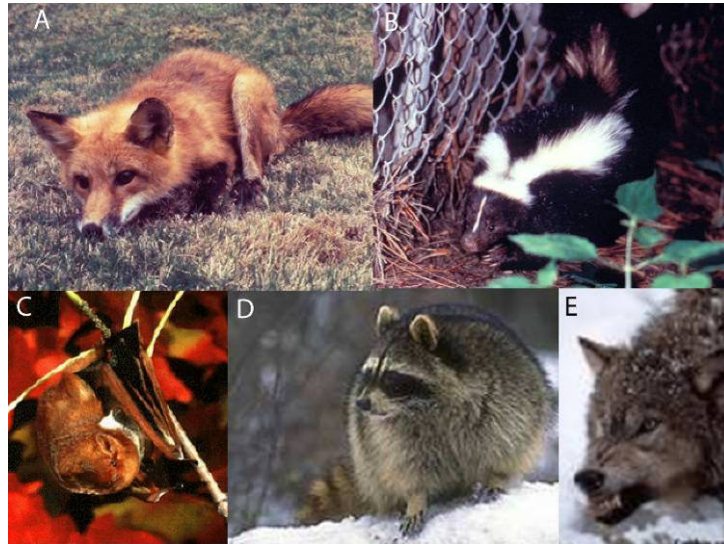


Figure 5 : Quelques animaux sauvages susceptibles d'être porteurs du virus de la rage.

A : Renard roux, B : Moufette, C : Chauve-souris, D : Raton-laveur, E : Loup.

Source : <http://www.who.int/rabies/en/>

3) Physiopathologie du virus de la rage

Le virus de la rage est transmis via la morsure d'un animal enragé mais aussi, par griffures ou léchage des plaies. Le virus est présent dans la salive des animaux infectés trois à cinq jours avant l'apparition des symptômes neurologiques. L'animal meurt dans les 15 jours qui suivent la déclaration de la maladie.

Pendant la période d'incubation, le virus se propage dans l'organisme en cheminant le long des neurones par voie axonale centripète. Cette période a une durée variable, dépendante de la zone de morsure. Le virus gagne des zones de l'organisme où il pourra se multiplier. La rage est un virus neurotrope : le virus remonte d'abord les nerfs périphériques et arrive au niveau de la moelle épinière pour aboutir dans le cerveau. Enfin, le virus passe par voie axonale centrifuge au niveau d'organes préférentiels (Figure 6). D'importantes doses virales sont alors retrouvées au niveau des glandes salivaires mais aussi des muqueuses ainsi que dans d'autres organes comme les yeux ou les reins. Le virus de la rage, qui a une forte affinité pour les zones du cerveau impliquées dans le comportement, induit souvent un comportement agressif chez l'animal ou la personne contaminée. Les modifications comportementales liées à cette

infection contribuent fortement à la transmission du virus puisque l'animal contaminé, devenu enragé, va lui-même disséminer le virus à un autre hôte par morsure.

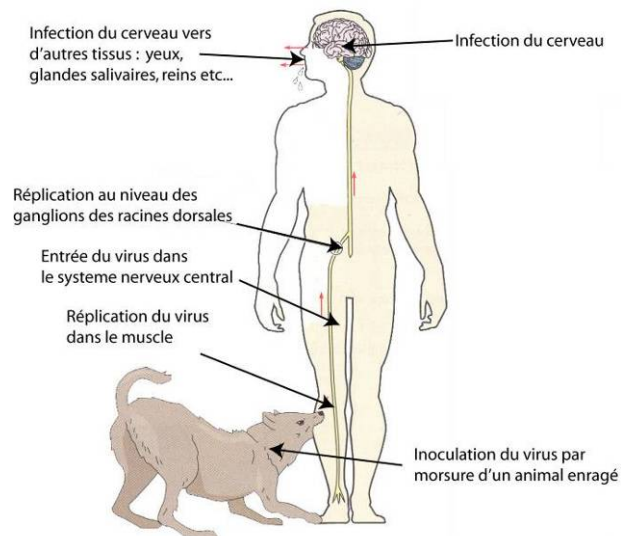


Figure 6 : Infection par le virus de la rage.

Adapté de Principles of Virology, Flint, Enquist, Racaniello, Skalka.

D. Le cycle viral des Rhabdovirus

Le cycle viral des Rhabdovirus se déroule entièrement dans le cytoplasme de la cellule hôte (Figure 7) à l'exception des Nucleorhabdovirus dont le cycle viral présente une phase nucléaire.

Entrée du virus : L'infection virale débute lorsque le virus se fixe à la surface des cellules hôtes. Les spicules de glycoprotéine à la surface du virus reconnaissent les récepteurs cellulaires (étape 1). Cette interaction induit l'internalisation du virus par endocytose (étape 2). L'abaissement du pH à l'intérieur de l'endosome active la glycoprotéine qui devient fusogénique, provoquant le processus de fusion membranaire entre les membranes virales et endosomales. Le complexe ribonucléoprotéique est alors libéré dans le cytoplasme de la cellule hôte (étape 3).

Multiplication virale : L'étape suivante est la transcription primaire. L'ARN viral est transcrit en premier lieu en ARN messagers par le complexe ARN-polymérase-ARN-dépendante - phosphoprotéine (étape 4). Ces ARN messagers sont polyadénylés et coiffés par l'ARN-polymérase virale puis traduits en protéines virales par la machinerie cellulaire (étape 5). L'ARN-polymérase virale passe en « mode réplication » lorsque la quantité de protéines néo-synthétisées atteint un niveau suffisant. Lors de la réplication, la polymérase virale ignore

les signaux de début et de fin de gènes; une molécule d'ARN positif totale est synthétisée. Cet ARN antisens va servir de matrice pour la synthèse de nouveaux génomes viraux. Ceux-ci peuvent alors servir de matrice soit pour la synthèse de protéines virales (mode transcription), soit pour la synthèse de nouveaux génomes viraux (mode réplication) (Banerjee, 1987). Le nombre d'ARN viraux augmente ainsi de manière exponentielle. Il est important de noter que les ARN viraux et ARN positifs sont toujours encapsidés par la nucléoprotéine qui joue un rôle structural crucial.

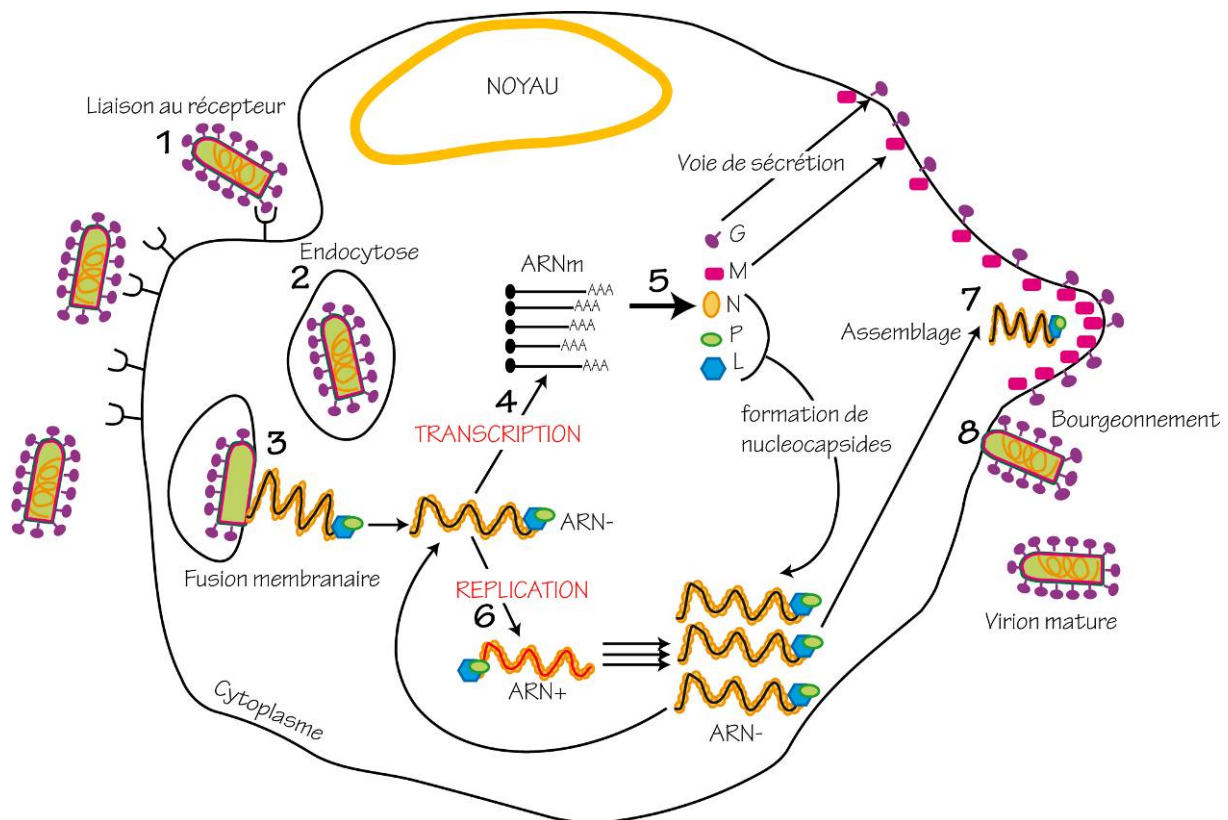


Figure 7 : Cycle viral du virus de la rage.

Le cycle viral se déroule entièrement dans le cytoplasme et peut être décomposé en 8 étapes : 1) Liaison au récepteur, 2) Endocytose, 3) Fusion membranaire et libération de la nucléocapside, 4) Transcription, 5) Traduction, 6) Réplication, 7) Assemblage et enfin 8) Bourgeonnement.

Sortie de la cellule :

Pour la formation de nouvelles particules virales, les ARN viraux néoformés associés à la nucléoprotéine et à l'ARN-polymérase-virale sont exportés vers la membrane plasmique. Les glycoprotéines sont transportées via l'appareil de Golgi vers la région baso-latérale de la membrane plasmique. La protéine de matrice M se place sur la face interne de la membrane plasmique où elle interagit avec les queues cytoplasmiques des spicules de glycoprotéine ainsi qu'avec les nucléocapsides dont elle assure la condensation en hélice. De nouvelles particules

virales compactes sont formées et peuvent alors bourgeonner hors des cellules hôtes (étapes 7 et 8).

II. La transcription et la réplication chez les Mononegavirales

L'ARN-polymérase-ARN-dépendante virale possède une activité réplacase et transcriptase responsable des processus de synthèse d'ARN durant le cycle viral. Une fois libéré dans le cytoplasme de la cellule hôte, l'ARN viral subit en premier la transcription puis la réplication, ceci s'appliquant à tous les virus de l'ordre des *Mononegavirales*. La plupart des connaissances sur le sujet dérivent d'études sur des virus modèles : le Virus de la Stomatite Vésiculaire (Rhabdovirus) et le virus Sendai (Paramyxovirus).

A. Le génome des Rhabdovirus

Le génome du virus de la rage et des Rhabdovirus est composé d'un unique brin d'ARN de polarité négative. Sa taille est comprise entre 11 et 15 kb. Cet ARN, toujours associé de manière étroite à la nucléoprotéine, code au moins 5 protéines virales.

La transcription du génome viral est initiée à l'extrémité 3' du génome par la synthèse d'un ARN court de 58 nucléotides appelé ARN leader (Colonna and Banerjee, 1976). Cet ARN n'est pas un ARN messenger : il n'est pas traduit, et n'est ni coiffé, ni polyadénylé. L'ARN leader est encapsidé en premier lors de la réplication virale (Kurilla et al., 1984; Tordo et al., 1986a), et est impliqué dans l'initiation de la transcription. Il a aussi été montré, dans le cas du VSV, que l'ARN leader s'accumulait dans les noyaux des cellules infectées inhibant ainsi la transcription chez la cellule hôte (Kurilla et al., 1982).

On trouve ensuite successivement sur le génome les gènes codant pour les protéines virales N, P, M, G et L. L'extrémité 5' de l'ARN viral est une région extra-cistronique appelée région «trailer». Cette séquence de 69 nucléotides est importante dans la phase de réplication virale et présente un signal pour le début de l'encapsidation de l'ARN par la nucléoprotéine (Bourhy et al., 1989).

Les gènes viraux sont séparés par des régions intergéniques de taille variable, ces zones étant délimitées par des signaux de début et fin de transcription. Dans le cas du VSV, ces régions intergéniques sont réduites à 2 nucléotides (Rose, 1980). En ce qui concerne le virus de la rage, les régions intergéniques entre les gènes N et P, les gènes P et M et les gènes M et G sont généralement courtes (2 à 5 nucléotides) et conservées. En revanche, la région

intergénique entre les gènes G et L est très longue (423 nucléotides) avec une séquence nucléotidique variable suivant les différentes souches de virus rabique (Figure 8).

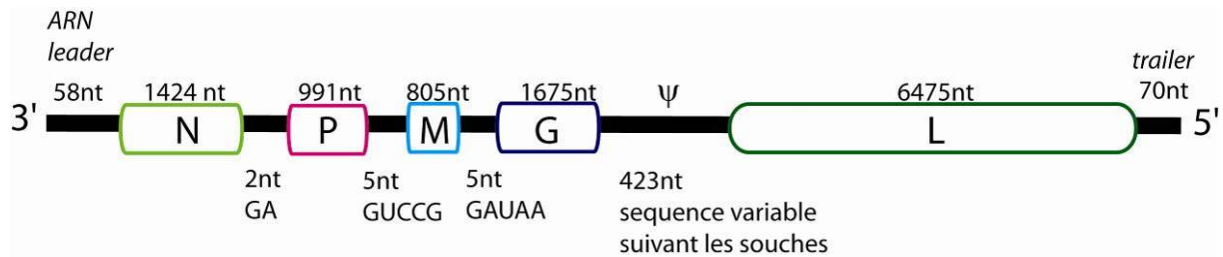


Figure 8 : Organisation du génome du virus de la rage.

La longueur des gènes, de l'ARN leader et de la région trailer est indiquée en nombre de nucléotides au dessus du schéma. La longueur et la séquence des régions intergéniques sont indiquées au dessous du schéma. La localisation du pseudogène ψ est également indiquée.

La région intergénique de 423 nucléotides située entre les gènes G et L est flanquée des signaux de début et de fin de transcription dégénérés. Parfois cette région est transcrite à la suite du gène de la protéine G. Cette zone intergénique n'est jamais traduite en protéine, elle constitue le pseudogène ψ (Tordo et al., 1986b) (Figure 8).

Il est possible que cette longue zone intergénique soit la conséquence de la perte d'un gène au cours de l'évolution des virus. Chez le VSV, cette région intergénique se limite à seulement deux nucléotides. En revanche, chez le virus IHNV (Virus de la Nécrose Hématopoïétique Infectieuse, Rhabdovirus touchant les poissons, Tableau 1) cette région constitue un sixième gène qui code pour la protéine NV (Non-Virion protein). Cette protéine de 12 kDa est exprimée dans les cellules infectées (Kurath and Leong, 1985), elle a été montrée comme étant un facteur de virulence (Thoulouze et al., 2004).

B. La transcription virale

La transcription est la première étape juste après l'entrée du virion dans la cellule hôte. Cette étape obligatoire précède le cycle de répllication virale. Elle s'effectue de manière séquentielle et aboutit à la synthèse d'ARNs messagers codant pour les protéines virales.

En tenant compte du fait que la transcription chez le VSV se déroule de manière polarisée et séquentielle, trois modèles ont été proposés (Banerjee, 1987) :

- Modèle clivable : la transcription s'initie au niveau d'un site unique d'entrée pour l'ARN-polymérase-ARN-dépendante virale (en 3') générant un transcrit unique qui

serait par la suite clivé en 6 ARNs : 1 ARN leader et 5 ARN messagers (Colonno and Banerjee, 1976).

- Modèle avec plusieurs sites d'initiation de la transcription : la polymérase peut accéder au génome au niveau de chaque signal d'initiation de transcription.

- Modèle dit « Start –Stop » qui semble le plus probable, selon lequel la transcription serait initiée au niveau d'un promoteur unique. L'accès au gène suivant dépend de la terminaison de la transcription du gène précédent (Emerson, 1982).

Le promoteur pour la transcription est situé au niveau de la région leader du génome. Les résidus 19-29 et 34-46 situés dans cette région leader sont essentiels pour la transcription (Whelan and Wertz, 2002).

Dans le génome, l'ordre des gènes détermine directement leur taux d'expression. La transcription de chaque gène ne se fait pas de manière indépendante, elle est régulée par un promoteur unique situé en 3' (Wertz et al., 2002). Les gènes proches du promoteur sont transcrits en quantité importante contrairement aux gènes éloignés du promoteur qui le sont moins. Le gène de la nucléoprotéine est le plus proche du promoteur. La production de nucléoprotéines doit être efficace et doit se faire en quantité stœchiométrique pour encapsider la totalité des génomes viraux. A l'inverse, le gène produisant l'ARN-polymérase-ARN-dépendante, nécessaire en quantité catalytique, est situé en position distale (Wertz et al., 1998) (Figure 9).

L'atténuation de la transcription ne se fait pas au milieu du gène mais au niveau des régions intergéniques. Grace à des mesures cinétiques de l'étape de transcription, il a été montré qu'il existait des pauses de quelques minutes (2.5 à 5.7 minutes) au niveau des régions intergéniques. Ces pauses se produisent durant les étapes de polyadénylation et la synthèse de la coiffe sur les ARNm viraux induisant une transcription discontinue du génome. Ceci a été montré en premier pour VSV (Iverson and Rose, 1981). En effet, au niveau de chaque jonction, la transcription ne redémarre pas dans 30 % des cas. Ceci aboutit à l'accumulation des ARN messagers suivant un gradient décroissant fonction de l'éloignement du gène de l'extrémité 3' du génome (Figure 9). La régulation du niveau d'expression par gradient de transcription est une caractéristique générale au sein de l'ordre des *Mononegavirales*.

Néanmoins, il existe des variations en termes de régulation de la transcription, en particulier dans le cas du virus de la rage. Les régions intergéniques du génome du VSV sont toutes semblables (2nt) alors que dans le cas du virus de la rage, ces régions sont variables en taille (Conzelmann et al., 1990; Tordo et al., 1986b) (Figure 8, Figure 9). Ceci a des conséquences

sur le taux de transcription des gènes, en particulier sur celui de l'ARN-polymérase situé en position distale, et permet au virus d'affiner le ratio de protéines produites.

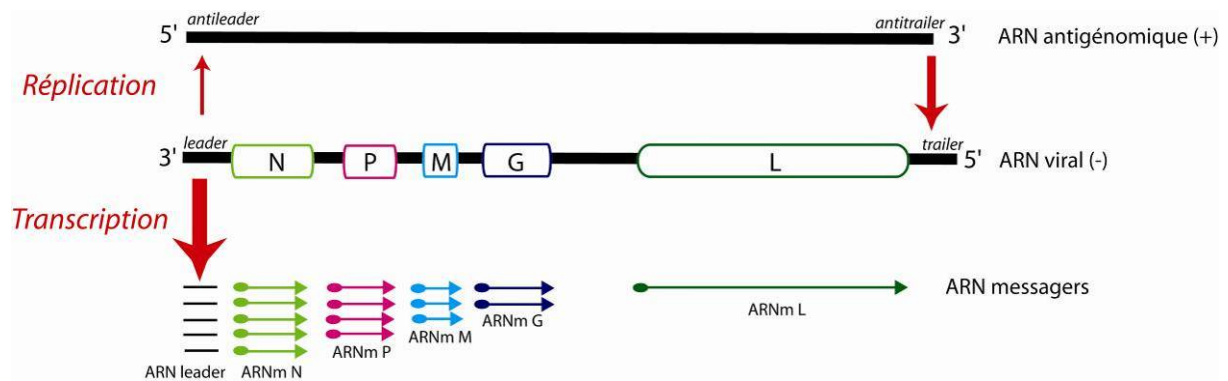


Figure 9 : Représentation schématique de la réplique et la transcription du virus de la rage.

Au cours de la transcription, l'ARN leader et les ARNm codant pour les protéines virales s'accumulent suivant un gradient de concentration décroissant. L'étape de réplique produit un ARN antigénomique de longueur totale. Les flèches rouges indiquent l'activité relative des promoteurs pour la réplique et la transcription.

La fixation de la coiffe des ARN messagers viraux se fait par l'intermédiaire d'une réaction propre au VSV (Abraham et al., 1975a ; Abraham et al., 1975b; Moyer et al., 1975). Il a été prouvé que les phosphates α et β de la coiffe proviennent d'un donneur GDP. Ceci est différent dans le cas des ARN cellulaires pour lesquels la fixation de la coiffe s'effectue habituellement *via* le transfert d'un GMP par la guanydyletransferase sur le 5' di-phosphate de l'ARNm neo-synthétisé. Combiné au fait que la synthèse des ARNm viraux se déroule dans le cytoplasme, ceci a permis de mettre en évidence l'activité de « capping » de la sous-unité L du complexe ARN-polymérase viral.

Après l'étape de transcription, la traduction protéique s'effectue en utilisant la machinerie cellulaire de la cellule hôte.

C. La réplique virale

Durant la réplique, l'ARN-polymérase-ARN-dépendante virale adopte un mode processif ignorant les signaux de transcription (fin et début de gène). Le processus de réplique débute lorsque les protéines virales ont déjà été synthétisées (après la traduction primaire). La réplique virale ne conduit pas à la synthèse d'un ARN libre mais à la production d'un complexe ribonucléoprotéique au sein duquel l'ARN antigénomique total est encapsidé par la nucléoprotéine. La réplique nécessite un apport constant en protéines N et P. En effet la polymérisation de l'ARN et son encapsidation sont deux mécanismes associés. La formation concomitante de complexes entre la nucléoprotéine et phosphoprotéine et de complexes entre

la polymérase virale et la phosphoprotéine sont nécessaires pour encapsider la chaîne naissante (Horikami et al., 1992).

Le passage de la transcription à la réplication est contrôlé par les protéines virales. Plus précisément, un certain taux de protéines N et P est nécessaire pour que le génome viral soit répliqué et non transcrit. La manière dont le passage de la phase de transcription à la phase de réplication se produit est toujours sujette à spéculations.

Le passage de l'ARN-polymérase-ARN-dépendante du mode transcriptase à réplique est corrélé à une forte concentration de protéines N à l'intérieur de la cellule infectée. Lorsque la quantité de nucléoprotéines dans la cellule dépasse un seuil critique, ceci autorise l'encapsidation de l'ARN leader au fur et à mesure de sa synthèse. La nucléoprotéine jouerait alors le rôle «d'anti-terminateur». En masquant les signaux de fin et début de transcription, la nucléoprotéine force la polymérase virale à conserver un cheminement processif lors de la réplication résultant en la synthèse d'un ARN antigénomique de polarité positive.

Cette molécule d'ARN antigénomique sert ensuite de matrice pour la synthèse de nouveaux ARN viraux. Les extrémités 3' et 5' du génome viral étant partiellement complémentaires, l'ARN-polymérase-ARN-dépendante virale peut amorcer la synthèse d'ARN au niveau de plusieurs promoteurs (Figure 9) :

- Au niveau de la région leader du génome pour la transcription et la réplication.
- Au niveau de l'anti-trailer de l'antigénome pour la réplication du brin positif.

La polymérase virale copie l'antigénome de façon processive, la séquence trailer constitue à son tour un signal d'encapsidation.

De façon logique, pour accentuer la production de virus, la réplication se fait de manière asymétrique. Elle produit 10 fois plus de génomes que d'antigénomes (Whelan and Wertz, 2002). En effet, la région leader est un promoteur moins fort que le promoteur situé sur la région anti-trailer (Whelan and Wertz, 1999).

III. Les protéines virales

A. Structures impliquées dans l'entrée et dans le bourgeonnement du virion

L'enveloppe virale de la rage est constituée de 50 % de protéines et 50 % de lipides. Deux protéines sont impliquées dans la formation de l'enveloppe du virus : la glycoprotéine transmembranaire qui est ancrée dans la membrane virale et la protéine de matrice localisée à l'intérieur de la particule virale.

1) La Glycoprotéine (G)

La glycoprotéine du virus de la rage est une protéine membranaire intrinsèque qui forme des spicules trimériques facilement visibles à la surface du virus en microscopie électronique (Figure 2 A) (Gaudin et al., 1992). Environ 400 spicules de glycoprotéines sont présentes sur la membrane plasmique du virus de la rage (Flamand et al., 1993).

La glycoprotéine contient les sites antigéniques responsables de la réponse immunitaire, capable d'induire la production d'anticorps. Elle joue un rôle décisif lors de l'établissement de l'immunité antirabique.

La glycoprotéine est reconnue par les récepteurs de la cellule hôte. Les récepteurs du virus de la rage reconnus par la glycoprotéine doivent être présents à la surface des nombreux types cellulaires impliqués lors de l'infection virale : cellules musculaires et cutanées, neurones, cellules glandulaires (glandes salivaires et cerveau). Ce récepteur n'a pas été identifié avec certitude pour l'instant. Cependant, plusieurs propositions de récepteurs cellulaires ont été faites. Les récepteurs nicotiques joueraient un rôle important dans l'infection des cellules musculaires et neuronales par le virus de la rage. Le récepteur N-acétylcholine avait également été proposé en se basant sur l'homologie existant entre la région 189-214 de la glycoprotéine et des neurotoxines de serpents (Lentz et al., 1987); (Gastka et al., 1996). Deux autres protéines ont été plus récemment identifiées comme récepteur potentiel : la molécule d'adhésion des cellules nerveuses (NCAM) (Thoulouze et al., 1998) et le récepteur de la neurotrophine p75 (Tuffereau et al., 1998a; Tuffereau et al., 1998b).

La protéine G est aussi responsable du phénomène de fusion membranaire entre la membrane du virus et celle de l'endosome. La glycoprotéine est produite sous forme d'un précurseur à l'intérieur du réticulum endoplasmique rugueux puis exportée à la surface de la cellule hôte par voie de sécrétion. Les 19 premiers acides aminés formant le peptide signal sont clivés à l'intérieur du réticulum endoplasmique, donnant naissance à une glycoprotéine virale mature de 505 acides aminés. La glycoprotéine est composée de trois domaines :

- La partie N-terminale (439 résidus) de la glycoprotéine constitue l'ectodomaine, qui est exposé à la surface du virus. L'ectodomaine contient les sites de glycosylation et la plupart des déterminants de l'immunogénicité.

- Le segment transmembranaire est composé de 22 acides aminés majoritairement hydrophobes. Ce domaine permet l'ancrage de la protéine dans la membrane cellulaire et virale.

-L'extrémité C-terminale de la protéine (44 résidus) constitue un segment cytoplasmique hydrophile qui peut interagir avec les autres protéines virales internes.

Au cours de sa maturation, la glycoprotéine du virus de la rage subit plusieurs modifications post-traductionnelles au niveau de son ectodomaine. Elle est glycosylée et acylée à l'intérieur de l'appareil de Golgi (Gaudin et al., 1991a). La glycosylation se fait sur des résidus asparagine (4 ou 5 selon la souche de rage), les résidus asparagine 319 et 37 étant les sites de glycosylation majeurs. Les autres sites de glycosylation sont variables suivant les souches de rage.

La glycoprotéine du virus de la rage existe sous trois conformations (Gaudin et al., 1991b) : la forme native présente à la surface du virus ($\text{pH} \geq 7$), une forme active permettant à la glycoprotéine de s'ancrer dans la membrane cellulaire pour amorcer la fusion membranaire et enfin une forme inactive, post-fusion ($\text{pH} < 6.4$).

Récemment, la structure cristallographique de la glycoprotéine du VSV dans sa conformation de post-fusion a été résolue (Roche et al., 2006) (Figure 10).

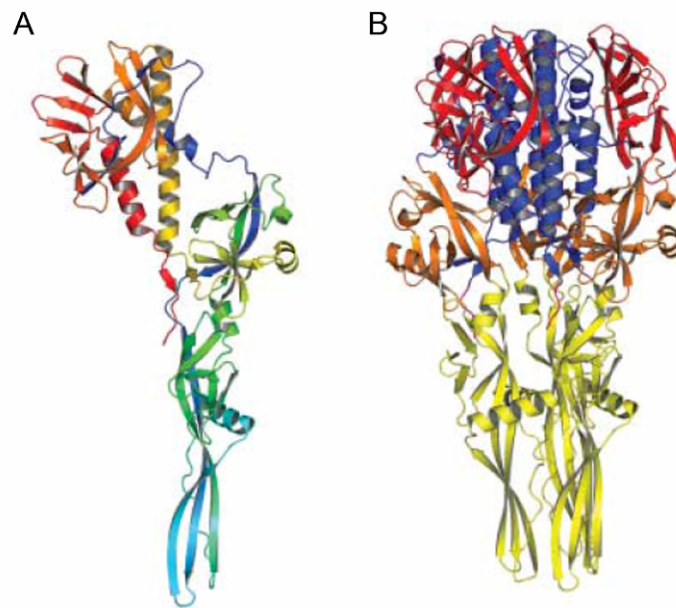


Figure 10 : Structure cristallographique de la protéine G du VSV à bas pH (Roche et al., 2006).

A : Représentation d'un seul protomère. Gradient de couleur : du bleu (N-terminal) au rouge (C-terminal)

B : Représentation du trimère de glycoprotéine, les différentes couleurs correspondent aux différents domaines de la protéine.

Le repliement protéique de cette structure combine des caractéristiques des protéines de fusion de classe I et de classe II. Cette structure présente une conformation en feuillets β caractéristique des protéines de fusion de classe II dans leur conformation de post-fusion. Le

domaine impliqué dans la multimérisation de la protéine présente une organisation en hélices α caractéristique des protéines de fusion de classe I.

La glycoprotéine du VSV présente des similitudes structurales avec la glycoprotéine B du virus Herpes Simplex 1 (famille des Herpesvirus dont le génome est un double brin d'ADN) (Heldwein et al., 2006). Ces deux structures, bien que leurs séquences en acides aminés soient différentes, sont organisées de façon identique. Ceci suggère que ces deux protéines ont un ancêtre commun et que les *Mononegavirales* sont capables d'acquérir un gène en le volant à leur hôte cellulaire.

2) La protéine Matrice (M)

La protéine matrice du virus de la rage est composée de 202 acides aminés (229 acides aminés chez le virus du VSV). C'est la plus petite des protéines virales. Elle est phosphorylée dans le cas du VSV et palmitoylée pour le virus de la rage (Gaudin et al., 1991a). Les protéines matrice sont globalement basiques, elles comportent des domaines hydrophobes mais elles ne sont ni insérées dans la membrane lipidique ni synthétisées à l'intérieur du réticulum endoplasmique. Pour le VSV, environ 1800 exemplaires de la protéine matrice sont présents au sein d'un virion (Thomas et al., 1985). Il a été montré pour VSV et le virus de Sendai que la protéine M interagissait de manière simultanée avec la membrane et la nucléocapside (Chong and Rose, 1993; Stricker et al., 1994). La protéine de matrice joue un rôle prépondérant dans le bourgeonnement des virus à ARN négatif et facilite la condensation hélicoïdale de la nucléocapside (Newcomb et al., 1982).

La position de la protéine matrice au sein de la structure du virion est cependant toujours sujette à spéculations. Il est possible que la protéine M fasse le lien entre la nucléocapside et la membrane virale au sein du virus (Pal et al., 1987). Cette association est due à une interaction entre la protéine M et les têtes hydrophiles de lipides chargés négativement (phosphatidylsérines). La protéine matrice a une forte affinité pour elle-même : à une concentration élevée, cette protéine a tendance à s'associer pour former des polymères (McCreedy et al., 1990). Il est également possible que les protéines M forment une sorte de squelette protéique en forme de cigare autour duquel la nucléocapside s'enroulerait en spirale (Barge et al., 1993) (Figure 3). Ces deux hypothèses sont compatibles avec les caractéristiques physico-chimiques de la protéine M.

Les structures atomiques de plusieurs parties de protéines matrices de virus à ARN négatif sont connues à ce jour (Figure 11) : dans le cas du VSV (Gaudier et al., 2002), du virus de la grippe M1 (Sha and Luo, 1997), et du virus Ebola VP40 (Dessen et al., 2000).

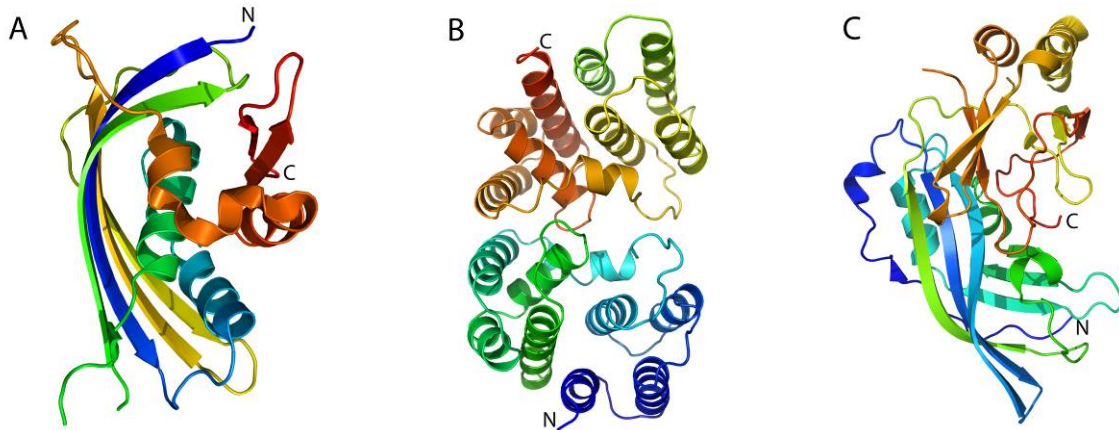


Figure 11 : Données structurales sur les protéines de matrice des virus à ARN négatif.

Gradient de couleur : du bleu (N-terminal) au rouge (C-terminal)

A : Structure de la protéine de matrice M du VSV, fragment résistant à la thermolysine (résidus 58 à 121 et résidus 128 à 227, code pdb : 1LG7).

B : Structure de la protéine de matrice M1 du virus de la grippe (résidus 2 à 158, code pdb : 1AA7).

C : Structure de la protéine de matrice VP40 du virus Ebola (résidus 31 à 326, code pdb : 1ES6).

Comme il est visible sur la Figure 11, les structures secondaires et tertiaires des protéines matrices diffèrent beaucoup selon les virus. Ceci pourrait expliquer pourquoi ces virus ont des morphologies variées (Tableau 1, Figure 1).

La protéine M joue aussi un rôle de régulation dans la cellule hôte. Dans le cas du VSV ainsi que d'autres Vesiculovirus, il a été montré que la protéine M inhiberait la synthèse des ARNm cellulaires (Her et al., 1997). La présence la protéine matrice dans le noyau bloquerait le transport bidirectionnel des protéines et des ARN cellulaires entre le noyau et le cytoplasme des cellules infectées (Glodowski et al., 2002). La protéine matrice du virus de la rage est aussi impliquée dans la régulation de la balance transcription/réplication; la protéine M inhiberait la transcription et stimulerait la réplication virale (Finke and Conzelmann, 2003; Finke et al., 2003)

B. Structures impliquées dans la multiplication du virus et la synthèse d'ARN

1) La Phosphoprotéine P

Tous les *Mononegavirales* possèdent une phosphoprotéine. Les séquences, tailles et degrés d'oligomérisation des phosphoprotéines de ces différents virus sont très variables (568 résidus pour le virus de Sendai, 297 résidus pour le virus de la rage). Malgré tout, les phosphoprotéines ont plusieurs caractéristiques communes :

- la protéine P a un rôle de molécule chaperonne pour la nucléoprotéine empêchant ainsi sa liaison aux ARNs cellulaires (cette forme de nucléoprotéine est dite « soluble », elle est notée N°).
- la protéine P se fixe sur le complexe nucléoprotéines-ARN (nucléocapside).
- la protéine P se lie à l'ARN-polymérase virale ARN-dépendante (L) et forme le complexe ARN-polymérase virale.
- toutes les phosphoprotéines étudiées à ce jour forment des homo-oligomères.

Ces caractéristiques se retrouvent au sein de la structure de la phosphoprotéine (Figure 12 A). La phosphoprotéine du virus de la rage présente deux sites d'interactions avec la nucléoprotéine. Le peptide minimal nécessaire à la formation d'un complexe nucléoprotéine-phosphoprotéine soluble a été identifié (Mavrakis et al., 2006). Les résidus 4 à 40 de la phosphoprotéine suffiraient à garder soluble le complexe formé entre la nucléoprotéine et la phosphoprotéine.

Un second site de fixation de nucléoprotéine existe sur la phosphoprotéine du virus de la rage, il s'agit de l'extrémité C-terminale (résidus 185 à 297). Cette partie de la protéine P interagit spécifiquement avec les complexes nucléoprotéines-ARN (Chenik et al., 1994; Jacob et al., 2000).

L'interaction avec l'ARN-polymérase virale L se produit au niveau des résidus 1 à 19 de l'extrémité N-terminale de la phosphoprotéine (Chenik et al., 1998).

Le domaine d'oligomérisation de la phosphoprotéine se situe entre les résidus 93 et 128 (Gigant et al., 2000; Mavrakis et al., 2004).

La phosphoprotéine est présente dans la cellule infectée ainsi qu'au sein du virion avec plusieurs degrés de phosphorylation. Concernant la phosphoprotéine du virus de la rage, deux kinases cellulaires sont impliquées dans cette phosphorylation. La PKC (protéine kinase C) et une autre kinase spécifique, la RVPK (Rabies Virus Protein Kinase) (Gupta et al., 2000). La phosphoprotéine du VSV est quant à elle phosphorylée par la caséine kinase II et par une

protéine kinase associée à l'ARN-polymérase virale (L) (Chattopadhyay and Banerjee, 1987; Chen et al., 1997).

La structure cristallographique du domaine C-terminal de la phosphoprotéine du virus de la rage a été résolue (Mavrakis et al., 2004). Ce domaine fixe la matrice nucléoprotéine-ARN (Figure 12 C). Le fragment cristallisé contient deux sites potentiels de phosphorylation : les sérines 210 et 271. Selon cette structure, la sérine 271 ne peut pas être accessible à la kinase RVPK. Il est donc possible que la phosphorylation de cette sérine soit liée à un changement de conformation de la phosphoprotéine permettant sa fixation à la matrice nucléoprotéine-ARN.

La structure cristallographique du domaine de dimérisation de la phosphoprotéine du VSV a été résolue (Ding et al., 2006) (Figure 12 B). Ce domaine est résistant à la protéolyse par la protéase K (Ding et al., 2004). Il est composé de deux brins β et d'une hélice α . La structure tridimensionnelle a révélé que des contacts hélices α - hélices α sont à l'origine de la formation du dimère. En se basant sur ces données structurales, on peut penser que l'entité biologique active de la phosphoprotéine du VSV est un dimère.

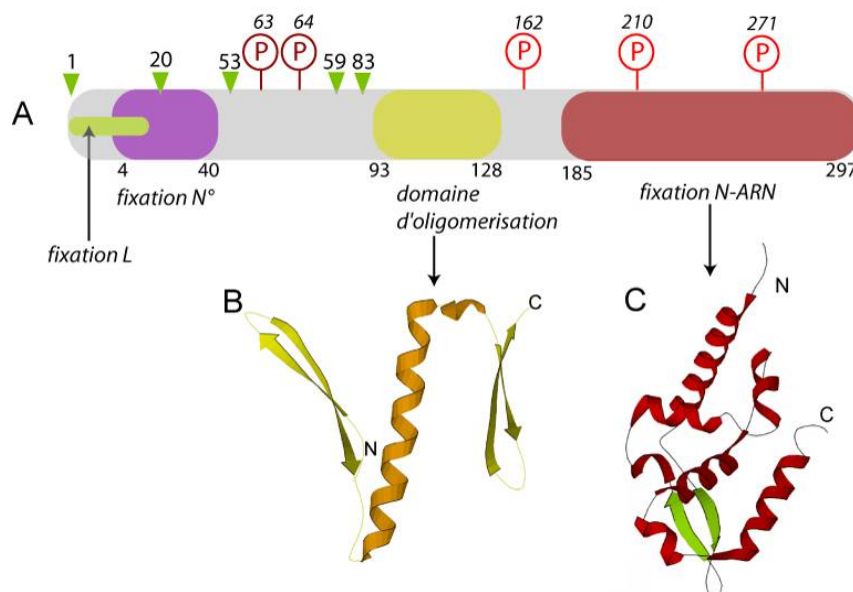


Figure 12 : La phosphoprotéine du virus de la rage.

A : Les différents domaines de la phosphoprotéine du virus de la rage.

La PKC phosphoryle les sérines en positions 63 et 64, la RVPK phosphoryle les sérines 162, 210 et 271 (Gupta et al., 2000).

La position des codons AUG des différents produits du gène de P (P complète et les sous-produits P2, P3, P4 et P5) est indiquée par des triangles verts.

B : Structure cristallographique du domaine d'oligomérisation de la phosphoprotéine du VSV (Ding et al., 2006).

C : Structure cristallographique du domaine C-terminal de la phosphoprotéine du virus de la rage (Mavrakis et al., 2004).

Récemment, dans notre laboratoire, une étude approfondie des phosphoprotéines du virus de la rage et du VSV a été menée en utilisant différentes techniques biophysiques (Gerard et al., 2006). Il a alors été montré grâce à des expériences de diffusion de lumière et de chromatographie d'exclusion de taille que les phosphoprotéines du virus de la rage et du VSV sont des dimères en solution. Ces deux protéines ont une forme allongée, et comportent plusieurs régions désordonnées. En se basant sur des analyses bioinformatiques, il a été montré que ces phosphoprotéines sont composées de trois domaines structurés et de deux régions désordonnées. Ceci suggère que les domaines des phosphoprotéines de Rhabdovirus pourraient être organisés d'une manière similaire (Figure 13).

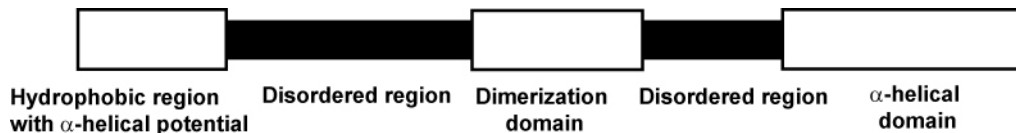


Figure 13 : Organisation générale de la phosphoprotéine des Rhabdovirus (Gerard et al., 2006). Les régions structurées sont représentées par des cadres blancs, les régions désordonnées sont en noir.

Interaction de la phosphoprotéine de la rage avec des partenaires cellulaires :

Des expériences d'immuno-précipitation ont permis de détecter la présence de 4 sous-produits du gène de la phosphoprotéine de la rage tronqués en N-terminal (Chenik et al., 1995) (Figure 12 A). Le gène de la phosphoprotéine est polycistronique, il code pour une protéine P ainsi que pour 4 sous-produits P2, P3, P4 et P5. Ces protéines sont issues d'une initiation de la traduction au niveau de codons AUG situés en phase à l'intérieur du gène de la protéine P par un mécanisme de « leaky scanning » du ribosome (Kozak, 1989). 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 localisés majoritairement au niveau du noyau des cellules infectées.

La protéine PML (promyelocytic leukaemia) est une protéine induite par l'interféron (α , β et γ). La protéine PML se trouve à l'intérieur du noyau cellulaire, ou associée à des corps nucléaires. L'interaction des produits du gène de la phosphoprotéine avec la protéine PML provoque une réorganisation des corps nucléaires. Ceci suggère que la phosphoprotéine joue un rôle important pour contrer la réponse immunitaire induite par l'interféron (Blondel et al., 2002).

La protéine STAT1 (*Signal Transducer and Activator of Transcription 1*) interagit avec la partie C-terminale de la phosphoprotéine. Cette interaction empêche l'accumulation de

STAT1 dans le noyau et bloque ainsi la réponse interféron (Brzozka et al., 2005; Vidy et al., 2005).

Des études de co-localisation en microscopie à fluorescence ont montré que la phosphoprotéine pouvait interagir de manière forte avec la chaîne légère de la dynéine (LC8), (Jacob et al., 2000). La dynéine est un moteur moléculaire impliqué dans les mouvements intracellulaires tels que le transport rétrograde le long des axones. Cette interaction se fait au niveau des résidus 140-150 de la phosphoprotéine (Poisson et al., 2001). L'interaction entre la phosphoprotéine et la chaîne légère de la dynéine suggère que la phosphoprotéine joue un rôle dans le transport rétrograde axonal de la particule virale dans le système nerveux central (Figure 14).

La phosphoprotéine n'est donc pas uniquement une protéine structurale du virus, elle possède de multiples fonctions aussi bien dans la réplication du virus que dans le transport ou dans la lutte contre la réponse immunitaire de la cellule (Brzozka et al., 2005).

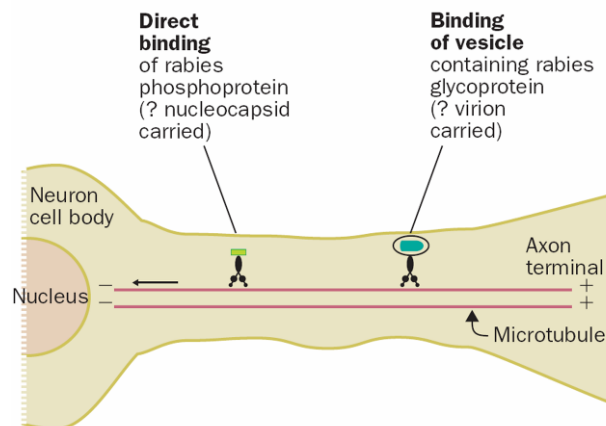


Figure 14 : Le transport viral rétrograde axonal (Warrell and Warrell, 2004).

La propagation du virus le long des axones pourrait se faire deux manières : soit en transportant une particule ribonucléoprotéique, soit en transportant le virus dans des vésicules.

2) La Nucléoprotéine (N)

La masse moléculaire des nucléoprotéines des virus à ARN négatif est généralement comprise entre 50 et 60 kDa. La nucléoprotéine des Filoviridae a une masse moléculaire bien plus importante (80 kDa) (Sanchez et al., 1992). Il n'y a que peu d'homologies de séquence entre les nucléoprotéines des différents virus à ARN négatif. La moitié N-terminale de la nucléoprotéine fixe l'ARN viral (Buchholz et al., 1993) et, chez la plupart des *Mononegavirales*, la partie C-terminale interagit avec la phosphoprotéine. Pour le virus de la

rougeole, il a été proposé que la partie C-terminale de la nucléoprotéine puisse être dans un état intrinsèquement désordonné (Longhi et al., 2003). Le repliement de ce fragment de la nucléoprotéine serait induit par le contact avec son partenaire, la phosphoprotéine. A la différence de celui de la rougeole, le domaine C-terminal de la nucléoprotéine du virus de la rage semble former une structure bien définie (Schoehn et al., 2001).

Dans les cellules infectées, la nucléoprotéine peut se trouver sous deux formes différentes : soit liée à l'ARN et formant ainsi la matrice ARN-nucléoprotéine utilisée pour la transcription et la réplication, soit « soluble », liée à phosphoprotéine formant un complexe nucléoprotéine-phosphoprotéine (N°-P).

La nucléoprotéine du virus de la rage est composée de 450 acides aminés (Figure 15). A l'inverse de celle du VSV, la nucléoprotéine du virus de la rage est phosphorylée au niveau de la sérine 389 par la caséine kinase II lorsqu'elle est engagée dans les complexes nucléoprotéine-ARN (Kawai et al., 1999).



Figure 15 : Les différents domaines de la nucléoprotéine du virus de la rage.

Les domaines de fixation de l'ARN (Kouznetzoff et al., 1998) et de la phosphoprotéine (Schoehn et al., 2001) sont indiqués. Le site de phosphorylation par la caséine kinase II se situe sur le résidu sérine 389.

La partie C-terminale (résidus 375-450) de la nucléoprotéine est indispensable à la fixation de la phosphoprotéine (Schoehn et al., 2001). La suppression de ce domaine abolit la formation de complexes nucléoprotéine-phosphoprotéine. Des expériences de « UV-LASER cross-link » ont prouvé que les 376 premiers acides aminés de la protéine N interagissent avec l'ARN viral (Kouznetzoff et al., 1998). Des expériences de chromatographie d'affinité ont permis l'identification d'un peptide de nucléoprotéine (résidus 298-352 de N) pouvant lier l'ARN viral. La comparaison de la séquence de ce peptide avec celles d'autres protéines liant l'ARN n'a révélé aucune homologie. Ceci suggère fortement que la nucléoprotéine s'associe avec l'ARN en utilisant un motif de fixation inconnu.

(a) Le complexe soluble Nucléoprotéine-Phosphoprotéine (N^o-P):

Comme mentionné précédemment, la nucléoprotéine du virus de la rage peut exister sous une forme dite « soluble », libre de tout ARN à condition d'être en complexe avec la phosphoprotéine. Il est possible de produire un complexe nucléoprotéine-phosphoprotéine recombinant en co-exprimant les deux protéines dans le système baculovirus-cellules d'insectes.

Il a été montré par ultracentrifugation analytique, par spectrométrie de masse et par microscopie électronique que ce complexe N^o-P est constitué d'une nucléoprotéine associée à deux phosphoprotéines (Mavrakis et al., 2003).

La nucléoprotéine contenue dans le complexe N^o-P n'est pas phosphorylée au niveau de la sérine 389 (Kawai et al., 1999). A l'aide d'anticorps épitope-spécifiques, Kawai et al., ont mis en évidence le fait que la nucléoprotéine neo-synthétisée n'est pas immédiatement phosphorylée mais qu'elle s'associe en premier avec la phosphoprotéine pour former le complexe N^o-P. Le complexe N^o-P participe ensuite sous cette forme à la formation de complexes nucléoprotéines-ARN. La nucléoprotéine non-phosphorylée lie plus efficacement la matrice ARN (Yang et al., 1998; Yang et al., 1999). La protéine N serait alors phosphorylée durant la phase d'encapsidation. Il est fort probable que la phosphorylation soit induite par un changement de conformation de la nucléoprotéine une fois liée à l'ARN. Cette phosphorylation est importante lors des phases de réplication et de transcription.

La nucléoprotéine du virus de la rage joue un rôle important dans la régulation de la transcription et la réplication et ceci probablement en modulant l'encapsidation de l'ARN (Yang et al., 1999).

(b) Les complexes N-ARN :

Associée à l'ARN viral pour former la nucléocapside, la nucléoprotéine protège l'ARN contre une digestion par des ribonucléases (RNases) et empêche la formation de structures secondaires au niveau des ARNs simples brins (Baudin et al., 1994; Iseni et al., 2000). La nucléoprotéine n'interagit pas avec l'ARN d'une manière classique. Les nucléocapsides des Rhabdovirus et Paramyxovirus sont très stables, elles peuvent supporter de fortes concentrations de sel et une force de gravité importante lors de longues ultracentrifugations (Blumberg et al., 1984; Heggeness et al., 1980). La formation d'un tel complexe entre la

nucléoprotéine et l'ARN est également nécessaire pour éviter la formation de structure ARN double brin entre l'ARN viral et les ARN messager lors de leur synthèse.

L'interaction entre l'ARN et la nucléoprotéine est très forte. Pour VSV, il a été montré que cette interaction résistait à un traitement dénaturant avec 8 M urée (Iseni et al., 1998).

La nucléoprotéine des virus à ARN négatif se fixe sur un nombre entier de nucléotides. Des reconstructions par microscopie électronique (pour le virus de Sendai et VSV) et des mesures d'absorbances à 280 et 260 nm (pour le virus de Marburg) ont permis de déterminer le nombre de nucléotides fixés par une nucléoprotéine pour plusieurs virus à ARN négatif. Ce nombre est différent et spécifique pour chaque famille de *Mononegavirales* : 6 nucléotides par nucléoprotéine pour les *Paramyxoviridae* (Egelman et al., 1989) et entre 12 et 15 nucléotides par nucléoprotéine pour les *Filoviridae* (Mavrakis et al., 2002). La nucléoprotéine des Rhabdovirus lie quant à elle 9 nucléotides (Flamand et al., 1993; Thomas et al., 1985). Pour les virus de la famille des *Paramyxoviridae*, ce ratio est extrêmement important. Il existe une règle appelée « la règle de six » (Calain and Roux, 1993), selon laquelle le nombre exact de nucléotides du génome du virus doit être un multiple de six pour que la polymérase virale réplique efficacement le génome viral. Cela signifie aussi que chaque N interagit avec exactement six nucléotides et que les points d'interaction de la nucléoprotéine avec les six nucléotides ne sont pas équivalents (Vulliamoz and Roux, 2001). La « règle de six » s'applique à tous les *Paramyxoviridae* sauf à ceux de la famille des Pneumovirus comme le virus respiratoire syncytial humain. En revanche, il n'a pas été mis en évidence de « règle de neuf » pouvant s'appliquer aux virus de la famille des Rhabdovirus.

Toutes les nucléocapsides des virus à ARN négatif ont une symétrie hélicoïdale mais leurs paramètres hélicoïdaux sont variables. La Figure 16 montre des clichés de microscopie électronique de nucléocapsides isolées à partir de différents virus. L'expression des nucléoprotéines des *Mononegavirales* en utilisant différents systèmes d'expression (eucaryotes ou bactériens) produit des complexes dont la morphologie est identique à celle des nucléocapsides virales. Les nucléoprotéines recombinantes s'associent de manière irréversible aux ARNs cellulaires pour former des complexes présentant la même stoechiométrie (nombre de nucléotides liés par une protéine N) que les nucléocapsides virales. La forme générale de la nucléoprotéine est relativement similaire pour ces virus (Figure 17), présentant dans tous les cas une forme en haricot avec deux domaines distincts.

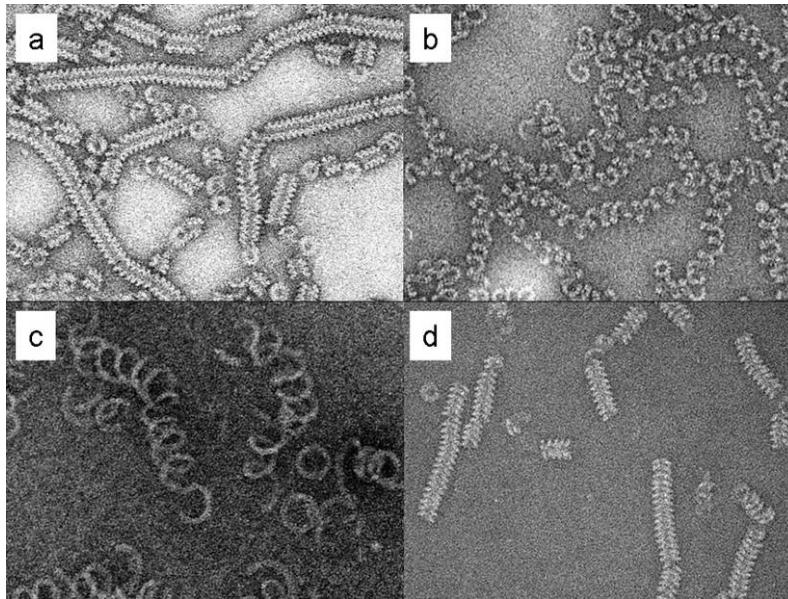


Figure 16 : Echantillons de nucléocapsides de *Mononegavirales* visualisés en microscopie électronique.

a- Nucléocapsides recombinantes du virus de la rougeole (*Paramyxoviridae*) b- Nucléocapsides virales du virus de la rage (*Rhabdoviridae*), c- Nucléocapsides recombinantes du virus de Marburg (*Filoviridae*), d- Nucléocapsides recombinantes du virus Nipah (*Henipahviridae*).

Le nombre de points de contacts entre deux protomères successifs de nucléoprotéine est une caractéristique remarquable chez les virus à ARN négatif. La nucléoprotéine du virus de la grippe n'établit qu'un seul contact avec son voisin et sa nucléocapside ne présente pas de structure hélicoïdale mais ressemble plutôt à une torsade. Les nucléoprotéines des virus de la rage et de la rougeole établissent deux contacts. Les contacts de la nucléoprotéine du virus de la rage sont situés aux deux extrémités de la protéine induisant une forme hélicoïdale lâche. En comparaison, la nucléocapside de la rougeole est plus régulière avec deux points d'interaction situés au niveau de la partie interne de l'hélice donnant une structure hélicoïdale plus régulière caractéristique de la famille des *Paramyxoviridae*. Des reconstructions hélicoïdales effectuées à partir d'images de microscopie électronique ont permis de déterminer le nombre de sous-unités de nucléoprotéines par tour d'hélice pour certains de ces virus. Pour le virus de la rougeole et le virus de Sendai, le nombre de protéines N par tour est de 13 (Egelman et al., 1989; Schoehn et al., 2004). Concernant le virus respiratoire syncytial (RSV) et le Virus Simien 5 (SV5), 10 à 14 nucléoprotéines seraient nécessaires pour réaliser un tour d'hélice (Bhella et al., 2002).

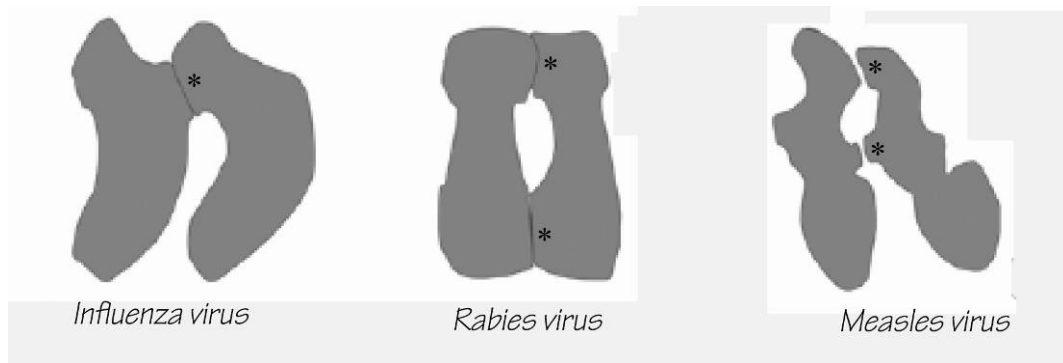


Figure 17 : Comparaison des nucléoprotéines de différents virus à ARN négatif (Schoehn et al., 2004).

Les points de liaison entre deux protomères voisins de nucléoprotéine sont signalés par une *.

A : Le virus Influenza, une seule zone d'interaction. B : Le virus de la rage, deux zones d'interactions.

C : Le virus de la rougeole, deux zones d'interactions proches.

L'hétérogénéité et la flexibilité de ces nucléocapsides excluent totalement leur cristallisation. La détermination de structures par cristallographie est donc impossible avec les échantillons de nucléocapsides. Les techniques d'analyses d'images utilisées en microscopie électronique nécessitent également des objets homogènes et réguliers et ne sont donc pas utilisables pour les nucléocapsides purifiées à partir de virus (sauf les nucléocapsides digérées de virus de la rougeole).

Les nucléocapsides des Rhabdovirus sont très irrégulières et ceci encore plus que celles des Paramyxovirus. Les nucléocapsides de Rhabdovirus sont par conséquent inutilisables pour réaliser une reconstruction tridimensionnelle à partir d'images obtenues en microscopie électronique.

Lorsque l'on exprime la nucléoprotéine du virus de la rage ou du VSV dans des cellules d'insectes ou dans des bactéries, une partie des protéines synthétisées se fixent à de longs ARN, probablement des ARN messagers de la cellule, menant à la formation de nucléocapsides recombinantes (Figure 18 B). D'autres nucléoprotéines peuvent s'associer à des ARNs cellulaires plus courts, probablement des ARN de transfert (Chen et al., 2004; Schoehn et al., 2001). Ceci mène à la formation des structures nucléoprotéine-ARN en forme d'anneaux avec un nombre de sous-unités de nucléoprotéine variable (de 9 à 15 N), et probablement avec une longueur d'ARN encapsidée également variable (Figure 18 C).

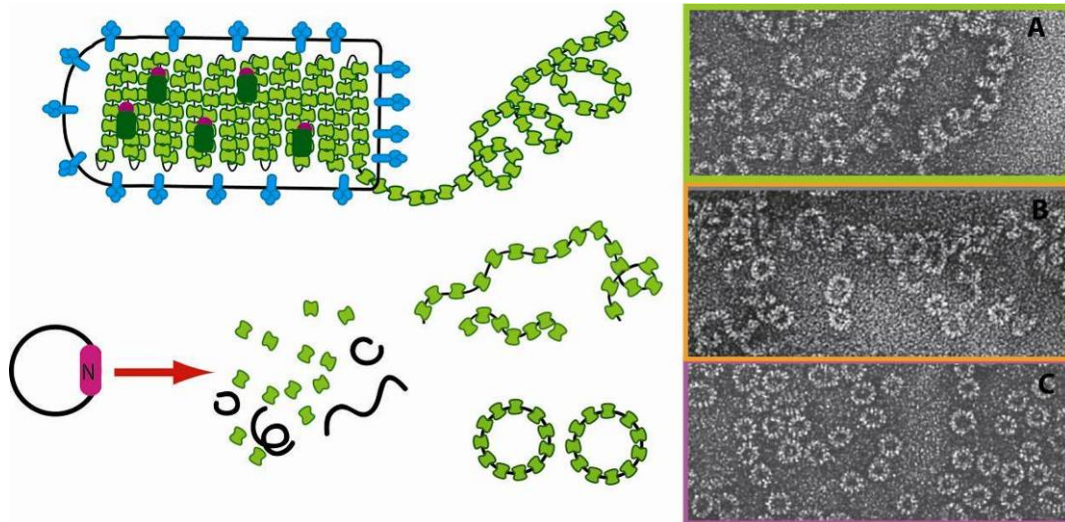


Figure 18 : Les différents complexes ARN-nucléoprotéine du virus de la rage observés en microscopie électronique (coloration négative) (Iseni et al., 1998).

A : Nucléocapsides virales. B : Nucléocapsides recombinantes produites en cellules d'insectes. C : Anneaux issus de l'association entre 9 à 15 sous unités de nucléoprotéine et une molécule d'ARN simple brin.

Une reconstruction en trois dimensions d'anneaux formés de 10 sous-unités de nucléoprotéine a été réalisée à partir d'images prises en cryomicroscopie électronique dans notre laboratoire (Schoehn et al., 2001) (Figure 19 A et B). Les sous-unités de nucléoprotéine ont une forme dite en « graine de haricot » (Figure 17) avec une extrémité supérieure plus volumineuse. Quand ces anneaux (ou les nucléocapsides) sont traités par la trypsine de manière prolongée, un clivage au niveau du résidu 376 de la nucléoprotéine élimine le domaine C-terminal de la protéine N. Ces nucléocapsides digérés avec la nucléoprotéine dépourvue d'extrémité C-terminale ne peuvent plus fixer la phosphoprotéine. Grâce à la détermination de la structure tridimensionnelle de ces anneaux et en la comparant avec la structure des anneaux natifs, le domaine d'attachement de la phosphoprotéine à la nucléoprotéine a pu être localisé dans la partie supérieure de l'anneau (Figure 19 A & B).

Dans le cas du VSV, la co-expression de la nucléoprotéine et de la phosphoprotéine chez *E. coli*, mène à la formation d'anneaux comportant 10 sous-unités de nucléoprotéine. La reconstruction tridimensionnelle de ces anneaux du VSV montre une structure qui est proche de celle des anneaux N-ARN du virus de la rage après digestion à la trypsine (Figure 19 B & C) (Chen et al., 2004; Green et al., 2000).

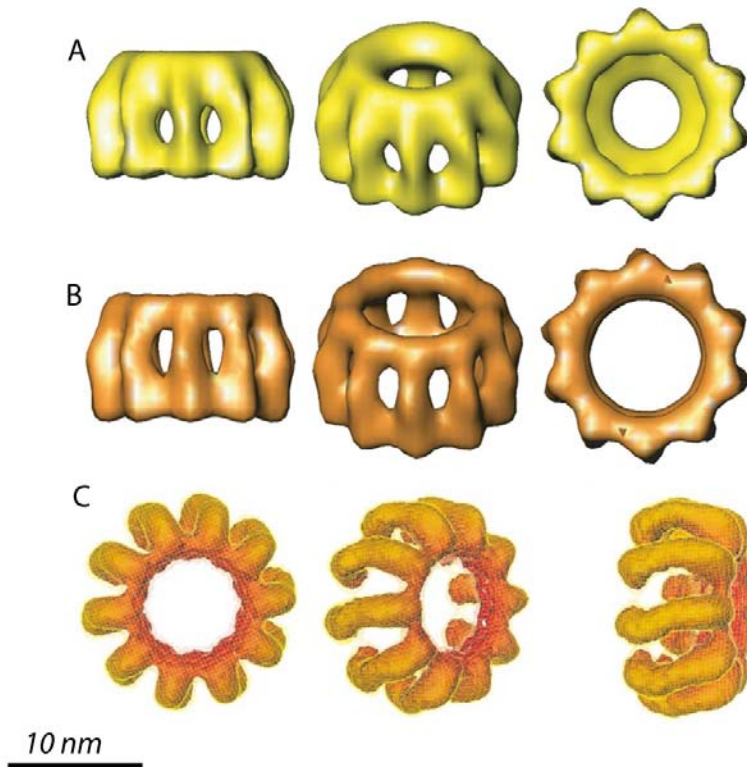


Figure 19 : Reconstructions de complexes nucléoprotéine-ARN en forme d’anneaux.

A : Reconstruction des anneaux nucléoprotéine-ARN du virus de la rage.

B : Reconstruction des anneaux nucléoprotéine-ARN du virus de la rage après protéolyse à la trypsine (Schoehn et al., 2001)

C : Reconstruction des anneaux nucléoprotéine-ARN du VSV (Chen et al., 2004).

(c) La nucléoprotéine du Bornavirus

La structure cristallographique de la nucléoprotéine du Bornavirus, virus à ARN négatif de l’ordre des *Mononegavirales* a été résolue (Rudolph et al., 2003) (Figure 20). Contrairement aux autres *Mononegavirales*, la nucléoprotéine de ce virus pouvait être obtenue sous une forme soluble, et a été cristallisée sans être associée avec un ARN ou à la phosphoprotéine.

La structure tridimensionnelle de cette protéine montre que la nucléoprotéine de ce virus est également composée de deux domaines distincts (Figure 20 A). A l’intérieur des cristaux, les sous-unités de nucléoprotéine sont arrangées en homo-tétramères (Figure 20 B). Les auteurs ont proposé deux modèles pour l’empaquetage de l’ARN : la molécule d’ARN s’enroulerait autour du tétramère de nucléoprotéine mais pourrait alternativement se faufiler à l’intérieur d’un canal formé par l’association des quatre sous-unités de nucléoprotéine. Ces conclusions sont essentiellement basées sur le calcul de surfaces électrostatiques (Figure 20 C). Sans connaître la structure du complexe nucléoprotéine-ARN, il est difficile de conclure définitivement sur la localisation de ce dernier.

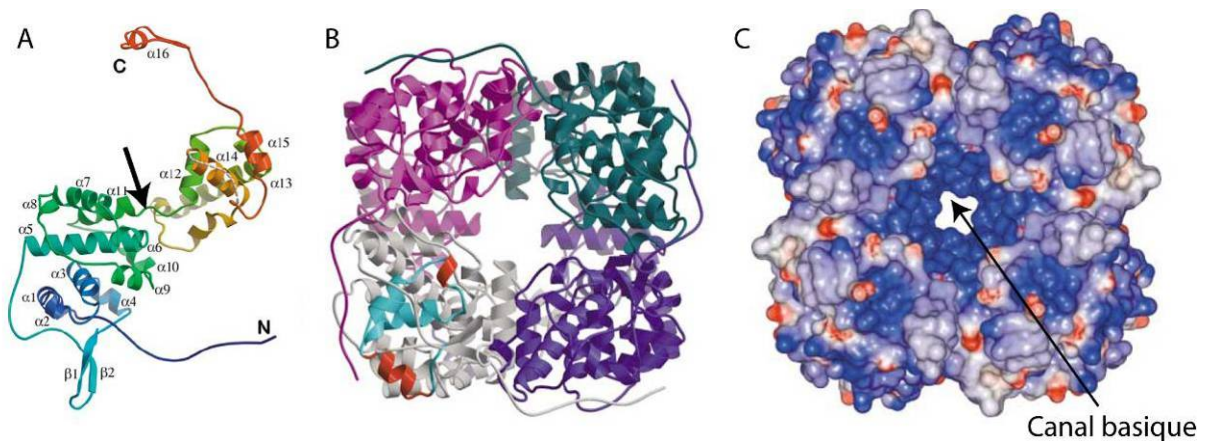


Figure 20 : Structure cristallographique de la nucléoprotéine du Bornavirus (Rudolph et al., 2003).

A : Protomère de nucléoprotéine du Bornavirus. La flèche indique la séparation des deux domaines (code pdb : 1N93). Gradient de couleur : du bleu (N-terminal) au rouge (C-terminal).

B : Tétramère de nucléoprotéine du Bornavirus. Chaque protomère est d'une couleur différente.

C : Surface électrostatique du tétramère de nucléoprotéine (en bleu : surface basique, en rouge : surface acide).

3) L'ARN-polymérase-ARN-dépendante virale (L)

Il existe plusieurs types d'ARN-polymérases virales. Les premières à avoir été caractérisées sont l'ARN-polymérase-ADN-dépendante du virus de la vaccine et l'ARN-polymérase-ARN-dépendante des Reovirus (virus à ARN double brin) (Lewandowski et al., 1969). Dans notre cas, la polymérase des *Mononegavirales* est capable de transcrire l'ARN viral simple brin de polarité négative en ARN messager. La première polymérase de ce type a été mise en évidence chez le VSV il y a plus de trente ans (Baltimore et al., 1970). Les ARN-polymérases virales des *Mononegavirales* sont de grosses molécules ayant un poids moléculaire qui varie de 250 à 300 kDa, elles sont notées L pour « Large protein ». Elles possèdent toutes une caractéristique unique : elles ne peuvent être actives que sur des matrices nucléoprotéine-ARN et non sur l'ARN viral nu (Emerson and Wagner, 1972). La présence de la protéine P est également indispensable pour que l'ARN-polymérase virale soit active en transcription (Emerson and Yu, 1975). Des analyses effectuées avec des constructions tronquées de protéine L ont permis d'identifier les zones d'interactions entre l'ARN-polymérase-ARN-dépendante et la phosphoprotéine. Concernant le virus de la rage, la fixation de la phosphoprotéine se fait sur les 566 derniers acides aminés de l'ARN-polymérase-ARN-dépendante virale (Chenik et al., 1998). Dans le cas du virus de Sendai (Paramyxovirus) la phosphoprotéine se fixe au niveau des 400 premiers acides aminés (Holmes and Moyer,

2002). Le nombre de molécules d'ARN-polymérase-ARN-dépendante virale par virion de rage est faible par rapport aux autres protéines : il est estimé à environ 50 molécules associées au complexe nucléoprotéine-ARN (Thomas et al., 1985).

La protéine L constitue la sous-unité principale du complexe enzymatique polymérase viral chez les *Mononegavirales*. Cet enzyme catalyse plusieurs activités durant le cycle viral : transcription, réplication, ainsi que l'ajout de la coiffe et la polyadénylation des ARN messagers.

A ce jour, aucune structure cristallographique d'une polymérase de virus à ARN négatif n'a été résolue. Par contre, ces polymérases possèdent une certaine homologie de séquence avec les polymérases virales des virus à ARN positif dont plusieurs structures sont connues (Poch et al., 1989). De nombreuses données structurales de polymérases de virus à ARN positif sont disponibles, et il est certain qu'une part de ces informations devrait aussi s'appliquer aux polymérases des virus à ARN négatif (Kolakofsky et al., 2004). L'interaction de la polymérase avec la matrice nucléoprotéine-ARN en général et entre la polymérase et la nucléoprotéine en particulier demeure pour le moment l'un des points obscurs de la transcription et de la réplication de ces virus.

La seule information structurale à notre disposition pour un virus à ARN négatif concerne la polymérase virale du virus de la grippe. Cette polymérase est très différente de celles des autres virus à ARN négatif car elle se présente sous la forme d'un hétérotrimère. Une reconstruction de cette ARN-polymérase-ARN-dépendante du virus de la grippe obtenue à partir d'images de microscopie électronique à une résolution de 23 Å a pu être établie (Area et al., 2004; Martin-Benito et al., 2001) (Figure 21). La reconstruction a été réalisée par l'intermédiaire de particules ribonucléoprotéiques recombinantes. Cet échantillon a été produit par amplification *in vivo* à partir des ADN complémentaires de la nucléoprotéine et des différentes sous-unités de l'ARN-polymérase virale (PB1, PB2 et PA) associée à un ARN de 248 nucléotides. La reconstruction montre une structure très compacte et ne laisse pas entrevoir de contacts extensifs entre les différentes sous-unités de la polymérase et les sous-unités de nucléoprotéine. Les différentes sous-unités de l'ARN-polymérase-ARN-dépendante du virus de la grippe ont pu être localisées à l'aide d'anticorps spécifiques sur la reconstruction.

Il s'agit ici du seul modèle de l'unité de réplication d'un virus à ARN négatif disponible à ce jour.

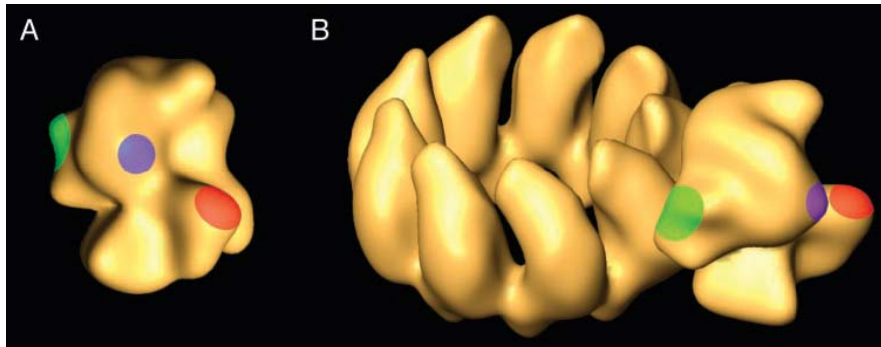


Figure 21 : Modèle tridimensionnel de l'ARN-polymérase du virus de la grippe (Area et al., 2004).

Les surfaces colorées indiquent la localisation des sous-unités de l'ARN polymérase virale.

En rouge : région N-terminale de la sous-unité PB2. En Bleu : région C-terminale de la sous-unité PA.

En vert : région C-terminale de la sous-unité PB1.

A : Reconstruction tridimensionnelle de l'ARN-polymérase virale du virus de la grippe.

B : Reconstruction tridimensionnelle d'une particule ribonucléique artificielle.

4) Cibles pour molécules antivirales

Les nucléocapsides des virus à ARN négatif sont des structures uniques dans la biologie des acides nucléiques. Dans notre cas, la nucléoprotéine qui couvre le génome est indispensable à au cheminement de l'ARN-polymérase virale. Ces matrices nucléoprotéine-ARN peuvent être des cibles parfaites pour le développement de molécules antivirales spécifiques qui seraient dépourvues d'effets secondaires nocifs du fait de leur caractère unique au monde viral. Les sites actifs de l'ARN polymérase virale, les interactions protéine-protéine comme N-P, P-P, P-L et L-L sont autant de très bonnes cibles potentielles pour la conception de médicaments antiviraux. La mise au point de molécules modifiant de façon définitive les caractéristiques hélicoïdales des nucléocapsides et empêchant donc le cheminement de l'ARN-polymérase le long de cette hélice est une très prometteuse voie de recherche de nouveaux antiviraux. Pour un tel développement, les structures tridimensionnelles et la compréhension des mécanismes moléculaires associés aux différents composants de la nucléocapside sont indispensables.

Objectifs du travail de Thèse

L'objectif de mon travail de thèse consistait à élucider par une approche structurale les mécanismes de synthèses de l'ARN pour les virus à ARN négatif et plus particulièrement du virus de la rage. Afin de mieux comprendre les mécanismes de réplication et transcription virale, la connaissance de la structure tridimensionnelle des protéines virales est nécessaire. Les virus à ARN négatif possèdent tous une protéine appelée nucléoprotéine (N) qui chaperonne l'ARN. La nucléoprotéine protège l'ARN de l'environnement extérieur tout en le rendant accessible au complexe ARN-polymérase viral.

La nucléoprotéine du virus de la rage existe sous deux formes. La première forme est dite « soluble » et résulte de l'association entre la nucléoprotéine et la phosphoprotéine. La seconde forme, résulte quant à elle de l'association irréversible entre la nucléoprotéine et un ARN.

Il est possible de produire le complexe nucléoprotéine-phosphoprotéine (N^o-P) de manière recombinante en co-exprimant les deux protéines dans le système baculovirus-cellules d'insectes. Durant la première année de ma thèse, j'ai ainsi pu purifier le complexe N^o-P et obtenir des cristaux du complexe protéolysé. Malheureusement, ces cristaux étaient trop petits pour être analysés sur les lignes de lumière du synchrotron. Il m'a de plus été impossible de reproduire ces cristaux.

Je me suis intéressée dans un second temps aux complexes formés par la nucléoprotéine avec de l'ARN cellulaire. Lorsque la nucléoprotéine est exprimée en cellules d'insectes sans son cofacteur, la phosphoprotéine, elle lie de manière forte et non spécifique les ARN cellulaires. La fixation de la nucléoprotéine sur des ARN courts (ARN de transfert) conduit à la formation de complexes en forme d'anneaux (un brin d'ARN associé à 9, 10, 11, 12, 13 sous-unités de nucléoprotéine et plus). Après mise au point, une méthode originale et efficace basée sur des études électrophorétiques en conditions natives a permis de purifier les différents types d'anneaux. Des études en microscopie électronique ont permis d'obtenir des reconstructions tridimensionnelles de ces anneaux à meilleure résolution. L'homogénéité des échantillons ainsi purifiés a également rendu possible leur cristallisation. Les anneaux de 9, 10 et 11 sous-unités de nucléoprotéine ont pu être cristallisés dans différentes conditions. Les cristaux d'anneaux contenant 10 nucléoprotéines n'ont conduit qu'à l'obtention de données de diffraction à une résolution moyenne (> à 8 Å). En revanche, avec les cristaux d'anneaux de 11 sous-unités de nucléoprotéines, après mise au point des conditions de cristallisation et de cryoprotection des cristaux, j'ai pu obtenir des données de diffraction à une résolution de 3.5

Å. La structure a ensuite été résolue par Single Anomalous Dispersion. La résolution de la structure a été réalisée en collaboration avec les scientifiques de l'EMBL-Grenoble, en particulier avec le groupe du Dr W. Weissenhorn.

Cette structure, la première du genre, permet de mieux comprendre le mode de fixation de la nucléoprotéine sur l'ARN. Nous avons ainsi mis en évidence le fait que les bases de l'ARN n'étaient pas exposées au solvant mais étaient étroitement séquestrées entre deux domaines de la nucléoprotéine. Ceci nous permet d'envisager que plusieurs conformations de ce complexe puissent exister. En effet, l'ARN doit pouvoir se détacher localement de la nucléoprotéine pour permettre aux mécanismes de transcription et de réplication virale de se produire. Ceci est une hypothèse nouvelle quant au mode de fonctionnement de la nucléoprotéine. Cette hypothèse conforterait un rôle de protection du génome par la nucléoprotéine vis-à-vis de l'environnement extérieur et plus particulièrement vis-à-vis des digestions par les RNases cellulaires.

Pour mieux comprendre la réplication et la transcription du virus, nous étudions actuellement le complexe formé entre les anneaux de nucléoprotéine et la phosphoprotéine.

Partie II : Protocoles pour l'expression en cellules d'insectes et la purification des complexes N-°P et N-ARN

Le système d'expression baculovirus-cellules d'insectes a été utilisé pour produire la nucléoprotéine et la phosphoprotéine du virus de la rage. Les cellules d'insectes sont cultivées en monocouche à 27 °C dans des flasques en plastique de 75 cm² à 300 cm². La production et le titrage des baculovirus recombinants sont réalisés sur des cellules *Sf21* (souche *Spodoptera frugiperda*). Les cellules HF (souche High Five™ Invitrogen) sont utilisées quant à elles pour l'expression des protéines recombinantes.

Les cellules d'insectes sont cultivées dans du milieu TC100 (Invitrogen) supplémenté avec 10 % de sérum de veau fœtal et des antibiotiques (pénicilline/streptomycine) pour l'expression de protéines natives et/ou la production de baculovirus recombinants.

Afin de surexprimer les protéines sélénométhionylées, la souche de cellules d'insectes *Sf9* est utilisée. Les cellules *Sf9* peuvent être cultivées, après adaptation, dans un milieu minimum SF900 (Invitrogen) sans sérum et supplémenté avec de la sénénométhionine et de la L-cystéine (Mazza et al., 2001).

I. Expression et purification du complexe N^o-P natif du virus de la rage

A. Expression du complexe N^o-P natif du virus de la rage en cellules d'insecte

Le complexe N^o-P est surexprimé en cellules d'insectes. Deux baculovirus recombinants sont utilisés : le premier contient le gène de la nucléoprotéine N fusionnée avec une étiquette histidine (N_{his}) et le second le gène de la phosphoprotéine P. Les cellules HF de 50 flasques de 300 cm² sont infectées avec les deux baculovirus recombinants à une m.o.i. de 5 pour la P et une m.o.i. de 2 pour la N_{his}.

Les cellules sont récoltées 65 heures *p.i.* puis rincées avec du PBS 1x et reprises dans 120 ml de tampon de lyse (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 2 mM β-mercaptoéthanol, 1 mM EDTA, 1x Complete-EDTA free (Roche Diagnostics™) puis stockées à -80°C jusqu'à une utilisation future.

La lyse des cellules est provoquée par congélation/décongélation en utilisant en alternance un bain à 37 °C et de l'azote liquide (-180 °C) (3 cycles). Le lysat est ensuite centrifugé à 12000 g pendant 15 min. Le surnageant clair contenant les protéines solubles est aussitôt utilisé pour la purification.

Lorsque la nucléoprotéine et la phosphoprotéine sont co-exprimées dans les cellules d'insectes, différents complexes protéiques sont produits (Figure 22). Lors de la purification du complexe N^o-P, il est nécessaire d'éliminer les complexes N-ARN et les oligomères de phosphoprotéine produits lors de la co-expression.

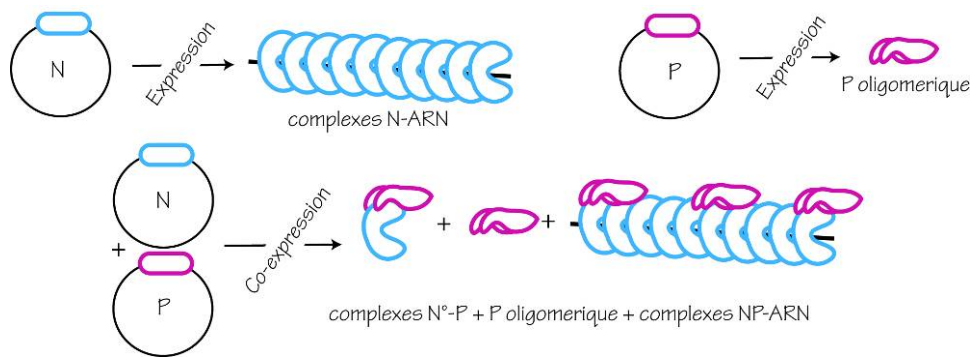


Figure 22 : Expression et co-expression des protéines N et P du virus de la rage.

Illustration des différents complexes obtenus en exprimant ou co-exprimant le gène de la nucléoprotéine et/ou de la phosphoprotéine du virus de la rage dans le système baculovirus-cellules d'insectes.

B. Purification du complexe N°-P du virus de la rage

Le lysat clair est chargé sur une colonne échangeuse d'anions (DEAE-Sepharose, Amersham). Le complexe N°-P est élué au cours d'un gradient de chlorure de sodium (120 mM NaCl, 20 mM Tris-HCl pH 7.5 - 375 mM NaCl, 20 mM Tris-HCl pH 7.5) à une concentration d'environ 150 mM NaCl. Les fractions contenant le complexe N°-P sont identifiées sur SDS-PAGE puis rassemblées. Une étape de précipitation au sulfate d'ammonium à 32 % permet de précipiter N°-P en excluant quelques contaminants cellulaires. Le précipité est repris dans 2 ml de tampon (150 mM NaCl, 20 mM Tris-HCl pH 7.5) pour être injecté sur une première colonne de chromatographie d'exclusion de taille (Superdex 200 Hiload 16/60, Amersham). Cette étape permet de se séparer des complexes N-ARN de haut poids moléculaire. Les fractions contenant N°-P et P oligomérique sont ensuite chargées sur une colonne d'affinité (NiNTA, Quiagen). Grâce à l'étiquette histidine sur la nucléoprotéine, le complexe N°-P natif est retenu, les oligomères de phosphoprotéine sont éliminés dans l'éluat.

C. Purification du complexe N°-P_{dig}

Le complexe N°-P est incubé 15 à 18 heures avec 1/500 p/p de trypsine. La protéolyse est ensuite stoppée par ajout de *péfabloc* (Roche diagnostics™). Le mélange de digestion est ensuite injecté sur une colonne de chromatographie d'exclusion de taille (Superdex 75 HR10/30, Amersham). Le complexe est élué dans du tampon 20 mM Tris-HCl pH 7.5 - 150 mM NaCl (flux d'éluat : 0.5 ml/min. Taille des fractions : 0.2 ml). Les fractions contenant le complexe N°-P_{dig} sont rassemblées pour l'étape de chromatographie suivante.

Le complexe digéré est dilué 5 fois dans du tampon 20 mM Tris-HCl pH 8.5 puis chargé sur une colonne échangeuse d'anions (Mono Q HR 5/5, Amersham). Après une étape de lavage à 50 mM NaCl, l'élution se fait grâce à un gradient 50 à 250 mM NaCl (flux d'élution : 0.5 ml/min. Taille des fractions : 0.5 ml). Les fractions des pics contenant le complexe N^o-P_{dig} sont rassemblées et concentrées pour la suite des expériences.

II. Expression et purification des complexes N-ARN du virus de la rage

A. Expression des complexes N-ARN du virus de la rage en cellules d'insectes

Les cellules HF de 18 flasques de 300 cm² sont infectées à une m.o.i. de 5 avec le baculovirus recombinant pour la protéine N. Les cellules sont récoltées 3 jours *p.i.* et rincées avec du PBS 1x puis re-suspendues dans 12 ml de tampon de lyse (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 2 mM β-mercaptoéthanol, 1 mM EDTA, 1x Complete-EDTA free (Roche Diagnostics™)) et stockées à -80°C jusqu'à une utilisation ultérieure. La lyse des cellules est provoquée par congélation/décongélation en utilisant en alternance un bain à 37 °C et de l'azote liquide (-180 °C) (3 cycles). Le lysat est centrifugé durant 15 min à 12000 g, le surnageant clair contenant les protéines solubles est aussitôt utilisé pour la purification.

B. Purification des complexes N-ARN : gradients de chlorure de Césium

Cette étape d'ultracentrifugation permet d'isoler les complexes N-ARN (RNP) constitués par un mélange d'anneaux et de nucléocapsides recombinantes.

Le surnageant est chargé sur un gradient de CsCl (40 % p/p CsCl dans 150 mM NaCl, 20 mM Tris-HCl pH 8 à 20 % CsCl p/p dans 150 mM NaCl, 20 mM Tris HCl pH 8) puis centrifugé à 30000 rpm pendant 16 heures à 8 °C (rotor sw41). Après ultracentrifugation, les bandes visibles sur le gradient sont ponctionnées à l'aide d'une aiguille en perçant directement la paroi du tube de polypropylène puis déposées sur SDS-PAGE. Les fractions contenant la nucléoprotéine sont rassemblées et dialysées contre un tampon 20 mM Tris-HCl pH 8 – 100 mM NaCl.

C. Purification des anneaux N-ARN : gradients de glycérol

Ce second gradient permet de séparer les anneaux des nucléocapsides. L'échantillon dialysé contenant les RNP est chargé sur un gradient continu de glycérol (15 % glycérol, 150 mM NaCl, 20 mM Tris HCl pH 8 v/v à 7.5 % glycérol, 150 mM NaCl, 20 mM Tris HCl pH 8 v/v).

Après une ultracentrifugation à 32000 rpm pendant 15 heures à 4 °C (rotor sw41), des fractions de 500 µl sont collectées depuis le fond à l'aide d'une sonde reliée à une pompe péristaltique. Les fractions contenant la nucléoprotéine pure sont identifiées par SDS-PAGE puis rassemblées et concentrées à environ 100 mg/ml avec une unité de concentration Millipore™ (c/o : 100 KDa) (concentration déterminée selon la méthode de Bradford).

**Partie III : Purification et cristallisation des complexes
contenant la nucléoprotéine**

I. Introduction

Que ce soit pour des études en microscopie électronique ou en cristallographie, l'obtention d'un échantillon pur et homogène est indispensable. La principale difficulté de l'étude d'une macromolécule par cristallographie réside dans l'obtention de cristaux. Hélas, la connaissance des propriétés physico-chimiques de la macromolécule n'est pas suffisante pour établir une méthode rationnelle permettant d'obtenir des cristaux de protéine. La cristallogenèse reste pour une grande part très empirique. La plateforme de cristallisation robotisée de l'EMBL-Grenoble a largement facilité cette étape de mes projets car elle permet de cribler rapidement un grand nombre de conditions de cristallisation avec une faible quantité de protéine. Les cristaux obtenus avec le robot peuvent ensuite être reproduits manuellement et leur qualité améliorée en affinant la condition de cristallisation. Le but est d'obtenir des données de diffraction à la plus haute résolution possible afin de résoudre la structure avec la plus grande précision.

La nucléoprotéine joue un rôle fondamental dans la réplication et la transcription des virus à ARN négatif. Il s'agit de la protéine majoritaire au sein de ces virus. La nucléoprotéine interagit d'une part avec la phosphoprotéine et d'autre part avec l'ARN donnant lieu à la formation de deux genres de complexes. La phosphoprotéine est une protéine chaperonne pour la nucléoprotéine, l'ARN étant quant à lui le « substrat » principal de la nucléoprotéine. L'objectif principal des expériences menées durant ma thèse consistait à obtenir un jeu de données à haute résolution sur la nucléoprotéine.

Pour un tel projet deux alternatives étaient possibles : étudier la nucléoprotéine sous sa forme dite « soluble » en complexe avec la phosphoprotéine (complexe N^o-P), ou en complexe avec l'ARN (complexe N-ARN). Chaque possibilité présente des intérêts différents. La structure du complexe N^o-P donnerait des informations sur le domaine d'interaction entre la nucléoprotéine et la phosphoprotéine. La structure du complexe nucléoprotéine-ARN permet quant à elle d'étudier le mode d'interaction de la protéine N avec elle-même et avec son substrat principal, l'ARN. Le mode d'empaquetage de l'ARN viral, ainsi que sa protection et son accessibilité à la polymérase virale seraient également visualisés.

J'ai travaillé sur le complexe entre la nucléoprotéine et la phosphoprotéine pendant la première année de mon travail doctoral dans la continuité de la thèse de Manos Mavrakis en vue d'obtenir des cristaux et de résoudre une structure cristallographique.

Un deuxième projet sur les complexes entre la nucléoprotéine et l'ARN qui avait pour but initial la réalisation d'une reconstruction tridimensionnelle en microscopie électronique a

débouché sur l'obtention de cristaux reproductibles. La purification des échantillons N-ARN basée sur une expérience d'électrophorèse préparative en conditions natives a été l'étape déterminante de ce projet. L'électrophorèse préparative a été mise au point en collaboration avec le Dr C. Clapier de l'EMBL-Grenoble.

Cette partie a pour but de présenter les résultats biochimiques et les cristaux obtenus avec la nucléoprotéine du virus de la rage.

II. Cristallisation du complexe entre la nucléoprotéine et la phosphoprotéine

La production du complexe natif est réalisée comme décrit précédemment (Mavrakis et al., 2003). Un protocole plus détaillé pour produire les complexes N^o-P natif et N^o-P_{dig} est rapporté en partie II.

Le complexe N^o-P natif semble posséder de nombreuses zones flexibles probablement à l'origine de l'échec de sa cristallisation. La production d'un complexe protéolysé a donc été envisagée pour obtenir un échantillon plus rigide et plus approprié à la cristallogénèse.

Des tests en utilisant plusieurs protéases, différents temps d'incubation et différents ratio quantité de protéase / quantité de protéine, ont permis la détermination d'une condition de protéolyse limitée adéquate. La digestion avec la trypsine à 1 /500^{ème} p/p pendant 15 à 18 heures semble donner un échantillon protéolysé stable. Après protéolyse, le mélange est injecté sur une colonne de chromatographie d'exclusion de taille (Superdex 75 HR10/30, Amersham). Un complexe N^o-P_{dig} est alors isolé (Figure 23 A et B). Cet échantillon censé être moins flexible n'a pourtant pas conduit à l'obtention de cristaux.

Une étape de chromatographie supplémentaire a donc été envisagée. En soumettant le complexe digéré à une colonne échangeuse d'anions (Mono Q HR 5/5, Amersham) nous avons remarqué que deux espèces différentes pouvaient être purifiées. La première espèce de complexe digéré est éluée à 130 mM NaCl (pic I), la seconde espèce apparaît à environ 150 mM NaCl (pic II). Sur SDS-PAGE, les deux échantillons semblent identiques (Figure 23 C et D).

Nous allons effectuer des expériences de spectrométrie de masse et de séquençage N-terminal sur les deux échantillons afin de caractériser plus précisément et avec certitude ces deux espèces. Il est probable que la différence entre les deux pics soit due à une petite variation au niveau de la charge du complexe, ou à une différence de quelques acides aminés seulement.

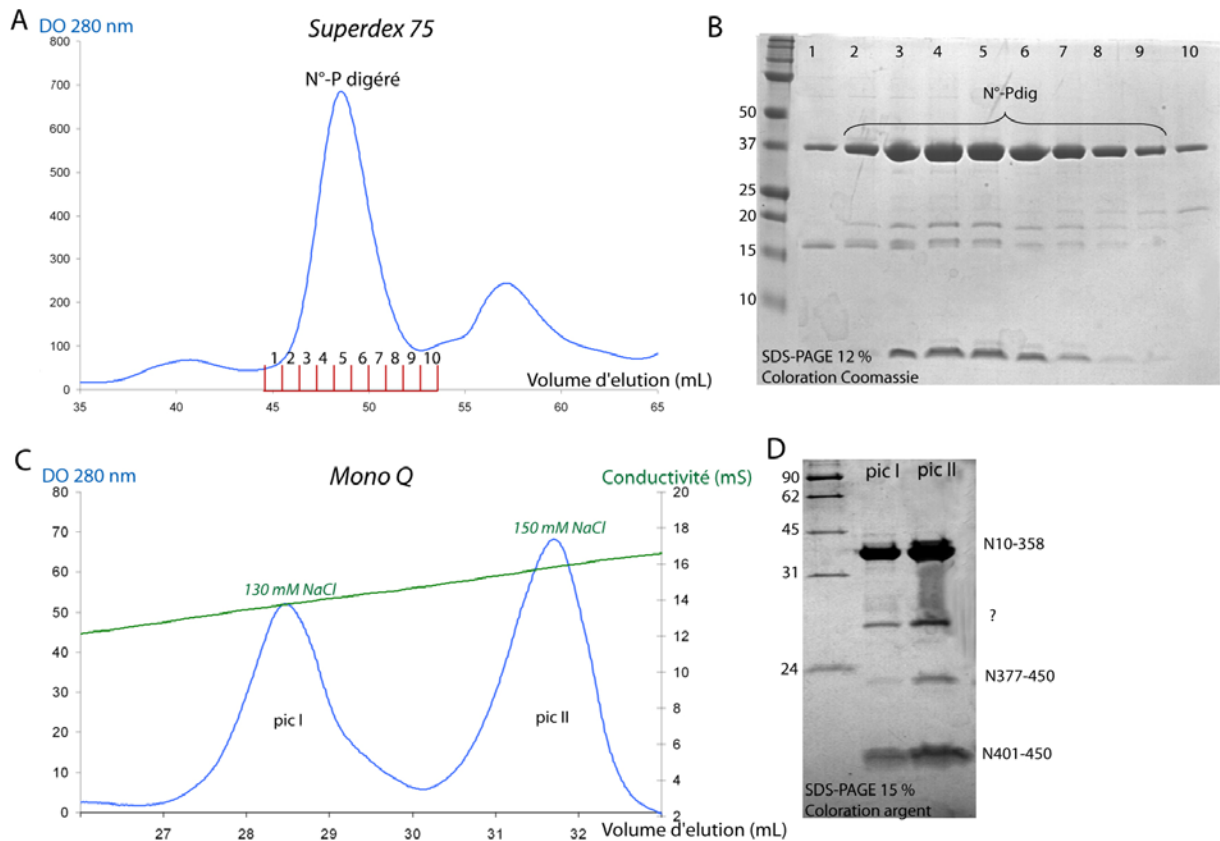


Figure 23 : Profils d'élution et gel SDS-PAGE de la purification du complexe N°-P digéré.

A : Profil d'élution de la chromatographie d'exclusion de taille (Superdex 75) le pic central contient le complexe N°-P digéré.

B : SDS-PAGE correspondant à l'étape de Superdex 75. Les fractions 2 à 9 contenant N°-P_{dig} sont rassemblées.

C : Profil d'élution de la chromatographie échangeuse d'anions (Mono Q), les pics I et II contiennent le complexe N°-P_{dig}.

D : SDS-PAGE des échantillons concentrés (pic I et pic II) après colonne échangeuse d'anions (Mono Q). Les bandes de protéines ont été identifiées d'après Mavrakis et al., 2006.

Après concentration jusqu'à 5 mg/ml, seul le complexe digéré contenu dans le pic II (Figure 24) a pu être cristallisé dans deux conditions du screen INDEX (Hampton research). Les cristaux obtenus ont la forme de petites aiguilles très fines. Du fait de leur taille, ils n'ont pas pu être testés sur les lignes de lumière du synchrotron. De plus, ces cristaux n'ont jamais pu être reproduits ni manuellement ni à l'aide du robot. Cette absence de reproductibilité est sûrement liée à de légères variations lors de la purification du complexe N°-P_{dig}. et notamment lors de sa protéolyse limitée. Cette étape est en effet la moins reproductible du processus de purification du fait des nombreuses variables à contrôler (température, concentration et volume d'échantillon à digérer et d'*aliquot* de trypsine utilisé).

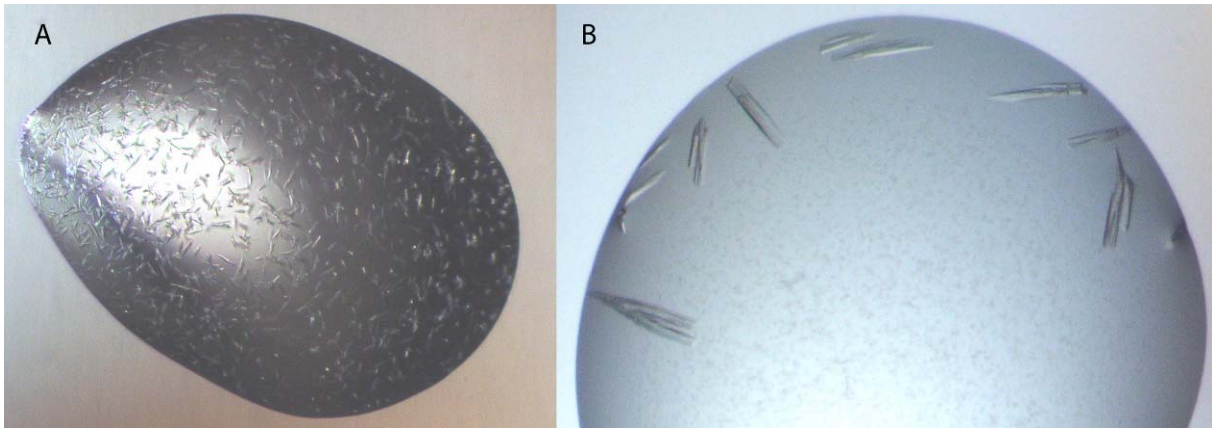


Figure 24 : Cristaux du complexe N°-P_{dig} obtenus en nanogouttes (0.1 µl + 0.1 µl).

A : 20 % PEG 3350, 150 mM acide DL malique.

B : 15 % PEG 3350, 100 mM magnesium formate.

III. Cristallisation des complexes nucléoprotéine-ARN

A. Production d'un mélange d'anneaux nucléoprotéine-ARN

Dans ce paragraphe, les étapes préliminaires à l'expérience d'électrophorèse préparative sont détaillées. Cette partie de la purification repose sur les résultats précédemment publiés (Iseni et al., 1998; Schoehn et al., 2001) auxquels quelques modifications ont été apportées. La purification des complexes N-ARN débute par deux étapes d'ultracentrifugations détaillées en partie II.

Un premier gradient de densité de chlorure de césium (20 % CsCl - 40 % CsCl p/p) permet d'isoler les complexes N-ARN : un mélange d'anneaux et de nucléocapsides recombinantes est ainsi purifié. Le surnageant issu de la lyse des cellules d'insectes est chargé sur ce gradient de CsCl. Les bandes visibles sur le gradient après ultracentrifugation sont récupérées (Figure 25 A). D'une manière générale trois bandes sont visibles sur le gradient (notées I, II et III). La bande III contient les complexes N-ARN (Figure 25 B).

Un second gradient de glycérol permet de séparer les anneaux des nucléocapsides suivant leur masse. Les nucléocapsides sont de très gros complexes de plusieurs mégadaltons qui sont culottés lors de l'ultracentrifugation. En revanche les anneaux, d'un poids moléculaire bien plus faible, migrent à l'intérieur du gradient vers une position intermédiaire.

D'un point de vue pratique, le mélange de complexes N-ARN est déposé à la surface d'un gradient continu de glycérol (7.5 % - 15 % v/v). Les fractions contenant de la nucléoprotéine pure sont identifiées par SDS-PAGE (Figure 25 C), rassemblées puis concentrées à environ 100 mg/ml.

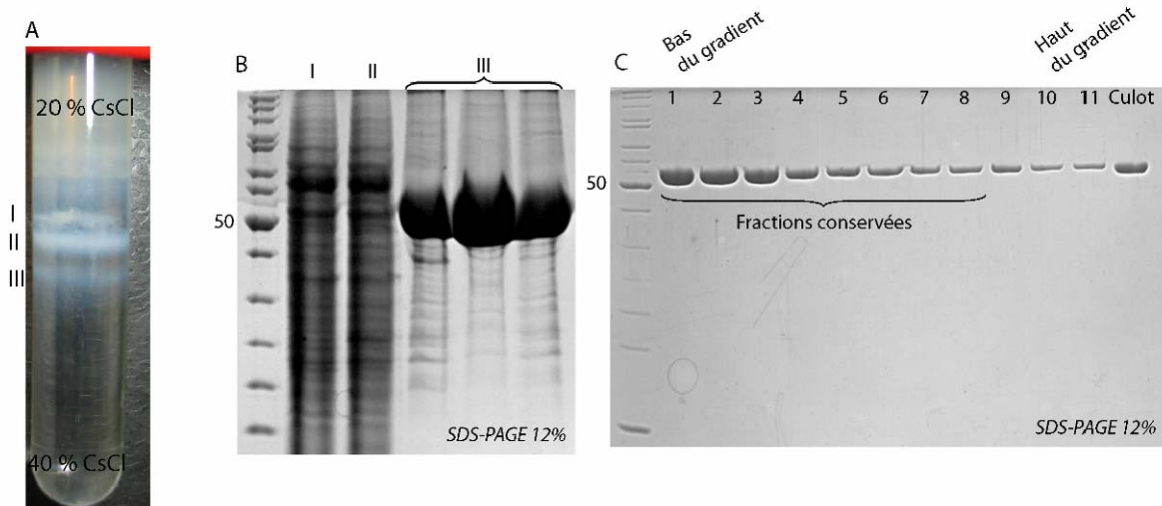


Figure 25 : Purification des complexes recombinants N-ARN du virus de la rage.

A : Gradient de CsCl 20 % - 40 % p/p après ultracentrifugation. Trois bandes sont visibles (I, II et III).

B : SDS-PAGE 12 %. Les fractions issues du gradient de CsCl ont été déposées. La bande III contient les complexes N-ARN.

C : SDS-PAGE 12 %. Après ultracentrifugation sur gradient de glycérol (7.5 % - 15 % v/v), les fractions sont collectées à partir du fond du tube. Les fractions contenant de la nucléoprotéine sont rassemblées puis concentrées.

La Figure 26 montre une image de microscopie électronique des anneaux N-RNA purifiés à l'issue des deux étapes d'ultracentrifugation.

Les sous-unités nucléoprotéiques sont visibles et il est possible de les dénombrer directement sur les clichés de microscopie électronique. Les anneaux sont composés de 9 à 15 sous-unités en général mais il n'est pas rare d'observer des complexes comportant davantage de nucléoprotéines. Dans ce cas, ces complexes sont considérés comme de mini-nucléocapsides plutôt que comme des anneaux. Il est intéressant de noter aussi la présence d'un contaminant protéique co-purifié avec les anneaux, le protéasome.

La reconstruction publiée récemment (Schoehn et al., 2001) a été réalisée à partir d'un échantillon enrichi en anneaux composés de 10 sous-unités de nucléoprotéine et semblable à celui de la Figure 26. Lors de la sélection des particules sur les images prises en cryomicroscopie (peu contrastées), il est très difficile de faire la différence entre les vues de côté des anneaux de 9, 10 ou 11 sous-unités de nucléoprotéine. Ceci a pour conséquence de limiter la résolution de la reconstruction à 30 Å.

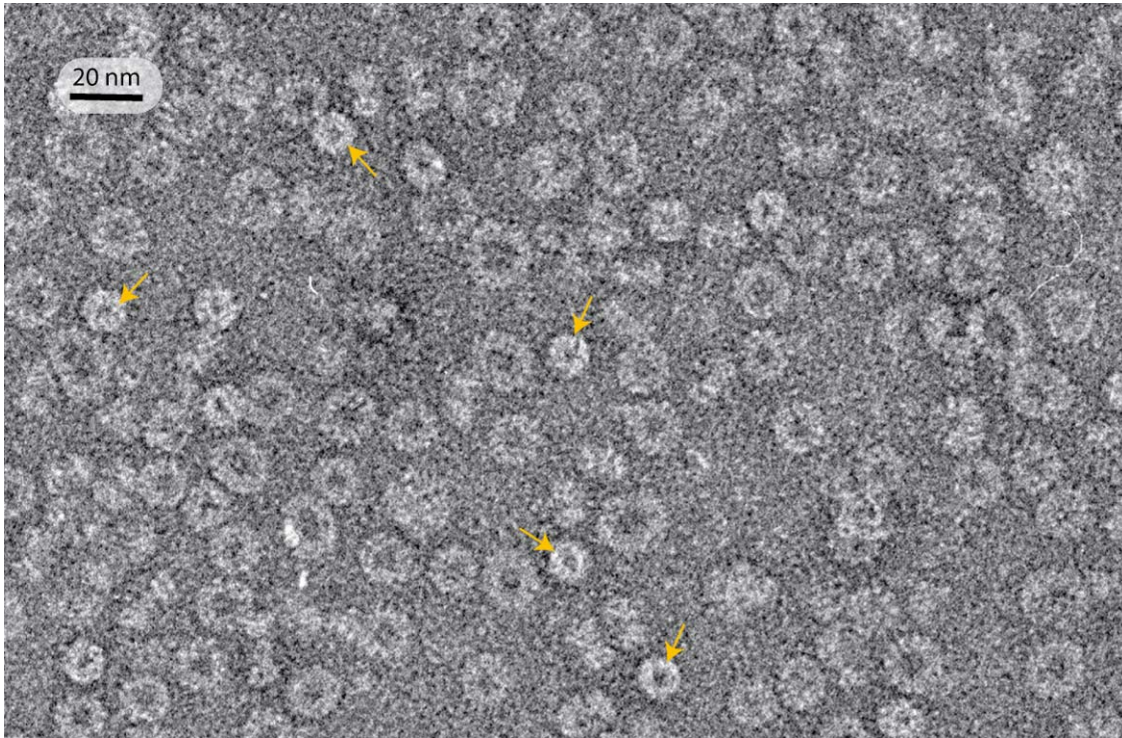


Figure 26 : Mélange d'anneaux nucléoprotéine-ARN observé par microscopie électronique (coloration négative).

Les flèches jaunes montrent la présence de proteasomes contaminants.

B. ARTICLE 1: Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein–RNA rings.

Article sous presse dans Journal of Structural Biology

Available online at www.sciencedirect.com

Journal of Structural Biology xxx (2006) xxx–xxx

Journal of
Structural
Biology

www.elsevier.com/locate/yjsbi

Crystallization notes

Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings

Aurélie A.V. Albertini^a, Cedric R. Clapier^b, Amy K. Wernimont^b, Guy Schoehn^a,
Winfried Weissenhorn^b, Rob W.H. Ruigrok^{a,*}

^a *Institut de Virologie Moléculaire et Structurale, FRE 2854 Université Joseph Fourier-CNRS, IVMS, clo EMBL, BP 181, 38042 Grenoble cedex 9, France*

^b *European Molecular Biology Laboratory (EMBL), BP 181, 38042 Grenoble, France*

Received 7 September 2006; received in revised form 10 October 2006; accepted 15 October 2006

Abstract

In order to study the packaging of rabies virus RNA inside the viral nucleocapsid, rabies nucleoprotein was expressed in insect cells. In the cells, it binds to cellular RNA to form long, helical or short circular complexes, depending on the length of the bound RNA. The circular complexes contained from 9 up to 13 N-protomers per ring. Separation of the rings into defined size classes was impossible through regular column chromatographies or gradient centrifugation. The size classes could be separated by native polyacrylamide gel electrophoresis. A large-scale separation was achieved with a 4% native gel using a preparative electrophoresis apparatus. Crystallisation trials were set up with N-RNA rings from three size classes and crystals were obtained in all cases. The best diffracting crystals, diffracting up to 6 Å, contained rings with 11 N-protomers plus an RNA molecule of 99 nucleotides. The diffraction limit was improved to 3.5 Å by air dehydration prior to flash freezing.

© 2006 Published by Elsevier Inc.

Keywords: Rabies virus; Protein–RNA complex; Native preparative gel electrophoresis; EM reconstruction; Purification large complexes

1. Introduction

The genome of negative strand RNA viruses is tightly associated with the viral nucleoprotein (N; MW between 55 and 60 kDa depending on the virus) forming helical N-RNA structures. These N-RNA structures are the templates for transcription and replication by the viral RNA-dependent RNA polymerase. Paramyxoviruses (such as Measles, Sendai and Respiratory Syncytial Viruses) have rather tight helical “herring-bone” structures typical for this family. Rhabdoviruses (such as Rabies and Vesicular Stomatitis Viruses) have ribbon-like N-RNA structures that wind into a tight helical structure inside the virus particles but the helices become rather loose when they are taken out of the virus (Nakai and Howatson, 1968; Schoehn et al., 2001). The nucleoprotein of rhabdoviruses has a bilobed

structure and each N-protomer binds nine ribonucleotides (Thomas et al., 1985). In order to understand how the RNA is packaged, we started a research project on Measles virus and Rabies virus N-RNA. We could determine that for both viruses the RNA is associated with the protein by its phosphate–ribose backbone but that the nucleotide bases were rather inaccessible for chemical modification (Iseni et al., 2000, 2002). However, for detailed knowledge on the interaction between the protein and the RNA, we needed more structural information. Unfortunately, the helical N-RNA structures are long and flexible and therefore impossible to crystallise. Recombinant Measles virus N-RNA was studied by cryo-electron microscopy after removal of the flexible C-terminal tail of N which made the helices more or less rigid. A 3D reconstruction of the complex to a resolution of 10 Å could be calculated but this resolution was not enough to visualise the RNA (Schoehn et al., 2004). We have not been able to find any method for rigidifying Rhabdovirus N-RNA complexes. However,

* Corresponding author. Fax: +33 4 76 20 94 00.

E-mail address: ruigrok@embl.fr (R.W.H. Ruigrok).

when Rabies virus nucleoprotein is expressed in insect cells, it binds to cellular RNAs and forms ring structures when the bound RNA is short and long flexible ribbons when the RNA is long (Iseni et al., 1998). Negative stain EM observation shows that the rings contain 9 to 13 N-protomers with molecular weights between 500 and 700 kDa. Subsequent glycerol gradient centrifugation allowed the isolation of a fraction enriched for rings with 10 protomers but the 3-D cryo-EM reconstruction was limited to about 30 Å because the preparation was still not homogeneous enough. The production of the protein, its purification over a CsCl gradient and subsequent glycerol gradient is described in Iseni et al. (1998) and Schoehn et al. (2001). Crystallisation trials with this enriched fraction did not lead to diffraction quality crystals.

2. Isolation of homogeneous fractions by preparative continuous elution native PAGE

We tried many different techniques in order to obtain a homogeneous fraction such as charge and hydrophobicity chromatographies but, because all the N-RNA rings have the same chemical composition and the same structure, we could not improve the separation of the rings. Moreover, because the size differences between rings of 10 and 11 protomers is only 10%, size exclusion chromatography was insufficient to generate a separation. Native polyacrylamide gel electrophoresis (PAGE) has been used to separate large complexes like intact mutant bacteriophages from wild type phages (Childs and Birnboim, 1975). Indeed, when we analysed the mixture of N-RNA rings by native 4% PAGE, the sample revealed a number of discrete bands, probably corresponding to the different population of rings with different numbers of N-protomers (Fig. 1A, input lane). The conditions for the native gel were varied until an optimal gel system was obtained: 4% (w/v) polyacrylamide (20:1 acrylamide to bisacrylamide w/w) in 0.2× TBE (1× TBE is 89 mM Tris-HCl, pH 7.5, 89 mM boric acid, 2.5 mM EDTA). The resolution for native gels is directly related to the height of the loaded sample. For the gel shown in Fig. 1 we used a loading height of 0.5 mm. The bands were separated by a minimal distance of 0.8 cm, indicating that the separation of the different populations of rings is possible using a preparative continuous elution native PAGE method. Therefore, in order to separate the N-RNA ringed complexes to homogeneity, we set up a method using a preparative electrophoresis apparatus (BioRad Model 491 Prep Cell apparatus; BioRad Laboratories, Richmond, CA) using the same conditions as those used for the slab analytical native gel in Fig. 1A.

A 60 ml volume of the 4% polyacrylamide gel was poured as a hollow cylindrical gel with an outer radius of 28 mm, an inner radius of 19 mm, and a height of 125 mm. The electrophoresis buffer was 0.2× TBE. The gel was pre-run for 90 min in 0.2× TBE at 4°C and 500 V. A concentrated sample of N-RNA ringed complexes (a maximum of

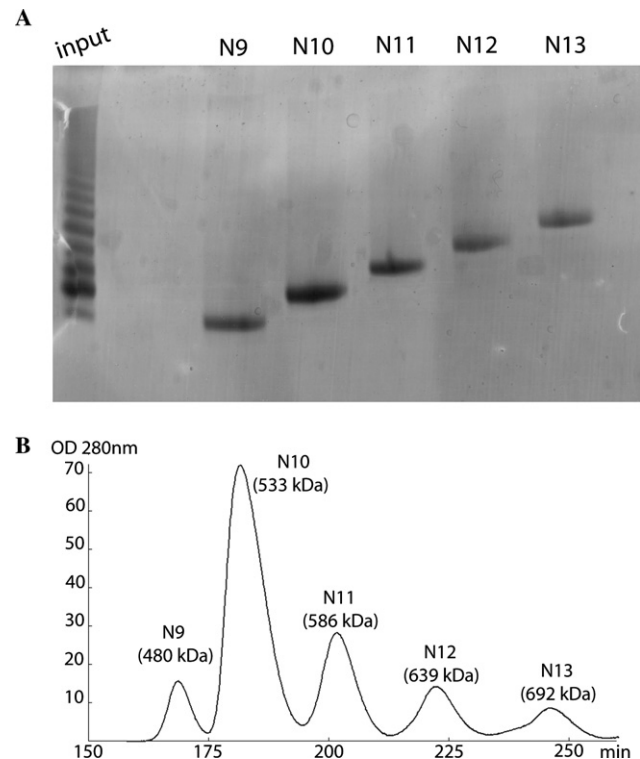


Fig. 1. Separation of N-RNA fractions by native gel electrophoresis. (A) Slab 4% native gel stained by Coomassie G dye showing the input material of the preparative electrophoresis (Input lane) and the fractions eluted from the preparative gel (lanes N9–N13). (B) Elution profile of the preparative gel recorded by monitoring the UV absorbance at an OD of 280 nm. Fractions N9 to N13 were analysed on the slab analytical gel in A. The N-RNA ring mixture was also slightly contaminated with proteasomes that eluted in the shoulder leading up to the peak of fraction N13. Molecular weights of the successively eluted N-RNA ring complexes are indicated.

150 µl or 15 mg) was mixed with sucrose to a final concentration of 2% (w/v) and loaded on the preparative gel. Electrophoresis was carried out at 4°C at constant voltage of 500 V and the different fractions were eluted with 20 mM Tris-HCl (pH 7.5) at a flow rate of 2 ml/min. The elution was recorded at an OD of 260 and 280 nm and 2 ml fractions were collected. Fractions from the eluted peaks (Fig. 1B) were pooled, concentrated by ultrafiltration (Amicon Ultra-15 centrifugal filter units, 10,000 NMWL, Millipore) to 5–10 mg/ml and loaded on a slab analytical native gel to control separation quality (Fig. 1A lanes N9–N13). Homogenous N-RNA rings were stored at 4°C. The nature of the rings in the fractions was verified by negative staining EM. Fig. 2A and B show fields of the rings in the N10 and N11 fractions plus averaged views of the rings clearly showing, respectively, 10 and 11 N-protomers per ring. Our successful purification in a large-scale range of closely migrating Rabies N-RNA ringed complexes demonstrates that it is possible to separate with high resolution complexes with very low molecular weight differences (below 10% in our case).

An almost identical preparative continuous elution native PAGE method is commonly used to isolate reconstituted recombinant nucleosomes from one position along a

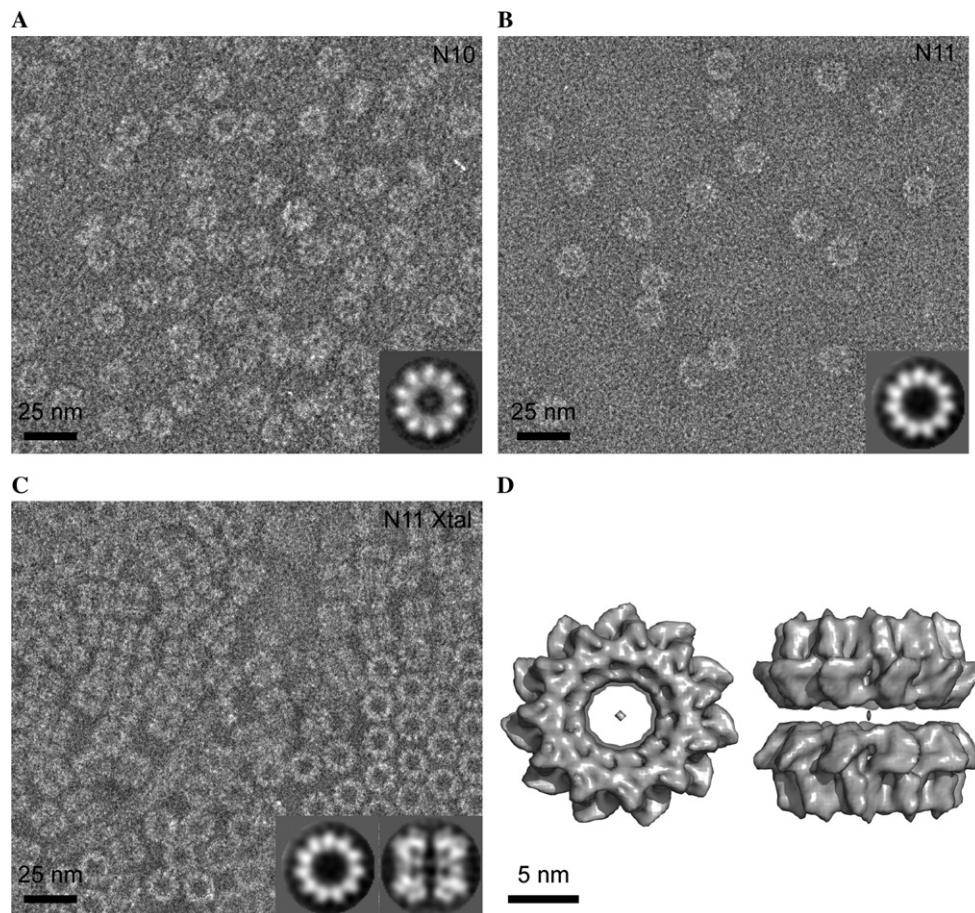


Fig. 2. Electron micrographs of negatively stained particles from the N10 (A) and N11 (B) fractions and from crushed N11 crystals (C). The inner and outer diameters for the N10 rings are about 60 and 160 Å and for N11 65 and 165 Å, respectively. Samples at approximately 0.1 mg/ml were applied to the clean side of carbon on mica (carbon/mica interface) and negatively stained with 1% sodium silicotungstate (pH 7.0). Micrographs were taken under low-dose conditions with a JEOL 1200 EX II microscope at 100 kV and a nominal magnification of 40,000 times. Electron micrographs were digitized using a Zeiss scanner with a pixel size of 14 μm (3.5 Å at the sample scale). Projection averages (insets) were calculated from top-views using the SPIDER image-processing package (Frank et al., 1996). Single images were manually selected from one micrograph, cut out, band pass filtered between 200 and 25 Å without CTF correction and then normalized to the same mean and standard deviation. Approximately 100 particle images were selected for each type of ring. The images from each set of particles were aligned to each other in translation and rotation and then averaged. For the crushed crystals in (C), 10 crystals were fished from the crystallisation drop, washed with buffer, resuspended in 20 μl and then crushed with a mini potter. Four microliter of this solution was used to make a grid as described above. (D) Top view and side view of a 3D-reconstruction of the double 11-mer rings from the crushed crystals. 1876 side-views were selected from 14 micrographs and corrected for the CTF as described in Fabry et al. (2005). The first 3D model was generated by back projection of one side view taking into account the 11-fold symmetry axis (Schoehn et al., 2001). A projection matching method implemented in SPIDER (Frank et al., 1996) was used to generate, after 50 cycles, the final 3D model. Twenty two class averages of side view were used for the final reconstruction. The resolution was determined by Fourier shell correlation (FSC; 0.5 level) to be 22 Å (20 Å for the 0.3 level).

DNA fragment from another (Dyer et al., 2004). In that particular application, the molecular weight is strictly identical between the species separated but the overall charge varies depending on the wrapping of the DNA fragment around the histone octamer. The main difference between applying the method to the nucleosomes or the Rabies RNA-nucleoprotein rings lies in varying the ratio acrylamide:bisacrylamide within the native gel: 59:1 and 20:1, for the nucleosomes and the Rabies rings, respectively. Therefore, it is possible to favour rather a separation mainly by mass or by charge and to tune the conditions of the preparative continuous elution native PAGE method for a broad range of applications involving large and heterogenic entities.

3. Crystallisation of the N-RNA rings

Crystallization screenings were carried out at the High Throughput Crystallization Laboratory of the EMBL Grenoble Outstation (HTX Lab.). Typically, 576 crystallization conditions were screened per sample using a PixSys4200 robot (Cartesian) and the vapour diffusion method in CrystalQuick™ (Greiner Bio-One) 96-well sitting drop crystallization plates with square wells. Crystallization experiments were set up by mixing 100 nl of protein solution with 100 nl of crystallization cocktail and were equilibrated against a reservoir containing 88 μl of crystallization cocktail. Screenings were set up of the N9, N10, and N11 fractions. All three fractions crystallised in 9.5% PEG 6K, 1.3 M LiCl

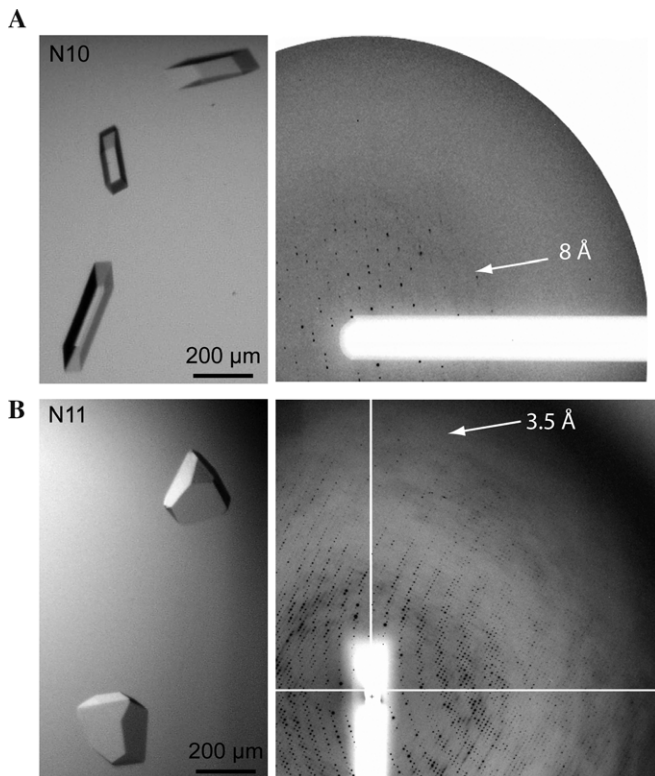


Fig. 3. Crystals of N10 and N11 N-RNA rings. (A) Crystals and diffraction pattern of N10 N-RNA rings. (B) Crystals and best diffraction pattern of N11 N-RNA rings grown in 60% TACSIMATE plus 10 mM L-cysteine.

and 0.1 M MES pH 6 (Fig. 3A). The crystals of the N9 fraction were too small to be analysed. The space group of the N10 crystal was C2 with $a=292 \text{ \AA}$, $b=314 \text{ \AA}$, $c=154 \text{ \AA}$, $\alpha=\gamma=90^\circ$, $\beta=118^\circ$. These crystals diffracted to 8 \AA and the resolution could not be improved by refinement screening. The N11 fraction also crystallised in 60% Tacsimate (Hampton research) plus 10 mM L-cysteine. Bigger crystals were grown manually using the hanging drop method, combining $1 \mu\text{l}$ of protein solution with a concentration of 5 mg/ml with $1 \mu\text{l}$ of the reservoir solution indicated above. These

crystals (Fig. 3B) belonged to space group $P2_12_12$ with $a=270 \text{ \AA}$, $b=281 \text{ \AA}$, $c=237 \text{ \AA}$, $\alpha=\beta=\gamma=90^\circ$. Fig. 2C shows a negative stain image of crushed crystals. Apart from end-on views there were many side-views of rings that were attached to each other with a twofold axis. A reconstruction of these double rings is shown in Fig. 2D. Indeed, in the crystal structure the asymmetric unit was found to be a 22-mer of two rings attached through their N-terminal domains (Albertini et al., 2006).

When the crystals were taken out of the drop and frozen directly by flash freezing in a stream of liquid nitrogen, the diffraction limit was only around 8 \AA (Fig. 4). However, when the crystals were kept in the air for 5 min before freezing the diffraction improved to 6 \AA ; after 10 min dehydration the best crystals diffracted to 4 \AA and, ultimately, after 15 min dehydration optimal diffraction was obtained at 3.5 \AA (Figs. 4 and 3B). When the crystals were kept in air for 20 min, the diffraction became anisotropic. The method used to solve the structure, the crystal structure of the ring of 11 N-protomers with its attached RNA (PDB Accession code No. 2GTT) and the statistics are published in Albertini et al., 2006. The residing RNA is clearly visible. It is indeed bound to the protein by the sugar-phosphate backbone but most of the bases are hidden by the protein that surrounds the RNA. This implies that the N-RNA structure will have to undergo conformational changes in order to allow access of the polymerase to the RNA. An antiviral compound could be designed that either blocks the protein in its closed conformation or force the protein to be continually open, allowing degradation of the RNA.

4. Comparison with the crystallisation of Vesicular Stomatitis Virus N-RNA rings

In parallel with our work on Rabies virus N-RNA, Ming Luo and co-workers crystallised the N-RNA rings of another virus of the same family as Rabies virus: Vesicular Stomatitis Virus (VSV) (Green and Luo, 2006) and solved the structure (Green et al., 2006). The structures of the rabies and VSV proteins are very similar and the way the

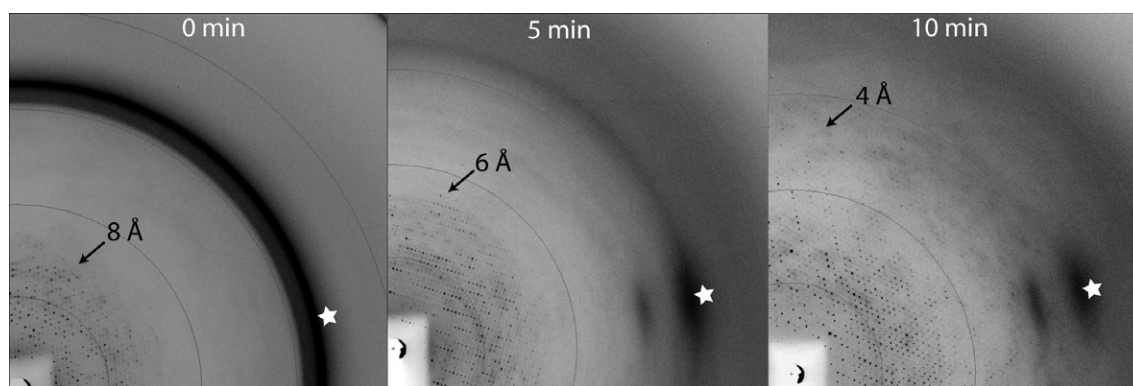


Fig. 4. Improvement of diffraction of the N11 crystals by dehydration before flash freezing. Crystals were held in air for 0, 5 or 10 min before flash freezing in a stream of liquid nitrogen and the diffraction limit improved from 8 to 4 \AA resolution. Optimal diffraction up to 3.5 \AA was obtained after 15 min of dehydration (Fig. 3B). On the different diffraction patterns, the rings mark 11.8^{-1} , 5.9^{-1} , 3.9^{-1} and 3^{-1} \AA^{-1} going from the inside to the outside and the star indicates the ice ring.

RNA is encapsidated is identical. For both systems it was essential that homogeneous populations were obtained for crystallization but, due to small biological differences between the proteins, the way this was achieved was very different. The VSV protein was expressed in *Escherichia coli* together with another viral protein, the phosphoprotein. In this system the main product was rings with 10 N-protomers plus bound phosphoprotein, which was removed by dialysis at pH 4 (Green and Luo, 2006). The fact that the phosphoprotein is a dimer (Ding et al., 2006) may be the reason why most rings had 10 N-protomers and not 9 or 11. However, a similar expression approach did not work for Rabies virus. The nucleoprotein of Rabies virus has to be phosphorylated in order to bind to RNA (Kawai et al., 1999) and this phosphorylation is necessary for the binding of P to N (Toriumi and Kawai, 2004). Perhaps for these reasons expression of N with or without P in bacteria did not give ordered N-RNA complexes (results not shown). Co-expression of N and P in insect cells did work but a mixture of free P-oligomers, RNA-free N-P complexes, N-RNA and P bound to N-RNA was produced (Mavrakis et al., 2003). However, the N-RNA complexes isolation from this double expression system were not more homogeneous than when N was expressed alone (data not shown).

Both the VSV and the Rabies virus N-RNA crystals diffracted poorly and different strategies were used in order to improve diffraction. As described above, the Rabies N-RNA crystals were improved by dehydration before freezing. The VSV N-RNA crystals produced by Green and Luo (2006) diffracted to about 6 Å resolution. Stepwise soaking of the crystals in uranyl acetate stabilised and improved the diffraction to a maximum of 2.9 Å. Soaking of the Rabies virus N-RNA crystals with uranyl acetate resulted in fracturing of the crystals. However, for both systems further soaking with a higher uranyl acetate concentration (VSV) and further dehydration (Rabies) both led to the crystals becoming anisotropic.

It is remarkable to notice how similar problems in closely related scientific projects have been overcome by completely different technical strategies. We can envision that detailed information on the original methods that successfully lead to obtain crystals and solve the structure of VSV and Rabies nucleoproteins will be useful for crystallographic projects involving large complexes in which heterogeneity is a major challenge.

Acknowledgments

We thank Florine Dupeux, Martin Rower, José-Antonio Marquez of The High Throughput Crystallization laboratory (HTX Lab) at the EMBL Grenoble Outstation for HT crystallization screening. This work was supported by the Université Joseph Fourier (R.W.H.R.), the CNRS (G.S.), the EMBL (A.K.W.; W.W.), the Deutsche Forschungsgemeinschaft SFB 593 (W.W.), an EMBO long term

postdoctoral fellowship (C.R.C.) and a Ph.D fellowship from the French Ministry for Education and Research (A.A.V.A.).

References

- Albertini, A.A.V., Wernimont, A.K., Muziol, T., Ravelli, R.B.G., Clapier, C.R., Schoehn, G., Weissenhorn, W., Ruigrok, R.W.H., 2006. Crystal structure of the rabies virus nucleoprotein–RNA complex. *Science* 313, 357–360.
- Childs, J.D., Birnboim, H.C., 1975. Polyacrylamide gel electrophoresis of intact bacteriophage T4D particles. *J. Virol.* 16, 652–661.
- Ding, H., Green, T.J., Lu, S., Luo, M., 2006. Crystal structure of the oligomerization domain of the phosphoprotein of Vesicular Stomatitis Virus. *J. Virology* 80, 2808–2814.
- Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., Luger, K., 2004. Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* 375, 23–44.
- Fabry, C.M.S., Rosa-Calatrava, M., Conway, J.F., Zubieta, C., Cusack, S., Ruigrok, R.W.H., Schoehn, G., 2005. A quasi-atomic model of human Adenovirus type 5 capsid. *EMBO J.* 24, 1645–1654.
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., Leith, A., 1996. SPIDER and WEB: processing and visualisation of images in 3D electron microscopy and related fields. *J. Struct. Biol.* 116, 190–199.
- Green, T.J., Luo, M., 2006. Resolution improvement of X-ray diffraction data of crystals of a vesicular stomatitis virus nucleocapsid protein oligomer complexed with RNA. *Acta Cryst. D62*, 498–504.
- Green, T.J., Zhang, X., Gail, W., Wertz, G.W., Luo, M., 2006. Structure of the vesicular stomatitis virus nucleoprotein–RNA complex. *Science* 313, 357–360.
- Iseni, F., Baudin, F., Blondel, D., Ruigrok, R.W.H., 2000. Structure of the RNA inside the Vesicular Stomatitis Virus nucleocapsid. *RNA* 6, 270–281.
- Iseni, F., Barge, A., Baudin, F., Blondel, D., Ruigrok, R.W.H., 1998. Characterization of Rabies Virus nucleocapsids and recombinant nucleocapsid-like structures. *J. Gen. Virol.* 79, 2909–2919.
- Iseni, F., Baudin, F., Garcin, D., Marq, J.B., Ruigrok, R.W.H., Kolakofsky, D., 2002. Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids. *RNA* 8, 1056–1067.
- Kawai, A., Toriumi, H., Tochikura, T.S., Takahashi, T., Honda, Y., Morimoto, K., 1999. Nucleocapsid formation and/or subsequent conformational change of Rabies Virus Nucleoprotein (N) is a prerequisite step for acquiring the phosphatase-sensitive epitope of monoclonal antibody 5–2–26. *Virology* 263, 395–407.
- Mavrakis, M., Iseni, F., Mazza, C., Schoehn, G., Ebel, C., Gentzel, M., Franz, T., Ruigrok, R.W.H., 2003. Isolation and characterisation of the rabies virus N^o-P complex produced in insect cells. *Virology* 305, 406–414.
- Nakai, T., Howatson, A.F., 1968. The fine structure of vesicular stomatitis virus. *Virology* 35, 268–281.
- Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D., Ruigrok, R.W.H., 2001. Structure of recombinant rabies virus N-RNA and identification of the phosphoprotein binding site. *J. Virol.* 75, 490–498.
- Schoehn, G., Mavrakis, M., Albertini, A., Wade, R., Hoenger, A., Ruigrok, R.W.H., 2004. 12 Å Structure of trypsin-treated Measles Virus N-RNA. *J. Mol. Biol.* 339, 301–312.
- 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.
- 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.

Précisions sur la méthode de purification des différentes espèces d'anneaux N-ARN et l'obtention des cristaux (complément à l'article I)

1) La méthode d'électrophorèse préparative

La masse moléculaire des complexes en forme d'anneaux entre la nucléoprotéine et l'ARN varie de 500 à 700 KDa. Leur forme générale diffère peu qu'ils soient composés de 9, 10, 11 ou plus de sous-unités de nucléoprotéine (notés respectivement N₉, N₁₀, et N₁₁). Des tentatives de séparation des différentes espèces par diverses méthodes biochimiques telles que des gradients plus fins ou des étapes supplémentaires de chromatographie ont été menées. Au mieux, elles ont permis l'obtention de fractions enrichies mais pas totalement pures en complexes N₁₀-ARN. Afin de résoudre ce problème d'hétérogénéité, nous avons étudié le comportement des complexes dans un gel natif afin de mettre au point une technique de séparation basée sur le principe de l'électrophorèse préparative en condition native.

(a) Mise au point du gel natif

La mise au point de la composition du gel natif analytique est importante en vue de l'utilisation de l'appareil d'électrophorèse préparative. Il s'agit de visualiser au préalable en «deux dimensions» la séparation des différents complexes avant de réaliser une expérience préparative en «trois dimensions». La migration du gel natif 2D s'effectue à 4 °C avec un recyclage continu du tampon de migration (0.2 x TBE) afin de pouvoir conserver des conditions similaires lors du passage en mode préparatif.

Le pourcentage d'acrylamide optimal est de 4 %. En effet, avec un gel à 5 % la résolution des bandes n'est pas satisfaisante, et en dessous de 4 % le gel ne possède pas une consistance suffisante pour être manipulé.

Le ratio acrylamide / bisacrylamide est aussi très important, il détermine le degré de réticulation du gel. Après en avoir testé plusieurs, le ratio 19:1 a été retenu car il permet d'obtenir la meilleure séparation sur les gels natifs.

La résolution des bandes sur le gel est directement reliée au volume d'échantillon chargé. La Figure 27 représente le gel obtenu lorsque deux volumes différents d'échantillon ont été déposés. Plus le volume déposé est important, plus le phénomène de diffusion est conséquent. Sur le gel de la Figure 27, ceci se traduit par une augmentation de l'épaisseur des bandes. Sur la ligne A, où un volume de 10 µl (1 mm de hauteur) a été chargé, le résultat est un «smear».

Ce volume d'échantillon est « non résolutif ». Différents essais de volumes (3 à 10 μ l) ont été réalisés et nous avons déterminé que le volume résolutif limite est de 5 μ l (Figure 27, ligne B). Les bandes sont alors bien séparées les unes des autres (0.8 cm). Il est important de tenir compte de ces volumes lors du dépôt de l'échantillon pour l'électrophorèse préparative en condition native.

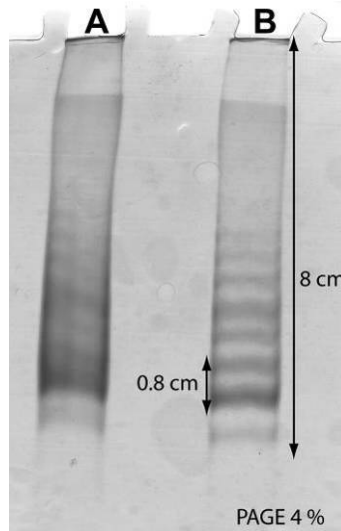


Figure 27 : Variation de la résolution des bandes protéiques en fonction du volume chargé sur le gel.

Gel natif 4 % Acrylamide : Bisacrylamide (19 : 1), Coloration au bleu de coomassie, temps de migration : 6 heures à 120 Volts à une température de 4 °C.

Ligne A : 10 μ l d'échantillon concentré chargé (hauteur de 1 mm dans le puits).

Ligne B : 5 μ l d'échantillon concentré chargé (hauteur de 0.5 mm dans le puits).

(b) Principe de l'électrophorèse préparative en conditions natives

Cette méthode repose sur le principe de la migration des macromolécules en conditions natives. Il s'agit de reproduire à l'échelle préparative en «trois dimensions» un résultat obtenu au préalable analytiquement avec un gel non dénaturant plat. A cette fin, un système de gel cylindrique a été élaboré par la société Biorad (Figure 28 A). A l'intérieur du gel, la vitesse de migration va être directement reliée à la masse moléculaire des complexes : plus la masse moléculaire du complexe est élevée moins sa migration est rapide. A la sortie du gel, les espèces protéiques sont entraînées par un flux continu de tampon et sont alors collectées de manière séquentielle.

L'expérience se déroule en trois temps (Figure 28, B) :

- **Le chargement de l'échantillon** : cette étape est cruciale car la hauteur du dépôt va déterminer directement la résolution des bandes de protéines (voir plus haut).
- **La migration** : elle se déroule à un voltage constant de 500 V.

- **L'éluion** : un débit d'éluion constant (2ml/min) permet de laver la base du gel natif et de collecter les différents complexes à leur sortie. La vitesse d'éluion doit être suffisante pour ne pas mélanger les différentes espèces oligomériques. La contrepartie de cette éluion est une importante dilution de l'échantillon (1500 x). Pour permettre la détection de la sortie du matériel par UV ($\lambda = 280$ et 260 nm), il est nécessaire que l'échantillon déposé soit très concentré. Les complexes N-ARN de la rage peuvent être concentrés à 100 mg/ml. La détection des fractions éluées reste alors efficace malgré une dilution conséquente.

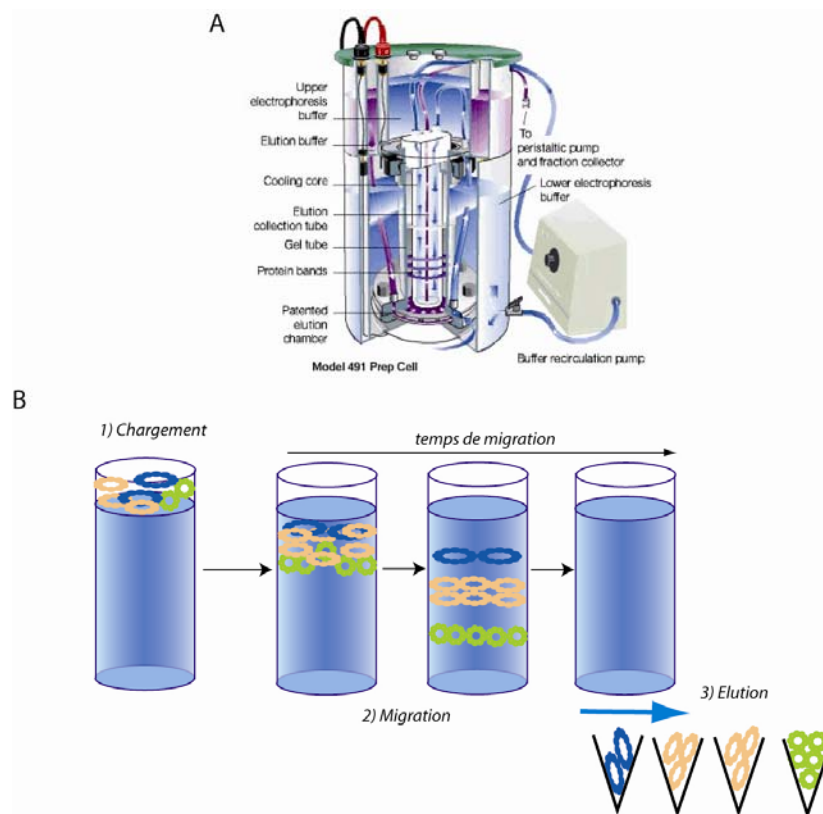


Figure 28 : Représentation schématique de l'expérience d'électrophorèse préparative.

A : Schéma de l'appareil d'électrophorèse préparative (<http://www.bio-rad.com>).

B : Les trois étapes principales sont 1) Chargement de l'échantillon, 2) Migration à voltage constant, 3) Elution et collecte séquentielle des complexes à leur sortie du gel natif.

(c) Mise au point de l'électrophorèse préparative

Après la mise au point du gel natif plat, il a été nécessaire d'affiner plusieurs paramètres pour l'étape d'électrophorèse préparative. Le voltage, la vitesse d'éluion, le diamètre, et la hauteur du gel sont des paramètres déterminants.

Le voltage est un paramètre important pour cette expérience. En réduisant le voltage, nous avons obtenu une meilleure séparation mais avec un temps de migration bien plus long (9

heures à 300 Volts au lieu de 6 heures à 500 Volts). Nous avons donc préféré faire varier d'autres paramètres pour des questions d'ordre pratique.

Changer la vitesse du flux d'élution pour récolter le matériel dès sa sortie du gel a permis d'améliorer de façon significative la séparation des espèces. Malgré les spécifications techniques de Biorad (débit d'élution < 1ml/min) nous avons augmenté le débit d'élution à 2 ml/min. Si la séparation de l'échantillon a été améliorée, nous avons également très significativement dilué l'échantillon : l'échantillon concentré d'un volume initial de 0.1 ml est dilué dans un volume d'environ 150 ml lors de la séparation.

Le paramètre le plus important demeure la hauteur du gel natif à l'intérieur du système. Ce gel natif est en forme de cylindre creux, il est traversé en son centre par un système de refroidissement et de collecte des échantillons élués. Les profils d'élution de deux électrophorèses préparatives avec des gels de hauteurs différentes sont présentés sur la Figure 29. Huit centimètres séparent la première de la dernière bande directement sur le gel natif 2D de la Figure 27 B. L'électrophorèse préparative 3D avec un gel de la même hauteur (8 cm) ne génère pas des pics bien séparés. Avec un gel de 12.5 cm (hauteur maximale permise par le dispositif préparatif), les pics sont séparés convenablement (Figure 29 B).

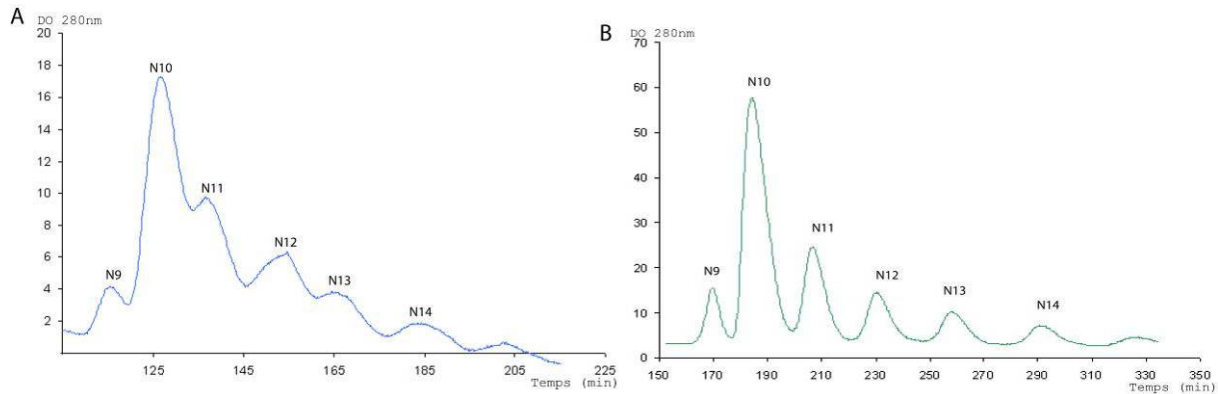


Figure 29 : Profils d'élution obtenus lors de deux expériences d'électrophorèse préparative qui diffèrent par la hauteur du gel utilisé.

A : Gel de 8 cm de haut. B : Gel de 12.5 cm de haut

Le contenu de chaque pic a fait l'objet d'une observation en microscopie électronique afin de vérifier la stœchiométrie de la nucléoprotéine dans l'échantillon. Le premier pic contient les anneaux à 9 sous-unités de nucléoprotéine, le second les anneaux à 10 sous-unités de nucléoprotéine et ainsi de suite. Après chaque électrophorèse, un gel natif plat permet d'apprécier la qualité de la purification et de vérifier qu'il n'y ait pas de contamination des

pics par les pics voisins. (cf. article I « Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein–RNA rings »).

2) Cristallogénèse

Les tests de cristallogénèse ont été réalisés sur les échantillons contenant 9, 10, et 11 sous-unités de nucléoprotéine (notés respectivement N₉, N₁₀, N₁₁) et ont permis l'obtention de nombreuses formes cristallines (Figure 30).

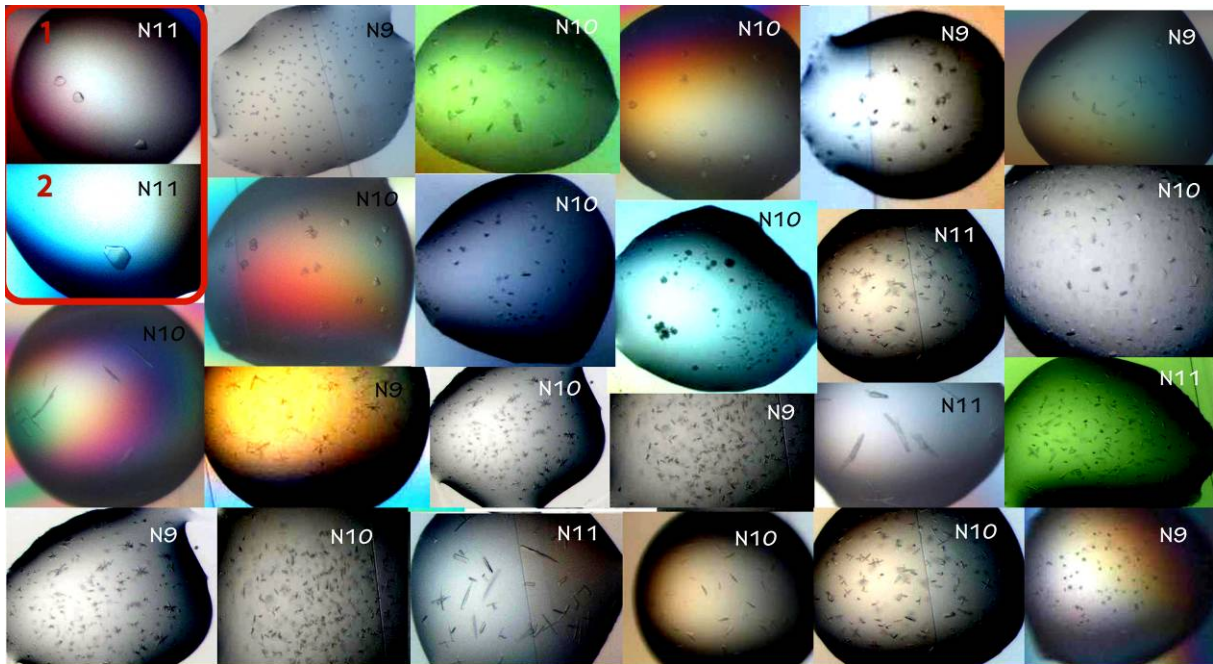


Figure 30 : Cristaux obtenus en nanogouttes (0.1 µl + 0.1 µl) avec le robot cartésien de cristallisation.

Toutes les conditions contiennent du PEG comme précipitant sauf celles de l'encadré rouge.

Cadre rouge : cristaux de N₁₁ qui ont pu être directement mesurés sur les lignes du synchrotron. Condition 1 : 60 % Tacsimate, condition 2 : 1.8 M Na/K Phosphate. La nature de l'anneau cristallisé est indiquée sur chaque image.

(a) Cristallogénèse des anneaux à 10 sous-unités de nucléoprotéine (N₁₀)

Nous avons commencé par travailler sur les N₁₀ car ils représentaient l'espèce majoritaire dans le mélange d'anneaux. Cet échantillon a cristallisé dans de nombreuses conditions. Ces conditions sont le plus souvent à base de PEG de masse moléculaire moyenne (4kD à 6kD) comme précipitant et de sel de lithium (LiCl, Li₂SO₄). La condition de base est : 9.5 % PEG 6k, 0.1 M MES pH 6, 1.3 M LiCl. Les premiers cristaux étaient de petites plaques dont la plus grande dimension mesurait 40 µm. Ces cristaux ont pu être testés en septembre 2004 sur la

ligne microfocuss ID23 à l'ESRF. La résolution limite des taches de diffraction atteignait environ 25 Å (17 taches). Le cristal subissait des dommages d'irradiation sévères dus à l'intensité du rayonnement conduisant à sa destruction rapide lors de la prise des premiers clichés de diffraction (3 images). La taille de ces cristaux a pu être augmentée en utilisant des additifs tels que l'EDTA, le BaCl₂ ou encore la glycine. Des cristaux dont la dimension principale était comprise entre 300 et 500 µm ont alors pu être obtenus. La qualité de la diffraction n'a cependant pas évolué aussi bien que la taille de ces cristaux. Dans le meilleur des cas, les cristaux de N₁₀ ont diffracté à une résolution de 8 Å.

Malgré le test et l'utilisation de plusieurs cryoprotectants (Tableau 2), la qualité de la diffraction n'a jamais pu être améliorée.

Tableau 2 : Agents cryoprotectants testés avec les cristaux de N₁₀.

Agent cryo-protectant	Pourcentage utilisé
Ethylène Glycol	20 % - 30 %
Glycérol	20 % - 30 %
PEG 400 et PEG 200	20 % - 30 %
MPD	30 % - 40 %
Sucrose	30 % - 50 %
Xylitol	30 % - 40 %
Erythritol	30 % - 40 %
Dioxane	40 %
LiCl	5 M

La fragilité des cristaux de N₁₀ vis-à-vis du faisceau a constitué un obstacle important à l'obtention d'un résultat satisfaisant. Cette fragilité était telle que nous n'avons jamais obtenu un jeu de données complet. Le groupe d'espace des cristaux de N₁₀ a tout de même pu être déterminé sur plusieurs jeux de données partiels : le groupe d'espace est C2 (monoclinique, 4 unités asymétriques dans la maille) avec pour paramètres de la maille $a = 292 \text{ \AA}$, $b = 314 \text{ \AA}$, $c = 154 \text{ \AA}$, et $\alpha = \gamma = 90^\circ$, $\beta = 118^\circ$.

Pour essayer d'expliquer la fragilité de nos cristaux, nous avons estimé le pourcentage de solvant présent à l'intérieur des cristaux grâce au coefficient de Matthews (Matthews, 1968) (Tableau 3). En effet plus ce pourcentage est élevé, plus le cristal contient une forte proportion de solvant et donc plus il sera susceptible d'être sensible aux dommages d'irradiation. Nous avons fait l'estimation du pourcentage de solvant avec différents nombres d'anneaux dans l'unité asymétrique (Tableau 3).

En général, un cristal de protéine contient entre 30 et 70 % de solvant ($1.62 < V_M < 3.53 \text{ \AA}^3/\text{Da}$). Nous connaissons la taille d'un anneau N₁₀, 160 Å x 160 Å x 80 Å (Schoehn et al., 2001), et

les paramètres de maille des cristaux. Ceci semble indiquer que l'unité asymétrique serait composée de deux anneaux N_{10} soit 8 anneaux en tout dans la maille élémentaire. Dans ce cas, le pourcentage de solvant est alors estimé à 60% (Tableau 3). Ce pourcentage de solvant n'est pas particulièrement élevé. Toutefois, nous avons pensé que des tests de déshydratation entraînant une diminution du pourcentage de solvant pourrait conduire à l'obtention de cristaux de meilleure qualité.

Tableau 3 : Contenu en solvant estimé par la méthode de Matthew.

Nombre d'anneaux dans l'unité asymétrique	Coefficient de Matthews en ($\text{\AA}^3 / \text{Da}$)	Pourcentage de solvant (%)
0.5	11.33	89.15
1	5.67	78.3
2	2.83	56.59

Des tentatives de déshydratation des cristaux ont donc été menées. Nous avons essayé la technique de déformation des cristaux («shrinkage»), qui consiste à tremper les cristaux par étapes dans des solutions de cryoprotectant contenant un PEG de poids moléculaire plus élevé (8k) et/ou à un pourcentage plus important (Schick and Jurnak, 1994 ; (Kawashima et al., 1996). Cette technique de déshydratation n'a pas permis l'amélioration des clichés de diffraction. Nous avons ensuite tenté de déshydrater les cristaux en utilisant des « cryo-sels » comme le LiCl, le NaCl, le Li_2SO_4 , et le LiOAc (Kawashima et al., 1996; Rubinson et al., 2000; Schick and Jurnak, 1994) . Ces composés sont solubles à de hautes molarités, l'addition de ces sels entraîne une déshydratation du cristal. De plus l'utilisation de tels composés comme cryoprotectant est une alternative à l'utilisation des composés organiques. Nous avons remarqué que l'utilisation du LiCl à 5M en particulier permettait de réduire l'anneau de glace sur les clichés de diffraction. En revanche, la résolution de diffraction ne présentait pas d'amélioration si ce n'est une sensibilité moindre du cristal aux dommages causés par les radiations du faisceau de rayons X.

Une autre hypothèse pour expliquer le faible pouvoir de diffraction des cristaux est la présence d'ARN hétérogène au sein de l'échantillon d'anneau N_{10} . Si le brin d'ARN est composé de plus de 90 nucléotides, un morceau d'ARN «flottant» hors de l'anneau pourrait entraîner un arrangement irrégulier à l'intérieur des cristaux. Nous avons alors entrepris la cristallisation des N_{10} après digestion à la ribonucléase A. Ce traitement n'a cependant pas non plus permis d'améliorer le pouvoir de diffraction des cristaux.

Malgré tout, la possibilité d'obtenir des cristaux de meilleure qualité avec les N_{10} n'est toujours pas exclue.

(b) Cristallogénèse des anneaux à 11 sous-unités de nucléoprotéines (N₁₁)

Les N₁₁ cristallisent dans des conditions similaires aux N₁₀ : PEG moyen, haut sel, et large gamme de pH (tampon : acide citrique pH 5, à bicine pH 9). La qualité de diffraction des cristaux N₁₁ testés ne présentait pas d'amélioration sensible. Nous avons alors décidé de nous focaliser sur deux conditions de cristallisation qui ont généré des cristaux suffisamment gros pour être testés directement sur les lignes de lumière. (Figure 30, encadré rouge). La première condition est uniquement à base de tampon Phosphate (1.8 M Na/K Phosphate). Le cristal contenu dans cette condition a diffracté autour de 7 Å mais il ne supportait pas très longtemps l'exposition aux rayons X.

La seconde condition est composée de 60 % Tacsimate. Le Tacsimate est un produit développé par la société Hampton Research qui est composé d'un mélange non équimolaire à pH 7 de malonate de sodium, d'acétate de sodium, de tri-ammonium citrate, d'acide succinique, d'acide DL-malique, de formate de sodium, et de di-ammonium tartrate. Ce mélange a été mis au point en se basant sur les travaux de cristallogénèse en condition salines (McPherson, 2001).

Cette condition a de plus l'avantage d'être cryoprotectante, ce qui rend le test des cristaux plus simple et plus rapide. Les minicristaux obtenus en nanogouttes avec 60 % de Tacsimate ont diffracté à 6 Å. Une fois reproduits manuellement, leur taille et forme ont été améliorées par l'ajout d'un additif adéquat (10 mM L-cystéine), et nous avons pu obtenir un jeu de donnée complet à 3.5 Å de résolution (cf. article I).

IV. Conclusion

L'obtention de cristaux avec le complexe N°-P digéré nous a donné de grands espoirs quant à la détermination d'une structure de la nucléoprotéine. Ils ont été rapidement évincés par l'impossibilité de reproduire ces cristaux. L'obtention de cristaux et la résolution d'une structure du complexe entre la nucléoprotéine et la phosphoprotéine reste un objectif important pour comprendre le mécanisme d'assemblage d'une nucléocapside et les changements de conformation que la nucléoprotéine peut subir. La protéolyse limitée n'est pas toujours le meilleur moyen pour obtenir un échantillon reproductible car elle met en jeu de nombreux facteurs. Une approche différente pourrait donner de meilleurs résultats. La construction de formes tronquées de nucléoprotéine et de la phosphoprotéine pourrait éventuellement améliorer l'homogénéité de l'échantillon N°-P_{dig.} Ces approches en combinaison avec des conditions de cristallogenèse supplémentaires maintenant disponibles sur la plateforme de cristallisation devraient permettre dans un avenir proche d'obtenir de nouveaux cristaux du complexe N°-P sous sa forme native ou protéolysée.

Concernant la purification des anneaux de nucléoprotéine, la mise au point de l'étape d'électrophorèse préparative a nécessité de nombreux essais. Cette partie du travail nous a permis de découvrir qu'il était possible de résoudre le problème de purification d'un mélange de gros complexes proches en taille. L'étape d'électrophorèse préparative est une méthode pouvant s'appliquer à la purification de gros complexes dont la stœchiométrie n'est pas homogène. La séparation d'oligomères n'est pas aisée dans la plupart des cas à cause d'un faible écart de masse moléculaire entre les différentes espèces de formes trop similaires, et de propriétés physico-chimiques voisines. Il est très probable que cette technique puisse s'appliquer à d'autres gros complexes et permettre ainsi la purification d'échantillons plus homogènes en vue d'une étude par microscopie électronique et/ou par la cristallographie aux rayons X.

**Partie IV : Structure cristallographique du complexe
ARN-Nucléoprotéine du virus de la rage**

I. Introduction

Au début de mon travail de thèse, aucune structure à haute résolution de nucléoprotéine de virus à ARN négatif en complexe avec l'ARN n'était disponible. Les seules informations structurales de ces complexes étaient basées sur des observations et des reconstructions par microscopie électronique. (Egelman et al., 1989; Jeng et al., 1989; Schoehn et al., 2001 ; Schoehn et al., 2004).

L'obtention d'un jeu de données natives à 3.5 Å de résolution nous a permis de résoudre la structure cristallographique des complexes composés de 11 sous-unités de nucléoprotéine et d'un brin d'ARN (N₁₁-ARN). L'obtention des phases nécessaires à cette résolution a posé de nombreux problèmes. Le marquage des cristaux avec les dérivés d'atomes lourds n'étant pas très concluant, plusieurs stratégies ont été envisagées pour résoudre cette structure. Nous avons en premier lieu, tenté sans succès de résoudre la structure par remplacement moléculaire avec la carte de microscopie électronique réalisée à partir d'images de cristaux de N₁₁ broyés. Ensuite, après obtention d'un jeu de données sur des cristaux marqués à l'or, la structure a finalement pu être résolue par la méthode SAD (Single Anomalous Diffraction). Une collaboration avec plusieurs cristallographes de l'EMBL a été mise en place afin de résoudre et d'affiner la structure. En effet, la résolution de cette structure a rassemblé plusieurs difficultés que l'on peut rencontrer en cristallographie :

- 1) la basse résolution (3.5 Å) ne facilite pas l'assignement des acides aminés et le positionnement des chaînes latérales.
- 2) Les 22 nucléoprotéines présentes dans l'unité asymétrique ne sont pas équivalentes.
- 3) Les différents protomères de nucléoprotéine présentent deux domaines interchangeables (« domain swapping ») très flexibles et difficiles à fixer dans la densité électronique.
- 4) Le nombre de résidus est très important ce qui augmente considérablement le temps de calcul, en particulier lors des étapes d'affinement.

La résolution de cette structure n'est que de 3.5 Å. Ceci ne permet pas d'analyser d'une manière poussée les contacts entre les chaînes latérales des acides aminés. En revanche, la localisation de l'ARN au sein de la structure ainsi que l'arrangement des protomères les uns par rapport aux autres reste sans aucune ambiguïté. Ceci nous a permis de réaliser des observations intéressantes pour expliquer l'accès, la protection de l'information génétique virale et l'assemblage des complexes entre la nucléoprotéine et l'ARN viral en général.

Ce travail a fait l'objet d'une publication dans le journal *Science* : Albertini, A.A., Wernimont, A.K., Muziol, T., Ravelli, R.B., Clapier, C.R., Schoehn, G., Weissenhorn, W. and Ruigrok, R.W. (2006). Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science*, **313**, 360-363.

II. ARTICLE II : Crystal structure of the rabies virus nucleoprotein-RNA complex

points of notable rotation. We estimate that about 38.5 maneuvered N-RNA complex units are required to complete one round of the RNP structure with a radius of 245 Å.

The sequence of the vRNA in the RNP must be read by the polymerase complex during RNA synthesis. There are three possible mechanisms for copying the RNA sequence: (i) The vRNA is completely exposed in the RNP, so Watson-Crick base-pairing can occur without any change of the RNP structure. (ii) The vRNA is completely dissociated from the RNP, so it serves as a template like a naked RNA molecule. (iii) A pronounced structural change occurs in the RNP to allow the sequence of the vRNA to be read by the polymerase complex without disrupting the integrity of the RNP. The conformational arrangement of the RNA in the N protein as revealed by our structure suggests that Watson-Crick base-pairing could not occur when the N protein is closed on the vRNA. Bases in positions 5, 7, and 8 are completely shielded by the N protein such that their backbone conformation is held rigidly by the N protein, thus preventing the formation of an RNA duplex. The second possibility is also unlikely because the RNP remains intact after one round of RNA

synthesis and may be used as the template again. Thus, it is likely that the vRNA is temporarily dissociated from the N protein within the active polymerase complex.

References and Notes

1. S. U. Emerson, R. R. Wagner, *J. Virol.* **10**, 297 (1972).
2. S. Naito, A. Ishihama, *J. Biol. Chem.* **251**, 4307 (1976).
3. A. S. Huang, R. R. Wagner, *J. Mol. Biol.* **22**, 381 (1966).
4. D. Thomas *et al.*, *J. Virol.* **54**, 598 (1985).
5. F. Iseni, F. Baudin, D. Blondel, R. W. Ruigrok, *RNA* **6**, 270 (2000).
6. S. U. Emerson, Y. Yu, *J. Virol.* **15**, 1348 (1975).
7. M. Howard, G. Wertz, *J. Gen. Virol.* **70**, 2683 (1989).
8. A. F. Howatson, G. F. Whitmore, *Virology* **16**, 466 (1962).
9. R. M. McCombs, M. B. Melnick, J. P. Brunschwig, *J. Bacteriol.* **91**, 803 (1966).
10. T. Nakai, A. F. Howatson, *Virology* **35**, 268 (1968).
11. G. Schoehn, F. Iseni, M. Mavrikis, D. Blondel, R. W. Ruigrok, *J. Virol.* **75**, 490 (2001).
12. J. Martin-Benito *et al.*, *EMBO Rep.* **2**, 313 (2001).
13. Z. Chen, T. J. Green, M. Luo, H. Li, *Structure* **12**, 227 (2004).
14. T. J. Green *et al.*, *J. Virol.* **74**, 9515 (2000).
15. T. J. Green, M. Luo, *Acta Crystallogr. D* **62**, 498 (2006).
16. J. D. Keene, B. J. Thornton, S. U. Emerson, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6191 (1981).
17. M. Carson, *Methods Enzymol.* **277**, 493 (1997).
18. W. L. Delano, *The PyMOL User's Manual* (Delano Scientific, San Carlos, CA, 2002).
19. We thank L. Andrew Ball for critical reading of the manuscript before submission. We thank the staff of the South East Regional Collaborative Access Team (SER-CAT)

and Bio-CARS at the Advanced Photon Source, Argonne National Laboratory, for their assistance in data collection. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract W-31-109-Eng-38. SER-CAT supporting institutions may be found at www.ser-cat.org/members.html. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the U.S. Department of Energy, Office of Biological and Environmental Research, and by the NIH, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. This work is supported in part by NIH grants AI050066 (to M.L.) and R37AI012464 (to G.W.W.). The atomic coordinates and reflection file have been deposited with the Research Collaboratory for Structural Bioinformatics Protein Data Bank. The access code for the coordinates is 2GIC.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1126953/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
Table S1
References

2 March 2006; accepted 5 June 2006

Published online 15 June 2006;

10.1126/science.1126953

Include this information when citing this paper.

Crystal Structure of the Rabies Virus Nucleoprotein-RNA Complex

Aurélien A. V. Albertini,^{1*} Amy K. Wernimont,^{2*} Tadeusz Muziol,² Raimond B. G. Ravelli,² Cedric R. Clapier,² Guy Schoehn,¹ Winfried Weissenhorn,^{2†} Rob W. H. Ruigrok¹

Negative-strand RNA viruses condense their genome into a helical nucleoprotein-RNA complex, the nucleocapsid, which is packed into virions and serves as a template for the RNA-dependent RNA polymerase complex. The crystal structure of a recombinant rabies virus nucleoprotein-RNA complex, organized in an undecameric ring, has been determined at 3.5 angstrom resolution. Polymerization of the nucleoprotein is achieved by domain exchange between protomers, with flexible hinges allowing nucleocapsid formation. The two core domains of the nucleoprotein clamp around the RNA at their interface and shield it from the environment. RNA sequestering by nucleoproteins is likely a common mechanism used by negative-strand RNA viruses to protect their genomes from the innate immune response directed against viral RNA in human host cells at certain stages of an infectious cycle.

Rabies virus, a member of the *Rhabdoviridae* family, is the causative agent of rabies, a fatal central nervous system disease (1), which constitutes a serious health problem in developing countries that lack effective vaccination programs (2). *Rhabdoviridae*, together with *Paramyxoviridae* (e.g., measles virus), *Filoviridae* (e.g., Ebola virus) and *Bornaviridae* (e.g., Borna disease virus),

are RNA-containing enveloped viruses that use nonsegmented negative sense RNA as their genome. The RNA is condensed by the nucleoprotein (N) into a helical nucleocapsid (NC) (3) and this N-RNA complex constitutes the essential template for replication by the RNA-dependent RNA polymerase complex (4). The polymerase complex selects for either transcription or replication and is composed of the enzymatic active L protein and the phosphoprotein P (5–7). Replication produces a full-length (+) copy of the viral RNA (vRNA), which is the specific target for encapsidation by N and serves as a template for (–) RNA replication. The switch from transcription to replication is in part regulated by the abundance of

free N in the cytoplasm (8). In order to understand the role of N-RNA complexes in the viral life cycle, we solved the crystal structure of a nucleoprotein oligomer from rabies virus complexed to a 99-nucleotide-long RNA segment.

Recombinant expression of most viral nucleoproteins from negative-strand RNA viruses leads to nonspecific host cell RNA encapsidation by N (9, 10), resulting in either helical or ring-like structures depending on the length of the RNA (10, 11). Rabies virus N expression in insect cells produced N-RNA rings containing 9, 10, 11, 12, or 13 copies of N as determined after purification (12). The undecameric ring produced the best diffracting crystals belonging to space group P2₁2₁2. The structure was solved by single anomalous dispersion (SAD) and refined to 3.5 Å resolution with an *R*-factor of 27.5 (*R*_{free} = 32.6) (table S1) (13). Two ~550-kDa large undecameric N-RNA rings pack head to head in the crystal asymmetric unit. Each ring has an outer diameter of 160 Å, an inner diameter of 60 Å, and a height of 74 Å (Fig. 1A). The N protomer consists of two main domains, which contact nine nucleotides of single-stranded RNA, as predicted (14), that are occluded in the center of the ring (Fig. 1, B, C, and D, and Fig. 2). The N-terminal core domain (NTD; residues 32 to 233) folds into a helical arrangement composed of 6 helices connected by large loops. The C-terminal core domain (CTD; residues 236 to 356 and 396 to 450) is composed of 11 helices joined by tighter loops (Figs. 1C and fig. S1). Two regions in N (NTD: 105 to 118 and CTD: 376 to 397) are

¹Institut de Virologie Moléculaire et Structurale, FRE 2854 Université Joseph Fourier-CNRS, ²European Molecular Biology Laboratory (EMBL), Boite Postale 181, 38042 Grenoble, France.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: weissen@embl-grenoble.fr

presumably flexible, because they are absent in the structure (Fig. 1C and fig. S1). In addition to the core domains, two smaller subdomains participate in domain exchange between protomers and stabilize polymerization (Figs. 1C and 3, A and B). Both the top (NTD) and bottom (CTD) domains act as “jaws” that clamp down onto the RNA strand and enclose it completely (Fig. 1D), an observation which is consistent with the fact that the RNA remains bound to N in CsCl gradients (15). The closest contact between the two jaws is between NTD residue N157 and CTD residue P435 (9 to 10 Å, depending on the protomer) (Fig. 1D). One likely reason for tight RNA sequestering is to prevent immune recognition. Tightly packaged RNA does not constitute a target for the innate immune system such as Toll-like receptors, which are present during transport of the nucleocapsid along the endosomal pathway into the cytoplasm to its replication site (16, 17) and is protected from exonuclease activities triggered by the interferon antiviral defense system (16, 18). In addition, complete genome protection could be crucial during transport of nucleocapsids to the site of virus assembly and budding.

The experimental electron density map showed continuous density representing the sugar-phosphate backbone and averaged densities for the bases indicating a 5′-3′ direction in clockwise orientation in the ring-structure (Figs. 1B and 2A). Although each N-RNA ring bound short random RNA from the expression host, differences between purines and pyrimidines could be discerned in the electron density map based on the angles and the sizes of the nucleotides, and they were modeled as either

adenine or cytosine (Fig. 2A). The RNA strand is twisted clockwise in an irregular left-handed helix along the inner perimeter of the ring within the continuous cleft made up by the NTD (top) and CTD (bottom) interface in the ring structure (Fig. 1B and fig. S2). For each protomer, the RNA strand is roughly split into two halves, which wrap around the NTD jaw; this arrangement provides most of the interactions and forces the RNA to bulge out at the tip of the NTD jaw (Figs. 1D and 2A). Each nucleotide makes polar contacts either via its phosphate group (seven out of nine phosphates are recognized), as predicted (19), or its ribose moiety (two out of nine are recognized), which is modeled in the C3′ *endo* sugar pucker conformation (Fig. 2, A and B). The first three bases point toward the solvent and stack onto each other and onto the last two bases of the preceding N protomer (M−1). Nucleotide 5 is involved in a kink that allows the base of nucleotide 4 to stack onto the bases from nucleotides 6 and 7 pointing toward the protein moiety, while its own base points away from the protein. The last two bases then point again away from the protein (Fig. 2, A and B). Most of the basic residues involved in RNA coordination are conserved between rabies virus and vesicular stomatitis virus (genus *Vesiculovirus*), indicating a conserved RNA coordination network (fig. S1).

There are no significant interactions between the NTDs within the ring structure (Fig. 1B). Their position is mainly determined by the connection to the CTD, the bound RNA, and extensive crystal contacts between NTDs of the two rings in the crystal asymmetric unit. In contrast, the CTDs share a large interaction

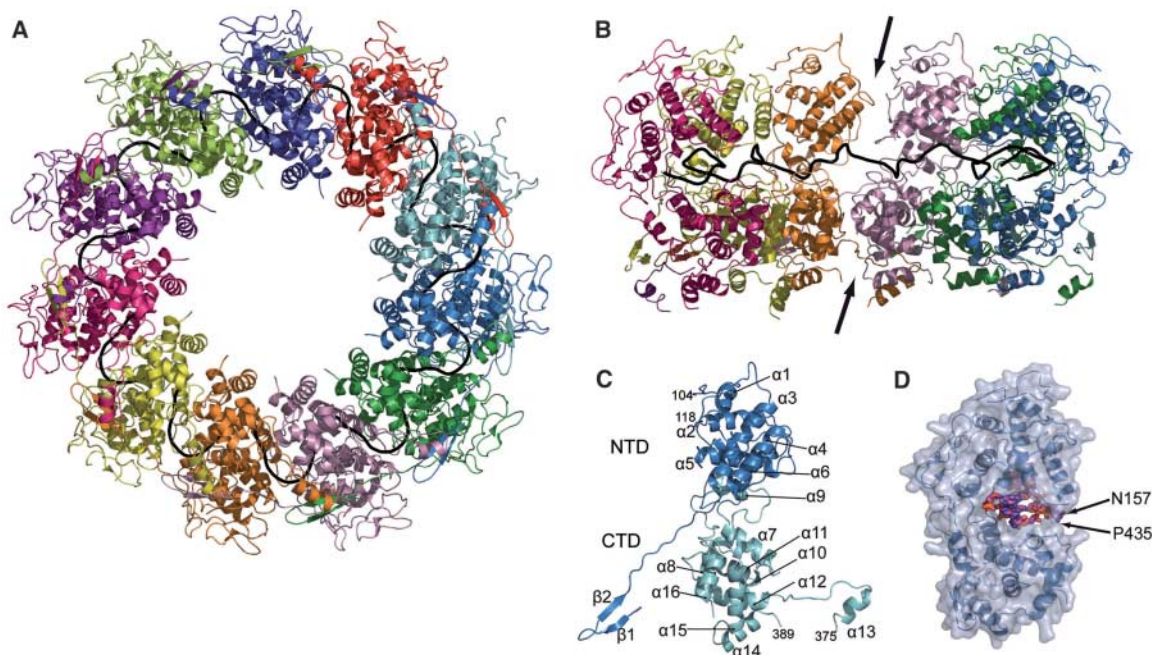
surface (2700 Å² total buried surface) (Fig. 1B) that is determined by hydrophobic and van der Waals contacts as well as multiple polar interactions (total of 15) involving many main chain contacts (total of nine contacts).

Two small subdomains emerge from NTD and CTD (Fig. 1C) and reach over to neighboring protomers, contacting them either clockwise (NTD; M+1) or counter clockwise (CTD; M−1) and so establish domain swapping (Fig. 3A). The extreme N terminus folds into a short β hairpin that nestles between the 3/10 helix η1 and helix 8 (fig. S1) of the CTD from M+1. The main contacts are made between the N11 and E266, as well as E20 and R254 in some monomers, in addition to hydrophobic interaction (V10, I22) (22). The β hairpin is followed by a stretch of residues (23 to 29) that are completely solvent exposed and constitute a potential hinge region between the subdomain and the core NTD, while Y28 and Y30 are used as anchor residues (Fig. 3B).

A second small subdomain emanates from the CTD as a coil region followed by helix 13 that reaches over to a neighboring protomer (M−1) in counterclockwise fashion (Fig. 3, A and B). The connection from helix 13 back to the core of the main CTD is disordered in all protomers (Fig. 1C). The CTD subdomain is much shorter and is firmly attached to the CTD by a hydrophobic core (F350, F349, F355), which is followed by a shorter potential hinge region (residues 351 to 356). The following loop region leading to helix 13 interacts with M−1 (R361 to E403; R357 to E403) and is in close contact with β strand 1 from M+1 (Fig. 3B).

Fig. 1. Overall structure of the N-RNA complex.

(A) Ribbon diagram of the 11-nucleotide oligomer N-RNA ring structure as viewed from the bottom. Each N protomer is colored differently. The RNA is shown as a black coil. (B) View of the inside of the ring structure; only six protomers are depicted in different colors with the NTD on the top and the CTD on the bottom. The path of the RNA is shown as a coil in a clockwise 5′ to 3′ orientation. The top arrow indicates that the NTDs do not interact with each other; the bottom arrow points toward the extensive interface between CTDs. (C) Ribbon diagram of the N protomer; the NTD is shown in dark blue and the CTD in light blue. The helical secondary structure elements are numbered consecutively. (D) Space-filling model of the N



protomer reveals that the RNA is completely clamped at the interface of the NTD and the CTD and thus is not accessible as a template for the polymerase.

The structural overlay of the Calpha atoms of the 11 protomers reveals that the two swapped domains and an NTD region present the highest level of flexibility (fig. S3); this supports the notion that the two subdomains could act as potential hinges and lead to lateral opening of the ring. This flexible domain linkage thus permits not only the formation of differently sized ring-like structures but also the assembly of the nucleocapsid that adopts a diameter of ~75 nm, containing ~53 protomers per helical turn (12).

The tight sequestering of the RNA suggests that the rabies virus genomic RNA has to

dissociate in order to become a template for the polymerase. However, simultaneously, the RNA must remain close to N, because the polymerase stays attached to N during its activity (4, 5). Although each protomer spans a distance of 22 Å between nucleotide ends (fig. S2), the nine nucleotides could be stretched out to a length of ~49 Å. Thus, a local dissociation of the RNA from one or several N protomers could provide sufficient space for binding of the polymerase complex (5) (fig. S4).

Phosphoprotein P links the polymerase complex to N and binding requires that S389 be

phosphorylated within a highly mobile region on the CTD of N (Fig. 1C and fig. S1) (21). This flexible region is disordered in the structure (Fig. 1C), but may become ordered upon P binding (22), as has been observed for the measles virus N-P interaction (23). P binding in the region of S389 (Fig. 3, A and B) may thus affect both subdomains including the hinge regions and could transfer a signal to the NTD, which would lead to the vertical opening of the NTD-CTD clamp, facilitating RNA exposure. The absence of any protein-protein contacts between NTDs (Fig. 1B) would facilitate a large NTD motion required for this process. In contrast, the CTD conformation is locked in by the large CTD-CTD interaction surface and by the tethering of the hinge loops. The latter interactions will thus maintain the contacts between the N protomers during the passage of the polymerase and the displacement of the RNA. Maintaining contact between N and RNA is required for rebinding of RNA and thus prevents the formation of double-stranded RNA during transcription and replication, which would otherwise require a helicase activity that is not encoded in the viral genome. Furthermore, an extensive production of double-stranded RNA would induce toll-like receptor 3 (TLR3) recognition and trigger immune reactions (16).

Even though N spontaneously associates with RNA in vivo in the absence of other viral proteins, a number of studies have shown that the 5' region of the viral genomic sequence regulates RNA packaging into a functional nucleocapsid (11, 24). Because the structure does not reveal any sequence specificity in RNA recognition, an alternative RNA-binding mechanism must exist. In the first protomer of the assembled nucleocapsid, the side extending the NTD hinge loop will be exposed and will not participate in polymerization (Fig. 1C); consequently, this region may provide a surface for sequence-specific RNA recognition. Conversely, at the 3' end of the packaged genome, the CTD hinge loop (Fig. 1C) may adopt a different conformation that may facilitate access of the polymerase to the 3' end,

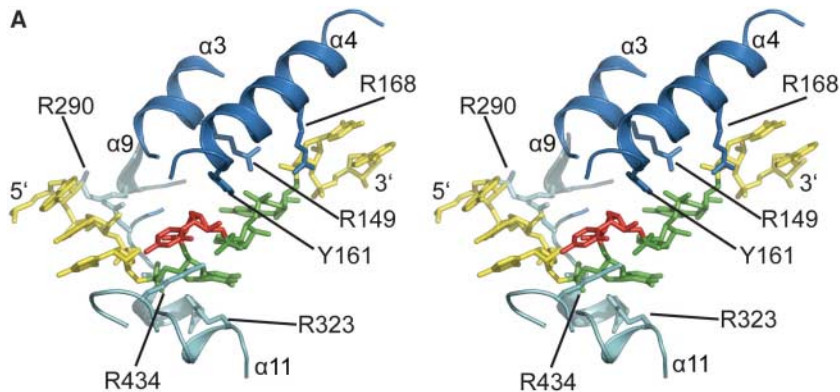
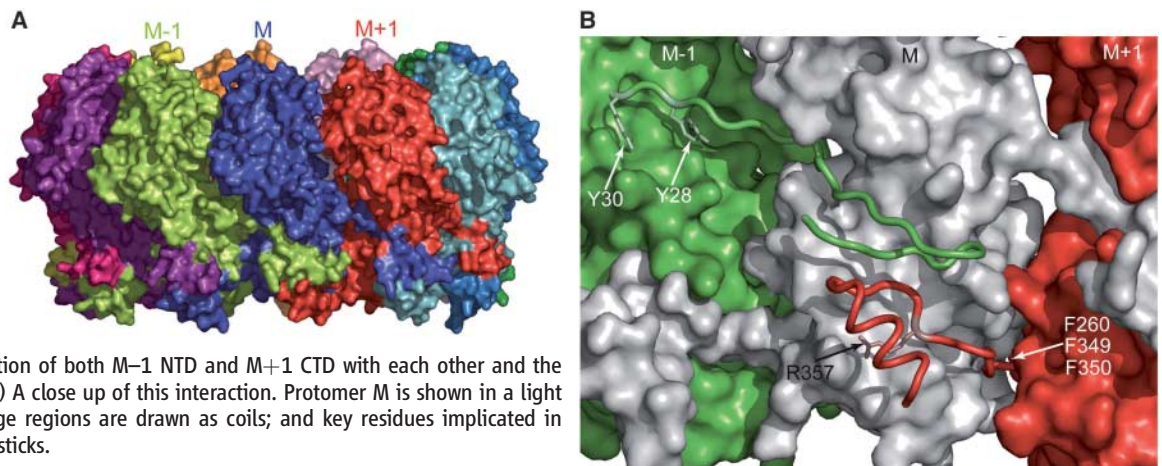


Fig. 2. Nucleoprotein-RNA interaction. (A) Stereo view of a close up of the interaction between N and the RNA. The RNA is shown as stick model and the protein moiety as a coil; the NTD is depicted in dark blue and the CTD in light blue. The nucleotides with bases facing the solvent are drawn in yellow and the bases facing the protein in green. The central nucleotide at the tip of the left-hand helical RNA segment is shown in red. Residues contacting the phosphates and ribose moieties are shown as sticks. **(B)** Schematic drawing highlighting the protein-RNA interactions with the phosphates and the ribose moieties. Color coding and orientations of the bases are the same as in (A). Most interactions are with the ribose-phosphate backbone.

Fig. 3. Two hinge regions stabilize polymerization of N. (A) Space-filling model of the N-RNA ring as viewed from the side. Each protomer is shown in a different color. This indicates that the NTD hinge from protomer M reaches over to M+1 and its CTD hinge reaches over to M-1. This arrangement leads to the interaction of both M-1 NTD and M+1 CTD with each other and the surface of protomer M. **(B)** A close up of this interaction. Protomer M is shown in a light gray for clarity; both hinge regions are drawn as coils; and key residues implicated in interaction are shown as sticks.



which is the start site for replication and transcription (25).

Electron micrographs of nucleocapsids from measles virus (*Paramyxoviridae*) and Marburg virus (*Filoviridae*), as well as an RNA-free crystal structure from the Borna virus (*Bornaviridae*) nucleoprotein, suggest that these nucleoproteins also adopt a two-domain structure (26, 27). This suggests that these enveloped viruses use an RNA-sequestering mechanism similar to that observed for the rabies virus N-RNA complex. The N-RNA polymer has thus evolved as the ideal template for the polymerase activity, which exposes the genomic RNA only temporarily to the host cell defense systems during replication. The tight sequestering of RNA observed in the crystal structure suggests further that the closed N-RNA conformation might be stabilized or frozen by small molecules, which could thus act as antiviral agents preventing rabies virus replication.

Note added in proof: This version of the manuscript is slightly changed relative to the version that was published *Science Express* on June 15. Improved refinement of the structure led to a better definition of the following regions, which have been changed accordingly: the N terminus, including the β hairpin leading to helix 1; Tyr¹⁶¹, Arg¹⁶⁸, and Arg⁴³⁴ contacting the RNA; as well as helices 13 to 16 of the C-terminal domain. None of

the changes made influences the overall conformation of the structure or any of the conclusions drawn from the structure.

References and Notes

1. B. Dietzschold, M. Schnell, H. Koprowski, *Curr. Top. Microbiol. Immunol.* **292**, 45 (2005).
2. D. L. Knobel *et al.*, *Bull. World Health Organ.* **83**, 360 (2005).
3. R. A. Lamb, D. Kolakofsky, in *Fields Virology*, D. M. Knipe and P. M. Howley, Eds. (Lippincott, Williams and Wilkins, Philadelphia, ed. 4, 2001), pp. 1305–1340.
4. H. Arnheiter, N. L. Davis, G. Wertz, M. Schubert, R. A. Lazzarini, *Cell* **41**, 259 (1985).
5. D. Kolakofsky, P. Le Mercier, F. Iseni, D. Garcin, *Virology* **318**, 463 (2004).
6. O. Poch, I. Sauvaget, M. Delarue, N. Tordo, *EMBO J.* **8**, 3867 (1989).
7. S. U. Emerson, M. Schubert, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5655 (1987).
8. D. Kolakofsky, *J. Virol.* **41**, 566 (1982).
9. D. Spohner, R. Drillien, P. M. Howley, *Virology* **232**, 260 (1997).
10. F. Iseni, A. Barge, F. Baudin, D. Blondel, R. W. Ruigrok, *J. Gen. Virol.* **79**, 2909 (1998).
11. B. M. Blumberg, C. Giorgi, D. Kolakofsky, *Cell* **32**, 559 (1983).
12. G. Schoehn, F. Iseni, M. Mavrikis, D. Blondel, R. W. Ruigrok, *J. Virol.* **75**, 490 (2001).
13. Supporting data are available on *Science Online*.
14. D. Thomas *et al.*, *J. Virol.* **54**, 598 (1985).
15. S. Lynch, D. Kolakofsky, *J. Virol.* **28**, 584 (1978).
16. S. Akira, S. Uematsu, O. Takeuchi, *Cell* **124**, 783 (2006).
17. I. Le Blanc *et al.*, *Nat. Cell Biol.* **7**, 653 (2005).
18. L. Espert *et al.*, *J. Biol. Chem.* **278**, 16151 (2003).
19. F. Iseni, F. Baudin, D. Blondel, R. W. Ruigrok, *RNA* **6**, 270 (2000).
20. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
21. B. Dietzschold *et al.*, *Virus Res.* **8**, 103 (1987).
22. H. Toriumi, A. Kawai, *Microbiol. Immunol.* **49**, 757 (2005).
23. S. Longhi *et al.*, *J. Biol. Chem.* **278**, 18638 (2003).
24. A. Kouznetsov, M. Buckle, N. Tordo, *J. Gen. Virol.* **79**, 1005 (1998).
25. S. U. Emerson, *Cell* **31**, 635 (1982).
26. G. Schoehn *et al.*, *J. Mol. Biol.* **339**, 301 (2004).
27. M. G. Rudolph *et al.*, *Structure* **11**, 1219 (2003).
28. We thank D. Kolakofsky (Geneva), Y. Gaudin, and D. Blondel (Gif-sur-Yvette), for many discussions and ideas; F. Iseni and M. Mavrikis for early contributions to the project; S. Cusack for discussions and comments on the text; and all members of the EMBL European Synchrotron Radiation Facility (ESRF) Joint Structural Biology Group (JSBG) for access to the ESRF beam lines. This work was supported by the EMBL (W.W.), the Deutsche Forschungsgemeinschaft SFB 593 (W.W.), the Université Joseph Fourier (R.W.H.R.), the CNRS (G.S.), an EMBO fellowship (C.R.C.), and a Ph.D. fellowship from the French Ministry for Education, Research, and Technology (A.A.V.A.). Coordinates have been deposited in the Protein Data Bank (accession code no. 2GTT).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1125280/DC1

Materials and Methods

Figs. S1 to S4

Table S1

References

23 January 2006; accepted 17 May 2006

Published online 15 June 2006;

10.1126/science.1125280

Include this information when citing this paper.

Characterization of the piRNA Complex from Rat Testes

Nelson C. Lau,^{1*} Anita G. Seto,^{1*} Jinkuk Kim,^{2,3} Satomi Kuramochi-Miyagawa,⁴ Toru Nakano,⁴ David P. Bartel,^{3,5} Robert E. Kingston^{1†}

Small noncoding RNAs regulate processes essential for cell growth and development, including mRNA degradation, translational repression, and transcriptional gene silencing (TGS). During a search for candidate mammalian factors for TGS, we purified a complex that contains small RNAs and Piwi, the rat homolog to human Piwi. The RNAs, frequently 29 to 30 nucleotides in length, are called Piwi-interacting RNAs (piRNAs), 94% of which map to 100 defined (≤ 101 kb) genomic regions. Within these regions, the piRNAs generally distribute across only one genomic strand or distribute on two strands but in a divergent, nonoverlapping manner. Preparations of piRNA complex (piRC) contain rRecQ1, which is homologous to *qde-3* from *Neurospora*, a gene implicated in silencing pathways. Piwi has been genetically linked to TGS in flies, and slicer activity cofractionates with the purified complex. These results are consistent with a gene-silencing role for piRC in mammals.

Gene-silencing pathways guided by small RNAs, essential for maintaining proper cell growth and differentiation, operate at either the transcriptional or posttranscriptional level (1). Posttranscriptional gene silencing acts through mRNA destabilization or inhibition of mRNA translation (1), whereas TGS represses gene expression by altering chromatin conformation (2). Each pathway uses a core complex containing small RNA associated with a member of the Argonaute (Ago) protein family; however, the different mechanistic needs of each pathway

require differences in complex composition. Although RNA-mediated TGS has been studied in fission yeast and other eukaryotes (2–5), the mechanism of this process in mammals remains elusive.

To identify candidate complexes for TGS in mammals, we exploited the previous observations that TGS might use small RNAs longer than the 21- to 23-nucleotide (nt) microRNAs (miRNAs). In *Arabidopsis*, *Tetrahymena*, *Drosophila*, and zebrafish, RNAs that are 24 nt and longer have been associated with TGS and/or genomic repeats, which are often silenced (6–11).

In *Drosophila*, these repeat-associated small interfering RNAs (rasiRNAs) are enriched in the testis (6, 12). Therefore, we prepared extract from rat testes and fractionated it on an ion-exchange Q column, monitoring the small RNAs. A peak of small RNAs longer than a 22-nt marker eluted in mild salt conditions, which suggested the presence of a novel ribonucleoprotein complex (Fig. 1A).

To characterize the small RNAs, we sequenced cDNA libraries made from flowthrough and eluate fractions, obtaining 61,581 reads from the eluate that matched perfectly to the *Rattus norvegicus* genome (13). In contrast to the flowthrough RNAs, which were mostly miRNAs (69%), the eluate RNAs derived primarily from regions of the genome not previously thought to be expressed (Fig. 1A). Some eluate reads matched expressed sequence tags (EST) (11%),

¹Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA.

²Harvard-MIT Division of Health Sciences and Technology, E18-435, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. ³Howard Hughes Medical Institute and Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. ⁴Department of Molecular Cell Biology, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamada-oka, Suita-shi, Osaka 565-0871, Japan. ⁵Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: kingston@molbio.mgh.harvard.edu

Revised 21 July 2006: see p. 6 for details.



www.sciencemag.org/cgi/content/full/1125280/DC1

Supporting Online Material for

Crystal Structure of the Rabies Virus Nucleoprotein-RNA Complex

Aurélie A. V. Albertini, Amy K. Wernimont, Tadeusz Muziol, Raimond B. G. Ravelli,
Cedric R. Clapier, Guy Schoehn, Winfried Weissenhorn,* Rob W. H. Ruigrok

*To whom correspondence should be addressed. E-mail weissen@embl-grenoble.fr

Published 16 June 2006 on *Science Express*

DOI: 10.1126/science.1125280

This PDF file includes

Materials and Methods
Figs. S1 to S4
Table S1
References

Materials and Methods

Protein production and crystallization. Expression and purification of the nucleoprotein was performed as described (1). The mixture of the gradient purified N-RNA rings was further separated by preparative electrophoresis, using a BioRad Model 491 Prep Cell. The purity and homogeneity was confirmed by negative staining electron microscopy. The undecameric ring was crystallized using the hanging drop method by combining 1 μ l of protein solution at 5 mg/ml in 20 mM Tris pH 7.5 with 1 μ l of reservoir solution containing 60% TACSIMATE (Hampton) and 10 mM L-cysteine. SDS-PAGE analysis of the crystal confirmed that the full-length N protein was crystallized. The crystals were flash cooled in a stream of liquid nitrogen for cryo-protection by harvesting directly from the drop.

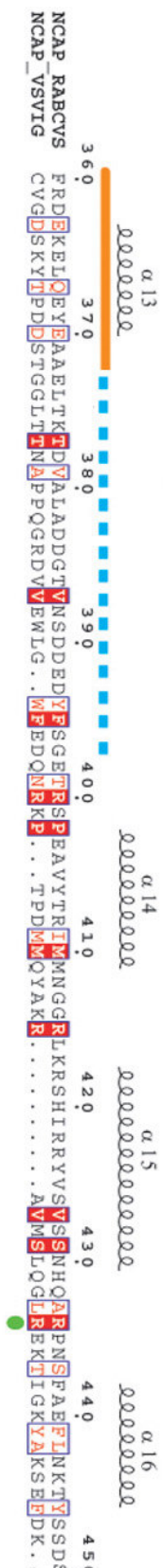
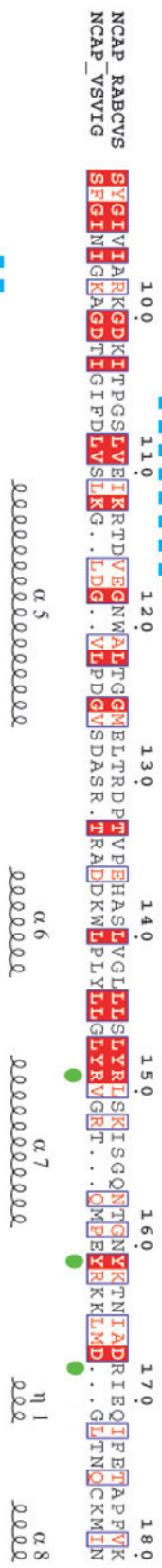
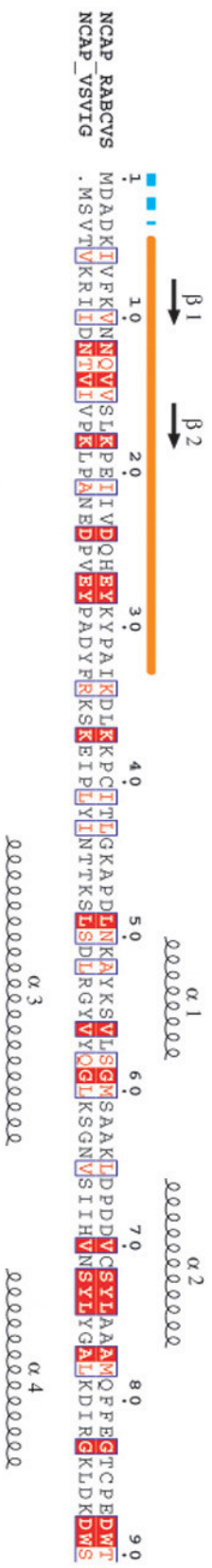
Data collection, structure determination and refinement. The N-RNA complex crystallizes in space group $P2_12_12$ with unit cell parameters of $a=270$, $b=281$, $c=237$ Å and two undecameric rings per asymmetric unit. Two native datasets extending to 3.9 Å resolution (nat1) and to 3.5 Å resolution (nat2) were collected at the ESRF (Grenoble) beamline ID14-4. A heavy ion derivative was obtained by soaking a crystal during 9 days in the harvesting buffer containing 0.1 mM $AuCl_2$ and one dataset was collected at the Au L-III edge, $\lambda=1.0372$ Å, to a resolution of 6 Å at the ID14-4 beamline. The data were processed using the program XDS (2). The positions of 22 Au sites forming two identical rings of 11 atoms were found using the program SHELXD using data limited to 10 Å (3). SHARP (4) was used to refine the sites and calculate phases to 6 Å with a figure of merit of 0.26.

The NCS operators were refined with DM while extending the phases to a native dataset of 4 Å (nat1 data set) (5). The resulting 4 Å, NCS averaged electron density map was continuous, and a polyaniline chain was traced for ~ 80 % of the protomer. This initial model showed a clear two-domain organization and the presence of a potential RNA chain spliced in between

both domains. The resolution of the crystals was further improved to 3.5 Å by dehydrating the crystals via air exposure for several minutes prior to flash cooling in a stream of liquid nitrogen. This second native dataset, nat2, was combined with the 4 Å phases and phase extension and 44-fold NCS was imposed to more precisely determine the positions of the two domains of each of the 22 protomers. This improved the figure of merit to 0.756 and the resulting electron density maps were of sufficient quality to finalize model building using the programs COOT (6) and O (7). The N-terminal region containing residues 35 to 78 revealed only few sequence markers. The orientation of helices 1 and 2 was determined using the program ArpWarp (8). CNS was used for positional (torsion angle simulated annealing and minimization) and group B-factor refinement (9). A final round included TLS refinement with REFMAC5 (5, 10). Phase restraints on the NCS averaged map were imposed throughout the refinement. The structure was refined to an R_{factor} of 27.5% and R_{free} of 32.6% (2260 reflections in resolution shells) and good stereochemistry (table S1). 98.1 % of the residues are in allowed regions of a Ramachandran plot (5). Figures were generated with programs ESPript (11) and PyMOL (<http://www.pymol.org>). Sequences were aligned using ClustalX (12). Secondary structure elements were assigned using the program DSSP (13).

Supplementary figures

Figure S1. Sequence conservation of Rhabdovirus nucleoproteins. The sequence similarity between two representative sequences of a rabies virus N (accession # Q7TBN9) and a vesicular stomatitis virus (VSV) N protein (accession # P11212) is 36 %. The secondary structure elements are shown above the sequence. Disordered regions, hinge regions, and residues implicated in RNA coordination are indicated.



- : residues implicated in RNA interaction
- : disordered regions
- ▬ : domain swapping

S2

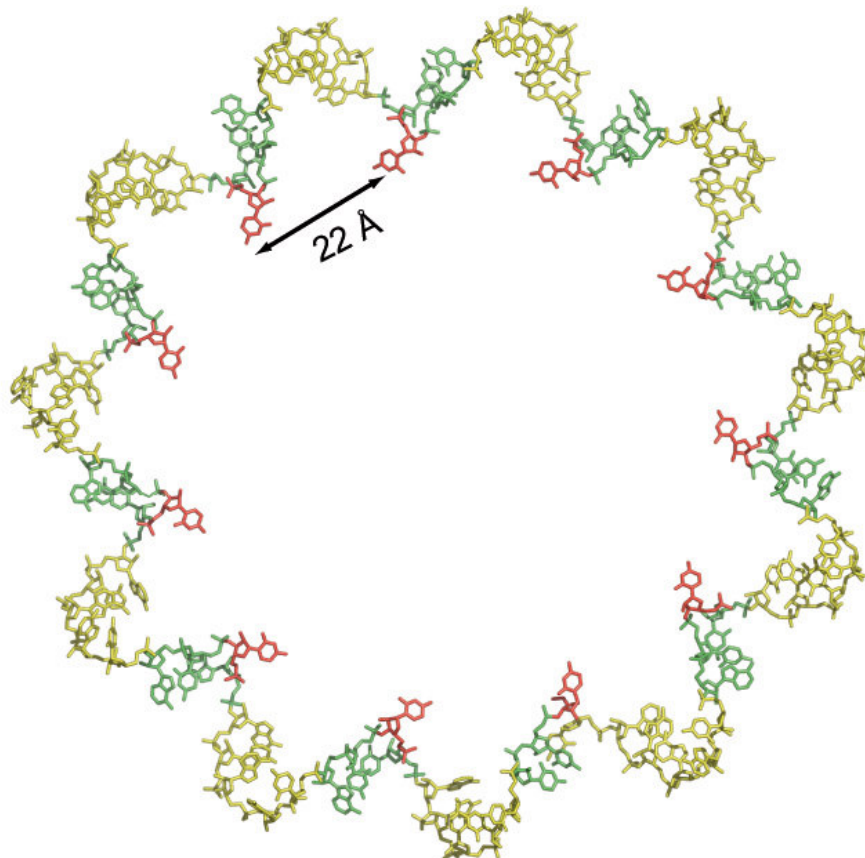


Figure S2. A; RNA conformation alone within the ring structure as an all atom model. The RNA strands are twisted clockwise in a left handed helix. This allows a maximum RNA to protein binding ratio and may thus also accommodate sufficient space for the polymerase complex during transcription and replication when the RNA is dissociated from N.

S3

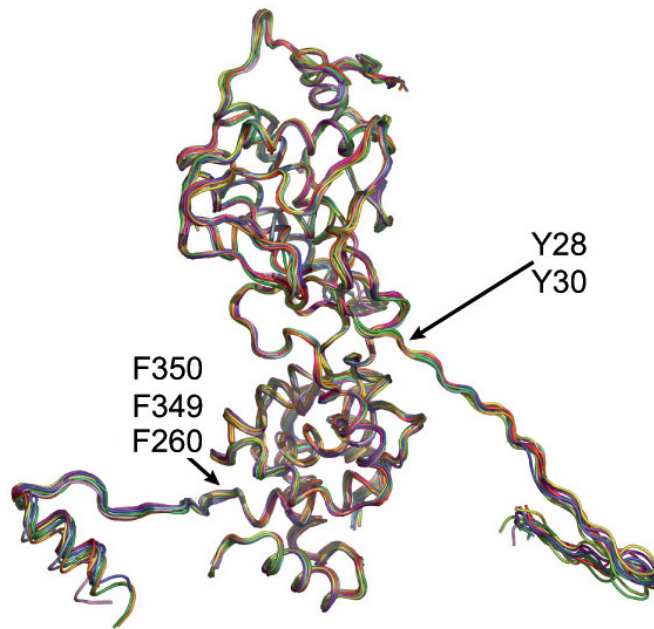


Figure S3. C alpha overlay of 11 protomers reveals an r.m.s.d of 1.4 Å. Notably the most prominent differences are within the NTD and the CTD hinge regions together with a poorly ordered region in the NTD (residues 103 to 118).

S4

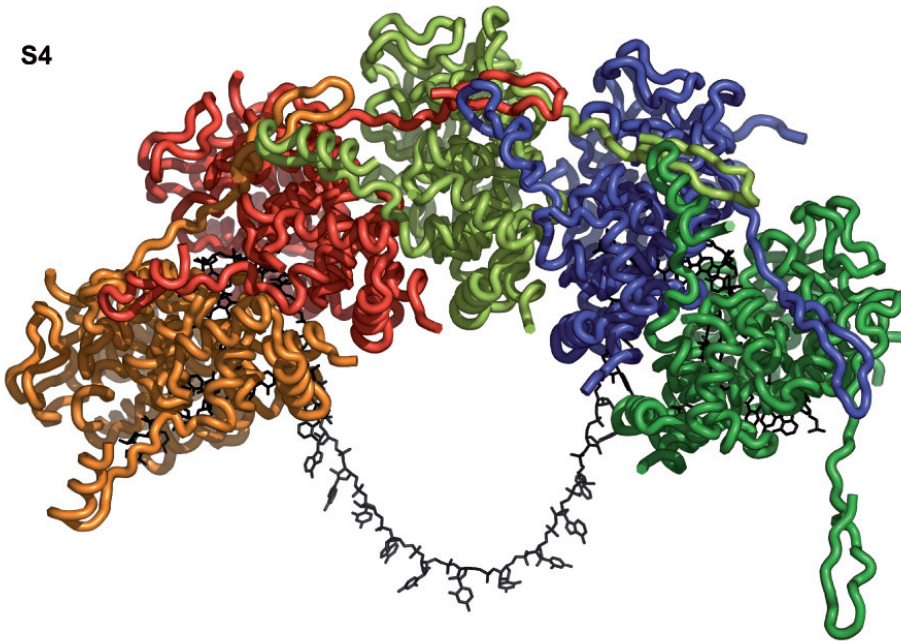


Figure S4. Hypothetical model of an 'open' N-RNA conformation showing RNA detachment from 3 N protomers. No potential conformational changes of N are taken into account. This provides enough space for the polymerase that is likely to surround the coding RNA strand.

Revision (21 July 2006): The current version has been revised to clarify procedures, to add a reference, and to revise refinement data to reflect changes made for the print publication.

Table S1 Data collection and refinement statistics

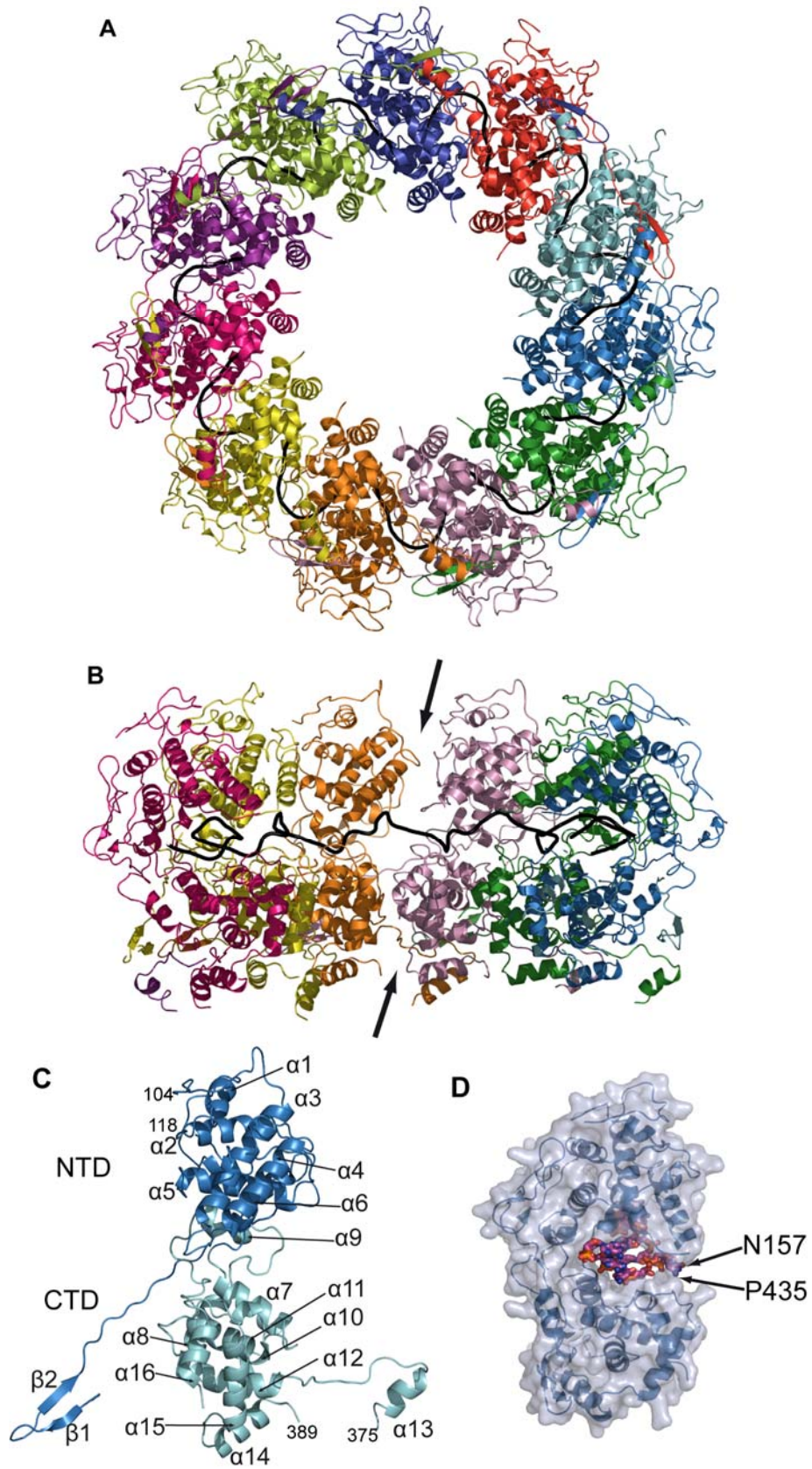
	nat1	nat2	Au
Data collection			
Space group	P 21 21 2	P 21 21 2	P 21 21 2
Cell dimensions			
a (Å)	267.58	270.43	269.44
b	283.86	281.00	284.20
c	235.18	236.90	236.62
Wavelength (Å)	0.9393	0.9794	1.0372
Resolution (Å)	50.0-3.9(4.04-3.9)	100-3.5 (3.63-3.5)	50.0-6.0(6.21-6.0)
R_{sym} (%)	16.1 (79.8)	9.6 (45.7)	6.1 (48.7)
$I / \sigma I$	9.44 (2.53)	12.15 (3.36)	10.8 (2.22)
Completeness (%)	99.2 (99.1)	99.9 (99.9)	97.9(97.9)
Refinement			
Resolution (Å)		25.0-3.5	
No. reflections		223,764	
$R_{\text{work}} / R_{\text{free}}$		27.5/32.6	
No. atoms		74,596	
Protein		70,492	
RNA		4104	
R.m.s deviations			
Bond lengths (Å)		0.019	
Bond angles (°)		1.75	

*Highest resolution shell is shown in parenthesis.

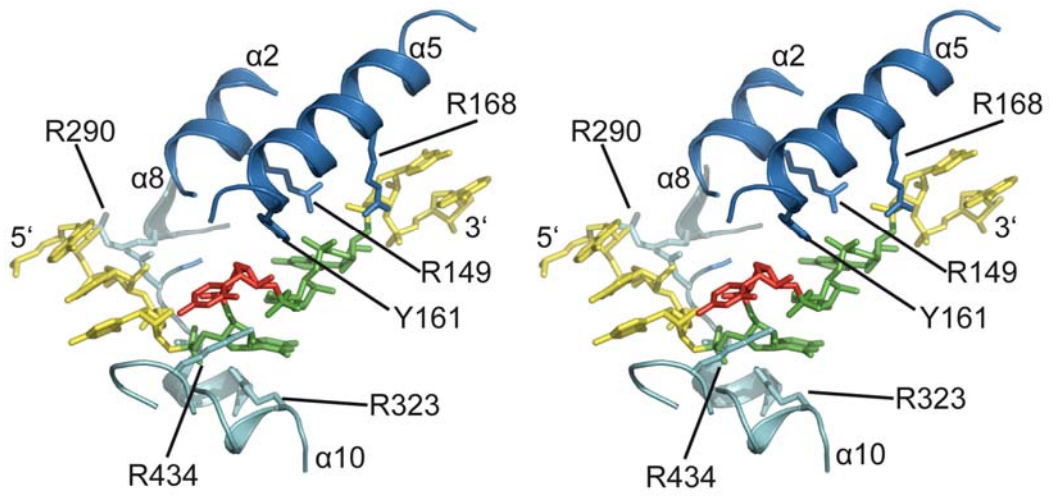
References

1. G. Schoehn, F. Iseni, M. Mavrakis, D. Blondel, R. W. Ruigrok, *J Virol* **75**, 490-8 (2001).
2. W. Kabsch, *J Applied Crystallogr* **21**, 916-924 (1988).
3. T. R. Schneider, G. M. Sheldrick, *Acta Crystallogr D* **58**, 1772-9 (2002).
4. E. d. La Fortelle, G. Bricogne, *SHARP: A maximum-likelihood heavy-atom parameter refinement program for the MIR and MAD Methods*. C. Carter, R. M. Sweet, Eds., Meth. Enzym. (Academic Press, Orlando, FL, USA, 1997), vol. 276.
5. CCP4, *Acta Crystallogr. D* **50**, 157-163 (1994).
6. P. Emsley, K. Cowtan, *Acta Crystallogr D* **60**, 2126-2132 (2004).
7. T. Jones, J. Zou, S. Cowan, M. Kjeldgaard, *Acta Crystallogr A*. **47**, 110-9 (1991).
8. R.J. Morris et al., *J. Synchrotron Radiat.* **11**, 56-59 (2004).
9. A. T. Brunger *et al.*, *Acta Crystallogr D* **54**, 905-921 (1998).
10. G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr D* **53**, 240-55 (1997).
11. P. Gouet, E. Courcelle, D. Stuart, F. Metoz, *Bioinformatics* **15**, 305-308 (1999).
12. J. Thompson, T. Gibson, F. Plewniak, F. Jeanmougin, D. Higgins, *Nucleic Acids Res* **25**, 4876-4882 (1997).
13. W. Kabsch, C. Sander, *Biopolymers* **22**, 2577-2637 (1983).

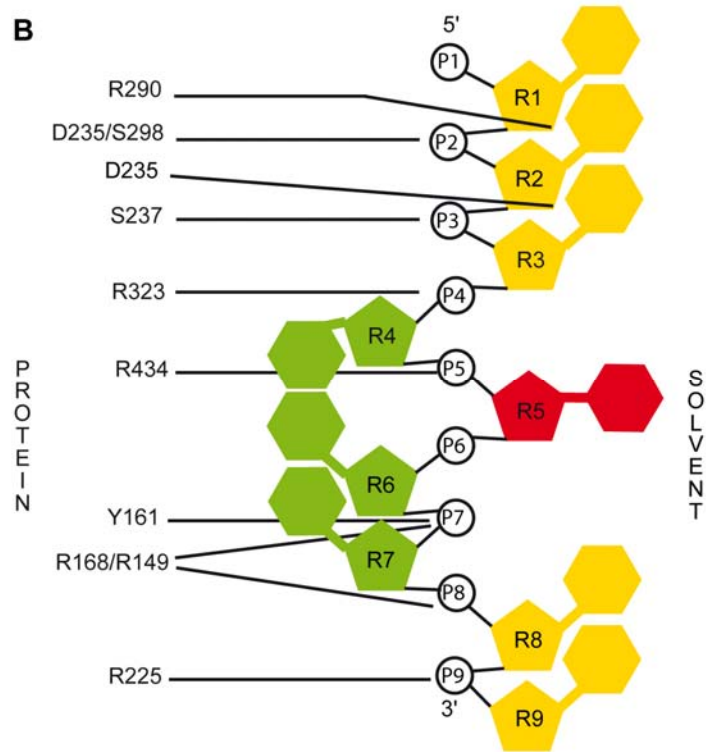
III. Figures de la publication (Albertini et al., 2006) à meilleure résolution

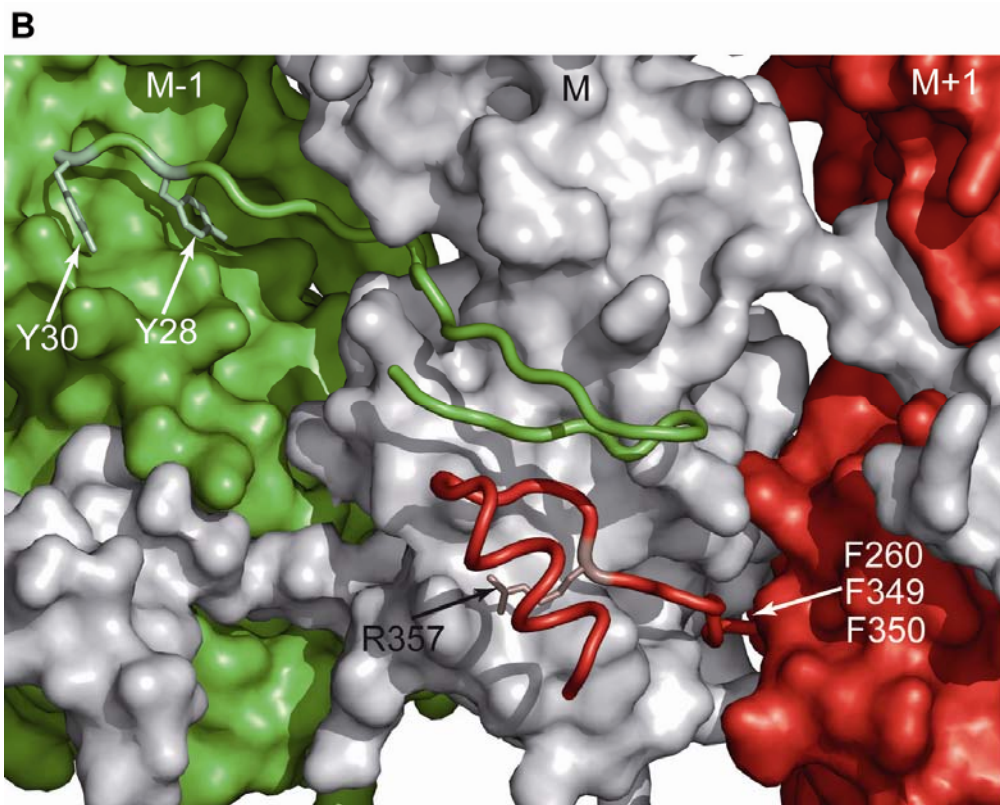
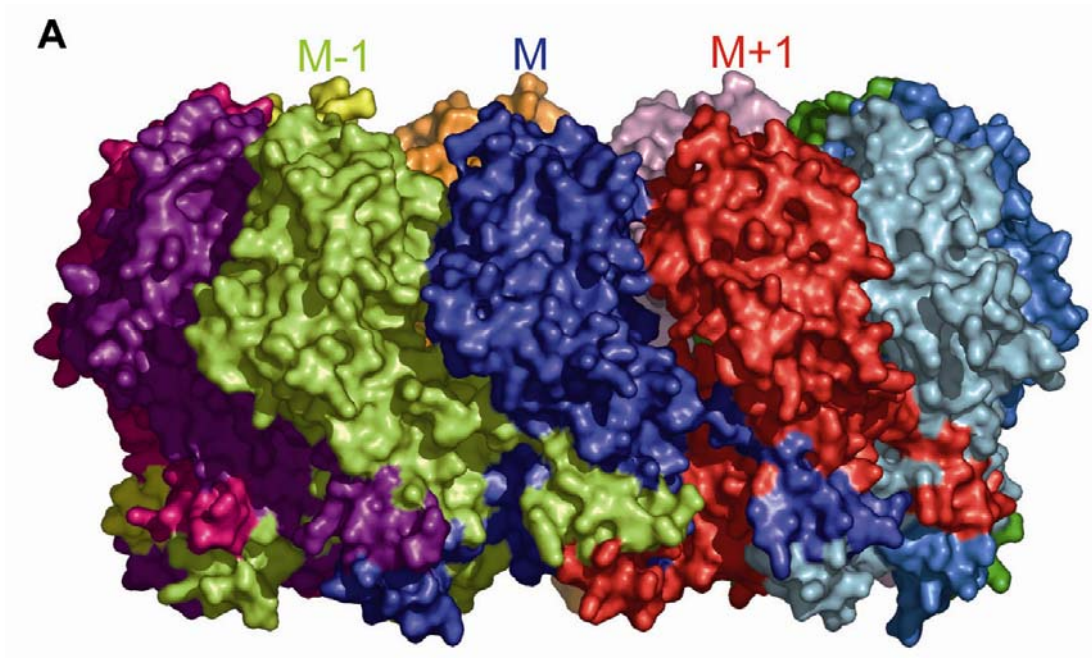


A



B





IV. Précisions sur la méthode de résolution de la structure en complément à l'article II

Les trois jeux de données utilisés pour la résolution de la structure ont été collectés à l'ESRF (Grenoble) sur la ligne ID14eh4. Les cristaux de N_{11} ont été directement pêchés, déshydratés puis congelés dans le flux d'azote liquide (100°K). Le faisceau est très énergétique avec une longueur d'onde variable 9.6 KeV à 14.5 KeV. Cette station peut aussi bien être utilisée pour la collecte de jeux de données natives que pour la collecte de jeux de données à haute énergie pour les expériences de MAD ou de SAD comme dans notre cas.

A. Détermination du groupe d'espace et collecte des jeux de données

Le premier jeu de données (Nat1) à haute résolution collecté sur les cristaux natifs montrait des taches de diffraction à une résolution maximale de 3.5 Å. Après une collecte complète et l'analyse des statistiques, la résolution globale s'est avérée être 4 Å à cause des dommages liés aux radiations. Le groupe d'espace a été déterminé avec le programme MOSFLM (Leslie, 2006) comme étant orthorhombique ($P222$ ou $P2_12_12$ ou $P222_1$ ou $P2_12_12_1$). La solution $P2_12_12_1$ a été choisie car l'extinction systématique des réflexions suivant les trois axes h,k et l permettait de croire en l'existence de trois axes hélicoïdaux dans la maille cristalline. Ultérieurement, les analyses de la fonction de self-rotation et des cartes de Patterson ont permis de se rendre compte que le groupe d'espace du cristal était en réalité $P2_12_12$. L'unité asymétrique est composée de deux anneaux face à face générant un axe pseudo-symétrique selon l'axe l.

Deux programmes peuvent être utilisés pour l'indexation et l'intégration des données : XDS (Kabsch, 1993) et MOSFLM. Le programme XDS a été retenu car les statistiques obtenues après intégration étaient de meilleure qualité. L'indexation se fait image par image tout en affinant la distance du cristal au détecteur, la position du centre du détecteur, l'orientation du cristal, et les paramètres de maille. Lors de l'intégration, l'intensité des taches de diffraction est mesurée en appliquant un «masque» autour de la tache. L'intérieur du masque délimite la surface à intégrer et l'extérieur permet d'estimer le bruit de fond à retrancher aux intensités intégrées.

Le programme Xscale (Kabsch, 1993) a ensuite été utilisé pour la mise à l'échelle et la réduction des données. Une mise à l'échelle des données est nécessaire car l'exposition aux rayons X et le pouvoir de diffraction du cristal sont différents sur chaque image. Lors de la réduction des données, le jeu de données est moyenné suivant les réflexions équivalentes

collectées plusieurs fois ou reliées entre elles par des opérations de symétries. La qualité du jeu de données est estimée grâce aux valeurs de R_{sym} entre les réflexions équivalentes reliées par symétrie et au rapport $I/\sigma I$ (signal/bruit).

Deux autres jeux de données ont été utilisés pour la résolution de la structure. Un jeu de données a été collecté sur un cristal marqué à l'or à une résolution de 6 Å (Au) à la longueur d'onde correspondant au pic d'absorption L III de l'or. Un troisième jeu de données natives complet à 3.5 Å de résolution (Nat2) a pu être obtenu en déshydratant les cristaux. Ces deux jeux de données supplémentaires ont été indexés et intégrés de la même façon que le premier. Les statistiques relatives à la collecte des données des trois jeux sont répertoriées dans le Tableau 4.

Tableau 4 : Statistiques relatives aux collectes de données.

Jeu de données	Nat 1	Nat 2	Au
Nombre d'images	240	240	600
Oscillation	0.5 °	0.5°	0.2°
Groupe d'espace	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Dimensions de la maille élémentaire:			
a (Å)	267.58	270.43	269.44
b (Å)	283.86	281.00	284.20
c (Å)	235.18	236.90	236.62
Longueur d'onde (Å)	0.9393	0.9794	1.0372
Limites de résolution (Å)	50.0-3.9 (4.04-3.9)	100-3.5 (3.63-3.5)	50.0-6.0 (6.21-6.0)
R_{sym} (%)	16.1 (79.8)	9.6 (45.7)	6.1 (48.7)
$I / \sigma I$	9.44 (2.53)	12.15 (3.36)	10.8 (2.22)
Complétude (%)	99.2 (99.1)	99.9 (99.9)	97.9 (97.9)

B. Résolution de la structure des anneaux N₁₁

1) Première stratégie pour l'obtention des phases : le remplacement moléculaire

La première tentative de résolution de la structure consistait à utiliser la carte de microscopie électronique obtenue à partir des cristaux broyés (Cf. partie III, article I) en utilisant le programme de remplacement moléculaire MOLREP (Vagin and Teplyakov, 2000).

Le remplacement moléculaire « direct » en utilisant la carte de microscopie et les données cristallographiques à 4 Å n'est pas possible car la carte de microscopie est une enveloppe vide. Afin de remédier à cela, un modèle protéique a été construit à l'intérieur même de la carte de microscopie. A cet effet, le fichier de coordonnées d'une protéine (penton de l'adénovirus humain) a été découpé et modifié pour remplir le plus précisément possible la

densité électronique de la carte issue de la microscopie électronique (Figure 31). Aucune solution de remplacement moléculaire n'a pu être établie.

Nous avons pensé que le remplacement moléculaire échouait car le modèle protéique ne contenait pas assez d'atomes. Pour remédier à cela le programme ARP/wARP (Morris et al., 2004) a été utilisé pour ajouter des molécules d'eau au fichier de coordonnées et ceci en vue de remplir le plus possible l'enveloppe de la carte de microscopie. Suite à cela, les statistiques du remplacement moléculaire ont commencé à converger vers une solution plausible ($R_{\text{fact}}=0.44$ et $CC=0.55$). Le programme RAVE (Kleywegt and Jones, 1994) a été utilisé pour calculer les phases et la carte de densité électronique. L'analyse des cartes de densité électronique a permis d'observer la position des canaux de solvant et des 22 protomères, sans toutefois mettre en évidence une continuité permettant de tracer une chaîne protéique (Figure 34 A). Les opérateurs de symétries non cristallographiques n'ont pas pu être déterminés avec une précision suffisante pour améliorer la qualité de la carte de densité électronique.

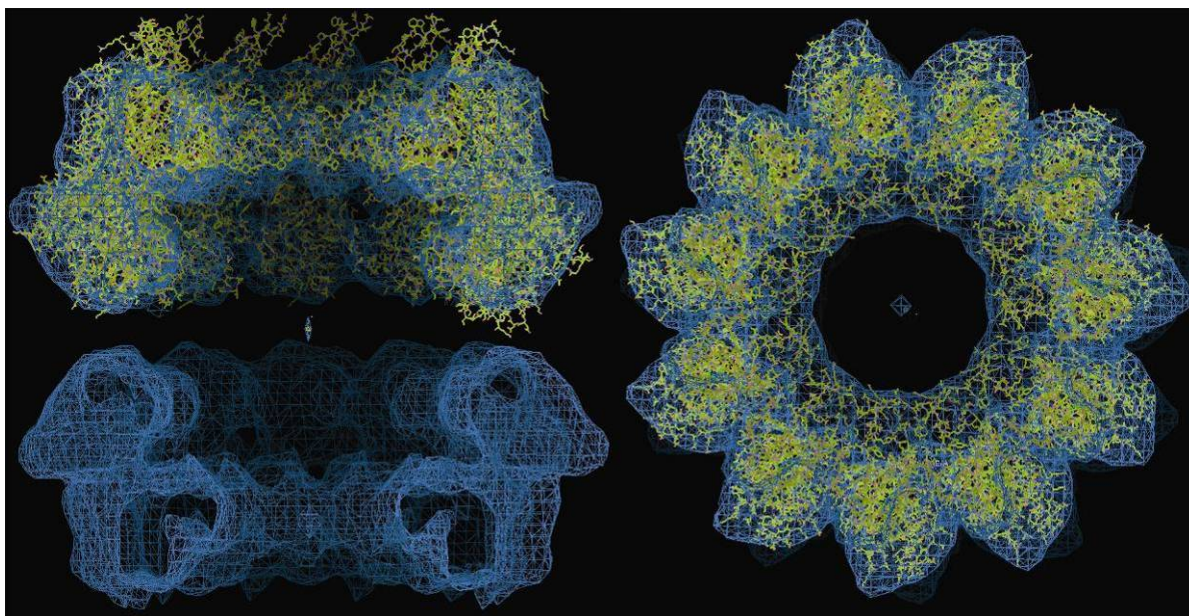


Figure 31 : « Faux pdb » recalé dans la densité de microscopie électronique.
Protéine utilisée : penton de l'adénovirus humain (code pdb : 1X9P)

2) Deuxième stratégie pour l'obtention des phases : Single Anomalous Diffraction

L'utilisation de la cystéine en tant qu'additif combinée à la quantité importante de sel dans la condition de cristallogénèse a rendu difficile les essais de marquage des cristaux avec des

dérivés d'atomes lourds (Petsko, 1985). La co-cristallisation avec différents atomes lourds (avec ou sans cystéine) s'est elle aussi révélée infructueuse. L'expression d'une protéine sélénométhionylée en système baculovirus/cellules d'insectes n'a pas fonctionné non plus, la sélénométhionine n'étant pas incorporée dans la nucléoprotéine lors de sa surexpression. Le marquage de la protéine par trempage demeurant la seule alternative, des tests en masse de différents dérivés lourds ont été effectués. Les essais de marquage les plus prometteurs sont répertoriés dans le tableau suivant.

Tableau 5 : Résumé des différents essais de marquage avec des dérivés d'atomes lourds.

Le site <http://www.sbg.bio.ic.ac.uk/had/heavyatom.html> a été consulté pour le choix de certains dérivés lourds.

Sel de métal testé	Résultat
Brome (NaBr, BrLi) Enrobe la protéine	Signal très faible
Uranium (acétate d'uranyle)	Détruit les cristaux
Mercure (EMTS, HgCl ₂ , CH ₃ HgCl) Interagit avec les cystéines et l'ARN	Pas de signal et non isomorphe (co-cristallisation) Signal anomal extrêmement faible (trempage)
Cadmium (CdCl ₂ , Cd(NO ₃) ₂)	Pas de signal (co-cristallisation et trempage)
Platine (PtCl ₂ , PtSO ₄ , cis-Pt)	Aucun signal
Or (AuCl ₂ , KAuCl ₄)	Présence d'un signal anomal

Après intégration et mise à l'échelle des intensités, la fixation du marqueur anomal testé est vérifiée avec le programme XPREP (Bruker AXS, Madison, WI). Ce programme calcule un rapport signal anomal / bruit. Si ce rapport est supérieur à 1, le signal anomal est significatif.

Dans le cas de l'or, l'énergie du faisceau de rayons X a été réglée au niveau du pic d'absorption LIII de l'or ($\lambda = 1.037$, Figure 32). Le marquage avec le dérivé d'or AuCl₂ s'est révélé positif (une seule fois). L'analyse du jeu de données Au (6Å de résolution) a permis de détecter la présence d'un pouvoir de phasage significatif de 1.44 jusqu'à 10 Å.

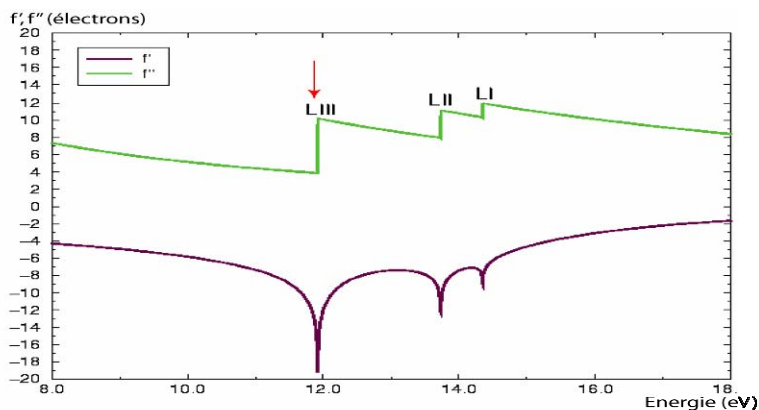


Figure 32 : Courbes d'intensité du signal anomal de l'or.

La flèche rouge indique à quelle énergie le jeu de données utilisé pour le SAD a été collecté.

Le programme ShelxD (Sheldrick, 1997) a permis de trouver 22 sites pour les atomes d'or. Ces 22 atomes lourds forment 2 anneaux parfaits de 11 atomes permettant d'affirmer que chaque protomère de nucléoprotéine fixe un atome d'or (Figure 33). L'ultrastructure du complexe étant un anneau, l'obtention d'un ou plusieurs cercles d'atomes était prévisible. Malgré les apparences, les deux cercles définis par les 22 atomes d'or ne sont pas superposables et ne peuvent être reliés par une simple opération de symétrie. Ceci nous a permis de confirmer que le groupe d'espace était bien $P2_12_12$ avec une pseudo-symétrie suivant l'axe 1 générée par la présence de deux anneaux protéiques au sein de l'unité asymétrique.

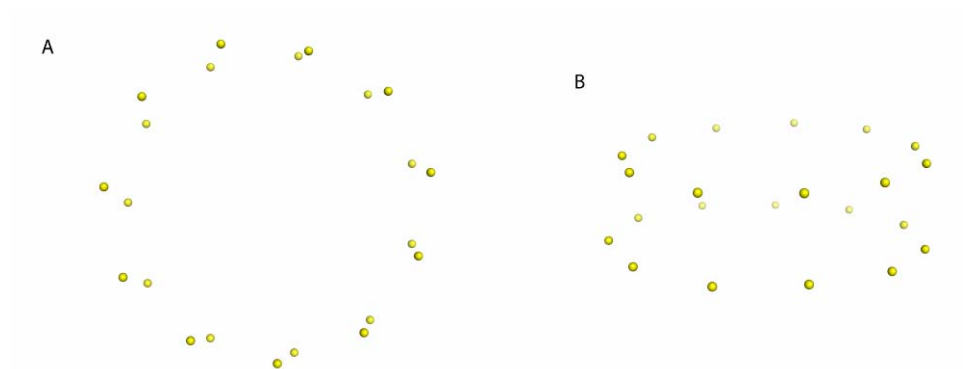


Figure 33 : Sites forts d'atomes d'or trouvés par ShelxD et affinés avec SHARP.

A : vue de dessus. B : vue de côté.

Le programme SHARP (Bricogne, 1997) a ensuite été utilisé pour affiner la position des sites d'atomes lourds et pour étendre les phases à 6 Å. Le programme DM a permis d'affiner les phases ainsi que les opérateurs de symétrie non cristallographique. La carte de densité électronique obtenue après ces étapes était très prometteuse et permettait d'observer des chaînes individuelles pour chaque protomère ainsi que de la densité électronique pour les hélices α . Ces phases, à 6 Å, ont ensuite été étendues en utilisant les données natives à 4 Å. La carte de densité électronique résultante était continue et a permis la modélisation d'une chaîne poly-alanine dans la densité électronique à l'aide du programme de construction COOT (Emsley and Cowtan, 2004) (Figure 34 B).

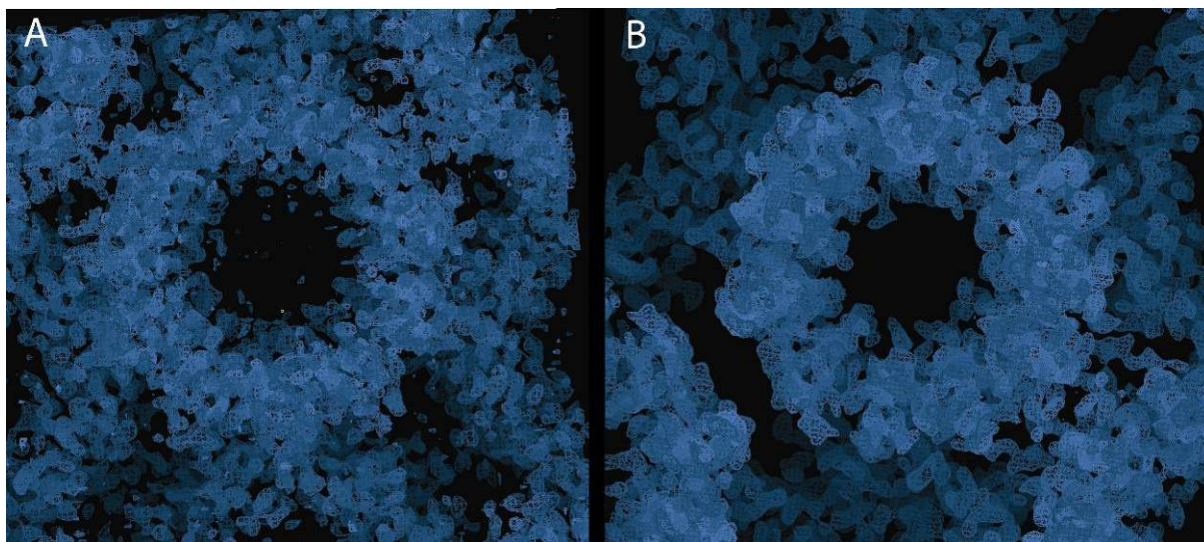


Figure 34 : Cartes de densité électronique expérimentales.

A : Carte obtenue suite au remplacement moléculaire avec le modèle issu de la microscopie électronique.

B : Carte obtenue après extension des phases à 3.5 Å.

C. Construction du modèle

La carte de densité à 4 Å de résolution était facilement interprétable. Au final 80 % d'un protomère a pu être construit manuellement en alanine. Les densités pour des gros acides aminés tels que des phénylalanines ou des tyrosines au niveau des cœurs hydrophobes visibles sur la carte nous ont donné des points de départ pour assigner la séquence polypeptidique.

Dans ce modèle initial, il était déjà clairement visible que le protomère était constitué de deux domaines distincts. Logé entre ces deux domaines, nous avons aussi détecté une densité électronique correspondant au squelette sucre-phosphate de l'ARN contenu dans les complexes.

L'obtention d'un troisième jeu de données natives à 3.5 Å de résolution a permis d'étendre les phases à 3.5 Å une nouvelle fois et d'améliorer la matrice de symétries non cristallographiques. De 22 opérateurs de symétrie, nous sommes passés à 44 opérateurs. L'utilisation de 44 opérateurs de symétrie est basée sur le fait que les deux domaines de la nucléoprotéine (mâchoires supérieure et inférieure) bougent légèrement latéralement et ceci de façon indépendante. La carte finale obtenue montrait bien plus de détails que la précédente et nous a permis de détecter d'une manière bien plus aisée d'autres éléments de la structure. Des feuillets β , des boucles ainsi qu'une densité plus évidente pour l'ARN ont pu être localisés sans trop d'ambiguïté. La densité de l'ARN entre les différents anneaux est moyennée et ceci ne permet pas d'assigner une séquence éventuelle à cet ARN. Les bases de

l'ARN placées dans la densité sont soit des adénines soit des cytosines afin de remplir au mieux la densité électronique (Figure 35).

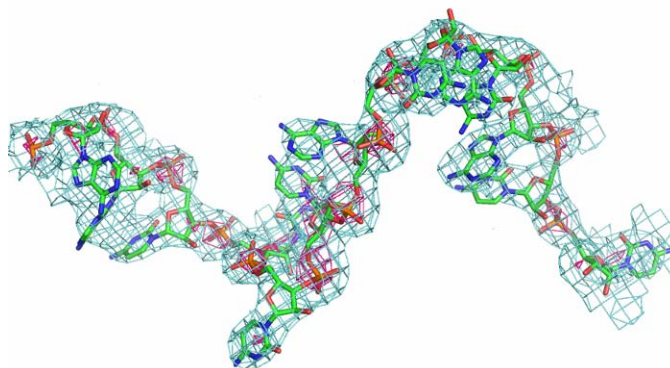


Figure 35 : Carte de densité électronique expérimentale autour de la molécule d'ARN.
La molécule d'ARN a pu être tracée en se basant sur la densité plus forte des atomes de Phosphate.
(Densité rouge : contour 3σ , densité bleue : contour 1σ).

A cette résolution, il a aussi été possible de détecter la présence de deux sous-domaines interchangeables («swapping domains») ce qui n'avait pas été possible dans la carte à 4 Å de résolution. Au final, nous avons donc pu placer 90 % des résidus dans la densité électronique et construire les deux brins d'ARN de 99 bases dans chaque anneau protéique. Deux régions de la protéine sont toujours désordonnées et ceci dans tous les protomères (résidus 104-116 et résidus 374-397).

V. Conclusion

L'obtention d'une structure de nucléoprotéine avec pour substrat l'ARN nous a permis de formuler plusieurs conclusions quant à son mode de fonctionnement.

L'unité asymétrique est composée de deux anneaux de 11 nucléoprotéines positionnés face à face (N-terminal contre N-terminal) et de deux molécules d'ARN simple brin de 99 bases. La forme générale d'un protomère de nucléoprotéine peut être comparée à celle d'une mâchoire qui bloque le brin d'ARN comme un étau. Cette structure nous a permis de confirmer que chaque nucléoprotéine interagit avec 9 bases d'ARN ainsi que cela avait été suggéré précédemment (Thomas et al., 1985). La présence de domaines interchangeables émergeant cotés N-terminal et C-terminal et liant les nucléoprotéines voisines contribue à la stabilité du complexe mais aussi à sa flexibilité. La flexibilité de ces domaines explique non seulement l'existence d'anneaux de différents diamètres mais également la possibilité de formation d'une nucléocapside hélicoïdale avec 53 sous-unités de nucléoprotéine par tour (Schoehn et al., 2001).

La présence de différences au niveau de la morphologie des nucléocapsides (Blumberg et al., 1984); (Heggeness et al., 1980) suggérait qu'il devait exister un genre de transition structurale pour «amener l'ARN à la surface» de la nucléocapside et le rendre accessible. Au vue de la structure cristallographique, il apparaît clairement que l'ARN est totalement séquestré au sein de la nucléoprotéine. Il s'agit donc dans ce cas d'une conformation « fermée » du complexe. L'existence d'une telle conformation de la nucléoprotéine est très importante. En effet, sous cette forme, l'ARN est alors protégé des RNases alors qu'habituellement, un ARN simple brin est automatiquement dégradé dans une cellule. L'ARN viral, enfermé dans la nucléoprotéine peut ainsi échapper à la réponse immunitaire innée car il n'est plus reconnu par les «Toll-Like Receptors» (Lopez et al., 2006).

Le mode de fixation de l'ARN à la nucléoprotéine est également inédit (Duarte and Pyle, 1998). L'ARN présente une structure «sous tension» et en quelque sorte condensée.

La nucléoprotéine constitue une cible antivirale intéressante. En bloquant la nucléoprotéine en conformation «fermée» à l'aide d'une drogue, l'ARN viral ne serait plus accessible pour la transcription, le virus serait alors incapable de se multiplier. Ce concept pourrait s'appliquer à d'autres virus à ARN négatif responsables eux aussi de pathologies humaines graves comme le virus Ebola et le virus de la rougeole.

**PARTIE V : Caractérisation biochimique des complexes
Nucléoprotéine-ARN-Phosphoprotéine**

I. Introduction

Lors du cycle de multiplication virale et à l'intérieur des particules virales, la phosphoprotéine associée à l'ARN-polymérase-ARN-dépendante (L) lie la matrice N-ARN et réalise ainsi un lien physique entre la protéine L et la nucléocapside virale. La phosphoprotéine interagit avec les complexes nucléoprotéine-ARN par l'intermédiaire de son extrémité C-terminale. Le mode d'interaction et la stœchiométrie de ce complexe sont pour l'instant inconnus. Nous nous sommes donc intéressés aux complexes formés entre la phosphoprotéine et les anneaux N-ARN. L'objectif à long terme de cette étude est d'obtenir une reconstruction en microscopie électronique ou une structure cristallographique d'anneaux N-ARN associés à la phosphoprotéine. Ceci permettrait de constater si les nucléoprotéines associées à l'ARN subissent des changements de conformation lorsqu'elles lient la phosphoprotéine.

Dans une première approche, nous avons observé l'interaction entre les complexes N-ARN et la phosphoprotéine par chromatographie d'exclusion de taille couplée à la diffusion statique de lumière à multi-angles (MALLS), puis par électrophorèse en gel natif.

Les expériences de MALLS ont été réalisées en collaboration avec le Professeur Marc Jamin et son étudiante en thèse Francine Gérard.

II. Expression et purification de la phosphoprotéine du virus de la rage

A. Expression de la phosphoprotéine du virus de la rage en cellules d'insecte

Les cellules HF de 6 flasques de 300 cm² sont infectées à une m.o.i. de 5 avec le baculovirus recombinant pour la phosphoprotéine P. Les cellules sont récoltées 3 jours *p.i.* et rincées avec du PBS 1x puis re-suspendues dans 40 ml de tampon de lyse (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 2 mM β -mercaptoéтанол, 1 mM EDTA, 1x Complete-EDTA free) et stockées à -80°C jusqu'à une utilisation ultérieure. La lyse des cellules est provoquée par congélation/décongélation en utilisant en alternance un bain à 37 °C et de l'azote liquide (-180 °C) (3 fois). Le lysat est centrifugé 15 minutes à 12000 g, puis le surnageant clair est aussitôt utilisé pour la purification.

B. Purification de la Phosphoprotéine P du virus de la rage

Le lysat est chargé sur une colonne échangeuse d'anions (DEAE-Sepharose, Amersham), la phosphoprotéine est éluée au cours d'un gradient de chlorure de sodium (120 mM NaCl, 20 mM Tris-HCl pH 7.5 – 375 mM NaCl, 20 mM Tris-HCl pH 7.5) à environ 150 mM NaCl.

Les fractions contenant de la protéine P identifiées par SDS-PAGE sont concentrées avec une unité de concentration Millipore™ (c/o : 10 KDa) puis chargées sur une colonne d'exclusion de taille (Superdex 200 Hi-load, HR16/60, Amersham). Les fractions contenant la phosphoprotéine pure sont identifiées sur SDS-PAGE puis conservées à 4°C pour une utilisation ultérieure.

III. Interaction entre les anneaux N-ARN (N₁₀) et la phosphoprotéine P

Les tests d'interactions ont été réalisés avec des anneaux N-ARN à 10 sous-unités de nucléoprotéine, purifiés selon la méthode décrite en partie III.

A. Etude de l'interaction entre les complexes N-ARN et la phosphoprotéine par MALLS

Le couplage de la chromatographie d'exclusion de taille à un système de détection par diffusion de lumière fournit une mesure absolue de la masse moléculaire (Wyatt, 1993).

La masse moléculaire théorique des anneaux N₁₀ est d'environ 533 kDa (10 protéines N + 90 nucléotides d'ARN). La mesure de la masse moléculaire de l'échantillon d'anneaux N₁₀ sur l'appareil de diffusion de lumière fournit une masse moléculaire de 520 kDa (+/- 1%). Cette mesure est donc aux erreurs de mesure expérimentales près, en accord avec la masse théorique prédite (Figure 36 A). D'après l'expérience de MALLS, la masse moléculaire de la phosphoprotéine du virus de la rage est de 68 kDa (+/- 2%). Cela correspond à un dimère de phosphoprotéine (P₂, 2 x 33 KDa) (Figure 36 B).

Une expérience de titrage a été réalisée afin de déterminer le nombre de dimères de phosphoprotéine pouvant fixer les anneaux N₁₀ (Figure 36 C). Une quantité croissante de phosphoprotéine a donc été incubée avec un échantillon d'anneaux N₁₀. L'addition de phosphoprotéine provoque un déplacement du pic de complexes N₁₀, indiquant que la phosphoprotéine se fixe sur les anneaux (Figure 36 C, D & E).

Deux espèces différentes de complexes ARN-N-P peuvent alors être visualisées : un complexe transitoire formé d'un anneau N₁₀ plus un dimère de phosphoprotéine (N₁₀-P₂, MM = 580 KDa) et un complexe formé par l'association d'un anneau N₁₀ et de deux dimères de P (N₁₀-2P₂, MM = 650 kDa). La masse moléculaire de ces complexes N₁₀-P₂ et N₁₀-2P₂ a été calculée grâce à la mesure de l'intensité de lumière diffusée et de l'indice de réfraction à différents temps d'élution. Il semblerait que le complexe N₁₀-2P₂ n'apparaisse que lorsque tous les anneaux N₁₀ ont préalablement fixé un dimère de phosphoprotéine. Les sites de fixation de dimères de phosphoprotéine sur les anneaux N₁₀ peuvent donc être saturés. Une

fois ce seuil dépassé, il est possible de visualiser sur les chromatogrammes l'apparition de dimères de phosphoprotéine non associés aux anneaux N_{10} .

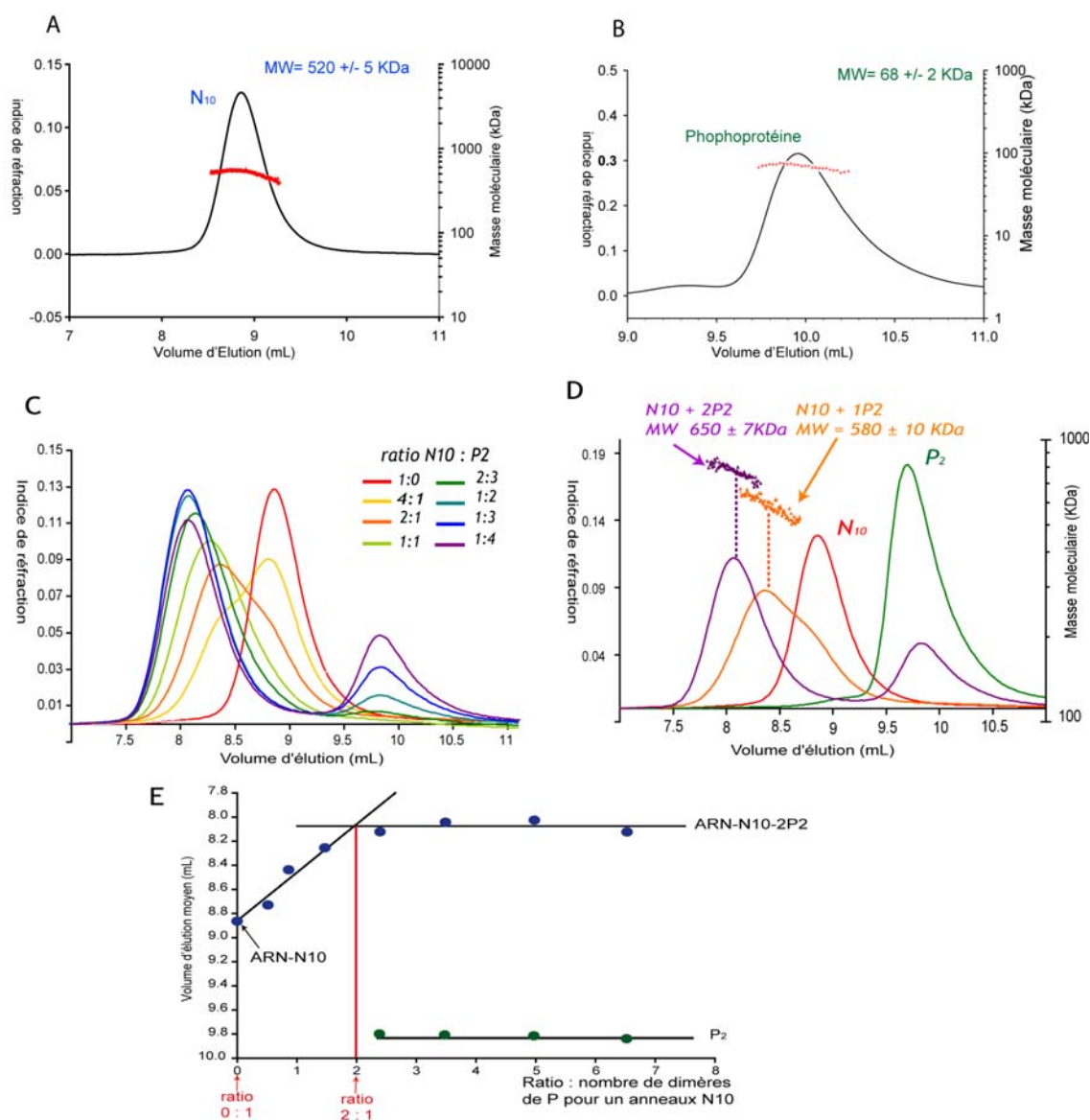


Figure 36 : Chromatographie d'exclusion de taille couplée à la diffusion statique de lumière et à la réfractométrie.

A : Profil d'élution des complexes N-ARN avec 10 sous-unités de nucléoprotéine N_{10} . La masse moléculaire de N_{10} est calculée à partir de mesures de l'intensité de lumière diffusée et de l'indice de réfraction à différents temps d'élution.

B : Profil d'élution de la phosphoprotéine de la rage. La masse moléculaire de P_2 est calculée à partir de l'intensité de lumière diffusée et de réfractométrie à différents temps d'élution.

C : Profil d'élution des complexes N_{10} avec différentes quantités de dimères de phosphoprotéine ajoutés.

D : Profil d'élution et calcul de la masse moléculaire à partir de l'intensité de lumière diffusée et de la réfractométrie à différents temps d'élution pour les complexes N_{10} - P_2 et N_{10} - $2P_2$.

E : Titrage des anneaux N_{10} -ARN avec la phosphoprotéine.

B. Etude de l'interaction entre les anneaux N-ARN à 10 sous-unités (N₁₀) et la phosphoprotéine sur gel natif

Afin de visualiser l'interaction entre les anneaux N₁₀ et la phosphoprotéine par électrophorèse en gel natif, les mêmes échantillons utilisés pour les expériences de MALLS ont été déposés sur un gel en conditions natives à 4 %. Un retard sur le gel natif est constaté lors de l'addition de phosphoprotéine (Figure 37).

Le simple fait d'observer un retard sur gel indique que les anneaux et la phosphoprotéine interagissent mais aussi que cette interaction est suffisamment forte pour persister lors de la migration à travers de la matrice du gel natif. Lorsque la quantité de phosphoprotéine ajoutée est augmentée, plusieurs formes de complexes N₁₀-ARN-P peuvent être visualisées sur le gel natif sous forme de bandes discrètes. Le premier complexe (***) est formé par l'association entre un anneau N₁₀ et un dimère de phosphoprotéine. Le second complexe (***) formé est composé d'un anneau N₁₀ et de deux dimères de phosphoprotéine.

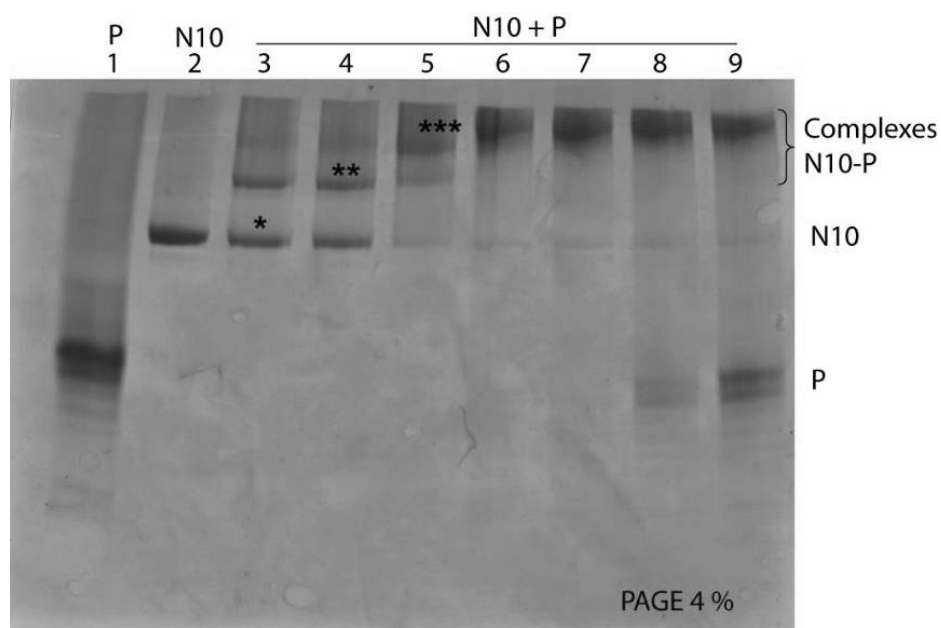


Figure 37 : Gel natif à 4 % mettant en évidence l'interaction entre N₁₀ et la phosphoprotéine et la formation des différents complexes N-ARN-P (Coloration bleu de coomassie).

* : complexe N₁₀ seuls. ** : complexe N₁₀ + 1 dimère de P. *** : complexe N₁₀ + 2 dimères de P.

Ligne 1 : phosphoprotéine seule.

Ligne 2 : complexe N-ARN avec 10 sous-unités de nucléoprotéines (N₁₀) seul.

Lignes 3 – 9 : complexes N₁₀ + P.

Ligne sur le gel natif :	2	3	4	5	6	7	8	9
Nombre théorique de dimères de P pour un anneau N-ARN :	0	0.25	0.5	1	2	1.5	3	4
Ratio N ₁₀ : dimère de P :	1:0	4:1	2:1	1:1	2:3	1:2	1:3	1:4

IV. Conclusion

Ces expériences montrent qu'un maximum de deux dimères de phosphoprotéine peuvent se fixer les anneaux N-ARN à 10 sous-unités de nucléoprotéine comme montré précédemment dans (Schoehn et al., 2001).

La Figure 38, présente un schéma de l'assemblage de la phosphoprotéine avec les anneaux ARN-N₁₀. On suppose pour cela qu'une unité de phosphoprotéine peut se fixer par nucléoprotéine. Il est possible que l'encombrement stérique généré par la fixation de deux dimères de phosphoprotéine sur l'anneau empêche la liaison éventuelle d'un troisième dimère de phosphoprotéine. Il est aussi probable que le changement conformationnel provoqué par la fixation de deux dimères de phosphoprotéine n'autorise pas la fixation d'un troisième dimère sur un anneau de nucléoprotéine. En revanche, sur une nucléocapside linéaire nettement moins enroulée qu'un anneau, l'espace disponible pourrait être suffisant pour permettre la fixation successive de davantage de dimères de phosphoprotéine.

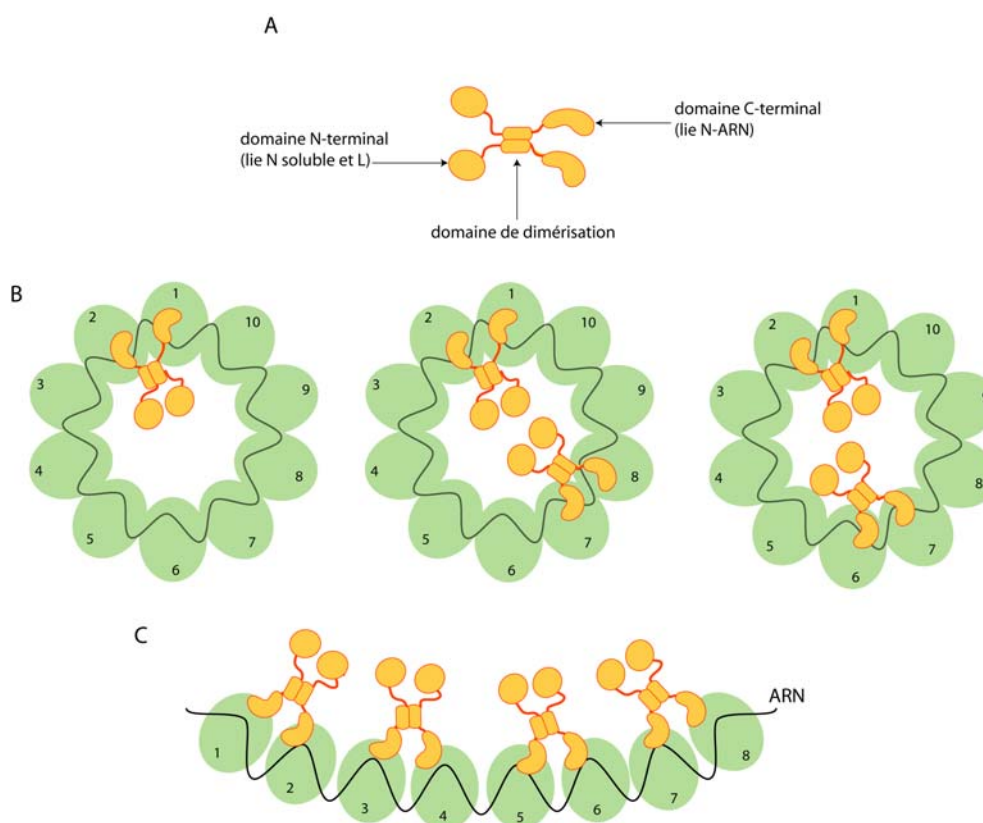


Figure 38 : Schéma de l'assemblage de la phosphoprotéine avec les anneaux ARN-N₁₀.

A : Dimère de phosphoprotéine. La phosphoprotéine est composée de trois domaines structurés : le domaine N-terminal capable de fixer la nucléoprotéine soluble, le domaine de dimérisation, le domaine C-terminal interagissant avec la nucléoprotéine engagée dans les complexes N-ARN.

B : Hypothèse pour l'assemblage entre un ou deux dimères de phosphoprotéine et un anneau N₁₀.

C : Hypothèse pour l'assemblage entre les dimères de phosphoprotéine et une nucléocapside.

Plusieurs possibilités d'arrangements existent pour fixer deux dimères de phosphoprotéine sur un anneau N_{10} . Le complexe formé par l'association de deux dimères de protéine P et un anneau N_{10} est par conséquent un mauvais candidat en vue d'obtenir des reconstructions par microscopie électronique ainsi que pour la cristallographie.

Il serait nécessaire de réaliser des expériences complémentaires de fixation de la phosphoprotéine sur les anneaux formés de 9, 11 et 12 nucléoprotéines. En effet, il est envisageable que les anneaux composés de 12 nucléoprotéines ou plus permettent la fixation de 3 dimères de protéines P, l'espace disponible au centre de l'anneau étant plus grand.

L'étude de ces complexes par microscopie électronique et cristallographie est essentielle afin de visualiser les changements structuraux que subit la nucléoprotéine lorsqu'elle fixe la phosphoprotéine.

Le domaine C-terminal isolé de la phosphoprotéine (P_{Cter}) de la rage se lie à la nucléoprotéine engagée dans la matrice N-ARN. Ce domaine de 145 acides aminés est monomérique (Mavrakis et al., 2004). Nous pensons l'utiliser afin d'obtenir un complexe N-ARN- P_{Cter} . Il est probable que ce complexe soit plus homogène et donc plus approprié à la microscopie électronique ou la cristallographie. En effet, la taille réduite de ce fragment devrait permettre de saturer les sites de fixation de la phosphoprotéine disponibles sur les complexes N-ARN (Figure 39).

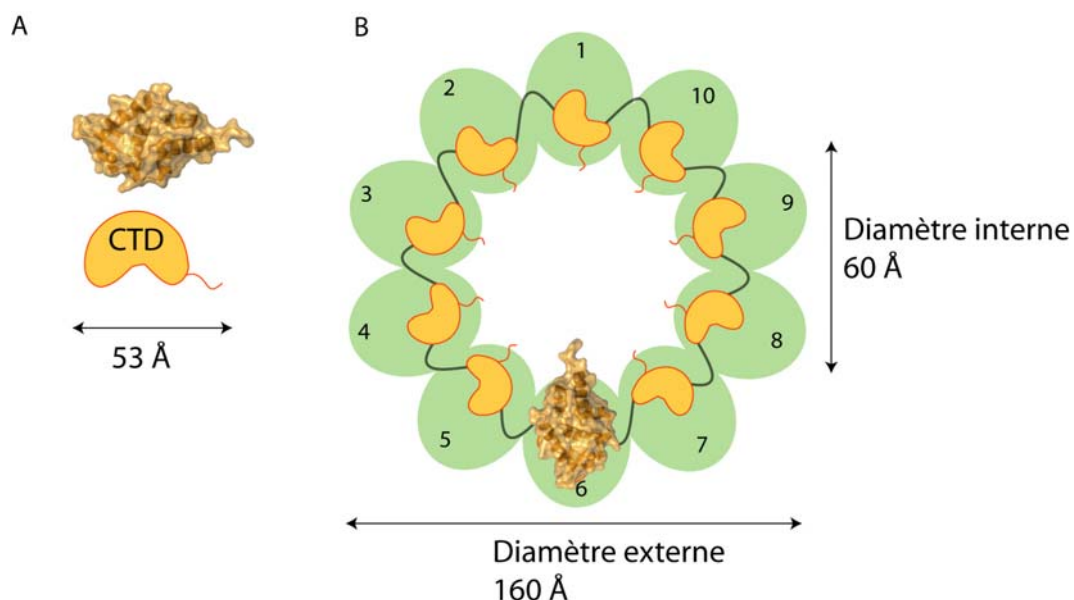


Figure 39 : Schéma de l'assemblage du domaine C-terminal de phosphoprotéine avec les anneaux ARN- N_{10} .

Chaque sous-unité de nucléoprotéine pourrait lier une sous-unité de P_{Cter} . CTD = domaine C-terminal
A : Structure cristallographique du domaine C-terminal de la phosphoprotéine.

B : Hypothèse selon laquelle chaque sous-unité de nucléoprotéine de l'anneau N_{10} peut lier une unité P_{Cter} .

Des expériences complémentaires sont nécessaires afin de vérifier cette hypothèse. La stœchiométrie du complexe ARN-N-P_{Cter} pourra être vérifiée en réalisant des mesures en diffusion de lumière similaires à celles mises en œuvre pour caractériser les complexes N-ARN avec la phosphoprotéine non protéolysée. Des expériences utilisant les complexes entre les anneaux N₁₀ et N₁₁ et le domaine C-terminal de la phosphoprotéine sont actuellement en cours.

Partie VI : Discussion des résultats scientifiques

I. Comparaison de la structure de la nucléoprotéine du virus de la rage avec les structures de nucléoprotéines des autres *Mononegavirales*

A. Comparaison entre la structure de la nucléoprotéine du virus de la rage et celle du VSV

La Figure 40 montre une reconstruction de microscopie électronique d'anneaux à 10 sous-unités de nucléoprotéine chez le virus de la stomatite vésiculaire (Chen et al., 2004). En 2004, les auteurs pensaient que l'ARN était localisé au niveau de l'extrémité C-terminale de la nucléoprotéine.

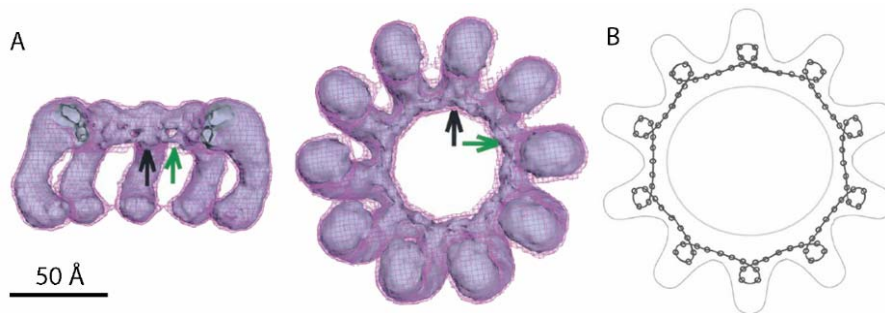


Figure 40 : Reconstruction en microscopie électronique des anneaux N-ARN du VSV (Chen et al., 2004).

A : Les flèches indiquent la localisation de l'ARN proposée par les auteurs.

B : Proposition d'une hypothèse pour l'empaquetage de l'ARN de 90 bases dans la structure en anneaux.

Le colorant utilisé pour les images de microscopie est l'acétate d'uranyle. Ce composé utilisé en microscopie électronique pour la coloration négative possède une forte affinité pour les acides nucléiques. Lors de la reconstruction en microscopie électronique, les auteurs ont localisé une densité plus intense en forme de cercle ressemblant à de l'ARN au sommet des sous-unités de nucléoprotéine. La localisation de l'ARN était inattendue mais néanmoins possible : l'ARN pouvait être à la fois protégé et se rendre accessible à l'ARN-polymérase virale par une transition structurale.

La Figure 41 montre une comparaison entre la structure de cryomicroscopie électronique des anneaux N₁₀-ARN de rage (Schoehn et al., 2001), et la structure aux rayons X des anneaux N₁₁-ARN. La Figure 41 C présente la superposition de la structure protéolysée et non protéolysée des anneaux de nucléoprotéine du virus de la rage. Il apparaît alors clairement que la partie digérée par la trypsine lors de la protéolyse est un petit groupe d'hélices α localisé au centre de l'anneau, et formant la structure qui a pu être confondue avec de l'ARN lors de l'étude des anneaux du VSV.

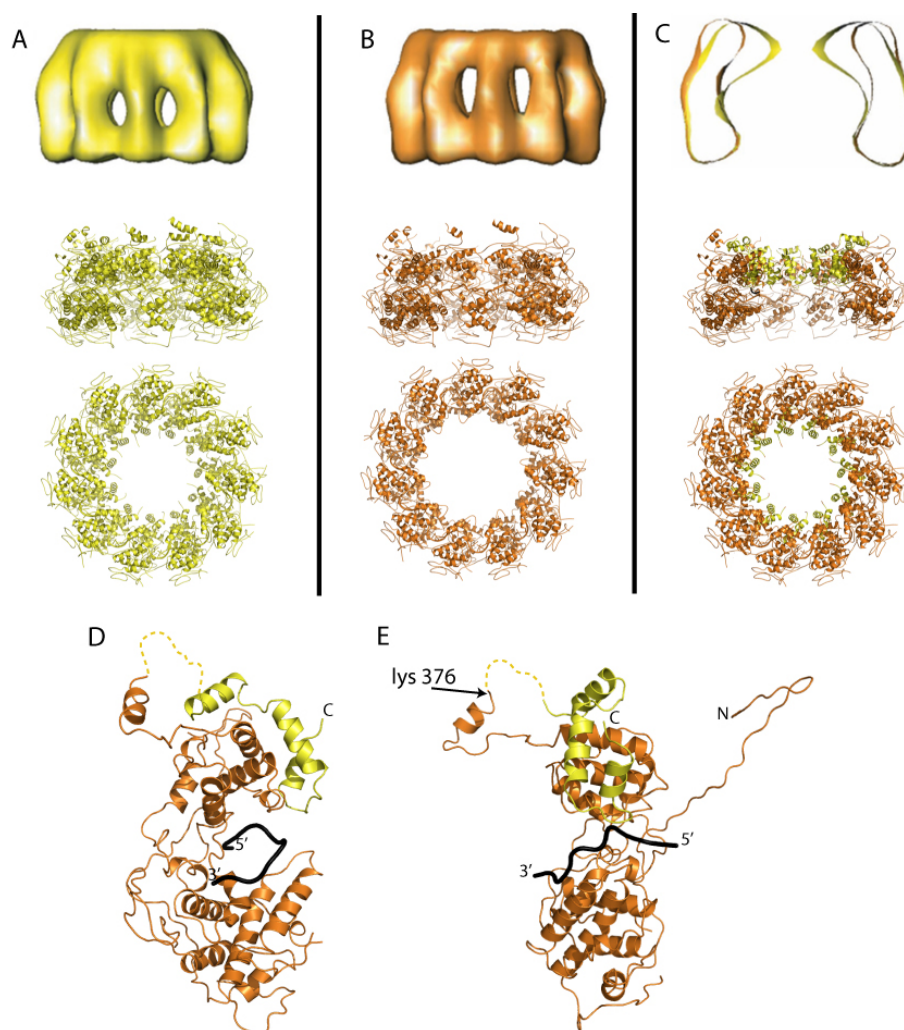


Figure 41 : Comparaison entre la structure obtenue par microscopie électronique des N10 et celle obtenue par cristallographie aux rayons X des N11.

En jaune : nucléoprotéine non protéolysée. En orange : nucléoprotéine protéolysée.

A : Anneaux non digérés. B : Anneaux digérés. C : Superposition des anneaux non digérés et digérés.

D : Vue latérale du protomère de nucléoprotéine du virus de la rage. E : Vue frontale du protomère de nucléoprotéine du virus de la rage. Le site de coupure au niveau du résidu lysine 376 est indiqué.

En pointillés : partie de la nucléoprotéine du virus de la rage non définie dans la densité électronique (résidus 374-397).

Le VSV est un proche « cousin » du virus de la rage car c'est un membre de la famille des Rhabdovirus, appartenant toutefois au genre Vesiculovirus (Tableau 1). La structure cristallographique de la nucléoprotéine du VSV a été résolue et publiée en même temps que celle du virus de la rage (Green et al., 2006). Cette structure à 2.9 Å de résolution, est un anneau composé de 10 sous-unités de nucléoprotéine et d'un ARN de 90 bases. Les séquences en acides aminés des nucléoprotéines du VSV et du virus de la rage ne sont identiques qu'à 30 %. Pourtant, il apparaît sur la Figure 42, que les deux structures sont très semblables et se superposent avec un rmsd de 2.18 Å sur les C α .

Le protomère de nucléoprotéine peut être divisé en deux domaines principaux (mâchoire supérieure et mâchoire inférieure) qui séquestrent l'ARN. Le sillon où se loge l'ARN se situe au même endroit sur ces deux nucléoprotéines. L'ARN présente une organisation quasi identique dans les deux structures : huit bases sur neuf sont empilées de la même manière, et chaque nucléoprotéine lie 9 nucléotides d'ARN (Figure 42 C).

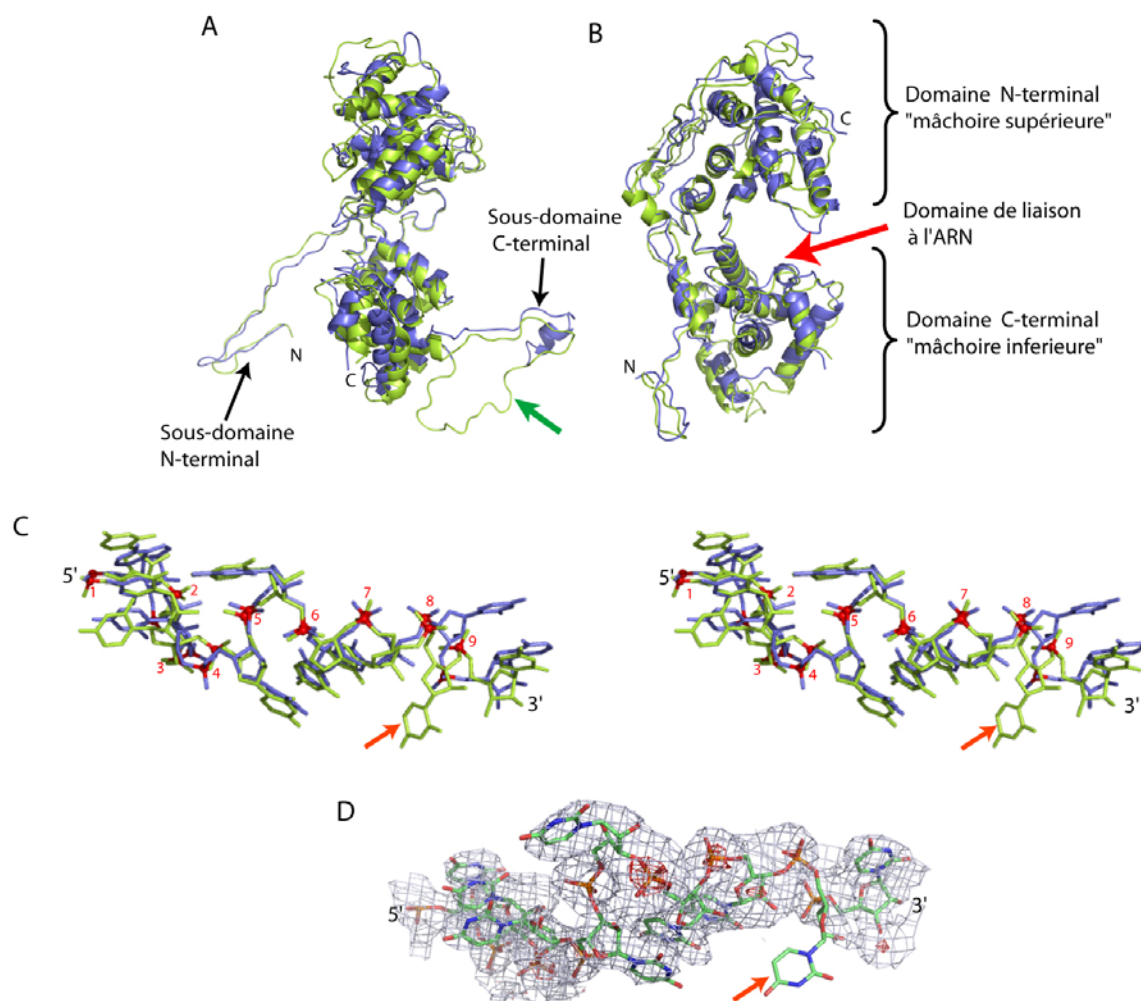


Figure 42 : Superposition des protomères de la nucléoprotéine du virus de la rage et de la nucléoprotéine du VSV.

La nucléoprotéine et l'ARN du virus de la rage sont colorés en bleu (code pdb : 2GTT), La nucléoprotéine et l'ARN du VSV sont colorés en vert (code pdb : 2GIC). Rmsd : 2.18 Å.

A : Vue de face, la flèche verte indique la partie du sous-domaine C-terminal qui a été construite dans la structure du VSV.

B : Vue de coté, la flèche rouge indique le domaine de liaison à l'ARN.

C : Stéréovue d'une superposition des 9 bases répétées d'ARN du VSV et du virus de la rage. Les atomes de phosphate sont numérotés et marqués en rouge. La base d'ARN 8 indiquée par la flèche de couleur orange ne réalise pas d'empilement avec la base 9 chez le VSV.

D : Densité électronique autour de l'ARN dans la structure des anneaux N-ARN du VSV. (Densité rouge : contour 3 σ , densité bleue : contour 1 σ).

Sur l'ARN, la base 8 de la structure du VSV pointe vers la protéine alors que dans la structure du virus de la rage, cette base 8 pointe en direction du solvant. Après analyse de la carte de densité électronique de la structure des anneaux de VSV calculée à partir des facteurs de structures déposés, il s'avère que la densité de cette base d'ARN numéro huit est très mal définie (Figure 42 D et Figure 35).

Deux sous-domaines en N-terminal et en C-terminal interagissent avec les deux protomères voisins et présentent, comme chez le virus de la rage une flexibilité importante. De plus, l'enchaînement des structures secondaires de ces deux nucléoprotéines est identique.

Contrairement à ce que l'on observe dans la structure de la nucléoprotéine du virus de la rage, dans la structure de la nucléoprotéine du VSV, la région comprise entre le domaine émergent C-terminal et la «mâchoire inférieure» de la nucléoprotéine a été construite dans deux protomères sur cinq. Ce domaine est très flexible comme le montre la représentation de chaque protomère en fonction des facteurs de température (Figure 43). Le facteur de température (B_{fact}) permet d'évaluer la présence éventuelle de désordre dans une structure. La région entre le sous domaine C-terminal et le corps de la nucléoprotéine ont des B_{fact} élevés que ce soit pour la nucléoprotéine du virus de la rage ou celle du VSV.

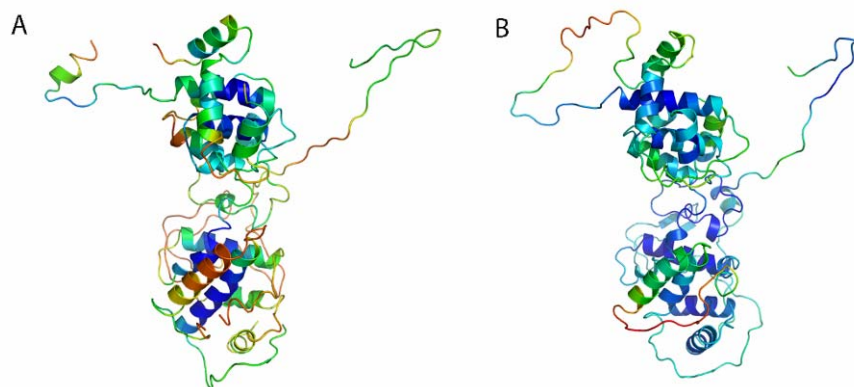


Figure 43 : Représentations des protomères de nucléoprotéine du VSV et du virus de la rage en fonction des facteurs de températures.

Gradient de couleur rouge à bleu : B_{fact} élevé à B_{fact} faible.

A : Protomère de nucléoprotéine du virus de la rage.

B : Protomère de nucléoprotéine du VSV.

L'unité asymétrique de la structure des anneaux de nucléoprotéine du VSV est un demi-anneau, elle est constituée de 5 protomères de nucléoprotéine et de 45 bases d'ARN (groupe d'espace P2₁2₁2₁). Dans le cas des anneaux de la rage, l'unité asymétrique de la structure est un double anneau, elle est composée de 22 protomères et de deux brins d'ARN de 99 bases (groupe d'espace P2₁2₁2). L'organisation cristalline de ces deux structures est très différente : pour le VSV, les anneaux sont empilés et interagissent par des contacts N-terminal – C-

terminal entre les nucléoprotéines. En revanche, dans le cas du virus de la rage, les anneaux sont arrangés deux à deux par des contacts N-terminal – N-terminal résultant en une unité asymétrique de 22 nucléoprotéines.

Au-delà des considérations purement cristallographiques, ces structures sont quasi identiques et délivrent la même information : que ce soit pour le VSV ou le virus de la rage, l'ARN est séquestré et la nucléoprotéine doit subir un changement de conformation significatif pour rendre l'ARN accessible à l'ARN-polymérase virale.

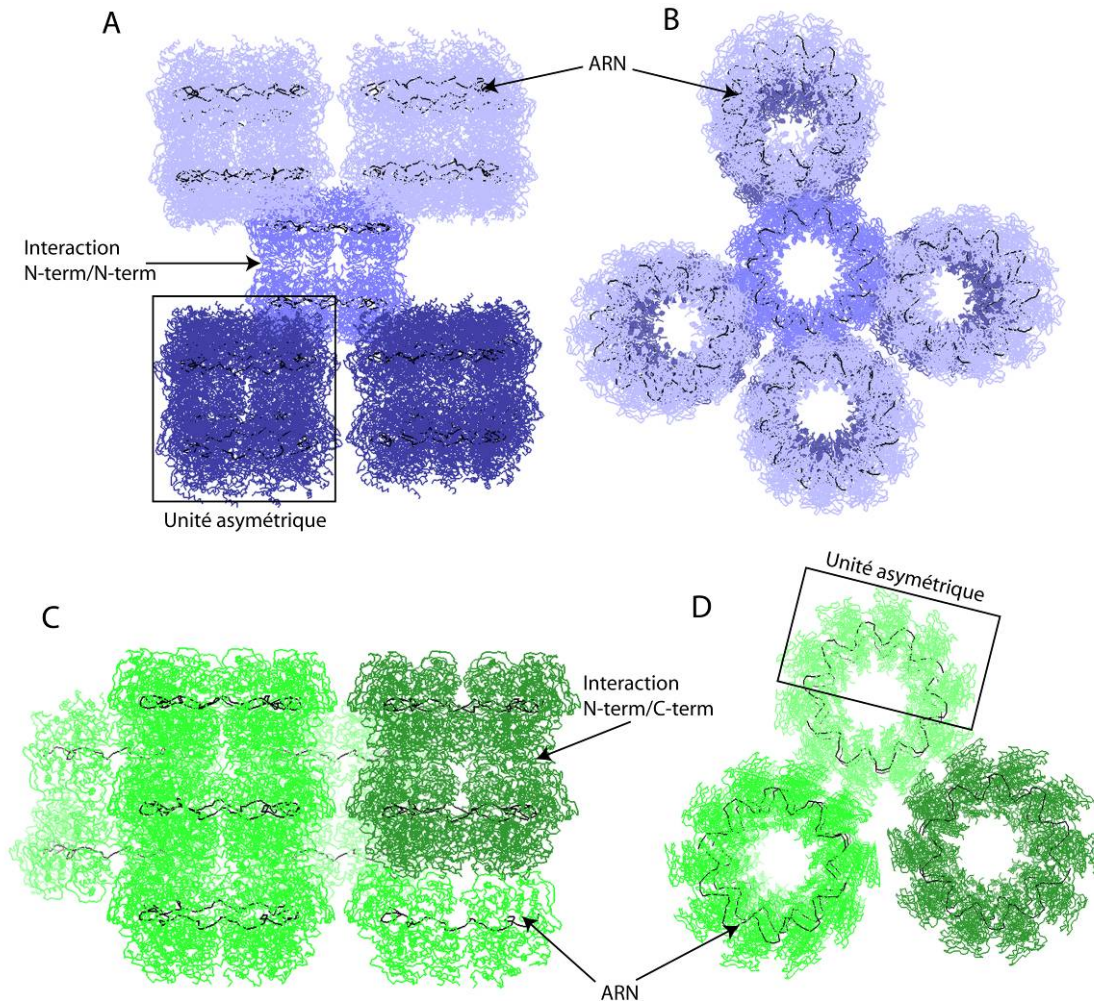


Figure 44 : Arrangement cristallin à l'intérieur des cristaux d'anneaux N-ARN du virus de la rage et du VSV.

A et B : Vue latérale (A) et vue supérieure (B) de l'arrangement cristallin à l'intérieur des cristaux N₁₁-ARN du virus de la rage.

C et D : Vue latérale (C) et vue supérieure (D) de l'arrangement cristallin à l'intérieur des cristaux N₁₀-ARN du VSV.

B. Comparaison des structures de la nucléoprotéine du virus de la rage et du Bornavirus

La nucléoprotéine du Bornavirus a été cristallisée et sa structure a été résolue sans ARN ou protéine chaperonne associée (Rudolph et al., 2003). Il n'existe pas ou très peu d'informations structurales sur la nucléocapside du Bornavirus. A la différence des virus de la rage et de la stomatite vésiculaire, il est possible de produire la nucléoprotéine du Bornavirus dans un système d'expression bactérien sans qu'elle soit liée à de l'ARN cellulaire. La structure de la nucléoprotéine du Bornavirus apparaît très différente de celle de la rage ou du VSV.

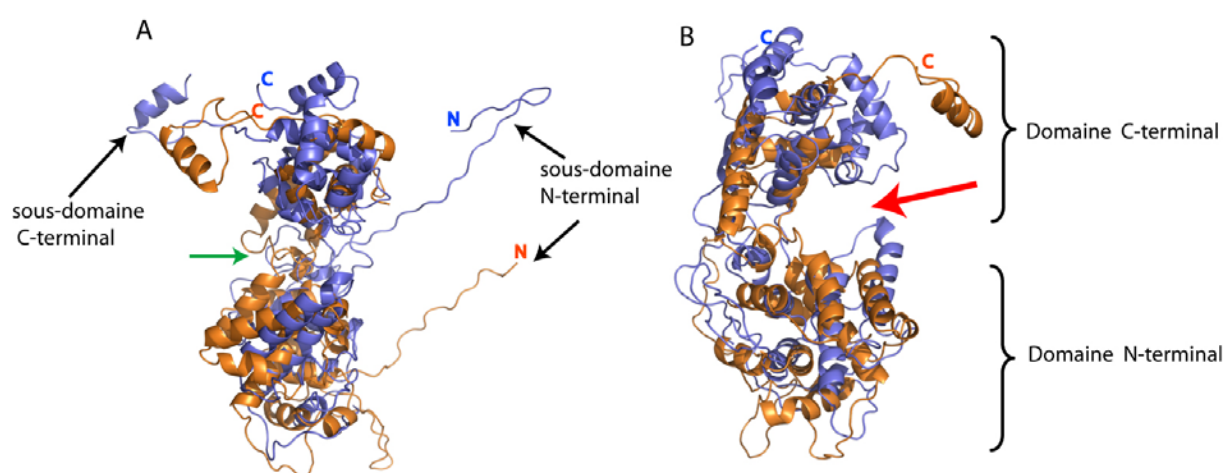


Figure 45 : Superposition des protomères de la nucléoprotéine du virus de la rage et de la nucléoprotéine du Bornavirus.

La nucléoprotéine de la rage est coloré en bleu (code pdb : 2GTT), celle de Bornavirus est colorée en orange (code pdb : 1N93). Rmsd : 12.06 Å.

A : Vue frontale. La flèche verte indique la séparation entre le domaine N-terminal et le domaine C-terminal.

B : Vue latérale. La flèche rouge indique le sillon où se loge l'ARN de la nucléoprotéine du virus de la rage.

Les structures des nucléoprotéines du Bornavirus et du virus de la rage se superposent approximativement (Figure 45) avec un rmsd (en utilisant les $C\alpha$) de 12.06 Å. En effet, l'organisation des éléments de structure secondaire est très différente. Néanmoins ces deux structures présentent des similitudes « morphologiques » en terme d'organisation au sein des protomères. Dans les deux cas, les structures secondaires sont principalement des hélices α , et la nucléoprotéine peut être divisée en deux domaines C-terminal et N-terminal distincts. Sur la vue latérale de la nucléoprotéine de Bornavirus (Figure 45 B), il est possible de distinguer un sillon contenant des résidus basiques qui pourraient être impliqués dans la liaison à l'ARN. La nucléoprotéine du Bornavirus présente elle aussi deux sous-domaines qui émergent du

corps de la nucléoprotéine. Il est probable que ces deux sous-domaines flexibles soient eux aussi impliqués dans la formation de complexes nucléoprotéine-ARN de dimensions variables.

II. Hypothèse pour le mécanisme de la réplication chez le virus de la rage

L'obtention de la structure de la nucléoprotéine complexée à l'ARN ne permet pas de compléter avec certitude le mécanisme de la réplication. Néanmoins cette structure apporte des données supplémentaires. La nucléoprotéine est capable de séquestrer l'ARN et de l'isoler de son environnement extérieur. Elle doit donc subir des changements de conformation drastiques lors du passage de l'ARN-polymérase virale. Nous avons émis l'hypothèse que la phosphoprotéine jouerait un rôle « d'interrupteur » pour ce changement de conformation. Ainsi, lorsqu'une sous-unité de nucléoprotéine serait liée par la phosphoprotéine, les « mâchoires » de cette nucléoprotéine se desserreraient, libérant ainsi l'ARN viral. Il est aussi nécessaire d'envisager que ce signal apporté par la phosphoprotéine soit transmis aux deux sous-unités de nucléoprotéines suivantes afin que la longueur d'ARN libérée soit suffisante.

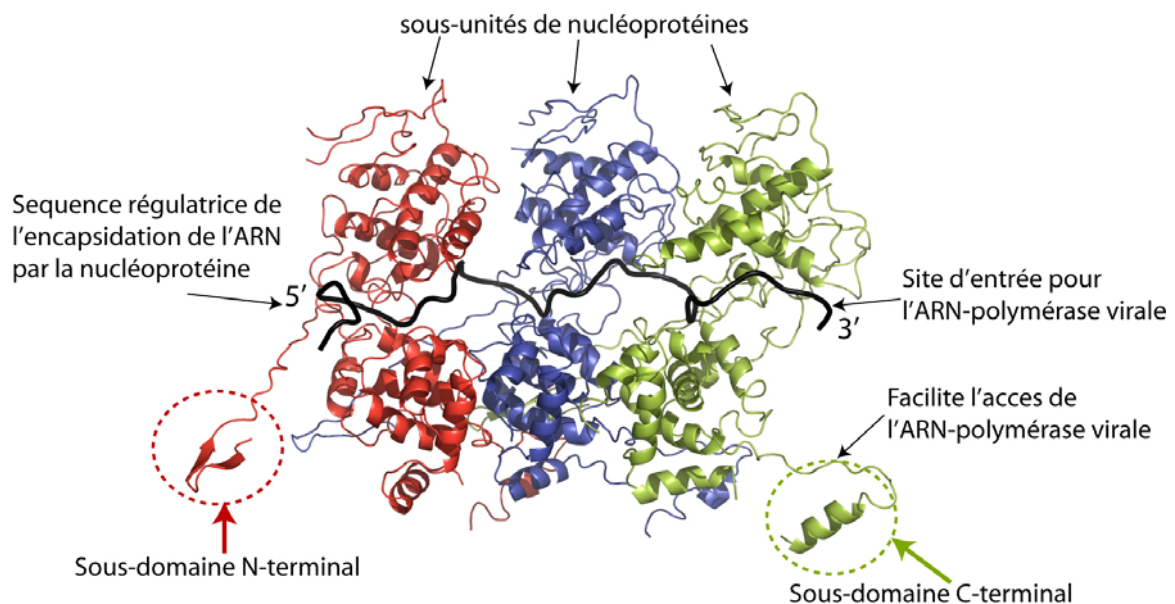


Figure 46 : Représentation de trois protomères de nucléoprotéine du virus de la rage liés à l'ARN. Les sous-domaines au niveau des extrémités 3' et 5' de l'ARN sont libres.

En analysant la structure de la nucléoprotéine, nous avons aussi émis des hypothèses sur le rôle fonctionnel des sous-domaines C- et N-terminaux. Au sein d'une nucléocapside, ces sous-domaines permettent aux nucléoprotéines de se maintenir en place tout en gardant une certaine flexibilité. Néanmoins, sur un ARN viral « linéaire », à chaque extrémité (3' et 5') un

sous-domaine est libre (Figure 46). Au niveau de l'extrémité 5' de l'ARN viral, le sous-domaine N-terminal de la nucléoprotéine n'est pas engagé dans un contact nucléoprotéine-nucléoprotéine. Nous pensons que ce sous-domaine N-terminal pourrait jouer un rôle important lors de l'encapsidation de l'ARN synthétisé, en reconnaissant peut-être de manière spécifique la séquence de l'ARN leader. Un domaine C-terminal est également libre à l'extrémité 3' de l'ARN viral. Le site d'entrée de l'ARN-polymérase virale se situant également en 3' de l'ARN viral, il est envisageable que le sous-domaine C-terminal facilite l'accès de l'ARN-polymérase virale à l'ARN viral.

La

Figure 47 présente une hypothèse expliquant le mécanisme de réplication du virus de la rage. Ce mécanisme mettrait en jeu le complexe phosphoprotéine-ARN-polymérase, le complexe nucléoprotéine-phosphoprotéine et une matrice d'ARN viral encapsidée par la nucléoprotéine. Nous supposons que la nucléoprotéine engagée dans les complexes N-ARN est phosphorylée, la phosphorylation se produisant lors de l'assemblage de la nucléocapside (Yang et al., 1999). Nous faisons aussi l'hypothèse que la stœchiométrie du complexe dimère de phosphoprotéine / ARN-polymérase est 1:1, ce qui n'a jamais été prouvé scientifiquement. Enfin, nous supposons également que le dimère de phosphoprotéine lie la nucléoprotéine soluble grâce à une interaction entre le domaine N-terminal de la phosphoprotéine et le domaine de fixation à l'ARN de la nucléoprotéine (Mavrakis et al., 2006).

Voici un bref descriptif des étapes successives du schéma fonctionnel proposé :

- 1) La phosphoprotéine interagit avec l'ARN-polymérase par son extrémité N-terminale (Chenik et al., 1998). Le sous-domaine C-terminal libre de la nucléoprotéine sur l'extrémité 3' de l'ARN viral constitue un signal d'entrée pour le complexe ARN-polymérase-phosphoprotéine.
- 2) Le domaine C-terminal de la phosphoprotéine interagit avec la nucléoprotéine contenue dans les nucléocapsides (Schoehn et al., 2001). Ceci provoque un changement de conformation au sein de la nucléoprotéine. On peut imaginer une transition structurale au niveau des deux domaines de la nucléoprotéine provoquée par la fixation de la phosphoprotéine. Cela aboutit localement à la libération de l'ARN par la nucléoprotéine.

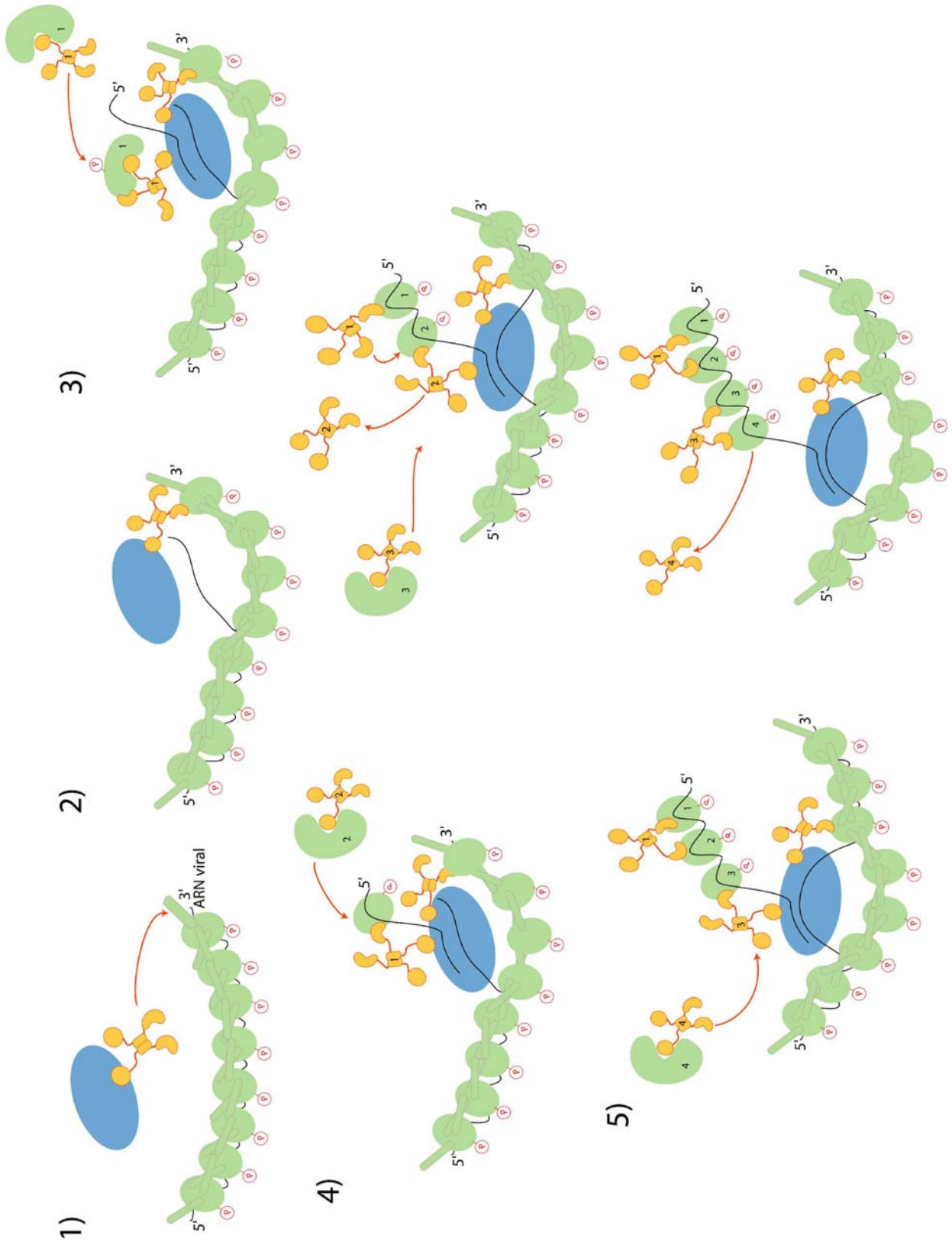


Figure 47 : Schéma proposé pour la réplication du virus de la rage.

3) La nucléoprotéine soluble, en complexe avec un dimère de phosphoprotéine, forme le complexe N^o-P soluble. Lorsque la concentration en N^o-P atteint un seuil critique, la phosphoprotéine interagit avec l'ARN-polymérase et conduirait ainsi la nucléoprotéine à proximité de la chaîne d'ARN naissante.

La nucléoprotéine contenue dans le complexe N^o-P est alors phosphorylée par la CKII et subie un changement conformationnel permettant sa liaison à l'ARN naissant (Toriumi and Kawai, 2004). Dès sa sortie du complexe ARN-polymérase virale, l'ARN serait ainsi lié à la nucléoprotéine.

4) Au fur et à mesure de sa synthèse, l'ARN requiert d'être encapsidé. La nucléoprotéine suivante, associée à un dimère de phosphoprotéine, lie l'ARN naissant. Il est envisageable qu'une fois liée, la phosphoprotéine fixée sur la sous-unité de nucléoprotéine précédente interagisse avec cette nucléoprotéine et libère alors un dimère de phosphoprotéine.

5) Ainsi, de manière séquentielle, les nucléoprotéines se lieraient à la matrice N-ARN en libérant une fois sur deux un dimère de phosphoprotéine.

Il manque de nombreuses données expérimentales afin de valider ces hypothèses. Des données structurales sur le complexe ARN-polymérase - phosphoprotéine en particulier permettront de compléter les connaissances. Orienter les recherches sur la stœchiométrie phosphoprotéine : nucléoprotéine dans les complexes avec l'ARN viral est aussi déterminant afin de mieux comprendre le mécanisme de la réplication. Enfin, des études exhaustives ayant pour objectif une meilleure caractérisation des zones d'interaction entre la nucléoprotéine et la phosphoprotéine lors du processus d'assemblage s'avèrent toujours nécessaires.

Références Bibliographiques

- Abraham, G., Rhodes, D.P. and Banerjee, A.K. (1975a) The 5' terminal structure of the methylated mRNA synthesized in vitro by vesicular stomatitis virus. *Cell*, **5**, 51-58.
- Abraham, G., Rhodes, D.P. and Banerjee, A.K. (1975b) Novel initiation of RNA synthesis in vitro by vesicular stomatitis virus. *Nature*, **255**, 37-40.
- Albertini, A.A., Wernimont, A.K., Muziol, T., Ravelli, R.B., Clapier, C.R., Schoehn, G., Weissenhorn, W. and Ruigrok, R.W. (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science*, **313**, 360-363.
- Area, E., Martin-Benito, J., Gastaminza, P., Torreira, E., Valpuesta, J.M., Carrascosa, J.L. and Ortin, J. (2004) 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc Natl Acad Sci U S A*, **101**, 308-313.
- Badrane, H., Bahloul, C., Perrin, P. and Tordo, N. (2001) Evidence of two Lyssavirus phylogroups with distinct pathogenicity and immunogenicity. *J Virol*, **75**, 3268-3276.
- Baltimore, D., Huang, A.S. and Stampfer, M. (1970) Ribonucleic acid synthesis of vesicular stomatitis virus, II. An RNA polymerase in the virion. *Proc Natl Acad Sci U S A*, **66**, 572-576.
- Banerjee, A.K. (1987) Transcription and replication of rhabdoviruses. *Microbiol Rev*, **51**, 66-87.
- Barge, A., Gaudin, Y., Coulon, P. and Ruigrok, R.W. (1993) Vesicular stomatitis virus M protein may be inside the ribonucleocapsid coil. *J Virol*, **67**, 7246-7253.
- Baudin, F., Bach, C., Cusack, S. and Ruigrok, R.W. (1994) Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *Embo J*, **13**, 3158-3165.
- Bhella, D., Ralph, A., Murphy, L.B. and Yeo, R.P. (2002) Significant differences in nucleocapsid morphology within the Paramyxoviridae. *J Gen Virol*, **83**, 1831-1839.
- Blondel, D., Regad, T., Poisson, N., Pavie, B., Harper, F., Pandolfi, P.P., De The, H. and Chelbi-Alix, M.K. (2002) Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies. *Oncogene*, **21**, 7957-7970.
- Blumberg, B.M., Giorgi, C., Rose, K. and Kolakofsky, D. (1984) Preparation and analysis of the nucleocapsid proteins of vesicular stomatitis virus and sendai virus, and analysis of the sendai virus leader-NP gene region. *J Gen Virol*, **65 (Pt 4)**, 769-779.
- Bourhy, H., Kissi, B. and Tordo, N. (1993) Molecular diversity of the Lyssavirus genus. *Virology*, **194**, 70-81.
- Bourhy, H., Tordo, N., Lafon, M. and Sureau, P. (1989) Complete cloning and molecular organization of a rabies-related virus, Mokola virus. *J Gen Virol*, **70 (Pt 8)**, 2063-2074.
- Bricogne, D.I.F.a. (1997) SHARP: A maximum-likelihood heavy-atom parameter refinement program for the MIR and MAD Methods.
- Brzozka, K., Finke, S. and Conzelmann, K.K. (2005) Identification of the rabies virus alpha/beta interferon antagonist: phosphoprotein P interferes with phosphorylation of interferon regulatory factor 3. *J Virol*, **79**, 7673-7681.
- Buchholz, C.J., Spehner, D., Drillien, R., Neubert, W.J. and Homann, H.E. (1993) The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly. *J Virol*, **67**, 5803-5812.
- Calain, P. and Roux, L. (1993) The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *J Virol*, **67**, 4822-4830.

- Chattopadhyay, D. and Banerjee, A.K. (1987) Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro. *Cell*, **49**, 407-414.
- Chen, J.L., Das, T. and Banerjee, A.K. (1997) Phosphorylated states of vesicular stomatitis virus P protein in vitro and in vivo. *Virology*, **228**, 200-212.
- Chen, Z., Green, T.J., Luo, M. and Li, H. (2004) Visualizing the RNA molecule in the bacterially expressed vesicular stomatitis virus nucleoprotein-RNA complex. *Structure (Camb)*, **12**, 227-235.
- Chenik, M., Chebli, K. and Blondel, D. (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**, 707-712.
- Chenik, M., Chebli, K., Gaudin, Y. and Blondel, D. (1994) In vivo interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein. *J Gen Virol*, **75 (Pt 11)**, 2889-2896.
- Chenik, M., Schnell, M., Conzelmann, K.K. and Blondel, D. (1998) Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J Virol*, **72**, 1925-1930.
- Chong, L.D. and Rose, J.K. (1993) Membrane association of functional vesicular stomatitis virus matrix protein in vivo. *J Virol*, **67**, 407-414.
- Chua, K.B., Bellini, W.J., Rota, P.A., Harcourt, B.H., Tamin, A., Lam, S.K., Ksiazek, T.G., Rollin, P.E., Zaki, S.R., Shieh, W., Goldsmith, C.S., Gubler, D.J., Roehrig, J.T., Eaton, B., Gould, A.R., Olson, J., Field, H., Daniels, P., Ling, A.E., Peters, C.J., Anderson, L.J. and Mahy, B.W. (2000) Nipah virus: a recently emergent deadly paramyxovirus. *Science*, **288**, 1432-1435.
- Colonno, R.J. and Banerjee, A.K. (1976) A unique RNA species involved in initiation of vesicular stomatitis virus RNA transcription in vitro. *Cell*, **8**, 197-204.
- Conzelmann, K.K., Cox, J.H., Schneider, L.G. and Thiel, H.J. (1990) Molecular cloning and complete nucleotide sequence of the attenuated rabies virus SAD B19. *Virology*, **175**, 485-499.
- Curran, J., Pelet, T. and Kolakofsky, D. (1994) An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation. *Virology*, **202**, 875-884.
- Dessen, A., Volchkov, V., Dolnik, O., Klenk, H.D. and Weissenhorn, W. (2000) Crystal structure of the matrix protein VP40 from Ebola virus. *Embo J*, **19**, 4228-4236.
- Dietzschold, B. and Koprowski, H. (2004) Rabies transmission from organ transplants in the USA. *Lancet*, **364**, 648-649.
- Ding, H., Green, T.J., Lu, S. and 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. and 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.
- Duarte, C.M. and Pyle, A.M. (1998) Stepping through an RNA structure: A novel approach to conformational analysis. *J Mol Biol*, **284**, 1465-1478.
- Egelman, E.H., Wu, S.S., Amrein, M., Portner, A. and Murti, G. (1989) The Sendai virus nucleocapsid exists in at least four different helical states. *J Virol*, **63**, 2233-2243.
- 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. and Wagner, R.R. (1972) Dissociation and reconstitution of the transcriptase and template activities of vesicular stomatitis B and T virions. *J Virol*, **10**, 297-309.

- Emerson, S.U. and Yu, Y. (1975) Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. *J Virol*, **15**, 1348-1356.
- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D*, **60**, 2126-2132.
- Finke, S. and Conzelmann, K.K. (2003) Dissociation of rabies virus matrix protein functions in regulation of viral RNA synthesis and virus assembly. *J Virol*, **77**, 12074-12082.
- Finke, S., Mueller-Waldeck, R. and Conzelmann, K.K. (2003) Rabies virus matrix protein regulates the balance of virus transcription and replication. *J Gen Virol*, **84**, 1613-1621.
- Flamand, A., Raux, H., Gaudin, Y. and Ruigrok, R.W. (1993) Mechanisms of rabies virus neutralization. *Virology*, **194**, 302-313.
- Gastka, M., Horvath, J. and Lentz, T.L. (1996) Rabies virus binding to the nicotinic acetylcholine receptor alpha subunit demonstrated by virus overlay protein binding assay. *J Gen Virol*, **77** (Pt 10), 2437-2440.
- Gaudier, M., Gaudin, Y. and Knossow, M. (2002) Crystal structure of vesicular stomatitis virus matrix protein. *Embo J*, **21**, 2886-2892.
- Gaudin, Y., Ruigrok, R.W., Tuffereau, C., Knossow, M. and Flamand, A. (1992) Rabies virus glycoprotein is a trimer. *Virology*, **187**, 627-632.
- Gaudin, Y., Tuffereau, C., Benmansour, A. and Flamand, A. (1991a) Fatty acylation of rabies virus proteins. *Virology*, **184**, 441-444.
- Gaudin, Y., Tuffereau, C., Segretain, D., Knossow, M. and Flamand, A. (1991b) Reversible conformational changes and fusion activity of rabies virus glycoprotein. *J Virol*, **65**, 4853-4859.
- Gerard, F., Albertini, A., Ramos, C., Garden, J., Guillou, H., Ruigrok, R. and Jamin, M. (2006) Unphosphorylated Rhabdivirus phosphoproteins are structurally disordered dimers. *en soumission*.
- Gigant, B., Iseni, F., Gaudin, Y., Knossow, M. and Blondel, D. (2000) Neither phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is required for its oligomerization. *J Gen Virol*, **81**, 1757-1761.
- Glodowski, D.R., Petersen, J.M. and Dahlberg, J.E. (2002) Complex nuclear localization signals in the matrix protein of vesicular stomatitis virus. *J Biol Chem*, **277**, 46864-46870.
- Gombart, A.F., Hirano, A. and Wong, T.C. (1995) Nucleoprotein phosphorylated on both serine and threonine is preferentially assembled into the nucleocapsids of measles virus. *Virus Res*, **37**, 63-73.
- Green, T.J., Macpherson, S., Qiu, S., Lebowitz, J., Wertz, G.W. and 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., Zhang, X., Wertz, G.W. and Luo, M. (2006) Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science*, **313**, 357-360.
- Gupta, A.K., Blondel, D., Choudhary, S. and Banerjee, A.K. (2000) The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C. *J Virol*, **74**, 91-98.
- Heggeness, M.H., Scheid, A. and Choppin, P.W. (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**, 2631-2635.
- Heldwein, E.E., Lou, H., Bender, F.C., Cohen, G.H., Eisenberg, R.J. and Harrison, S.C. (2006) Crystal structure of glycoprotein B from herpes simplex virus 1. *Science*, **313**, 217-220.
- Her, L.S., Lund, E. and Dahlberg, J.E. (1997) Inhibition of Ran guanosine triphosphatase-dependent nuclear transport by the matrix protein of vesicular stomatitis virus. *Science*, **276**, 1845-1848.

- Hogenhout, S.A., Redinbaugh, M.G. and Ammar el, D. (2003) Plant and animal rhabdovirus host range: a bug's view. *Trends Microbiol*, **11**, 264-271.
- Holmes, D.E. and Moyer, S.A. (2002) The phosphoprotein (P) binding site resides in the N terminus of the L polymerase subunit of sendai virus. *J Virol*, **76**, 3078-3083.
- Horikami, S.M., Curran, J., Kolakofsky, D. and Moyer, S.A. (1992) Complexes of Sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication in vitro. *J Virol*, **66**, 4901-4908.
- Iseni, F., Barge, A., Baudin, F., Blondel, D. and Ruigrok, R.W. (1998) Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. *J Gen Virol*, **79** (Pt 12), 2909-2919.
- Iseni, F., Baudin, F., Blondel, D. and Ruigrok, R.W. (2000) Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *Rna*, **6**, 270-281.
- Iverson, L.E. and Rose, J.K. (1981) Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. *Cell*, **23**, 477-484.
- Jacob, Y., Badrane, H., Ceccaldi, P.E. and Tordo, N. (2000) Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein. *J Virol*, **74**, 10217-10222.
- Jeng, T.W., Crowther, R.A., Stubbs, G. and Chiu, W. (1989) Visualization of alpha-helices in tobacco mosaic virus by cryo-electron microscopy. *J Mol Biol*, **205**, 251-257.
- Kabsch, W. (1993) Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *Journal of Applied Crystallography*, **26**, 795-800.
- Kawai, A., Toriumi, H., Tochikura, T.S., Takahashi, T., Honda, Y. and Morimoto, K. (1999) Nucleocapsid formation and/or subsequent conformational change of rabies virus nucleoprotein (N) is a prerequisite step for acquiring the phosphatase-sensitive epitope of monoclonal antibody 5-2-26. *Virology*, **263**, 395-407.
- Kawashima, T., Berthet-Colominas, C., Cusack, S. and Leberman, R. (1996) Interconversion of crystals of the Escherichia coli EF-Tu.EF-Ts complex between high- and low-diffraction forms. *Acta Crystallogr D Biol Crystallogr*, **52**, 799-805.
- Kleywegt, G.J. and Jones, T.A. (1994) Halloween ... masks and bones. "From First Map to Final Model" , *SERC Daresbury Laboratory*, pp. 59-66.
- Kohno, T., Goto, T., Takasaki, T., Morita, C., Nakaya, T., Ikuta, K., Kurane, I., Sano, K. and Nakai, M. (1999) Fine structure and morphogenesis of Borna disease virus. *J Virol*, **73**, 760-766.
- Kolakofsky, D., Le Mercier, P., Iseni, F. and Garcin, D. (2004) Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, **318**, 463-473.
- Kouznetzoff, A., Buckle, M. and 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** (Pt 5), 1005-1013.
- Kozak, M. (1989) The scanning model for translation: an update. *J Cell Biol*, **108**, 229-241.
- Kurath, G. and Leong, J.C. (1985) Characterization of infectious hematopoietic necrosis virus mRNA species reveals a nonvirion rhabdovirus protein. *J Virol*, **53**, 462-468.
- Kurilla, M.G., Cabradilla, C.D., Holloway, B.P. and Keene, J.D. (1984) Nucleotide sequence and host La protein interactions of rabies virus leader RNA. *J Virol*, **50**, 773-778.
- Kurilla, M.G., Piwnica-Worms, H. and Keene, J.D. (1982) Rapid and transient localization of the leader RNA of vesicular stomatitis virus in the nuclei of infected cells. *Proc Natl Acad Sci U S A*, **79**, 5240-5244.
- Lentz, T.L., Hawrot, E. and Wilson, P.T. (1987) Synthetic peptides corresponding to sequences of snake venom neurotoxins and rabies virus glycoprotein bind to the nicotinic acetylcholine receptor. *Proteins*, **2**, 298-307.

- Leslie, A.G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr*, **62**, 48-57.
- Lewandowski, L.J., Kalmakoff, J. and Tanada, Y. (1969) Characterization of a Ribonucleic Acid Polymerase Activity Associated with Purified Cytoplasmic Polyhedrosis Virus of the Silkworm *Bombyx mori*. *J Virol*, **4**, 857-865.
- Longhi, S., Receveur-Brechot, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D., Yeo, R., Finet, S. and Canard, B. (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.
- Lopez, C.B., Yount, J.S. and Moran, T.M. (2006) Toll-like receptor-independent triggering of dendritic cell maturation by viruses. *J Virol*, **80**, 3128-3134.
- Madhusudana, S.N., Nagaraj, D., Uday, M., Ratnavalli, E. and Kumar, M.V. (2002) Partial recovery from rabies in a six-year-old girl. *Int J Infect Dis*, **6**, 85-86.
- Martin-Benito, J., Area, E., Ortega, J., Llorca, O., Valpuesta, J.M., Carrascosa, J.L. and Ortin, J. (2001) Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO Rep*, **2**, 313-317.
- Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.
- Mavrakis, M., Iseni, F., Mazza, C., Schoehn, G., Ebel, C., Gentzel, M., Franz, T. and 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., Kolesnikova, L., Schoehn, G., Becker, S. and Ruigrok, R.W. (2002) Morphology of Marburg virus NP-RNA. *Virology*, **296**, 300-307.
- Mavrakis, M., McCarthy, A.A., Roche, S., Blondel, D. and 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., Mehoulas, S., Real, E., Iseni, F., Blondel, D., Tordo, N. and Ruigrok, R.W. (2006) Rabies virus chaperone: Identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology*.
- Mazza, C., Ohno, M., Segref, A., Mattaj, I.W. and Cusack, S. (2001) Crystal structure of the human nuclear cap binding complex. *Mol Cell*, **8**, 383-396.
- McCree, B.J., Jr., McKinnon, K.P. and Lyles, D.S. (1990) Solubility of vesicular stomatitis virus M protein in the cytosol of infected cells or isolated from virions. *J Virol*, **64**, 902-906.
- McPherson, A. (2001) A comparison of salts for the crystallization of macromolecules. *Protein Sci*, **10**, 418-422.
- Morris, R.J., Zwart, P.H., Cohen, S., Fernandez, F.J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A. and Lamzin, V.S. (2004) Breaking good resolutions with ARP/wARP. *J Synchrotron Radiat*, **11**, 56-59.
- Moyer, S.A., Abraham, G., Adler, R. and Banerjee, A.K. (1975) Methylated and blocked 5' termini in vesicular stomatitis virus in vivo mRNAs. *Cell*, **5**, 59-67.
- Neumann, G., Castrucci, M.R. and Kawaoka, Y. (1997) Nuclear import and export of influenza virus nucleoprotein. *J Virol*, **71**, 9690-9700.
- Newcomb, W.W., Tobin, G.J., McGowan, J.J. and Brown, J.C. (1982) In vitro reassembly of vesicular stomatitis virus skeletons. *J Virol*, **41**, 1055-1062.
- Pal, R., Barenholz, Y. and Wagner, R.R. (1987) Vesicular stomatitis virus membrane proteins and their interactions with lipid bilayers. *Biochim Biophys Acta*, **906**, 175-193.
- Petsko, G.A. (1985) Preparation of isomorphous heavy-atom derivatives. *Methods Enzymol*, **114**, 147-156.

- Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *Embo J*, **8**, 3867-3874.
- Poisson, N., Real, E., Gaudin, Y., Vaney, M.C., King, S., Jacob, Y., Tordo, N. and Blondel, D. (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**, 2691-2696.
- Porras, C., Barboza, J.J., Fuenzalida, E., Adaros, H.L., Oviedo, A.M. and Furst, J. (1976) Recovery from rabies in man. *Ann Intern Med*, **85**, 44-48.
- Roche, S., Bressanelli, S., Rey, F.A. and Gaudin, Y. (2006) Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science*, **313**, 187-191.
- Rose, J.K. (1980) Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. *Cell*, **19**, 415-421.
- Rubinson, K.A., Ladner, J.E., Tordova, M. and Gilliland, G.L. (2000) Cryosalts: suppression of ice formation in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **56 (Pt 8)**, 996-1001.
- Rudolph, M.G., Kraus, I., Dickmanns, A., Eickmann, M., Garten, W. and Ficner, R. (2003) Crystal structure of the borna disease virus nucleoprotein. *Structure (Camb)*, **11**, 1219-1226.
- Sanchez, A., Kiley, M.P., Klenk, H.D. and Feldmann, H. (1992) Sequence analysis of the Marburg virus nucleoprotein gene: comparison to Ebola virus and other non-segmented negative-strand RNA viruses. *J Gen Virol*, **73 (Pt 2)**, 347-357.
- Schick, B. and Jurnak, F. (1994) Extension of the diffraction resolution of crystals. *Acta Crystallogr D Biol Crystallogr*, **50**, 563-568.
- Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D. and Ruigrok, R.W. (2001) Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J Virol*, **75**, 490-498.
- Schoehn, G., Mavrakis, M., Albertini, A., Wade, R., Hoenger, A. and Ruigrok, R.W. (2004) The 12 Å structure of trypsin-treated measles virus N-RNA. *J Mol Biol*, **339**, 301-312.
- Schwemmler, M., Salvatore, M., Shi, L., Richt, J., Lee, C.H. and Lipkin, W.I. (1998) Interactions of the borna disease virus P, N, and X proteins and their functional implications. *J Biol Chem*, **273**, 9007-9012.
- Sha, B. and Luo, M. (1997) Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1. *Nat Struct Biol*, **4**, 239-244.
- Sheldrick, G., and Schneider, T. (1997) High-resolution Structure Refinement. *Methods Enzymol.*, **277**, 319-343.
- Stricker, R., Mottet, G. and Roux, L. (1994) The Sendai virus matrix protein appears to be recruited in the cytoplasm by the viral nucleocapsid to function in viral assembly and budding. *J Gen Virol*, **75 (Pt 5)**, 1031-1042.
- Thomas, D., Newcomb, W.W., Brown, J.C., Wall, J.S., Hainfeld, J.F., Trus, B.L. and Steven, A.C. (1985) Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. *J Virol*, **54**, 598-607.
- Thoulouze, M.I., Bouguyon, E., Carpentier, C. and Bremont, M. (2004) Essential role of the NV protein of Novirhabdovirus for pathogenicity in rainbow trout. *J Virol*, **78**, 4098-4107.
- Thoulouze, M.I., Lafage, M., Schachner, M., Hartmann, U., Cremer, H. and Lafon, M. (1998) The neural cell adhesion molecule is a receptor for rabies virus. *J Virol*, **72**, 7181-7190.
- Tordo, N., Poch, O., Ermine, A. and Keith, G. (1986a) Primary structure of leader RNA and nucleoprotein genes of the rabies genome: segmented homology with VSV. *Nucleic Acids Res*, **14**, 2671-2683.

- Tordo, N., Poch, O., Ermine, A., Keith, G. and Rougeon, F. (1986b) Walking along the rabies genome: is the large G-L intergenic region a remnant gene? *Proc Natl Acad Sci U S A*, **83**, 3914-3918.
- Toriumi, H. and 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.
- Tuffereau, C., Benejean, J., Alfonso, A.M., Flamand, A. and Fishman, M.C. (1998a) Neuronal cell surface molecules mediate specific binding to rabies virus glycoprotein expressed by a recombinant baculovirus on the surfaces of lepidopteran cells. *J Virol*, **72**, 1085-1091.
- Tuffereau, C., Benejean, J., Blondel, D., Kieffer, B. and Flamand, A. (1998b) Low-affinity nerve-growth factor receptor (P75NTR) can serve as a receptor for rabies virus. *Embo J*, **17**, 7250-7259.
- Vagin, A. and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr*, **56 Pt 12**, 1622-1624.
- Vidy, A., Chelbi-Alix, M. and Blondel, D. (2005) Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways. *J Virol*, **79**, 14411-14420.
- Vulliemoz, D. and Roux, L. (2001) "Rule of six": how does the Sendai virus RNA polymerase keep count? *J Virol*, **75**, 4506-4518.
- Warrell, M.J. and Warrell, D.A. (2004) Rabies and other lyssavirus diseases. *Lancet*, **363**, 959-969.
- Wertz, G.W., Moudy, R. and Ball, L.A. (2002) Adding genes to the RNA genome of vesicular stomatitis virus: positional effects on stability of expression. *J Virol*, **76**, 7642-7650.
- Wertz, G.W., Perepelitsa, V.P. and Ball, L.A. (1998) Gene rearrangement attenuates expression and lethality of a nonsegmented negative strand RNA virus. *Proc Natl Acad Sci U S A*, **95**, 3501-3506.
- Whelan, S.P. and Wertz, G.W. (1999) The 5' terminal trailer region of vesicular stomatitis virus contains a position-dependent cis-acting signal for assembly of RNA into infectious particles. *J Virol*, **73**, 307-315.
- Whelan, S.P. and 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.
- Wyatt, P. (1993) Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta*, **272** 1-40.
- Yang, J., Hooper, D.C., Wunner, W.H., Koprowski, H., Dietzschold, B. and Fu, Z.F. (1998) The specificity of rabies virus RNA encapsidation by nucleoprotein. *Virology*, **242**, 107-117.
- Yang, J., Koprowski, H., Dietzschold, B. and Fu, Z.F. (1999) Phosphorylation of rabies virus nucleoprotein regulates viral RNA transcription and replication by modulating leader RNA encapsidation. *J Virol*, **73**, 1661-1664.

Articles supplémentaires

The 12 Å Structure of Trypsin-treated Measles Virus N-RNA

Schoehn, G., Mavrakis, M., Albertini, A., Wade, R., Hoenger, A. and Ruigrok, R.W. (2004)
The 12 Å structure of trypsin-treated measles virus N-RNA. *J Mol Biol*, 339, 301-312.

Le virus de la rougeole est un *Mononegavirales*, de la famille des paramyxovirus. Lorsque la nucléoprotéine du virus de la rougeole est exprimée dans des cellules d'insectes, elle lie les ARN cellulaires et forme des nucléocapsides recombinantes. Ces nucléocapsides ont une structure hélicoïdale. Ces complexes N-ARN sont assez flexibles, et ne sont pas assez rigides pour réaliser une reconstruction par cryomicroscopie électronique. En digérant la partie C-terminale de la nucléoprotéine avec de la trypsine, les complexes N-ARN deviennent plus réguliers et plus rigides. L'analyse de ces nucléocapsides révèle qu'elles sont en fait composées de plusieurs états hélicoïdaux. Deux reconstructions hélicoïdales ont été obtenues pour ces nucléocapsides avec des paramètres hélicoïdaux différents représentant l'état majoritaire de ces nucléocapsides. Sur la reconstruction, il est clairement visible que la nucléoprotéine est composée de trois domaines. Les protomères de nucléoprotéine dans un même tour d'hélice sont connectés entre eux par deux points. Ces zones d'interactions sont situées près de l'axe hélicoïdal et au milieu du protomère. En revanche, il n'y a pas de contact visible entre les protomères de deux tours d'hélice successifs.

L'ARN des nucléocapsides a été marqué avec du cis-platine. Le cis-platine est un composé qui lie de manière covalente les acides nucléiques. Sur la reconstruction hélicoïdale de nucléocapsides marquées au cis-platine, la densité de la zone de connexion proche de l'axe hélicoïdale est plus importante. Ceci a permis de faire une hypothèse quant à la localisation de l'ARN au sein de la nucléocapside.

The 12 Å Structure of Trypsin-treated Measles Virus N–RNA

Guy Schoehn^{1,2,3*}, Manos Mavrikis¹, Aurélie Albertini^{1,3}
Richard Wade², Andreas Hoenger⁴ and Rob W. H. Ruigrok^{1,3}

¹EMBL Grenoble Outstation
6 rue Jules Horowitz, B.P. 181
38042 Grenoble Cedex 9
France

²Institut de Biologie Structurale
UMR 5075 CEA-CNRS-UJF
41 rue Jules Horowitz, 38027
Grenoble Cedex 1, France

³Laboratoire de Virologie
Moléculaire et Structurale, EA
2939, Université Joseph Fourier
Grenoble, France

⁴EMBL, Structure Programme
Meyerhofstrasse 1, 69117
Heidelberg, Germany

Recombinant measles virus nucleoprotein (N) was produced in insect cells where it bound to cellular RNA to form helical N–RNA structures. These structures were observed by electron microscopy but were too flexible for high-resolution image analysis. Removal of the C-terminal tail of N by trypsin treatment resulted in structures that were much more rigid and seemed more regular. Several methods of image analysis were employed in order to make a helical reconstruction of the digested N–RNA. During this analysis, it became clear that the apparently regular coils of digested N–RNA consisted of a series of closely related helical states. The iterative helical real space reconstruction method allowed the identification of two helical states for which a reconstruction could be calculated. The model with the highest resolution shows N monomers that consist of three domains and that are connected to their neighbours by two narrow connections, one close to the helical axis and another toward the middle of the monomers. There are no connections between N molecules in subsequent helical turns. After labelling the RNA in the structure with *cis*-platinum, the connection closest to the helical axis increased in density, suggesting the position of the RNA. The shapes of the monomers of the nucleoproteins of influenza virus, rabies virus (both determined before) and that of measles virus (determined here) are all similar, whereas the overall shapes of their respective N–RNAs (nucleocapsids) is very different. This is probably due to the position and number of the connections between the N subunits in the N–RNA, one for influenza virus allowing much flexibility, two for rabies virus at either end of the N molecules leading to ribbons and two for measles virus leading to the typical paramyxovirus helical nucleocapsid.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: measles virus; replication; electron cryo-microscopy; nucleocapsid; helical reconstruction

*Corresponding author

Introduction

Paramyxoviruses are negative-strand RNA viruses. This means that the viral RNA is in the sense opposite to that of mRNA and cannot be translated directly into protein, as is the case for the RNA of positive-strand RNA viruses. The

RNAs of positive-strand RNA viruses have specific three-dimensional (3D) structures at their 5' ends to which cellular translation initiation factors can bind to initiate the production of viral proteins. One of the viral proteins that are produced in this way is the RNA-dependent RNA polymerase necessary for replication of the viral RNA (vRNA). The RNA of negative-strand viruses first needs to be transcribed into mRNA before viral proteins can be made. These viruses do not introduce naked vRNA into the infected cell but a nucleocapsid that consists of vRNA that is stoichiometrically bound to a protein (N), plus a viral RNA-dependent RNA polymerase complex. This polymerase complex consists of the large protein

Abbreviations used: vRNA, viral RNA; RSV, respiratory syncytial virus; IHRSR, iterative helical real space reconstruction method; AcNPV, *Autographa californica* nuclear polyhedrosis virus; EM, electron microscope; CTF, contrast transfer function.

E-mail address of the corresponding author: schoehn@embl-grenoble.fr

(L) that contains the enzymatic activity and a polymerase cofactor, the phosphoprotein (P), that binds L to the N-vRNA complex. Naked vRNA cannot serve as a template for the polymerase complex.¹ For paramyxoviruses such as Sendai virus (a mouse virus that causes pneumonia) and measles virus (a human virus that causes about one million deaths per year in the third world), each N monomer binds exactly six ribonucleotides.²⁻⁴ N binds irreversibly to the phosphate sugar backbone of the RNA and, in doing so, exposes the nucleotide bases so that these can be read by the polymerase.^{5,6}

Different paramyxovirus N proteins have been expressed in various expression systems, such as in chick embryo fibroblasts using a vaccinia vector,⁷ in human fibroblasts using a recombinant adenovirus vector,⁸ in insect cells using a baculovirus vector⁹ and in *Escherichia coli*.¹⁰ In all these systems N binds to different sized cellular RNAs and forms helical N-RNA polymers with the same morphology as viral nucleocapsids. Optical density spectra of recombinant measles virus N structures show that these complexes indeed contain RNA.¹¹

Nucleocapsids of paramyxoviruses are helical structures with a characteristic herringbone appearance¹² that is often used for the identification of viruses in this family. The helical parameters have been studied by negative stain and metal shadowing electron microscopy (EM).^{2,9,13} For Sendai virus and measles virus there are about 13 subunits of N per helical turn but the helical pitch seems to be variable.² Paramyxovirus nucleoproteins (58–60 kDa) consist of a conserved N-terminal domain that makes up three-quarters of the protein and a non-conserved, acidic C-terminal domain. The N-terminal domain is essential for forming the helical N-RNA structures^{14,15} and for respiratory syncytial virus (RSV) it was shown that the N-terminal 92 amino acid residues were sufficient for the formation of nucleocapsid-like structures.¹⁰ The C-terminal domain of rabies viruses and measles virus N is the binding site for the phosphoprotein, the polymerase cofactor.^{16,17} When the C-terminal domain of N is expressed alone it appears to be unfolded.

However, this domain seems to fold in the presence of the N-terminal domain of N and in the presence of the C-terminal domain of the phosphoprotein that binds to N.^{17,18} The presence of the charged C-terminal domain of N seems to be connected with the loose coiling of the nucleocapsids and the observed variation in helical parameters, since removal of this domain by trypsin¹⁹ or preparation of nucleocapsids in high salt²⁰ stiffens the structure and makes it more regular.

Here, we have produced measles virus nucleoprotein (58 kDa) in insect cells where it forms N-RNA complexes with cellular RNA. We describe the 3D EM reconstruction of recombinant measles virus N-RNA after trypsin treatment that resulted in more or less regular helical structures with an N-terminal domain of the nucleoprotein of 43 kDa. In the N-RNA, two density bridges, close to the helical axis of the structure, connect the subunits. In a second reconstruction where the RNA was labelled with *cis*-platinum, a density increase was observed for the continuous density closest to the helical axis, suggesting that this is the position for the RNA. We compared the shapes of the nucleoproteins of measles, rabies and influenza virus and found that these shapes are similar apart from differences in inter-subunit contacts that determine the helical characteristics of the different N-RNA structures.

Results

Recombinant measles virus N-RNA was produced in insect cells and purified over a CsCl gradient. PAGE analysis of the material that banded in the CsCl gradient showed a single protein component with a molecular mass corresponding to that of measles N (58 kDa; [Figure 1a, lane 1](#)). The material contained nucleic acid, since the ratio of the absorbance at 260 nm/280 nm was 1.17 (not shown).¹¹ Observation of the purified material by negative stain and electron cryomicroscopy (cryo-EM) showed the characteristic herringbone structure of a helical paramyxovirus nucleocapsid ([Figure 1b, intact](#)). However, the

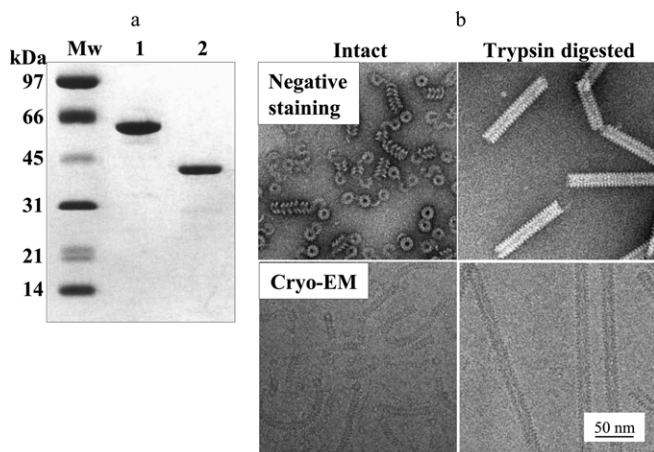


Figure 1. Intact and trypsin-digested recombinant measles virus N-RNA. (a) Lane 1, SDS-PAGE of intact, purified N-RNA and lane 2, trypsin digested N-RNA. (b) Electron micrographs of negatively stained (top images) and frozen hydrated (bottom images) preparations of intact (left) and trypsin-digested (right) N-RNA.

N–RNA complexes were of various, rather short lengths and flexible, and the images as such could not be used for image analysis. Therefore, we decided to treat the N–RNA with trypsin, which removed the first 13 amino acid residues and the last 120 amino acid residues, leaving a 43.5 kDa protease-resistant fragment, residues 14–405 (see Figure 1a, lane 2, and Materials and Methods). SDS-PAGE of digested N–RNA that had been sedimented through a glycerol cushion showed that the 15 kDa fragment was cleaved and removed from the remaining helical N–RNA complex (Figure 1a and data not shown). Negative stain and cryo-EM of this material showed very regular and straight complexes (Figure 1b, trypsin digested). However, the helices were much longer than those of intact material, suggesting that several short segments have locked together into longer helices.

Subsequently, we prepared the digested N–RNA complexes by high-resolution metal shadowing for EM in order to determine the handedness of the coil (Figure 2). This method revealed straight and tight coils but also some unwound stretches of N–RNA (Figure 2a, arrow). The measles virus N–RNA forms a left-handed helix, like that of Sendai virus.² From image analysis of the particles it appeared that there were at least two helical forms present and sometimes the two forms could be found in the same particle (not shown). This happens either because the helix within a single, contiguous N–RNA structure can adopt different contacts and form helices with different parameters or because during the locking-on process of shorter helices to form the long helices observed after trypsin treatment, short helices with different parameters can still associate to form longer structures. One of the differences between types I and II (except for the pitch) is the orientation of the line that connects subunits in subsequent turns; see Figure 2b.

For the following image analyses we have used frozen hydrated samples. First, we attempted a helical analysis of isolated, straight particles. Twenty helices were selected from the negatives using Ximdisp,²¹ and were rotated in-plane so that the helical axis was parallel with the vertical axis.

After this the helices were masked in order to reduce the noise (Figure 3(a)). These images were Fourier transformed using SUPPRIM²² (Figure 3(b)). For all Fourier transforms, the phases were checked along all layer-lines in order to see if these layer-lines had odd or even orders of a Bessel function. Only one particle was regular enough to assign some of the layer-lines and, for this particle, we calculated a 3D reconstruction using the Fourier Bessel algorithm (Figure 3(c)).²³ The resulting model showed a helix with a pitch of 50 Å and with 12.3 subunits per helical turn. However, the power spectrum for this particle (Figure 3(b)) was strongly asymmetric with higher resolution diffraction coming from the long direction of the particle. This could be due to the fact that in the particle that was analysed, several shorter fragments were associated with the same pitch but with a break in the helical symmetry at the contact points. Since the helical analysis did not lead to a high-resolution reconstruction (around 25 Å), we decided to use hybrid methods combining helical and single-particle analysis.

For the first hybrid approach, 3746 boxes sections that were corrected for the contrast transfer function (CTF) (Materials and Methods) were selected from six different micrographs. We used square boxes of 128 × 128 pixels after interpolation, in order to obtain short helical fragments that, we hoped, contained only a single helical structure. In order to increase the chance of obtaining a single helical structure we selected the particles with an overlap of half a box (in a long helix the centre of the boxes was chosen to be at approximately every tenth repeat). As a starting model for the image alignment and search for image classes we used the helical reconstruction described above (Figure 3(c)). This model was reprojected in 50 views (Figure 3(d); 29 side views rotated by 1° steps, 14 views tilted by 2° increments and another seven further tilted by 4° increments). All 3746 phase flipped images were aligned against the 50 reprojections using a projection matching method¹⁶ (Figure 3). The corresponding deconvoluted images were then averaged into class averages (Figure 3(e)). The 29 side views were Fourier transformed (Figure 3(f)) and analysed using

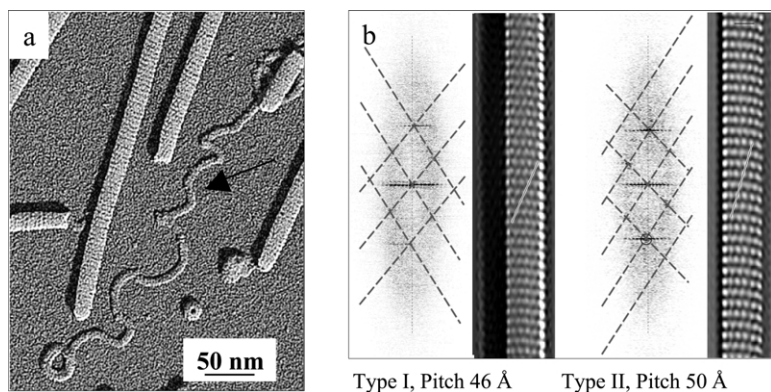


Figure 2. Electron micrographs of metal shadowed, trypsin-digested N–RNA. a, Field of tightly coiled and unwound (arrow) N–RNA. b, Image analysis on two independent coils present on the same sample grid as the picture shown in a and indicating the presence of at least two different helical structures. Both structures consist of a left-handed helix. Lines on the filtered model show the orientation of the line that connects subunits in subsequent turns.

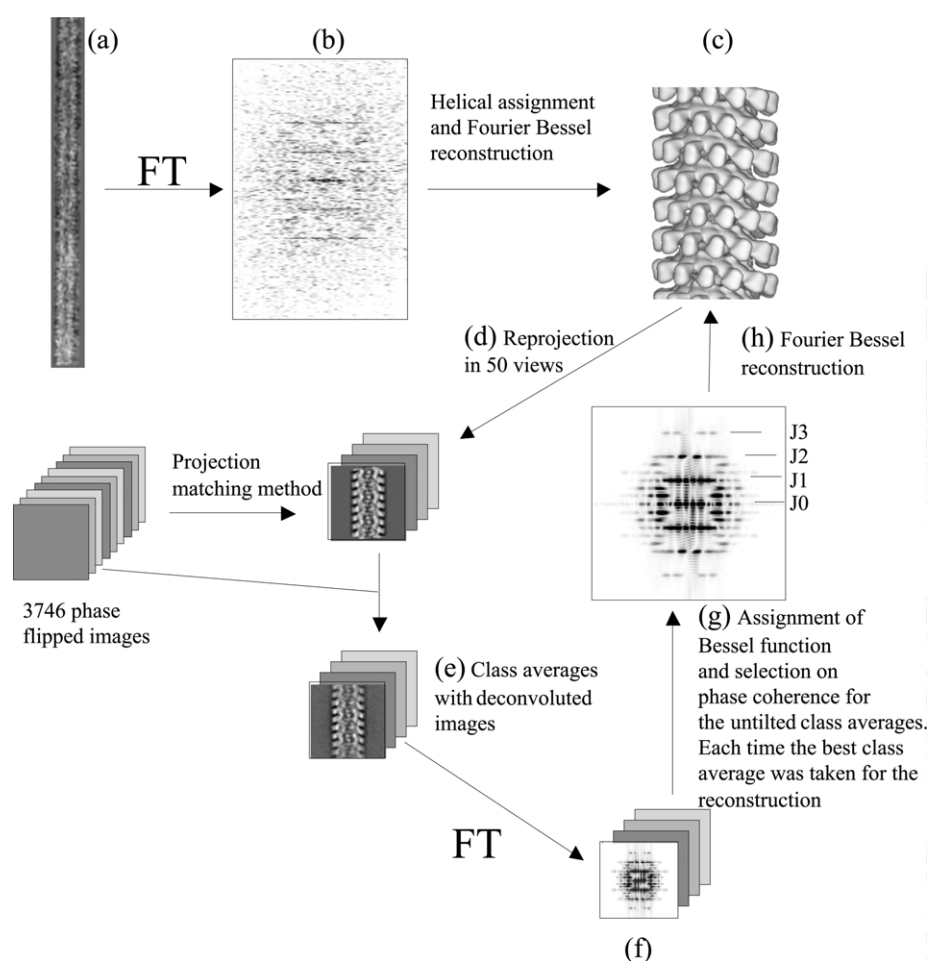


Figure 3. Image analysis using helical symmetry or combined helical and single-particle analysis on electron micrographs of frozen hydrated, trypsin-digested N-RNA. (a)–(c) Helical analysis was attempted on 20 straight and long particles but only one particle (a) gave an interpretable diffraction image (b) from which a helical reconstruction could be calculated (c). (c)–(h) Hybrid image analysis was started with the reprojection of the helical reconstruction (c) and projection matching of 3746 phase flipped, selected images. The images that were averaged into classes were used as deconvoluted images (e). The Fourier transforms of these class averages (f) were submitted to helical analysis and the best transform (g) was used for the reconstruction of the model (h) for the next cycle of analysis.

SUPPRIM.²² It was then possible to assign the diffraction patterns from some of the class averages (J0, J1, J2, J3, J11, J12, J13 and J14, only J0–J3 shown in Figure 3(g) for clarity reasons). The phases and the positions of the layer-lines were consistent with the chosen selection rule (37/3). The best class average (average of 250 raw images) was then used to calculate a new 3D reconstruction using the Fourier Bessel algorithm (Figure 3(h)). This new 3D model was then used in a new cycle of the projection matching method. After eight cycles of this process the resulting model was stable, with a pitch of 50.0 Å and 12.33 subunits per turn. The resolution along the helical axis was about 17 Å but only 25 Å in the other direction. This suggested that this method did not lead to separation of images with different helical parameters and, therefore, this approach cannot lead to a high-resolution model.

Subsequently, we used a modified version of the iterative helical real space reconstruction method

(IHRSR).²⁴ We used the phase-flipped images to do the alignment plus parameter search and the parameters that were determined were applied to the deconvoluted images that were selected in the alignment procedure. This method does allow the separation of different classes of helical particles. We used different helical models and starting parameters for the analysis procedure. Starting model A: the helical model derived above with a pitch of 50.0 Å and 12.33 subunits per turn. Starting model B: a continuous helix with a pitch of 50.0 Å and constrained initially to 14 subunits per turn. Starting model C: a continuous helix with the same pitch but initially constrained to ten subunits per turn (Figure 4). After 120 cycles of the IHRSR procedure, starting models A and B converged to the same solution with a pitch of 50.8 Å and 12.35 subunits per turn (reconstruction 1). Starting model C led to a solution with a pitch of 50.7 Å and 11.64 Å subunits per turn (reconstruction 2). In the final reconstruction, only 460 images (12% of starting

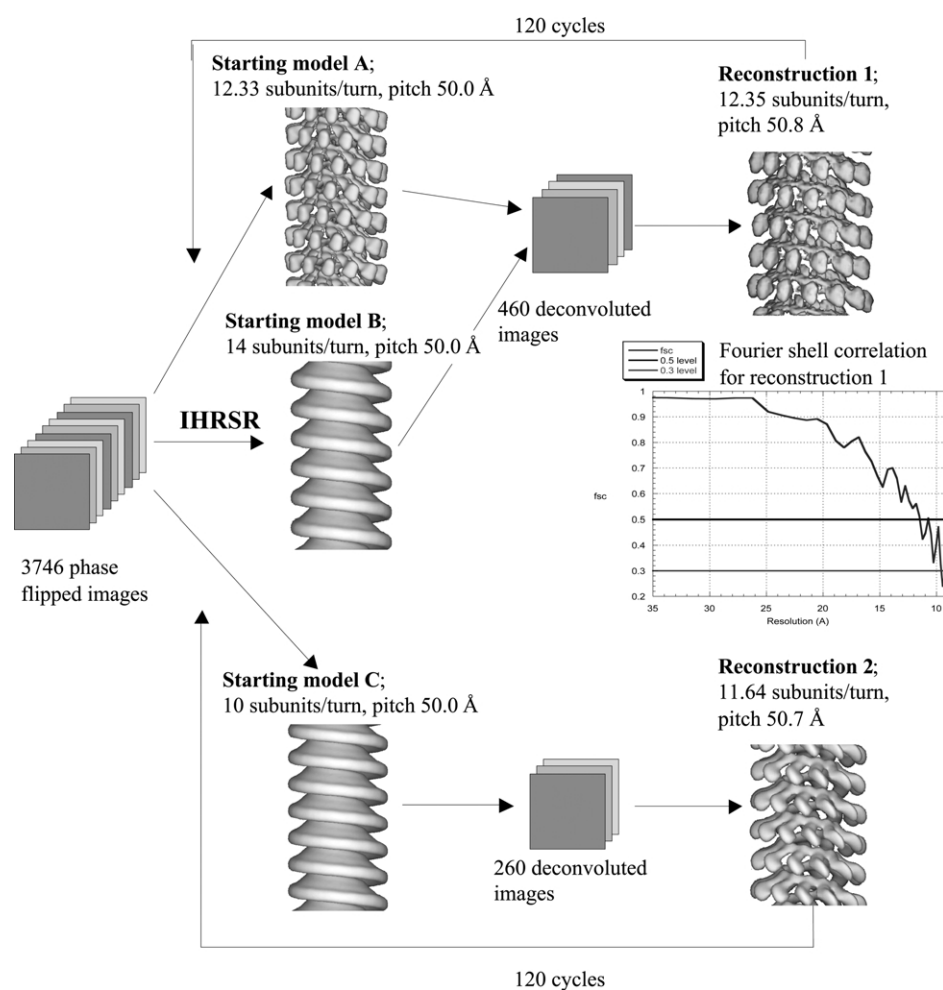


Figure 4. Iterative helical real space reconstruction on electron micrographs of frozen hydrated, trypsin-digested N-RNA. The same 3746 phase flipped images were used as for the analysis in Figure 3. The images were aligned according to the method described by Egelman²⁴ against three starting models with parameters as indicated. Models A and B independently gave the same solution (reconstruction 1), whereas starting model C led to another solution, reconstruction 2.

pictures) were retained for reconstruction 1 and 260 (7%) for reconstruction 2. Reconstruction 1 had a resolution of 11.8 Å and reconstruction 2 had a resolution of 25 Å (see Materials and Methods). The main differences between the two structures are the resolution, the number of subunits per turn and the fact that in reconstruction 2 the terminal lobe of density seems to be tilted compared to reconstruction 1 (see details in Figure 6). Another difference between the two structures is the orientation of the right-handed 12-start helix (lines in Figure 6). When we compare these orientations with the corresponding ones in the two models obtained by shadowing (Figure 2), it is clear that reconstruction 1 corresponds to type II. When we used the 260 images that were selected for reconstruction 2 and we recalculated a model but now with reconstruction 1 as starting model, we again got reconstruction 2 as a result. We do not have an interpretation for the differences in shape between the two reconstructions and we also do not know why reconstruction 2 is so bad

compared to reconstruction 1 (the small number of images selected for this model will have some influence on the resolution) but all our results are consistent with the presence of at least two different helices. Figure 5 shows a higher magnification image of the structure of reconstruction 1, both from the outside of the coil (Figure 5a) and when the front half of the helix is peeled away so that the inner surface is shown (Figure 5b). Figure 5c shows the resulting power spectrum for the final model. In contrast to the power spectrum shown in Figure 3, diffraction is now isotropic. Figure 5d–f show a top view, a front view and a view from the underside of a number of subunits. The subunits are attached to each other by two connections, one near the middle of the subunits and one towards the end of the subunits closest to the helical axis. There are no contacts between subunits in subsequent turns of the helix. The monomer appears to consist of three domains: a globular domain closest to the helical axis, then a rather flat and elongated domain at the middle of

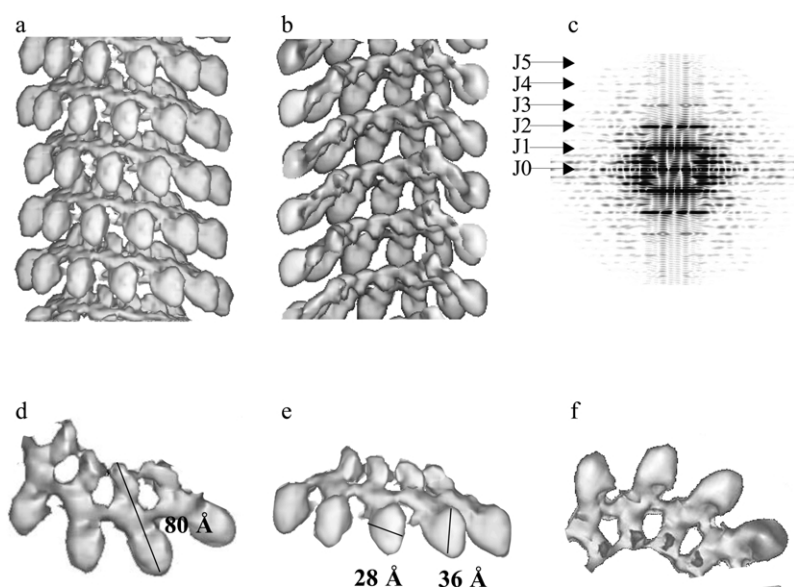


Figure 5. Details of reconstruction 1. a, Front view of the helix, the back part of the helix has been removed to make it clearer; (b) view of the inside of the helix with the front half cut away; and (c) power spectrum of the reprojection of reconstruction 1 showing equal details in all directions. d–f, Top, front and under-side views of a single helical turn showing the three domains of the monomer, a globular domain towards the helical axis, a flat connector domain and the terminal and largest domain. Each monomer is connected at two positions with its neighbours. The length of the N subunit and the width and the height of the terminal domain are indicated in d and e.

the structure that is connected through a thin neck to the largest, distal domain. The N subunits are 80 Å long and the distal domain is 36 Å high and 28 Å wide (Figure 5d and e).

As mentioned above, each subunit binds six ribonucleotides and somewhere in the structure the continuous RNA molecule must constitute a continuous density. From the results in Figure 5 it is not clear which of the densities connecting the adjacent subunits contains the RNA. Therefore, we decided to treat the trypsin-treated N–RNA with *cis*-platinum that binds covalently to the N7 positions of adjacent purines^{25–27} but is not known to bind to protein. *cis*-Platinum-treated N–RNA was purified over a glycerol cushion, dialysed and prepared for cryo-EM. More than 5000 images were selected and analysed with the IHRSR

method as described for untreated N–RNA. The starting model was reconstruction 1 for the untreated complex. After 100 cycles we obtained a stable solution from 700 images (14%) with a pitch of 48.1 Å and 12.56 subunits per turn and with a resolution of 15 Å (reconstruction 3). Figure 6 shows the three reconstructions for the untreated and *cis*-Pt treated complexes.

Reconstruction 1 and 3 differ in the helical pitch but the morphologies of the subunits are similar. The main difference between the subunits is that the open space between the two densities that connect the subunits in reconstruction 1 is filled up in reconstruction 3 (Figure 7). This indicates that the RNA is present in this region but does not clearly implicate either of the two connecting densities. The contour level that we used for the

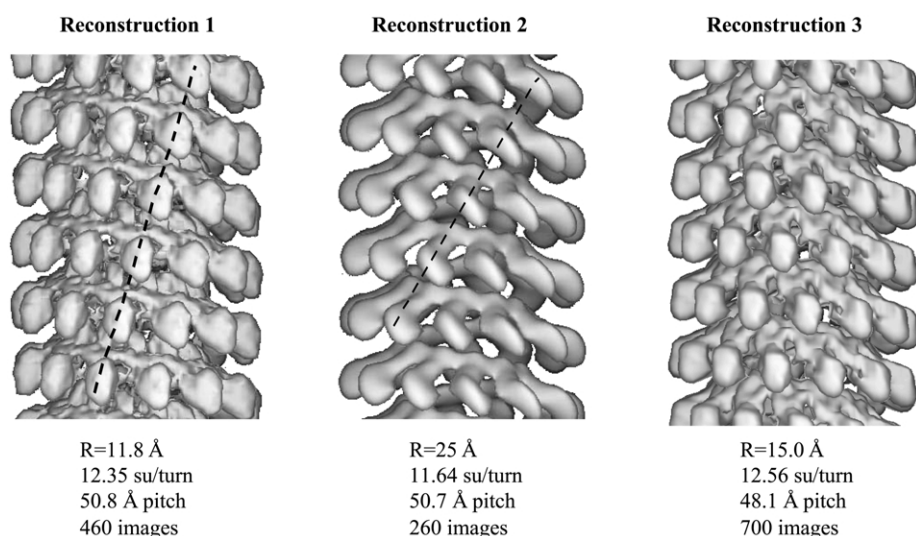


Figure 6. Comparison of reconstructions 1, 2 and 3. Reconstructions 1 and 2 are from trypsin-digested N–RNA and reconstruction 3 is from trypsin-digested N–RNA that was subsequently treated with *cis*-platinum. For reconstructions 1 and 2, the black broken lines show the orientation of the line that connects subunits in subsequent turns.

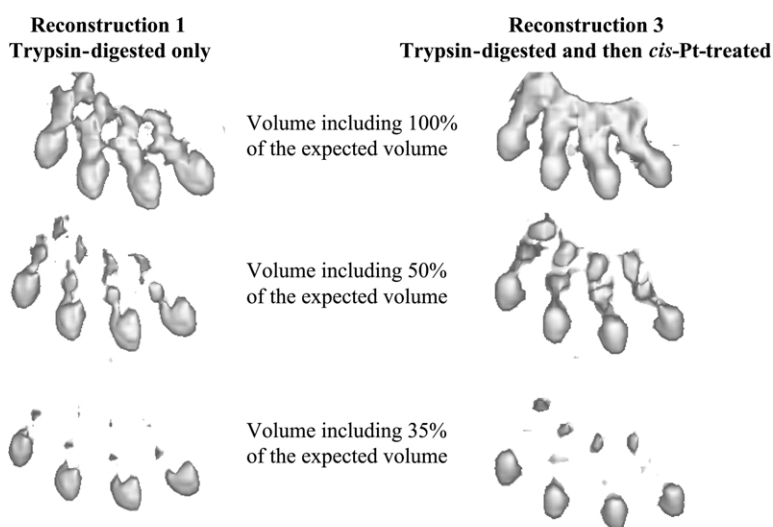


Figure 7. Lowering the cut-off volume indicates the increase in density caused by the *cis*-platinum treatment. Subunits in reconstructions 1 and 3 at 100%, 50% and 35% of their volume. Whereas in reconstruction 1 only the end domain remains visible at the 35% volume level, in reconstruction 3 the domain closest to the helical axis also remains visible, suggesting that *cis*-platinum attachment occurred preferentially to this domain.

reconstructions was $0.84 \text{ Da}/\text{\AA}^3$ for an N subunit of 43 kDa, i.e. $50,000 \text{ \AA}^3$. When the volume is reduced to only 50 or even 35% of the original volume, only the parts of the reconstruction with the highest density remain visible. Whereas in the untreated N-RNA the only density that is still present when the volume is reduced to 35% is the distal domain, in the *cis*-Pt-treated sample there remains an area of density closer to the helical axis, indicating that *cis*-Pt attaches to the domain closest to the helical axis and suggesting that this is the region where the RNA is bound. The centre-to-centre distance between the high-density domains nearest the helical axis is about 26 Å. A stretched-out nucleotide (ribose plus phosphate) is about 6.5 Å long and there are 6 nt per N

monomer, making a maximum possible distance of 39 Å between N monomers at the position of the RNA. This calculation indicates that the RNA must be near the helical axis.

Reconstruction 1 has a resolution of better than 12 Å although the surface of the helical structure does not seem to show this detail. This is probably due to the fact that the low resolution totally masks out the high resolution (Figure 8a). To correct this, we applied a Gaussian high-pass filter limited to 40 Å.²⁸ The corresponding power spectrum became more equal in spot intensity (Figure 8b) and the 12 Å details in the structure were reinforced (Figure 8c and d). After correction, the three domains appear more distinct and it can be seen that the connection between the second,

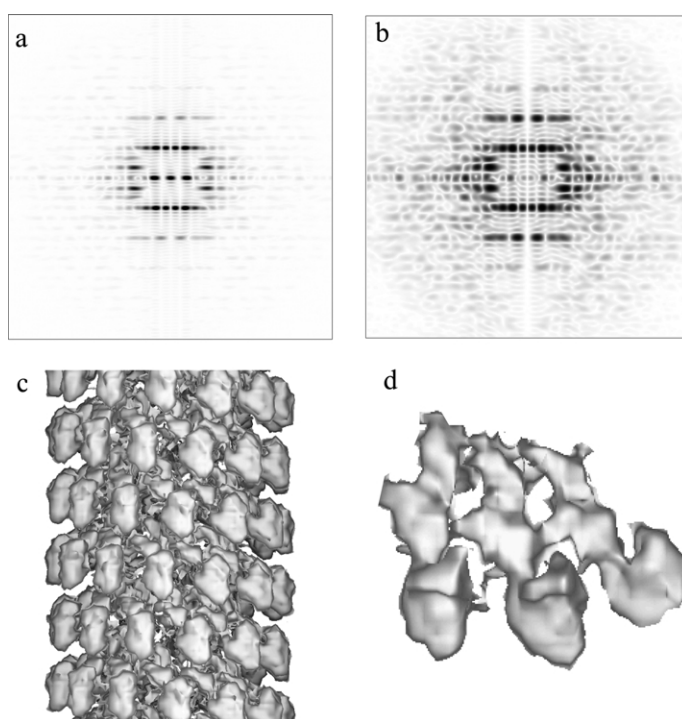


Figure 8. Correction of the contrast transfer function in order to visualise the fine details of reconstruction 1. a, Power spectrum of the reprojection of reconstruction 1 before correction. b, Power spectrum of the reprojection of reconstruction 1 after applying a 40 Å high-pass filter. c and d, Helix and part of a single helical turn of the reconstruction 1 after filtering.

flat domain and the distal domain is very narrow. This may explain why the distal domain in reconstruction 2 is turned in another direction from that in reconstruction 1. It can now also be seen that the middle connection between the monomers consists of density that comes from one monomer and goes over to the next monomer, covering part of this next monomer.

Discussion

The first structural study on a paramyxovirus nucleocapsid was performed by Egelman,² who studied negatively stained and metal shadowed Sendai virus nucleocapsids. Egelman identified at least four different helical states with pitches of 53 Å (two discrete states), 68 Å and 375 Å. The Sendai virus nucleocapsids were supposed to be intact although no SDS/polyacrylamide gels were shown. Sendai nucleoprotein is very easily cleaved by contaminating proteases, like measles virus N, and it is possible that some of the strong variations in the pitch values (53 Å *versus* 68 Å) were due to the presence of cleaved and uncleaved N subunits in the nucleocapsid. In the analysis presented here, intact measles virus N–RNA was clearly too flexible even to begin a high-resolution structural analysis. In a preliminary experiment we had observed that recombinant N–RNA that was harvested too late (beginning of apoptosis of the infected insect cells), had both a tighter and more regular appearance and was also strongly degraded by cellular proteases. Subsequently, we removed the negatively charged C-terminal tail from N by trypsin cleavage, resulting in a preparation of tight coils. From our image analysis results it is obvious that this preparation is still very heterogeneous with probably a large number of helical states.

The helical analysis was only successful for a single helix and, apart from the resolution, the result of this analysis was very similar to reconstruction 1. The combined helical–single-particle analysis presented in Figure 3 also gave a similar result but, again, the resolution was anisotropic and rather poor. Only the third method of analysis, IHRSR, gave a satisfactory result with a good resolution. However, for reconstruction 1 only about 10% of the images was used. A second model derived with this method was of much lower quality. The fact that we found only two solutions does not mean that there are not many more intermediate helical states, but that these intermediates are not strongly represented in the population. The resolution of the reconstructions presented here may be limited by the quality and the number of images. When images can be taken using a field emission gun as electron source one of these limitations may be reduced. A significant increase in the amount of images included in the reconstruction may also increase the resolution. The third limiting factor is the pixel size that we

had to use because of limits in computing power. Even if all these technical problems are solved, the final limiting factor may be the level to which the complexes included into the reconstruction are identical.

One of the reasons for the observed flexibility in this helix may be due to the fact that there are no connections between successive helical turns. A similar situation is encountered with filaments consisting of DNA and the recA protein^{24,29} where the IHRSR method of analysis was also much more successful than more classical helical analysis methods. These helical structures are very different from those of tobacco mosaic virus or bacteriophage tails where there are extensive contacts between subunits in all directions. The contact surface area between subunits in the measles virus N–RNA is less than 1% of the total surface area per subunit.

The position of the RNA in the N–RNA structure was suggested from an increase in density after labelling with *cis*-Pt that has an affinity for neighbouring purines. The reconstructions of the untreated and the *cis*-Pt-treated N–RNAs were very similar but not identical. The density connecting neighbouring protein subunits had increased but the clear bridge towards the middle of the proteins (Figures 5 and 7) was less pronounced. In addition, the thin neck connecting the terminal lobe of density with the domains more interior to the helix seems to be slightly different between untreated and *cis*-Pt-treated reconstructions. We cannot exclude that the treatment has led to slight changes in the protein–RNA structure other than labelling the RNA. The fact that the increase in density after labelling is at the region of the structure closest to the helical axis, agrees with the length of the helical path of the RNA at this radius. The promoter for replication of Sendai virus and human parainfluenza virus type 3 is made up by two elements in the genomic RNA. The first element consists of the first 12 nucleotides, bound to the first two N protein subunits, whereas the second element is a hexameric sequence repeat that is bound to subunits 14–16.^{30,31} The polymerase complex binds to both RNA elements in the helical nucleocapsid where subunits 14–16 are close neighbours to subunits 1–2 by being in the next turn. This would probably imply that the RNA is situated at either tip of the elongated N subunit, since it is difficult to see how the polymerase could interact with the two RNA elements if the RNA were bound in the middle of N, above or beneath the protein. Since the outside diameter of the helical structure is too large for the length of six nucleotides, this also suggests that the RNA is situated close to the helical axis.

The open space inside the helix has a diameter of about 65 Å. The physical size of the measles virus polymerase complex, L plus a tetramer of P,³² is not known but the molecular mass of the L subunit is similar to that of the influenza virus polymerase complex that has a diameter of about 110 Å.³³ This

would mean that the complex of L plus a P-tetramer is too large to be situated inside the helical coil. The N–RNA may have to change its structure for the large polymerase complex to have access to the RNA. It is possible that the complex with the large helical pitch observed by Egelman² on Sendai virus nucleocapsids or the unwound measles N–RNA shown in Figure 2a is the open helical form to which the polymerase complex can attach and on which it is active. Such big changes in the helical structure are possible because there are no connections between subsequent helical turns of the N–RNA.

This polymerase complex binds to the negatively charged C-terminal tail of N that was removed by the trypsin treatment. We cannot say directly where this domain would have been in our structure. However, the shape of measles virus N is similar to the shape of trypsin-treated rabies virus N (see Figure 9) where the C-terminal domain was also removed.¹⁶ Since the structures of both intact and trypsin-treated rabies virus N were determined, we know the position of the rabies virus C-terminal domain of N. If the measles and rabies virus proteins were indeed homologous, this would imply that the C-terminal domain of measles virus N is situated close to the helical axis and this would position the polymerase complex also close to this axis where the RNA is bound.

For RSV it was shown that the N-terminal 92 amino acid residues were sufficient for the formation of nucleocapsid-like structures¹⁰ and could form a helical structure. This domain probably corresponds to the domain closest to the helical axis in our reconstructions. It would imply that, like the C-terminal domain, most of the rest of the structure, probably contained in the terminal lobe,

could be removed without a major effect on the helical parameters. This would agree with our reconstructions that show interactions between adjacent monomers are limited to the one-third of the structure that is closest to the helical axis and not between subsequent turns of the helix. It would also imply that the C-terminal domain folds back upon the N-terminal domain.

At a first glance, the helical nucleocapsids formed by the paramyxoviruses, rhabdoviruses and orthomyxoviruses are very different (Figure 9, top). However, the overall shapes of the N-protein subunits are actually quite similar (Figure 9, bottom). A major difference between the nucleoprotein molecules in the various N–RNAs is that influenza virus nucleoprotein has only one contact area with neighbouring molecules, leading to a very flexible helical structure and the possibility of coiling back on itself in order to make a double coil, the rabies virus N proteins have contact areas at each end of the molecule leading to the formation of coiled ribbons, whereas the measles virus N molecules have two contact areas towards one side on the molecule leading to the characteristic paramyxovirus helices.

Materials and Methods

Expression and purification of recombinant measles virus N–RNA

Recombinant AcNPV (*Autographa californica* nuclear polyhedrosis virus) coding for full-length measles virus N, Halle strain³⁴ was used to infect *Sf21* insect cells at an m.o.i. of 5. Insect cells were grown in TC100 medium (Gibco BRL, Life Technologies) supplemented with 10% (v/v) foetal calf serum and antibiotics (penicillin/streptomycin) in monolayer cultures at 27 °C. For each batch

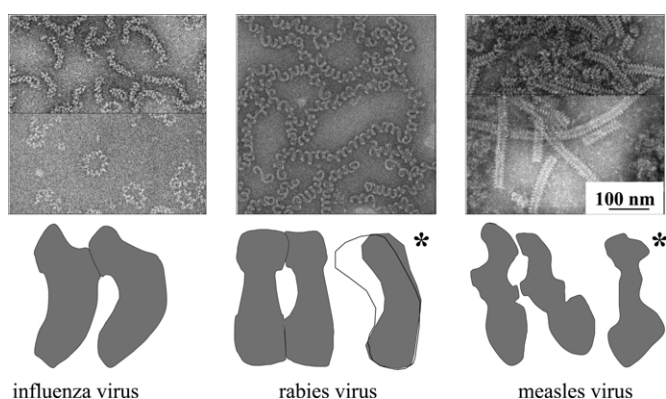


Figure 9. Comparison of nucleocapsids and individual N subunits from influenza, rabies and measles virus. Influenza virus nucleocapsids are circular and form a double-coiled helix resulting in flexible rods. The open circular form can be observed under low-salt conditions.³⁹ There is only a single point of contact between influenza virus nucleoprotein subunits^{33,40} leading to this highly flexible architecture. Rabies virus nucleocapsids are coiled ribbons and there are two points of contact between the N subunits, at the

extreme ends of the molecules. The coiling of the ribbons in free nucleocapsids is different from that inside a virion where the rigid coil gives the shape to the particle.¹⁶ Measles virus nucleocapsids are rather tight helices where the number of subunits per turn is determined by the double contacts between the nucleoprotein monomers. Despite the differences in nucleocapsid morphologies, the shapes of the N subunits of the three virus groups are rather similar. The top part of this Figure consists of electron micrographs from negatively stained nucleocapsids isolated from virus, whereas the lower part indicates the N subunit shape derived from various image analysis methods (influenza virus;⁴⁰ rabies virus¹⁶ and this work for measles virus). The N protein of influenza virus was intact, that of rabies virus both intact (black outline) and trypsin treated to remove the C-terminal domain (filled figure), whereas for measles virus only trypsin treated N was determined. The asterisks show side views of rabies and measles virus N protein, indicating the similarities between the trypsin-treated proteins.

of protein, six 175 cm² flasks were infected and the cells harvested at 48 hours post infection. The cells were pelleted at 350g for ten minutes and resuspended in a hypotonic buffer (50 mM NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM EDTA in the presence of the protease inhibitor cocktail complete™-EDTA free from Roche Diagnostics), and lysed by three cycles of freezing and thawing (in liquid nitrogen and at 37 °C, respectively). Cell debris was pelleted at 12,000g for 15 minutes at 4 °C and the cleared lysate was loaded onto the top of a continuous 20% to 40% (w/w) CsCl gradient in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl and centrifuged for 16 hours at 30,000 rpm and 12 °C in an SW41 rotor (Beckman; corresponds to 111,000g). The visible band in the gradient was collected by puncturing the tube and the sample was dialysed in 150 mM NaCl, 20 mM Tris-HCl (pH 7.5). The presence of the nucleoprotein was verified by SDS/10% (w/w) PAGE and the presence of RNA by taking a UV absorption spectrum.

Trypsin treatment of recombinant measles virus N-RNA

Purified N-RNA (200 µl) was digested with trypsin at a final trypsin concentration of 0.1 mg/ml for ten minutes at 37 °C. The digestion mixture was deposited on top of a 450 µl 30% (v/v) glycerol cushion and centrifuged for one hour at 45,000 rpm and 4 °C in an SW55 rotor (Beckman; 192,000g). The pellet was resuspended in 200 µl of buffer and the remaining glycerol dialysed away in 150 mM NaCl, 20 mM Tris-HCl (pH 7.5). The dialysed sample was used for negative staining, cryo-EM or metal shadowing.

N-terminal sequencing and mass spectrometry after trypsin treatment

Measles virus N (Halle strain; SWISSPROT ID NCAP_MEASH) is 525 amino acid residues long and has a calculated molecular mass of 58,159 Da. Purified N-RNA was analysed by electrospray mass spectrometry, giving a mass of 58,155 Da for the protein, confirming that the nucleoprotein was intact. After trypsin treatment, the digested N-RNA was used for N-terminal sequencing and electrospray mass spectrometry. The resulting sequence (¹⁴NKDKPPITSG²³) and mass (43,515 Da) indicate that the part of the nucleoprotein remaining after digestion comprises of amino acid residues 14–405 (calculated molecular mass 43,573 Da).

Labelling of RNA in N-RNA

cis-Platinum diamine dichloride (cisplatin, SIGMA) was dissolved in water to 5 mM and left for one hour at room temperature to promote hydrolysis of the molecule. After one hour, this solution was added to purified, trypsin-digested measles virus N-RNA at a final concentration 0.25 mM and incubated for two hours at room temperature. Before EM analysis, the labelled N-RNA was separated from the electron dense reagent (*cis*-platin) by sedimentation through a glycerol cushion.

Electron microscopy: negative staining

N-RNA at a concentration of about 0.1 mg/ml was adsorbed onto the clean face of a carbon film on mica, negatively stained with 1% (w/v) neutral sodium silicotungstate and observed under low-dose conditions with

a JEOL 1200 EX II microscope at 100 kV and a nominal magnification of 40,000 ×.

Unidirectional surface shadowing

An N-RNA solution of 0.5 mg/ml was adsorbed to carbon-coated EM grids for 30 s, washed once in distilled water and subsequently quick-frozen in liquid nitrogen. The grids were then mounted on a special specimen holder, designed for the so-called "Midilab" freeze-drying and shadowing unit. The advantage of this unit is that it is directly connected to an FEI-Philips (Eindhoven, NL) CM12 electron microscope (EM), which allows a direct transfer of specimens into the EM without exposing them to atmospheric conditions.³⁵ This procedure essentially allows omitting carbon-backing which is otherwise used to avoid oxidation when tantalum/tungsten (Ta/W) is used for shadowing. Samples were freeze-dried for one hour at 190 K. Metal shadowing was performed unidirectionally at an elevation angle of 45° using Ta/W. The sample was then transferred to a modified GATAN-626 cryo-holder, pre-positioned within the EM, for imaging with a GATAN-794 CCD camera (GATAN Inc., Pleasanton, CA).

Frozen hydrated samples

Vitrified samples were prepared according to the method of Dubochet.³⁶ Digested measles virus N-RNA (3 µl) at 1 mg/ml were applied to a thin holey carbon film on a 400-mesh copper grid (the same method was used for the *cis*-Pt-treated sample but on R2/2 quantifoil grids (Quantifoil Micro Tools GmbH, Germany)). After ten seconds the excess liquid was blotted with filter paper for 1–2 s and rapidly plunged into liquid ethane at –175 °C. Specimens were observed with a GATAN cryotransfer stage on a Philips CM20 microscope at 200 kV. Images were recorded on Kodak SO 163 film under low-dose conditions (<10 e⁻/Å²) at a nominal magnification of 38,000 × (39,500 according to TMV calibration) with defocus values between 1.2 µm and 2.5 µm. Negatives were checked on an optical bench.

Scanning of images and correction of CTF

Micrographs were digitised on a Zeiss Photoscan TD scanner with a pixel size of 7 µm (corresponding to a pixel size of 1.77 Å at the specimen). Particles were selected and interactively boxed into boxes with 256 × 256 pixels. The resulting images were binned into boxes with 128 × 128 pixels, thus with a pixel size of 3.44 Å. These images were used to determine the defocus using SUMPS and CTFZEROS.³⁷ The data were then corrected for the CTF corresponding to the defocus value of the image in two different ways: either the phases were merely flipped or the images were deconvoluted as far as possible for the CTF.³⁷ The resolution limit in the images of the digested N-RNA complex was considered to be 9 Å, as this was the distance of the last visible zero of the CTF. The images of the trypsin-digested and *cis*-platinum-treated N-RNA complexes were corrected in the same way but filtered to 14 Å because of their lower quality.

Determination of the resolution of the reconstructions

The resolution of the resulting structures was

determined by Fourier shell correlation³⁸ using a threshold value of 0.5 (Figure 4). For determination of the resolution of reconstruction 1 (Figure 4) we compared the two structures (which turned out to be identical) that were determined using starting models A and B. We also split the 460 remaining images used for reconstruction 1 into halves in order to calculate two independent reconstructions. The Fourier shell correlations for these two reconstructions gave the same value for the resolution. For reconstructions 2 and 3 we only compared two independent structures using half of the selected images.

Acknowledgements

We thank Dr Heinz Gross and Peter Tittmann for having granted us access to the Midilab setup at the ETH-Zuerich Hoenggerberg and Dr James Conway for access to the cryo-EM at the Institut de Biologie Structurale, Grenoble. We thank Drs Fabian Wild and Robin Buckland for the recombinant baculovirus coding for measles virus N and for discussions. We are also grateful to Professor Dan Kolakofsky for extensive discussions about replication and transcription of negative strand RNA viruses and to Dr Florence Baudin for her suggestion to label the RNA with *cis*-Pt. Finally, we are greatly indebted to Professor Ed Egelman for his suggestions and for providing the programs for the IHRSR method of analysis.

References

- Emerson, S. U. (1987). Transcription of vesicular stomatitis virus. In *The Rhabdoviruses* (Wagner, R. R., ed.), pp. 245–269, Plenum, New York.
- Egelman, E. H., Wu, S. S., Amrein, M., Portner, A. & Murti, G. (1989). The Sendai virus nucleocapsid exists in at least four different helical states. *J. Virol.* **63**, 2233–2243.
- Calain, P. & Roux, L. (1993). The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *J. Virol.* **67**, 4822–4830.
- Kolakofsky, D., Pelet, T., Garcin, D., Hausmann, S., Curran, J. & Roux, L. (1998). Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. *J. Virol.* **72**, 891–899.
- Iseni, F., Baudin, F., Blondel, D. & Ruigrok, R. W. H. (2000). Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *RNA*, **6**, 270–281.
- Iseni, F., Baudin, F., Garcin, D., Marq, J. B., Ruigrok, R. W. H. & Kolakofsky, D. (2002). Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids. *RNA*, **8**, 1056–1067.
- Spehner, D., Kirn, A. & Drillien, R. (1991). Assembly of nucleocapsid like structures in animal cells infected with a vaccinia virus recombinant encoding the measles virus nucleoprotein. *J. Virol.* **65**, 6296–6300.
- Fooks, A. R., Schadeck, E., Liebert, U. G., Dowsett, A. B., Rima, B. K., Steward, M. *et al.* (1995). High-level expression of the measles virus nucleocapsid protein by using a replication-deficient adenovirus vector: induction of an MHC-1-restricted CTL response and protection in a murine model. *Virology*, **210**, 456–465.
- Bhella, D., Ralph, A., Murphy, L. B. & Yeo, R. P. (2002). Significant differences in nucleocapsid morphology within the Paramyxoviridae. *J. Gen. Virol.* **83**, 1831–1839.
- Murphy, L. B., Loney, C., Murray, J., Bhella, D., Ashton, P. & Yeo, R. P. (2003). Investigations into the amino-terminal domain of the respiratory syncytial virus nucleocapsid protein reveal elements important for nucleocapsid formation and interaction with the phosphoprotein. *Virology*, **307**, 143–153.
- Mavrakis, M., Kolesnikova, L., Schoehn, G., Becker, S. & Ruigrok, R. W. H. (2002). Morphology of Marburg virus NP-RNA. *Virology*, **296**, 300–307.
- Finch, J. T. & Gibbs, A. J. (1970). Observations on the structure of the nucleocapsids of some paramyxoviruses. *J. Gen. Virol.* **6**, 141–150.
- Hosaka, Y. & Hosoi, J. (1983). Study of negatively stained images of Sendai virus nucleocapsids using minimum-dose system. *J. Ultrastruct. Res.* **84**, 140–150.
- Buchholz, C. J., Spehner, D., Drillien, R., Neubert, W. J. & Homann, H. E. (1993). The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly. *J. Virol.* **67**, 5803–5812.
- Bankamp, B., Horikami, S. M., Thompson, P. D., Huber, M., Billeter, M. & Moyer, S. A. (1996). Domains of the measles virus N protein required for binding to P protein and self-assembly. *Virology*, **216**, 272–277.
- Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D. & Ruigrok, R. W. H. (2001). Structure of recombinant rabies virus N-RNA and identification of the phosphoprotein binding site. *J. Virol.* **75**, 490–498.
- Longhi, S., Receveur-Brechot, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D. *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.
- Johansson, K., Bourhis, J. M., Campanacci, V., Cambillau, C., Canard, B. & Longhi, S. (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.
- Mountcastle, W. E., Compans, R. W., Lackland, H. & Choppin, P. W. (1974). Proteolytic cleavage of subunits of the nucleocapsid of the paramyxovirus simian virus 5. *J. Virol.* **5**, 1253–1261.
- Heggeness, M. H., Scheid, A. & Choppin, P. W. (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.
- Crowther, R. A., Henderson, R. & Smith, J. M. (1996). MRC image processing programs. *J. Struct. Biol.* **116**, 9–16.
- Schroeter, J. P. & Bretaudiere, J. P. (1996). SUPRIM: easily modified image processing software. *J. Struct. Biol.* **116**, 131–137.
- DeRosier, D. J. & Moore, P. B. (1970). Reconstruction of three-dimensional images from electron

- micrographs of structures with helical symmetry. *J. Mol. Biol.* **52**, 355–369.
24. Egelman, E. H. (2000). A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy*, **85**, 225–234.
 25. Girault, J. P., Chottard, J. C., Guittet, E. R., Lallemand, J. Y., Huynh-Dinh, T. & Igolen, J. (1982). Specific platinum chelation by the guanines of the deoxyhexanucleotide d(T-G-G-C-C-A) upon reaction with cis-[Pt(NH₃)₂(H₂O)₂](NO₃)₂. *Biochem. Biophys. Res. Commun.* **109**, 1157–1163.
 26. Sherman, S. E., Gibson, D., Wang, A. H. & Lippard, S. J. (1985). X-ray structure of the major adduct of the anticancer drug cisplatin with DNA: cis-[Pt(NH₃)₂(d(pGpG))]. *Science*, **230**, 412–417.
 27. Burnouf, D., Gauthier, C., Chottard, J. C. & Fuchs, R. P. (1990). Single d(ApG)/cis-diamminedichloroplatinum(II) adduct-induced mutagenesis in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **87**, 6087–6091.
 28. Kronenberg, S., Kleinschmidt, J. A. & Bottcher, B. (2001). Electron cryo-microscopy and image reconstruction of adeno-associated virus type 2 empty capsids. *EMBO Rep.* **2**, 997–1002.
 29. VanLoock, M. S., Yu, X., Yang, S., Lai, A. L., Low, C., Campbell, M. J. & Egelman, E. H. (2003). ATP-mediated conformational changes in the RecA filament. *Structure*, **11**, 187–196.
 30. Tapparel, C., Maurice, D. & Roux, L. (1998). The activity of Sendai virus genomic and antigenomic promoters requires a second element past the leader template regions: a motif (GNNNNN)₃ is essential for replication. *J. Virol.* **72**, 3117–3128.
 31. Hoffman, M. A. & Banerjee, A. K. (2000). Analysis of RNA secondary structure in replication of human parainfluenza virus type 3. *Virology*, **272**, 151–158.
 32. Tarbouriech, N., Curran, J., Ruigrok, R. W. H. & Burmeister, W. P. (2000). Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nature Struct. Biol.* **7**, 777–781.
 33. Martín-Benito, J., Area, E., Ortega, J., Llorca, O., Valpuesta, J. M., Carrascosa, J. L. & Ortín, J. (2001). Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO Rep.* **2**, 313–317.
 34. Ravanel, K., Castelle, C., Defrance, T., Wild, T. F., Charron, D., Lotteau, V. & Rabourdin-Combe, C. (1997). Measles virus nucleocapsid protein binds to FcγRII and inhibits human B cell antibody production. *J. Expt. Med.* **186**, 269–278.
 35. Gross, H., Krusche, K. & Tittmann, P. (1990). Recent progress in high-resolution shadowing for biological samples. In *Proceedings of the XIIIth International Congress for Electron Microscopy* (Peachey, L. D. & Williams, D. B., eds), vol. 1, pp. 510–511, San Francisco Press Inc., San Francisco.
 36. Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowell, W. & Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Quart. Rev. Biophys.* **21**, 129–228.
 37. Conway, J. F. & Steven, A. C. (1999). Methods for reconstructing density maps of “single” particles from cryoelectron micrographs to subnanometer resolution. *J. Struct. Biol.* **128**, 106–118.
 38. Yang, S., Yu, X., Galkin, V. E. & Egelman, E. H. (2003). Issues of resolution and polymorphism in single-particle reconstruction. *J. Struct. Biol.* **144**, 162–171.
 39. Klumpp, K., Ruigrok, R. W. H. & Baudin, F. (1997). Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *EMBO J.* **16**, 1248–1257.
 40. Ruigrok, R. W. H. & Baudin, F. (1995). Structure of influenza virus RNP. II. Purified, RNA-free influenza ribonucleoprotein forms structures that are indistinguishable from the intact viral ribonucleoprotein particles. *J. Gen. Virol.* **76**, 1009–1014.

Edited by Sir A. Klug

(Received 16 February 2004; received in revised form 24 March 2004; accepted 26 March 2004)

**Revue : Structures impliquées dans la réplication et la transcription des
virus à ARN non segmenté de sens négatif**

Structures impliquées dans la réplication et la transcription des virus à ARN non segmenté de sens négatif

A.A.V. Albertini^{1,2}

G. Schoehn^{1,2}

R.W.H. Ruigrok^{1,2}

¹ Laboratoire de virologie moléculaire et structurale, FRE 2854 CNRS-UJF, c/o EMBL, BP 181, 38042 Grenoble Cedex 9

² EMBL, Antenne de Grenoble, 6, rue Jules-Horowitz, BP181, 38042 Grenoble Cedex 9

Résumé. Les virus à ARN négatif non segmenté ont un génome composé d'un ARN simple brin ayant un sens complémentaire de celui de l'ARN messenger. L'ARN viral forme un complexe stable et hélicoïdal en s'associant avec la nucléoprotéine virale. La polymérase virale (ARN polymérase ARN-dépendante) de ces virus se fixe sur cette matrice (nucléoprotéine plus ARN) par l'intermédiaire d'un cofacteur, la phosphoprotéine. Cette polymérase est uniquement active sur cette matrice et pas sur l'ARN viral nu. Les caractéristiques moléculaires et structurales de ces composants viraux et leurs interactions sont examinées dans cette revue. La connaissance de leur structure détaillée est indispensable pour la mise au point de molécules ayant une activité antivirale.

Mots clés : virus à ARN négatif, réplication, nucléocapside, antiviral, virus de la rage, virus de la rougeole

Abstract. Non-segmented negative strand RNA viruses have an RNA genome that is in the opposite sense of that of messenger RNA. The viral RNA forms a tight helical complex with the viral nucleoprotein. The viral RNA dependent RNA polymerase binds to this helical complex with the help of a cofactor, the phosphoprotein and is only active on this RNA-protein complex and not on naked RNA. In this review we will describe the molecular and structural characteristics of all these molecules and the interactions between the various components. Detailed structural knowledge of this viral replication system will be necessary for the directed development of antiviral drugs.

Key words: negative strand RNA viruses, replication, nucleocapsid, antiviral drug, rabies virus, measles virus

Les virus à ARN simple brin de polarité négative sont des virus enveloppés dont l'ARN viral est complémentaire de l'ARN messenger. Ils possèdent soit un unique brin d'ARN viral (ordre des *Mononegavirales*), soit un génome segmenté. Les familles des *Rhabdoviridae* (virus de la rage), *Paramyxoviridae* (virus de la rougeole, virus respiratoire syncytial, Hendra et Nipah), *Filoviridae* (virus Ebola) et *Bornaviridae* appartiennent à l'ordre des *Mononegavirales*. Le génome des virus des familles des *Arenaviridae* (2 segments d'ARN), *Bunyaviridae* (3 segments, virus de la fièvre de la vallée du Rift) et *Orthomyxoviridae* (7-8 seg-

ments, virus de la grippe) est aussi un ARN simple brin de polarité négative ; en revanche cet ARN ne se présente pas sous la forme d'une seule molécule mais est segmenté. Tous ces virus sont à l'origine de nombreuses pathologies humaines (maladies des voies respiratoires inférieures et supérieures causées par le virus de la grippe ou par le virus respiratoire syncytial). Les virus de la rage, de Nipah et certains *Bunyaviridae* sont quant à eux responsables d'encéphalites. D'autres virus au sein de cette famille (virus Ebola, virus de la fièvre de Crimée-Congo, virus Lassa) sont également à l'origine de fièvres hémorragiques. Cette revue se limitera uniquement aux *Mononegavirales*.

Lors de l'infection, après leur entrée dans la cellule hôte, la première activité de ces virus est la transcription de l'ARN

Tirés à part : R.W.H. Ruigrok

simple brin négatif en ARN messenger et non sa traduction directe comme dans le cas des virus à ARN positif. L'ARN des *Mononegavirales* n'est jamais nu, ni dans les particules virales, ni dans les cellules infectées mais toujours sous forme de complexe nucléoprotéique. La protéine majoritaire de ce complexe est la nucléoprotéine N qui, en décorant entièrement l'ARN viral, forme une nucléocapside N-ARN hélicoïdale (aussi appelée matrice N-ARN). Celle-ci est caractéristique de ce groupe de virus et la séquence de son gène est souvent utilisée pour leur classification. Deux autres protéines sont associées à la nucléocapside : la polymérase virale (ARN polymérase ARN-dépendante) et son cofacteur, la phosphoprotéine P. Cette dernière est indispensable à la fixation de la polymérase sur la nucléocapside [1]. Cet ensemble (nucléocapside – polymérase – phosphoprotéine) est l'unité de réplication et forme la partie infectieuse du virus.

La nucléoprotéine des *Mononegavirales*

Le poids moléculaire des nucléoprotéines se situe généralement autour de 60 kDa, sauf celui des *Filoviridae* qui est de 80 kDa [2]. Si l'homologie entre les trois quarts N-terminaux de toutes ces nucléoprotéines est significative, les parties C-terminales sont pour leur part extrêmement diverses. Le domaine conservé N-terminal de la nucléoprotéine fixe l'ARN viral [3] et, pour la plupart des *Mononegavirales*, la partie C-terminale interagit avec la phosphoprotéine. Pour le virus de la rougeole, il a été proposé que la partie C-terminale ne soit pas, contrairement aux protéines normales, dans un état replié mais plutôt désordonné [4]. Le repliement de ce fragment de N serait induit par le contact avec son partenaire, la phosphoprotéine. À la différence de celui de la rougeole, le domaine C-terminal de la nucléoprotéine du virus de la rage semble former une structure bien définie [5].

Fixation sur l'ARN viral

La nucléoprotéine s'associe à l'ARN viral par l'intermédiaire du squelette ribose-phosphate, laissant ainsi les bases des nucléotides accessibles à la polymérase virale [6, 7]. Initialement, la nucléoprotéine était supposée n'avoir qu'un rôle protecteur au sein de la matrice N-ARN. En effet, elle protège l'ARN contre une digestion par des ribonucléases (ARNases). L'évolution des connaissances concernant la fonction de N est similaire à celle concernant les nucléosomes de la chromatine des cellules eucaryotes : leur rôle initial de compactage et de protection s'est enrichi d'une composante de régulation de l'expression des gènes. Les nucléoprotéines modulent le niveau de transcription des ARN messagers viraux et influencent la manière dont les signaux de transcription et de réplication codés par l'ARN viral sont interprétés par la polymérase [7, 8].

Structure

Les nucléoprotéines se fixent sur un nombre entier de nucléotides. Ce nombre est différent et spécifique pour chaque famille de *Mononegavirales* : 6 nucléotides par nucléoprotéine pour les *Paramyxoviridae* [9], 9 par nucléoprotéine pour les *Rhabdoviridae* [10, 11] et entre 12 et 15 par nucléoprotéine pour les *Filoviridae* [12]. Pour être viables, certains *Paramyxoviridae* (ceux qui ont une activité d'édition de leurs ARNm) doivent avoir un génome viral dont la longueur totale est un multiple de six nucléotides [8, 13].

Du point de vue ultra-structural, toutes les nucléocapsides des virus à ARN négatif sont hélicoïdales mais avec des paramètres hélicoïdaux différents. La *figure 1* montre des clichés de microscopie électronique de nucléocapsides isolées à partir de différents virus. Elle représente respectivement les vues des nucléocapsides du virus de la rougeole (*Paramyxoviridae*), du virus de la rage (*Rhabdoviridae*) et du virus de Marbourg (*Filoviridae*). Seule la nucléocapside du virus de Marbourg n'a pas été isolée à partir du virus, vu sa pathogénicité, mais produite sous la forme d'une protéine recombinante dans des cellules d'insectes [12]. L'expression des nucléoprotéines des *Mononegavirales* dans différents systèmes d'expression (aussi bien eucaryotes que bactériens) produit des complexes dont la morphologie est identique à celle des nucléocapsides virales (*figures 1 et 2*). Les nucléoprotéines recombinantes s'associent en effet de manière irréversible aux ARN cellulaires pour former des complexes N-ARN ayant la même structure et présentant la même stœchiométrie ($N = nt$) que les nucléocapsides virales [12-17].

La *figure 1* montre que, bien qu'ayant une forme plutôt allongée, les nucléocapsides des virus à ARN négatif sont relativement irrégulières et ont surtout des morphologies très différentes selon les virus. Cette hétérogénéité entre les différentes nucléocapsides d'un même virus et même au sein d'une nucléocapside exclut leur cristallisation, rendant impossible la détermination de leur structure à une échelle atomique. Les techniques d'analyses d'images utilisées en microscopie électronique nécessitent également des objets homogènes et réguliers et ne sont donc pas utilisables pour les nucléocapsides purifiées à partir de virus. Pour pallier ce manque d'information structurale, des systèmes plus réguliers ont été développés. Ces nouveaux objets permettent de calculer, grâce à l'analyse des images obtenues en microscopie électronique, des reconstructions tridimensionnelles (3D) à une résolution plus haute.

Paramyxoviridae

Dans le cas du virus de la rougeole, il est difficile, à partir de virus ou de cellules infectées, d'isoler des quantités suffisantes de nucléocapsides pour permettre des analyses structurales. Ce problème a été résolu en exprimant la nucléo-

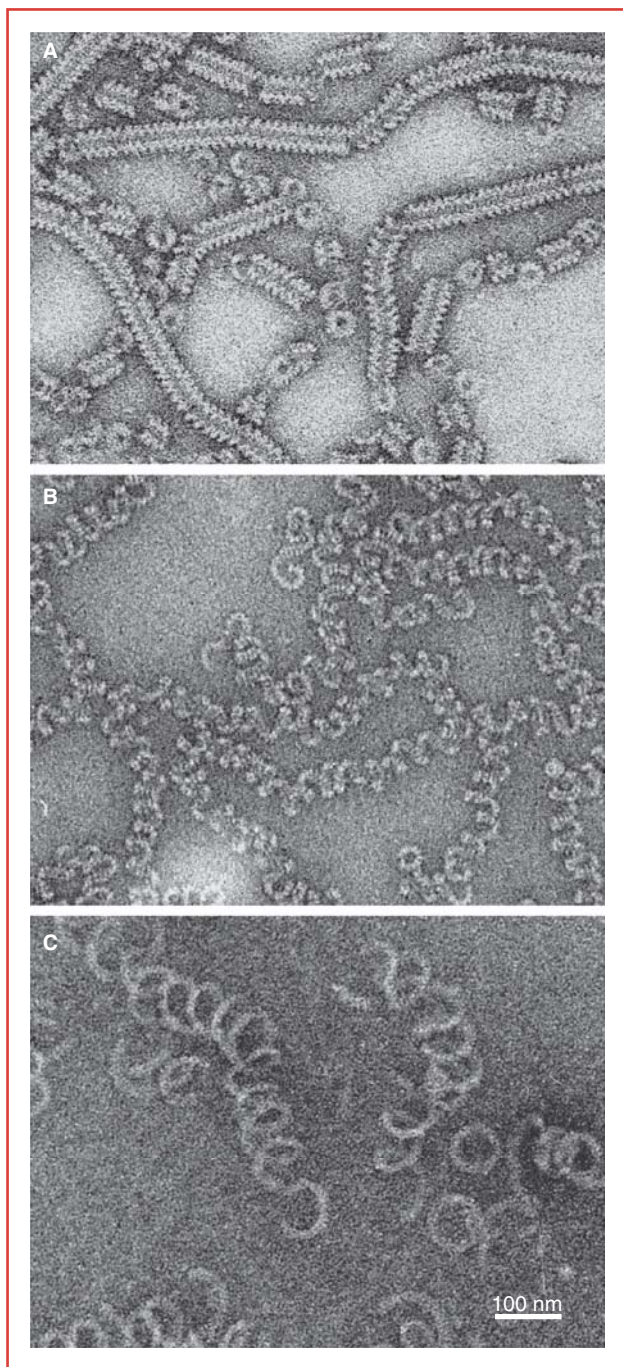


Figure 1. Clichés de microscopie électronique de nucléocapsides préparées en coloration négative : **A**) Nucléocapside du virus de la rougeole. **B**) Nucléocapside du virus de la rage. **C**) Nucléocapside recombinante du virus de Marbourg.

protéine dans des cellules d'insectes. La protéine exprimée se fixe alors sur les ARN cellulaires pour former de longs complexes N-ARN qui sont irréguliers comme ceux isolés à partir du virus (*figures 1 et 2*). Une digestion avec une

protéase, la trypsine, coupe et élimine entièrement le domaine C-terminal des nucléoprotéines (élimination des résidus 406-525) [18]. Cette opération condense les nucléocapsides N-ARN en hélices plus rectilignes et plus régulières (*figure 2, B*). Même si ces structures rigides sont mieux adaptées à une analyse structurale, elles gardent une certaine variabilité, aussi bien au niveau du nombre de monomères de N par tour d'hélice que dans le pas de l'hélice [18]. Une structure de l'état majoritaire de cette nucléocapside a été déterminée à partir d'images obtenues en cryomicroscopie électronique à une résolution de 1,2 nm (*figure 2, C*). Cette structure hélicoïdale est caractérisée par un pas de 5,1 nm et un nombre de monomères de N de 12,35 par tour d'hélice [18]. À cette résolution, l'ARN n'est toujours pas localisable au sein de la structure. Nous avons donc traité ces complexes avec du cisplatine. Ce réactif est connu pour se lier aux bases des nucléotides. Après analyse des images obtenues en cryomicroscopie, nous avons observé une augmentation de densité près de l'axe de l'hélice (pointillé rouge sur la *figure 2, D*). L'ARN se situe donc vraisemblablement dans cette zone, à l'intérieur de la nucléocapside. Nous savons également que la queue C-terminale de N est probablement localisée près de l'ARN (voir plus loin, complexe N-ARN du virus de la rage), dans une position clé permettant au complexe phosphoprotéine-polymérase d'interagir avec l'ARN. La flexibilité des nucléocapsides est probablement indispensable au passage de la polymérase pendant la répllication et la transcription de l'ARN. En effet, même si pour le moment nous ne disposons pas de la structure d'une polymérase d'un virus de la famille des *Mononegavirales*, leur taille et leur forme peuvent être proches de celles de la polymérase du virus de la grippe (11,5 x 10 x 8,5 nm) [19]. Dans ce cas, la distance entre deux tours successifs de l'hélice condensée (5,1 nm) est trop petite pour permettre le passage du complexe formé par la polymérase et son cofacteur, le tétramère de P, qui le fixe sur la matrice N-ARN. La nucléocapside doit donc pouvoir se déformer pour permettre le cheminement de la polymérase.

Le virus Nipah appartient également à la famille des *Paramyxoviridae*. Nous avons étudié le complexe N-ARN de ce virus qui infecte normalement des chauves-souris frugivores en Malaisie. Il a franchi, en 1998, la barrière interespèces et entraîné dans un premier temps une épidémie de maladies respiratoires chez des porcs. Les humains en contact avec ces animaux malades ont à leur tour développé des encéphalites mortelles. Ce virus et sa nucléocapside infectieuse sont trop dangereux pour être étudiés directement dans notre laboratoire. Nous avons donc, comme pour le virus de la rougeole, exprimé la protéine N du virus Nipah dans des cellules d'insectes (système Baculovirus). Les nucléocapsides recombinantes purifiées ont été observées en microscopie électronique dans leur état natif puis,

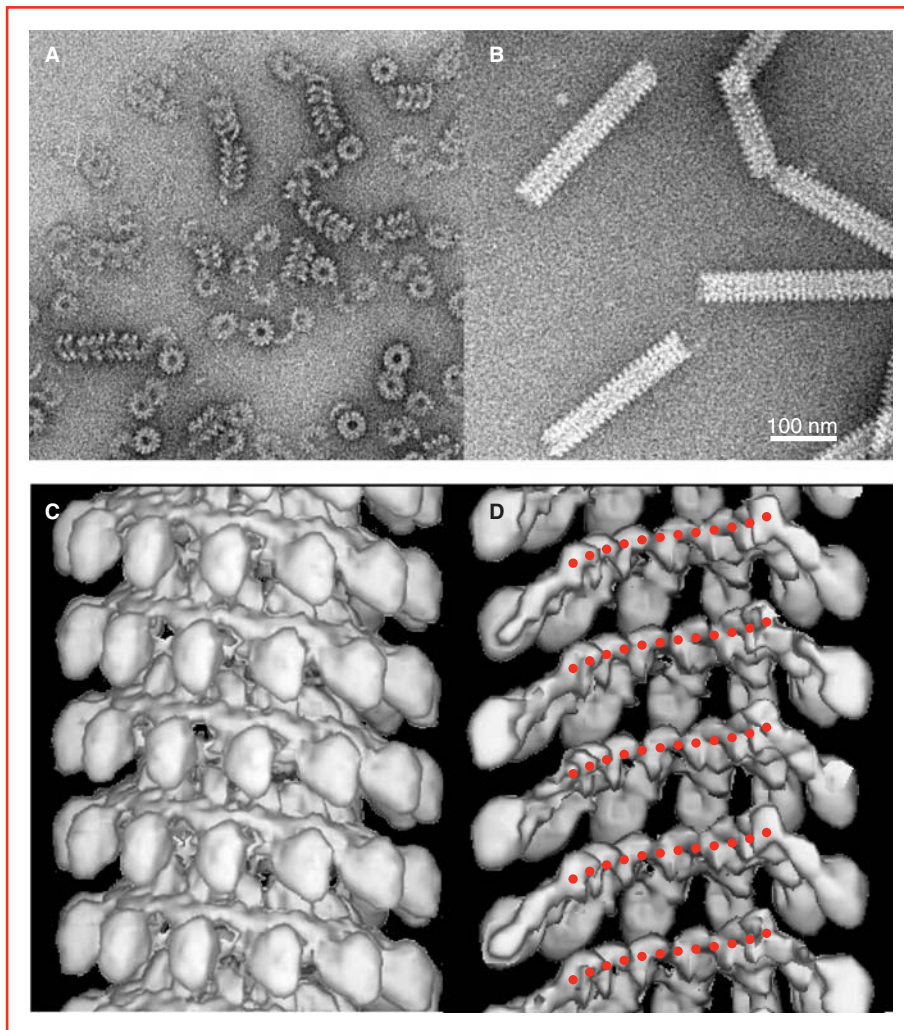


Figure 2. Nucléocapside du virus de la rougeole : **A)** Nucléocapside non digérée observée en coloration négative. **B)** Nucléocapside digérée à la trypsine observée en coloration négative. **C et D)** Structure 3D à 1,2 nm de résolution de la nucléocapside digérée (12,35 monomères de N par tour d'hélice et un pas de 5,1 nm) calculée à partir d'images de cryomicroscopie électronique. Les points rouges visualisent l'endroit où se trouve l'ARN. **C)** Vue de l'extérieur de l'hélice et **D)** vue de l'intérieur après élimination de la moitié de l'hélice.

après traitement par l'élastase, une autre protéase (*figure 3, A*). Bien que le complexe traité paraisse plus ordonné que le natif, nous n'avons vraisemblablement pas réussi à découper d'une façon homogène les extrémités C-terminales de toutes les molécules de N. Les images de microscopie électronique obtenues de l'échantillon après coloration négative et leur reconstruction tridimensionnelle (*figure 3, B*) représentent probablement un mélange de molécules de N intactes et digérées. Cette structure, dont la résolution est plus faible que celle du virus de la rougeole (à cause de l'hétérogénéité de l'échantillon), a presque les mêmes paramètres hélicoïdaux que ce dernier : 13 monomères de N par tour d'hélice et un pas de 6 nm.

D'autres analyses de nucléocapsides des *Paramyxoviridae* ont toutes donné des résultats similaires : des hélices rela-

tivement flexibles comprenant 12 à 14 sous-unités de N par tour. Seule la nucléocapside du virus respiratoire syncytial semble n'avoir que 10 monomères de N par tour d'hélice [17].

Rhabdoviridae

Les nucléocapsides des *Rhabdoviridae* sont encore plus irrégulières que celles des *Paramyxoviridae* et par conséquent inutilisables pour une analyse d'images obtenues par microscopie électronique. Quand les nucléoprotéines du virus de la rage ou du virus de la stomatite vésiculaire (VSV) sont exprimées dans des cellules d'insectes ou dans des bactéries, une partie des protéines se fixe à de longs ARN, probablement des ARN messagers de la cellule. Le reste (environ la moitié des protéines) s'associe aux ARN

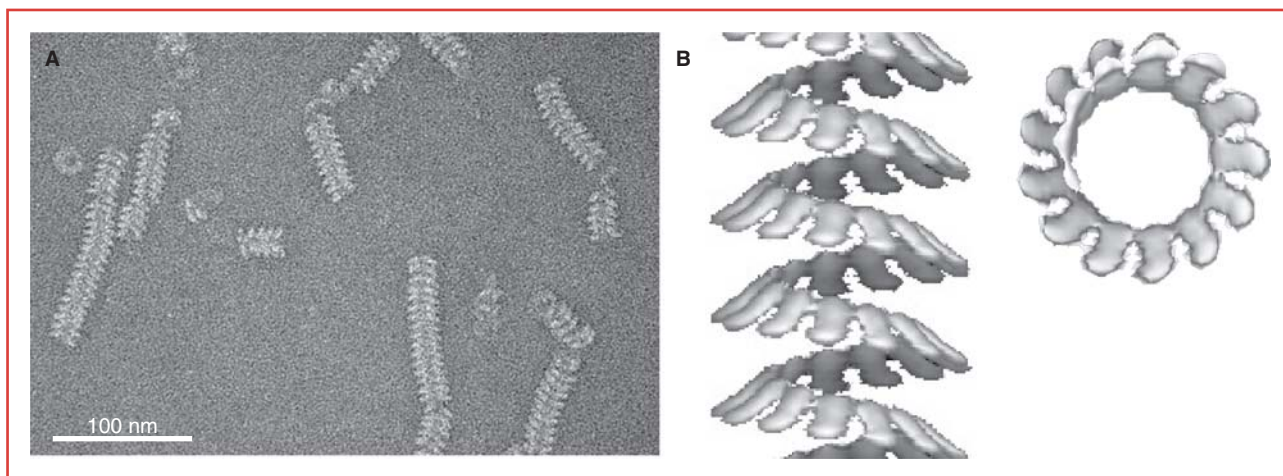


Figure 3. Nucléocapside du virus Nipah : **A)** Coloration négative d'un échantillon de nucléocapside recombinant du virus de Nipah digéré de façon non homogène. **B)** Structure 3D de la nucléocapside du virus de Nipah en vue de côté et de dessus (13 monomères par tour, pas de 6,0 nm) obtenue à partir d'images en coloration négative. Le sens de l'enroulement de l'hélice n'a pas été déterminé expérimentalement mais imposé par comparaison avec la nucléocapside du virus de la rougeole.

de transfert de la cellule, plus courts [16, 20]. Cette dernière fraction forme des anneaux de N-ARN avec un nombre de monomères de N dépendant de la longueur de ces ARN (de 9 à 12 N). La *figure 4* présente les reconstructions 3D d'anneaux formés de 10 monomères de N calculées à partir d'images de cryomicroscopie électronique. Les monomères de N ont une forme de graine de haricot avec une extrémité plus volumineuse (située en haut de la molécule). Quand ces anneaux (ou les nucléocapsides) sont traités avec la trypsine, une coupure après le résidu 376 de la nucléoprotéine [21] élimine le domaine C-terminal de N (résidus 377-450) [5]. Ces nucléocapsides, sans C-terminal, ne fixent plus la phosphoprotéine. Nous avons donc émis l'hypothèse que la phosphoprotéine du virus de la rage, comme celle de la plupart des *Paramyxoviridae*, se fixe au domaine C-terminal de la nucléoprotéine. D'un point de vue structural, la digestion de ce domaine entraîne la diminution du volume de la plus grosse extrémité de la protéine N (*figure 4*). Cette diminution de volume correspond bien au volume prédit du domaine C-terminal éliminé. Le domaine C-terminal est donc localisé au niveau de cette partie manquante et semble être bien replié, contrairement au domaine correspondant dans la protéine N du virus de la rougeole. La co-expression de N et de P de VSV dans *E. coli* induit également la formation d'anneaux de N associés à P [20]. Cet échantillon, gardé à 4 °C pendant quelques mois, a été observé en coloration négative. Pendant ce laps de temps, des protéases contaminantes ont digéré ses protéines P. Le modèle 3D de ces anneaux, réalisé grâce à l'analyse de clichés de microscopie électronique, montre une structure qui ressemble beaucoup à celles des anneaux N-ARN du virus de la rage après digestion à la trypsine [20] (*figure 4*). Les clichés d'origine ont

été pris sur un échantillon coloré négativement à l'acétate d'uranyl. Ce colorant se fixe davantage sur les acides nucléiques et, en les rendant plus denses que les protéines, permet leur localisation. La position de l'ARN ainsi visualisé dans les anneaux N-ARN de VSV est indiquée sur la structure des complexes de N-ARN du virus de la rage dans la *figure 4* par un trait ou par des points rouges. La localisation de l'ARN se superpose à celle du domaine C-terminal de la protéine N. Cela permet une hypothèse intéressante : le domaine C-terminal de N dans les nucléocapsides viraux se trouve dans une position favorable pour couvrir, protéger et interagir avec l'ARN viral. Selon cette hypothèse, la fixation de la phosphoprotéine qui amène la polymérase sur la nucléocapside déplacerait ce domaine et, en dégageant l'ARN, permettrait la réaction de polymérisation des nucléotides lors de la réplication et de la transcription. Cette hypothèse expliquerait également pourquoi l'ARN viral dans les nucléocapsides des *Paramyxo-* et *Rhabdoviridae* réagit si pauvrement avec des petites molécules chimiques [6, 7], alors que l'ARN viral dans les nucléocapsides du virus de la grippe (la grippe n'a ni P, ni domaine C-terminal de N) réagit très fortement avec ces mêmes réactifs [22].

Les phosphoprotéines

Tous les *Mononegavirales* possèdent une phosphoprotéine. Ces protéines ont au moins deux fonctions distinctes lors du cycle viral : elles permettent la fixation de la polymérase sur la nucléocapside et fonctionnent aussi comme molécules chaperons pour N, empêchant la fixation non spécifique des nucléoprotéines nouvellement produites sur les ARN

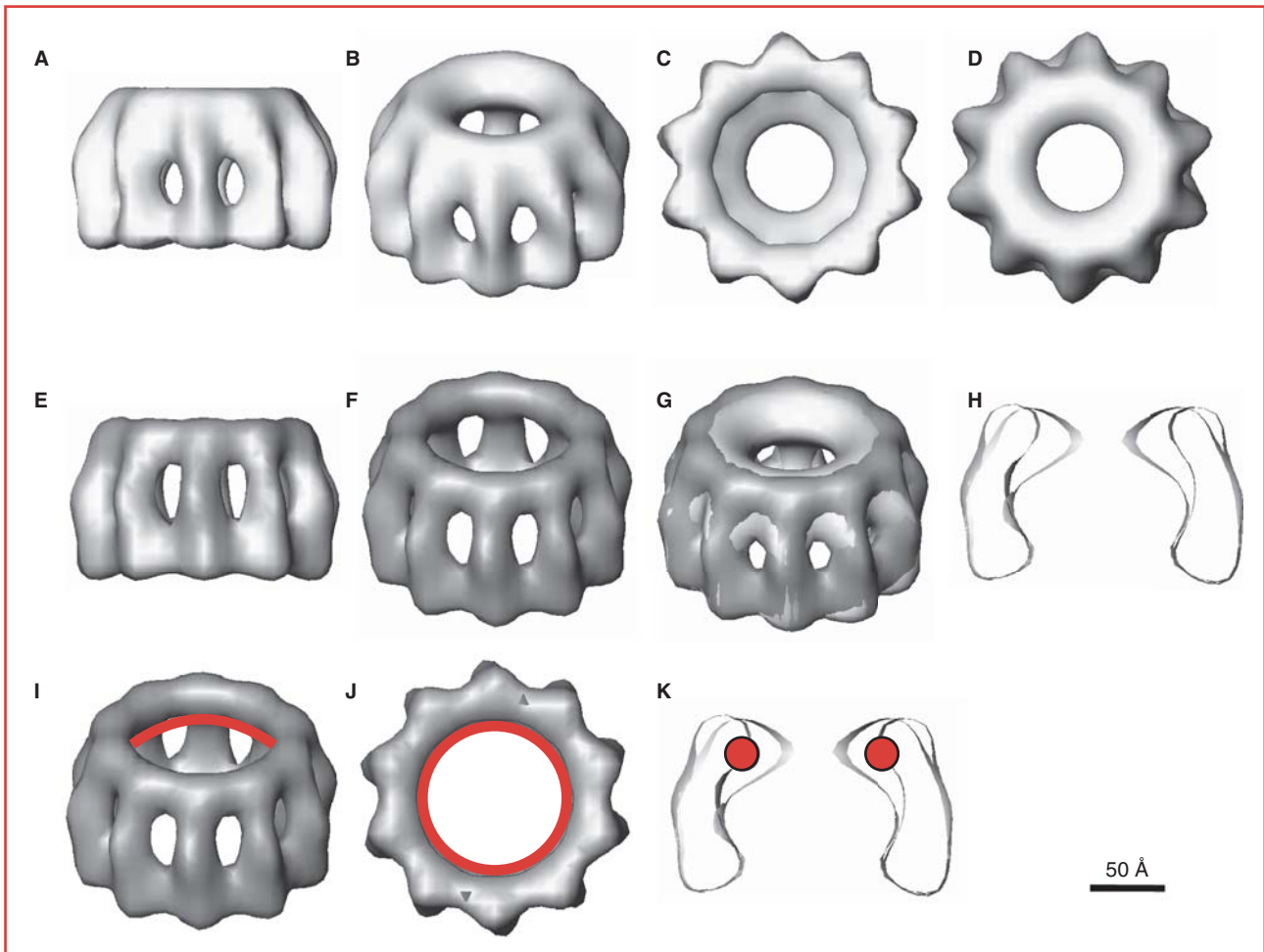


Figure 4. Structures tridimensionnelles des anneaux N-ARN du virus de la rage produits en cellules d'insectes. Les reconstructions 3D ont été calculées à partir d'images de cryomicroscopie électronique. **A à D)** Anneau non digéré. **E et F)** Anneau digéré par la trypsine. **G et H)** Comparaison des anneaux natifs (gris clair) et digérés (gris foncé). **I, J et K)** Localisation de l'ARN par comparaison avec les structures tridimensionnelles des anneaux de VSV obtenues en coloration négative [20].

cellulaires. Il existe également d'autres rôles pour P qui ont surtout été décrits pour le virus de la rage [23].

Les séquences et les longueurs des phosphoprotéines des différents virus sont très variables (568 aa pour Sendai, 297 aa pour la rage), mais, toutes ces protéines ont au moins quatre caractéristiques communes : P a un rôle de molécule chaperon pour N empêchant ainsi sa liaison aux ARN cellulaires (cette forme de N est notée N°) ; elle se fixe sur le complexe N-ARN (nucléocapside) et se lie à la polymérase (L). Toutes les phosphoprotéines étudiées à ce jour forment des homo-oligomères.

La phosphoprotéine la mieux caractérisée est celle du virus de Sendai (figure 5). Cette protéine de 568 acides aminés se divise en deux parties N et C-terminales. La partie N-terminale, non conservée au sein des paramyxovirus, ne semble pas être repliée de façon traditionnelle [24]. Une

seule partie de ce domaine est estimée, avec une forte probabilité, de former une structure en hélice α . Cette courte séquence comprenant les résidus 33 à 41 est impliquée dans la fixation de N au sein du complexe N°-P [25]. La partie C-terminale a été exprimée, puis étudiée par diffusion de neutrons [26]. Elle a une longueur totale de 16 nm. D'un point de vue structural, elle contient deux domaines distincts : un premier domaine permettant l'oligomérisation et un second responsable de la fixation de P sur le complexe N-ARN. Les structures atomiques des deux domaines, déterminées indépendamment [26, 27], sont indiquées dans la figure 5. La structure de la séquence qui fait la jonction entre ces deux domaines (résidus 455 à 517) n'est pas connue mais n'est probablement pas totalement repliée [28]. Le domaine d'oligomérisation induit la formation d'un tétramère d'hélices α torsadées. Cette oligomérisation

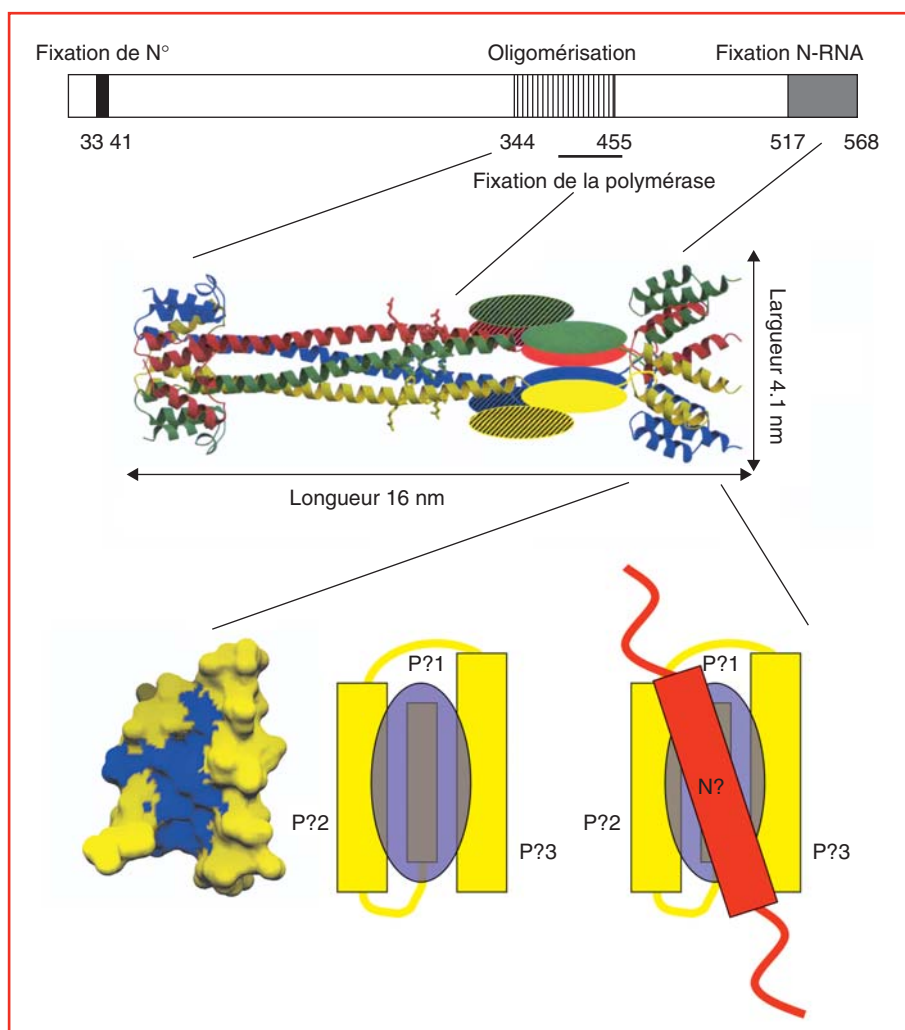


Figure 5. La phosphoprotéine du virus de Sendai. En haut une représentation linéaire de la protéine avec indication de trois domaines fonctionnels. Au milieu, la structure du domaine C-terminal de P : La structure du domaine d'oligomérisation (résidus 344-455) a été déterminée par cristallographie aux rayons X [27], la structure de fixation de la nucléocapside (résidus 517-568) a été déterminée par RMN [26] et les dimensions générales de ce domaine sont dérivées des résultats de diffusion de neutrons aux petits angles. En bas la structure d'un des domaines de fixation de la nucléocapside. Ce domaine consiste en trois hélices ($P\alpha 1$ - $P\alpha 3$, en jaune) avec entre eux une surface hydrophobe (en bleu). Une hélice du domaine C-terminal de N ($N\alpha$) se fixe avec sa surface hydrophobe sur la surface hydrophobe de P [31].

sation de P est indispensable pour permettre son interaction avec la polymérase par l'intermédiaire des chaînes latérales des acides aminés de monomères différents (figure 5). Le domaine de fixation de P sur le complexe N-ARN est constitué de trois courtes hélices α antiparallèles formant une petite structure relativement flexible (figure 5) [26]. La structure atomique de ce même domaine de P du virus de la rougeole a aussi été déterminée, elle est identique à la structure du domaine équivalent du virus de Sendai [29, 30].

Fixation sur la nucléocapside

Les phosphoprotéines, seules ou en complexe avec leur polymérase, se fixent sur le domaine C-terminal des nucléoprotéines. Les domaines de fixation des phosphoprotéines des paramyxovirus Sendai et rougeole sur les nucléoprotéines sont petits, flexibles et relativement instables (figure 5). Les domaines C-terminaux de la nucléoprotéine N de Sendai et de la rougeole sont, eux aussi, très flexibles et seule une petite partie de ce domaine aurait tendance à former une hélice α [4, 30, 31]. Récemment, un complexe

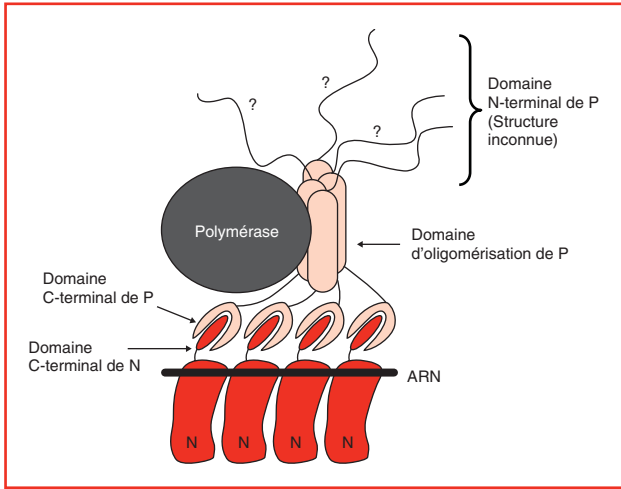


Figure 6. Modèle pour la fixation d'un tétramère de P du virus de Sendai avec une molécule de la polymérase associée sur les domaines C-terminaux de N. Pour l'activité de réplication ou de transcription, la molécule P (très flexible) doit basculer pour mettre la polymérase en contact avec l'ARN viral. Dans le texte nous avons émis l'hypothèse que le domaine C-terminal de N couvre et cache l'ARN viral quand P n'est pas fixée.

entre le domaine C-terminal de P et le domaine C-terminal de la nucléoprotéine N du virus de la rougeole a été cristallisé. Cette partie de N forme une hélice α avec un côté hydrophile et un côté hydrophobe. Le côté hydrophobe de cette hélice entre en contact avec une autre zone hydrophobe formée par les trois hélices α du domaine C-terminal de P (figure 5) [31]. L'affinité extrêmement faible ($kd = 3,5 \times 10^{-5} M$ à $37^\circ C$) [30] entre ces deux domaines explique l'absence de fixation de P du virus de Sendai sous forme monomérique sur la nucléocapside. La figure 6 montre un schéma récapitulatif de l'unité de réplication du virus de Sendai : un tétramère de phosphoprotéine du virus de Sen-

dai avec sa polymérase fixée sur la nucléocapside. La fixation de P expose les bases des nucléotides de l'ARN. Après le basculement de P, la polymérase peut lire la séquence de nucléotides et produire une copie de l'ARN (ARNm ou ARN viral complémentaire).

Comparaison des phosphoprotéines des Paramyxoviridae

Une étude plus fonctionnelle de P de Sendai nous a appris que, si la délétion de toute la partie N-terminale préserve uniquement l'activité de transcription de cette protéine, une délétion des résidus 79-319 produit en revanche une protéine active pour la réplication et la transcription [32] (figure 7). La plus grande partie de cette région N-terminale n'est donc pas indispensable pour ces deux activités de P mais la partie 5' du cadre ouvert de lecture serait nécessaire pour la production des protéines C, V ou W, codée par la même séquence du génome viral [33].

La structure et la fonction de la phosphoprotéine du virus de la rage sont décrites dans une revue de Blondel *et al.* [23]. Il n'est pas, pour le moment, prouvé que tous les aspects structuraux de P du virus de la rage soient identiques à ceux de P de VSV. Pour VSV, la phosphorylation de P apparaît indispensable à son oligomérisation et seule cette forme oligomérique semble active pour la transcription et la réplication [34]. Cela n'est pas le cas pour P du virus de la rage qui semble être active pour la transcription sous forme monomérique [35, 36].

Les différents domaines fonctionnels (fixation de N°, oligomérisation et interaction avec N-ARN) des phosphoprotéines des virus de Sendai et de la rage sont comparés dans la figure 7. Les régions qui peuvent être éliminées sans perte de fonctions sont également indiquées. Pour le virus de la rage, le domaine d'oligomérisation est situé entre les résidus 52 et 185 [23]. Dans ce cas, si l'oligomérisation se

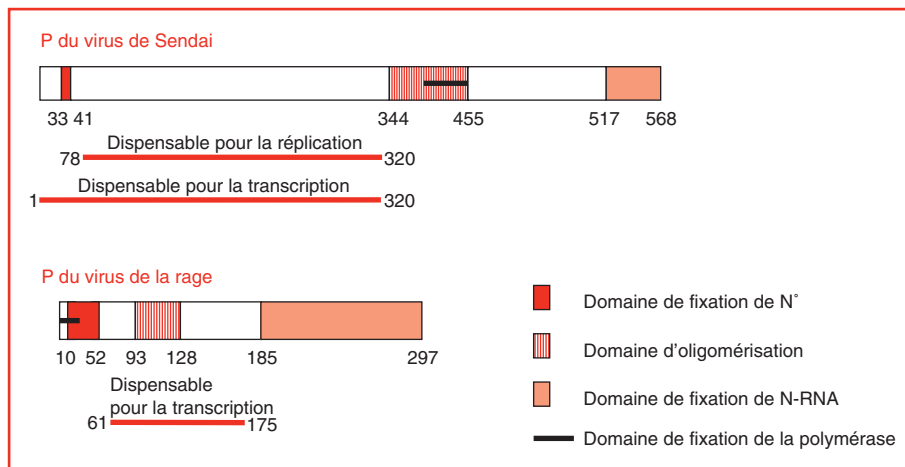


Figure 7. Organisation linéaire des phosphoprotéines du virus de Sendai et du virus de la rage et fonctions connues de ces domaines.

fait par formation d'un domaine constitué d'hélices torsadées, le domaine responsable de cette oligomérisation ne peut être situé, selon les prédictions d'hélices α , qu'entre les résidus 93 et 128. En comparaison avec la protéine P de Sendai, qui doit être sous une forme oligomérique pour être active, le domaine d'oligomérisation de P de la rage n'est pas nécessaire à l'activité de transcription [35]. La polymérase du virus de Sendai se fixe au niveau du domaine d'oligomérisation de P (figure 6), ce qui n'est pas le cas pour le virus de la rage. En plus, la protéine P de la rage se fixant sur la nucléocapside sous une forme monomérique, l'affinité de P du virus de la rage pour sa nucléocapside devrait être plus grande que celle du virus de Sendai [36]. À première vue, l'organisation des domaines des deux molécules de P (domaines de fixation de N°, d'oligomérisation et de fixation à N-ARN) semble suggérer une évolution similaire. Pourtant, les structures des domaines de fixation des deux protéines sur leurs nucléocapsides sont très différentes et il est très peu probable qu'elles aient un ancêtre commun [36].

De plus, la nature nous présentant toujours des solutions alternatives compliquant les règles générales, nous avons le cas du virus des oreillons. La phosphoprotéine de ce virus ne se fixe pas sur le domaine C-terminal de sa nucléoprotéine mais sur le domaine N-terminal (qui fixe aussi l'ARN) [30].

Les ARN polymérases ARN-dépendantes

Les polymérases des *Mononegavirales* sont de grandes molécules ayant un poids moléculaire qui varie de 250 à 300 kDa. Aucune structure de polymérase de virus à ARN de sens négatif, ARN-, n'a encore été résolue car ces polymérases sont très difficiles à purifier ou même à produire. Ces molécules ont des séquences communes, non seulement au sein des virus ARN- mais également, dans une certaine mesure, avec les polymérases des virus à ARN de sens positif ARN+ [37]. De nombreuses données structurales des polymérases des virus à ARN+ sont disponibles et il est certain qu'au moins une part de ces informations devrait aussi s'appliquer aux polymérases des virus à ARN- [32]. L'interaction de la polymérase avec la matrice N-ARN en général et entre la polymérase et la nucléoprotéine en particulier reste pour l'instant l'un des points obscurs de la transcription et de la réplication de ces virus. Les structures des ARN polymérases des virus à ARN+ et des transcriptases inverses des rétrovirus ne semblent pas montrer une architecture permettant une interaction avec une nucléoprotéine [38].

Les polymérases des virus à ARN+, comme le virus de la polio, forment des réseaux extensifs [39] qui induisent une coopération entre les unités enzymatiques. Les polyméras-

ses des *Paramyxoviridae* forment également des oligomères [40] indispensables à l'activité enzymatique.

Cibles pour molécules antivirales

Les nucléocapsides des virus à ARN- sont des structures uniques dans la biologie des acides nucléiques, de la réplication et de la transcription. Dans tous les autres cas connus, les matrices servant à la réplication ou à la transcription sont dépourvues de leurs protéines structurales et la polymérase chemine le long d'une matrice nue d'acide nucléique. Ces objets uniques dans le monde animal sont des cibles parfaites pour le développement de molécules antivirales dépourvues d'effet secondaire nocif. Les sites actifs de la polymérase, les interactions protéine-protéine comme N-P, P-P, P-L et L-L sont autant de très bonnes cibles potentielles pour des médicaments antiviraux. La mise au point de molécules modifiant de façon définitive les caractéristiques hélicoïdales des nucléocapsides et empêchant donc le cheminement de la polymérase le long de cette hélice est une seconde voie de recherche antivirale très prometteuse. Pour un tel développement, les structures atomiques des différents composants de la nucléocapside seront indispensables.

Références

1. 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.
2. Sanchez A, Kiley MP, Klenk HD, Feldmann H. Sequence analysis of the Marburg virus nucleoprotein gene : comparison to Ebola virus and the other non-segmented negative-strand RNA viruses. *J Gen Virol* 1992 ; 73 : 347-57.
3. 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.
4. 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.
5. Schoehn G, Iseni F, Mavrakis M, Blondel D, Ruigrok RWH. Structure of recombinant rabies virus N-RNA and identification of the phosphoprotein binding site. *J Virol* 2001 ; 75 : 490-8.
6. Iseni F, Baudin F, Blondel D, Ruigrok RWH. Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *RNA* 2000 ; 6 : 270-81.
7. Iseni F, Baudin F, Garcin D, Marq J-B, Ruigrok RWH, Kolakofsky D. Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids. *RNA* 2002 ; 8 : 1056-67.
8. Roux L. Dans le génome des paramyxovirinae, les promoteurs et leur activité sont façonnés par la règle des six. *Virologie* 2005 ; 9 : 19-34.
9. 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.
10. 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.

11. Flamand A, Raux H, Gaudin Y, Ruigrok RWH. Mechanisms of rabies virus neutralisation. *Virology* 1993 ; 194 : 302-13.
12. Mavrakis M, Kolesnikova L, Schoehn G, Becker S, Ruigrok RWH. Morphology of Marburg virus NP-RNA. *Virology* 2002 ; 296 : 300-7.
13. Calain P, Roux L. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *J Virol* 1993 ; 67 : 4822-30.
14. Spohner D, Kirn 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.
15. Fooks AR, Schadeck E, Liebert UG, *et al.* High-level expression of the measles virus nucleocapsid protein by using a replication-deficient adenovirus vector : induction of an MHC-I-restricted CTL response and protection in a murine model. *Virology* 1995 ; 210 : 456-65.
16. Iseni F, Barge A, Baudin F, Blondel D, Ruigrok RWH. Characterization of Rabies Virus nucleocapsids and recombinant nucleocapsid-like structures. *J Gen Virol* 1998 ; 79 : 2909-19.
17. Bhella D, Ralph A, Murphy LB, Yeo RP. Significant differences in nucleocapsid morphology within the Paramyxoviridae. *J Gen Virol* 2002 ; 83 : 1831-9.
18. Schoehn G, Mavrakis M, Albertini A, Wade R, Hoenger A, Ruigrok RWH. 12 Å Structure of trypsin-treated Measles Virus N-RNA. *J Mol Biol* 2004 ; 339 : 301-12.
19. Area E, Martin-Benito J, Gastaminza P, *et al.* 3D structure of the influenza virus polymerase complex : localization of subunit domains. *Proc Natl Acad Sci USA* 2004 ; 101 : 308-13.
20. Chen Z, Green TJ, Luo M, Li H. Visualizing the RNA molecule in the bacterially expressed vesicular stomatitis virus nucleoprotein-RNA complex. *Structure* 2004 ; 12 : 227-35.
21. Kouznetsov A, Buckle M, Tordo N. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. *J Gen Virol* 1998 ; 79 : 1005-13.
22. Klumpp K, Ruigrok RWH, Baudin F. Roles of the Influenza Virus Polymerase and Nucleoprotein in forming a functional RNP structure. *EMBO J* 1997 ; 16 : 1248-57.
23. Blondel D, Ruigrok R, Chelbi-Alix M, Tordo N. La protéine P du virus rabique : une protéine multifonctionnelle à l'interface entre le virus et son hôte. *Virologie* 2005 ; (sous presse).
24. 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.
25. Curran J, Marq JB, 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* 1995 ; 69 : 849-55.
26. 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.
27. Tarbouriech N, Curran J, Ruigrok RWH, Burmeister WP. Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nature Struct Biol* 2000 ; 7 : 777-81.
28. 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.
29. 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.
30. 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 USA* 2004 ; 101 : 8301-6.
31. Kingston RL, Baase WA, Gay LS. Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins. *J Virol* 2004 ; 78 : 8630-40.
32. 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.
33. Curran J, Kolakofsky D. Replication of paramyxoviruses. *Adv Virus Res* 1999 ; 54 : 403-22.
34. Gao Y, Lenard J. Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II. *EMBO J* 1995 ; 14 : 1240-7.
35. Jacob Y, Real E, Tordo N. Functional interaction map of lyssavirus phosphoprotein : identification of the minimal transcription domains. *J Virol* 2001 ; 75 : 9613-22.
36. Mavrakis M, McCarthy AA, Roche S, Blondel D, Ruigrok RWH. Structure and function of the C-terminal domain of the polymerase cofactor of Rabies virus. *J Mol Biol* 2004 ; 343 : 819-31.
37. Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* 1989 ; 8 : 3867-74.
38. Bressanelli S, Tomei L, Roussel A, *et al.* Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc Natl Acad Sci USA* 1999 ; 96 : 13034-9.
39. Lyle JM, Bullitt E, Bienz K, Kirkegaard K. Visualization and functional analysis of RNA-dependent RNA polymerase lattices. *Science* 2002 ; 296 : 2218-22.
40. Smallwood S, Cevik B, Moyer SA. Intragenic complementation and oligomerization of the L subunit of the sendai virus RNA polymerase. *Virology* 2002 ; 304 : 235-45.

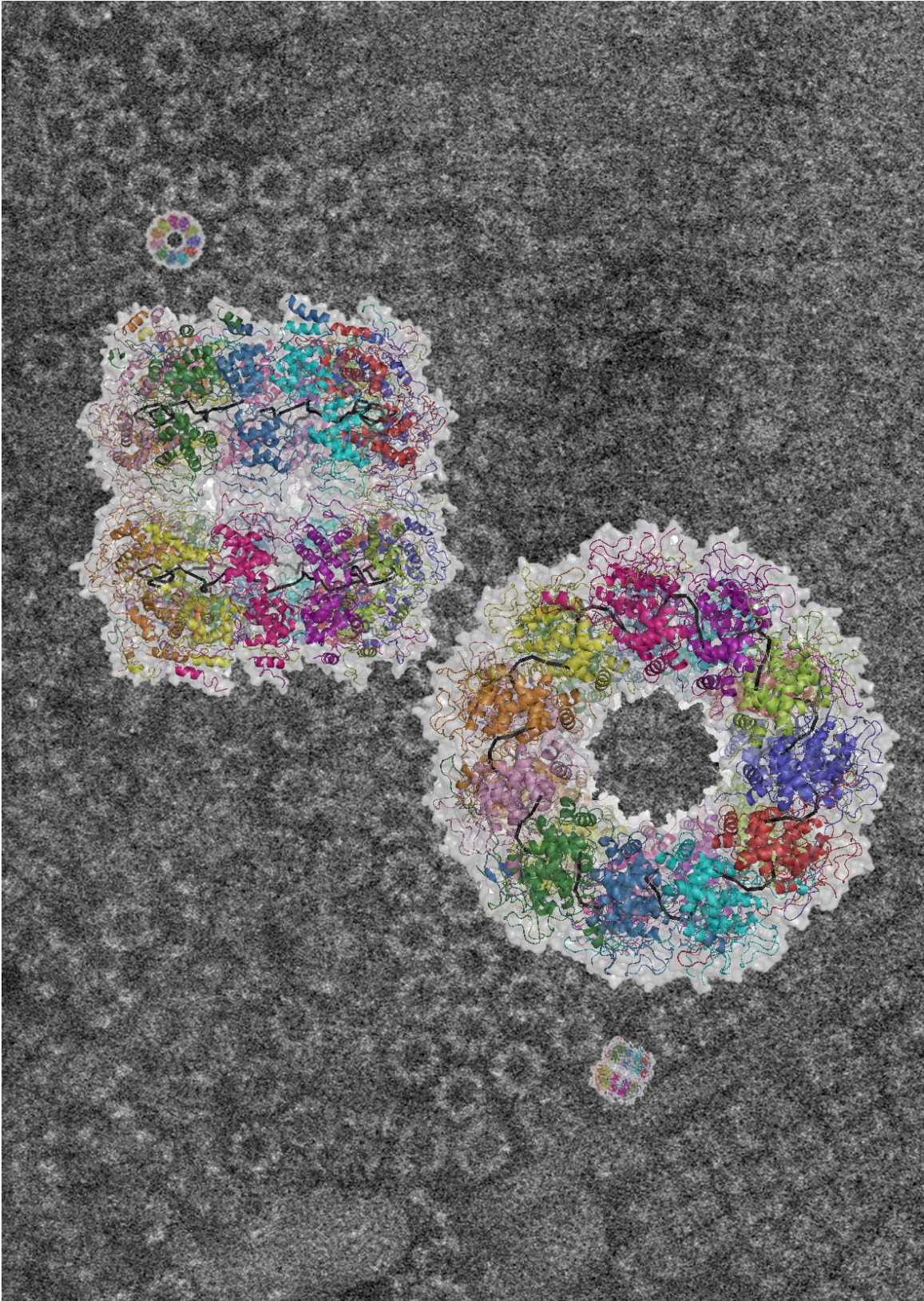


Figure 48 : Couverture du congrès “Thirteenth International Conference Negative Strand Viruses 2006”.

Fond : Cliché de microscopie électronique en coloration négative de cristaux d’anneaux à 11 sous-unités de nucléoprotéines broyés.

Premier plan : structure cristallographique des anneaux à 11 sous-unités de nucléoprotéines du virus de la rage.

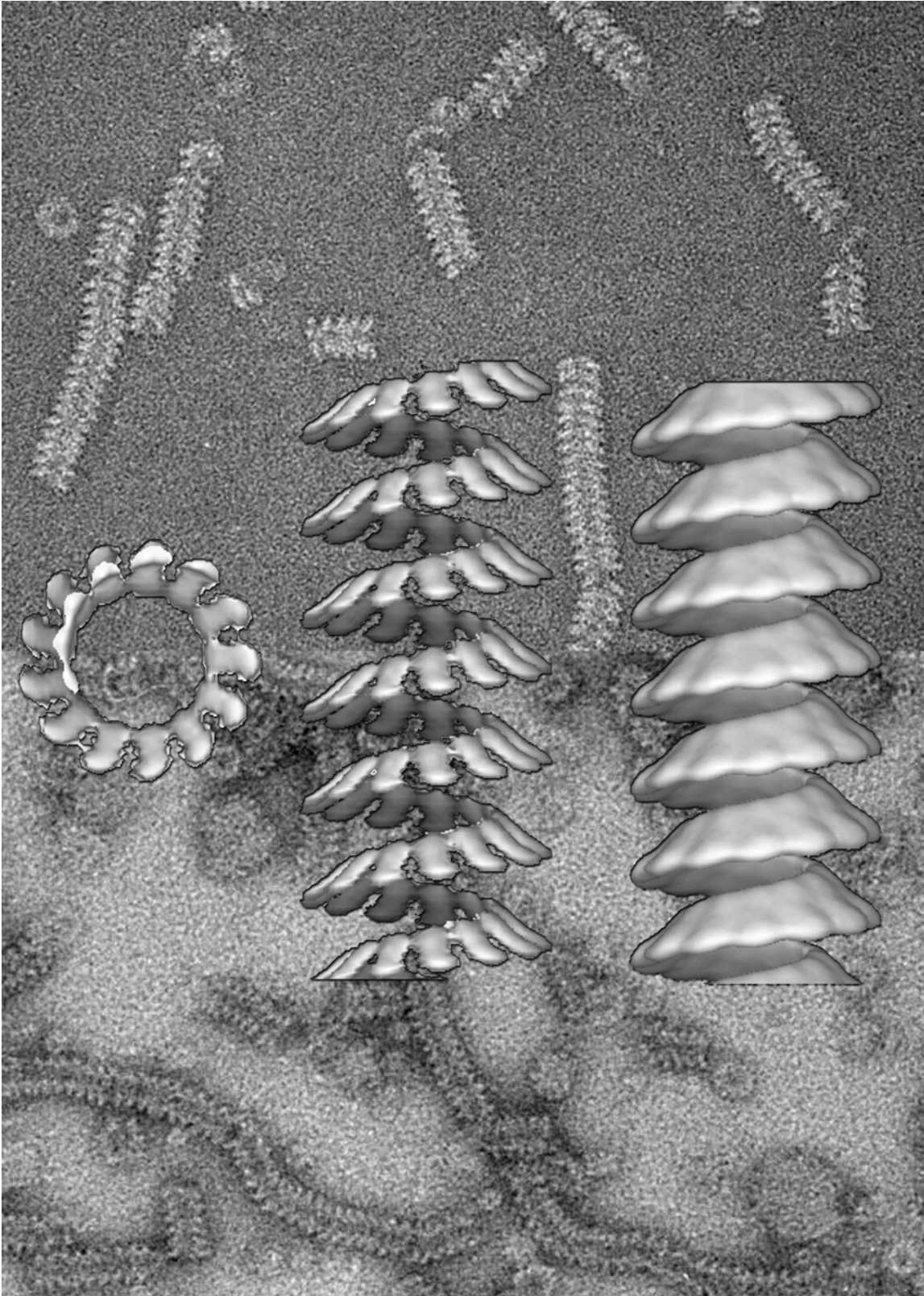


Figure 49 : Twelfth International Conference Negative Strand Viruses 2003.

Micrographies de complexes Nucléoprotéine-ARN recombinants du virus Nipah colorés négativement. Fond, partie inférieure : complexes N-ARN intacts. Fond, partie supérieure : complexes N-ARN après traitement protéasique.

Premier plan : Reconstructions 3D à basse résolution de la nucléocapside recombinante de Nipah traitée à l'élatase. De gauche à droite : vue de dessus de l'hélice, hélice représentée avec deux valeurs différentes de seuils de densité.