



**HAL**  
open science

# Caractérisations biophysiques et structurales du complexe de réplication des Rhabdoviridae

Francine C. A. Gerard

► **To cite this version:**

Francine C. A. Gerard. Caractérisations biophysiques et structurales du complexe de réplication des Rhabdoviridae. Biochimie [q-bio.BM]. Université Joseph-Fourier - Grenoble I, 2008. Français. NNT : . tel-00343710

**HAL Id: tel-00343710**

**<https://theses.hal.science/tel-00343710>**

Submitted on 2 Dec 2008

**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.

*Ecole Doctorale Chimie et Sciences du Vivant*

**THESE**

Pour l'obtention du diplôme de  
**Docteur de l'Université Joseph Fourier – Grenoble 1**

Spécialité : Biologie structurale et nanobiologie

Présentée et soutenue publiquement par  
**Francine GERARD**

Le vendredi 28 novembre 2008

---

**CARACTERISATIONS BIOPHYSIQUE ET STRUCTURALE  
DU COMPLEXE DE REPLICATION DES *RHABDOVIRIDAE***

---

Devant le jury composé de :

**Dr. José A. Melero**  
**Pr. Daniel Kolakosfsky**

Rapporteur  
Rapporteur

**Dr. Danielle Blondel**  
**Dr. Noël Tordo**  
**Pr. Uwe Schlattner**

Examineur  
Examineur  
Examineur

**Pr. Rob Ruigrok**  
**Pr. Marc Jamin**

Co-directeur de thèse  
Co-directeur de thèse

Thèse réalisée à l'**UVHCI** (Unit of **V**irus **H**ost **C**ell Interaction UMR 5233 UJF-EMBL-CNRS)  
D'octobre 2005 à septembre 2008.

## Remerciements

L'histoire d'une vie peut parfois se résumer à des rencontres et à des opportunités. Mon désir de devenir chercheur s'est très largement développé lors de mon arrivée à l'Université, mais je ne remercierai jamais assez le Pr. Marc JAMIN de m'avoir baigné dans le milieu de la recherche. J'ai eu l'opportunité de réaliser un stage dès ma licence, et de rester constamment connectée à un laboratoire depuis plus de 5 ans maintenant. Cet environnement a été décisif pour moi quant à mes espérances de carrière.

J'ai découvert le monde de la virologie grâce au Pr. Rob RUIGROK et au Pr. Patrice MORAND, qui, dans un module proposé en Master 1 à l'Université Joseph Fourier (Grenoble), enseignent avec habileté et grand intérêt les aspects structuraux et fondamentaux, mais aussi l'aspect médical de la virologie.

Ma chance a été de pouvoir effectuer ma thèse sous la direction des Prs. Rob RUIGROK et Marc JAMIN, où j'ai pu combiner les approches biophysiques de la recherche fondamentale avec la virologie. Je tiens à leur témoigner toute ma gratitude pour m'avoir dirigé dans mes recherches. Je les remercie pour leur patience, leur soutien pendant ces toutes ces années de travail, pour leur énergie et aussi pour la transmission de leur savoir.

Marc, merci d'avoir suivi et aidé à mon évolution d'étudiante à doctorante, merci de m'avoir toujours soutenue, même pendant des moments difficiles.

Rob, ta passion pour les virus est une vraie chance pour nous tous, l'encadrement et l'intérêt que tu prête à tous tes étudiants est incroyable.

Je remercie le Pr. Dan KOLAKOFSKY et le Dr. José MELERO d'avoir accepté d'être les rapporteurs de ce travail. Je suis aussi extrêmement reconnaissante envers le Dr. Danielle BLONDEL, le Pr. Uwe SCHLATTNER et le Dr. Noël TORDO pour leur participation à mon jury de thèse.

Je remercie le directeur de l'EMBL Grenoble et directeur de l'UVHCI, le Dr. Stephen Cusak pour m'avoir accueilli lors de mon arrivée en Master 2 à l'EMBL, où les membres de l'IVMS étaient depuis peu « *visitor EMBL* », mais aussi pour l'environnement scientifique qui règne dans le CIBB, regroupant l'EMBL I'ILL, l'ESRF, et l'IBS. C'est un environnement propice aux développements de collaborations, à l'éveil scientifique et à la diversité des techniques utilisables.

Je tiens également à remercier nos collaborateurs :

- Dr. Danielle BLONDEL et Dr. Yves GAUDIN au laboratoire de Virologie Moléculaire et Structurale à Gif-sur-Yvette (plasmides et échantillons de nucléocapsides virales),
- Jean-Luc GARDEN et Dr. Hervé GUILLOU au Centre de Recherches sur les Très Basses Températures (CNRS, UJF et INPG) à Grenoble (expériences de DSC),
- Dr. Christine EBEL à l'Institut de Biologie Structurale à Grenoble (Ultracentrifugation analytique),
- Dr. Guiseppe ZACCAI à l'Institut Laue Langevin à Grenoble (SANS).
- Bertrand Lefèbvre, Denis Daveloos et Bernard Brasmes du CRSSA à Grenoble (spectrométrie de masse),
- Adrien FAVIER, Dr. Martin BLACKLEDGE au laboratoire de RMN à l'Institut de Biologie Structurale à Grenoble,
- Dr. Noël TORDO, Dr. Bernadette HEYD et Guillaume CASTEL de l'Institut Pasteur à Paris (collaboration projet ANR, ANRAGE),
- Pr. Pascal DUMY du département de chimie moléculaire de Grenoble (collaboration projet ANR, ANRAGE, synthèse de peptides),
- Dr. Nicole THIELENS à l'Institut de Biologie Structurale à Grenoble (BIAcore),
- Dr. Carlos RAMOS du Laboratorio Nacional de Luz Sincrotron à Campinas, Brésil.

Je voudrais dire un grand merci au Dr. Aurélie ALBERTINI, qui m'a coaché lors de mon arrivée à l'IVMS. Sans tous ses conseils, je n'aurais jamais fait de biochimie de qualité. Aurélie, je tiens à te dire merci pour ton soutien permanent, même si quelques kms nous séparent. Ton amitié est précieuse.

Un grand merci au Dr. Euripedes RIBEIRO, qui a totalement boosté les projets de recherche initiés au laboratoire et qui a pu me guider pendant ces deux dernières années. Ta bonne humeur et ton sourire brésilien nous ont très largement conquis !

Je voudrais remercier très chaleureusement nos microscopistes électroniques, le Dr. Irina GUSTCHE et le Dr. Guy SCHOEHN, qui ont tous deux regardé nombre de mes échantillons. Guy, merci pour ta bonne humeur

légendaire (non, mais tu fais un peu peur aux gens, mais tu es un vrai gentil, je le sais...) et Irina, merci pour toutes ces conversations professionnelles et personnelles...surtout merci d'avoir regardé tous ces échantillons d'anneaux.

Merci au Dr. Nicolas TARBOURIECH, qui a été mon « *coach cristallographie* ». Merci à toi Nicolas pour ta gentillesse, pour tous ces conseils, pour ta patience avec ceux qui « ont du mal avec la cristallographie ». Merci pour m'avoir emmené sur les lignes, même le soir où Annabelle allait donner naissance à votre « *first DNA experiment!* ». Merci pour tout le temps que tu m'as accordé pour qu'on puisse traiter les données...

Je tiens à remercier mes deux voisins de paillasse, qui ont très largement facilité ma vie dans le laboratoire : Sébastien Boulo, mon binôme, et Julien Perard, avec qui j'ai passé de supers moments.

Je tiens à remercier tous les membres du laboratoire, que ce soit les anciens de l'EMBL ou les nouveaux IVMS/UVHCI, merci à vous tous. Je tiens à remercier plus particulièrement Céline, Majida, Lucy, Marlyse, Alex, Priscilla, Julien, Pierre-Alain, Thibault, Jeanne, Sophie, Cédric, Cyril, Monique, Hatice, Florence, Ivan, Xavier, Denis...et bien d'autres...

Je tiens à remercier le personnel de l'EMBL et de l'UVHCI qui nous rend la vie si simple : merci à Annie, Jocelyne, Virginie, Françoise, Sandrine, Christelle, Pierre, Lionel, José, Benjamin, Aymeric, Franck...

Je voudrais aussi dire merci à ma famille et mes amis, qui m'ont toujours soutenue. Merci à vous de m'avoir encouragée même si les protéines et les virus, ça ne vous parle pas plus que ça.

Merci à mes parents, Conchita et Georges, qui m'ont laissé le choix de faire des études, de longues études et qui m'ont toujours aidée dans mes démarches. Merci à vous deux du fond du cœur.

Je tiens également à remercier ma grand-mère, 96 ans, qui résiste toujours et encore à la vie, même après les souffrances qu'elle a endurées. *Abuela*, tu resteras toujours un modèle de fierté pour moi.

Enfin, un merci du plus profond de mon cœur à mon chéri, et futur époux, qui me soutient depuis près de 10 ans maintenant, et qui part à l'aventure canadienne avec moi...

A mes parents, Conchita et Georges,  
A mi abuela, Ana,  
A Sébastien.

# TABLE DES MATIERES

<b>TABLE DES MATIERES</b>	<b>1</b>
<b>ABREVIATIONS</b>	<b>5</b>
<b>ABSTRACT</b>	<b>7</b>
<b>CHAPITRE I : INTRODUCTION</b>	<b>8</b>
<b>I – INTRODUCTION GENERALE</b>	<b>9</b>
A – LA VIROLOGIE, DES ORIGINES A NOS JOURS	9
B – LES VIRUS A ARN NEGATIF	12
C – LES <i>RHABDOVIRIDAE</i>	13
1 - Classification	13
2 – la rage	15
3 – Le virus de la stomatite vésiculaire (VSV)	20
<b>II – LA TRANSCRIPTION ET LA REPLICATION CHEZ LES <i>RHABDOVIRIDAE</i></b>	<b>23</b>
A – MORPHOLOGIE ET COMPOSITION DES <i>RHABDOVIRIDAE</i>	23
B – GENOME DES <i>RHABDOVIRIDAE</i>	25
C – LE CYCLE VIRAL	27
D – STRUCTURES ET FONCTIONS DES PROTEINES VIRALES	29
1 – Protéines impliquées dans le bourgeonnement	29
La Glycoprotéine	29
La protéine Matrice (M)	33
2 – Protéines impliquées dans la multiplication du virus et la synthèse d'ARN	35
L'ARN polymérase ARN-dépendante (L pour Large protéine)	35
La Nucléoprotéine (N)	36
La Phosphoprotéine (P)	43
E – LA TRANSCRIPTION VIRALE	48
F – LA REPLICATION VIRALE	52
<b>III – OBJECTIFS DU TRAVAIL DE THESE</b>	<b>53</b>
<b>CHAPITRE II : MATERIELS ET METHODES DE CARACTERISATION DES PROTEINES DE <i>RHABDOVIRIDAE</i></b>	<b>55</b>
<b>I – CLONAGE, EXPRESSION ET PURIFICATION DES PROTEINES DE <i>RHABDOVIRIDAE</i> ETUDIEES</b>	<b>56</b>
A - RECAPITULATIF DES PROTEINES DE <i>RHABDOVIRIDAE</i> ETUDIEES	56
B – CLONAGE, EXPRESSION ET PURIFICATION DE LA PHOSPHOPROTEINE DU VIRUS DE LA RAGE (P-MBP) ET DE SON DOMAINE C-TERMINAL (CTERP-MBP) EN FUSION AVEC LA MBP ( <i>MALTOSE BINDING PROTEIN</i> )	58
1 – Clonage et expression des phosphoprotéines fusions P-MBP et CterP-MBP	58
2 - Purification des phosphoprotéines fusions P-MBP et CterP-MBP	59
C – EXPRESSION ET PURIFICATION DES COMPLEXES NUCLEOCAPSIDES VIRALES DU VIRUS DE LA RAGE (NC RV) ET DU VIRUS DE LA STOMATITE VESICULAIRE (NC VSV)	60
<b>II – METHODES DE CARACTERISATION DES PROTEINES</b>	<b>61</b>
A – ELECTROPHORESE	61
1 – Gel d'électrophorèse en conditions dénaturantes	61
2 – Gel d'électrophorèse en conditions natives	61
B – SPECTROSCOPIE D'ABSORBANCE	61
C – DICHROÏSME CIRCULAIRE (CD)	62

1 – Principe	62
2 – Analyse des courbes de dénaturation	63
D – SPECTROSCOPIE DE FLUORESCENCE	65
1– Fluorescence intrinsèque des résidus Trp et Tyr	65
2– Anisotropie de fluorescence	65
E – CALORIMETRIE A BALAYAGE DIFFERENTIEL (DSC)	67
F – DIFFUSION DE LUMIERE MULTI-ANGLES COUPLEE A LA CHROMATOGRAPHIE D'EXCLUSION DE TAILLE ET A LA REFRACTOMETRIE (SEC-MALLS-RI)	68
G – RESONANCE PLASMONIQUE DE SURFACE (SPR)	71
1– Principe	71
2– Préparation des puces	72
3 – Enregistrement des données	73
4– Analyse des données	74
<b>III – ETUDES STRUCTURALES</b>	<b>75</b>
A – DIFFUSION AUX PETITS ANGLES ( <i>SMALL ANGLE SCATTERING</i> )	75
1 – Introduction à la diffusion aux petits angles	75
2 – Détermination des paramètres structuraux	76
3 – Collecte et traitement des données	78
4 – Modélisations <i>ab initio</i>	80
5 – Modélisation de la phosphoprotéine du virus de la stomatite vésiculaire ( <i>P VSV</i> ) par reconstruction à partir de données structurales partielles	81
6 – Superposition des modèles	82
B – CRISTALLOGRAPHIE DES RAYONS X	82
1 – Cristallisation	82
2 – Traitement des données	83
C – MICROSCOPIE ELECTRONIQUE	84
1 – Préparations des échantillons	84
2 – Reconstruction tridimensionnelle de molécules isolées à partir d'images prise en microscopie électronique	86

### **CHAPITRE III : CARACTERISATIONS BIOPHYSIQUES ET STRUCTURALES DES PHOSPHOPROTEINES DE RHABDOVIRIDAE** **88**

<b>I – LES PHOSPHOPROTEINES DES RHABDOVIRIDAE SONT DES DIMERES ALLONGES EN SOLUTION</b>	<b>89</b>
A – INTRODUCTION	89
B – ARTICLE 1: UNPHOSPHORYLATED <i>RHABDOVIRIDAE</i> PHOSPHOPROTEINS FORMS ELONGATED DIMERS IN SOLUTION	90
C – CONCLUSION	91
<b>II – LA PHOSPHOPROTEINE DU VIRUS DE LA RAGE EST MODULAIRE</b>	<b>92</b>
A – INTRODUCTION	92
B – ARTICLE 2: MODULAR ORGANIZATION OF RABIES VIRUS PHOSPHOPROTEIN	93
C – CONCLUSION	94
<b>III – STRUCTURE PAR RESONANCE MAGNETIQUE NUCLEAIRE (RMN) DU DOMAINE C-TERMINAL DE LA PHOSPHOPROTEINE DU VIRUS DE LA STOMATITE VESICULAIRE (VSV)</b>	<b>96</b>
A – INTRODUCTION	96
B – ARTICLE 3: SOLUTION STRUCTURE OF THE C-TERMINAL N-RNA BINDING DOMAIN OF THE VESICULAR STOMATITIS VIRUS PHOSPHOPROTEIN	97
C – DONNEES SUPPLEMENTAIRES	97
D – CONCLUSION	98
<b>IV – ANALYSE DE LA STABILITE DE LA PHOSPHOPROTEINE DU VIRUS DE LA STOMATITE VESICULAIRE (P VSV) ET DU VIRUS DE LA RAGE (P RV)</b>	<b>99</b>
A – SPECTROSCOPIE DE DICHROÏSME CIRCULAIRE (CD) ET DE FLUORESCENCE	99
B – TRANSITION DE DENATURATION INDUITE PAR L'UREE	100

C - DENATURATION THERMIQUE	101
D - CALORIMETRIE DIFFERENTIELLE A BALAYAGE (DSC)	102
E - CHANGEMENT DE SURFACE ACCESSIBLE A L'AGENT DENATURANT	103
F - CONCLUSIONS	104
<b>V – ETUDES STRUCTURALES DES PHOSPHOPROTEINES DU VIRUS DE LA STOMATITE VESICULAIRE (P VSV) ET DU VIRUS DE LA RAGE (P RV)</b>	<b>105</b>
A – EXTRACTION DES PARAMETRES STRUCTURAUX	106
B – MODELISATIONS <i>AB INITIO</i>	107
1 – <i>Modélisations ab initio de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV)</i>	107
2 – <i>Modélisations ab initio de la phosphoprotéine du virus de la rage (P RV)</i>	110
C – MODELISATION DE LA PHOSPHOPROTEINE DU VIRUS DE LA STOMATITE VESICULAIRE (P VSV) AVEC LE LOGICIEL BUNCH	112
D – CONCLUSION	114

## **CHAPITRE IV : BASES STRUCTURALES DE L'INTERACTION ENTRE LE COMPLEXE NUCLEOPROTEINE – ARN ET LA PHOSPHOPROTEINE** **115**

<b>I – INTERACTION ENTRE LES ANNEAUX NUCLEOPROTEINE-ARN ET LA PHOSPHOPROTEINE</b>	<b>116</b>
A – CARACTERISATIONS DES COMPLEXES ANNEAUX (NUCLEOPROTEINE-ARN) ET DE LA PHOSPHOPROTEINE PAR SEC-MALLS-RI	116
1 – <i>Caractérisation des anneaux N9, N10 et N11 par SEC-MALLS-RI</i>	116
2 – <i>Stoechiométrie des complexes anneaux-phosphoprotéine par SEC-MALLS-RI</i>	118
3 – <i>Stoechiométrie des complexes anneaux-CterP RV par SEC-MALLS-RI</i>	120
4 – <i>Stoechiométrie des complexes anneaux-P<sub>Δ91-131</sub> par SEC-MALLS-RI</i>	122
5 – <i>Conclusions</i>	123
B – ANALYSE DES COMPLEXES ANNEAUX N-ARN ET PHOSPHOPROTEINE PAR ELECTROPHORESE EN CONDITIONS NATIVES	123
1 – <i>Complexes N10-phosphoprotéine</i>	123
2 – <i>Complexes N10-domaine C-terminal de la phosphoprotéine</i>	124
3 – <i>Complexes N10-P<sub>Δ91-131</sub></i>	125
4 – <i>Conclusions</i>	125
C – ETUDES DES COMPLEXES N-ARN-P ET N-ARN-CTERP PAR SAXS	126
1 – <i>Données de diffusions des rayons X</i>	126
2 – <i>Modélisations ab initio des anneaux</i>	129
3 – <i>Superposition de la structure cristallographique des N11 avec le modèle filtré des N11 (damfilt)</i>	132
4 – <i>Modélisations ab initio des complexes N11, N11-phosphoprotéine et N11-domaine C-terminal de la phosphoprotéine</i>	132
D – ESSAI DE CO-CRISTALLISATION	134
1 – <i>Trempage de cristaux de N11 et N10</i>	134
2 – <i>Co-cristallisation des anneaux N10 et N11 avec le domaine C-terminal de la phosphoprotéine et le mutant P<sub>Δ91-131</sub></i>	138
E – ANALYSE EN MICROSCOPIE ELECTRONIQUE DU COMPLEXE ANNEAU DE NUCLEOPROTEINE-ARN ET LA PHOSPHOPROTEINE	139
1 – <i>Microscopie électronique sur des N9, N10 et N11 et reconstruction d'images</i>	140
2 – <i>Microscopie électronique sur le complexes N9-P RV et reconstruction d'images</i>	141
3 – <i>Protéines fusions à la Maltose Binding Protein (MBP)</i>	143
4 – <i>Electrophorèse en conditions natives des complexes N10-P-MBP et N10-CterP-MBP</i>	146
5 – <i>Microscopie électronique du complexe N11-P-MBP et reconstruction d'images</i>	147
6 – <i>Conclusions</i>	149
F – DETERMINATION DE L'AFFINITE DU COMPLEXE N-ARN-P	150
1 – <i>Expériences de résonance plasmonique de surface sur le complexe anneaux-phosphoprotéine</i>	150



2 – Analyse des données	151
3 - Conclusion	153
G – CONCLUSIONS	153
<b>II – INTERACTION ENTRE LES NUCLEOCAPSIDES VIRALES (NC) ET LA PHOSPHOPROTEINE (P)</b>	<b>154</b>
A – CARACTERISATION STRUCTURALE DES NUCLEOCAPSIDES VIRALES (NC VSV ET NC RV)	154
B – COMPACTION DES NUCLEOCAPSIDES VIRALES INDUITE PAR LA PHOSPHOPROTEINE	156
C – AFFINITE DE LA PHOSPHOPROTEINE ET DE SON DOMAINE C-TERMINAL POUR LES NUCLEOCAPSIDES (NC)	158
1 – Anisotropie de fluorescence	158
2 – Résonance plasmonique de surface	160
D – CONCLUSIONS	161
<b>CHAPITRE V : DISCUSSION</b>	<b>162</b>
<i>I – ORGANISATION MODULAIRE DE LA PHOSPHOPROTEINE DES RHABDOVIRIDAE</i>	<b>163</b>
<i>II – INTERACTION DE LA PHOSPHOPROTEINE ET DES ANNEAUX DE NUCLEOPROTEINES-ARN</i>	<b>165</b>
<i>III – INTERACTION DE LA PHOSPHOPROTEINE ET DES NUCLEOCAPSIDES</i>	<b>167</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b>	<b>168</b>
<b>REFERENCES ELECTRONIQUES</b>	<b>181</b>
<b>ARTICLES SUPPLEMENTAIRES</b>	<b>182</b>

## ABREVIATIONS

2D : deux dimensionnel  
3D : trois dimensionnel  
Å : Angström  
aa : acide aminé  
ABL : *Australian Bat Lyssavirus*  
ADN : Acide DésoxyriboNucléique  
ARN : Acide RiboNucléique  
ARNm : ARN messenger  
AUC: *Analytical UltraCentrifugation*  
CD: Dichroïsme Circulaire  
CK-II: Caséine Kinase II  
CVS: *Challenge Virus Strain*  
Da / kDa : Dalton / kiloDalton  
EBL: *European Bat Lyssavirus*  
*E. coli*: *Escherichia coli*  
EDTA : *Ethylene Diamine Tetraacetic Acid*  
ESRF : *European Synchrotron Radiation Facility*  
G: Glycoprotéine  
H: heure  
HF: HighFive®, dérivés de la lignée parentale de cellules *Trichopulsia ni*. (Invitrogen)  
ICTV : *International Committee on Taxonomy of Viruses*  
kb : kilobases  
« Large » protéine: ARN polymérase ARN dépendante virale  
M / mM: Molaire / milliMolaire  
M: protéine Matrice  
MALLS : *Multi Angle Laser Light Scattering*  
*moi*: *multiplicity of infection*  
MM: Masse Moléculaire  
nm : nanomètre.  
nt : nucléotide  
N : Nucléoprotéine  
OMS : Organisation Mondiale de la Santé  
P : Phosphoprotéine  
*pi* : *post infection*  
PAGE : *PolyAcrylamide Gel Electrophoresis*  
PBS : *Phosphate Buffered Saline*  
PCR: *Polymerase Chain Reaction*  
PDB : *Protein Data Bank*  
PEG : *Polyethylene Glycol*  
*pfu*: unité formant une plage de lyse  
pH : potentiel Hydrogène  
pI : point Isoélectrique  
PM : Poids Moléculaire  
RER : réticulum endoplasmique rugueux  
Rg : Rayon de gyration  
RI : réfractométrie  
RMN: Résonance Magnétique Nucléaire  
RNP: RiboNucléoProtéine ou RiboNucléocapside  
*rmsd*: *root mean square deviation*  
rpm: tour par minute  
RSV : virus Respiratoire Syncytial humain

RV : virus de la rage  
SANS: *Small Angle Neutron Scattering*  
SAXS: *Small Angle X-ray Scattering*  
Sf21: *Spodoptera frugiperda* (Invitrogen #0)  
SDS: Sodium Docecyle Sulphate  
SEC: *Size Exclusion Chromatography*  
SPR : résonance plasmonique de surface (*surface plasmon resonance*)  
SNC: Système Nerveux Central  
TBE : Tris Borate EDTA  
UV : Ultraviolet  
VSV : virus de la stomatite vésiculaire  
v/v : volume/volume  
w/w : weight/weight (poids/poids)

Code à une et trois lettres pour les acides aminés :

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

## ABSTRACT

### ***Biophysical and structural studies of the replicative complex of Rhabdoviridae***

*Vesicular stomatitis virus (Vesiculovirus, VSV) and rabies virus (Lyssavirus, RV) belong to the Rhabdoviridae family. VSV serves as a model for studying the multiplication of viruses from the Mononegavirales order, whereas rabies is still a serious health problem in developing countries where it kills more than 50 000 people per year. The genome of these viruses is composed of one single stranded RNA molecule of negative polarity, and encodes five proteins. The nucleoprotein (N) binds to the viral genome forming an N-RNA complex that is packaged into virions that serves as a matrix for viral transcription and replication. The phosphoprotein (P) is a cofactor of the viral RNA-dependant RNA polymerase (L), and a chaperone of the nucleoprotein. The C-terminal domain of P binds to the N-RNA matrix and its N-terminal domain binds to the polymerase, making a physical link between the viral genome and the polymerase. The stoichiometry, the structure and the exact functions of P oligomers are controversial or unknown. The aim of this work was to undertake a structural and biophysical characterization of the phosphoprotein and of the complexes that it forms with the N-RNA complexes in order to understand the dynamic of the replicative complex of these viruses.*

*Hydrodynamic radius and molecular mass of the RV and VSV P, measured by SEC-MALLS-RI ((Size Exclusion Chromatography coupled with Multi Angle Laser Light Scattering and refractometry), show that these proteins behave as elongated dimers in solution. Bioinformatics analysis of their amino acid sequences indicates a modular organisation of P, similar to the one observed for Paramyxovirus phosphoprotein. Biochemical and structural studies of RV P mutants confirmed these results. Structure of the C-terminal domain of VSV P was solved by NMR showing a homology with the C-terminal domain of RV P. Ab initio modeling using SAXS (Small Angle X-ray Scattering) data confirmed the elongated shape of the phosphoprotein.*

*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. SEC-MALLS-RI and native electrophoresis analysis of N-RNA-P complexes showed two forms of N-RNA-P complexes, with one and two dimers of phosphoprotein. Binding of P to N-RNA complexes is mediated by one C-terminal domain of P in each dimer. An electron microscopy study allows us to visualise an additional electron density that could correspond to the phosphoprotein. Structural analysis with electron microscopy of the nucleocapsid-phosphoprotein complexes revealed an important conformational change.*

*In the nucleocapsid context, the viral genome must dissociate locally from the nucleoprotein to allow the viral polymerase to access to the genome information. This specific process may exist for other viruses such as measles virus, Ebola virus or influenza virus, also responsible for human diseases. The interaction between the phosphoprotein and the N-RNA complex is an interesting target for drug design as it only exists in viruses.*

**Keywords:** *vesicular stomatitis virus (VSV), rabies virus (RV), negative strand RNA viruses, nucleoprotein, phosphoprotein, RNA protection, Multi Angle Laser Light Scattering coupled with Size Exclusion Chromatography and refractometry (MALLS-SEC-RI), Small Angle Scattering (SAS), ab initio modelisation, electron microscopy, fluorescence anisotropy, surface plasmon resonance (SPR).*

## **CHAPITRE I : INTRODUCTION**

## I – Introduction générale

### A – La virologie, des origines à nos jours

Les premières traces de maladies infectieuses, notamment de maladies virales telles que la rage, la poliomyélite, les oreillons, le zona ou la variole, remontent à l'antiquité même si les agents pathologiques eux-mêmes n'avaient pas été découverts (figure 1). Les premières vaccinations ont été réalisées sans aucune connaissance du type d'agent infectieux responsable de la maladie, en 1796 par Edward JENNER pour la variole, puis en 1885 par Louis PASTEUR pour la rage (Cann, 1997). Dimitri L. IWANOVSKI et Martinus W. BEIRJERINCK, en 1892 et 1898, furent les premiers à montrer de façon indépendante que l'agent pathogène responsable de la mosaïque du tabac était un « agent filtrable » plus petit qu'une bactérie. En 1898, LÖFFLER et FROSCHE découvrent le caractère « filtrant » de la fièvre aphteuse, une épizootie redoutable chez les bovins.



**Figure 1** - Toile exposée dans la *Freer Gallery of Art, Washington DC*, décrivant le mode de contamination de la rage. Traduction : « On dit que des personnes ont vu un ou deux hommes mordus, ce qui leur a été fatal » (Dr. Noél TORDO).

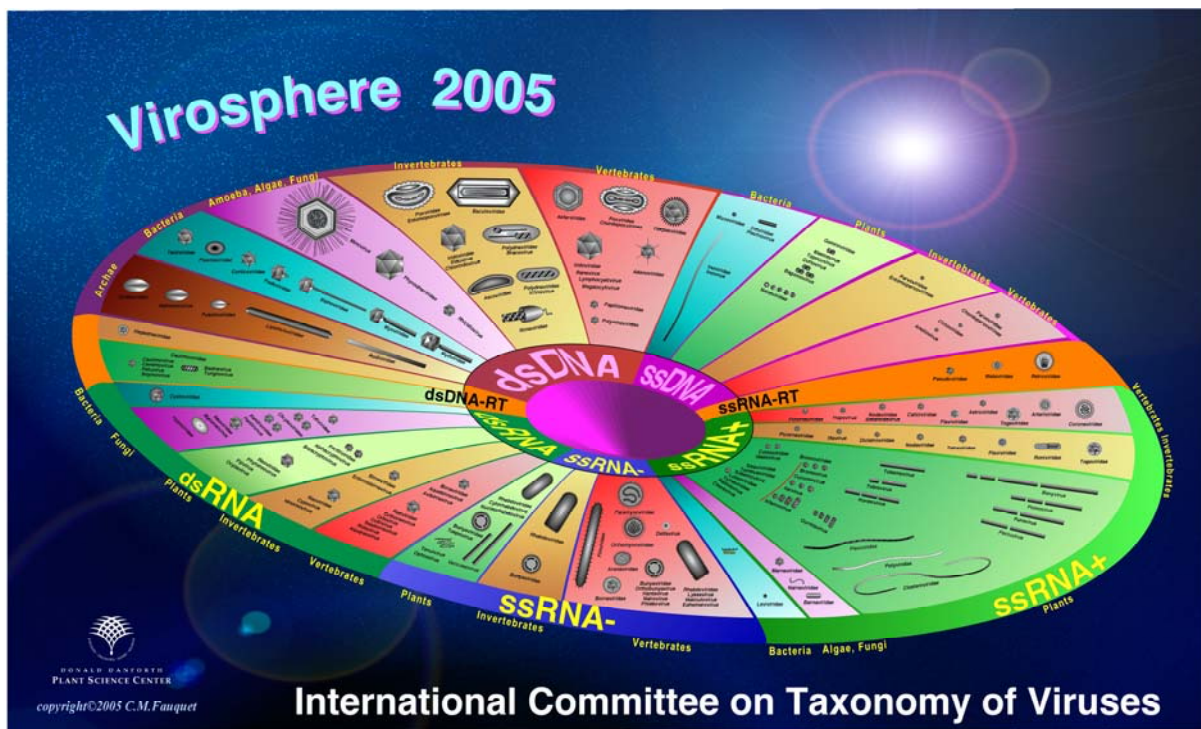
Le terme virus est issu du latin et il signifie « poison ». Louis PASTEUR décrit à l'époque (XIX<sup>ième</sup> siècle) ces « agents filtrables », puis il utilise plus spécifiquement le terme de « virus filtrables » pour les agents infectieux qui ne rentrent pas dans les autres catégories (bactéries, champignons, parasites). C'est André LWOFF qui en 1953 donne une définition du virion:

- le virion ne possède qu'un seul type d'acide nucléique : soit ADN, soit ARN,
- il se reproduit uniquement à partir de son matériel génétique,
- il est incapable de croître et de se diviser seul,
- son génome n'a aucune information pour la synthèse d'enzymes du métabolisme intermédiaire,
- c'est un parasite absolu de la cellule hôte.

Cependant, cette définition pourrait être remise en cause avec la découverte récente des *Mimivirus* (*Mimicking microbe virus*) (Suzan-Monti & al., 2006) (Claverie & al., 2006). Ces virus possèdent le génome le plus grand parmi les virus à ADN et celui-ci code pour des enzymes du métabolisme intermédiaire. Seules deux de ces affirmations restent vraies, le virus est un parasite absolu de la cellule hôte qui est incapable de se multiplier par division.

L'histoire de la virologie est fortement liée au développement de la microscopie électronique dans les années 1930 (Ernst RUSKA). La mise au point des techniques de visualisations, notamment la coloration négative a permis, dans les années 1960 de visualiser directement les agents infectieux (Horne & Wildy 1963). Leur structure et leur morphologie ont alors pu être analysées. Ainsi, une première classification a pu être élaborée, selon la nature de l'acide nucléique, la symétrie de la capsid, la présence ou non d'une enveloppe, la taille du virion et de la capsid (McKinley, 1932). Les progrès fondamentaux réalisés dans la mise au point des techniques de cultures de virus, chez l'animal par Louis PASTEUR, puis sur les embryons de poulet par Max THEILER, et enfin sur les cultures cellulaires *in vitro* ont largement contribué au développement de cette science. Enfin, l'apport des études biochimiques et génétiques des bactériophages dans les années 1950 et 1960 a été l'un des moteurs de recherche conduisant au développement de la biologie moléculaire moderne, qui en retour, a permis une étude plus approfondie des virus. L'explosion de la biologie moléculaire, de la génétique et l'apparition de la génétique inverse ces dernières années a aussi permis un développement considérable de la recherche en virologie. Plus récemment, l'étude des virus a été aussi grandement facilitée par le développement de la cristallographie aux rayons X, avec l'étude de cristaux de protéines virales et de virus entiers, qui ont permis de mieux comprendre l'organisation et la fonction des protéines présentes au sein de la particule virale. Actuellement, la recherche en virologie vise à comprendre gène par gène, protéine par protéine comment ces virus sont capables de s'adapter et d'évoluer avec leur hôte. Aujourd'hui, on dénombre plus de 4 000 virus, capables d'infecter des organismes appartenant à tous les domaines d'organismes vivants, aux Archaeobactéries, aux Bactéries et aux Eucaryotes (figure 2). Ils possèdent des génomes (ADN monocaténaire ou bicaténaire, circulaires ou linéaires, ARN monocaténaire (polarité négative ou positive) ou bicaténaire, linéaire ou segmenté) (classification de Baltimore) et des morphologies variées, qui prennent en compte la symétrie de la capsid ou de la nucléocapsid (icosaédrique par exemple pour l'adénovirus, ou hélicoïdale pour les nucléocapsides des virus à ARN négatif), et la présence ou l'absence d'enveloppe. Toutefois, il apparaît clairement aujourd'hui que la compréhension des processus de réplication requière d'intégrer des approches biologiques, biochimiques, biophysiques et structurales.

La notion de virus est à peine vieille d'un siècle et la virologie est un domaine de recherche qui suscite énormément d'intérêt dans la communauté scientifique, en raison de la variabilité des virus d'un point de vue structural, génomique et pathologique. L'étude des virus constitue également un outil pour l'étude des mécanismes moléculaires de la cellule.

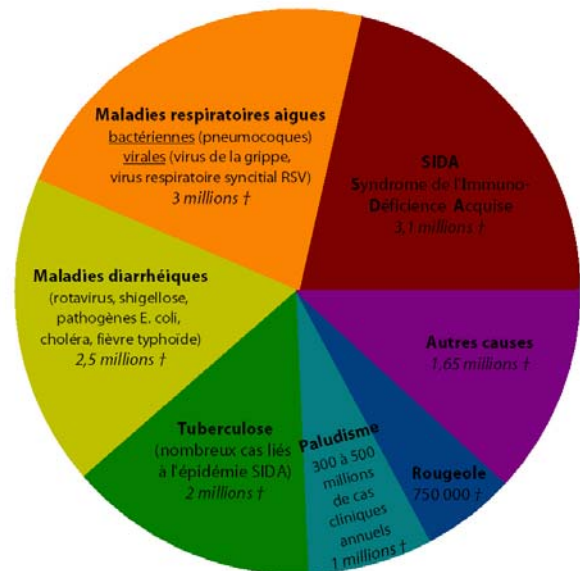


**Figure 2** - Virosphère d'après ICTV (*International Committee on Taxonomy of Viruses*), classifiant les virus selon leurs propriétés génomiques, leurs formes ainsi que leurs hôtes.

Les maladies infectieuses sont provoquées par la transmission de virus, de bactéries, de parasites, de mycoses ou de prion, et sont responsables de 14 millions de décès dans le monde par an (un tiers de la mortalité mondiale) (figure 3), le plus souvent dans les pays en voie de développement. Parmi ces maladies meurtrières, on retrouve de nombreux virus tels que le virus de la grippe, le virus respiratoire syncytial humain (RSV) qui provoquent des maladies respiratoires (bronchiolites), les rotavirus mais aussi les adénovirus, les astrovirus et les entérovirus qui provoquent des maladies diarrhéiques, le virus de la rougeole ainsi que le virus de l'Immunodéficience humaine (VIH). Alors que des vaccins efficaces existent contre plusieurs de ces virus, de nombreuses personnes, et plus particulièrement les enfants, continuent de mourir chaque année, par manque d'accès aux traitements préventifs ou curatifs. Aussi, certains cancers sont maintenant associés à la présence de virus, comme le cancer du col de l'utérus, qui est la deuxième cause de mortalité par cancer dans le monde chez la femme (environ 230 000 décès et 500 000 nouveaux cas par an), et, qui dans plus de 95% des cas est relié à la persistance de *papillomavirus* (deux souches sont très oncogènes, HPV-6 et 8 (*Human PapillomaVirus*)), ou encore, le cancer du foie lié à la présence des virus des hépatites B et C. On estime aujourd'hui que plus de 15% des cancers seraient d'origine infectieuse (Institut Pasteur).



**Figure 3** - Répartition des causes de mortalité par maladies infectieuses dans le monde, d'après l'Institut Pasteur (<http://www.pasteur.fr/ip/index.jsp>). Les maladies infectieuses sont la cause de 43 % des décès dans les pays en voie de développement contre 1% dans les pays industrialisés. Les six maladies présentées représentent 90 % des décès par maladies infectieuses dans le monde.



## B – Les virus à ARN négatif

Les virus à ARN négatif sont des virus enveloppés, leur membrane dérivant de la membrane de la cellule hôte infectée, dont l'ARN viral génomique simple brin de polarité négative est complémentaire de l'ARN messager. Les virus à ARN négatif ont des morphologies variables (figure 4), ils infectent un large spectre d'hôtes allant des plantes aux invertébrés et aux mammifères et leurs vecteurs sont le plus souvent des arthropodes. Ces virus possèdent un génome variant de 9 à 19 kb, codant entre 5 et 11 protéines seulement (tableau 1) et ils sont à l'origine de nombreuses pathologies humaines. En effet, certains sont responsables d'infections respiratoires aiguës (virus de la grippe, virus respiratoire syncytial humain), d'encéphalites graves (virus de la rage, virus Nipah), ou de fièvres hémorragiques (virus Ebola, Lassa, de la fièvre de Crimée Congo).

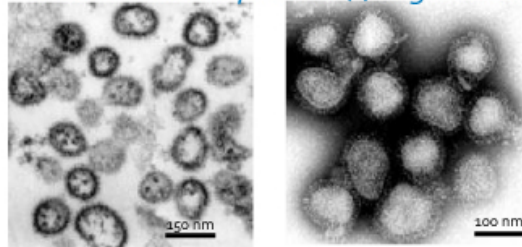
On distingue les virus dont le génome est segmenté et ceux dont le génome est non segmenté. Les premiers, divisés en trois familles, ont un génome constitué de plusieurs molécules ou « segments » d'ARN simple brin de polarité négative :

- *Arenaviridae* (2 segments, virus Lassa)
- *Bunyaviridae* (3 segments, virus de la fièvre de la vallée du Rift)
- *Orthomyxoviridae* (7 à 8 segments, virus de la grippe).

Les virus dont le génome est non segmenté sont regroupés dans l'ordre des *Mononegavirales* sous la forme de quatre familles :

- *Paramyxoviridae* (virus de la rougeole, Sendai, Nipah, virus respiratoire syncytial)
- *Rhabdoviridae* (virus de la rage, de la stomatite vésiculaire)
- *Filoviridae* (virus Marburg, Ebola)
- *Bornaviridae* (virus Borna).

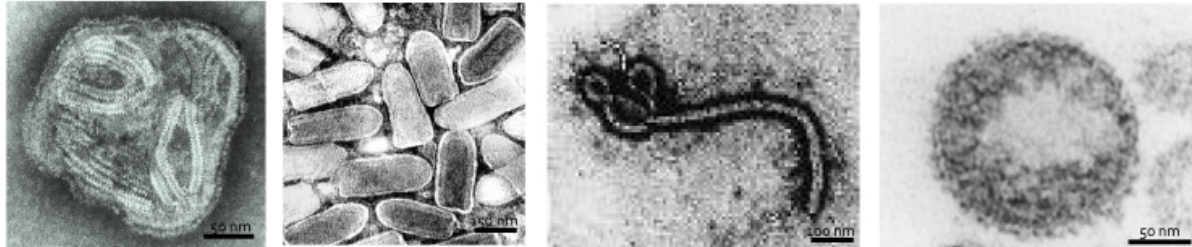
### Virus à ARN simple brin (-) segmenté



Virus Lassa (Arenaviridae)

Virus de la grippe (Orthomyxoviridae)

### Mononegavirales



Virus parainfluenza  
(Paramyxoviridae)

Virus de la Stomatite Vésiculaire  
(VSV) (Rhabdoviridae)

Virus Ebola (Filoviridae)

Virus Borna (Bornaviridae)

**Figure 4** - Images en microscopie électronique de particules virales de virus à ARN négatif. Les images proviennent de la « picture gallery » de ICTV (*International Committee on Taxonomy of Viruses*, <http://www.ncbi.nlm.nih.gov/ICTVdb/>). Virus Lassa (Arenaviridae), virus de la grippe (A/Hong Kong/1/68), virus de la stomatite vésiculaire (VSV), virus Ebola (Zaire Ebola, image prise en octobre 1976), virus Borna: courtoisie du Dr. F. A. MURPHY (School of Veterinary Medicine, University of California, Davis. Virus parainfluenza: courtoisie du Dr. Linda STANNARD (Department of Medical Microbiology, University of Cape Town, South Africa).

Les *Mononegavirales* regroupent un grand nombre de virus, certains connus depuis des siècles comme le virus de la rage, d'autres découverts plus récemment, les virus dits « émergents ». Ils s'adaptent rapidement à de nouveaux hôtes et sont responsables d'épidémies à forte mortalité, comme celle attribuée au virus Nipah en 1998 (Chua & al., 2000).

## C – Les *Rhabdoviridae*

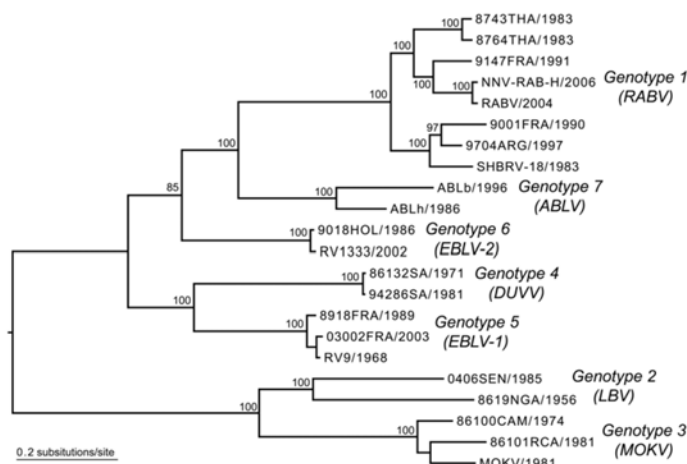
### 1 - Classification

La famille des *Rhabdoviridae* comporte plus de 200 virus infectant des mammifères, des poissons, des crustacés, des reptiles et des plantes. Parmi les cinq genres appartenant à cette famille, les genres *Vesiculovirus*, *Lyssavirus* et *Ephemerovirus* infectent des animaux alors que les genres *Cytorhabdovirus* et *Nucleorhabdovirus* infectent des végétaux. Il semblerait qu'un grand nombre d'entre eux soit véhiculé par des insectes (Hogenhout & al., 2003). Ainsi, différentes espèces de mouches sont impliquées dans la transmission d'espèce à espèce de certains *Vesiculovirus*, des moustiques transmettent le virus de la fièvre éphémère bovine, et des pucerons sont responsables de la transmission de *Rhabdoviridae* aux végétaux.

Famille	Sous-Famille	Genre	Espèce	Abréviations	Principales caractéristiques
<b>Bornaviridae</b>		<b>Bornavirus</b>	virus de la maladie de Borna	BDV	<b>Génome</b> : 9 kb. <b>Protéines structurales</b> : 5. <b>Hôte</b> : vertébrés.
<b>Filoviridae</b>		<b>Ebolavirus</b>	virus Ebola	/	<b>Génome</b> : 19 kb. <b>Protéines structurales</b> : 7. <b>Morphologie</b> : filamenteuse. <b>Hôtes</b> : chauve-souris.
		<b>Marburgvirus</b>	virus Marburg	/	
<b>Paramyxoviridae</b>	<b>Paramyxovirinae</b>	<b>Avulavirus</b>	<i>Paramyxovirus</i> aviaire (2, 3, 4 et 6)	APVM2	<b>Génome</b> : de 15 à 18 kb. <b>Protéines structurales</b> : 10/11. <b>Morphologie</b> : pléiomorphe. <b>Hôte</b> : vertébrés (oiseaux, primates, bovins...)  <b>Décès annuel dans le monde</b> : Rougeole : 750 000 morts.
			virus de la maladie de Newcastle	NDV	
			<i>Avulavirus</i> non classifié	/	
		<b>Henipavirus</b>	virus Hendra	HeV	
			virus Nipah	NIv	
		<b>Morbillivirus</b>	virus de la maladie de Carré	CDV	
			<i>Morbillivirus</i> de cétacés	DMV & PMV	
			virus de la rougeole	MeV	
			peste-des-petits-ruminants virus	PPRV	
			<i>phocine distemper virus</i>	PDV	
		<b>Respirovirus</b>	virus Rinderpest	RPV	
			virus bovin parainfluenza 3	bPIV-3	
			virus humain parainfluenza 1	hPIV-1	
			virus humain parainfluenza 3	hPIV-3	
			virus Sendai	SeV	
	<i>Respirovirus</i> non classifié		/		
	<b>Rubulavirus</b>		virus humain parainfluenza 2	hPIV-2	
			virus humain parainfluenza 4	hPIV-4	
		virus mapuera	/		
		virus des oreillons	MuV		
<i>Rubulavirus</i> porcine		/			
virus simian 5		SV5			
virus simian 41		SV41			
<b>Paramyxovirinae non classifié</b>			/		
<b>Pneumovirinae</b>	<b>Metapneumovirus</b>	<i>Metapneumovirus</i> aviaire	/		
		<i>Metapneumovirus</i> humain	/		
	<b>Pneumovirus</b>	Virus humain respiratoire syncytial	hRSV		
		Virus bovin respiratoire syncytial	bRSV		
		Virus murin pneumonia	/		
<i>Pneumovirus</i> non classifié	/				
<b>Rhabdoviridae</b>	<b>Cytorhabdovirus</b>	virus nécrotique de la laitue jaune	LNYV		
		virus de la mosaïque céréales (Nord)	/		
		<i>strawberry crinkle virus</i>	/		
		<i>Cytorhabdovirus</i> non classifié	/		
	<b>Ephemerovirus</b>	virus de la rivière Adalaide	/		
		virus Berrimah	/		
		virus de la fièvre éphémère bovine	BEFV		
		virus Kimberley	/		
	<b>Vesiculovirus</b>	<i>Ephemerovirus</i> non classifié	/		
		virus Carajas	/		
		virus Chandipura	CV		
		virus coccal	/		
		virus isfahan	/		
		virus piry	/		
		virus stomatite vésiculaire Alagoas	VSV A		
		virus stomatite vésiculaire Indiana	VSV I		
	virus stomatite vésiculaire New Jersey	VSV NJ			
	<b>Lyssavirus</b>	<i>Vesiculovirus</i> non classifié	/		
		<i>Australian bat lyssavirus</i> (1 et 2)	ABL		
		virus Duvenhage	/		
		<i>European bat lyssavirus</i> (1 & 2)	EBL		
		<i>Lagos bat virus</i>	/		
		virus Mokola	/		
		virus de la rage	RV		
		<i>Lyssavirus</i> non classifié	/		
	<b>Novirhabdovirus</b>	<i>Rhabdovirus</i> Hiramé	/		
		Virus de la nécrose hématopoïétique infectieuse	IHNV		
		<i>Rhabdovirus</i> Snakehead	/		
	<b>Nucleorhabdovirus</b>	virus de la septicaémie virale hémorragique	/		
		virus de la mosaïque du maïs	/		
		virus de la patate jaune naine	/		
		virus du riz jaune	/		
		<i>Sonchus yellow net virus</i>	SYNV		
	<b>Rhabdoviridae non classifié</b>			/	

Tableau 1 - Classification des différents virus appartenant à l'ordre des *Mononegavirales*. Adapté de « Taxonomy » Pubmed (<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root>).

Le genre *Lyssavirus* est divisé en sept génotypes, sur la base d'une analyse de séquence de la nucléoprotéine (Bourhy & al., 1993) (Badrane & al., 2001), mais aussi plus récemment sur la base d'une analyse des génomes entiers (Delmas & al., 2008). Le génotype 1 englobe toutes les souches de rage alors que les autres génotypes correspondent à des virus apparentés à la rage (*Lagos Bat* : génotype 2 ; *Mokola* : génotype 3 ; *Duvenhage* : génotype 4 ; *EBVL1* (*European Bat LyssaVirus 1*) : génotype 5 ; *EBLV2* : génotype 6 ; *ABLV* (*Australian Bat LyssaVirus*) : génotype 7). Tous les génotypes excepté le génotype 3, pour lequel l'espèce hôte reste encore inconnue, ont des chiroptères comme réservoir naturel. Plus récemment, quatre génotypes additionnels, qui infectent des chauves-souris en Asie centrale et du sud-est, ont été ajoutés : virus Aravan, Khujand, Irkut et West Caucacian Bat (Kuzmin & al., 2008).



**Figure 5** - Relation phylogénétique de 22 régions codantes complètes de génomes de *Lyssavirus*, représentant les 7 génotypes. D'après (Delmas & al., 2008).

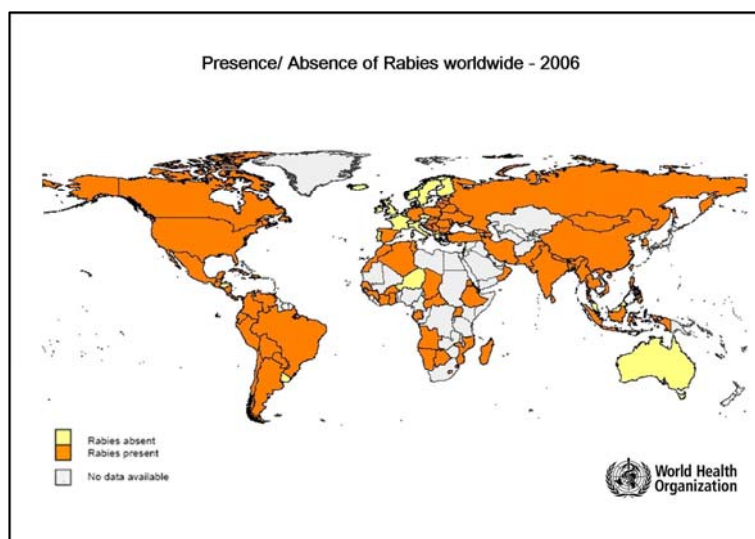
Dans ce travail, nous nous sommes intéressés aux protéines du virus de la rage (RV) (genre *Lyssavirus*) et du virus de la stomatite vésiculaire (VSV) (genre *Vesiculovirus*).

## 2 – la rage

La rage est une maladie très ancienne, décrite depuis l'Antiquité, dont l'origine infectieuse n'a été proposée qu'au début du XIX<sup>ème</sup> siècle (Georg Gottfried ZINKE, 1804). L'étiologie virale a été démontrée dans les années 1880 par Louis PASTEUR. En juillet 1885, PASTEUR réussit à immuniser un jeune garçon, Joseph MEISTER, qui avait été mordu plusieurs fois par un chien enragé, en inoculant à plusieurs reprises un vaccin constitué de virus atténué, préparé à partir de moelles épinières de lapin.

### **La rage, zoonose virale**

La rage est une zoonose virale affectant les mammifères terrestres à sang chaud qui est très largement répandue dans le monde (plus de 2,5 milliards de personnes vivent dans des régions endémiques de rage) (figure 6). Il existe deux cycles épidémiologiques majeurs du virus de la rage : la **rage terrestre** et la **rage sauvage**.

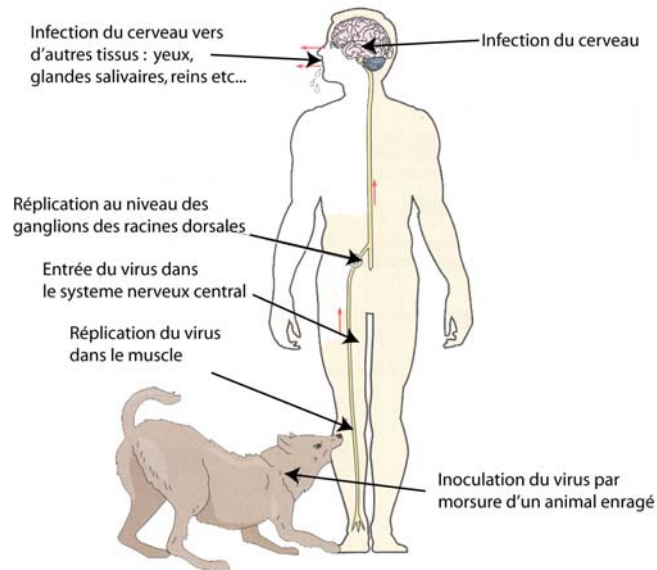


**Figure 6** - Répartition mondiale de la rage terrestre dans le monde (Organisation Mondiale de la Santé, source Rabnet, <http://www.who.int/GlobalAtlas/>).

### ***La rage terrestre (urbaine, des rues ou canine)***

Elle est responsable de 90% des cas de rage dans le monde. Chez le renard, le chien et l'homme, la rage est un virus neurotrope qui provoque une infection du cerveau toujours mortelle lorsqu'elle est déclarée. La rage terrestre est transmise à l'homme par les chiens qui constituent le réservoir et le vecteur principal du virus dans le monde. Chez l'homme, les contaminations rabiques sont le plus souvent dues à un contact (morsure) avec la salive de l'animal enragé, toutefois, le virus peut pénétrer l'organisme à la faveur de griffures, d'excoriations cutanées ou de léchage d'une plaie ouverte. En effet, le virus de la rage ne pénètre pas les téguments intacts, mais peut franchir les muqueuses saines. D'autres contaminations plus exceptionnelles ont été décrites : une contamination interhumaine s'est produite lors d'une greffe de cornée prélevée chez des personnes décédées d'encéphalites (le diagnostic rabique n'avait pas été effectué).

Pendant la période d'incubation, le virus se propage dans l'organisme en cheminant le long des neurones par **voie axonale centripète** (figure 7). Cette période a une durée variable, dépendante de la zone de morsure. Le virus remonte le long des nerfs périphériques puis de la moelle épinière pour finalement aboutir au cerveau. Dans une deuxième phase, le virus se répand par **voie axonale centrifuge** vers divers organes préférentiels tels que les reins, la cornée et les glandes salivaires (figure 7). Chez les personnes décédées par suite d'une infection par le virus de la rage, de hauts titres viraux sont retrouvés dans ces organes ainsi que dans les muqueuses. Le virus de la rage a une forte affinité pour les zones du cerveau impliquées dans le comportement, ce qui induit souvent un comportement agressif chez l'animal ou l'homme, permettant ainsi la transmission du virus. L'animal contaminé, devenu enragé, va lui-même disséminer le virus à un autre hôte par morsure.



**Figure 7** - Infection par le virus de la rage chez l'homme. Adapté de *Principles of virology* (Flint, Enquist, Racaniello, Skalka).

Deux formes cliniques de la rage existent chez l'homme. La **forme spastique dite rage furieuse** représente 70 % des cas. Elle se caractérise par des troubles du comportement, une hyperactivité, des hallucinations et une hydrophobie (impliquant aérophobie, dysphagie et salivation abondante). La phase aigue dure de 2 à 7 jours puis le patient entre dans un coma et meurt par paralysie du système cardiorespiratoire en environ une semaine. La **forme paralytique dite rage muette** représente 30 % des cas. Seul 50% des personnes infectées présentent les symptômes hydrophobie et aérophobie, mais des paresthésies apparaissent au niveau de la région de la morsure, suivies de paralysie puis de paraplégie ou quadriplégie. La mort se produit par paralysie respiratoire au bout de 14 jours. Souvent, la rage transmise par les chiroptères est paralytique.

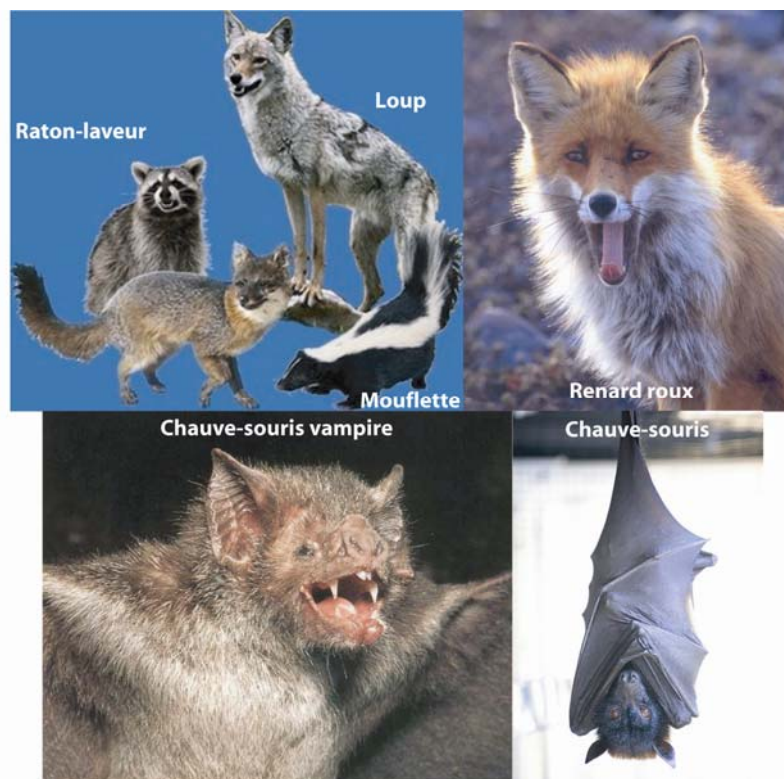
### **La rage sauvage (ou selvatique) des mammifères et des chiroptères**

De nombreuses espèces de chauve-souris sont concernées (McColl & *al.*, 2000) par la rage des chiroptères : par exemple, la grande chauve souris marron (*Eptesicus fuscus*), la chauve souris à tête d'argent (*Lasionycteris noctivagans*) et aussi la chauve souris pipistrelle de l'est de l'Amérique (*Pipistrellus subflavus*). Cependant, seuls les chiroptères semblent être porteurs asymptomatiques de la rage. Ce virus « des chiroptères » est très rarement transmis aux humains, sauf lors de contact direct avec les animaux ou par contact avec des aérosols dans les grottes, où vivent les chauves-souris (Fooks, Brookes & *al.*, 2003) (Fooks, McElhinney & *al.*, 2003). Les espèces insectivores sont des vecteurs de rage dans toutes les régions du globe y compris dans les pays indemnes de rage terrestre tels que la Grande-Bretagne et l'Australie. En Amérique latine, ce sont surtout les chauves-souris hémato-phages (vampires) (figure 8) qui transmettent la rage au bétail. Les chiroptères, pouvant excréter le virus dans leur salive pendant de longues périodes sont des agents de contamination potentiellement redoutables. Comme le réservoir chauve souris est

important, le contrôle de la rage terrestre est seulement possible par la vaccination des animaux réservoirs.

Divers animaux sauvages forment le réservoir permanent de la maladie. A leur contact, les animaux domestiques, les chiens errants voire l'homme peuvent se contaminer. Selon les régions du monde considérées, de nombreux vecteurs différents sont retrouvés (figure 8), tels que le chacal et la mangouste en Afrique, la mangouste et le coyote en Amérique du Sud, le raton-laveur et la mouffette en Amérique du Nord, le renard au Moyen Orient, le loup en Europe occidentale, de l'Est et du Nord, le renard roux en Europe de l'Est et du Nord mais aussi le chien viverrin en Europe. Le virus est présent dans la salive des animaux infectés 3 à 5 jours avant l'apparition des symptômes neurologiques. L'animal meurt dans les 15 jours qui suivent la déclaration de la maladie.

En Eurasie, de nombreux siècles auparavant, il est possible que le virus de chauve souris soit entré dans la population des renards roux et ait établi un cycle répliatif dans ce nouvel hôte par une série de mutations (Badrane & al., 2001). Des changements similaires pourraient avoir eu lieu aux Etats-Unis, où la rage terrestre affecte les ratons laveurs et les mouffettes (Jesse D. BLANTON).



**Figure 8** - Quelques animaux susceptibles d'être porteurs du virus de la rage.

### **Diagnostic de la rage**

Adelchi NEGRI en 1903 fut le premier à identifier la présence du virus grâce à des ultrastructures présentes dans le cytoplasme des cellules, accumulation de protéines virales que l'on a appelées les corps

de Negri (Miyamoto & Matsumoto, 1965) (De Brito & Araujo, 1973). Cependant, la détection de ces structures n'est pas fiable à 100% et d'autres techniques de détection pour le diagnostic efficace et rapide telles que l'immunofluorescence directe, la culture cellulaire, les tests immunoenzymatiques ainsi que la détection par PCR sont actuellement utilisées.

### **Prévention, vaccination et traitement de la rage**

Il n'existe pas de traitement préventif de la rage. La seule façon de se protéger contre le virus est de se faire vacciner. Le vaccin original contre le virus de la rage a été développé par Louis PASTEUR à partir de moelles épinières de lapin. Cependant, le vaccin préparé à partir de tissu cérébral, encore utilisé dans les pays en voie de développement, est encéphalitogène à cause de la présence de myéline et d'impuretés dans les préparations. Aujourd'hui, des progrès considérables ont été réalisés dans la préparation des vaccins anti-rabiques. Ceux-ci sont mis au point en culture cellulaire (*human diploid cell vaccine* (HDCV), *rabies vaccine absorbed (rhesus monkey cells)* (RVA)) ou sur œufs embryonnés de poulets (*purified chick embryo cell vaccine* (PCEC)). Le vaccin est ensuite concentré puis inactivé. Le plus souvent l'injection est intramusculaire.

La vaccination pré exposition n'est utilisée que pour les personnes en contact avec le virus comme les vétérinaires, les fermiers et le personnel de laboratoire.

Le traitement post exposition consiste en une vaccination et une injection d'immunoglobuline anti-rabique (*rabies immunoglobulin* (RIG)) à haut titre d'anticorps afin d'anticiper la réponse immunitaire de la personne infectée. Ce traitement reste cher mais efficace lorsque la maladie est traitée à temps. Cependant, il n'existe aucun médicament pour empêcher la multiplication virale et aucun traitement efficace une fois que les premiers symptômes sont apparus.

En Europe, la vaccination systématique (vaccination orale avec des vaccins atténués) des chiens puis des animaux sauvages (renards) a largement fait régresser la rage terrestre. La France a été déclarée indemne de rage terrestre en 2001. Le dernier cas de rage humaine contractée en France remonte à 1924. Toutefois, en 2006, près de 8500 consultations pour suspicion de rage ont été réalisées et le nombre de traitements post exposition chez des voyageurs mordus en zone d'endémie a plus que doublé en 20 ans (Institut Pasteur). En 2003, un enfant est décédé après avoir contracté le virus de la rage lors d'un voyage en Afrique du Nord et en 2004, trois cas de rage canine ont été diagnostiqués (chiens importés). A la fin de l'année 2007, en Vendée, un chat est mort du virus transmis par une chauve-souris. Début 2008, une fillette de 12 ans a été mordue par son American Staffordshire (jeune chien illégalement importé du Maroc) pour lequel un diagnostic de rage a été établi, en Seine et Marne.

Quelques cas de guérisons miraculeuses ont été décrits, notamment le cas de Jeanna GIESE, adolescente de 15 ans vivant dans le Wisconsin (Etats-Unis) qui a survécu grâce à un nouveau traitement



(Willoughby & al., 2005) (Willoughby, 2007). Les médecins l'ont plongé dans le coma (*kétamine et midazolam*) et lui ont administré des antiviraux (*ribavirine et amantadine*) afin de stimuler son immunité innée, sans toutefois lui administrer de vaccination anti-rabique. Même si la jeune fille a gardé des troubles de la parole et de la stature plusieurs mois après son hospitalisation, elle a survécu. Toutefois, après analyse, l'agent infectieux était un variant de virus de la rage associé à la chauve-souris à tête en argent *Lasionycteris noctivagans*, ce qui peut expliquer la survie de cette jeune fille, puisqu'elle n'a pas été infecté par le virus de la rage dite terrestre.

D'après l'Organisation Mondiale de la Santé (OMS), la rage reste un problème important de santé publique puisqu'elle cause plus de 50 000 décès par an dans le monde. Même si toutes les classes d'âges sont sensibles, les enfants de 5 à 15 ans sont particulièrement exposés, puisque plus de 50 % des sujets mordus par des chiens enragés sont des enfants, le plus souvent des garçons de moins de 10 ans. L'incidence de la rage, humaine et animale, est bien plus importante dans les pays en voie de développement, où 99 % des cas humains sont mortels : l'Asie concentre la majorité des ces cas (Inde, Pakistan, Bangladesh et Birmanie), mais des centaines de personnes meurent aussi de la rage en Afrique (Ouganda, Togo, Ghana, Egypte, Maroc). Chaque année, entre 6 à 7 millions de personnes reçoivent un traitement post exposition. Quelques cas sont répertoriés en Amérique du sud, et des dizaines de cas sporadiques autochtones ou de cas importés sont signalés en Amérique du nord ou en Europe.

Malgré tous les progrès techniques de diagnostic, la rage humaine reste de nos jours au dixième rang des maladies infectieuses mortelles (Institut Pasteur). Il n'existe aujourd'hui aucun traitement de la maladie une fois que celle-ci est déclarée et donc la rage reste un problème majeur de santé publique. Pour traiter la rage, il semble essentiel de bloquer le virus avant qu'il ne puisse pénétrer dans le système nerveux central ou de bloquer sa multiplication massive dans le cerveau. La connaissance des détails structuraux des processus de réplication de ces virus pourrait aider au développement de nouvelles cibles thérapeutiques et antivirales.

### 3 – Le virus de la stomatite vésiculaire (VSV)

Cette maladie du bétail a été décrite aux Etats-Unis en 1821 (Hanson, 1981). Elle se localise principalement en Amérique du Nord et du Sud, toutefois, des épidémies de plus faibles amplitudes ont été décrites chez des chevaux en France et en Afrique du Sud au début du XX<sup>ième</sup> siècle. Le virus a initialement été isolé par W. E. COTTON (Cotton, 1927) et sa morphologie a été établie par Chow et ses collaborateurs (Chow & al., 1954).

Le virus de la stomatite vésiculaire (VSV) touche principalement le bétail, les chevaux, et les porcs, et cause une maladie bénigne qui se caractérise par l'apparition de fièvre, de salivation excessive et de lésions vésiculaires sur les gencives, la langue (figure 9, A), la bouche (figure 9, B), sur les pis (vaches laitières) et les sabots (figure 9, C) de ces animaux. Cette maladie a des symptômes similaires à ceux de la fièvre aphteuse (*Picornaviridae, Aphthovirus*) et de la maladie vésiculaire du porc (*Picornaviridae, Enterovirus*). Les animaux éprouvent des difficultés pour se nourrir et se mouvoir entraînant une perte de poids et un arrêt de la lactation chez les vaches laitières. Cette maladie a des conséquences économiques importantes dans les grandes fermes américaines. La maladie est rarement fatale sauf chez les porcs. Les épidémies de stomatite vésiculaire apparaissent soudainement (approximativement en 2 semaines) et se propagent rapidement dans le troupeau mais ne se propagent pas nécessairement aux fermes adjacentes. Des vaccins ont été utilisés avec peu de succès (virus atténués, vaccin recombinant spécifique). Leur développement n'a pas été poursuivi, en raison du caractère bénin de la maladie.



**Figure 9** - Images montrant les vésicules dues à une infection par le VSV, sur la langue d'un veau (A), sur la bouche (B) et sur le sabot (C) d'un cheval.

L'infection est transmise par voie cutanée ou par contact avec des muqueuses infectées et le virus se trouve dans les fluides des vésicules et dans l'épithélium. Il ne persiste pas après la guérison. Le virus n'est excrété ni dans les urines, ni dans les fèces, ni dans le lait. La diffusion par contact direct avec les fluides est possible, et le transfert des lésions des mamelons par le matériel de tirage de lait est occasionnellement signalé. Cependant, les études expérimentales n'ont pas pu démontrer une transmission directe d'animal à animal, et on a remarqué de nombreuses épidémies où la maladie ne se propageait pas aux animaux voisins dans une exploitation (Sellers & Maarouf, 1990) (Sellers, 1984). La maladie chez les animaux est brève, avec une virémie transitoire et minimale. La transmission par des arthropodes piqueurs au bétail semble être la plus probable. De plus, de nombreux *Vesiculovirus* ont été isolés chez les moustiques, les mouches, les moucheron, les mites et les tiques (Tesh & al., 1972) (Johnson & al., 1969). Il a été suggéré que certaines souches pouvaient être des souches de plantes, et

que les animaux sont infectés directement, soit en mangeant les plantes infectées, soit en mangeant les insectes.

Le virus peut parfois infecter les humains, causant généralement des symptômes de type grippaux et des lésions vésiculaires sur les lèvres et les mains, mais le plus souvent la maladie est asymptomatique. La période d'incubation est d'environ 9 jours après l'exposition et les symptômes peuvent persister 3 à 4 jours. De façon occasionnelle, il peut y avoir des rechutes (Sellers, 1984). Toutefois, les infections ont été observées chez des personnes en contact avec un haut titre viral (inhalation d'aérosols, ou contamination accidentelle des yeux) comme les vétérinaires, le personnel travaillant dans les fermes en contact avec les animaux infectés ou encore les laborantins. Des cas d'encéphalites associées au VSV ont été signalé (Quiroz & *al.*, 1988). Cependant, dans certaines zones de l'Amérique centrale et du Nord, de 25 à 90 % des fermiers ont développé des anticorps anti-VSV (Hanson, 1981).

Deux sérotypes majeurs de *Vesiculovirus* existent, les sérotypes Indiana et New Jersey. Il existe d'autres *Vésiculovirus* capables de provoquer des lésions similaires chez les animaux et d'infecter l'homme : le virus Piry (isolé chez un opossum au Brésil), le virus Chandipura (initialement isolé chez l'homme en Inde) et le virus Isfahan (isolé à partir de mouches et d'hommes en Iran) (Tesh & *al.*, 1977). Aux Etats-Unis et en Amérique Centrale, ce sont principalement les sérotypes Indiana et New Jersey qui sont présents, alors qu'on trouve en Amérique du Sud, les souches Indiana Cocal et Alagoas. Les épidémies sont plus fréquentes dans les zones tropicales et subtropicales et apparaissent tous les ans ou tous les 2 à 3 ans, et les épidémies dans les zones tempérées apparaissent à des intervalles de 5 à 10 ans, pendant les mois d'été. Il existe une variation saisonnière pour la disparition de stomatite vésiculaire chez les animaux : elle disparaît à la fin de la saison des pluies dans les régions tropicales, et aux premières gelées dans les régions tempérées.

### ***VSV, un outil thérapeutique potentiel ?***

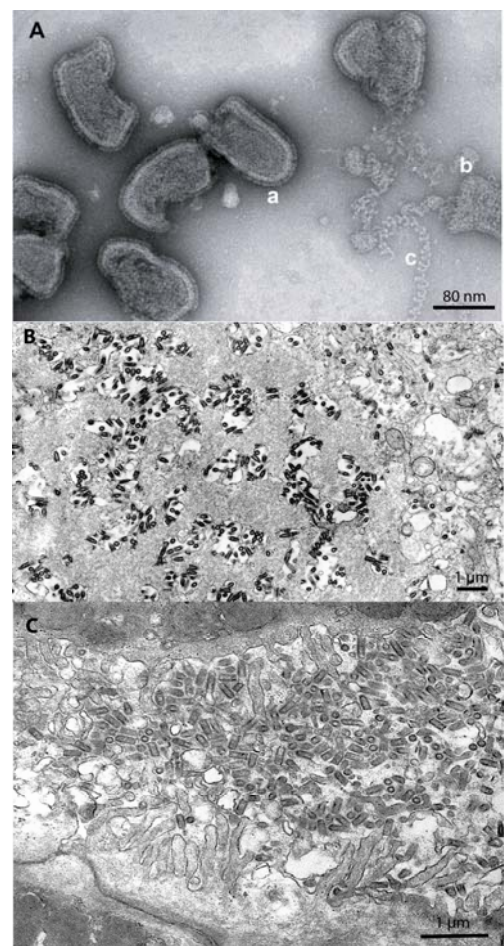
Le virus de la stomatite vésiculaire (VSV) constitue un excellent modèle pour la réplication et la transcription car c'est le virus le plus étudié de l'ordre des *Mononegavirales*. Il ne comporte que cinq gènes dont les signaux de régulation de leur expression sont plus simples que pour les autres virus de cet ordre. Ce virus se multiplie facilement dans le cytoplasme dans de nombreuses souches cellulaires, et des tests de transcription et réplication *in vitro* ont été établis. Il est aussi très largement utilisé comme outil de laboratoire pour étudier divers aspects de la biologie cellulaire, pour comprendre l'immunité innée et acquise. Depuis quelques années, son potentiel thérapeutique a été étudié : en effet, le VSV semble adapté comme agent de vaccination (Lichty & *al.*, 2004) (Bukreyev & *al.*, 2006) mais aussi comme virus oncolytique (Lichty & *al.*, 2004). Le VSV a été testé dans des modèles animaux comme agent de vaccination, et son application à l'homme pourrait être possible. En effet, le VSV pourrait être utilisé pour

immuniser la quasi-totalité de la population puisque la plupart des personnes ne possède pas d'immunité contre ce virus. Aussi, de nombreux types cellulaires peuvent être infectés, avec un niveau élevé et stable d'expression des gènes, même étrangers, une réplication rapide et efficace *in vitro*. La réplication de ce virus dans le cytoplasme des cellules est un plus puisqu'il ne va pas affecter le génome de ces cellules. Toutefois, peu d'informations sont disponibles quant à l'effet de l'administration de ce virus à l'homme, et il n'est pas clair quels types de tissus sont infectés. Plus récemment encore, le VSV a été utilisé comme vecteur oncolytique, et il semblerait qu'il soit capable de cibler les cellules cancéreuses. Il serait alors possible d'appliquer un traitement (par exemple interféron) à ces cellules.

## II – La transcription et la réplication chez les *Rhabdoviridae*

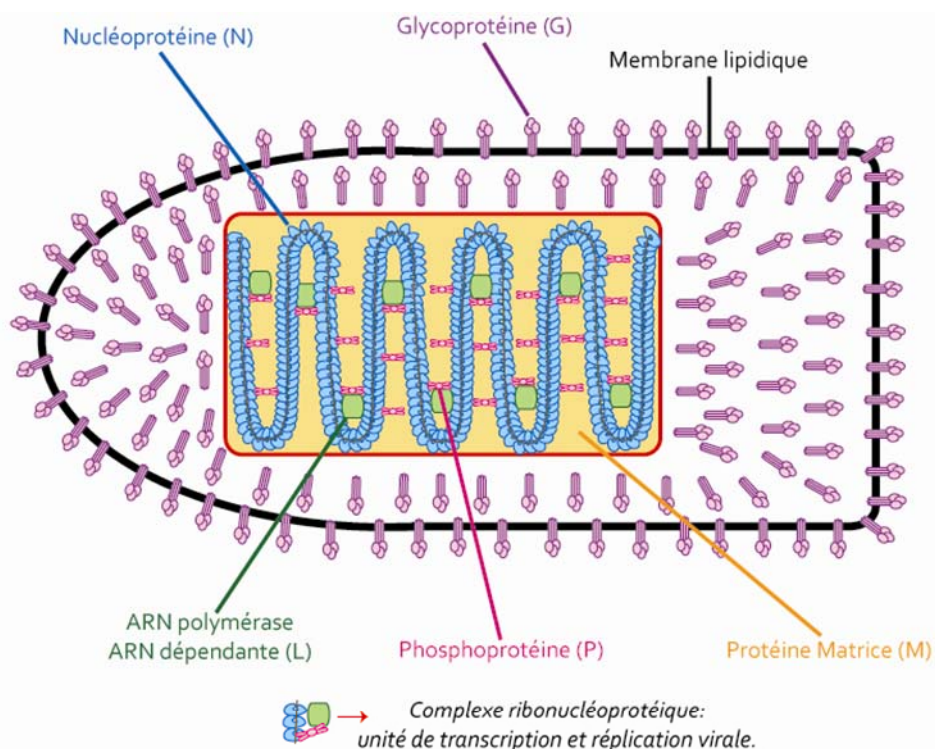
### A – Morphologie et composition des *Rhabdoviridae*

Les virus appartenant à l'ordre des *Mononegavirales* montrent une grande diversité en termes de morphologie (figure 4) et d'interactions avec leurs hôtes. Les *Rhabdoviridae* sont des virus enveloppés dont la membrane lipidique dérive de la membrane de la cellule hôte infectée. Le génome de ces virus est composé d'une molécule **d'ARN simple brin de polarité négative** (complémentaire de l'ARNm). La particule virale des *Rhabdoviridae* a une forme caractéristique ressemblant à une balle de fusil (« *bullet shape* ») (figure 10). L'une des extrémités du virus est plate alors que l'autre est en forme d'ogive (figure 10). Le diamètre des particules virales avoisine les 75 nm et leur longueur varie entre 100 et 300 nm. Ces virus sont composés de 74 % de protéines, 20 % de lipides, 3 % de carbohydrates et de 3 % d'ARN (Thomas & al., 1985).



**Figure 10** - Visualisation en microscopie électronique (coloration négative) du virus de la rage. (A) : particules virales de rage (souche Pasteur), photo prise par le Pr. Rob W.H. RUIGROK, a : particule virale en forme de balle de fusil, b : nucléocapside virale se déroulant, c : nucléocapside libre. (B) : virus de la rage présent dans une cellule neuronale. (C) : virus de la rage présent dans les cellules de glandes salivaires d'un renard. (B & C) : photo prises par le Dr. Frederick A. MURPHY.

Les virus appartenant à l'ordre des *Mononegavirales* partagent la même organisation structurale et génomique ainsi que des modes similaires de répllication et de transcription de leur ARN (David M Knipe). Il existe en tout **cinq protéines virales majeures** avec des modifications post-traductionnelles qui ont été attribuées soit à l'enveloppe de la particule virale soit à la nucléocapside virale, par une analyse biochimique par dissections du virion avec des protéases et des détergents. La **glycoprotéine (G)** est enchâssée dans la membrane lipidique et forme des spicules trimériques que l'on peut observer en microscopie électronique (Doms & *al.*, 1987) (Gaudin & *al.*, 1992). La **protéine matrice (M)** est responsable de la condensation de la nucléocapside. La structure interne de la particule virale est constituée de trois protéines virales qui s'associent à l'ARN viral afin de former un complexe hélicoïdal nommé **nucléocapside**, l'unité infectieuse du virus (Gaudin & *al.*, 1992). L'ARN viral n'est jamais nu, ni dans les particules virales ni dans les cellules infectées, mais il est étroitement associé avec la **nucléoprotéine (N)** (Albertini & *al.*, 2006) (Green & *al.*, 2006) qui sert de matrice pour la répllication et la transcription (Arnheiter & *al.*, 1985). En effet, il a été montré que lors de l'infection, après l'entrée du virus dans la cellule, la synthèse des ARN messenger n'est possible qu'à partir de cette matrice nucléoprotéine-ARN viral (Ding & *al.*, 2006). Deux protéines codées par le génome viral sont nécessaires pour synthétiser l'ARN viral de façon efficace et régulée (Emerson & Yu, 1975) (Pattnaik and Wertz 1990) : l'**ARN polymérase ARN dépendante virale (L)** qui possède les activités de transcriptase et répllicase, responsable de la synthèse des ARNs pendant le cycle viral, son cofacteur, la **phosphoprotéine (P)**, essentielle pour la transcription et la répllication du génome viral.



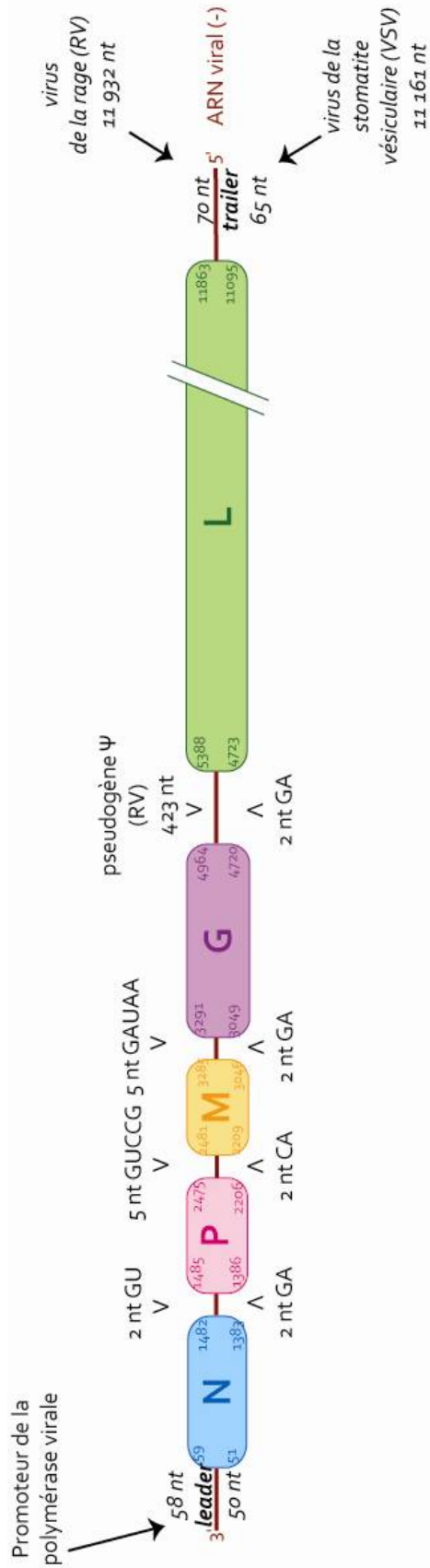
**Figure 11** - Représentation schématique de la particule virale en forme de balle de fusil des *Rhabdoviridae*. L'encadrer rouge nous permet de voir, schématiquement, l'intérieur de la particule et la nucléocapside virale hélicoïdale.

## B – Génome des *Rhabdoviridae*

Le génome des *Rhabdoviridae* est constitué d'un unique brin d'ARN de polarité négative (figure 12). Cet ARN viral code au minimum pour cinq protéines virales qui sont communes à tous les *Rhabdoviridae*, et dont les gènes sont organisés de façon similaire le long de la molécule d'ARN virale génomique (figure 12). On trouve ainsi successivement les gènes codant pour la nucléoprotéine (N), la phosphoprotéine (P), la protéine matrice (M), la glycoprotéine (G) et l'ARN polymérase ARN-dépendante (L) (Tordo, Poch, Ermine & Keith, 1986). La taille du génome est comprise entre 11 et 15 kb pour les *Rhabdoviridae* : 11 161 nucléotides (nt) pour le virus de la stomatite vésiculaire (VSV), et 11 932 nt pour le virus de la rage (RV). Il existe des séquences spécifiques à l'extrémité 3' du génome (région *leader*) et à l'extrémité 5' (région *trailer*), qui possèdent les signaux pour l'encapsidation du génome viral.

Les gènes viraux sont séparés par des régions intergéniques de tailles variables pour le virus de la rage, mais réduites à 2 nt pour le virus de la stomatite vésiculaire (Rose & Iverson, 1979). Pour le virus de la rage, les régions intergéniques entre les gènes N/P, P/M et M/G sont généralement courtes (2 à 5 nt) et conservées, par contre la région intergénique entre les gènes G et L est plus longue (423 nt) et de longueur variable chez les différentes souches de virus rabique. Cette séquence supplémentaire présente chez certains *Rhabdoviridae* dont le virus de la rage est nommée pseudogène  $\psi$  (Tordo, Poch, Ermine, Keith & al., 1986). Chez les *Novirhabdovirus*, notamment chez le virus de la nécrose hématopoïétique infectieuse (IHNV virus infectant les poissons), cette région constitue un sixième gène, codant pour une protéine NV (protéine *Non-Virion*). Cette protéine de petite taille (12 kDa) est exprimée dans les cellules infectées (Kurath & al., 1985) et constitue un facteur de virulence (Thoulouze & al., 2004).

Les mécanismes de l'expression du génome viral des Mononegavirales ont été largement exploré sur le modèle du virus de la stomatite vésiculaire mais aussi sur le virus de Sendai (*Paramyxovirus*) (Banerjee, 1987) (Tordo & al., 1988) (Vidal & Kolakofsky, 1989) (Banerjee & Barik, 1992) (Tordo & Kouknetzoff, 1993) (Barr & al., 2002) (Kolakofsky & al., 2004) (Whelan & al., 2004).



**Figure 12 - Organisation schématique du génome des *Rhabdoviridae*, notamment du virus de la rage (RV) (*Lyssavirus*) et du virus de la stomatite vésiculaire (VSV) (*Vesiculovirus*). La longueur des séquences *leader* et *trailer* sont indiquées en nombre de nucléotides. La position de chaque gène est précisée. La localisation du pseudogène  $\psi$  est indiquée : cette séquence est variable suivant les souches de *Lyssavirus*.**

## C – Le cycle viral

La plupart des *Rhabdoviridae* sont transmis mécaniquement par blessure, morsure ou aérosol. Les tropismes tissulaires et la pathogénicité varient individuellement avec chaque virus. Les *Rhabdoviridae* ont un cycle viral exclusivement cytoplasmique, à l'exception des *Nucleorhabdovirus* dont le cycle viral présente une phase nucléaire, bien que certaines protéines virales pénètrent dans le noyau pour interagir avec les partenaires cellulaires. Le cycle viral se décompose en huit étapes (figure 13) qui sont l'adhésion, l'endocytose, la fusion membranaire, la transcription, la traduction, la réplication, l'assemblage et le bourgeonnement et que l'on peut regrouper dans trois processus principaux.

### **Entrée du virus**

Le cycle viral débute par l'**adhésion** du virus à la membrane plasmique cellulaire. Les spicules de glycoprotéine présente à la surface de la particule virale reconnaissent les récepteurs cellulaires. Cette interaction induit l'internalisation du virus par **endocytose**. L'abaissement du pH dans le compartiment endosomal active la glycoprotéine qui devient alors fusogénique. Le processus de **fusion membranaire** entre les membranes cellulaires endosomales et virales est déclenché, libérant ainsi le complexe ribonucléoprotéique dans le cytoplasme de la cellule hôte.

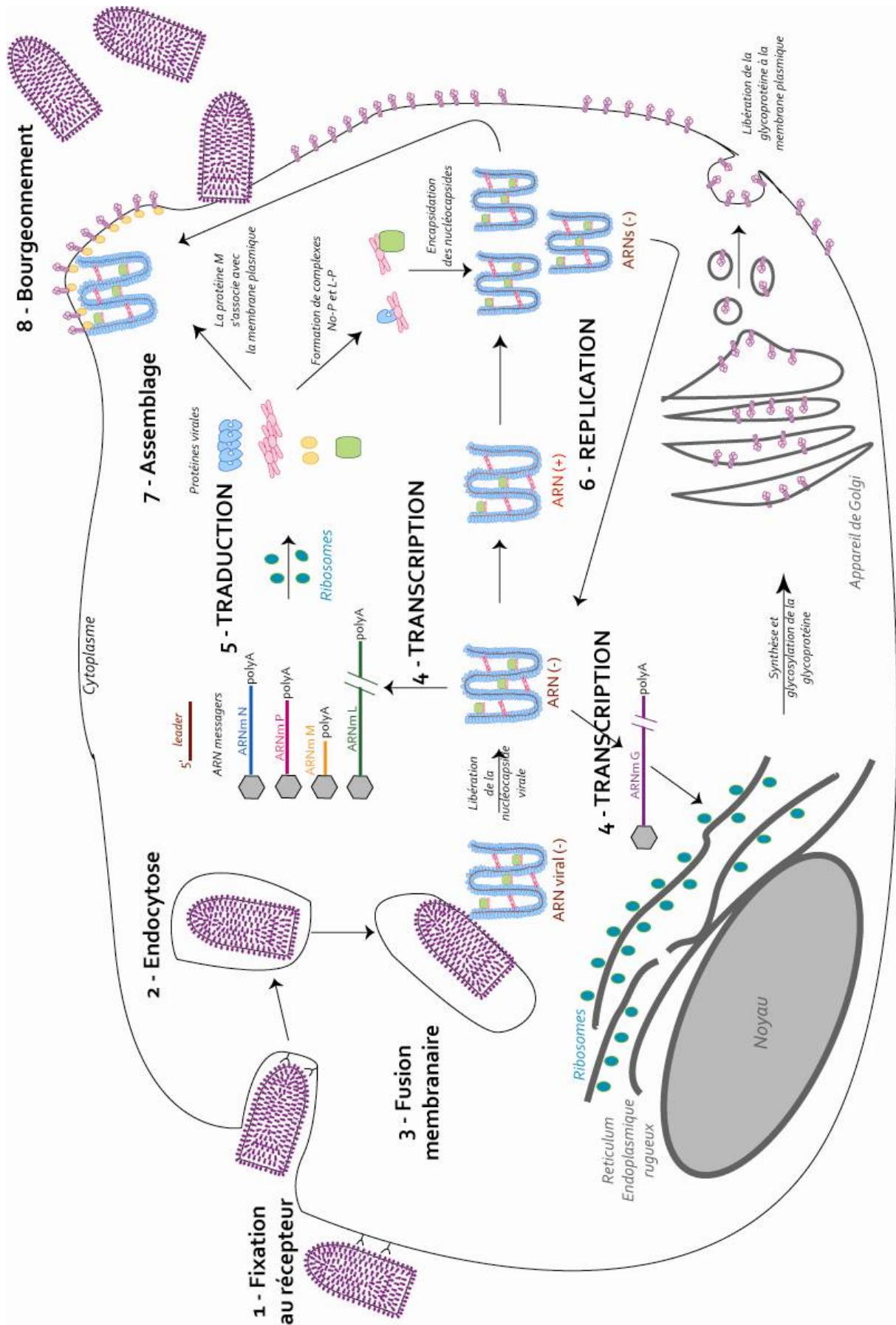
### **Multipliation du virus**

L'étape suivante est la **transcription**. L'ARN viral est transcrit en ARN messagers (ARNms) coiffés et polyadénylés par le complexe ARN polymérase ARN-dépendante – phosphoprotéine. Ces ARNms sont ensuite **traduits** en protéines virales par la machinerie cellulaire. Lorsque la quantité de protéines virales néo-synthétisées atteint un niveau suffisant, la polymérase virale passe en « mode **réplication** ». Elle ignore alors les signaux de début et de fin de gène et synthétise une molécule d'ARN positif, l'anti-génome, qui servira de matrice pour la synthèse de nouveaux génomes viraux. Ceux-ci peuvent alors servir soit de matrice pour la synthèse des ARNm viraux (transcription) soit pour la synthèse de nouveaux génomes viraux (réplication) (Banerjee, 1987). Dès le début de leur synthèse, les génomes et anti-génomes viraux sont encapsidés par la nucléoprotéine néo-synthétisée, qui joue un rôle structural crucial.

### **Sortie de la cellule**

Pour l'**assemblage** des nouvelles particules virales, les ARN viraux néoformés associés à la nucléoprotéine et à la polymérase virale sont exportés vers la membrane plasmique. Les glycoprotéines sont transportées *via* l'appareil de Golgi. La protéine matrice se place sur la face interne de la membrane plasmique où elle interagit avec les queues cytoplasmiques des spicules de glycoprotéines ainsi qu'avec les nucléocapsides dont elle assure la condensation en hélice. De nouvelles particules virales compactes sont formées et peuvent alors **bourgeonner** hors de la cellule hôte, à l'aide de protéines cellulaires.





**Figure 13** - Cycle viral des *Rhabdoviridae*. Le cycle viral se déroule dans le cytoplasme de la cellule hôte et il est décomposé en 8 étapes : 1- Fixation au récepteur, 2 – Endocytose, 3 – Fusion membranaire, 4 – Transcription, 5 – Traduction, 6 – Réplication, 7 – Assemblage et 8 – Bourgeonnement.

## D – Structures et fonctions des protéines virales

### 1 – Protéines impliquées dans le bourgeonnement

L'enveloppe virale est constituée à 50% en masse de lipides et à 50% de protéines. La composition lipidique de la membrane virale est relativement proche de la membrane cellulaire dont elle dérive, mais elle montre cependant un enrichissement en cholestérol. Cet enrichissement peut s'expliquer par un bourgeonnement du virus à partir de microdomaines lipidiques particuliers, tels que les radeaux lipidiques (*rafts*). Deux protéines virales sont impliquées dans la formation de l'enveloppe virale : la protéine de matrice (M), localisée à l'intérieur de la particule virale et la glycoprotéine (G), ancrée dans la membrane virale.

#### La Glycoprotéine

La glycoprotéine est l'unique protéine membranaire intrinsèque des *Vesiculovirus* et *Lyssavirus*. Elle forme des spicules trimériques visibles à la surface du virus par microscopie électronique (figure 10, A) (Gaudin & al., 1992). Chez le virus de la rage (RV), il existe environ 400 spicules de glycoprotéines sur la membrane virale (Flamand & al., 1993). La glycoprotéine joue de nombreuses fonctions au cours du cycle viral. Lors de l'entrée du virus dans la cellule hôte, elle assure les fonctions de reconnaissance du récepteur du virus, puis elle catalyse la fusion des membranes virales et cellulaires, permettant ainsi la libération du génome viral dans le cytoplasme de la cellule infectée. Aussi, elle a un rôle essentiel lors de l'assemblage et du bourgeonnement viral, en association avec la protéine de matrice (M) (Rolls & al., 1994). La glycoprotéine porte des sites antigéniques responsables de la réponse immunitaire, capable d'induire la production d'anticorps.

La glycoprotéine est reconnue par des récepteurs spécifiques à la surface de la cellule hôte. Pour le virus de la rage, ce récepteur doit être présent à la surface de nombreux types cellulaires impliqués lors de l'infection virale : les cellules musculaires et cutanées, les neurones et les cellules du cerveau. Pour le virus de la stomatite vésiculaire (VSV), le tropisme du virus semble diriger celui-ci vers les cellules épithéliales, bien que le VSV soit capable d'infecter un large spectre de types cellulaires. Le récepteur cellulaire n'a été identifié ni pour RV ni pour VSV. La nature de ces récepteurs reste un sujet de débat, et l'accumulation des données montre une réelle complexité du problème. De plus, la plupart des études ont été réalisées avec des souches qui ont été adaptées pour la multiplication en culture cellulaire. Cette adaptation peut être due à la capacité de souches de RV stables d'utiliser des récepteurs ubiquitaires présents sur tous les types cellulaires testés. Ces récepteurs ubiquitaires pourraient être des molécules telles que des phospholipides (Superti & al., 1984), des gangliosides (Conti & al., 1988) (Superti & al., 1986) ou des protéines (Wunner & al., 1984) (Broughan & Wunner, 1995) (Gastka & al., 1996). Plusieurs études

indiquent que le récepteur à l'acétylcholine nicotinique (nAChR) est un récepteur pour le virus de la rage (Lentz & *al.*, 1987) (Baer & *al.*, 1990) (Gastka & *al.*, 1996). Néanmoins, le nAChR est principalement présent sur les cellules musculaires, et le virus de la rage infecte les neurones qui eux n'expriment pas ce récepteur. Il a été proposé que, chez les animaux, ce récepteur permet au virus de la rage des rues de se multiplier localement dans les myotubes, uniquement sur la zone d'inoculation du virus (Burrage & *al.*, 1985). Cela faciliterait la pénétration du virus vers les neurones. Plus récemment, il a été montré que la molécule d'adhésion des cellules neuronales NCAM (*neuronal cell adhesion molecule*) facilite l'entrée du virus dans des souches de cellules de laboratoire (Thoulouze & *al.*, 1998), et il a ainsi été proposé comme autre récepteur pour le virus de la rage. Pour finir, une approche d'expression utilisant la glycoprotéine soluble de RV a menée à l'identification du récepteur de neurotrophine p75 (récepteur de basse affinité du facteur de croissance neuronale p75NTR, membre de la famille des récepteurs du TNF, *tumor necrosis factor*) comme possible récepteur (Tuffereau & *al.*, 1998). Des études sur la glycoprotéine du *European Bat lyssavirus 2* ont révélés que la protéine (mais pas celle des autres génotypes) peut aussi se fixer sur p75 (Tuffereau & *al.*, 2001). La phosphatidylsérine, un phospholipide chargé négativement a été proposé comme récepteur pour le VSV (Schlegel & *al.*, 1983). Cependant, seules les cellules en apoptose exposent ce lipide sur leur monocouche externe, suggérant que ce récepteur doit être considéré avec attention. Peu de choses sont connues concernant le récepteur des autres membres de la famille. Récemment, il a été suggéré que la fibronectine est le récepteur pour le virus de la septicémie hémorragique (VHSV), un *Rhabdoviridae* de poisson (Bearzotti & *al.*, 1999). Ce récepteur a été identifié en utilisant des anticorps monoclonaux (mAbs) générés contre les cellules de gonades des truites arc-en-ciel. Ces mAbs ont été sélectionnés pour leur capacité à protéger les cellules d'une infection par le VHSV. De façon similaire, la fibronectine est le récepteur pour l'entrée du virus de la rage dans des cellules BHK21. Ainsi, la fibronectine pourrait jouer un rôle général dans l'entrée des *Rhabdoviridae*.

La protéine G est synthétisée sous la forme d'un précurseur de 511 acides aminés (aa) chez VSV et 524 chez RV. Ce précurseur possède des aa en N-terminal qui constituent un peptide signal (16 pour VSV et 19 pour RV) et mènent à l'insertion de la protéine dans le réticulum endoplasmique rugueux (RER) (figure 13) pendant la traduction (ensuite la protéine est exportée vers la membrane virale par voie de sécrétion). Ce peptide signal est clivé par la signalase dans le RE et la glycoprotéine acquiert sa longueur définitive de 495 aa chez VSV et 505 aa chez RV. La protéine G mature est composée de trois domaines :

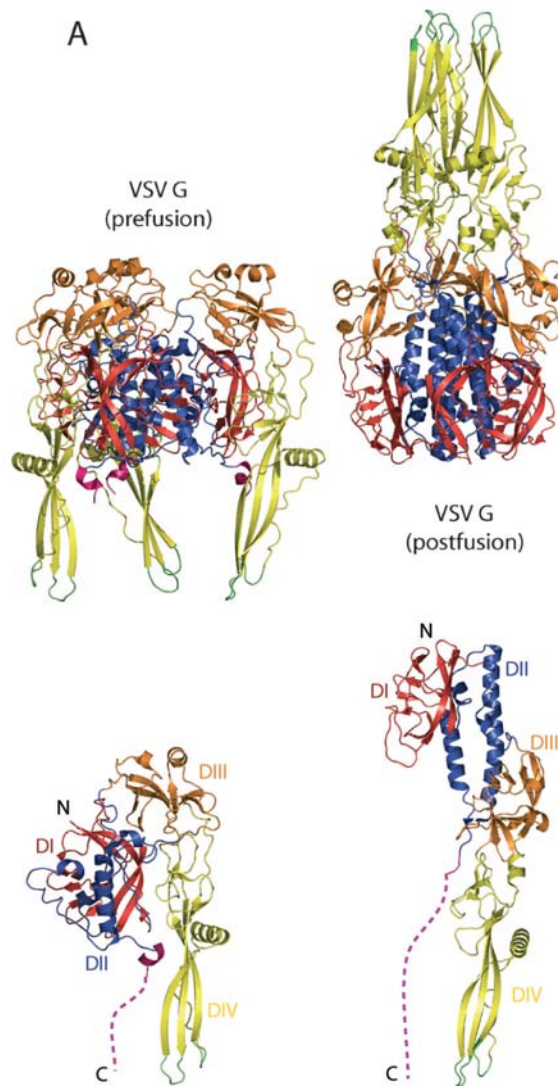
- La partie N-terminale de la protéine G constitue l'ectodomaine, exposé à la surface de la particule virale. Cet ectodomaine contient les sites de glycosylations et les déterminants de l'immunogénicité.
- Le domaine transmembranaire composé d'acides aminés majoritairement hydrophobes, replié en hélice  $\alpha$  (449 à 470 chez VSV, 440 à 460 chez RV),
- L'extrémité C-terminale de la protéine qui constitue un segment cytoplasmique hydrophobe et qui peut interagir avec les autres protéines virales internes.

La protéine G subit plusieurs modifications post-traductionnelles. Elle est glycosylée à l'intérieur de l'appareil de Golgi avec l'acide palmitique (Schmidt & Schlesinger, 1979) (Gaudin, Tuffereau, Benmansour & *al.*, 1991). La glycosylation se fait sur des résidus asparagines (aa 319 et 37 étant les sites de glycosylations majeurs chez RV), un de ces sites étant strictement conservé. La protéine G est également palmitoylée (Rose & *al.*, 1984) (Gaudin, Tuffereau, Benmansour & *al.*, 1991).

La glycoprotéine est aussi responsable de la fusion de la membrane du virus avec celle de l'endosome. La fusion du virus de la rage a été étudiée en détail (Gaudin, Tuffereau, Segretain & *al.*, 1991) (Gaudin & *al.*, 1993). La fusion est seulement déclenchée à pH bas et ne nécessite pas une composition spécifique de la membrane. La fusion est optimale à un pH aux alentours de 6,0, mais n'est plus détectée lorsque le pH est au dessus de 6,4. Ces valeurs de pH sont similaires à celle déterminées pour VSV (Crimmins & *al.*, 1983) (Carneiro & *al.*, 2003) et pour un *Rhabdoviridae* de poisson (Gaudin & *al.*, 1999). La préincubation des *Rhabdoviridae* à bas pH en absence d'une membrane cible inhibe les propriétés de fusion virale. Cependant, la perte de ces propriétés de fusion est réversible en réajustant le pH au dessus de 7. Ceci est la principale différence entre les *Rhabdoviridae* et les autres virus qui fusionnent aussi à bas pH, pour lesquels la fusion inhibée à bas pH est irréversible (Gaudin, 2000).

A la surface du virion, la protéine G est présente sous trois états conformationnels différents, caractérisé par microscopie électronique, sensibilité aux protéases et reconnaissance avec des anticorps monoclonaux (Gaudin, Tuffereau, Segretain & *al.*, 1991) (Gaudin & *al.*, 1993) (Gaudin, Tuffereau & *al.*, 1995) (Roche & Gaudin, 2002). L'état natif (N) est détecté à la surface du virus pour un pH supérieur à 7. L'état actif (A) est détecté immédiatement après l'acidification, et permet à la glycoprotéine de s'ancrer dans la membrane cible pour amorcer la fusion (Durrer & *al.*, 1995). Après une incubation prolongée à pH acide (pH < 6,4), la protéine G est dans une conformation inactive pour la fusion (I) qui est antigéniquement distincte de l'état N. Il existe un équilibre pH-dépendant entre ces états qui est décalé vers l'état I à bas pH (Roche & Gaudin, 2002). Des résultats similaires ont été obtenus pour la protéine G de VSV (Carneiro & *al.*, 2003) (Clague & *al.*, 1990) (Pak & *al.*, 1997). Finalement, les résultats obtenus jusqu'à présent indiquent, que pour les *Rhabdoviridae*, la séquence des événements qui aboutissent à la fusion membranaire sont similaires à ceux des autres virus enveloppés et que le complexe de fusion est composé de plusieurs trimères (Roche & Gaudin, 2002).

Récemment, les structures cristallographiques de la glycoprotéine de VSV dans ses conformations préfusion (Roche & *al.*, 2007) (code pdb : 2j6j) et postfusion (Roche & *al.*, 2006) (code pdb : 2cmz) ont été résolue (figure 14). Les réarrangements structuraux des différents domaines de la glycoprotéine sont très importants entre les deux états (figure 14).



**Figure 14** - Structure d'ensemble de l'ectodomaine de la glycoprotéine du VSV dans ses conformations pré et post fusions. VSVG préfusion, code pdb : 2j6j, residus 1–413) ; VSV G postfusion, code pdb : 2cmz, residus 1–410). La glycoprotéine est coloré en fonction de ces domaines : domaine I (DI) : rouge ; domaine II (DII) : bleu ; domaine III (DIII) : orange et domaine IV (DIV) : jaune ; les boucles responsables de la fusion en vert, et les lignes pointillées violettes correspondent au domaine C-terminal. Partie supérieure de l'image : trimère de protéine G, partie inférieure : monomère de protéine G. D'après (Roche & *al.*, 2008).

La structure atomique de la conformation post-fusion révèle que la protéine G de VSV est homologue avec la glycoprotéine B du virus Herpes Simplex (famille des *Herpesvirus* dont le génome est un double brin d'ADN) (Heldwein & *al.*, 2006), et qu'elle combine des caractéristiques des protéines de fusions de classe I et de classe II. La structure postfusion présente une conformation en feuilletts  $\beta$  caractéristique des protéines fusions de classe II, alors que le domaine de trimérisation de la protéine présente une organisation en hélices  $\alpha$  caractéristique des protéines de fusion de classe I. Lorsqu'on compare les états pré et post fusions de la protéine G de VSV, on remarque une réorganisation considérable des domaines de la molécule, qui évoque la protéine fusion (F) des *Paramyxovirus*.

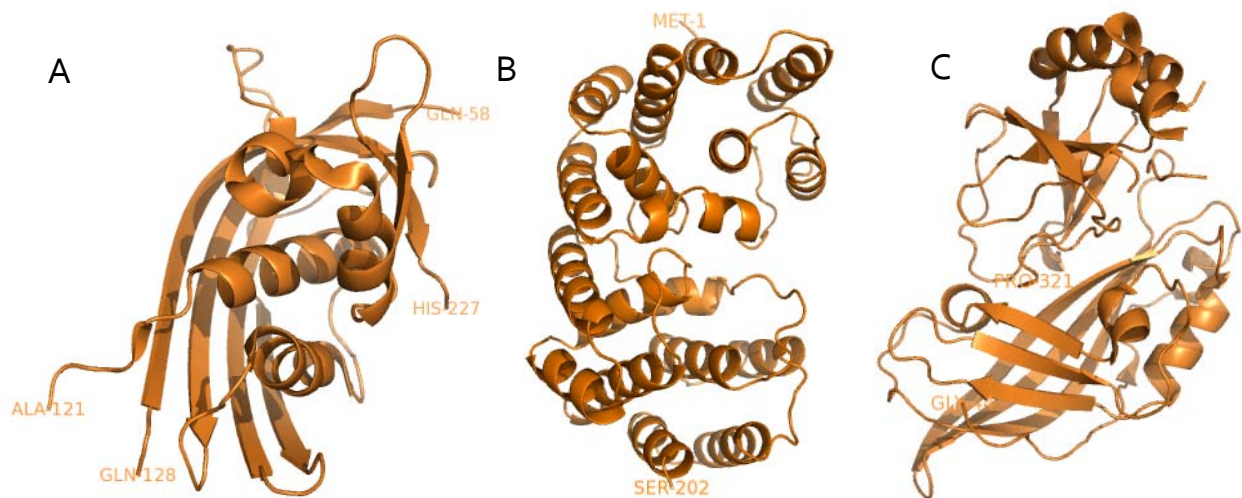
## La protéine Matrice (M)

La protéine Matrice (M) est principalement impliquée dans les étapes d'assemblage et de bourgeonnement viral. Elle facilite la condensation hélicoïdale de la nucléocapside (Newcomb & *al.*, 1982). Elle est composée de 202 acides aminés (aa) pour le virus de la rage, 229 aa pour le VSV. Chez RV, la protéine M est palmitoylée (Gaudin, Tuffereau, Benmansour & *al.*, 1991) alors que chez VSV, elle est phosphorylée (Kaptur & *al.*, 1995). Une sous population de protéine M est ubiquitinylée dans le but de recruter la machinerie cellulaire pour le bourgeonnement (Harty & *al.*, 2001). En effet, l'extrémité N-terminale de la protéine M contient deux motifs dits « tardifs », impliqués dans le bourgeonnement viral. Le premier de ces motifs, PPPY, est observé pour de nombreux virus (Harty & *al.*, 2000). Il interagit avec la protéine Nedd4, une ubiquitine ligase. Des essais réalisés *in vitro* ont montré que Nedd4 est capable de fixer des ubiquitines à la protéine M (Harty & *al.*, 2001). L'ubiquitination des protéines peut avoir des effets divers : la polyubiquitination (ajout d'ubiquitine en chaîne sur un même résidu) permet l'adressage de la protéine cible au protéasome, pour sa dégradation ; la multi-ubiquitination (ajout d'ubiquitine sur des résidus différents) permet l'adressage de la protéine cible au niveau de compartiments cellulaires tels que les corps multivésiculaires. Le second motif, PSAP, est également répandu parmi les virus, a un rôle encore inconnu (Irie & *al.*, 2004).

Pour VSV, il existe environ 1800 exemplaires de la protéine M au sein d'un virion (Thomas & *al.*, 1985). La protéine M est une protéine membranaire extrinsèque, associée à la face interne de la membrane virale. Pour VSV, il a été proposé que la protéine M formerait des structures en forme de cigare autour desquelles la nucléocapside s'enrouleraient (Barge & *al.*, 1993). La protéine M serait alors en contact avec la membrane virale qu'au niveau des extrémités de cette structure en cigare. Les protéines matrices sont globalement basiques, elles comportent des domaines hydrophobes mais elles ne sont ni insérées dans la membrane lipidique ni synthétisées à l'intérieur du RE. La protéine M semble interagir avec les têtes hydrophiles de lipides chargées négativement, comme la phosphatidylsérine. La protéine M a une forte affinité pour elle-même, ce qui se traduit par une forte polymérisation (et/ou agrégation) à haute concentration en protéine (McCreedy & *al.*, 1990), ou basse concentration saline. Ces polymères ont une forme allongée caractéristiques, rappelant ces structures « en cigare » autour desquelles s'organise la nucléocapside virale (Barge & *al.*, 1993). Il a été proposé que la polymérisation de la protéine M serait directement impliquée dans la compaction de la nucléocapside virale (Gaudin, Barge & *al.*, 1995). Pour VSV et le virus de Sendai (*Paramyxovirus*), il a été montré que la protéine matrice interagissait en même temps avec la membrane et la nucléocapside (Chong and Rose 1993) (Stricker & *al.*, 1994). Cependant, la position exacte de la protéine matrice (M) reste controversée.

Les structures atomiques de plusieurs protéines matrices de virus à ARN négatif sont connues (figure 15) : celle du VSV (Gaudier & *al.*, 2002), celle du virus de la grippe (Sha and Luo 1997) et celle du

virus Ebola (Dessen & *al.*, 2000). Plus récemment, la protéine M du virus Lagos Bat a été résolue (Assenberg & *al.*, 2008), et ce même groupe a obtenu la structure de la protéine M du virus de la rage (VIZIER/SPINE2 Workshop on Structural Biology, Juillet 2008, Vienne, Autriche) mais les structures ne sont toujours pas disponible sur la *Protein Data Bank* (PDB).



**Figure 15** - Structures atomiques des protéines matrices des virus à ARN négatifs. (A) Structure de la protéine M du VSV, fragment résistant à la thermolysine (58-121 et 128-227, code pdb : 1LG7). (B) Structure de la protéine M1 du virus de la grippe (résidus 2-158, code pdb : 1AA7). (C) Structure de la protéine M de la protéine matrice VP40 du virus Ebola (résidus 31-326, code pdb : 1ES6).

La digestion de la protéine M de VSV par la thermolysine permet d'éliminer l'extrémité N-terminale flexible et de cliver une boucle hydrophobe entre les positions 121 et 128. Les fragments 58-121 et 128-229 obtenus restent associés et ne s'agrègent pas, ce qui a permis leur cristallisation (Gaudier & *al.*, 2001). Ils forment un unique domaine constitué principalement d'un large feuillet  $\beta$  à 5 brins contre lesquels sont plaqués deux hélices  $\alpha$ . Cette structure constitue un nouveau repliement (Gaudier & *al.*, 2002), différent des repliements des protéines matrices du virus de la grippe et du virus Ebola (figure 15). Il semblerait, que le repliement de la protéine M de VSV (*Vesiculovirus*) et celle de la protéine M du virus Lagos Bat (*Lyssavirus*), sont les mêmes (Stephen GRAHAM, VIZIER/SPINE2 Workshop on Structural Virology, 14-16 juillet 2008, Vienne, Autriche).

La protéine M joue un rôle important dans la régulation de la cellule hôte. Pour les *Vesiculovirus*, notamment le VSV, il a été montré que la protéine M inhiberait la synthèse des ARNm cellulaires (Her & *al.*, 1997). La présence de la protéine M dans le noyau bloquerait le transport bidirectionnel des protéines et ARN cellulaires entre le noyau et le cytoplasme des cellules infectées (Glodowski & *al.*, 2002). La protéine M, en interagissant avec les pores nucléaires, améliore la disponibilité de la machinerie cellulaire de traduction en inhibant l'export des ARN messagers cellulaires provenant du noyau (Her & *al.*, 1997) (Petersen & *al.*, 2000). La protéine matrice du virus de la rage serait impliquée dans la régulation de la balance transcription/réplication : la protéine M inhiberait la transcription et stimulerait la réplication virale (Finke & Conzelmann, 2003) (Finke & *al.*, 2003).

## 2 – Protéines impliquées dans la multiplication du virus et la synthèse d'ARN

### L'ARN polymérase ARN-dépendante (L pour Large protéine)

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 & al., 1969). La première ARN polymérase ARN-dépendante (L pour Large protéine) mise en évidence fut celle du virus de la stomatite vésiculaire (VSV) (Baltimore & al., 1970). Chez les *Rhabdoviridae*, le nombre de molécules d'ARN polymérase ARN-dépendante virale par virion est faible, estimé à environ 50 molécules qui sont associées au complexe nucléoprotéine-ARN (Thomas & al., 1985). La polymérase est une protéine de 2 142 aa chez le virus de la rage (RV) et de 2 109 aa chez le VSV. Son gène représente ainsi 60% du génome virale (figure 12). Cette protéine est la plus fortement conservée au sein des *Rhabdoviridae*, et même au sein de l'ordre des *Mononegavirales* : les polymérases de VSV et du virus de la rage présentent ainsi 33 % de similarité de séquence.

L'ARN polymérase ARN-dépendante (L) constitue la sous-unité principale du complexe enzymatique polymérase virale chez les *Mononegavirales*. Associée à son cofacteur, la phosphoprotéine (P), la polymérase assure à la fois les activités de transcription et réplication du génome viral (Emerson & Yu, 1975), en utilisant exclusivement la matrice nucléoprotéine-ARN viral (Emerson & Wagner, 1972). Elle est aussi responsable des activités de méthyltransférase et de polyadénylase pour la traduction efficace des ARN messagers viraux. Cinq domaines particulièrement conservés ont été identifiés, ce qui a permis de repérer les sites probables d'interaction avec les matrices de réplication et les nucléotides (Tordo & al., 1988). Aussi, des analyses effectuées avec des constructions tronquées de la protéine L ont permis d'identifier les zones d'interactions entre l'ARN polymérase ARN-dépendante et la phosphoprotéine (P). Pour le virus de la rage (RV), la fixation de la phosphoprotéine s'effectue sur les 566 derniers résidus de la protéine L (Chenik & al., 1998). Pour le virus de Sendai, membre de la famille des *Paramyxovirus*, la phosphoprotéine se fixe au niveau des 400 premiers aa (Holmes & Moyer, 2002).

Il n'y a que très peu d'informations structurales sur les polymérases des virus à ARN négatif. Cependant, ces polymérases possèdent une certaine similarité de séquence avec les polymérases virales des virus à ARN positif dont plusieurs structures sont connues (Poch & al., 1989) notamment celle du virus de l'hépatite C (Lesburg & al., 1999) (Ago & al., 1999) mais plus récemment celles du virus de la Dengue (Yap & al., 2007) et du virus West Nile (Malet & al., 2007) (*Flaviviridae*). Il semble donc que certaines de ces données structurales puissent s'appliquer aux polymérases des virus à ARN négatifs (Kolakofsky & al., 2004). Cependant, la seule information structurale disponible pour un virus à ARN négatif concerne la



polymérase du virus de la grippe (*Orthomyxoviridae*) (Area & al., 2004) (Torreira & al., 2007) (He & al., 2008) (Obayashi & al., 2008).

Aucune connaissance structurale sur les ARN polymérases ARN-dépendantes des *Mononegavirales* n'est disponible à ce jour. Qanungo propose en 2004 qu'il est possible d'isoler deux complexes enzymatique contenant la polymérase virale à partir du virus de la stomatite vésiculaire (VSV) : un complexe « **transcriptase** » composé de la protéine L, de son cofacteur ainsi que de deux protéines cellulaires (le facteur d'élongation de la traduction 1 et la protéine de choc thermique HSP 60), et un complexe « **réplicase** » composé uniquement de la polymérase virale, de la phosphoprotéine et de la nucléocapside, mais sans la présence des partenaires cellulaires (Qanungo & al., 2004).

### La Nucléoprotéine (N)

La nucléoprotéine (N) est la protéine la plus abondante dans la particule virale. Sa fonction première est d'encapsider le génome et l'antigénome viral. La masse moléculaire de la nucléoprotéine du virus de la rage est de 50,5 kDa (450 aa) et celle du virus de la stomatite vésiculaire (VSV) de 47,4 kDa (422 aa). Peu de similitudes de séquences entre les nucléoprotéines des différents virus à ARN négatifs existent. La nucléoprotéine du virus de la rage et de VSV ne partagent que 36 % de similarité de séquence. Chez la plupart des *Mononegavirales*, la nucléoprotéine est organisée en 2 domaines : un domaine N-terminal qui fixe l'ARN viral (Buchholz & al., 1993) (Kouznetzoff & al., 1998) et un domaine C-terminal qui interagit avec la phosphoprotéine (Schoehn & al., 2001). Ce domaine C-terminal est intrinsèquement désordonné chez le virus de la rougeole (Longhi & al., 2003), et son repliement est induit par l'interaction avec la phosphoprotéine. Cependant pour le virus de la rage, ce domaine C-terminal semble former une structure définie (Schoehn & al., 2001).

Dans les cellules infectées, on retrouve la nucléoprotéine sous deux formes : soit liée à l'ARN et formant la matrice nucléocapside utilisée pour la transcription et réplication virale, soit liée à la phosphoprotéine formant un complexe nucléoprotéine-phosphoprotéine N<sup>0</sup>-P, le <sup>0</sup> indiquant que la nucléoprotéine est libre d'ARN. La nucléoprotéine du virus de la rage est phosphorylée au niveau de la sérine 389 par la caséine kinase II, kinase ubiquitaire des cellules, lorsqu'elle est présente sous la forme de complexes N-ARN (Kawai & al., 1999). Chez la nucléoprotéine de VSV, aucune phosphorylation n'a été identifiée.

### **Le complexe soluble nucléoprotéine-phosphoprotéine : N<sup>0</sup>-P**

Lorsque l'on coexprime dans un système de production recombinante la nucléoprotéine et la phosphoprotéine, une fraction de la nucléoprotéine forme un complexe N<sup>0</sup>-P, où la nucléoprotéine est présente sous une forme soluble, sans ARN. L'ultracentrifugation analytique, la spectrométrie de masse

et la microscopie électronique ont montré que ce complexe pour le virus de la rage était composé d'une nucléoprotéine associée à deux phosphoprotéines (Mavrakis & *al.*, 2003). La nucléoprotéine de ce complexe soluble n'est pas phosphorylée au niveau de la sérine 389 (Kawai & *al.*, 1999). En effet, la nucléoprotéine néo-synthétisée n'est pas immédiatement phosphorylée, mais elle s'associe à la phosphoprotéine. Ce complexe N<sup>0</sup>-P participe ensuite à la formation de complexes nucléoprotéine-ARN. Il semblerait que la nucléoprotéine non phosphorylée fixe plus efficacement la matrice ARN (Yang & *al.*, 1998) (Yang & *al.*, 1999), et que sa phosphorylation ait lieu pendant l'encapsidation. Pour le VSV, le complexe N<sup>0</sup>-P soluble produit de façon recombinante inhibe la transcription *in vitro* (Gupta and Banerjee 1997). Cependant, l'addition de complexe N<sup>0</sup>-P provenant de cellules infectées à un système reconstitué de transcription résulte dans la synthèse du génome. Ceci laisse à penser que le complexe N<sup>0</sup>-P joue le rôle de régulateur et permet, avec des facteurs cellulaires, le passage du mode transcription au mode réplication.

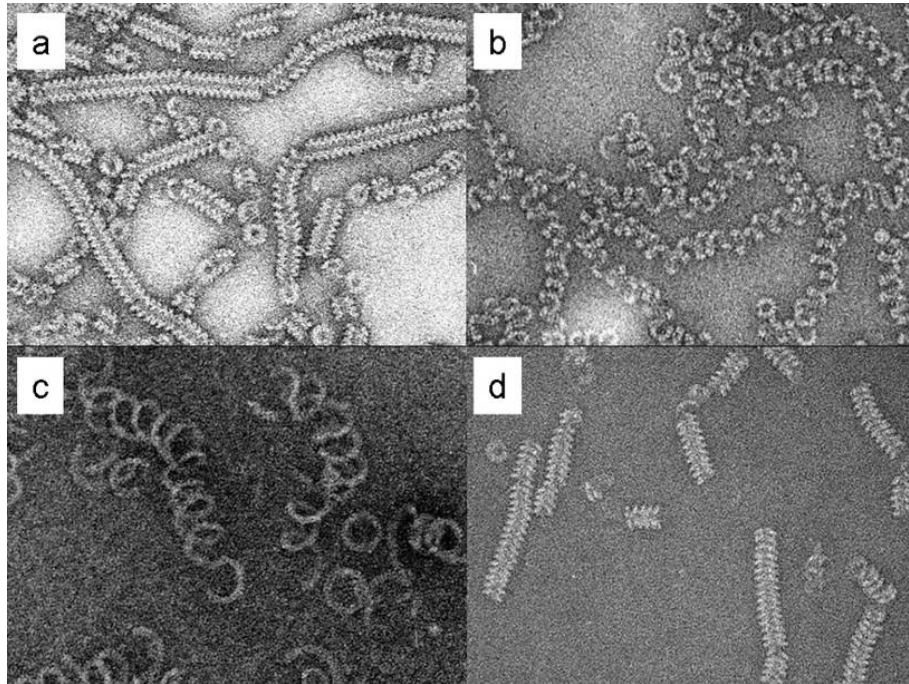
### **Les complexes nucléoprotéines – ARN**

#### La nucléocapside

Lorsque la nucléoprotéine est associée à l'ARN viral, sous la forme de nucléocapside, la nucléoprotéine protège l'ARN viral des ribonucléases (RNases), empêche la formation de structures secondaires au niveau des ARNs simples brins (Baudin & *al.*, 1994) (Gigant & *al.*, 2000) et évite la formation d'ARN double brin (ARN viral et messager) lors de la synthèse virale. Les nucléocapsides des virus à ARN négatifs ont une symétrie hélicoïdale, mais avec des paramètres variables, comme le montre les clichés de microscopie électronique (figure 16). L'interaction avec l'ARN est forte : il a été montré que les nucléocapsides de *Paramyxoviridae* et de *Rhabdoviridae* sont stables, et peuvent résister à des hautes concentrations en sel et une force de gravité importante (Blumberg & *al.*, 1984) (Heggeness & *al.*, 1980). Pour le virus de la stomatite vésiculaire (VSV), les nucléocapsides résistent à une centrifugation en gradient de chlorure de césium (CsCl) en présence de 8M d'urée, agent dénaturant fort des protéines (Iseni & *al.*, 1998). Les changements importants de ces nucléocapsides en réponse à la force ionique indiquent la plasticité de cette structure hélicoïdale. Notamment, ces changements pourraient être importants lors de la transcription et de la réplication virale, ou lors du bourgeonnement, un processus dans lequel l'enroulement et la flexibilité de ces nucléocapsides changeraient. L'expression de la nucléoprotéine des *Mononegavirales* dans différents systèmes d'expression (eucaryote ou procaryote) produit des complexes dont la morphologie et les propriétés biochimiques sont identiques à celles des nucléocapsides virales.

Des reconstructions hélicoïdales effectuées à partir d'images de microscopie électronique ont permis de déterminer les paramètres hélicoïdaux des nucléocapsides de certains virus, comme le nombre de sous-unités de nucléoprotéines par tour d'hélice. Pour le virus de la rougeole et de Sendai

(*Paramyxovirus*), le nombre de nucléoprotéines par tour est de 13 (Egelman & al., 1989) (Schoehn & al., 2004). Pour le virus respiratoire syncytial humain (RSV) et le virus Simien 5 (SV5), de 10 à 14 (Bhella & al., 2002). L'hétérogénéité et la flexibilité de ces nucléocapsides excluent totalement leur cristallisation. De plus, la détermination des paramètres hélicoïdaux de ces nucléocapsides à partir de clichés de microscopie électronique nécessite une homogénéité et une rigidité que l'on n'observe pas pour les nucléocapsides des *Rhabdoviridae*.



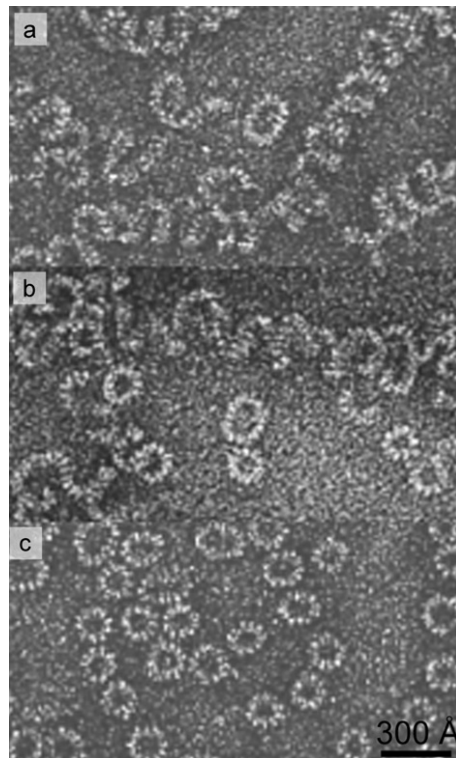
**Figure 16** - Images de microscopie électronique en coloration négative présentant la symétrie hélicoïdale de quelques nucléocapsides de *Mononegavirales*. (a) Nucléocapsides recombinantes du virus de la rougeole (*Paramyxoviridae*). (b) Nucléocapsides virales du virus de la rage (*Rhabdoviridae*). (c) Nucléocapsides recombinantes du virus de Marburg (*Filoviridae*). (d) Nucléocapsides recombinantes du virus Nipah (*Henipahviridae*).

#### Fixation à l'ARN

La nucléoprotéine des virus à ARN négatifs se fixe sur un nombre entier de nucléotides. Des études par mesures d'absorbances à 280 et 260 nm (virus de Marburg) et de microscopie électronique (VSV et Sendai) ont permis de déterminer le nombre de nucléotides fixés par une nucléoprotéine. Ce nombre est spécifique pour chaque famille au sein de l'ordre des *Mononegavirales* : la nucléoprotéine des *Paramyxoviridae* fixe six nucléotides (Egelman & al., 1989), celle des *Filoviridae* fixe entre 12 et 15 nt (Mavrakis & al., 2002) et celle des *Rhabdoviridae* en fixe neuf (Flamand & al., 1993) (Thomas & al., 1985). Pour les *Paramyxovirus*, excepté les *Pneumovirus*, il existe une règle appelée « **règle de six** » (Calain and Roux 1993), qui indique que le nombre exact de nucléotides du génome viral doit être un multiple de six pour que la polymérase virale réplique efficacement le génome. Ceci signifie que chaque nucléoprotéine interagit avec six nucléotides, et que les points d'interactions de la nucléoprotéine avec les six nucléotides ne sont pas équivalents (Vulliamoz and Roux 2001) (Iseni & al., 2000). En revanche, aucune « règle des neuf » n'a été mise en évidence chez les *Rhabdoviridae*.

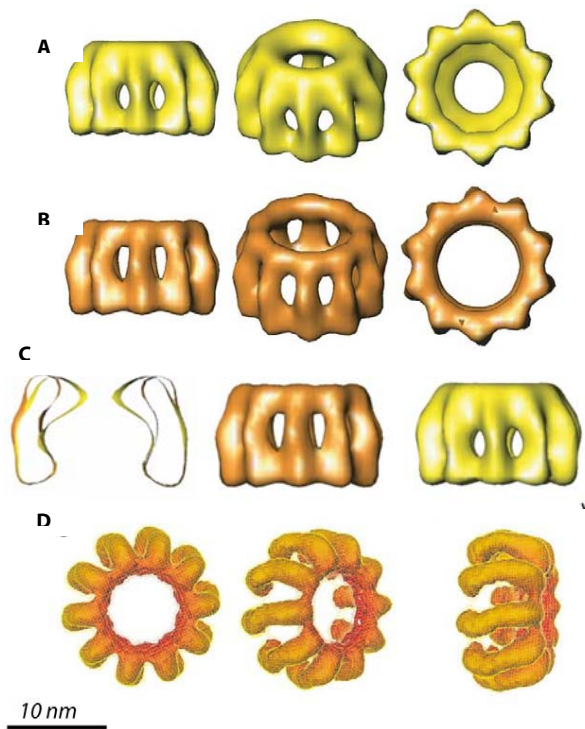
### Les anneaux de nucléoprotéine

Lorsque l'on exprime la nucléoprotéine du virus de la rage ou du VSV dans des cellules d'insectes ou des bactéries, celle-ci encapside des ARNs cellulaires de façon non-spécifiques pour former de longues structures hélicoïdales de type nucléocapside (vraisemblablement en fixant des ARN messagers de la cellule), ainsi que des structures circulaires comportant un nombre variable de nucléoprotéines (9 à 15) (fixation d'ARN courts de la cellule) (figure 17).



**Figure 17** - Les différents complexes nucléoprotéine-ARN du virus de la rage observés en microscopie électronique (coloration négative). (a) Nucléocapsides virales. (b) Nucléocapsides recombinantes produites en cellules d'insectes. (c) Anneaux issus de l'association de 9 à 15 sous-unités de nucléoprotéine et une molécule d'ARN simple brin. D'après (Iseni & *al.*, 1998) (Albertini & *al.*, 2008).

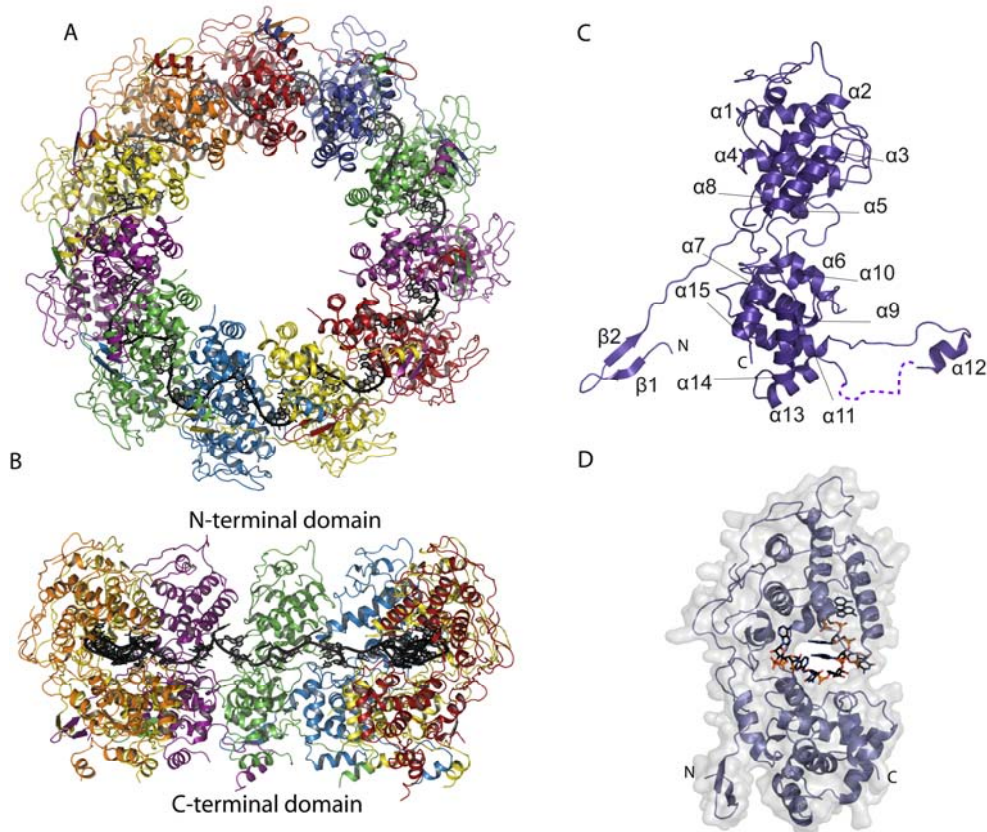
Les travaux réalisés dans notre laboratoire sur la nucléoprotéine du virus de la rage, ont permis, tout d'abord, d'obtenir une reconstruction 3D des anneaux formés de 10 nucléoprotéines à partir d'images de cryomicroscopie électronique (Schoehn & *al.*, 2001) (figure 18, A). La nucléoprotéine a une forme dite en « graine de haricot ». Le traitement de ces anneaux avec la trypsine entraîne un clivage au niveau du résidu 376 de la nucléoprotéine et élimine le domaine C-terminal, empêchant la fixation de la phosphoprotéine. La comparaison de la structure 3D en microscopie électronique des anneaux « natifs » N10 et des anneaux protéolysés (figure 18, B), qui ont perdu le domaine d'attchement de la phosphoprotéine a pu être identifié sur la partie supérieure de l'anneau (figure 18, C). Dans le cas du VSV, les travaux réalisés par Chen (Chen & *al.*, 2004) ont permis d'obtenir une reconstruction 3D à partir d'image de microscopie électronique des anneaux N10 de VSV, qui est très proche de celle des anneaux du virus de la rage (figure 18, D).



**Figure 18** - Reconstruction 3D de complexes nucléoprotéine-ARN en forme d'anneaux à partir d'images de microscopie électronique. (A) Reconstruction des anneaux N<sub>10</sub> du virus de la rage. (B) Reconstruction des anneaux N<sub>10</sub> du virus de la rage après protéolyse du domaine C-terminal à la trypsine. (C) Identification du domaine d'interaction de la nucléoprotéine avec la phosphoprotéine (Schoehn & *al.*, 2001). (D) Reconstruction des anneaux N<sub>10</sub> du virus de la stomatite vésiculaire (VSV) (Chen & *al.*, 2004).

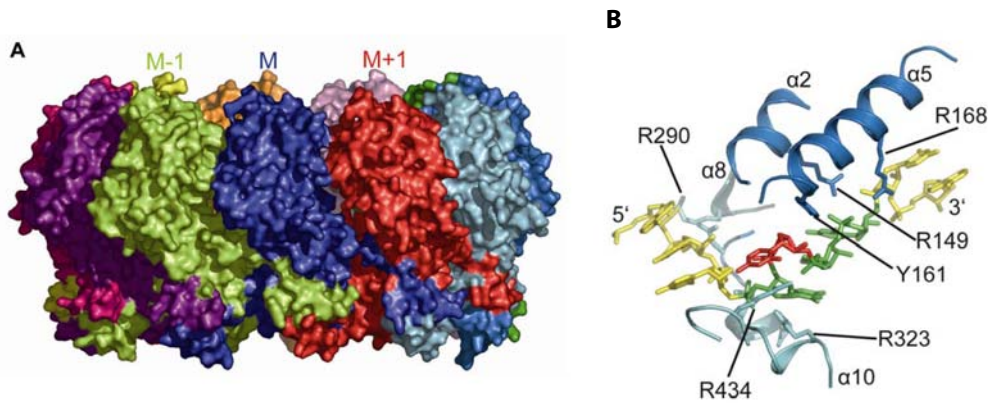
En 2006, la structure cristallographique des anneaux à 11 sous-unités de nucléoprotéine du virus de la rage a été obtenue à une résolution de 3,5 Å (Albertini & *al.*, 2006) et la structure cristallographique des anneaux à 10 sous-unités de nucléoprotéine du virus de la stomatite vésiculaire (VSV) a été obtenue à une résolution de 2,9 Å (Green & *al.*, 2006). Ce sont les premières structures de nucléoprotéines de virus à ARN négatifs où la nucléoprotéine est associée à l'ARN. Pour le virus de la rage, la nucléoprotéine a cristallisé dans le groupe d'espace P2<sub>1</sub>2<sub>1</sub>2, et la structure a été résolue par SAD (*single anomalous dispersion*). Les anneaux N11 (environ 550 kDa) sont positionnés face à face dans l'unité asymétrique et chaque anneau à un diamètre externe de 160 Å, un diamètre interne de 60 Å et une hauteur de 74 Å (figure 19, A & B). La protomère de nucléoprotéine consiste en deux domaines N et C-terminal (NTD et CTD) qui sont en contact avec neuf nucléotides d'un ARN simple brin, comme prédit (figure 19, C & D). Le domaine N-terminal (NTD : résidus 32 à 233) est replié en 6 hélices  $\alpha$  reliées par des longues boucles. Le domaine C-terminal (CTD : résidus 236 à 356 et 396 à 450) est composé de 11 hélices  $\alpha$  liés par des boucles plus courtes. Deux régions sont flexibles, et absentes dans la structure cristallographique (NTD : 105 à 118 et CTD : 376 à 397). Les deux domaines NTD et CTD agissent comme des mâchoires en séquestrant complètement la molécule d'ARN (figure 19). Les contacts les plus proches avec l'ARN sont entre le résidu N-terminal N157 et le résidu C-terminal P435, cependant la majorité des contacts avec l'ARN impliquent des résidus présents dans le domaine NTD. Les contacts entre protomères sont médiés

par une large surface d'interaction entre les CTDs (environ 2700 Å<sup>2</sup>), et par deux sous-domaines qui émergent du protomère en N et C-terminal et créent un contact entre les protomères N-1 et N+1 (figure 20, A), assurant la polymérisation mais aussi la flexibilité des structures formées par le complexe nucléoprotéine-ARN. En effet, cela permet de comprendre comment la nucléoprotéine peut former des structures en formes d'anneaux mais aussi en forme de nucléocapsides.



**Figure 19** - Structure cristallographique du complexe nucléoprotéine-ARN du virus de la rage sous la forme d'anneaux à 11 sous-unités de nucléoprotéine. (A) Vue du dessus de l'anneau N<sub>11</sub> : chaque nucléoprotéine est colorée différemment et l'ARN est schématisé en noir. (B) Vue de coté de l'anneau montrant l'intérieur de six protomères de nucléoprotéines. L'ARN est lié dans la cavité centrale formée par les deux domaines de la nucléoprotéine. (C) Protomère de nucléoprotéine indiquant la présence de deux domaines principaux avec deux autres domaines qui s'étendent du côté N et C-terminal. (D) Protomère de nucléoprotéine qui fixe les 9 bases d'ARN. D'après (Albertini & al., 2008).

L'ARN adopte dans l'anneau une conformation d'hélice gauche qui tourne dans le sens des aiguilles d'une montre (figure 20, B). Chaque nucléotide fait des contacts polaires par ses groupements phosphates (7 des 9 nucléotides (nt)) ou son ribose (2 des 9 nt). Les trois premières bases pointent vers le solvant et sont empilées avec les deux dernières bases liées au protomère de nucléoprotéine. Les nucléotides 4, 6 et 7 sont empilés entre eux et leurs bases pointent vers la protéine. La base du nucléotide 5 pointe quand à elle vers le solvant (figure 20, B).

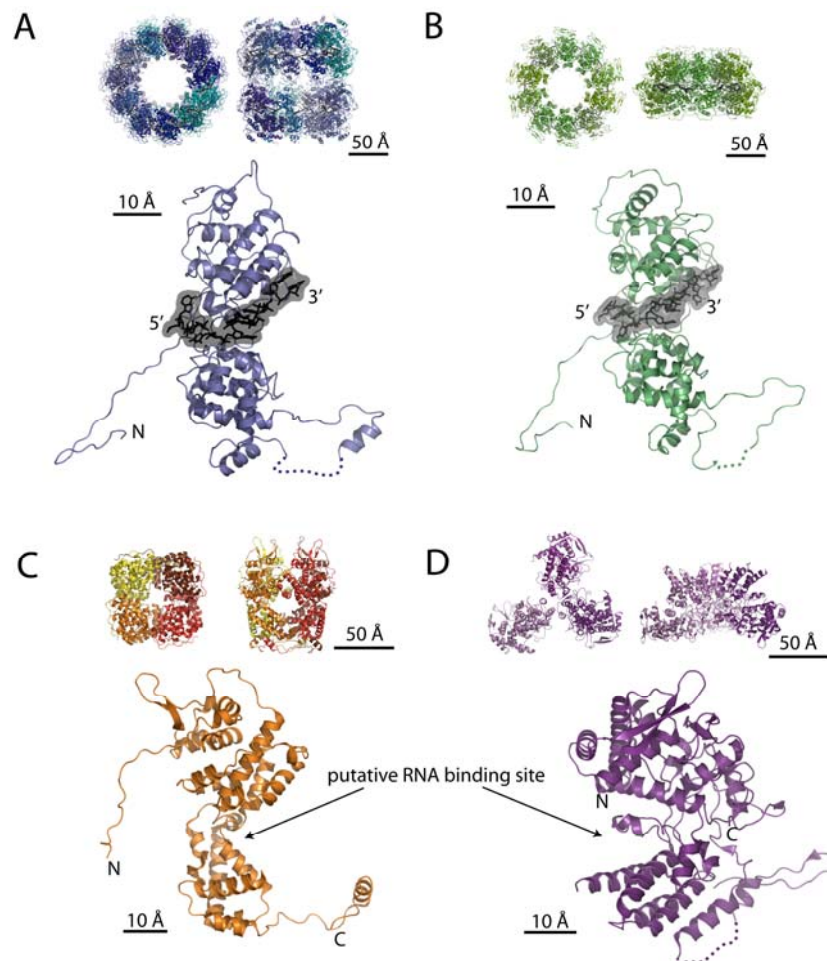


**Figure 20** - (A) Régions charnières responsables de la polymérisation de la nucléoprotéine. Modèle avec la représentation de surface de l'anneau N11, vue de côté. Chaque monomère est coloré différemment. La région charnière NTD du protomère M atteint le protomère M+1, et la région charnière CTD atteint le protomère M-1. Cet arrangement aboutit à l'interaction de la région charnière du NTD M-1 et du CTD M+1 à la surface du protomère M. (B) Interaction entre la nucléoprotéine et l'ARN. L'ARN est schématisé en batonnêt. La région NTD est colorée en bleu foncé et le CTD en bleu clair. Les nucléotides dont les bases sont accessibles au solvant sont indiqués en jaune et les bases qui font face à la protéine en vert. Le nucléotide central à la pointe de l'hélice gauche formé par le segment d'ARN est en rouge. Les résidus qui contactent les phosphates et les riboses du segment d'ARN sont schématisés en batonnêt. D'après (Albertini & *al.*, 2006).

#### Comparaison des structures des nucléoprotéines de virus à ARN négatifs

A l'heure actuelle, quatre structures de nucléoprotéine sont disponibles : deux nucléoprotéines de *Rhabdoviridae*, où chaque protomère est en liaison avec 9 nucléotides d'ARN simple brin (Albertini & *al.*, 2006) (Green & *al.*, 2006) (figure 21, A & B), la nucléoprotéine du virus de la maladie de Borna où la nucléoprotéine a été cristallisée sous la forme d'homo-tétramères, sans la présence d'ARN (Rudolph & *al.*, 2003) (figure 21, C) et la nucléoprotéine du virus de la grippe, où la nucléoprotéine a été cristallisée sous la forme d'homo-trimère, sans ARN (Ye & *al.*, 2006) (figure 21, D). Même si il est difficile de superposer les structures de nucléoprotéines des *Rhabdoviridae* avec celle des virus à ARN négatifs segmentés, on observe la présence de deux domaines principaux NTD et CTD, principalement composés d'hélices  $\alpha$ , responsables de l'interaction avec l'ARN. Cependant, la sensibilité de la nucléocapside aux nucléases est très différente entre les *Rhabdoviridae* et les virus à ARN négatifs dont le génome est segmenté. La nucléoprotéine du virus de la grippe fixe 24 nucléotides, et l'ARN est facilement éliminé par des traitements à hautes molarités en sel. Aussi, chez le virus de la grippe, les bases sont plus exposées au solvant (Baudin & *al.*, 1994) que chez les *Rhabdoviridae* et les *Paramyxoviridae* (Iseni & *al.*, 2000) (Iseni & *al.*, 2002). Les résidus basiques responsables de l'interaction avec l'ARN sont conservés chez la nucléoprotéine du virus de la rage et de la stomatite vésiculaire. Le monomère de nucléoprotéine du virus de la rage et du virus de la stomatite vésiculaire peuvent être superposés avec un *rmsd* de 2,18 Å (thèse d'Aurélien ALBERTINI 2006) (Luo & *al.*, 2007) (Luo & *al.*, 2007), ce qui indique une homologie des structures de ces nucléoprotéines, alors que leurs séquences primaires sont peu similaires.

La découverte de cette conformation fermée pour les anneaux de nucléoprotéines de *Rhabdoviridae* pose la question : comment l'ARN polymérase virale va accéder au génome viral ? En effet, la nucléoprotéine séquestre totalement l'ARN, ce qui est essentiel lors de l'infection virale pour la protection du génome, pour ne pas provoquer une reconnaissance par les facteurs cellulaires et entraîner une réponse immunitaire.



**Figure 21 - Comparaison des structures de nucléoprotéines de virus à ARN négatifs dans les états de polymères et de monomères.** (A) Nucléoprotéine du virus de la rage (RV). Anneaux N11 dans l'unité asymétrique et monomère. (B) Nucléoprotéine du virus de la stomatite vésiculaire (VSV). Anneaux N10 et monomère. (C) Nucléoprotéine du virus de la maladie de Borna (BVD). La forme tétramérique retrouvée en solution et dans le cristal est présentée. Le site de fixation possible de l'ARN est indiqué sur le monomère. (D) Nucléoprotéine du virus de la grippe. La protéine a cristallisé sous la forme d'un trimère. Le site de fixation possible de l'ARN est indiqué sur le monomère. D'après (Albertini & al., 2008).

### La Phosphoprotéine (P)

Tous les *Mononegavirales* possèdent une phosphoprotéine (P). Les séquences, tailles et degrés d'oligomérisation des phosphoprotéines de ces différents virus sont très variables (568 résidus pour le virus de Sendai, 297 pour le virus de la rage et 265 résidus pour le virus de la stomatite vésiculaire). Cependant, les phosphoprotéines ont des caractéristiques communes :

- elles ont le rôle de **chaperonne pour la nucléoprotéine** empêchant ainsi sa liaison aux ARN cellulaires,



- la phosphoprotéine se fixe sur le **complexe nucléoprotéine-ARN**,
- la phosphoprotéine est le **cofacteur de l'ARN polymérase virale ARN-dépendante**,
- les phosphoprotéines forment des **homo-oligomères**.

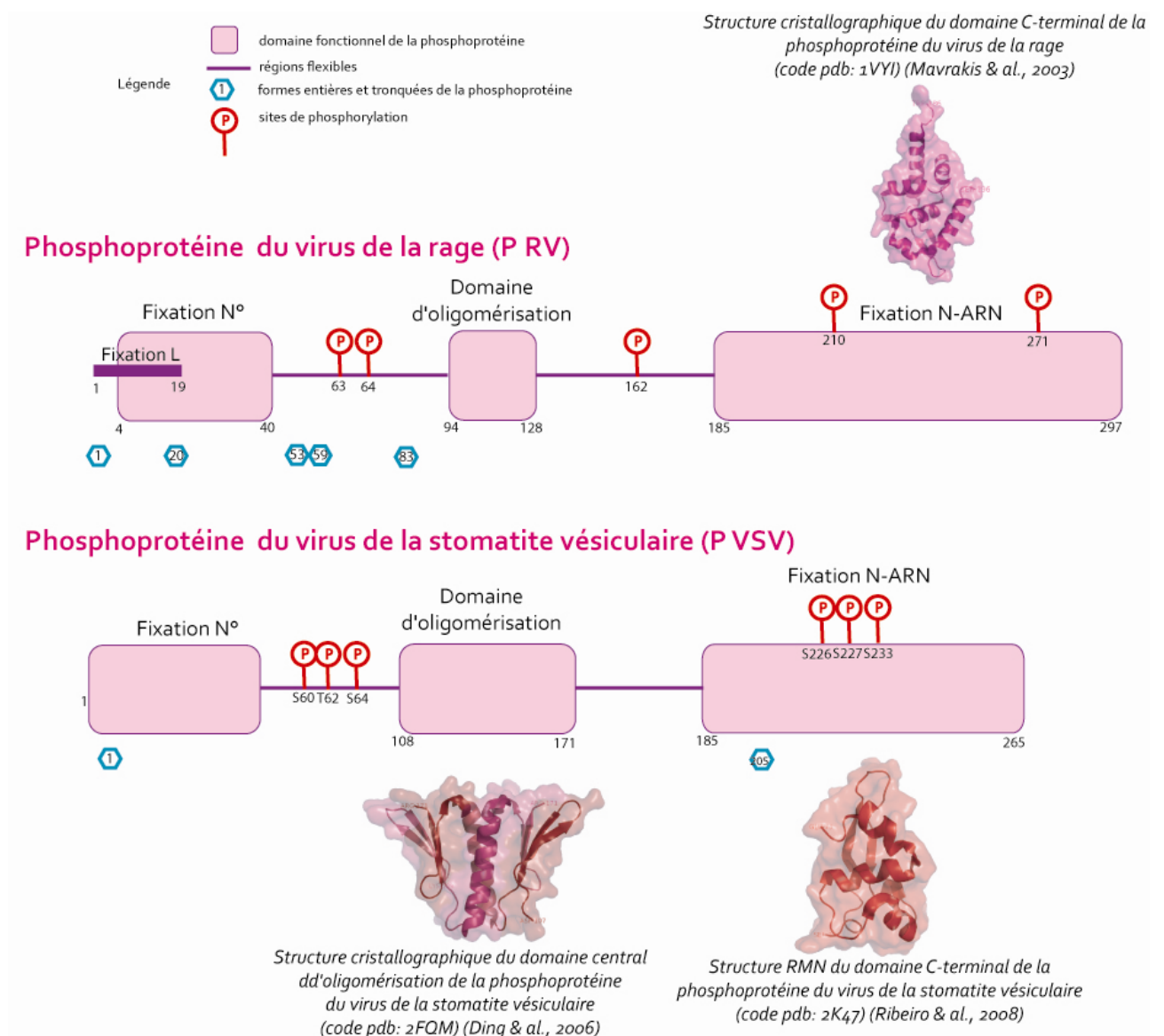
La phosphoprotéine des *Rhabdoviridae* a de multiples fonctions durant la réplication et est nécessaire pour la synthèse d'ARN et son encapsidation virale par la nucléoprotéine. Elle est présente dans la cellule infectée ainsi qu'au sein du virion avec plusieurs degrés de phosphorylation dont le rôle est à l'heure actuelle mal compris. La phosphorylation de résidus Ser et Thr en N-terminal de la protéine semble essentielle pour la multiplication virale, alors que le site de phosphorylation secondaire en C-terminal de la protéine reste encore inconnu. 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 & 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 & al., 1997). Cependant, cette kinase associée à la polymérase virale n'a pas été identifiée dans les complexes de transcription et de réplication du VSV (Qanungo & al., 2004).

### **Rôles majeurs de la phosphoprotéine**

La phosphoprotéine fixe la nucléoprotéine naissante en formant un complexe libre d'ARN nommé N<sup>0</sup>-P (avec une stoechiométrie N<sup>0</sup>-P<sub>2</sub>), nécessaire pour l'encapsidation virale (Masters and Banerjee 1988) (Masters and Banerjee 1988) (Gupta and Banerjee 1997). Chez le virus de la rage, la protéine P empêche l'encapsidation d'ARNs cellulaires non-spécifiques en se fixant à la nucléoprotéine par ses résidus 4 à 40 (Mavrakis & al., 2006) pour former des complexes N<sup>0</sup>-P solubles lors de la phase de transcription.

La phosphoprotéine est le cofacteur de la polymérase virale, formant un complexe à deux sous-unités dans lequel l'ARN polymérase ARN-dépendante possède les activités enzymatiques pour la transcription et réplication, incluant la synthèse de la coiffe et la polyadénylation de l'ARNm (Emerson and Wagner 1972) (Emerson and Yu 1975) (Banerjee 1987). En liant la nucléoprotéine et la polymérase virale, la phosphoprotéine permet à la polymérase de se fixer aux nucléocapsides et donc à la matrice nucléoprotéine-ARN. Pendant la réplication, la phosphoprotéine du virus de la rage peut fixer simultanément la polymérase par son domaine N-terminal (résidus 1 à 19) (Chenik & al., 1998), et les complexes nucléoprotéine-ARN par son domaine C-terminal (résidus 186-297) (Chenik & al., 1994) (Mavrakis & al., 2004), permettant ainsi l'interaction du complexe N-ARN avec la polymérase. La structure cristallographique du domaine C-terminal de la phosphoprotéine du virus de la rage, qui fixe la matrice nucléoprotéine-ARN, a été résolue (Mavrakis & al., 2004) (figure 22). Ce domaine est compact avec une

forme en « poire ». Il est composé de six hélices  $\alpha$ , d'une hélice  $3^{10}$  ainsi que de deux brins  $\beta$  formant un feuillet antiparallèle. 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. Cependant, peu d'informations sont connues sur les mécanismes d'action de cette protéine, et la protéine P pourrait aussi être impliquée dans le remodelage des nucléocapsides lors de la synthèse de l'ARN. La phosphoprotéine du virus de la stomatite vésiculaire possède la même organisation structurale, et les structures atomiques des domaines central (Ding & al., 2006) et C-terminal (Ribeiro & al., 2008) sont connues (figure 22).



**Figure 22** - Organisation schématique de la phosphoprotéine du virus de la rage (P RV) et de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV). Sur les schémas sont indiqués les domaines fonctionnels de la protéine ainsi que les régions flexibles. Les sites de phosphorylations majeures sont précisés. Les structures cristallographiques du domaine C-terminal de la phosphoprotéine du virus de la rage (code pdb : 1VYI) (Mavrakis & al., 2004) et de la phosphoprotéine du virus de la stomatite vésiculaire (code pdb : 2K47) (Ribeiro & al., 2008) et du domaine central d'oligomérisation de la phosphoprotéine du virus de la stomatite vésiculaire (code pdb : 2FQM) (Ding & al., 2006) sont représentés.

## **Structure de la phosphoprotéine**

L'oligomérisation de la phosphoprotéine (P) est une propriété commune aux *Mononegavirales*. Pour certains membres de la famille des *Paramyxoviridae* (sous-famille des *Paramyxovirinae* : virus Sendai et de la rougeole), le domaine d'oligomérisation est requis pour la transcription et la réplication, et l'oligomérisation n'est pas contrôlée par la phosphorylation (Curran, Marq & al., 1995) (Kolakofsky & al., 2004). Pour les autres membres de cette famille (RSV), la phosphorylation de la protéine P est requise pour les activités de transcription et de réplication (Barik & al., 1995) (Asenjo and Villanueva 2000). Dans la famille des *Rhabdoviridae*, les études réalisées avec VSV et RV n'ont apporté que des résultats peu clairs sur l'état d'oligomérisation et de phosphorylation nécessaires pour les activités de la phosphoprotéine.

Les études avec le VSV (genre *Vesiculovirus*) indiquent que la phosphorylation de la phosphoprotéine est nécessaire pour la transcription. L'oligomérisation de la phosphoprotéine est nécessaire pour la fixation à la polymérase virale et à la matrice N-ARN (Gao and Lenard 1995) (Raha & al., 2000) (Chen & al., 2006) et pour l'activité de transcription (Barik and Banerjee 1992) (Gao and Lenard 1995). L'oligomérisation de la phosphoprotéine est contrôlée par la phosphorylation dans la région N-terminale de la protéine (Gao and Lenard 1995) (Barik and Banerjee 1992) et par la concentration en protéine (Spadafora & al., 1996). De façon contradictoire, la phosphorylation dans la région N-terminale n'est pas nécessaire pour la formation des complexes libres d'ARN N<sup>0</sup>-P et pour la réplication de l'ARN viral (Masters and Banerjee 1988) (Spadafora & al., 1996) (Peluso and Moyer 1988). La structure cristallographique du domaine central de VSV (Ding & al., 2006), fragment résistant à la protéolyse par la protéase K (Ding & al., 2004), montre que ce domaine est dimérique et que son assemblage se fait par des interactions entre hélices  $\alpha$  et la formation de feuillets  $\beta$  intermoléculaires. La structure révèle que des contacts hélices  $\alpha$  - hélices  $\alpha$  sont à l'origine de la formation du dimère (figure 22). En se basant sur ces données structurales, on peut penser que l'entité biologique active de la phosphoprotéine du VSV est un dimère. Cependant, les études biochimiques ont suggéré que la forme non phosphorylée de la protéine P est monomérique et que les formes phosphorylées étaient dimériques, trimériques ou encore tétramériques (Gao and Lenard 1995) (Gao & al., 1996) (Das & al., 1995). Des données récentes suggèrent l'existence de complexes différents pour la transcription et la réplication (Gupta & al., 2003) (Qanungo & al., 2004), et il serait possible que la phosphorylation de la protéine P joue le rôle d'interrupteur entre ces deux activités. La substitution des sites de phosphorylation montre que la phosphorylation n'est pas strictement nécessaire pour la transcription et la réplication de l'ARN, mais stimulerait ces activités en favorisant l'oligomérisation de la protéine (Gao and Lenard 1995) (Spadafora & al., 1996). Aussi, le remplacement des résidus N-terminaux Ser 60 et Thr 62 par des Asp mime l'état phosphorylé de la protéine par sa capacité à induire l'oligomérisation et la transcription (Gao and Lenard 1995). Les études avec la phosphoprotéine du virus Chandipura (CHPV), un autre membre des *Vesiculovirus*, montrent que la forme non phosphorylée est inactive en transcription et existe dans un équilibre monomère-dimère-

tétramère (Basak & *al.*, 2004). La phosphorylation d'un résidu en N-terminal de la protéine P du CHPV induit un important changement de conformation dans la protéine (Chattopadhyay & *al.*, 1997) (Raha & *al.*, 1999), stabilise la forme dimérique (Raha & *al.*, 2000) et active la transcription (Chattopadhyay and Chattopadhyay 1994).

Les études avec le virus de la rage montrent que des formes oligomériques différentes de la protéine P non phosphorylée ont été trouvées en solution, suggérant que la phosphorylation n'est pas nécessaire pour l'oligomérisation (Gigant & *al.*, 2000). Pour la phosphoprotéine du virus de la rage, le domaine d'oligomérisation de la phosphoprotéine se situe entre les résidus 93 et 128 (Gigant & *al.*, 2000) (Mavrakis & *al.*, 2004). L'absence de test *in vitro* pour la transcription du virus de la rage rend difficile la caractérisation fonctionnelle de la phosphoprotéine dans les activités de transcription et de réplication virale. Parmi toutes ces études, la stœchiométrie des différentes formes non phosphorylées et phosphorylées de la protéine P de ces deux virus est peu claire, et des résultats contradictoires ont été obtenus en utilisant différentes techniques.

### ***Interaction de la phosphoprotéine avec des partenaires cellulaires***

La phosphoprotéine du virus de la rage (RV) joue un rôle fondamental aussi bien pendant la transcription et la réplication du virus qu'au cours des processus de transport rétrograde axonal de la particule virale dans le système nerveux central ou dans la lutte contre la réponse immunitaire (Blondel & *al.*, 2002). En effet, des expériences de colocalisation en microscopie à fluorescence ont montré que les résidus 140 à 150 de la phosphoprotéine du RV pouvaient interagir de manière forte avec la chaîne légère de la dynéine (LC8). La dynéine est un moteur moléculaire impliqué dans les mouvements intracellulaires (Jacob & *al.*, 2000) (Poisson & *al.*, 2001), suggérant que la dynéine joue un rôle dans le transport des nucléocapsides virales (Jacob & *al.*, 2000) (Raux & *al.*, 2000). Cependant, comme un mutant du virus de la rage, possédant une phosphoprotéine mutée incapable de lier LC8, se comporte comme un virus sauvage même chez les animaux, le rôle exact de ces interactions n'est pas clair (Mebatsion 2001).

Des expériences d'immuno-précipitation ont permis de détecter la présence de quatre formes de la phosphoprotéine du virus de la rage tronquées en N-terminal. Le gène de la phosphoprotéine code la phosphoprotéine ainsi que quatre sous-produits P2 (aa 20 à 297), P3 (aa 53 à 297), P4 (aa 69 à 297) et P5 (aa 69 à 297) (Chenik & *al.*, 1995). Ces protéines sont issues d'une initiation de la traduction alternative 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, dépendamment de leurs signaux d'export et d'import disponibles. Ces produits du gène de P sont capables d'interagir avec les corps nucléaires PML

(*promyelocytic leukemia protein*). Ces structures sont impliquées dans la résistance de nombreuses cellules à des infections virales, et cette interaction pourrait moduler la réponse cellulaire de l'immunité innée (Blondel & al., 2002). La protéine PML, qui se localise à l'intérieur du noyau cellulaire ou s'associe à des corps nucléaires, est induite par l'interféron ( $\alpha$ ,  $\beta$  et  $\gamma$ ). 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 & 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 & al., 2005) (Vidy & al., 2005). Ces protéines sont très rarement retrouvées dans la particule virale, ce qui suggère leurs implications dans les fonctions de réplication et d'interaction avec la cellule hôte. De façon contradictoire, plusieurs protéines cellulaires ont été détectées comme étant encapsidées dans les particules virales de VSV et RV, comme par exemple des kinases cellulaires qui jouent un rôle important pour l'activation de la phosphoprotéine pour ses fonctions répliquatives (Gao and Lenard 1995) (Gupta & al., 1995).

De la même façon, chez le virus de la stomatite vésiculaire (VSV), l'ARNm de P est traduit en plusieurs protéines : la protéine P, produit de la traduction initiée par l'AUG présent en 5' de l'ARNm P, une forme tronquée de la protéine P, et deux autres petites protéines de 55 et 65 aa, résultat d'une initiation de la traduction alternative à des AUGs en aval de l'extrémité 5' de l'ARNm P. Le premier AUG est dans le même cadre de lecture (Herman 1986), produisant une forme tronquée de la protéine P, et les deux autres AUG sont en décalage par rapport au cadre de lecture initial (Spiropoulou and Nichol 1993). Il est proposé comme pour les formes tronquées de la phosphoprotéine du virus de la rage que ces protéines joueraient aussi un rôle important pour l'interaction avec des partenaires cellulaires pour aider à la multiplication virale. Cependant, leurs fonctions ainsi que leurs partenaires cellulaires ne sont pas connus.

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 & al., 2005). Elle joue donc un rôle essentiel de régulateur pour la multiplication virale.

## E – La transcription virale

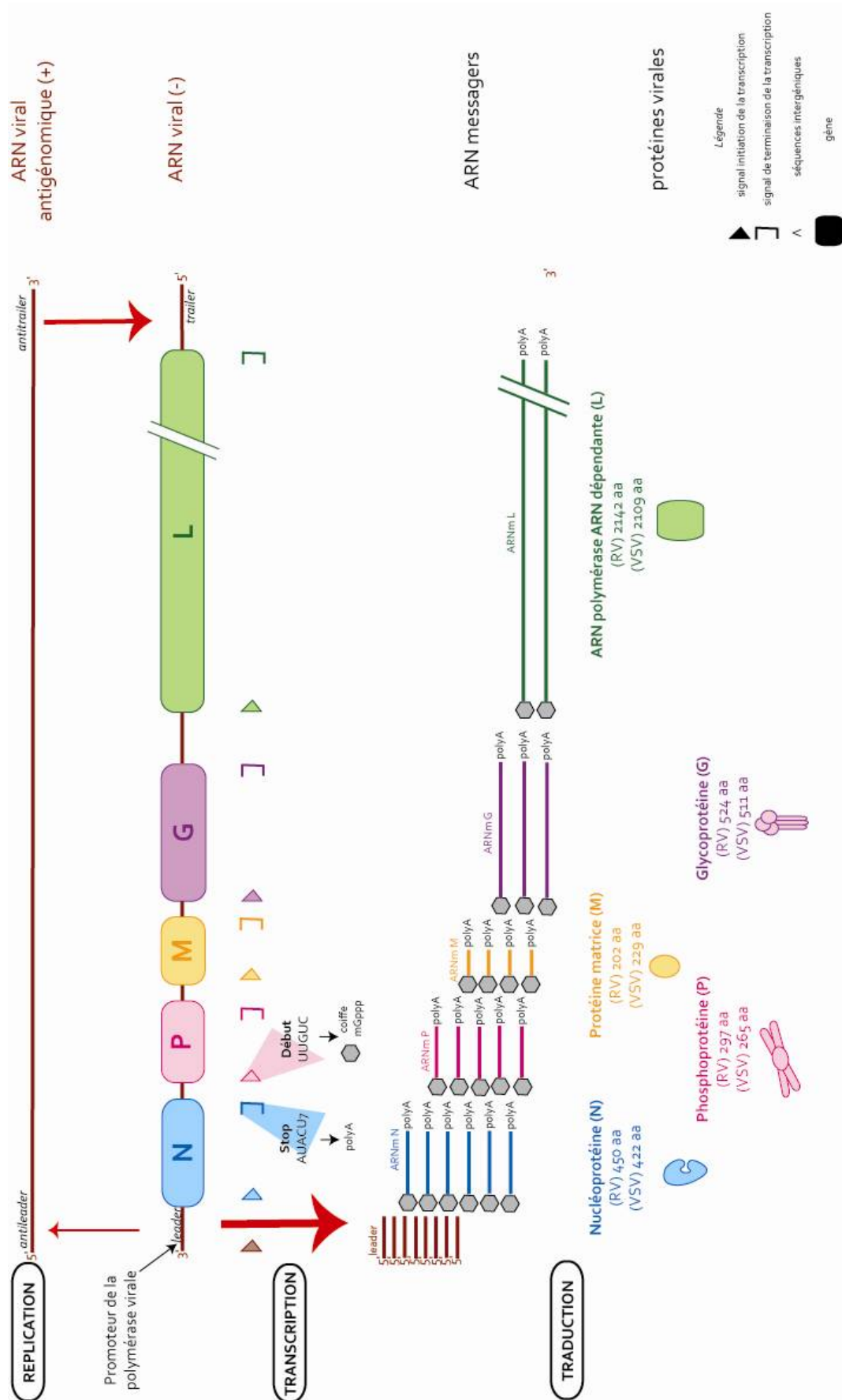
Pour tous les virus à ARN négatif, la première étape après l'entrée du virion dans la cellule hôte est la transcription, c'est-à-dire la synthèse des ARN messagers (ARNms). Cette étape obligatoire précède la réplication virale et s'effectue de manière séquentielle. Sur la base des données obtenues chez le VSV, le

déroulement de la transcription suit vraisemblablement le modèle dit du « *start-stop* » (figure 23, TRANSCRIPTION). La transcription du génome viral est initiée au niveau d'un site unique d'entrée pour la polymérase virale situé à l'extrémité 3' du génome viral (figure 23, promoteur de la polymérase virale) et l'accès au gène suivant dépend du signal de terminaison du précédent (Emerson 1982) (Whelan and Wertz 2002) ont montré que les nucléotides 19-29 et 34-46 situés dans cette région *leader* sont essentiels pour la transcription. Celle-ci débute par la synthèse d'un ARN court de 48 nt pour le VSV et de 58 nt pour RV, appelé *ARN leader* (Colonna and Banerjee 1976), qui n'est ni coiffé ni polyadénylé et n'est donc pas traduit. Après la libération de l'*ARN leader*, la synthèse du premier ARNm, celui codant pour la protéine N, est initié. La polymérase virale coiffe l'ARNm naissant dès l'initiation de la transcription. L'addition de la coiffe sur les ARNm viraux naissants s'effectue par un mécanisme propre aux *Rhabdoviridae*, identifié chez le VSV (Abraham & al., 1975) (Moyer & al., 1975). Les phosphates  $\alpha$  et  $\beta$  de la coiffe proviennent d'un donneur GDP. Ceci est différent de l'addition de la coiffe aux ARNm cellulaires pour lesquels la fixation de la coiffe s'effectue habituellement par le transfert d'un GMP par la guanydyltransférase sur le 5' diphosphate de l'ARNm néo synthétisé (Ogino and Banerjee 2007) (Ogino and Banerjee 2008). A la fin du gène codant pour la protéine N, la transcriptase atteint une séquence conservée : AUACUUUUUUU ( $AUACU_7$ ), où elle commence à glisser de la matrice nucléotidique et produit une queue polyA à l'extrémité 3' de l'ARNm N. La polymérase va s'arrêter uniquement sur cette séquence consensus de terminaison AUAC et créer la polyA grâce à la répétition  $U_7$ . L'ARNm N mature, coiffé et polyadénylé est libéré. La transcriptase passe alors la région intergénique, et initie la transcription du deuxième ARNm, celui de P grâce à une séquence d'initiation conservée : UUGUC. Puis, le signal de polyadénylation  $AUACU_7$  permet la polyadénylation de cet ARNm, la polymérase virale glisse sur la région intergénique et commence la transcription de l'ARNm suivant, celui codant pour la protéine M. Ce processus est poursuivi pour les gènes suivants conduisant à la production de cinq ARNm coiffés et polyadénylés qui sont toujours produits dans un ordre séquentiel strict. La polymérase termine la synthèse à environ 50 nt avant l'extrémité 5' du génome viral.

L'atténuation qui se produit à chaque région intergénique crée ainsi un gradient d'ARNm produits. La transcription des gènes viraux est décroissante en fonction de leur position dans le génome. En conséquence, la position relative des gènes permet de réguler la synthèse des protéines virales. Les gènes proches du promoteur sont transcrits en quantité importante contrairement aux gènes éloignés du promoteur. Le gène de la nucléoprotéine est le plus proche : en effet, la production des ARNm N doit être efficace et doit se faire en quantité suffisante pour encapsider la totalité des génomes viraux. A l'inverse le gène codant pour l'ARN polymérase ARN dépendante (L), nécessaire en quantité catalytique est situé en position distale (Wertz & al., 1998). Des mesures cinétiques de l'activité de transcription pour le VSV (Iverson and Rose 1981) ont montré que la polymérase réalise des pauses de 2,5 à 5,7 min au niveau de ces régions intergéniques. Ces pauses se déroulent lors des étapes de polyadénylation et d'addition de la coiffe sur les ARNm viraux, d'où une transcription discontinue du génome. Dans 30% des

cas, la transcription ne redémarre pas au niveau de ces régions intergéniques. Ce type de régulation est une caractéristique chez les *Mononegavirales*. La réorganisation des gènes pour changer leur ordre permet d'obtenir des virus viables, mais le rendement obtenu pour la multiplication de ces virus en culture cellulaire est beaucoup plus faible.

La traduction des ARNm viraux en protéine s'effectue en utilisant la machinerie cellulaire hôte (figure 23, TRADUCTION). Ils sont traduits par les ribosomes libres présents dans le cytoplasme de la cellule hôte. Les ARNm codant pour les protéines N, M, G et L sont traduits en une seule protéine. Il existe une traduction alternative pour l'ARNm codant pour la protéine P, chez le virus de la stomatite vésiculaire (VSV) et chez le virus de la rage (RV), décrit auparavant.



**Figure 23** - Représentation schématique de la réplication, de la transcription et de la traduction chez les *Rhabdoviridae*. Au cours de la transcription, l'ARN leader et les ARNm viraux codant pour les protéines virales s'accumulent suivant un gradient de concentration décroissant (par rapport à l'extrémité 3' du génome viral). La traduction des ARNm viraux est prise en charge par la machinerie cellulaire de la cellule hôte. La réplication produit un ARN antigénomique (ARN positif) complet. Les flèches rouges indiquent l'activité relative des promoteurs pour la réplication et la transcription.



## F – La réplication virale

La réplication de l'ARN négatif nécessite la production d'une copie complémentaire du génome, appelée antigénome ou ARN complémentaire viral (ARNcv) (figure 23), qui est distinct des ARNm. L'ARN génomique, tout comme l'ARN antigénomique, est produit sous la forme du complexe ribonucléoprotéique. La polymérisation de l'ARN est associée à l'encapsidation. La réplication ne peut avoir lieu que lorsque la synthèse des nouvelles protéines virales, notamment de complexe entre la nucléoprotéine et la phosphoprotéine et de complexes entre la polymérase virale et la phosphoprotéine, nécessaires pour encapsider la chaîne naissante d'ARN viral génomique ou antigénomique (Vidal and Kolakofsky 1989) (Horikami & al., 1992) est suffisante. Au cours de la réplication, la séquence *leader* est encapsidée en premier (Kurilla & al., 1984) (Tordo, Poch, Ermine and Keith 1986), et est impliqué dans l'initiation de la transcription. Il a été montré que chez VSV, l'ARN *leader* s'accumulait dans les noyaux des cellules infectées, inhibant la transcription de la cellule hôte (Kurilla & al., 1982). Il semblerait que ce facteur limitant soit responsable du changement vers le mode processif de la polymérase virale. Aussi, cette encapsidation immédiate des génomes et antigénomes viraux indique que la synthèse de la nucléoprotéine, lors de la réplication, est continue. Toutefois, la manière dont la polymérase virale change du mode transcriptase au mode réplacase reste sujette à des spéculations.

Dans le mode réplacatif, la polymérase virale ignore tous les sites d'initiation, de terminaison et de polyadénylation pour la synthèse des ARNm présents dans le génome viral. Les séquences présentes aux extrémités 3' et 5' de l'ARN viral génomique étant partiellement complémentaires, la polymérase virale peut amorcer la synthèse au niveau de plusieurs promoteurs :

- au niveau de la région *leader*, pour la transcription et la réplication,
- au niveau de l'*anti-trailer* de l'antigénome pour la réplication du brin positif. La séquence *trailer* de 69 nt et 64 nt respectivement chez RV et VSV, constitue à son tour un signal d'encapsidation par la nucléoprotéine (Bourhy & al., 1989) lorsque la polymérase virale copie l'antigénome de façon processive.

Pour accentuer la production de virus, la réplication se fait de manière asymétrique (figure 23, REPLICATION). 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). L'ARN antigénomique peut alors être répliqué par la polymérase virale pour produire plus de copies génomiques. Le nouvel ARN génomique peut être utilisé pour augmenter la réplication de l'ARN viral ou plus tard dans l'infection pour produire de nouveaux virions.

### **III – Objectifs du travail de thèse**

L'objectif de travail décrit dans cette thèse est comprendre la dynamique des mécanismes de réplication et de transcription des virus à ARN négatifs, plus particulièrement pour le virus de la rage (RV) et de la stomatite vésiculaire (VSV), en utilisant une approche biophysique et structurale. L'interaction entre la matrice nucléoprotéine-ARN et la phosphoprotéine est une des clés de la transcription et de la réplication virale. La phosphoprotéine, cofacteur de l'ARN polymérase ARN-dépendante, fait le lien physique entre la matrice nucléoprotéine-ARN et la polymérase virale, ce qui permettrait à la polymérase d'accéder au génome viral. Cependant, la connaissance des structures atomiques des anneaux de nucléoprotéine nous a montré une conformation fermée de la nucléoprotéine sur l'ARN, qui séquestre ce dernier dans la cavité formée par les deux machoires (domaines NTD et CTD) de la nucléoprotéine. Dans le laboratoire, notre hypothèse de recherche était que la phosphoprotéine pourrait être l'interrupteur assurant la transition entre la transcription et la réplication en induisant un changement de conformation de la nucléoprotéine, la fixation de la phosphoprotéine sur les nucléocapsides permettrait à la polymérase d'accéder au génome viral.

Dans un premier temps, nous avons entrepris une caractérisation biophysique et structurale approfondie de la phosphoprotéine des *Rhabdoviridae*, notamment de la phosphoprotéine du virus de la rage (P RV) et de celle du virus de la stomatite vésiculaire (P VSV). Nous avons utilisé plusieurs systèmes d'expressions (procaryote et eucaryote) pour produire ces protéines, ou des fragments de ces protéines, et nous avons utilisé plusieurs techniques biophysiques afin de caractériser leur état d'oligomérisation en solution, leur forme, leur stabilité et leur organisation. La technique de SEC-MALLS-RI (*Size Exclusion Chromatography coupled with Multi-Angle Laser Light Scattering and Refractometry*) ou chromatographie d'exclusion de taille couplée à la diffusion de lumière statique et à la refractométrie, nous a permis de montrer que les phosphoprotéines du virus de la rage (P RV) et de la stomatite vésiculaire (P VSV) sont des dimères en solution. La caractérisation de mutants (mutants P VSV : P VSV S60D/T62D : mutants P RV : P<sub>ΔN52</sub>, P<sub>Δ91-131</sub>, P<sub>ΔC120</sub>) et de certains des domaines de la phosphoprotéine (domaine C-terminal de P VSV, domaine C-terminal de P RV et domaine central de P RV) nous a permis de confirmer que le domaine central de la phosphoprotéine du virus de rage était responsable de l'oligomérisation de la protéine, tout comme pour la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) (Ding & al., 2006). L'analyse bioinformatique des séquences primaires de ces phosphoprotéines, en combinant les prédictions de structures secondaires, de régions désordonnées, d'alignement de séquence a montré que ces phosphoprotéines sont composées de trois domaines structurés et de deux régions désordonnées. Nous avons confirmé par Dichroïsme Circulaire (CD) et fluorescence intrinsèque des Trp que les phosphoprotéines étaient peu stables en solution. L'analyse en domaines de la phosphoprotéine a mis en évidence une organisation modulaire de la phosphoprotéine des *Rhabdoviridae*. Enfin, une étude par

diffusions aux petits angles et d'ultracentrifugation analytique a permis d'obtenir des informations sur la taille de ces protéines. Les phosphoprotéines du virus de la rage et de la stomatite vésiculaire sont allongées selon un axe, leur forme est assimilable à celle d'un cylindre d'environ 18 nm de longueur et 5 nm de diamètre.

Dans un second temps, nous nous sommes intéressés à l'interaction nucléoprotéine-ARN et phosphoprotéine en utilisant deux approches. Nous avons étudié quelle était l'action de la phosphoprotéine et notamment de son domaine C-terminal sur les anneaux de nucléoprotéine (N) du virus de la rage, pour lesquels nous disposons de l'information structurale atomique, et sur les nucléocapsides virales (NC). Nous avons caractérisé l'interaction anneaux N-ARN et phosphoprotéine par plusieurs techniques. Le SEC-MALLS-RI nous a permis d'identifier cette interaction et notamment de déterminer la stoechiométrie des complexes formés. Ainsi, deux dimères de phosphoprotéines se fixent sur les anneaux (de 9 à 11 nucléoprotéines), grâce à la fixation de deux domaines C-terminaux de la phosphoprotéine. Ensuite, nous avons voulu identifier si l'interaction provoquait un changement de conformation de la nucléoprotéine. Les essais de co-cristallisation du complexe anneaux-phosphoprotéine et anneaux-domaine C-terminal de la phosphoprotéine n'ont conduit qu'à l'obtention de données de diffractions à une résolution moyenne (environ 6 Å). La diffusion des rayons X aux petits angles n'ont pas permis de modéliser clairement le complexe anneaux-phosphoprotéine. Enfin, la classification des images de cryomicroscopie électronique du complexe N11-phosphoprotéine en fusion avec la MBP (*Maltose Binding Protein*) montre la présence de quatre densités électroniques supplémentaires à l'extérieur de l'anneau, et nous donne bon espoir d'obtenir une reconstruction tridimensionnelle de ce complexe. Nous avons aussi entrepris une caractérisation structurale de l'effet de la phosphoprotéine sur les nucléocapsides virales, travail réalisé en collaboration avec le Dr. Danielle BLONDEL (Laboratoire de Virologie Moléculaire et Structurale, Gif-sur-Yvette), qui nous a fourni des nucléocapsides virales du virus de la rage et de la stomatite vésiculaire (NC RV et NC VSV). L'analyse par microscopie électronique des nucléocapsides virales nous a cependant montré un changement de conformation important. En effet, l'ajout de phosphoprotéine ou de son domaine C-terminal sur les nucléocapsides du virus de la rage (NC RV) et de la stomatite vésiculaire (NC VSV) entraîne une compaction de ces nucléocapsides.

Nous avons aussi utilisé l'anisotropie de fluorescence ainsi que la technologie résonance plasmonique de surface (SPR) en collaboration avec le Dr. Nicole THIELENS (Institut de Biologie Structurale, Grenoble) afin de déterminer quelle est l'affinité de la phosphoprotéine ou de son domaine C-terminal pour le complexe nucléoprotéine-ARN. Le marquage chimique de la phosphoprotéine par une sonde fluorescente, et le développement d'un test basé sur l'anisotropie de fluorescence permettrait de suivre la formation de complexes dans le but de réaliser un criblage à haut débit pour la recherche d'inhibiteurs spécifiques.

**CHAPITRE II : MATERIELS ET METHODES DE CARACTERISATION DES  
PROTEINES DE *RHABDOVIRIDAE***

## ***I – Clonage, expression et purification des protéines de Rhabdoviridae étudiées***

### **A - Récapitulatif des protéines de *Rhabdoviridae* étudiées**

Nous avons utilisé deux systèmes pour la production de protéines recombinantes de *Rhabdoviridae*, un système eucaryote (baculovirus/cellules d'insectes) et un système procaryote (*E. coli*). De nombreux clones ont été réalisés, qui ont permis d'exprimer différents domaines ou mutants des phosphoprotéines du virus de la stomatite vésiculaire (VSV) et du virus de la rage (RV), ainsi que de la nucléoprotéine du RV, avec ou sans étiquette en N-terminale (Nter) ou C-terminale (Cter) de la protéine. Le [tableau 2](#) reprend les différentes protéines qui ont été produites en indiquant le système d'expression.

Le système d'expression eucaryote baculovirus-cellules d'insectes (*baculo*) a été utilisé pour produire la nucléoprotéine et la phosphoprotéine du virus de la rage. Les baculovirus recombinants *AcNPVN* (*Autographa californica nuclear polyhedrosis virus*) codant pour la nucléoprotéine (souche CVS) et *AcNPVM1* codant pour la phosphoprotéine (souche CVS) ont été fournis par le Dr. Christophe PREHAUD (Prehaud & al., 1990) (Prehaud & al., 1992). Dans ces virus, le gène codant la nucléoprotéine est sous le contrôle du promoteur de la polyhédrine, et celui codant la phosphoprotéine sous le contrôle du promoteur p10. Les cellules d'insectes sont cultivées en monocouche à 27 °C dans des flasques en plastique de 75 cm<sup>2</sup> à 300 cm<sup>2</sup> (BD Falcon), dans du milieu TC100 (Invitrogen) supplémenté avec 10 % de sérum de veau foetal et des antibiotiques (pénicilline/streptomycine). La réamplification et le titrage des baculovirus recombinants sont réalisés sur des cellules *Sf21* (souche *Spodoptera frugiperda*) alors que les cellules HF (souche High Five™ Invitrogen) sont utilisées quant à elles pour l'expression des protéines recombinantes.

Les plasmides de type pET22b contenant les gènes codant pour la phosphoprotéine de VSV (souche Orsay), la phosphoprotéine de RV (souche Pasteur) ainsi que des mutants P<sub>ΔN52</sub>, P<sub>Δ91-131</sub> et P<sub>ΔC120</sub> de la phosphoprotéine de RV (souche Pasteur) ont été fournis par le Dr. Danielle BLONDEL. Toutes les autres constructions ont été réalisées au laboratoire, le plus souvent par amplification spécifique du gène ou de la partie de gène d'intérêt par PCR.

Les protocoles d'expression et de purification des protéines de *Rhabdoviridae* étudiées sont précisés dans les publications indiquées dans le [tableau 2](#). Je ne décrirais ici que le clonage, l'expression et la purification de la phosphoprotéine et du domaine C-terminal de la phosphoprotéine du virus de la rage en fusion avec la Maltose Binding Protein (MBP), et l'expression et la purification des nucléocapsides viraux (Dr. Danielle BLONDEL).

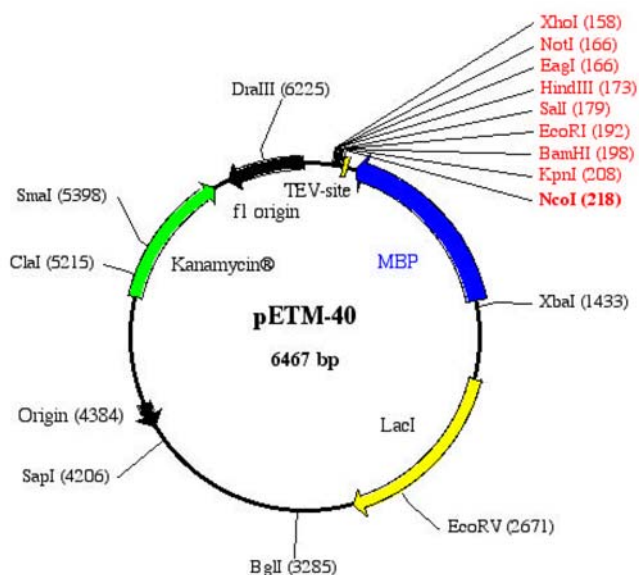
Protéines	Virus Souche	Système Vecteur	Etiquette	Naa MM (Da)	Purification
Nucléoprotéine	RV CVS	Baculo (Dr. C. Préhault)	/	450 50 578	(Iseini & <i>al.</i> , 1998) (Albertini & <i>al.</i> , 2007)
Phosphoprotéine	RV CVS	Baculo (Dr. C. Préhault)	/	297 33 616	(Mavrakis & <i>al.</i> , 2004)
P <sub>CDT3</sub> : Domaine C-terminal P (185-265)	VSV Orsay	Bactérien pET22B	His <sub>6</sub> en Cter	90 10 258	(Ribeiro & <i>al.</i> , 2008)
P <sub>CDT2</sub> Domaine C-terminal P (195-265)	VSV Orsay	Bactérien pET22B	His <sub>6</sub> en Cter	79 9 075	(Ribeiro & <i>al.</i> , 2008)
P <sub>CDT1</sub> Domaine C- terminal P (215-265)	Vsv Orsay	Bactérien pET22b	His <sub>6</sub> en Cter	60 6 983	(Ribeiro & <i>al.</i> , 2008)
P <sub>ΔN52</sub>	RV Pasteur	Bactérien pET22B (Dr. D. Blondel)	His <sub>6</sub> en Cter	253 28 950	(Gerard & <i>al.</i> , 2008)
P <sub>Δ91-131</sub>	RV Pasteur	Bactérien pET22B (Dr. D. Blondel)	His <sub>6</sub> en Cter	263 29 305	(Gerard & <i>al.</i> , 2008)
P <sub>ΔC120</sub>	RV Pasteur	Bactérien pET22B (Dr. D. Blondel)	His <sub>6</sub> en Cter	186 21 244	(Gerard & <i>al.</i> , 2008)
Domaine C-terminal (186-297)	RV CVS/Pasteur	Bactérien pET22B	His <sub>6</sub> en Cter	122 13 989	(Gerard & <i>al.</i> , 2008)
Domaine Central (91-134)	RV Pasteur	Bactérien pET22B	His <sub>6</sub> en Cter	54 6 501	(Gerard & <i>al.</i> , 2008)
Phosphoprotéine S6oD T62D	VSV Orsay	Bactérien pET22b	His <sub>6</sub> en Cter	273 31 018	(Gerard & <i>al.</i> , 2007)
P-MBP	RV Pasteur	Bactérien pETM40	MBP en Nter His <sub>6</sub> en Cter	698 77 361	1 et 2 – chromatographie d'affinité (MBP et Ni <sub>2</sub> <sup>+</sup> ) 3 – chromatographie d'exclusion de taille (S200)
CterP-MBP	RV Pasteur	Bactérien pETM40	MBP en Nter His <sub>6</sub> en Cter	515 57 136	1 et 2 – chromatographie d'affinité (MBP et Ni <sub>2</sub> <sup>+</sup> ) 3 – chromatographie d'exclusion de taille (S200)

**Tableau 2** - Récapitulatif des protéines, mutants, et domaines de protéines de *Rhabdoviridae* produites au laboratoire. (Naa : nombre d'acides aminés, MM : Masse Moléculaire, His<sub>6</sub> : étiquette 6 Histidines).

## B – Clonage, expression et purification de la phosphoprotéine du virus de la rage (P-MBP) et de son domaine C-terminal (CterP-MBP) en fusion avec la MBP (*Maltose Binding Protein*)

### 1 – Clonage et expression des phosphoprotéines fusions P-MBP et CterP-MBP

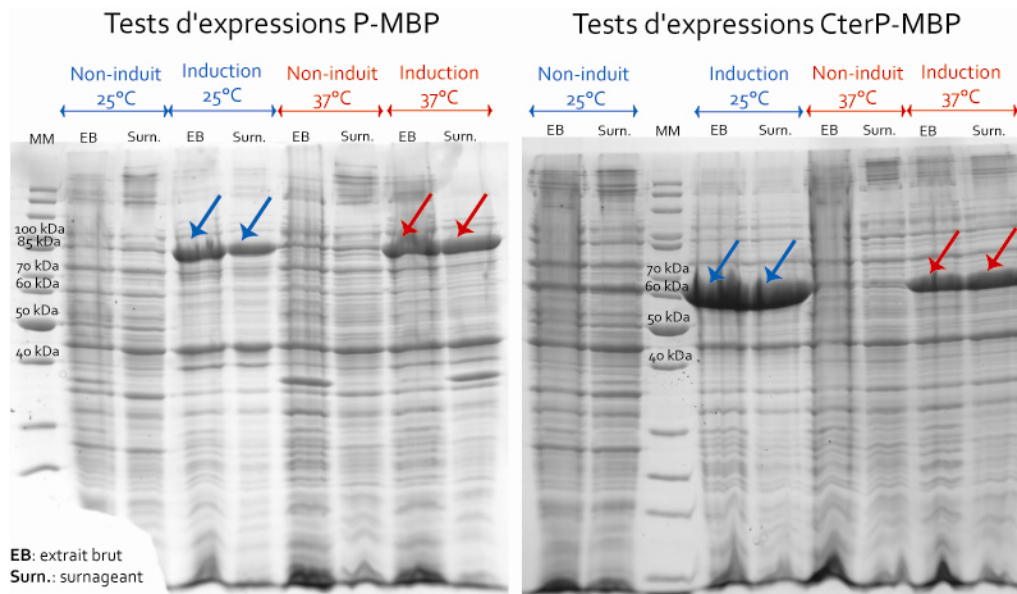
Nous avons réalisé deux nouvelles constructions dans le plasmide pETM40 (figure 24) où les gènes de la phosphoprotéine et du domaine C-terminal de la phosphoprotéine du RV ont été cloné entre les sites de restrictions *NcoI* et *XhoI*. Les constructions ont été réalisées par PCR en utilisant comme matrice les plasmides pET22b – PRV (Pasteur) (Dr. Danielle BLONDEL) et pET22b-CterRVP (Pasteur). Les constructions ont été vérifiées par PCR et par séquençage (méthode de Sanger).



**Figure 24** - Carte du plasmide pETM<sub>40</sub> utilisé pour la production de protéine fusion P-MBP et CterP-MBP. Les gènes codant pour la phosphoprotéine P et son domaine C-terminal ont été cloné entre les sites de restrictions *XhoI* et *NcoI*. Les protéines fusions possèdent la MBP en N-terminale et l'étiquette 6 Histidines en C-terminal.

Les cellules d'*E. coli* BL21(DE3) STAR ont été transformées avec ces plasmides pETM40-P-MBP et pETM40-CterP-MBP et sélectionnées sur boîte LB Kanamycine (60 µg/mL). Dans un premier temps, des premiers tests d'expressions en *E. coli* BL21 (DE3) STAR en petits volumes, à deux différentes températures d'induction, 25°C et 37°C ont été réalisés. L'induction a été réalisée avec 0,5 mM IPTG lorsque les cellules atteignent le milieu de la phase exponentielle de croissance ( $DO_{600nm} \sim 0,6/0,7$ ). Les bactéries sont culotées par centrifugation après 4H d'induction. Les cellules sont ensuite lysées (Bugbuster® et benzonase, Novagen) et les extraits bruts (EB) et les fractions solubles (Surn.) des ces tests d'expressions (figure 25) sont déposés sur gel d'électrophorèse en conditions dénaturantes (SDS-PAGE). Les protéines sont fortement exprimées, pour une température d'induction de 25°C comme à 37°C (puits correspondant à

induction). De plus, la protéine est produite de façon soluble, comme on peut le voir dans les puits Surn (figure 25). En effet, on retrouve la quasi totalité du dépôt de l'extrait brut dans les surnageants. Les protéines fusions P-MBP et CterP-MBP possède deux étiquettes pour faciliter la purification : une étiquette MBP en N-terminal qui aide clairement à la solubilisation et une étiquette 6 Histidines en C-terminal.



**Figure 25** - Analyse par SDS-PAGE des tests d'expressions des protéines fusions P-MBP et CterP-MBP. Deux températures *post induction* ont été testées : 25°C (bleu) et 37°C (rouge). Les flèches indiquent la surexpression de la protéine fusion. MM : marqueur de poids moléculaire, EB : extrait brut et Surn. : surnageant (fraction soluble).

Pour l'expression des protéines P-MBP et CterP-MBP, deux litres de cultures ont été induits avec de l'IPTG 0,5 mM à une  $DO_{600nm}$  de 0,6/0,8 et incubées 4H à 25°C après induction. Les cellules sont ensuite centrifugées à 7 500 g pendant 10 min. Elles sont reprises dans un tampon de lyse contenant 20 mM de Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7,5 et un mélange d'anti-protéases (Complete™ Protease Inhibitor Cocktail Tablets, Roche). Le lysat est incubé sur glace en présence de lysozyme, soniqué, et ensuite centrifugé 1h à 20 000 g afin d'éliminer les débris cellulaires et composants non solubles. Les surnageants contenant les phosphoprotéines fusions solubles sont utilisés pour la purification.

## 2 - Purification des phosphoprotéines fusions P-MBP et CterP-MBP

La première étape de purification est une chromatographie d'affinité. L'extrait soluble à été incubée avec une résine amylose liant la MBP, pendant 2H en chambre froide sous agitation. Ce mélange est ensuite chargé sur une colonne gravimétrique. Après lavage de la colonne avec un tampon 20 mM Tris-HCl, 1,5 M NaCl, pH 7,5 (élimination des acides nucléiques et complexes macromoléculaires contaminants), la protéine à été éluée dans un tampon 20 mM Tris-HCl, 150 mM NaCl, 1.5 M Maltose, pH



7,5. Les fractions ont été analysées par SDS-PAGE. Les fractions contenant les phosphoprotéines fusions sont alors regroupées pour subir une seconde chromatographie d'affinité afin d'éliminer une éventuelle contamination avec la protéine MBP seule. Les surnageants sont incubés avec une résine d'agarose chargée de Nickel (résine Ni-NTA, Qiagen) pendant 2H en chambre froide. Ce mélange est ensuite chargé sur une colonne gravimétrique. Un lavage pour éliminer les contaminations éventuelles par des acides nucléiques ainsi que des interactions non spécifiques avec la résine est réalisé avec un tampon 20 mM Tris-HCl, 1,5 M NaCl, 20 mM imidazole, pH 7,5. Puis, l'élution est réalisée avec un tampon contenant 20 mM Tris-HCl, 150 mM NaCl, 400 mM d'imidazole, pH 7,5. Les fractions sont analysées par SDS-PAGE et celles contenant les phosphoprotéines fusions sont rassemblées. Une chromatographie d'exclusion de taille Hiload 16/60 Superdex 200 (GE Healthcare) permet d'obtenir les phosphoprotéines fusions pures. L'élution est réalisé avec 1,5 volume colonne dans un tampon 20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7,5 à un débit de 1 mL/min. Les fractions contenant les phosphoprotéines fusions pures sont identifiées sur SDS-PAGE puis sont concentrées avec une unité de concentration Millipore™ (membrane de cellulose régénérée avec un seuil de coupure de 10 kDa) et sont conservées à 4°C pour une utilisation ultérieure.

## C – Expression et purification des complexes nucléocapsides viraux du virus de la rage (NC RV) et du virus de la stomatite vésiculaire (NC VSV)

Les nucléocapsides viraux sont préparées par le Dr. Danielle BLONDEL à Gif-sur-Yvette, et nous sont envoyés, conservés à 4°C. Les cellules BSR (cellules de mammifères BHK (*Baby Hamster Kidney*) modifiées) sont cultivées à 37°C dans une étuve à CO<sub>2</sub> dans un milieu MEM (Milieu Essentiel Minimum de Eagle) supplémenté avec 20% de FBS (serum de veau fœtal) et 10% de tryptone. 36 boîtes de Pétri de cellules BSR sont infectées à une *moi* de 0,1 avec le virus de la stomatite vésiculaire (sérotypage Indiana souche Mudd-Summers) et le virus de la rage (souche Pasteur).

Les cellules sont récoltées 24h *post infection* pour l'infection avec le virus de la stomatite vésiculaire et 72h *post infection* avec le virus de la rage, et rincées avec du tampon EPT (136 mM NaCl, 2,7 mM KCl, 3 mM Na<sub>2</sub> HPO<sub>4</sub>, 1,5 mM KH<sub>2</sub>PO<sub>4</sub>) puis resuspendues dans 1,5 à 2 mL de tampon hypotonique (10 mM Tris-HCl, 50 mM NaCl, 1mM EDTA, pH 7,4). La lyse des cellules est provoquée par trois cycles de congélation/décongélation en utilisant en alternance un bain à 37 °C et de l'azote liquide (- 180 °C). Le lysat est homogénéiser au Dounce puis est centrifugé durant 15 min à 12 000 g, le surnageant clair contenant les protéines solubles est aussitôt utilisé pour la purification.

Le surnageant est chargé sur un gradient de chlorure de césium (40 % w/w CsCl dans 150 mM NaCl, 20 mM Tris-HCl, pH 7,4 à 20 % CsCl w/w dans 150 mM NaCl, 20 mM Tris HCl, pH 7,5) puis centrifugé

à 35 000 rpm pendant 15 à 16H à 12 °C (rotor SW41, Beckman). 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, 150 mM NaCl, pH 7,4.

## ***II – Méthodes de caractérisation des protéines***

### **A – Electrophorèse**

#### *1 – Gel d'électrophorèse en conditions dénaturantes*

Les gels d'électrophorèse en conditions dénaturantes (SDS-PAGE) de 8, 12 et 15% avec tampon Tris-Tricine (TT) ou Tris-Glycine (TG) ont été utilisés pour analyser les fractions contenant les différentes protéines. La détection des protéines est réalisée par coloration au bleu de Coomassie, et par illumination sous lumière UV dans le cas de la phosphoprotéine et du domaine C-terminal de la phosphoprotéine du virus de la rage marquées avec le fluorophore IAEDANS.

#### *2 – Gel d'électrophorèse en conditions natives*

Les échantillons d'anneaux de nucléoprotéine du virus de la rage, ainsi que les complexes formés avec la phosphoprotéine du virus de la rage et ses mutants ont été analysés par électrophorèse en gel de polyacrylamide en conditions natives 4% (acrylamide/bisacrylamide 19:1 w/w) dans un tampon de migration TBE 0,2x (TBE 1X : 89 mM Tris-HCl, pH 7,5, 89 mM acide borique, 2,5 mM EDTA) comme décrit dans (Albertini & al., 2007). Le gel est pré-migré 30 min à 200 V et les échantillons sont chargés en sucrose. L'élution se déroule pendant 3 à 4H à 200 V en chambre froide. Le retard sur gel provoqué par la formation du complexe anneaux-phosphoprotéine a été visualisé directement sur le gel, coloré au bleu de Coomassie.

### **B – Spectroscopie d'absorbance**

La concentration des solutions de protéines a été déterminée par mesure du spectre d'absorbance. Le coefficient d'extinction molaire à 280 nm des protéines est calculé à partir de leur

contenu en résidus Trp, Tyr et Cys en utilisant le coefficient d'extinction molaire de ces résidus à cette longueur d'onde (Pace & al., 1995) :

$$\epsilon_{280\text{nm}} (\text{M}^{-1} \text{cm}^{-1}) = N_{\text{Trp}} * \epsilon_{\text{Trp}, 280\text{nm}} + N_{\text{Tyr}} * \epsilon_{\text{Tyr}, 280\text{nm}} + N_{\text{Cys}} * \epsilon_{\text{Cys}, 280\text{nm}}$$

avec  $\epsilon_{\text{Trp}, 280\text{nm}} = 5\,500 \text{ M}^{-1} \text{cm}^{-1}$  ;  $\epsilon_{\text{Tyr}, 280\text{nm}} = 1\,490 \text{ M}^{-1} \text{cm}^{-1}$  ;  $\epsilon_{\text{Cys}, 280\text{nm}} = 125 \text{ M}^{-1} \text{cm}^{-1}$ , et  $N_{\text{Trp}}$ ,  $N_{\text{Tyr}}$ ,  $N_{\text{Cys}}$  respectivement le nombre d'acides aminés Trp, Tyr et Cys de la protéine.

Protéines	P RV	P VSV	P <sub>ΔN52</sub>	P <sub>Δ91-131</sub>	P <sub>ΔC120</sub>	P <sub>CTD</sub>	P <sub>CED</sub>	P-MBP	CterP-MBP	N RV
$\epsilon_{280\text{nm}} (\text{M}^{-1} \text{cm}^{-1})$	25 565	38 390	25 565	17 085	9 970	15 595	8 480	93 280	83 310	46 300

**Tableau 3** - Récapitulatif des coefficients d'extinctions molaires des protéines de *Rhabdoviridae* étudiées.

L'absorption de la sonde fluorescente IAEDANS ( $\epsilon_{\text{IAEDANS}, 336\text{nm}} = 5700 \text{ M}^{-1} \text{cm}^{-1}$ ) a été utilisé pour estimer les concentrations en phosphoprotéine et domaine C-terminal de la phosphoprotéine du virus de la rage marquée par cette sonde fluorescente.

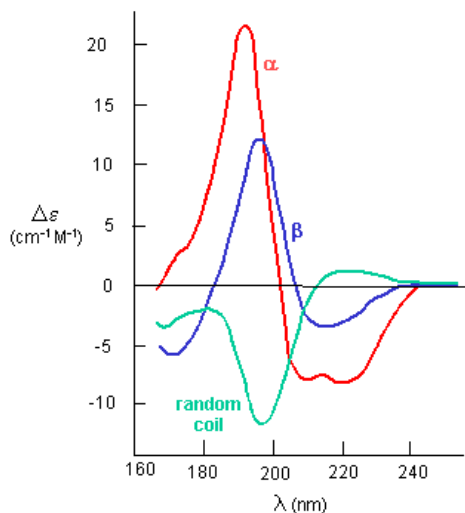
Pour les complexes nucléoprotéine-ARN, la concentration de l'échantillon a également été déterminée par la méthode de Bradford (Bio-Rad Protein Assay) par mesure de l'absorption à 595 nm.

## C – Dichroïsme Circulaire (CD)

### 1 – Principe

Le Dichroïsme Circulaire (CD) repose sur l'interaction entre une lumière circulairement polarisée et des molécules asymétriques. La forme du spectre de CD dans l'UV lointain permet de déterminer le contenu en structures secondaires de la protéine étudiée. On mesure l'absorbance d'un échantillon chiral avec de la lumière polarisée successivement à droite et à gauche et on observe une différence  $\Delta A = A_g - A_d$ . Avec un trajet optique « l » et une concentration « c », le pouvoir dichroïque molaire est considéré comme étant  $\Delta\epsilon = \epsilon_g - \epsilon_d = \Delta A / \Delta l$  ( $\Delta A < 1\%$  de l'absorbance totale). Dans le cas des protéines, les éléments de structure secondaire possèdent un pouvoir dichroïque particulier : une hélice  $\alpha$  (3,6 résidus/tour) possède un pouvoir dichroïque négatif à 222 nm et 208 nm et un positif vers 190-195 nm tandis que le feuillet  $\beta$  possède un pouvoir dichroïque négatif plus faible à 211 nm. Les entités non structurées possèdent une absorption caractéristique négative (figure 26).

Les spectres de dichroïsme circulaire (CD) ont été mesurés à l'aide d'un spectropolarimètre JASCO J-810 équipé d'un système de thermostatisation par effet Peltier. L'échantillon était contenu dans une cuve en quartz de 1 mm. La concentration en protéine était de l'ordre de 10  $\mu\text{M}$  dans un tampon 20 mM Tris-HCl, 150 mM NaCl, pH 7,5. Les spectres sont convertis en pouvoir dichroïque molaire ou ellipticité molaire par résidu en utilisant l'équation suivante :  $\Theta = \frac{0,1 \times CD_{\text{signal}}}{N_{\text{aa}} \times C \times l}$  où  $N_{\text{aa}}$  est le nombre d'acides amines, C la concentration molaire de l'échantillon, l la longueur de la cuve en cm et  $CD_{\text{signal}}$ , le signal de CD mesuré en millidegrés.



**Figure 26** - Schéma théorique du dichroïsme circulaire (CD). L'absorption caractéristique des feuillets  $\beta$  est représentée en bleue avec un minimum entre 210 et 212 nm ( $\beta$ ), les hélices  $\alpha$  possèdent une absorption caractéristique représentée en rouge avec deux minima à 208 et 222 nm ( $\alpha$ ) et les entités non structurées (random coil) possèdent une absorption caractéristique représentée en vert.

## 2 – Analyse des courbes de dénaturation

### Dénaturation induite par l'urée

Les dénaturations induites par l'urée ont été utilisées afin de caractériser les états natifs et dépliés et d'observer le chemin de dépliement de la protéine d'intérêt. La concentration en urée des solutions stocks a été déterminée par mesure de l'indice de réfraction (Pace, 1989). Tous les ajustements ont été réalisés en utilisant le logiciel Sigma Plot SPW10 (SYSTAT Software Inc.). Les transitions de dépliement induites par l'urée et mesurées par CD à 222 nm ou par fluorescence des Trp ont été analysées sur la base d'un modèle réversible à deux états (Tanford, 1970) (Pace, 1989) qui peut être schématisé par :



où U et N représentent les formes dépliées et natives, respectivement. La constante d'équilibre est donnée par :

$$K = \frac{[U]_{\text{eq}}}{[N]_{\text{eq}}} \quad (\text{Eq. 2})$$

Dans les conditions d'équilibres, la stabilité de N relative à U est définie par la différence d'énergie libre entre N et U et est donnée par :

$$\Delta G^\circ = -RT \ln K \quad (\text{Eq. 3})$$

Pour de nombreuses protéines,  $\Delta G^\circ$  varie de façon linéaire avec la molarité en urée dans la région de transition, où K et donc  $\Delta G^\circ$  peuvent être déterminés de façon précise. Si l'on suppose une dépendance linéaire à travers toute la gamme d'urée, la constante d'équilibre et la différence d'énergie libre sont exprimées en fonction de la molarité en urée ([Urée]) en utilisant les relations standards (Tanford, 1970) :

$$\Delta G^\circ = \Delta G^\circ(\text{H}_2\text{O}) - m[\text{Urea}] \quad (\text{Eq. 4a})$$

$$K = e^{\left(\frac{-\Delta G^\circ(\text{H}_2\text{O})}{RT} + \frac{m[\text{Urea}]}{RT}\right)} \quad (\text{Eq. 4b})$$

où  $\Delta G^\circ(\text{H}_2\text{O})$  est la stabilité de N en absence d'urée, et  $m = \frac{\partial \Delta G^\circ}{\partial [\text{Urea}]}$ , correspondant à la pente du graphe de  $\Delta G^\circ$  en fonction de [Urée].

Les dénaturations à l'équilibre induites par l'urée, suivies par CD et fluorescence des Trp ont été ajustées en utilisant les équations suivantes pour les données de CD et fluorescence (Santoro and Bolen 1988) :

$$[\theta]_{\text{obs}} = \frac{[\theta]_{\text{N}} + [\theta]_{\text{U}} e^{\left(\frac{-\Delta G^\circ(\text{H}_2\text{O})}{RT} + \frac{m[\text{Urea}]}{RT}\right)}}{1 + e^{\left(\frac{-\Delta G^\circ(\text{H}_2\text{O})}{RT} + \frac{m[\text{Urea}]}{RT}\right)}} \quad (\text{Eq. 5})$$

$$F_{\text{obs}} = \frac{F_{\text{N}} + F_{\text{U}} e^{\left(\frac{-\Delta G^\circ(\text{H}_2\text{O})}{RT} + \frac{m[\text{Urea}]}{RT}\right)}}{1 + e^{\left(\frac{-\Delta G^\circ(\text{H}_2\text{O})}{RT} + \frac{m[\text{Urea}]}{RT}\right)}} \quad (\text{Eq. 6})$$

où  $[\theta]_{\text{obs}}$  et  $F_{\text{obs}}$  sont les ellipticités molaires et les valeurs de fluorescence des Trp mesurées,  $[\theta]_{\text{N}}$ ,  $[\theta]_{\text{U}}$ ,  $F_{\text{N}}$  et  $F_{\text{U}}$  sont les ellipticités molaires et les valeurs de fluorescence des Trp pour N et U respectivement.

### **Dénaturation induite par la température**

Les dénaturations induites par la température ont été mesurées dans une cuve de 10 mm avec un spectropolarimètre JASCO J-810 équipé d'un système de thermostatisation par effet Peltier. Le CD est mesuré à 222 nm et la température est variée de 278 K et 338 K à des vitesses de 1 à 4 K/min. Les différences d'énergie libre standard ont été calculées en fonction de la température (Pace, 1989) en utilisant:

$$\Delta G^\circ = -RT \ln \left( \frac{[\theta]_N - [\theta]_{obs}}{[\theta]_{obs} - [\theta]_U} \right) \quad (\text{Eq. 7})$$

Le point médian de la transition induite par la température ( $T_m$ ) a été déterminé pour  $\Delta G^\circ = 0$ . En supposant une valeur de  $\Delta C_p$  négligeable, l'enthalpie ( $\Delta H$ ) entre N et U a été calculée en ajustant les courbes de transitions avec les équations suivantes :

$$[\theta]_{obs} = \frac{[\theta]_N + [\theta]_U e^{\left( \frac{-\Delta H \left( 1 - \frac{T}{T_m} \right)}{RT} \right)}}{1 + e^{\left( \frac{-\Delta H \left( 1 - \frac{T}{T_m} \right)}{RT} \right)}}. \quad (\text{Eq. 8})$$

## D – Spectroscopie de fluorescence

### *1– Fluorescence intrinsèque des résidus Trp et Tyr*

La fluorescence intrinsèque des résidus tyrosine (Tyr) et tryptophane (Trp) de la protéine est sensible à la polarité de l'environnement, et permet de suivre les changements de conformation des protéines. Les spectres de fluorescence ont été mesurés à l'aide d'un spectrofluorimètre Hitachi F2500 en utilisant une cuve en quartz de 1 mm. L'excitation était à 280 nm. Pour les dénaturations induites par l'urée, les mêmes échantillons mesurés en CD ont été mesurés en fluorescence.

### *2– Anisotropie de fluorescence*

L'anisotropie est une grandeur sans dimension liée au temps de corrélation de rotation des fluorophores. L'anisotropie de fluorescence fournit un signal indépendant de la concentration qui augmente avec le volume de la particule. Cette technique peut être utilisée pour suivre la formation de complexes engendrant une augmentation de la taille de la molécule marquée (Lundblad & *al.*, 1996).

### ***Marquage de la phosphoprotéine du virus de la rage et de son domaine C-terminal***

La phosphoprotéine et le domaine C-terminal de la phosphoprotéine du virus de la rage (P RV et CterP RV) ont été marquées par la sonde fluorescente 1,5-IAEDANS ( $\lambda_{excitation} = 334 \text{ nm}$ ,  $\lambda_{émission} \sim 480 \text{ nm}$ ). Celle-ci a été choisie pour sa sensibilité à la polarité de l'environnement et à son temps de vie long de l'état excité (de l'ordre de 10-15 ns), favorable aux mesures d'anisotropie. Une solution de protéine purifiée à environ 100  $\mu\text{M}$  est réduite en présence de 10 équivalents molaires de tris(2-

carboxyethyl)phosphine (TCEP) puis incubée 2H à l'obscurité en présence de 5 équivalents d'(((2-iodoacetyl)amino)ethyl)amino)-naphtalène-1 acide sulfonique (1,5-IAEDANS) (Invitrogen). Le mélange réactionnel est ensuite chargé sur une colonne de chromatographie d'exclusion de taille (S75, HR 10/30, GE Healthcare) dans un tampon 20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7,5 afin d'éliminer l'excès de réactifs. L'intégrité de la protéine marquée est analysée par SDS-PAGE, et la fluorescence est vérifiée par illumination du gel aux UV. Enfin, le rendement de la réaction est estimé par spectrométrie d'absorption ( $\epsilon_{1,5\text{-IAEDANS}, 336 \text{ nm}} = 5700 \text{ M}^{-1} \text{ cm}^{-1}$ ). Les temps d'incubation des échantillons étaient de 10 min ( $[P_2]_{\text{IAEDANS}} = 1 \text{ }\mu\text{M}$ ,  $[C\text{ter}P]_{\text{IAEDANS}} = 0,5 \text{ }\mu\text{M}$  et  $0,1 \text{ }\mu\text{M} < [\text{nucléoprotéine}] < 2,5 \text{ }\mu\text{M}$  dans un tampon 20 mM Tris-HCl, 150 mM NaCl, pH 7,5). En effet, si la concentration de protéine marquée est inférieure à  $0,5 \text{ }\mu\text{M}$ , le bruit expérimental rend les mesures imprécises; de même avec des concentrations élevées en nucléoprotéine (ajout de concentration en  $[N] > 8 \text{ }\mu\text{M}$ ), la contamination du signal par de la lumière diffusée hautement polarisée devient importante, ce qui est visible par l'augmentation de l'intensité du pic de diffusion Raman de l'eau ( $\sim 375 \text{ nm}$  pour une excitation à  $334 \text{ nm}$ ).

### **Mesure de l'anisotropie de fluorescence**

Toutes les mesures ont été réalisées à température ambiante dans des cuvettes à fluorescence en quartz (volume de  $200 \text{ }\mu\text{L}$ ) à l'aide d'un spectrofluorimètre de marque Photon Technologies International. Pour tous les points expérimentaux, l'intensité de fluorescence a été mesurée avec les polariseurs d'excitation et d'émission dans les quatre configurations (parallèles et perpendiculaires par rapport au plan de polarisation de la lumière incidente), de manière à calculer le facteur G (constante instrumentale) et l'anisotropie A selon les équations (A) et (B) :

$$G = I_{HV} / I_{HH} \quad (\text{Equation A})$$

$$A = (I_{VV} - GI_{VH}) / (I_{VV} + 2GI_{VH}) \quad (\text{Eq. B})$$

Les données ont été enregistrées à l'aide du logiciel d'acquisition Felix 32 TM (Photon Technology International).

### **Analyse des données d'anisotropie de fluorescence**

Les données ont été ajustées à l'aide d'une expression analytique décrivant l'interaction entre la phosphoprotéine et la nucléoprotéine à l'équilibre et prenant compte la déplétion des deux protéines ( $[P_2] \neq [P_2]_t$ ,  $[N] \neq [N]_t$ ). Classiquement, l'interaction entre un ligand L et son récepteur R est décrite par :



où  $R_f$  représente le récepteur libre,  $L_f$  le ligand libre et  $RL$  le complexe formé par l'association ligand-récepteur. De plus, la constante de dissociation du complexe  $K_D$  est liée aux concentrations de ces différentes espèces par l'équation (D) :

$$K_D = [R_f] + [L_f] / [RL] \quad (\text{Eq. D})$$

Si l'on note  $[R_t]$  et  $[L_t]$  les concentrations totales respectivement en récepteur et ligand, cette équation peut s'écrire sous la forme :

$$K_D = \{([R_f] - [RL]) ([L_f] - [RL])\} / [RL] \quad (\text{Eq. E})$$

La concentration du complexe,  $[RL]$ , est alors la solution de l'équation quadratique (F) :

$$[RL] = \{([L_t] + [R_t] + K_D) - \{([L_t] + [R_t] + K_D)^2 - 4[R_t][L_t]\}^{1/2}\} / 2 \quad (\text{Eq. F})$$

En utilisant la relation

$$[RL] = [L_t]\{A - A_f\} / (A_b - A_f) \quad (\text{Eq. G})$$

Où  $[L_t]$  est la concentration totale en ligand,  $A$  l'anisotropie mesurée,  $A_f$  l'anisotropie du ligand libre et  $A_b$  l'anisotropie du complexe  $RL$ . Il est possible de réécrire l'équation sous la forme :

$$A = A_f + (A_b - A_f) \{([L_t] + [R_t] + K_D) - \{([L_t] + [R_t] + K_D)^2 - 4[R_t][L_t]\}^{1/2}\} / 2 [L_t] \quad (\text{Eq. H})$$

L'équation (H) permet d'obtenir une estimation des paramètres  $K_D$ ,  $A_f$  et  $A_b$  par régression non linéaire à l'aide du logiciel Sigma Plot SPW10 (SYSTAT Software Inc.).

## E – Calorimétrie à balayage Différentiel (DSC)

La capacité de chaleur molaire à été mesurée en fonction de la température par calorimétrie à balayage différentiel (DSC : *Differential Scan Calorimetry*). Le calorimètre de type NanoDSCIII (Setaram) a été conditionné en remplissant les deux cellules avec du tampon (5 mM phosphate de sodium, pH 7), de façon à équilibrer les capillaire en platine avec le tampon. La température a été augmentée puis réduite à une vitesse de 1 K/min, entre 278 K et 348 K, jusqu'à ce que la ligne de base soit stabilisée. La ligne de base a ensuite été enregistrée et utilisée pour déterminer la capacité de chaleur de la protéine. La solution de protéine avait une concentration de 1,53 mg.mL<sup>-1</sup> et a été chargée dans le capillaire (cellule de l'échantillon) à 293 K, et du tampon a été injecté dans la cellule de référence. La température a été diminuée à 278 K, et gardée constante pendant 10 min pour que le système s'équilibre. Le cycle de température utilisé pour le conditionnement a été reproduit plusieurs fois afin de détecter des événements irréversibles. Les capacités de chaleur ont été extraites des mesures brutes de la puissance de compensation en utilisant la formule suivante, qui tient compte de la quantité de solvant dans la solution de protéine :

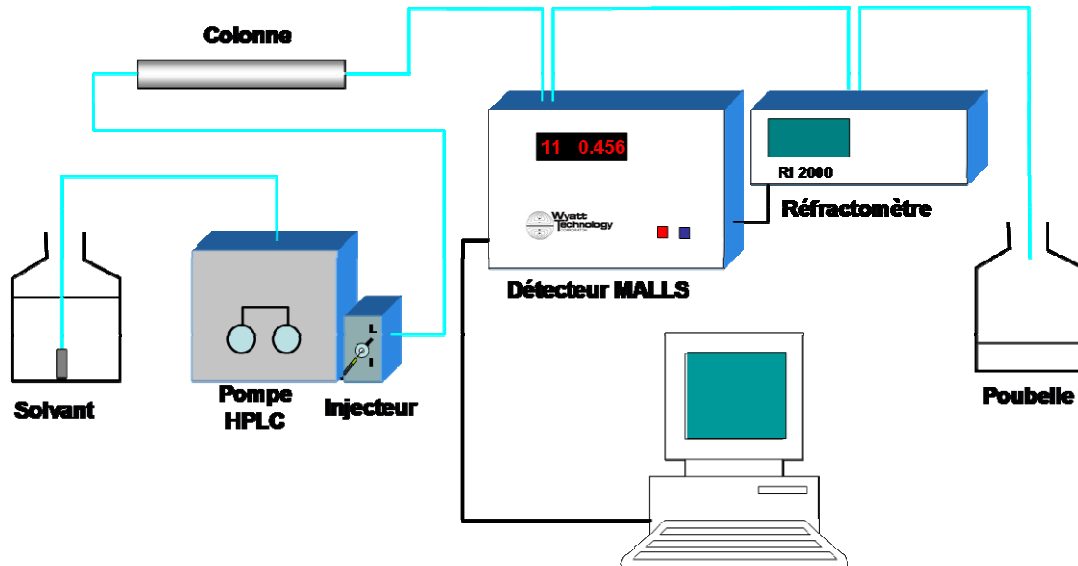


$$C_{prot} = \frac{M(\Delta P / (dT / dt) + \rho_{solvant} V_{prot} C_{solvant})}{V_{cell} [C_0]} \quad (\text{Eq. 9})$$

où  $V_{prot}$  est le volume spécifique de protéine calculé à partir de sa composition en acides aminés ( $0.73 \text{ mL.g}^{-1}$ ),  $\rho_{solvant}$  et  $C_{solvant}$  sont les densités et chaleur spécifique du solvant respectivement, en supposant que les paramètres suivants s'appliquent à l'eau ( $\rho_{solvant} = 1 \text{ g.ml}^{-1}$ ;  $C_{solvant} = 4.19 \text{ J.g}^{-1}.\text{K}^{-1}$ ) et  $[C_0]$  est la concentration en protéine. Le volume de la cellule sensible  $V_{cell}$  ( $299 \mu\text{l}$ ) a été calibré par le fabricant.

## F – Diffusion de lumière multi-angles couplée à la chromatographie d'exclusion de taille et à la réfractométrie (SEC-MALLS-RI)

La technique chromatographie d'exclusion de taille (SEC : *size exclusion chromatography*) couplée à la détection par diffusion de lumière plusieurs angles (MALLS : *Multi-Angle Laser Light Scattering*) et à la réfractométrie (RI : *refractometry*) (SEC-MALLS-RI) fournit une mesure absolue de la masse moléculaire des particules (figure 27). De plus, par calibration de la colonne de chromatographie d'exclusion de taille (SEC : *Size Exclusion Chromatography*) avec des protéines de rayon hydrodynamique connu (Uversky 1993), cette technique fournit également une mesure du rayon hydrodynamique des particules. Enfin, cette technique d'analyse permet d'estimer l'homogénéité et la dispersité des échantillons.



**Figure 27** - Schématisation de l'appareillage de diffusion de lumière statique couplé à la chromatographie d'exclusion de taille et réfractométrie (SEC-MALLS-RI : *Size Exclusion Chromatography coupled with Multi Angle Laser Light Scattering and refractometry*). Un système de chromatographie de type HPLC (chromatographie en phase liquide à haute performance) est associé à un détecteur de diffusion de lumière (18 angles) et à un réfractomètre (mesure de la différence d'indice de réfraction). L'échantillon est injecté sur la colonne de chromatographie d'exclusion de taille, et suit une élution normale.

La chromatographie d'exclusion de taille est réalisée à l'aide d'une chaîne HPLC (chromatographie en phase liquide haute performance). La diffusion de lumière statique est mesurée à

18 angles à l'aide d'un détecteur DAWN-EOS (Wyatt Technology Corp., Santa Barbara, CA) en utilisant un laser qui émet à 690 nm, et l'excès d'indice de réfraction est mesuré à l'aide d'un réfractomètre différentiel RI2000b (Schambeck). Les données ont été analysées avec le logiciel ASTRA V (Wyatt Technology Corp.).

Les mesures ont été réalisées avec différentes colonnes (S75, S200 HR 10/30 (GE Healthcare), Shodex, Protein Pak (Waters)) à 20°C en utilisant typiquement un tampon 20 mM Tris-HCl, 150 mM NaCl, pH 7,5 et un débit allant de 0,5 à 1 mL/min. Des échantillons de 20 à 50 µL ont été injectés à des concentrations allant de 1 à 10 mg/mL.

Sur la base de la théorie électromagnétique de la lumière, il est possible de démontrer que l'intensité de la lumière diffusée par des particules en solution est reliée au produit de la masse moléculaire des particules et de leur concentration. Pour une solution diluée de particules de taille largement inférieur à celle de la longueur d'onde de la lumière (Rayon de gyration  $R_g \ll \frac{1}{10\lambda}$ ) se comportant de manière idéale, l'intensité de lumière diffusée en solution est donnée par l'équation suivante (Schofield & al., 1995) (Wyatt, 1993) :

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

où  $R_\theta$  est rapport de Rayleigh mesuré,  $C$  est la concentration en protéine,  $M$  est la masse molaire ( $\text{g.mol}^{-1}$ ),  $P(\theta)$  est le facteur de forme qui dépend de la structure des particules qui diffusent la lumière et décrivent une dépendance angulaire à la lumière diffusée,  $A_2$  est le second coefficient virial, et  $K^*$  est une constante optique dépendant de caractéristiques de l'instrument et de la solution, dont l'équation est donnée par :

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

où  $N_A$  est le nombre d'Avogadro,  $n_0$  est l'indice de réfraction du solvant à la longueur d'onde incidente (1,333 pour les tampons aqueux dilués),  $dn/dc$  ( $\text{mL.g}^{-1}$ ) est l'incrément d'indice de réfraction du soluté et  $\lambda$  est la longueur d'onde de la lumière incidente. Pour des particules plus petite que la longueur d'onde incidente, aucune dépendance angulaire est observée ( $P(\theta) = 1$ ) et pour une solution suffisamment diluée,  $A_2$  est négligeable ( $A_2 = 0$ ). L'équation peut alors être simplifiée et le rapport de Rayleigh ne dépend plus que de la concentration en protéine et de la masse molaire :

$$R_\theta = K^* \times C \times M$$

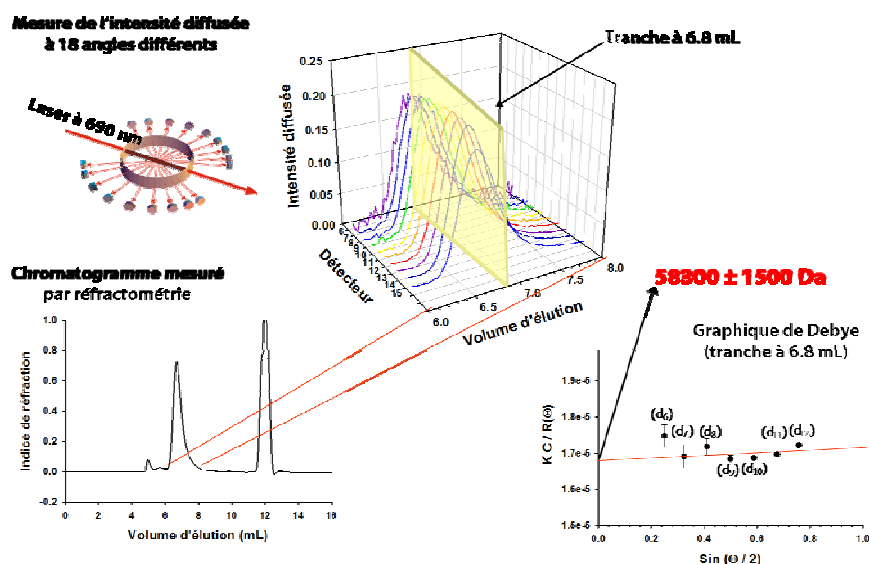
où  $R_\theta$  est le rapport de Rayleigh,  $C$  est la concentration massique ( $\text{g.mL}^{-1}$ ),  $M$  est la masse molaire ( $\text{g.mol}^{-1}$ ) et  $K^*$  est une constante optique dépendant de caractéristiques de l'instrument et de la solution.

La mesure de réfractométrie permet de déterminer la concentration avec précision en s'affranchissant de l'utilisation du coefficient d'extinction molaire à 280 nm. La mesure de l'intensité de lumière à plusieurs angles fournit autant de mesure de la masse molaire qui peuvent être moyennées. Pour des particules de tailles inférieures à celle de la longueur d'onde de la lumière incidente, l'intensité de lumière diffusée ne varie pas avec l'angle de mesure.

Les avantages du couplage des détections par diffusion de lumière plusieurs angles (MALLS) et réfractométrie (RI) avec la chromatographie d'exclusion de taille (SEC) sont :

- de permettre d'éliminer des éventuels agrégats de très grande taille, qui même s'ils sont présents en quantités infimes peuvent contaminer de manière importante la mesure de diffusion de lumière. En effet, à concentration égale, l'intensité de lumière diffusée augmente avec la taille des particules,
- d'obtenir une mesure de la masse moléculaire à chaque point du chromatogramme, et donc, de déterminer la masse molaire de différentes espèces qui sont présentes dans la solution et peuvent être résolues par séparation sur chromatographie d'exclusion de taille.

Au cours de la séparation chromatographique, un chromatogramme est obtenu pour chaque détecteur et en utilisant l'équation  $R_{\theta} = K * \times C \times M$ , il est possible de calculer la masse molaire en chaque point (figure 28). En portant sur un graphique la masse molaire en fonction du volume d'élution il est possible d'estimer l'homogénéité d'une espèce moléculaire s'éluant sous la forme d'un seul pic chromatographique. Une espèce monodisperse s'éluant de manière homogène à travers le pic d'élution et donc la masse moléculaire d'une espèce monodisperse est constante à travers la pic. Un mélange contenant plusieurs espèces de tailles différentes (polydisperse) peut s'éluant de la colonne de chromatographie d'exclusion de taille sous la forme d'un seul pic si la résolution de la colonne ne permet pas de séparer les différentes espèces. Toutefois une séparation en fonction de la taille moléculaire s'opère, les espèces de grande taille éluant en tête de pic et les espèces de plus petites tailles éluant en fin de pic. La masse moléculaire d'une espèce polydisperse varie à travers le pic.



**Figure 28** - Combinaison de données obtenues par SEC-MALLS-RI pour mesurer la masse en solution d'un échantillon. On mesure différents chromatogrammes : un chromatogramme en réfractométrie (nécessaire pour connaître la concentration de l'échantillon) et plusieurs chromatogrammes correspondant à la diffusion de lumière pour différentes positions de détecteurs. Le plus souvent les détecteurs 8 à 13 sont utilisés. Le calcul de la masse moléculaire est solution est alors obtenue. La longueur d'excitation de l'échantillon est de 690 nm.

## G – Résonance plasmonique de surface (SPR)

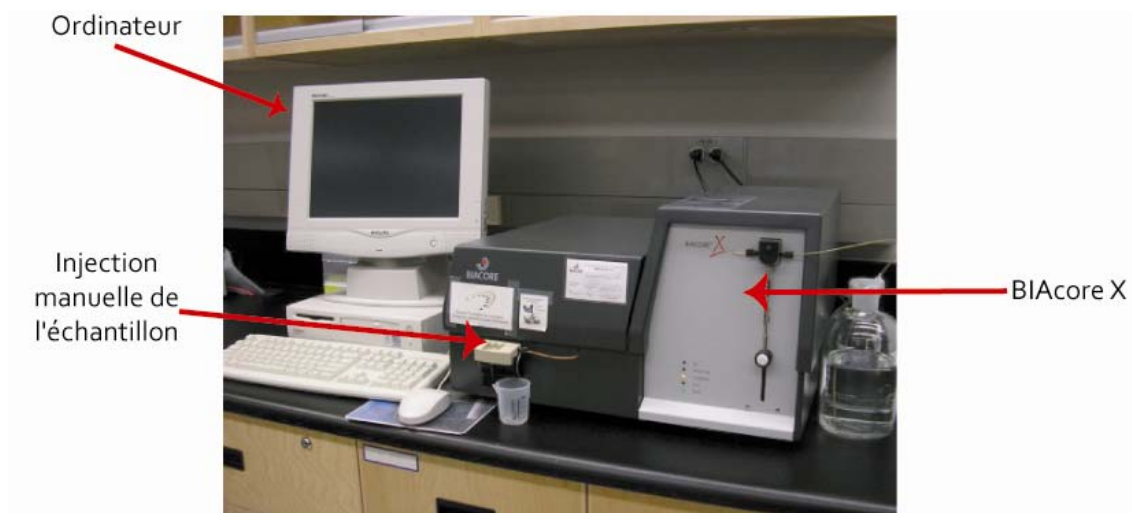
### 1– Principe

La technologie de résonance plasmonique de surface (SPR : *surface plasmon resonance*) permet d'étudier des interactions moléculaires en temps réel sans marquage d'un des deux interactants. L'appareil fonctionne sur le principe d'un biocapteur basé sur l'utilisation de la résonance plasmonique de surface qui détecte des variations de masse à la surface d'une puce détectrice sur laquelle un des deux interactants (le ligand) est immobilisé de façon covalente ou non. L'autre interactant (l'analyte) est injecté par un système microfluidique dans un flux continu de tampon à la surface de la puce. Le suivi en temps réel permet de déterminer les constantes cinétiques d'association ( $k_a$ ) et de dissociation ( $k_d$ ), et d'en déduire la valeur de la constante d'affinité ( $K_D$ ). La concentration en molécules fonctionnelles, la stœchiométrie, la cartographie des sites d'interaction ainsi que certains paramètres thermodynamiques peuvent également être déterminés par cette technologie. Elle permet la caractérisation des interactions moléculaires impliquant les petites molécules et toutes les classes de macromolécules biologiques (protéines, polysaccharides, lipides et acides nucléiques). Elle s'applique à l'étude d'analytes purifiés mais aussi à celle d'analytes présents dans des milieux complexes (lysats cellulaires ou bactériens, extraits tissulaires, sérums et autres liquides biologiques). Elle peut par conséquent être utilisée en association avec la spectrométrie de masse pour identifier dans un milieu biologique les partenaires d'une cible

moléculaire définie immobilisée à la surface de la puce. Par l'injection séquentielle de plusieurs interactants, elle permet de définir la composition et le mécanisme d'assemblage de complexes multiprotéiques ou de machines moléculaires.

## 2– Préparation des puces

Toutes les expériences de résonance plasmonique de surface (BIAcore) ont été réalisées en collaboration avec le Dr. Nicole THIELENS de l'Institut de Biologie Structurale (IBS, Grenoble) sur un appareil BIAcore X (GE Healthcare) (figure 29). Nous avons préparé quatre puces de type CM5 (*sensor chip CM5*, BIAcore, GEHealthcare), qui possède une surface traitée (dextran) sur laquelle la phosphoprotéine du virus de la rage, le mutant P<sub>Δ91-131</sub>, le domaine C-terminal (186-297) de la phosphoprotéine ou les anneaux de nucléoprotéine –ARN (N10) ont été fixés de façon covalente.



**Figure 29** - Montage expérimental lors de l'expérience de résonance plasmonique de surface (SPR) sur le BIAcore X (GE Healthcare).

L'immobilisation de la protéine d'intérêt sur la *sensor chip CM5* s'effectue dans un tampon HBS-EP (10 mM HEPES, 150 mM NaCl, 3,4 mM EDTA et 0,005% P20 (surfactant)), tampon ne contenant pas de groupement amine, selon les protocles fourni par GE Healthcare. A un débit de 5 $\mu$ L/min, sur la piste n°2 de la *sensor chip CM5*, on active la surface de la *sensor chip* en ajoutant 40  $\mu$ L du mélange EDC (1-ethyl-3-(dimethylaminopropyl) carbodiimine) / NHS (N-hydroxy-sulfosuccinimide) (v/v) (BIAcore, GE Healthcare) (figure 30, A). Le NHS est un agent pontant qui, immobilisé sur la surface de la *sensor chip CM5*, va fixé de façon covalente les groupements amines de la protéine à immobiliser. On prépare une solution diluée de protéine à immobiliser de l'ordre de 20  $\mu$ g/mL dans un tampon à un pH inférieur au point isolélectrique de

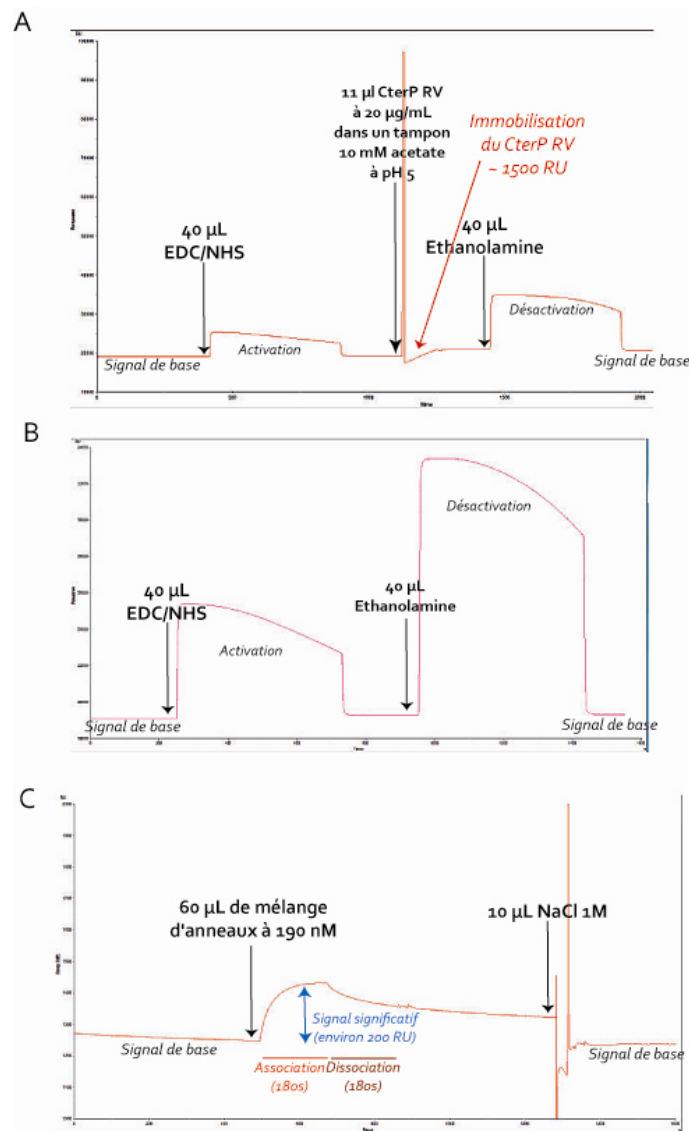
la protéine, de façon à ce que la charge globale de la protéine soit positive et qu'elle puisse fixer l'agent pontant (par exemple, pour le domaine CterP RV, un tampon 10 mM acétate pH 5). On injecte alors un volume de protéine et on vérifie que le changement de signal (RU, unité arbitraire) est suffisant. Pour le domaine CterP RV, présenté sur la [figure 30, A](#), nous avons fixée un équivalent de 1 500 RU. La surface de la *sensor chip CM5* est ensuite traitée avec 40 µL d'éthanolamide afin de désactiver la surface. Sur la piste n° 1 de la *sensor chip CM5*, on active et désactive la surface sans immobilisée de protéine. Ce traitement permettra de vérifier qu'il n'existe pas de signal non spécifique dû à l'interaction du ligand avec la *sensor chip* ([figure 30, B](#)). Le récapitulatif des immobilisations réalisées est présenté dans le [tableau 4](#).

### 3 – Enregistrement des données

Toutes les données ont été enregistrées sur un appareil de type BIAcore X (GE Healthcare), et visualisées en temps réel grâce au logiciel BIAmanual (GE Healthcare). Les injections sont réalisées de façon manuelle à l'aide d'une pipette ([figure 29](#)). Nous avons testé dans un premier temps si l'interaction entre le domaine C-terminal de la phosphoprotéine du virus de la rage (CterP RV) et un mélange d'anneaux de nucléoprotéine-ARN (Nmix avec des N9, des N10, des N11 et des N12) pouvait être suivi par la technologie de résonance plasmonique de surface, nous avons réalisé une injection « test ». A un débit de 20 µL/min, nous avons injecté 60 µL d'une solution d'un mélange d'anneaux à une concentration de 190 nM. Nous avons laissé le processus d'association avoir lieu sur 180s ([figure 30, C](#)) et nous avons pu observer une variation de signal significative (quelques centaines de RU). La dissociation est réalisée sur le même temps, mais sans avoir un retour au signal de base, indiquant que les complexes ne sont pas tous dissociés. Pour régénérer la *sensor chip CM5* avec le CterP RV immobilisé, on injecte alors 10 µL d'une solution de NaCl 1M. Par la suite, nous avons gardé ce volume d'injection et ces paramètres pour étudier l'association et la dissociation du complexe nucléoprotéine-ARN-phosphoprotéine. Nous avons réalisé pour chacune des *sensor chip CM5*, une cinétique de l'interaction, c'est-à-dire que nous avons injecté plusieurs fois le même échantillon mais à des concentrations différentes.

Protéine immobilisées	Concentration en µg/mL	Tampon	Volume injecté en µL	Immobilisation en RU
Phosphoprotéine du virus de la rage (P RV)	20	10 mM acétate, pH 4	22	≈ 6 800
Mutant P <sub>Δ91-131</sub>		10 mM acétate, pH 3,5	19	≈ 4 000
Domaine C-terminal de P RV		10 mM acétate, pH 5	11	≈ 1 500
Anneaux N10 du virus de la rage		10 mM acétate, pH 4	50	≈ 9 400

[Tableau 4](#) - Récapitulatif des immobilisations réalisées pour les expériences de résonance plasmonique de surface.



**Figure 30 - Mise au point de l'expérience de résonance plasmonique de surface (SPR).** Les courbes présentent la valeur RU (unité arbitraire du changement d'indice de réfraction du milieu) en fonction du temps. (A) Immobilisation du domaine C-terminal de la phosphoprotéine du virus de la rage (CterP RV) sur la piste n°2 de la *sensor chip CM5*. (B) Activation et désactivation de la surface de la *sensor chip* sur la piste de n°1, utilisé ensuite comme piste de référence. (C) Mesure par résonance plasmonique de surface (BIAcore, GE Healthcare) de l'interaction entre un mélange d'anneaux (60 µL à 190 nM) et le domaine CterP RV immobilisé sur la *sensor chip CM5*. L'association est mesurée pendant 180s et la dissociation aussi. Le signal étant de l'ordre de la centaine de RU (unité arbitraire), il est significatif.

#### 4- Analyse des données

Toutes les données ont été analysées avec le logiciel BIAevaluation (GE Healthcare). Ce logiciel, permet en plus du traitement brut des données (soustraction du tampon, moyenne, ajustement...), de réaliser des ajustements des courbes de résonance plasmonique de surface dans leur globalité, c'est-à-dire d'obtenir simultanément les paramètres  $k_a$  et  $k_d$  (constante d'association et de dissociation). Plusieurs modèles pour l'ajustement des courbes sont disponibles. Le modèle simple de fixation 1 :1 classique, avec transfert de masse, mais aussi les modèles qui mettent en avant la présence d'un analyte ou d'un ligand hétérogène (avec des réactions de compétition), ou encore le modèle caractérisant une

réaction à deux états, avec un changement de conformation. L'ajustement des courbes pour les valeurs à l'équilibre a été réalisée en utilisant une équation du type :  $y = \frac{RU_{\max} \times x}{k_D + x}$ , à l'aide du logiciel Sigma Plot SPW10 (SYSTAT Software Inc.).

### III – Etudes structurales

#### A – Diffusion aux petits angles (*Small Angle Scattering*)

##### 1 – Introduction à la diffusion aux petits angles

La diffusion aux petits angles (SAS : *Small Angle Scattering*) est une méthode d'analyse des structures et notamment des macromolécules biologiques. Nous avons utilisé deux techniques de SAS : la diffusion des neutrons aux petits angles (SANS : *Small Angle Neutrons Scattering*) et la diffusion des rayons X aux petits angles (SAXS : *Small Angle X-rays Scattering*). Ainsi, ces techniques donnent des informations sur la taille et la forme des macromolécules en solution. La [figure 31](#) montre une expérience type : un faisceau monochromatique de neutrons ou de rayons X illumine un échantillon, et il diffuse élastiquement cette radiation incidente. L'intensité diffusée aux petits angles est mesurée sur un détecteur et on obtient une courbe de diffusion. Cette courbe (intensité diffusée en fonction de l'angle de diffusion) est souvent utilisée pour créer des modèles à basse résolution de protéine.

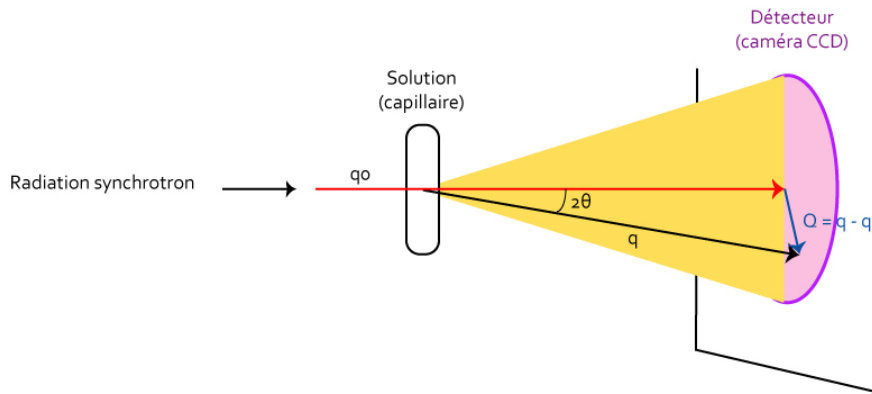
Il est possible de définir les vecteurs suivants caractéristiques d'une expérience de diffusion de neutrons ou de rayons X aux petits angles :

- $\mathbf{q}_0$  le vecteur d'onde incident pour lequel  $|\mathbf{q}_0| = \frac{2\pi}{\lambda}$
- $\mathbf{q}$  le vecteur d'onde caractéristique de l'onde diffusée, dont la direction indique la direction d'observation. Dans le cas d'une diffusion élastique,  $|\mathbf{q}| = |\mathbf{q}_0|$
- $\mathbf{Q}$  le vecteur de diffusion,  $\mathbf{Q} = \mathbf{q} - \mathbf{q}_0$  et son module  $|\mathbf{Q}| = \frac{4\pi \sin \Theta}{\lambda}$  où  $2\theta$  est l'angle de diffusion mesuré par rapport à la direction du faisceau incident et  $\lambda$  la longueur d'onde de la radiation incidente.

L'amplitude de l'onde diffusée par une particule de densité  $\rho(r)$  est donnée par :

$$A(\mathbf{Q}) = \iiint_v \rho(r) e^{i\vec{Q} \cdot \vec{r}} d^3 r .$$





**Figure 31** - Schéma de la diffusion élastique d'un rayonnement lors d'une expérience de diffusion des neutrons ou des rayons X aux petits angles (SANS/SAXS: *Small Angle Neutrons Scattering /Small Angle X-rays Scattering*).

L'approximation de Debye, valable dans les cas des solutions diluées et isotropes, repose sur deux hypothèses. La première postule qu'il n'y a pas de corrélation entre l'orientation des différentes molécules en solution et la seconde, qu'il n'y a pas d'interférence entre deux ondes diffusées par deux molécules différentes. Dans ce cas, si N est le nombre de particules identiques diffusantes, l'intensité observée est égale à :  $I(Q) = N \langle |A(Q)|^2 \rangle$ , en moyennant sur toutes les orientations possibles.

Dans le cas d'une particule en suspension dans un solvant, la diffusion aux petits angles dépend de la différence entre la densité électronique de la particule  $\rho(r)$  et celle du solvant  $\rho(s)$ . Dans ce cas,

l'expression finale est donnée par  $I(Q) = \left\langle \left| \int \int \int_v (\rho(r) - \rho(s)) e^{i\vec{Q}\cdot\vec{r}} d^3r \right|^2 \right\rangle$  (Equation 10).

## 2 – Détermination des paramètres structuraux

Aux très petits angles, dans le domaine de Guinier pour lesquels  $\vec{Q}\cdot\vec{r} \ll 1,3$ , en supposant qu'une seule espèce soit présente en solution, on peut réaliser un développement limité de l'exponentielle 10 pour conduire à la loi de Guinier :  $I(Q) = I(0) \exp\left[-\frac{1}{3} Rg^2 Q^2\right]$  (Equation 11), où Rg est le rayon de giration de la particule, c'est-à-dire la distance quadratique moyenne de tous les points de la particule au centre de gravité pondéré par la masse diffusante (le contraste de densité électronique) en chaque point :

$$Rg^2 = \frac{\int_{V_r} r^2 (\rho(r) - \rho(s)) dr}{\int_{V_r} (\rho(r) - \rho(s)) dr}$$

La loi de Guinier s'écrit le plus souvent sous sa forme logarithmique :

$$\ln I(Q) \cong \ln I(0) \cdot \left( -\frac{1}{3} Rg^2 Q^2 \right).$$

Une simple régression linéaire permet ainsi de déterminer deux paramètres:  $\ln I(0)$  qui est l'ordonnée à l'origine et  $Rg^2/3$  qui est la pente.

### **La masse moléculaire**

L'intensité  $I(0)$  diffusée à  $Q = 0$ , normalisée par la concentration  $c$  de la solution est proportionnelle à la masse moléculaire  $M$  de la particule étudiée :

$$\frac{I(0)}{c} = \frac{N_A M}{\mu^2} (1 - \rho(s)\psi)$$

avec  $N_A$  le nombre d'Avogadro,  $\rho(s)$  la densité électronique du solvant,  $\psi = \frac{1}{\rho(r)}$  est le volume

électronique spécifique moyen et  $\mu = \frac{M}{n}$  avec  $n$ , le nombre d'électron. En théorie, la valeur de  $\mu$  pour

une protéine vaut 1,87 g/mol/électron. Puisque nous ne connaissons pas tous les paramètres de l'équation ci-dessus, l'expérience est calibrée avec une protéine de masse moléculaire connue, le lysozyme. Si on considère que  $\mu$  et  $\psi$  sont à peu près les mêmes pour toutes les protéines,

$\frac{N_A}{\mu^2} (1 - \rho(s)\psi)^2$  peut être considéré comme une constante  $k$  et on peut ainsi déterminer la masse

moléculaire de la protéine étudiée grâce à l'équation  $I(0) = kcM$ .

### **Le rayon de giration (Rg)**

La pente de la droite  $\ln(I(Q))$  en fonction de  $Q^2$  permet de calculer le rayon de giration de l'espèce diffusante. On considère que cette relation n'est valable que dans un domaine restreint de  $Q$ , typiquement  $Q Rg < 1,3$ . Ce rayon de giration fournit une première indication sur la taille de la particule. En effet, pour un volume donné et de masse moléculaire donnée, une molécule avec un rayon de giration le plus petit possible correspond à une sphère de même volume. Plus le  $Rg$  est grand, à volume constant, moins la particule est sphérique.

### **Fonction de distribution des distances $p(r)$**

La fonction  $p(r)$  est un histogramme de distribution des distances entre paires d'atomes dans la molécule étudiée. Elle contient les mêmes informations que  $I(Q)$  et se définit comme :

$$p(r) = \frac{1}{\pi} \int_0^\infty r \frac{Q}{2\pi} I(Q) \sin rQ \cdot dQ.$$

La fonction  $p(r)$  est égale à zéro pour les distances supérieures à l'extension maximale de la particule  $D_{\max}$ .

Le calcul de la fonction  $p(r)$  à partir des données d'intensité de diffusion, s'effectue selon des méthodes de transformation de Fourier indirecte introduite pour la première fois par O. GLATTER (Laggner & al., 1977) et dont plusieurs versions sont proposées dans différents programmes de calculs. Nous avons utilisé le programme GNOM, mis au point par le Dr. Dimitri SVERGUN (Svergun 1991) (Svergun 1992). De la fonction de distribution des distances  $p(r)$ , nous pouvons calculer les valeurs de  $I(0)$  et de  $Rg$  :

$$I(0) = 4\pi \int_0^{D_{\max}} p(r) dr$$

$$Rg^2 = \frac{\int_0^{D_{\max}} r^2 p(r) dr}{2 \int_0^{D_{\max}} r^2 p(r) dr}.$$

La détermination des paramètres  $I(0)$  et  $Rg$  est moins sensible que par la méthode de Guinier en ce qui concerne les interactions avec les molécules et la présence d'oligomères en solution, puisque la méthode tient compte de toute la courbe de diffusion et non pas des points aux très petits angles.

L'allure de la fonction  $p(r)$  permet d'évaluer quelques caractéristiques de la forme globale de la molécule étudiée : les molécules globulaires ont une  $p(r)$  en forme de cloche, dont le maximum correspond à  $D_{\max}/2$  ; les particules allongées ont un maximum de  $p(r)$  pour  $r < D_{\max}/2$  ; la fonction  $p(r)$  des particules aplaties présente un maximum élargi pour  $r < D_{\max}/2$  alors que celle de particules trouées présente un maximum pour  $r > D_{\max}/2$ .

### 3 – Collecte et traitement des données

#### **Données de diffusion de neutrons aux petits angles (SANS)**

Les expériences de diffusion des neutrons aux petits angles ont été réalisées en collaboration avec le Dr. Guiseppe ZACCAI sur la ligne D22 à l'ILL (Institut Laue Langevin, Grenoble) à une longueur d'onde de 0,6 nm. La distance entre l'échantillon et le détecteur était de 4 m, couvrant ainsi une gamme de  $Q$  de  $0,11 \text{ nm}^{-1}$  à  $2,4 \text{ nm}^{-1}$ .

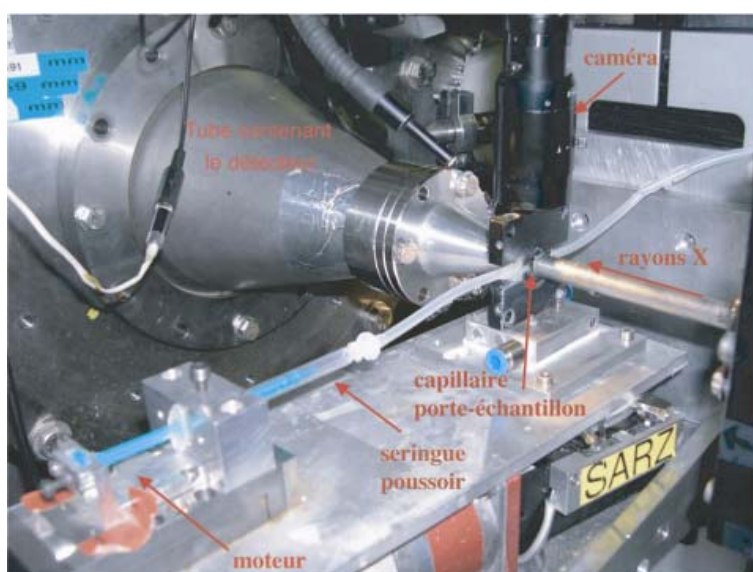
Le traitement des données de diffusions des neutrons aux petits angles a été réalisée avec la suite de logiciel disponible en ligne (<http://www.ill.fr/Computing/prop>; écrit par Ron GHOSH) et a consisté en la création d'un masque afin de retirer l'influence du faisceau incident non diffusé (programme Rmask). Le signal obtenu par convolution avec le masque a ensuite été moyenné radialement autour du centre du faisceau (programme Rnils) et corrigé de la transmission du système (programme Spolly). La diffusion

du solvant et de la cuve vide ont été de même retirées du signal (programme Spolly), et calibrée sur une échelle absolue en utilisant la diffusion de l'eau.

### **Données de diffusion des rayons X aux petits angles (SAXS)**

Les expériences de diffusion des rayons X aux petits angles ont été réalisées sur la ligne de lumière ID02 de l'ESRF (*European Synchrotron Radiation Facility*, Grenoble) à une longueur d'onde de 0,995 Å. Comme le montage expérimental le montre sur la [figure 32](#), l'échantillon est injecté dans un capillaire en quartz d'environ 2 mm de diamètre. Typiquement une série de 10 spectres est enregistrée sur un échantillon d'environ 30 µL de volume en poussant de la solution fraîche face au faisceau entre chaque spectre grâce à un moteur relié à la seringue poussoir. Le détecteur est situé à l'intérieur d'un tube sous vide et peut être déplacé de façon automatique (<http://www.esrf.eu/UsersAndScience/Experiments/SCMatter/ID02>). On réalise des séries de mesures des intensités de diffusion pour une distance échantillon-détecteur de 5 m ( $0,1 \text{ nm}^{-1} < Q < 4,7 \text{ nm}^{-1}$ ) et 1 m ( $0,01 \text{ nm}^{-1} < Q < 0,95 \text{ nm}^{-1}$ ) sur le lysozyme puis sur nos échantillons pour plusieurs concentrations. Les temps d'acquisition varient entre 0,2 s et 1 s par images. Aussi, pour chaque échantillon, on enregistre les images de la diffusion du tampon 20 mM Tris-HCl, 150 mM NaCl, pH 7,5. Pour limiter le bruit de fond produit par le tampon, tous les échantillons ont été dialysés contre ce tampon. Les données brutes obtenues sont présentés  $I(Q)$  (intensité de lumière diffusée) ou  $\log(I(Q))$  en fonction de  $Q$ , le vecteur de diffusion.

Les courbes de diffusions des rayons X sont analysées par le programme BHPlot (*ESRF SAXS Utilities*; Michael SZTUCKI, 2007). Cette première étape permet de moyenniser les courbes de diffusions pour chaque échantillon (protéine ou solvant) puis de soustraire la courbe moyenne de diffusion de protéine avec la courbe moyenne du solvant.



**Figure 32** - Montage expérimental utilisé lors de l'enregistrement des données de diffusion des rayons X aux petits angles sur la ligne ID02 de l'ESRF.

## Détermination expérimentale des paramètres structuraux

Les paramètres globaux ont été extraits des courbes de diffusion en utilisant les programmes développés par le Dr. Dimitri SVERGUN (<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/>; EMBL Hamburg, Allemagne). Les courbes sont alors analysées avec le programme PRIMUS (Konarev 2003). Ainsi, nous avons mis à l'échelle les courbes correspondant aux différentes concentrations utilisées et nous avons fait coïncider les courbes de diffusions enregistrées avec différentes distances échantillon-détecteur (1 m et 5 m pour les expériences de SAXS). Ce programme permet, parmi bien d'autres fonctions, de déterminer les valeurs du rayon de giration ( $R_g$ ) et de  $I(0)$  par le calcul de la courbe de Guinier. Enfin, en utilisant le programme GNOM (Svergun 1991) (Konarev 2003), il est possible d'évaluer la fonction de distributions des distances atomiques  $p(r)$  nous donnant accès au rayon de giration de la molécule ainsi que de la valeur  $D_{\max}$ .

### 4 – Modélisations *ab initio*

#### Principe

Les données de diffusion de rayons X aux petits angles ont été utilisées pour déterminer une enveloppe de la phosphoprotéine du virus de la rage (P RV) et de la stomatite vésiculaire (P VSV) mais aussi des anneaux de nucléoprotéine-ARN, grâce au programme DAMMIN (Svergun 1999). Le point de départ de la méthode consiste à remplir une sphère de rayon  $D_{\max}$  déterminée précédemment avec le programme GNOM, par un ensemble de pseudo-atomes disposés sur un réseau hexagonal compact. Chaque pseudo-atome peut appartenir soit à la protéine (index = 1) soit au solvant (index = 0). Une procédure de recuit simulé est utilisée pour déterminer la combinaison binaire (chercher la position des boules), permettant de correspondre à la courbe de diffusion expérimentale en minimisant une fonction du type  $E(r) = \chi^2 + P(x)$  où  $\chi^2$  mesure l'écart entre les données expérimentales et les données calculées selon l'équation suivante :

$$\chi^2 = \frac{1}{n-1} \sum_{j=1}^n \left[ \frac{\mu \langle c(Q) \rangle I_{\text{mod}}(sQ) - I_{\text{exp}}(sQ)}{\sigma(sQ)} \right]^2$$

avec  $\sigma(sQ)$  représentant les erreurs expérimentales,  $\mu$  un coefficient de mise à l'échelle,  $n$  le nombre de points sur lesquels on effectue la mesure,  $\langle c(Q) \rangle$  un facteur de correction,  $I_{\text{mod}}$  et  $I_{\text{exp}}$  étant les valeurs d'intensité du modèle et expérimentales respectivement.  $P(x)$  correspond à un terme de pénalité permettant d'assurer la compacité et la connectivité de la molécule. Le rayon des sphères est ajusté pour chaque modélisation.

### **Application à la phosphoprotéine du virus de la rage (P RV) et du VSV (P VSV)**

Comme la méthode de détermination de l'enveloppe de la particule en utilisant le programme DAMMIN décrite plus haut utilise le recuit simulé, nous avons réalisé 50 cycles indépendants du programme DAMMIN pour la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) et 15 cycles indépendants pour la phosphoprotéine du virus de la rage (P RV) en appliquant comme seule contrainte la symétrie P2. Pour évaluer la stabilité de la solution et obtenir le modèle le plus probable, nous avons utilisé le programme DAMAVER (Volkova 2003). Ce programme aligne les différents modèles deux à deux et détermine une pseudo-distance (*normalized Spatial Discrepancy* ou NSD) entre les modèles et un modèle choisi comme référence. Les régions communes des ces alignements de modèles sont représentés dans le modèle « *damfilt* » qui nous renseigne sur la forme générale de la protéine.

### **Application aux anneaux du virus de la rage (N9, N10 et N11) et aux complexes anneaux-phosphoprotéine, anneaux-domaine C-terminal de la phosphoprotéine**

Pour les anneaux de nucléoprotéine-ARN, nous avons réalisé 15 cycles indépendants avec DAMMIN en imposant la symétrie associée à l'anneau (P9, P10, P11). De la même façon nous avons évalué la stabilité des solutions avec le programme DAMAVER. Pour les anneaux en complexes avec la phosphoprotéine ou avec le domaine C-terminal de la phosphoprotéine, deux procédures ont été testées : en imposant la symétrie des anneaux lors des 15 cycles indépendants de DAMMIN, ou pas.

### *5 – Modélisation de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) par reconstruction à partir de données structurales partielles*

Nous avons utilisé le programme BUNCH (Petoukhov & *al.*, 2002) (Petoukhov and Svergun 2005) pour modéliser la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) en partant des structures cristallographiques du domaine central (résidus 107-171) (Ding & *al.*, 2006) (code pdb : 2FQM) et du domaine C-terminal de la protéine (résidus 195-265) (Ribeiro & *al.*, 2008) (code pdb : 2K47). Ce programme représente la protéine en utilisant la description atomique à haute résolution des domaines connus et des pseudo-résidus pour les parties inconnues. Les amplitudes partielles de diffusion des domaines connus sont calculés avec le programme CRY SOL (Svergun 1995). Les amplitudes des pseudo-résidus de la partie non connue sont calculées en utilisant le facteur de forme d'un résidu moyen. Le programme BUNCH détermine alors la position et l'orientation relative des domaines ainsi que la forme probable des parties inconnues de telle manière que la courbe de diffusion calculée de l'ensemble soit aussi proche que possible de la courbe expérimentale sous des contraintes de connectivité et de compacité (pour la partie inconnue).

## 6 – Superposition des modèles

La superposition des modèles à basse résolution obtenus avec la structure atomique connue a été réalisée en utilisant le programme SUPCOMB (Kozin 2001). Ce programme permet de trouver la meilleure superposition des modèles *ab initio* avec les structures à haute résolution. BUNCH introduit une mesure NSD entre les objets tridimensionnels à comparer, qu'il va chercher à minimiser afin d'avoir une estimation quantitative de la similarité des objets.

Le [tableau 5](#) résume les différents programmes que nous avons utilisés.

Programmes	Fonction
BHPlot	Moyenne et soustraction des courbes de diffusions des rayons X.
Rmask, Rnils, Spolly	Masque, moyenne du signal et correction de la transmission sur les courbes de diffusions des neutrons.
PRIMUS	Mise à l'échelle des courbes. Détermination du Rg à partir de l'approximation de Guinier.
GNOM	Calcul de la $p(r)$ , du Rg et du $D_{max}$ .
DAMMIN	Modélisation <i>ab initio</i> .
BUNCH	Modélisation d'un ou plusieurs domaines manquants dans une ou plusieurs structures cristallographiques.
CRYSOL	Calcul de la courbe de diffusion théorique correspondant à une structure cristallographique.
DAMAVR	Superposition des modèles deux à deux et détermine les valeurs de NSD.
SUPCOMB	Superposition des modèles obtenus à partir de la courbe de diffusion avec une structure atomique.

[Tableau 5](#) - Résumé des programmes utilisés pour analyser les données de diffusions aux petits angles.

## B –Cristallographie des rayons X

La cristallographie aux rayons X est la méthode la plus utilisée pour déterminer la structure atomique des macromolécules biologiques comme les protéines ou les acides nucléiques. Les rayons X interagissent avec les électrons qui constituent l'enveloppe des atomes. L'interaction entre ces rayons et une molécule génère une diffusion de faible intensité, c'est pour cela que nous utilisons un cristal dans lequel les molécules sont ordonnées dans les trois directions de l'espace, pour amplifier le signal. Les rayons X interagissent alors avec le cristal, ils sont alors diffractés et le détecteur enregistre le cliché de diffraction du cristal.

### 1 – Cristallisation

Pour tester de nombreuses conditions de cristallisation, notamment pour les complexes anneaux-phosphoprotéine, nous avons accès à un robot de cristallisation (*EMBL High Throughput Crystallisation*

Laboratory; <https://htxlab.embl.fr/>). Celui-ci crible automatiquement jusqu'à 6 x 96 conditions commerciales par la méthode de diffusion de vapeur en goutte assise avec une quantité de protéine très limitée (0,1 µL de solution protéique par goutte). Il est souvent nécessaire dans une seconde étape d'affiner les conditions expérimentales en testant la concentration en précipitant, en protéine, le pH, la température, la force ionique, l'effet d'additifs, de substrat, de ligand, de partenaire, le cryoprotecteur.

Afin d'obtenir des cristaux du complexe anneaux-phosphoprotéine et anneaux-domaine C-terminal de la phosphoprotéine, des essais de trempage ont été réalisés. La cristallisation des anneaux à 10 et 11 sous-unités de nucléoprotéine a été reproduite d'après (Albertini, 2007 #400), dans la condition 9,5% PEG 6k, 1,3 M LiCl et 0,1 M MES, pH 6 pour les N10 et 60% tascimate (Hampton Research), 10 mM L-cystéine pour les N11. La méthode utilisée est celle de la diffusion de vapeur en goutte suspendue (McPherson 2001) (goutte : 1 µL de solution protéique + 1 µL de la solution de cristallisation, 600 µL dans le réservoir. Ces cristaux ont ensuite été trempés dans des gouttes contenant la phosphoprotéine ou le domaine C-terminal de la phosphoprotéine.

Les cristaux ont été soumis au rayonnement synchrotron (*European Radiation Synchrotron Facility*, ESRF, Grenoble) sur la ligne de lumière ID23eh1 ( $\lambda \sim 1 \text{ \AA}$ ) dédiée à la cristallographie des protéines. On effectue donc une étape de congélation de ces cristaux pour éviter les dommages radiatifs. La condition de cristallisation des anneaux N11 est cryoprotectrice grâce à la présence de 60% tascimate mais pour les cristaux de N10 (condition : 9,5% PEG 6k, 1,3 M LiCl et 0,1 M MES, pH 6) nous avons utilisé du glycérol (20%) qui empêchent l'eau de cristalliser et d'entraîner une forte dégradation du cristal lors de l'expérience de diffraction.

## 2 – Traitement des données

Les clichés de diffraction des données collectées pour les cristaux de N11 trempés avec la phosphoprotéine ou avec le domaine C-terminal sont traités en trois étapes : indexation, intégration et mise à l'échelle. La répartition des taches de diffraction sur le cliché permet donc d'extraire les informations concernant le réseau cristallin: c'est l'**indexation**. On détermine ainsi le groupe d'espace et les paramètres du cristal. Lors de l'étape d'**intégration**, l'intensité des réflexions est déterminée en appliquant un masque autour de la réflexion. Ces intensités contiennent une partie de l'information sur la densité électronique et la position des atomes. L'ensemble des taches de diffraction est ensuite **mise à l'échelle** afin de constituer un jeu de données complet et homogène. En effet, les réflexions équivalentes par symétrie ou enregistrées plusieurs fois sont moyennées. Cependant, chaque image diffère par la dose de rayons X reçue et par la baisse du pouvoir de diffraction du cristal (dommages dus aux radiations). On définit le rapport signal sur bruit,  $I/\sigma$  qui permet de donner une estimation de l'intensité moyennes des réflexions mesurées en fonction de la résolution, où  $I$  est l'intensité des réflexions mesurées et  $\sigma$  l'écart-



type des intensités de réflexions (bruit de fond). Aussi, la qualité du jeu de données peut être alors estimée par une valeur du facteur  $R_{sym}$  qui permet de comparer l'intensité de chaque réflexion équivalente par la symétrie du cristal, à la valeur moyenne de ces réflexions :

$$R_{sym} = \frac{\sum_{hkl} |I_{hkl} - \langle I \rangle|}{\sum_{hkl} I_{hkl}}$$

Nous avons utilisée la méthode de remplacement moléculaire afin de calculer les phases de nos données. Cette méthode est utilisée lorsque l'on dispose d'une structure cristallographique d'une protéine possédant typiquement plus de 25% d'identité de séquence en acides aminés. Brièvement, cette méthode permet de calculer les fonctions représentant l'ensemble des vecteurs interatomiques d'une molécule, pour le modèle cristallographique et pour la structure recherchée. Le modèle est ajusté en translation et en rotation et la meilleure solution subit une première étape d'affinement en corps rigide.

L'indexation et l'intégration des données à été réalisées avec le programme MOSFLM (Leslie 2006). Le programme SCALA (CCP4) a été utilisé pour effectuer la mise à l'échelle des données. Enfin, le remplacement moléculaire a été réalisé avec le programme MOLREP (CCP4) et l'affinement a été réalisé avec REFMAC (CCP4) (1994) (<http://www.ccp4.ac.uk/index.php> ). Le programme Coot (Emsley and Cowtan 2004) nous a permis de visualiser les cartes de densités obtenues.

## C – Microscopie électronique

La microscopie électronique en transmission (MET) est une technique très couramment utilisée en biologie et permet de réaliser des études structurales à moyenne résolution sur des virus ou des complexes protéiques isolés.

### 1 – Préparations des échantillons

Les électrons constituent un rayonnement très destructeur pour les échantillons biologiques et il règne un vide poussé au sein de la colonne d'un microscope électronique. Si on veut observer un échantillon biologique en MET, celui-ci doit être préparé afin d'en préserver la structure. Divers méthodes de préparation des échantillons biologiques ont été mises au point. Cependant, je ne présenterai ici que les techniques de préparation que nous avons utilisées pour observer les protéines de *Rhabdoviridae*.

## ***La coloration négative***

La coloration négative consiste à observer un échantillon biologique par l'intermédiaire d'un colorant (sel de métaux lourds). Les métaux lourds sont des atomes possédant un noyau très dense qui va dévier fortement les électrons du faisceau. En conséquence, le contraste des images prises en coloration négative est important entre les zones où le colorant est présent et celles où il est exclu par l'échantillon. Elle permet l'observation d'échantillons biologiques d'assez petite taille (jusqu'à 20 kDa) mais ne donne pas accès à la structure interne de la particule observée. Cette méthode est la plus simple à mettre en oeuvre et permet l'observation rapide de l'échantillon avec un fort contraste, pour juger de son homogénéité et de sa pureté. L'échantillon est déposé sur une couche de carbone qui ensuite va venir recouvrir une grille en cuivre. L'échantillon est fixé par séchage de la solution de sels de métaux lourds. L'échantillon va laisser une empreinte dans les sels de métaux lourds en séchant et cette empreinte va être observée. Cependant, l'ajout de colorant provoque quelques artefacts limitant les applications de la coloration négative. Ainsi, lors du séchage de l'échantillon préalable à l'observation sous le microscope, il se produit une inévitable déshydratation déformant les particules biologiques.

### *Observation des protéines de Rhabdoviridae*

Un microscope JEOL 1200 EX II à 100 kV avec un grossissement de 39 750 (calibré sur le tour d'hélice de la capsid du virus de la mosaïque du tabac) a été utilisé pour observer les échantillons en coloration négative. Les négatifs ont ensuite été scannés (scanner Zeiss, Photoscan TD) avec un pas de 7  $\mu\text{m}$  correspondant à 1,75  $\text{\AA}/\text{pixel}$  (ou 3,5  $\text{\AA}/\text{pixel}$ ). Les échantillons dans un tampon 20 mM Tris-HCl, 150 mM NaCl, pH 7,5 ont été dilués à 0,1 mg/mL pour l'observation et absorbés sur une surface de carbone déposé sur une feuille de mica, et coloré avec 2% acétate uranyle. Pour les échantillons nucléoprotéine-ARN (anneaux et nucléocapsides), nous avons testé plusieurs colorants mais nous avons travaillé principalement avec 2% acétate d'uranyle, pH 4,5, car ce colorant permet d'obtenir des vues de côté des anneaux, ce qui est essentiel pour obtenir une reconstruction 3D à une résolution moyenne à partir des images de microscopie.

## ***La cryo-microscopie électronique ou technique des échantillons « congelés hydratés »***

Cette méthode développée dans les années 1980 (Adrian & al., 1984), consiste à déposer l'échantillon seul en solution dans son tampon sur une grille à trou en cuivre recouverte d'une fine couche de carbone trouée puis, à éliminer l'excès de solution avec un papier filtre afin que ne subsiste par capillarité qu'une fine couche de solution dans les trous de carbone. Enfin la grille est congelée dans l'éthane liquide. Le but est d'obtenir au final une très fine couche hydratée congelée de l'échantillon. La congélation quasi instantanée de l'échantillon dans l'éthane liquide provoque la formation de glace vitreuse ne possédant pas de structure ordonnée (contrairement à la glace cristalline). L'échantillon est donc figé dans un état proche de ses conditions physiologiques et est très peu déformé si la température

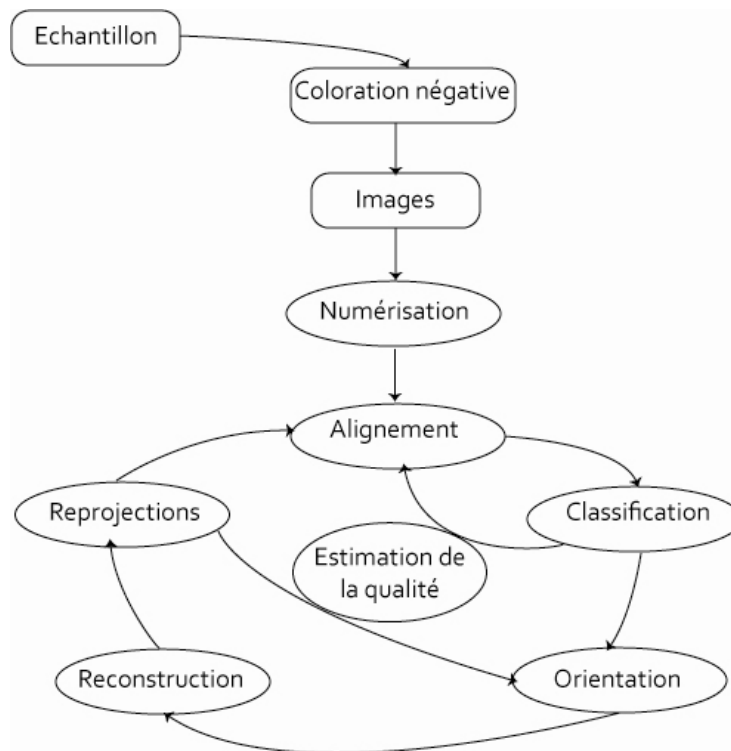
ne passe pas au-dessus de  $-160^{\circ}\text{C}$  (transition glace amorphe/glace cristalline). La cryo-microscopie électronique est la méthode de choix si on envisage de faire une reconstruction tridimensionnelle de la particule observée à haute résolution (8 – 15 Å) puisque à la fois des informations sur la structure interne et externe sont accessibles tout en maintenant la structure native de l'objet observé.

## *2 – Reconstruction tridimensionnelle de molécules isolées à partir d'images prise en microscopie électronique*

En prenant des images d'un échantillon pur, homogène et concentré, on obtient sur les micrographies de nombreuses et distinctes projections bidimensionnelles (2D) d'objets ayant la même structure tridimensionnelle. La structure 3D d'un objet peut être alors obtenue à partir de la combinaison de l'information contenue dans plusieurs projections 2D du spécimen.

### **Reconstructions des anneaux N9, N10 et N11 et du complexe N9-phosphoprotéine**

Après la numérisation des images (scanner Zeiss, Photoscan TD) avec un pas de  $7\mu\text{m}$  correspondant à  $1,75\text{ \AA}/\text{pixel}$ , leur qualité a été évaluée. En effet, pour chaque image le calcul de la fonction de transfert de contraste (CTF) a permis d'éliminer les images pour lesquelles il y avait une dérive et/ou un astigmatisme. Les images avec la meilleure qualité optique ont été conservées. Les particules (anneaux) ont été sélectionnées à l'aide du programme BOXER (EMAN). Pour chaque reconstruction, entre 10 000 et 20 000 particules ont été sélectionnées. Ensuite, ces particules ont été classifiées selon leur orientation. Pour chacune des classes obtenues, une moyenne des classes a été calculée afin d'augmenter le rapport signal sur bruit. Ensuite, l'orientation pour chaque moyenne de classes a été attribuée et une première reconstruction tridimensionnelle a été réalisée par rétroprojection. Ce premier modèle est ensuite projeté selon toutes orientations et ces reprojections servent de nouvelles références pour effectuer un nouveau cycle d'alignement, de classification et d'orientation des images (figure 33). Ce cycle est répété plusieurs fois afin d'améliorer la qualité de la reconstruction. Trois programmes ont principalement été utilisés : IMAGIC ([www.ImageScience.de](http://www.ImageScience.de)), EMAN (<http://blake.bcm.edu/EMAN/>) et SPIDER ([http://www.wadsworth.org/spider\\_doc/spider/docs/spider.html](http://www.wadsworth.org/spider_doc/spider/docs/spider.html)). Les reconstructions que nous avons obtenues ont une résolution entre 16 et 18 Å.



**Figure 33** - Diagramme de la procédure de traitement d'images de particules isolées utilisée pour obtenir une reconstruction tridimensionnelle à partir d'images de microscopie électronique.

**CHAPITRE III : CARACTERISATIONS BIOPHYSIQUES ET STRUCTURALES DES  
PHOSPHOPROTEINES DE *RHABDOVIRIDAE***

# ***I – Les phosphoprotéines des Rhabdoviridae sont des dimères allongés en solution***

## **A – Introduction**

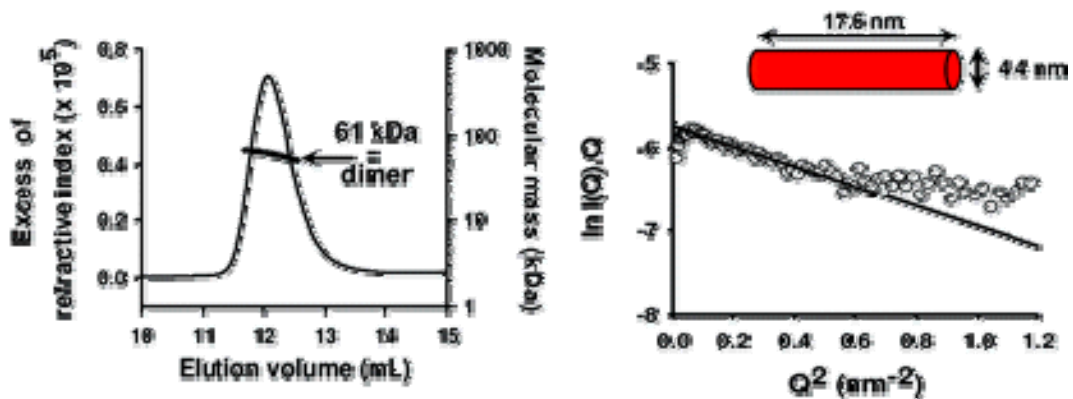
Des résultats dans la littérature suggéraient que l'oligomérisation de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) et du virus de la rage (P RV) serait nécessaire pour la fixation de la nucléoprotéine et de l'ARN polymérase virale ARN-dépendante, mais la stœchiométrie des oligomères formés, leur organisation structurale et leurs rôles exacts dans les processus de transcription et de réplication, n'étaient pas clairement établis. Les sites de phosphorylation nécessaire à la transcription se situent dans la partie N-terminale de la protéine. Toutefois, pour la phosphoprotéine du virus de la rage (P RV), la phosphorylation ne serait pas nécessaire pour l'oligomérisation (Gigant & *al.*, 2000) contrairement à la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) (Gao and Lenard 1995) (Chen & *al.*, 1997).

Dans cet article, nous avons caractérisé l'état d'oligomérisation des phosphoprotéines du VSV et RV, mais aussi celui du mutant de P VSV S60D/T62D. Ce dernier, a été décrit auparavant comme mimant la phosphorylation N-terminale de P VSV, et comme nécessaire pour l'oligomérisation de la protéine et son activité de transcription *in vitro* (Gao and Lenard 1995). La chromatographie d'exclusion de taille couplée à la diffusion de lumière statique et à la réfractométrie (SEC-MALLS –RI: *Size Exclusion Chromatography coupled with Multi-Angle Laser Light Scattering and Refractometry*), la diffusion de neutrons aux petits angles (SANS : *Small Angle Neutrons Scattering*), l'ultracentrifugation analytique (AUC : *Analytical UltraCentrifugation*) et la DLS (*Dynamic Light Scattering*) nous ont permis de caractériser biophysiquement les phosphoprotéines de *Rhabdoviridae*.

B – Article 1: Unphosphorylated *Rhabdoviridae* phosphoproteins forms elongated dimers in solution

Auteurs: **Francine C.A. GERARD**, Euripedes de Almeida Ribeiro Jr., Aurélie A. V. Albertini, Irina Gutsche, Guiseppe Zaccai, Rob W. H. Ruigrok et Marc Jamin.

Publié en 2007 dans le journal *Biochemistry*, volume 46, pages : 10328-10338.



[Figure 34](#) - Images des résultats principaux du papier, mis en ligne sur *Biochemistry* lors de la parution de la publication.

## Unphosphorylated *Rhabdoviridae* Phosphoproteins Form Elongated Dimers in Solution<sup>†</sup>

Francine C. A. Gerard,<sup>‡</sup> Euripedes de Almeida Ribeiro Jr.,<sup>‡</sup> Aurélie A. V. Albertini,<sup>‡,§</sup> Irina Gutsche,<sup>‡</sup> Guiseppe Zaccai,<sup>||</sup> Rob W. H. Ruigrok,<sup>‡</sup> and Marc Jamin<sup>\*‡</sup>

Unit of Virus Host Cell Interactions, UMR 5233 UJF-EMBL-CNRS, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France, and Institut Laue Langevin, 6 Rue Jules Horowitz BP 156, 38042 Grenoble Cedex 9, France

Received April 25, 2007; Revised Manuscript Received June 27, 2007

**ABSTRACT:** The phosphoprotein (P) is an essential component of the replication machinery of rabies virus (RV) and vesicular stomatitis virus (VSV), and the oligomerization of P, potentially controlled by phosphorylation, is required for its function. Up to now the stoichiometry of phosphoprotein oligomers has been controversial. Size exclusion chromatography combined with detection by multiangle laser light scattering shows that the recombinant unphosphorylated phosphoproteins from VSV and from RV exist as dimers in solution. Hydrodynamic analysis indicates that the dimers are highly asymmetric, with a Stokes radius of 4.8–5.3 nm and a frictional ratio larger than 1.7. Small-angle neutron scattering experiments confirm the dimeric state and the asymmetry of the structure and yield a radius of gyration of about 5.3 nm and a cross-sectional radius of gyration of about 1.6–1.8 nm. Similar hydrodynamic properties and molecular dimensions were obtained with a variant of VSV phosphoprotein in which Ser60 and Thr62 are substituted by Asp residues and which has been reported previously to mimic phosphorylation by inducing oligomerization and activating transcription. Here, we show that this mutant also forms a dimer with hydrodynamic properties and molecular dimensions similar to those of the wild type protein. However, incubation at 30 °C for several hours induced self-assembly of both wild type and mutant proteins, leading to the formation of irregular filamentous structures.

Viruses of the *Rhabdoviridae* family form enveloped particles that possess a nonsegmented single-stranded negative sense RNA genome. Together with the *Paramyxoviridae* (measles, parainfluenza, respiratory syncytial, Hendra, and Nipah viruses), *Bornaviridae*, and *Filoviridae* (Ebola and Marburg viruses) families, they constitute the order of *Mononegavirales* (MNV) that includes many human, animal, and plant pathogens (1). MNV viruses exhibit a great diversity in terms of morphology and interactions with their hosts, but they share similar genomic and structural organizations as well as similar modes of RNA replication and transcription

(1). The linear viral RNA genome is tightly associated with the viral nucleoprotein (N)<sup>1</sup> in a helical ribonucleoprotein complex (RNP) (2, 3) that serves as template for both transcription and replication (4). The genomes comprises five common genes, which are similarly organized along the RNA molecule for all MNV viruses, encoding successively the N, the phosphoprotein (P), the matrix protein (M), the glycoprotein (G), and the RNA-dependent RNA polymerase (L). For rabies virus (RV) and vesicular stomatitis virus (VSV), only three of these proteins, N, P, and L, are necessary for synthesizing viral RNA in an efficient and regulated manner (5, 6). Among those three, P has multiple functions during the replication cycle and is required for RNA synthesis and encapsidation of the viral RNA by N. P binds to nascent nucleoproteins, acting as a chaperone by forming RNA-free N<sup>o</sup>-P complexes necessary for encapsidation of new viral genomes (7, 8). P also acts as a cofactor of the viral RNA polymerase, forming a two-subunit RNA polymerase complex in which L carries out all the enzymatic activities of RNA-dependent RNA transcription and replication, including mRNA cap synthesis and mRNA polyadenylation (5, 9, 10). By binding to both L and N, P allows the RNA polymerase complex to bind to RNPs. However, little is known about the mechanisms of action of this protein, and P could also be involved in the remodeling of RNPs during RNA synthesis (2).

Oligomerization of P is a common property among viruses of the MNV order. For some members of the *Paramyxoviridae* family (subfamily of the *Paramyxovirinae*: Sendai virus,

<sup>†</sup> This work was supported by the interdisciplinary program “Maladies Infectieuses Emergentes” from the CNRS. F.C.A.G. and A.A.V.A. were supported by MENRT fellowships from the French government. E.d.A.R. was supported by a postdoctoral fellowship from the Université Joseph Fourier, Grenoble.

\* Corresponding author: Unit of Virus Host Cell Interactions (UVHCI), UMR 5233 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France. E-mail: jamin@embl.fr. Phone: + 33 4 76 20 94 62. Fax: + 33 4 76 20 94 00.

<sup>‡</sup> Unit of Virus Host Cell Interactions, UMR 5233 UJF-EMBL-CNRS.

<sup>§</sup> Present address: CNRS, UMR2472, INRA, UMR1157, IFR115, Laboratoire de Virologie Moléculaire et Structurale, 91198, Gif sur Yvette, France.

<sup>||</sup> Institut Laue Langevin.

<sup>1</sup> Abbreviations: AUC, analytical ultracentrifugation; CHPV, Chandipura virus; DLS, dynamic light scattering; HPLC, high-pressure liquid chromatography; MALLS, multiangle laser light scattering; N, nucleoprotein; P, phosphoprotein; RV, rabies virus; SANS, small-angle neutron scattering; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SEC, size-exclusion chromatography; VSV, vesicular stomatitis virus.



measles virus), the oligomerization domain is required for transcription and replication, and oligomerization is not controlled by phosphorylation (11, 12). For other members of this family, phosphorylation of P is required for its oligomerization and transcriptional activity (13, 14). In the *Rhabdoviridae* family, studies with RV and VSV have provided unclear pictures about the state of oligomerization and of phosphorylation required for the activities of the phosphoprotein. Studies with VSV (*Vesiculovirus* genus) indicate that oligomerization of P is required for binding to L polymerase and to N-RNA template (15–17) and for transcriptional activity (18–20). Oligomerization is controlled by phosphorylation in the N-terminal region of the protein (18–20) and by protein concentration (21). Conversely, phosphorylation in the N-terminal region is not required for the formation of RNA-free N<sup>o</sup>-P complexes and for RNA replication (8, 21, 22). The three-dimensional structure of the central domain of VSV P is dimeric (23), but biochemical studies have suggested that unphosphorylated P is monomeric and that phosphorylated P forms dimers, trimers, or tetramers (19, 24, 25). Recent data support the existence of different RNA polymerase complexes involved in transcription and replication (26, 27), and it is possible that phosphorylation of P acts as a switch between these activities. Substitution of the phosphorylation sites indicates that phosphorylation is not strictly required for transcription and replication of the RNA, but that it stimulates these activities by favoring oligomerization (19, 21). Also, the replacement of two phosphorylation sites by Asp in the N-terminal region was shown to mimic phosphorylation in its ability to induce oligomerization and transcriptional activity (19). Studies with Chandipura virus (CHPV), another *Vesiculovirus*, show that unphosphorylated P is inactive in transcription and exists in a concentration-dependent monomer-dimer-tetramer equilibrium (28). Phosphorylation of a single residue in the N-terminal region induced a large conformational change in the protein (29, 30), stabilized the dimeric form (16), and activated transcription (31). In studies with RV, belonging to another genus (*Lyssavirus* genus) of the *Rhabdoviridae* family, different oligomeric forms of unphosphorylated P were found in solution, suggesting that phosphorylation of P is not required for oligomerization (32). The absence of an *in vitro* transcription system for rabies virus precluded, however, the characterization of the role of P in RNA transcription and replication. Throughout this ensemble of studies, the stoichiometry of the different phosphorylated and unphosphorylated forms of P from both viruses remained unclear, and controversial results have been obtained by different approaches (19, 25, 32).

In order to characterize the stoichiometry and molecular dimensions of the unphosphorylated form of VSV and RV P, we used size-exclusion chromatography, multiangle static light scattering, dynamic light scattering, analytical ultracentrifugation, and small-angle neutron scattering. To investigate the effect of phosphorylation on the oligomerization of VSV P, we characterized the properties of a variant of this protein in which two phosphorylation sites (S60 and T62) were substituted by aspartate residues, substitutions that have been shown previously to mimic a constitutive phosphorylated state of the protein. Our results indicated that, at 20 °C, both WT and mutant proteins form highly asymmetric

dimers and convert to large oligomers (>700 kDa) in a temperature-dependent cooperative process.

## MATERIALS AND METHODS

**Production and Purification of Recombinant P Proteins.** The phosphoprotein from the Orsay strain (Indiana serotype) of VSV (VSV P) was expressed using a recombinant plasmid pET22b+ containing the full length gene with a His-Tag at its C-terminal end (a gift from Dr. Blondel, Gif-sur-Yvette). A second plasmid for expressing the S60D/T62D double mutant of VSV P was constructed by introducing mutations at positions 60 and 62 by PCR site-directed mutagenesis using an inverse PCR method (Quickchange, Stratagene). The sequences of both P genes were verified by standard dideoxy sequencing. Each plasmid was transformed into *Escherichia coli* BL21 (DE3) pLysS strain, and the cells were grown in Luria-Bertoni medium at 37 °C until OD 600 nm reached 0.3. Expression of the recombinant protein was then induced with 0.5 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) for 5 h at 25 °C. Cells were harvested by centrifugation and then suspended in lysis buffer containing 20 mM Tris-HCl, 150 mM NaCl at pH 7.5. The cells were disrupted by sonication on ice. The cell debris was removed by centrifugation for 1 h at 20000g, 4 °C. The supernatant was filtered over a 0.2  $\mu$ m membrane and loaded onto a Ni<sup>2+</sup> charged affinity column (Sigma). The column was washed with 20 mM Tris-HCl, 500 mM NaCl buffer at pH 7.5, and P was eluted with 400 mM imidazole. Fractions containing P were pooled and concentrated on a unit concentration cell (Amicon, cutoff 10 kDa, Millipore) and then applied onto a Hiload 16/60 Superdex 200 column equilibrated with 20 mM Tris-HCl, 150 mM NaCl buffer at pH 7.5 (GE Healthcare). The purity of the protein preparation was checked by 12% Tris-Tricine SDS-PAGE. Identification by immunodetection was done with a rabbit polyclonal antibody (a gift from Dr. J. Curran, Geneva) and a secondary anti-rabbit antibody coupled to alkaline phosphatase. Circular dichroism spectra for both proteins were identical and similar to those published previously (data not shown) (25). The identity and integrity of both VSV proteins were confirmed by electrospray mass spectrometry. The experiments were performed on a Quattro II mass spectrometer (Micromass, Altricham, U.K.) by continuously injecting the sample with a Type 22 pump from Harvard Apparatus at a flow rate of 5  $\mu$ L.min<sup>-1</sup>. VSV P and ST VSV P were at a concentration of 1 mg mL<sup>-1</sup> in H<sub>2</sub>O containing 1% formic acid. The measured molecular masses are 30 975  $\pm$  7 Da and 31 016  $\pm$  5 Da for expected molecular mass calculated from the amino acid sequence including the His-tag of 30 976 Da and 31 018, respectively.

The phosphoprotein from the CVS strain of rabies virus (RV P) was produced in insect cells (High Five) and purified according to the protocol described previously (33, 34). The purity of the protein preparation was checked by 12% Tris-Tricine SDS-PAGE, and identification by immunodetection was done with a mouse monoclonal antibody (25E6) (a gift from Dr. D. Blondel, Gif-sur-Yvette) and a secondary anti-mouse antibody coupled with alkaline phosphatase. Electrospray mass spectrometry revealed a major species with a molecular mass of 33 520  $\pm$  7 Da for and a minor species with a molecular mass of 33 596 Da that could correspond to a phosphorylated form of the protein (expected molecular mass calculated from the amino acid sequence: 33 616 Da).

*Analytical Size Exclusion Chromatography and Multiangle Laser Light Scattering.* Size exclusion chromatography (SEC) was performed with a S200 Superdex column (GE Healthcare). The column was equilibrated in 20 mM Tris-HCl, 150 mM NaCl buffer at pH 7.5. Separations were performed at 20 °C with a flow rate of 0.5 mL min<sup>-1</sup>. Typically, 20 μL of protein solution at a concentration of 3–5 mg mL<sup>-1</sup> was injected. The excluded ( $V_0$ ) and total volumes ( $V_t$ ) were measured with blue dextran and thymidine, respectively. The partition coefficients ( $K_{av}$ ) were calculated by using

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \quad (1)$$

where  $V_e$  is the elution volume of the protein. The column was calibrated using proteins of known Stokes radii ( $R_S$ ) and molecular mass ( $M_w$ ) (35): myoglobin ( $R_S = 2.10$  nm,  $M_w = 17$  kDa), chymotrypsinogen ( $R_S = 2.28$  nm,  $M_w = 25$  kDa), bovine serum albumine ( $R_S = 3.37$  nm,  $M_w = 67$  kDa), aldolase ( $R_S = 4.65$  nm,  $M_w = 158$  kDa), catalase ( $R_S = 5.21$  nm,  $M_w = 232$  kDa), ferritine ( $R_S = 6.80$  nm,  $M_w = 440$  kDa), and thyroglobulin ( $R_S = 7.85$  nm,  $M_w = 669$  kDa) (35).

On-line MALLS detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting at 690 nm. For each point on the HPLC chromatogram, the amount of light scattered is directly proportional to the product of protein concentration and molar mass, according to Zimm's formula for a diluted polymer solution (36, 37):

$$\frac{R_\theta}{K^*C} = MP(\theta) - 2A_2CM^2P^2(\theta) \quad (2)$$

where  $R_\theta$  is the measured excess Rayleigh's ratio,  $C$  is the protein concentration (g mL<sup>-1</sup>),  $M$  is the molar mass (g mol<sup>-1</sup>),  $P(\theta)$  is the form factor, which depends on the structure of the scattering particles and describes the angular dependence of the scattered light, and  $A_2$  is the second virial coefficient.  $K^*$  is an optical constant given by the following equation:

$$K^* = \frac{1}{N_A} \left( \frac{2\pi n_0}{\lambda^2} \right)^2 \left( \frac{dn}{dc} \right)^2 \quad (3)$$

where  $N_A$  is Avogadro's number,  $n_0$  is the refractive index of the solvent at the incident radiation wavelength (1.33 for a diluted aqueous buffer),  $dn/dc$  (mL g<sup>-1</sup>) is the specific refractive index increment of the solute, and  $\lambda$  is the wavelength of the incident light in void. For particles smaller than the incident wavelength, no angular dependence is observed ( $P(\theta) = 1$ ), and for a sufficiently diluted solution,  $A_2$  is negligible ( $A_2 = 0$ ). Equation 2 then simplifies and Rayleigh's ratio only depends on protein concentration and molar mass:

$$\frac{R_\theta}{K^*C} = M \quad (4)$$

The molar mass,  $M$ , can thus be calculated if  $C$  is known. Protein concentration was measured on-line by refractive index measurements using an RI2000 detector (Schambeck SFD) and refractive index increment  $dn/dc = 0.185$  mL g<sup>-1</sup>.

Within the elution peak, the chromatogram is divided in slices, and for each slice, MALLS and refractive index measurements are used to calculate the molar masses using eq 4. The number-averaged ( $M_n$ ) and weight-averaged ( $M_w$ ) molar mass are obtained from the molar mass distribution across the elution peak. Analysis of the data was performed with the ASTRA software (Wyatt Technology Corp., Santa Barbara, CA).

*Dynamic Light Scattering.* Samples with protein concentrations ranging from 0.5 to 3.0 mg mL<sup>-1</sup> were centrifuged for 10 min at 13000g and filtered through a 0.1 μm cutoff membrane filter in order to eliminate dust and large aggregates. Dynamic light scattering was measured at a scattering angle of 90° on a Dynapro apparatus (ProteinSolutions) using a 20 μL cuvette and a laser emitting at  $\lambda = 835.6$  nm. Data were analyzed with the Dynamics v5.26.41 software. The "baseline limit" parameter was used to reject initial measurements. The apparent translational diffusion coefficients,  $D$ , obtained from the autocorrelation function, were converted into hydrodynamic radii ( $R_S$ ) according to the Stokes–Einstein equation:

$$R_S = \frac{RT}{N_A 6\pi\eta_0 D} \quad (5)$$

where  $R$  is the gas constant,  $T$  is the temperature,  $N_A$  is Avogadro's number, and  $\eta_0$  is the solvent viscosity. Control experiments carried out with lysozyme yielded the expected apparent translational diffusion coefficient.

*Sedimentation Velocity Measurements.* Sedimentation velocity experiments were performed in a Beckman XL1 ultracentrifuge at rotor speed of 42 000 rpm and at 20 °C using Epon charcoal-filled 12 or 3 mm double-sector centerpieces. Radial absorbance scans ( $\lambda = 280$  nm) were acquired every 3 or 10 min. The program SEDFIT (38) (available from the RASMB Web site, <http://www.bbri.org/rasmb/rasmb.html>) was used to model the sedimentation profiles using the integrated Lamm equation solutions:

$$A(r,t) = \int c_0(s) \chi(s,D,r,t) ds + \epsilon + \delta \quad (6)$$

where  $A(r,t)$  is the absorbance at radius  $r$  and time  $t$ ,  $s$  and  $D$  are sedimentation and diffusion coefficients,  $c_0$  is the loading concentration of protein,  $\chi$  is the sedimentation profile (i.e., boundary), and  $\epsilon$  and  $\delta$  are signal offsets (the first term accounts for systematic errors, and the second term accounts for random errors). Sedimentation coefficient values between 0.5 and 15 S were resolved using maximum entropy regularization with a 0.68 confidence level. The relationships between the molecular mass,  $M_w$ , the sedimentation coefficient,  $s$ , the diffusion coefficient,  $D$ , and the Stokes radius,  $R_S$ , are obtained using the Svedberg equation (37, 39),

$$D = \frac{sRT}{M_w(1 - \bar{v}\rho_0)} \quad (7)$$

and by combining this equation with eq 5:

$$R_S = \frac{M_w(1 - \bar{v}\rho_0)}{6\pi\eta_0 s} \quad (8)$$

Knowledge of  $s$  and  $M_w$  or of  $s$  and  $D$  allows the evaluation of the translational friction ratio  $f/f_0$ , where  $f_0$  is the

translational friction coefficient for a sphere of the same mass and anhydrous volume, according to the following equations (37):

$$\frac{f}{f_0} = \frac{M_w(1 - \bar{v}\rho_0)}{N_A 6\pi\eta_0 s \left(\frac{3M_w\bar{v}}{4\pi N_A}\right)^{1/3}} \quad (9a)$$

$$\frac{f}{f_0} = \frac{RT}{N_A 6\pi\eta_0 D \left(\frac{3M_w\bar{v}}{4\pi N_A}\right)^{1/3}} \quad (9b)$$

Solvent density ( $\rho_0$ ) of 1.005 g mL<sup>-1</sup>, solvent viscosity ( $\eta_0$ ) of 1.0214 cP, and the partial specific volume ( $\bar{v}$ ) of VSV P and RV P of 0.726 and 0.732 mL g<sup>-1</sup> were calculated at 20 °C with Sednterp (<http://www.bbri.org/RASMB/rasmb.html>). Intrinsic sedimentation and diffusion coefficients ( $s_{20,w}^0$  and  $D_{20,w}^0$ ) corrected to water at 20 °C were calculated from the experimental  $s_{mes}$  and  $D_{mes}$  values obtained in the experimental solvent and temperature conditions according to

$$s_{20,w}^0 = s_{mes} \left(\frac{\eta_{20,w}}{\eta_{T,b}}\right) \left(\frac{1 - \bar{v}\rho_{T,b}}{1 - \bar{v}\rho_{20,w}}\right) \quad (10a)$$

$$D_{20,w}^0 = D_{mes} \left(\frac{\eta_{20,w}}{\eta_{T,b}}\right) \quad (10b)$$

where  $\eta_{20,w}$  and  $\eta_{T,b}$  are the water viscosity at 20 °C and buffer viscosity at the experimental temperature and  $\rho_{20,w}$  and  $\rho_{T,b}$  are the water density at 20 °C and the buffer density at the experimental temperature.

**Small-Angle Neutron Scattering (SANS).** SANS experiments were performed on the D22 beam-line at the Institut Laue-Langevin (ILL, Grenoble, France). Neutron wavelength of 0.6 nm and a sample-to-detector distance of 4 m were used to cover the range  $0.11 \text{ nm}^{-1} \leq Q \leq 2.4 \text{ nm}^{-1}$ . Protein concentration was 5 mg/mL. Scattering data were radially averaged about the beam center, corrected for transmission, corrected by subtracting empty cell and solvent scattering, and calibrated on an absolute scale by using the scattering of 0.100 cm of water (40).

Analysis of the scattering curves at small  $Q$  values using the Guinier approximation gives the radius of gyration ( $R_g$ ) and the forward scattering intensity at zero  $Q$  ( $I_0$ ) according to (37, 41)

$$I(Q) = I_0 \exp\left(\frac{-R_g^2 Q^2}{3}\right) \quad (11)$$

The scattering amplitude,  $Q$ , is given by the following equation:

$$Q = \frac{4\pi \sin \theta}{\lambda} \quad (12)$$

where  $2\theta$  is the scattering angle and  $\lambda$  is the wavelength. The radius of gyration was calculated from the initial slope of the plot of  $\ln(I)$  versus  $Q^2$ . The molecular mass of the particles,  $M_w$ , was directly calculated from the forward intensity,  $I_0$ , knowing the protein concentration and the scattering length density (40).

$$M_w = \frac{I(0)N_A}{c(\Delta\rho\bar{v})^2} \quad (13)$$

where  $N_A$  is Avogadro's number,  $c$  is the protein concentration in mg mL<sup>-1</sup>,  $\Delta\rho$  is the difference,  $\rho - \rho_0$ , between the scattering length density of the molecule,  $\rho$ , and the scattering length density of the solvent,  $\rho_0$ , and  $\bar{v}$  is the partial specific volume of the protein.

If one of the molecular dimensions is longer than the other two, the radius of gyration of the cross-section,  $R_c$ , can be obtained from the plot of the  $\ln(I \cdot Q)$  versus  $Q^2$  (41), according to the following equation:

$$\ln[I(Q \cdot Q)] = \ln[I(Q \cdot Q)]_{Q \rightarrow 0} - \frac{R_c^2 Q^2}{2} \quad (14)$$

and triaxial molecular dimensions were calculated by using  $R_g$  and  $R_c$  values derived from Guinier and cross-sectional analysis (41). Assuming a cylindrical shape for P proteins, the length of the cylinder,  $L$ , is given by

$$L = \sqrt{12(R_g^2 - R_c^2)} \quad (15)$$

and the radius,  $r$ , is given by

$$r = \sqrt{2}R_c \quad (16)$$

The data for  $Q$  values ranging from 0.11 to 1.0 nm<sup>-1</sup> were used with the GNOM program (42) to determine the distance distribution function,  $P(r)$ , the radius of gyration,  $R_g$ , and the forward intensity,  $I(0)$ . The maximum dimension,  $D_{max}$ , required in this procedure and beyond which  $P(r) = 0$ , was selected such that the  $P(r)$  function remained stable.

**Oligomerization of P and Electron Microscopy.** Samples of VSV P and ST VSV P proteins at 2.6 mg mL<sup>-1</sup> were incubated in 20 mM Tris/HCl, 150 mM NaCl, pH 7.5 at 30 °C for different times. The kinetics of disappearance of the dimers and of appearance of the large oligomers of P at 30 °C were fitted to second-order rate equation using SigmaPlot (SPSS Inc., Chicago, IL). Samples obtained after incubation in those conditions for 1 or 3.5 h were used for electron microscopy. These samples were diluted to a concentration of about 0.1 mg mL<sup>-1</sup>, adsorbed onto the clean face of a carbon film deposited on a mica sheet, negatively stained with 1% (w/v) uranyl acetate, and observed under low-dose conditions with a JEOL 1200 EX II microscope at 100 kV and a nominal magnification of 40000 $\times$ .

## RESULTS

**Stoichiometry of P Protein Oligomers in Solution.** Recombinant VSV P and its S60D/T62D variant (ST VSV P) were expressed in *E. coli*, and RV P was expressed in insect cells. All three proteins were purified as full-length proteins as shown by mass spectrometry. They elute from a Superdex 200 (GE Healthcare) SEC column as single peaks (Figure 1). The absolute molecular mass was determined at each point of the chromatograms by combining detections by MALLS and refractometry (eq 4). For all of these P proteins, the molecular mass is constant across the elution peak (Figure 1), and the polydispersity factor ( $M_w/M_n$ ) is equal to  $1.00 \pm 0.05$ , indicating monodisperse species. The weight-average

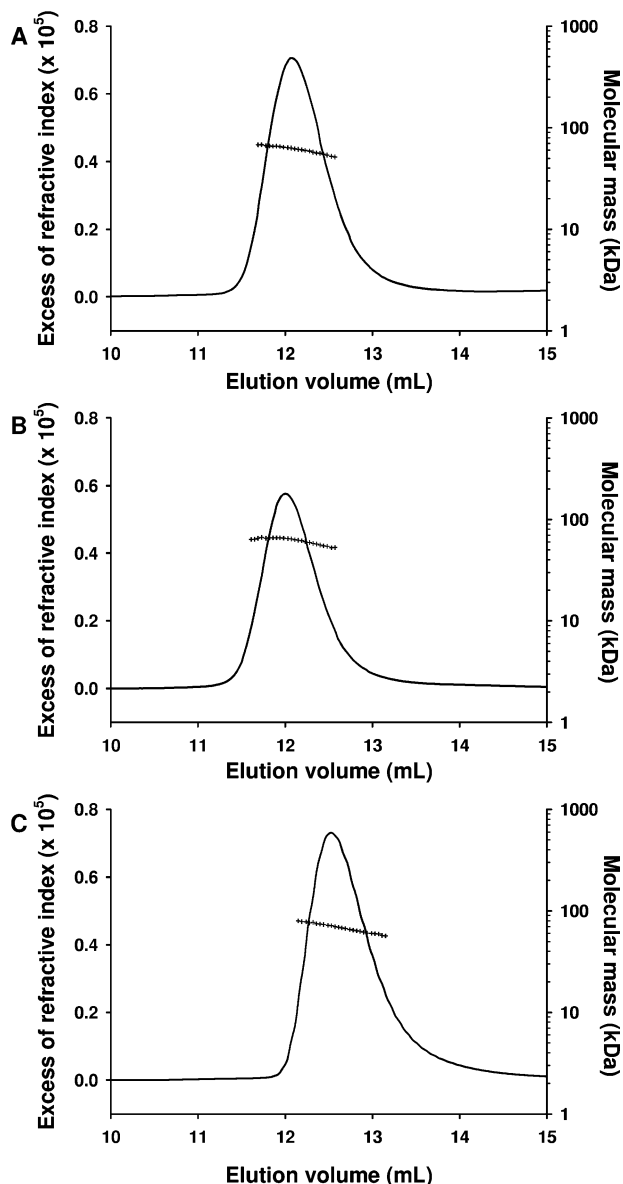


FIGURE 1: Molecular mass of phosphoprotein determined by multiangle laser light scattering and refractometry combined with size-exclusion chromatography. (A) VSV P, (B) ST VSV P, and (C) RV P. SEC was performed on a Superdex 200 column (300 mm  $\times$  10 mm) (GE Healthcare), equilibrated in 20 mM Tris-HCl, 150 mM NaCl buffer at pH 7.5. Separations were performed at 20  $^{\circ}$ C with a flow rate of 0.5 mL  $\text{min}^{-1}$ . 50  $\mu\text{L}$  of a protein solution at 3–5 mg  $\text{mL}^{-1}$  was injected. The line shows the elution profile monitored by excess refractive index (left ordinate axis). The crosses (+) show the molecular mass distribution (right ordinate axis) determined from MALLS and refractometry data. Detectors at 60 $^{\circ}$ , 69 $^{\circ}$ , 80 $^{\circ}$ , 90 $^{\circ}$ , 100 $^{\circ}$ , 111 $^{\circ}$ , and 121 $^{\circ}$  were used for the analysis, and the light scattering profile displays no angular dependence.

( $M_w$ ) molecular mass values are  $61 \pm 2$  kDa,  $63 \pm 3$  kDa, and  $69 \pm 2$  kDa for VSV P, ST VSV P, and RV P, respectively, demonstrating that the three P proteins form dimers in solution (molecular masses of dimeric P proteins calculated from the sequence are 61 952 Da, 62 018 Da, and 67 232 Da, respectively) (Table 1 and Figure 1). No significant variation of the average molecular mass is observed for concentrations ranging from 100 to 645  $\mu\text{M}$ , and the elution volume is identical for protein concentrations ranging from 20  $\mu\text{M}$  to 645  $\mu\text{M}$  (data not shown). Similar results (see below) confirming that P proteins form dimers

were obtained from small-angle neutron scattering experiments (SANS) and also by combining results from sedimentation velocity, SEC, and DLS (Table 1).

Some of our samples of RV P contained a small population (near 5%) of other species eluting from the SEC column at smaller volumes than RV P dimers. These species migrate differently than the major species on SDS-PAGE in the absence of a reducing agent. RV P contains two cysteine residues, and most of these additional species disappear upon addition of a reducing agent, suggesting that they result from the formation of intermolecular disulfide bonds (data not shown).

**Hydrodynamic Properties.** The Stokes radii of the three P proteins were determined by SEC using globular proteins of known  $R_S$  for calibration (Figure 2). Values of  $5.2 \pm 0.1$  nm,  $5.3 \pm 0.1$  nm, and  $4.8 \pm 0.1$  nm were obtained for VSV P, ST VSV P, and RV P, respectively (Table 2). These values are larger than those expected for a globular protein of the size of P monomers or even of P dimers. The  $R_S(\text{measured})/R_S(\text{globular})$  ratio, where  $R_S(\text{globular})$  is the Stokes radius for a globular protein of about 65 kDa (dimer), is of the order of 1.65–1.70, indicating that P dimers are nonspherical and/or partially disordered. Similar results were obtained with a Superose 12 column (data not shown).

Dynamic light scattering, measured at a protein concentration of 80  $\mu\text{M}$ , shows that both VSV P and ST VSV P are monodisperse species with translational diffusion coefficient ( $D_{20,w}$ ) values of  $4.27 \pm 0.08 \cdot 10^{-7}$   $\text{cm}^2 \text{s}^{-2}$  and  $4.15 \pm 0.08 \cdot 10^{-7}$   $\text{cm}^2 \text{s}^{-2}$ , respectively (Table 2). Hydrodynamic radii,  $R_S$ , calculated according to the Stokes–Einstein equation (eq 5), are in good agreement with the values measured by SEC (Table 2). With RV P, the DLS measurements indicate the presence of several species, with a major component exhibiting a  $D$  value of  $4.48 \pm 0.10 \cdot 10^{-7}$   $\text{cm}^2 \text{s}^{-2}$ , also in good agreement with the value measured by SEC.

The hydrodynamic properties of the three P proteins were finally characterized by sedimentation velocity experiments at three different protein concentrations. Small amounts of large aggregates corresponding to 5 to 15% of the absorbance signal are detected at low protein concentrations (0.25 mg  $\text{mL}^{-1}$  and 1 mg  $\text{mL}^{-1}$ ) as species that sediment very rapidly ( $s$  of about 60 S), whereas they are not detected in the sample of the highest concentration (4 mg  $\text{mL}^{-1}$ ). These large species are not further considered in the following analysis. The continuous sedimentation distribution ( $c(s)$  versus  $s$ ) was calculated using numerical solutions of the Lamm equation to fit the boundaries (37, 38). In a first approach, the translational friction ratio ( $f/f_0$ ) was treated as a fitting parameter and was determined through data analysis. Values of  $f/f_0$  ranging from 1.7 to 2.1 were obtained for the different protein concentrations together with size distributions of the sedimentation coefficient centered on  $s_{20,w}$  values ranging from 2.9 to 3.5 S. A spherical protein ( $f/f_0 = 1.25$ ) of 65 kDa would have a  $s_{20,w}$  value of 4.7 S, indicating that the P proteins are elongated. The  $s_{20,w}$  value corresponding to the maximum of the size distribution varies little with protein concentration (data not shown), and thus, in a second approach, a value of  $f/f_0$  for each protein was calculated by combining the most probable value of  $s_{20,w}$ , with the translational diffusion coefficient,  $D_{20,w}$ , measured by DLS or with the molecular mass,  $M_w$ , measured by MALLS (eqs 9a and 9b). This  $f/f_0$  value was then used as a fixed parameter

Table 1: Molecular Mass of P Proteins Obtained by Different Methods<sup>a</sup>

	$M_w$ (kDa)					
	from the sequence (for a dimer)	from MALLS	from AUC ( $s$ ) and SEC ( $R_S$ )	from AUC ( $s$ ) and DLS ( $D$ )	from SANS-Guinier	from SANS- $P(r)$
VSV P	61.952	61 ± 2	63 ± 5	61 ± 5	69 ± 4 (0.216 ± 0.008)	69 ± 5 (0.210 ± 0.017)
ST VSV P	62.018	63 ± 3	69 ± 5	67 ± 5	79 ± 5 (0.180 ± 0.012)	75 ± 5 (0.170 ± 0.012)
RV P	67.232	69 ± 2	71 ± 5	71 ± 5	81 ± 3 (0.255 ± 0.018)	78 ± 5 (0.253 ± 0.017)

<sup>a</sup> A value of the weight average molecular mass was obtained from MALLS data by averaging the molecular mass values calculated at each point of the chromatographic peak (Figure 1) using protein concentration measured by refractometry and the intensity of scattered light at 7 different angles (eq 4). A value was obtained from the sedimentation coefficient,  $s$ , measured by analytical ultracentrifugation (AUC) and either the Stokes radius,  $R_S$ , measured by SEC according to eq 8 or the diffusion coefficient measured by DLS according to eq 7. A value was obtained from the protein concentration, the scattering length density, and the intensity of neutron scattering at zero angle,  $I(0)$  (SANS), either measured by extrapolating the Guinier plot to zero angle or from the analysis of the distance function distribution ( $P(r)$ ), according to eq 13. Values in parentheses show  $I(0)$  values in  $\text{cm}^{-1}$ .

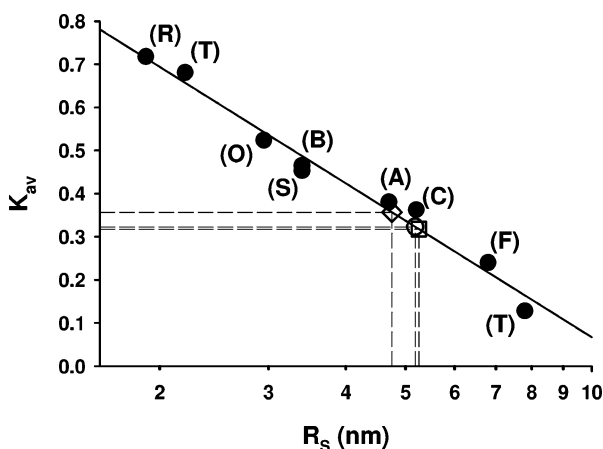


FIGURE 2: Stoke's radius ( $R_S$ ) of VSV P, ST VSV P, and RV P proteins measured by size-exclusion chromatography. The figure shows the plot of  $K_{av}$  versus  $R_S$ . SEC was performed as described in Figure 1. Standard proteins (filled circles) used to calibrate the column were (T) thyroglobulin, (F) ferritine, (C) catalase, (A) aldolase, (S) serumalbumine, (O) ovalbumine, (T)  $\alpha$ -chymotrypsinogen, and (R) RNase A. Stokes radii are taken from Uversky (35). The line shows the linear fit ( $K_{av} = 0.964 - 0.897 \log(R_S)$ ) with a correlation coefficient of 0.978. The open symbols show the measured  $K_{av}$  and calculated Stokes radii of VSV P (circle), ST VSV P (square), and RV P (diamond).

in the analysis of the data (Table 3). Figure 3 shows the plots of  $c(s)$  versus  $s$  for the three P proteins obtained from this second analysis. For VSV P (Figure 3A) and ST VSV P (Figure 3B) the analysis yields a single symmetrical peak with  $s$  values of  $2.8 \pm 0.1$  S ( $s_{20,w} = 2.9$  S) and  $3.0 \pm 0.1$  S ( $s_{20,w} = 3.1$  S), respectively (Table 2), confirming the existence of monodisperse species. With RV P (Figure 3C), several species are observed. The major species that represents 65% of the molecules has an  $s$  value of  $3.4 \pm 0.1$  S ( $s_{20,w} = 3.5$  S). These experiments were performed in the absence of a reducing agent and several species with  $s$  values ranging from 4 to 12 S (35%) could correspond to the disulfide bridged forms observed by SEC. Combining  $s$  values measured by analytical centrifugation with values of  $M_w$ ,  $D$ , or  $R_S$  determined by other methods yields self-consistent estimates of these different parameters (Tables 1 and 2). The values obtained with VSV P, ST VSV P, and the major form of RV P support the conclusion that P proteins form elongated dimers. The hydrodynamic param-

eters measured by the different methods are summarized in Table 2.

**Overall Shape of P by Small-Angle Neutron Scattering.** SANS experiments were used to characterize the molecular dimensions of the three P proteins. Guinier plots (Figure 4A) are linear for  $Q$  values ranging from 0.10 to 0.28  $\text{nm}^{-1}$  ( $0.5 < Q \cdot R_g < 1.5$ ) and yield model-independent values for the radii of gyration of 5.3 to 5.5 nm (Table 3). The scattered intensity extrapolated to zero angle,  $I(0)$ , provides an estimate of the molecular mass of the protein (Table 1) that is slightly higher than that measured by light scattering but remains in agreement with the molecular mass of dimers (Table 1). Similar  $R_g$  and  $M_w$  values are obtained from the size distribution function ( $P(r)$ ) calculated for  $Q$  values ranging from 0.15 to 0.76  $\text{nm}^{-1}$  (Figure 4C and Table 2). The discrepancy between molecular mass values measured by MALLS and SANS is larger for RV P than for VSV P and ST VSV P, possibly because of the presence of small amounts of larger species in the RV P samples. Cross-sectional Guinier plots (Figure 4B) are linear for  $Q$  value ranging from 0.22 to 0.75  $\text{nm}^{-1}$  ( $0.3 < Q \cdot R_c < 1.2$ ), an indication that the P proteins are elongated in one direction. The slope of these plots yielded the cross-sectional radii of gyration,  $R_c$ , of about 1.6–1.8 nm (Table 3). From  $R_g$  and  $R_c$  values, we calculated triaxial parameters for a cylinder (Table 3) (41). Values for the partial specific volumes calculated from the cylinder models and the molecular mass measured by MALLS are close to the values calculated from the amino acid composition of the proteins using Sedenterp.

**Formation of Large Oligomers at 30 °C.** Incubation of VSV P and ST VSV P at 30 °C induces the formation of larger oligomers that elute in the exclusion volume of our SEC column ( $M_w > 700$  kDa) (Figure 5A). The conversion occurs in about 8 h, and no intermediate forms accumulate during the reaction. The kinetics of disappearance of dimeric P (Figure 5B) is identical to the kinetics of appearance of the large oligomeric species. Electron microscopy showed that both VSV P and ST VSV P form short oligomeric species after incubation at 30 °C for 1 h (Figure 6A) that grow into irregular filamentous structures after incubation for 3.5 h (Figure 6B). Both oligomers and filaments have an irregular diameter of about 6–8 nm, and it is not possible to visualize the organization of the P dimers in these filaments. Incubation of RV P at 30 °C also leads to the

Table 2: Hydrodynamic Properties of P Proteins Obtained by Different Methods<sup>a</sup>

	$s_{20,w}$ (S)	$ff_0$	$D_{20,w}$ ( $10^{-7}$ cm s <sup>-2</sup> )			$R_S$ (nm)		
			from DLS	from AUC (s) and MALLS ( $M_w$ )	from AUC (s) and SEC ( $R_S$ )	from SEC	from DLS	from AUC (s) and MALLS ( $M_w$ )
VSV P	2.9 ± 0.1	1.92	4.27 ± 0.08	4.27 ± 0.10	4.13 ± 0.10	5.2 ± 0.1	5.1 ± 0.1	5.0 ± 0.1
ST VSV P	3.1 ± 0.1	1.88	4.15 ± 0.08	4.41 ± 0.10	4.05 ± 0.10	5.3 ± 0.1	5.2 ± 0.1	4.9 ± 0.2
RV P	3.5 ± 0.1 (65%)	1.74	4.48 ± 0.10	4.60 ± 0.10	4.47 ± 0.10	4.8 ± 0.1	4.8 ± 0.1	4.7 ± 0.2

<sup>a</sup> The  $s_{20,w}$  was obtained from the measured  $s$  value using eq 10a to correct for the viscosity and density of the buffer solution. The number within parentheses indicates the fraction of the major component of RV P. The  $ff_0$  is the frictional coefficient ratio used to fit the sedimentation velocity data. It corresponds to an average of the values calculated from a gross estimation of  $s$  and from  $M_w$  (MALLS) or  $D$  (DLS), according to eqs 9a and 9b, respectively. The diffusion coefficient,  $D_{20,w}$ , was measured by DLS and was calculated from the sedimentation coefficient,  $s_{20,w}$ , measured by analytical ultracentrifugation (AUC) and from either the molecular mass measured by MALLS or the Stokes radius measured by SEC, by combining eqs 7 and 8. The Stokes radius was measured by SEC, was calculated from the diffusion coefficient,  $D_{20,w}$ , according to eq 5, and was calculated by combining the sedimentation coefficient,  $s_{20,w}$ , measured by analytical ultracentrifugation (AUC) with the molecular mass measured by MALLS, according to eq 8.

Table 3: Molecular Dimensions of P Proteins Calculated from SANS Experiments<sup>a</sup>

	$R_g$ (nm)		$R_c$ (nm) from SANS-Guinier	$D_{max}$ (nm)	$L$ (nm)	$r$ (nm)
	from SANS-Guinier	from SANS- $P(r)$				
VSV P	5.3 ± 0.1	5.4 ± 0.1	1.55 ± 0.02	17.0	17.6 ± 0.3	2.2 ± 0.1
ST VSV P	5.4 ± 0.1	5.5 ± 0.1	1.63 ± 0.02	17.5	17.8 ± 0.4	2.3 ± 0.1
RV P	5.5 ± 0.1	5.5 ± 0.1	1.76 ± 0.02	19.0	18.1 ± 0.3	2.5 ± 0.1

<sup>a</sup> The radius of gyration,  $R_g$ , was calculated from the initial slope of the Guinier plot ( $\ln(I(Q))$  versus  $Q^2$ ) or from the analysis of the distance distribution function ( $P(r)$ ). The cross-sectional radius of gyration,  $R_c$ , was calculated from the slope of the plot of ( $\ln(I(Q) \cdot Q$ ) versus  $Q^2$ ).  $D_{max}$  represents the optimal value obtained in the calculation of the distance distribution function that gave a  $R_g$  value in agreement with that from Guinier analysis.  $L$  and  $r$  represents the length and radius of model cylinder calculated from  $R_g$  and  $R_c$  values according to eqs 15 and 16.

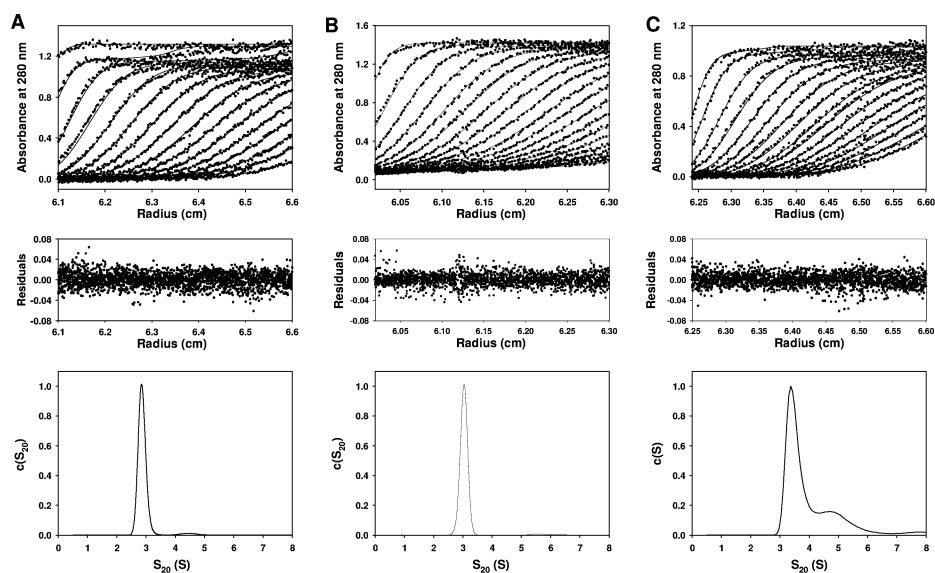


FIGURE 3: Sedimentation coefficient distribution calculated by modeling of sedimentation velocity concentration profiles. The figure shows the scans and fitting curves resulting from the analysis with Sedfit software (top panels), residuals of the fit (middle panels), and sedimentation coefficient distribution (bottom panels) for (A) VSV P, (B) ST VSV P, and (C) RV P. The sedimentation coefficients at the maximum of the distributions are 2.8 ± 0.1 S (VSV P WT), 3.0 ± 0.1 S (VSV P double mutant), and 3.7 ± 0.1 S (RV P WT). For all samples, the ultracentrifuge was run at 42 000 rpm and 20 °C. The buffer contained 20 mM Tris/HCl, 150 mM NaCl (pH 7.5). Cells were scanned at 280 nm.

formation of large filamentous oligomers but much more slowly, such that after incubation for 3.5 h only few were observed by electron microscopy.

## DISCUSSION

*Structural Properties of P Oligomers.* Combining static light scattering detection with SEC, we clearly demonstrated that, in solution near neutral pH, the recombinant P proteins from VSV and RV form dimers. Since the VSV phosphoprotein was produced in bacteria and since its mass determined

by mass spectrometry was that of the unphosphorylated protein, these observations suggest that phosphorylation is not necessary for dimerization. Static laser light scattering is a powerful tool for measuring absolute molecular mass of macromolecules and of multimolecular complexes (36, 37). For particles smaller than the wavelength of the incident light, the intensity of scattered light extrapolated to zero angle is directly proportional to the molecular mass of the particle and to its concentration. Because it does not depend on calibration with globular proteins, this detection method is

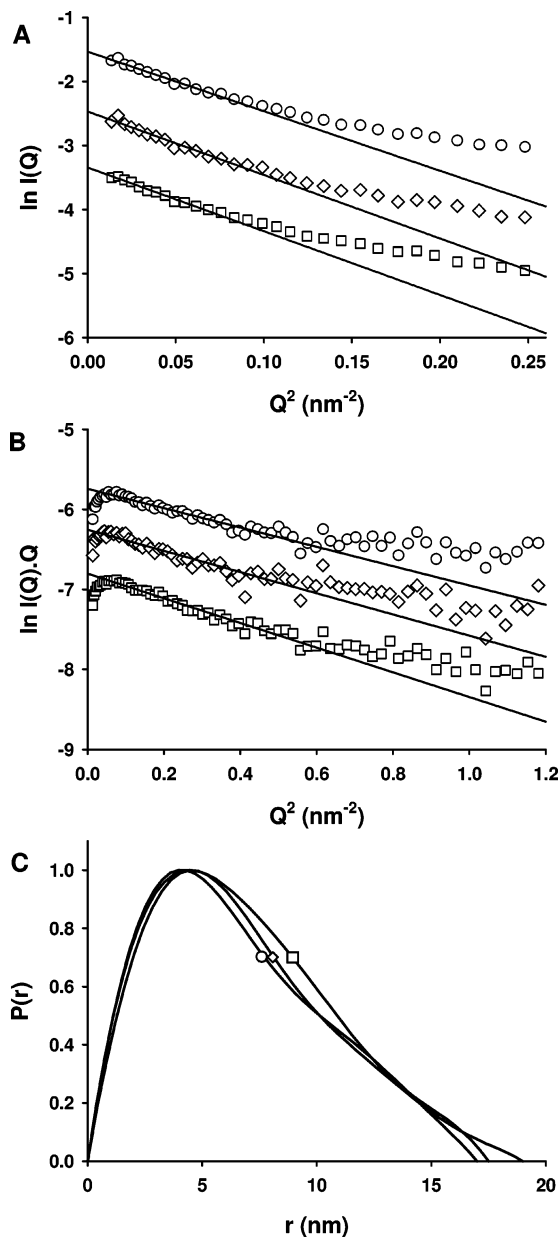


FIGURE 4: Small-angle neutron scattering analysis of phosphoproteins. (A)  $R_g$  Guinier plot ( $\ln(I(Q))$  versus  $Q^2$ ) for VSV P (circles), ST VSV P (diamonds), and RV P (squares). The fitted lines using  $0.10 < Q < 0.28 \text{ nm}^{-1}$  yielded values for the radii of gyration of  $5.3 \pm 0.1 \text{ nm}$  for VSV P,  $5.4 \pm 0.1$  for ST VSV P, and  $5.5 \pm 0.1$  for RV P. The fit corresponds to a range of  $0.5 < Q \cdot R_g < 1.5$ . The  $I(0)$  value calculated from the intercept, together with protein concentrations, corresponds to molar mass of  $69\,000 \pm 4\,000$  for VSV P,  $79\,000 \pm 5\,000$  for ST VSV P, and  $81\,000 \pm 3\,000$  for RV P. For clarity, each plot is shifted along the  $\ln(I(Q))$  axis. (B) Cross-sectional  $R_c$  Guinier plot ( $\ln(I(Q) \cdot Q)$  versus  $Q^2$ ) for VSV P (circles), ST VSV P (diamonds), and RV P (squares). The fitted lines using  $0.22 < Q < 0.75 \text{ nm}^{-1}$  yielded values for the cross-sectional radii of gyration of  $1.55 \pm 0.02 \text{ nm}$  for VSV P,  $1.63 \pm 0.02$ , for ST VSV P, and  $1.76 \pm 0.01$  for RV P. The fit thus corresponds to a range of  $0.3 < Q \cdot R_c < 1.3$ . For clarity, each plot is shifted along the  $\ln(I(Q) \cdot Q)$  axis. All samples were prepared in 20 mM Tris/HCl, 150 mM NaCl, pH 7.5, in 100%  $\text{H}_2\text{O}$ . Protein concentrations were close to  $5 \text{ mg mL}^{-1}$ . (C) Distance distribution function for VSV P (circle), ST VSV P (diamond), and RV P (square). The curve are calculated using the GNOM software, data for  $Q$  values ranging from 0.15 to  $0.76 \text{ nm}^{-1}$ , and  $D_{\text{max}}$  values reported in Table 3.

particularly suitable for determining the molecular mass of multimolecular complexes, or of elongated or partially

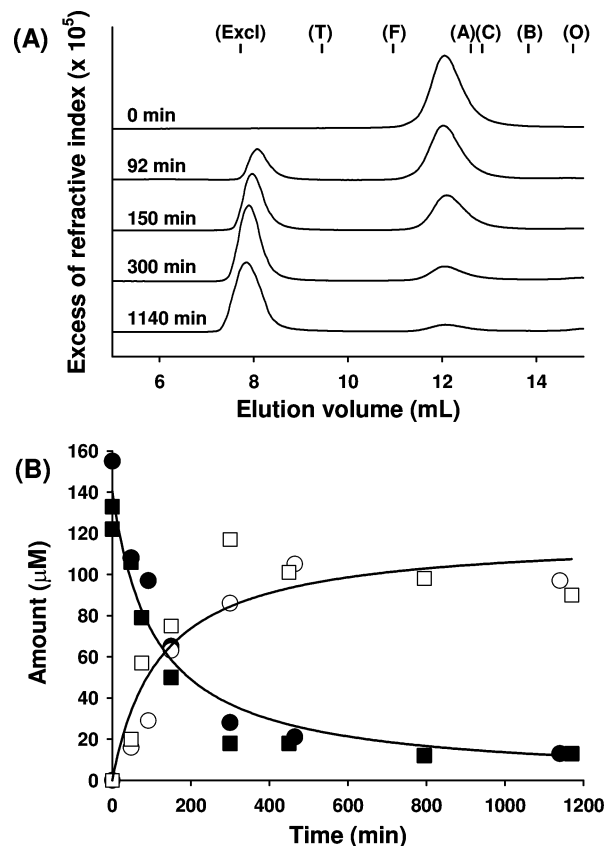


FIGURE 5: Kinetics of the oligomerization of VSV P at 30 °C. (A) Size exclusion chromatography profiles after different incubation times at 30 °C. VSV and ST VSV phosphoproteins at  $2.6 \text{ mg/mL}$  were incubated in 20 mM Tris/HCl, 150 mM NaCl at 30 °C pH 7.5. After different incubation times,  $50 \mu\text{L}$  aliquots were injected on the column and monitored by refractometry. The elution volumes of standard globular proteins are shown at the top of the figure with labels as in Figure 2. (B) Kinetic curves. Peaks areas were converted to concentrations and plotted as a function of incubation time. Closed symbols are for dimers of VSV P (circles) and ST VSV P (squares), and open symbols are for oligomers eluting in the exclusion volume of the column (molar mass  $> 1 \times 10^6 \text{ g mol}^{-1}$ ). The data were fitted to a second-order equation (solid lines).

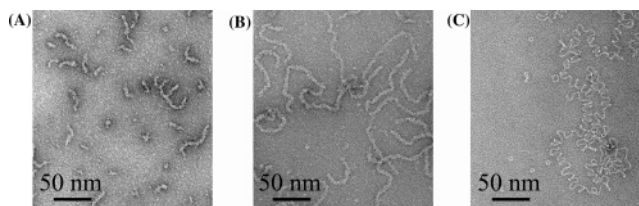


FIGURE 6: Electron microscopy of filamentous VSV P. Electron micrographs of negatively stained VSV P after 1 h (A) and 3.5 h (B) of incubation in 20 mM Tris/HCl, 150 mM NaCl, at pH 7.5 and 30 °C. Protein concentration during incubation at 30 °C was  $2.6 \text{ mg mL}^{-1}$ . (C) VSV ribonucleoprotein complex purified from infected cells and shown for comparison.

disordered proteins for which elution from size-exclusion chromatography columns is different from that of a globular protein. The combination with separation by SEC ensures that the measurement is not contaminated by the presence of small amounts of large aggregates. These results were confirmed by SANS experiments, in which the intensity extrapolated at zero angle is proportional to the molecular mass, which can thus be calculated on an absolute scale when the protein composition and concentration are known (37). Although SANS data were obtained without on-line separa-

tion by SEC, the molecular masses calculated from intensity at zero angle also show that P proteins form dimers (Table 1). We obtained similar results with ST VSV P (S60D/T62D variant) that was made to resemble a constitutively phosphorylated P protein, suggesting that phosphorylation does not induce oligomer formation or further oligomerization. The existence of dimers of P is in agreement with observations that two molecules of RV P bind per molecule of N in N<sup>o</sup>-P complexes (43), with the recent crystal structure of the central dimeric domain of VSV P protein (23), as well as with a previous study in which a combination of SEC and analytical centrifugation data indicated that RV P is larger than a monomer (32) and is consistent with a dimer (Y. Gaudin, personal communication). The Stokes radius of RV P (4.3 nm) measured in this previous study (32) is, however, significantly smaller than that measured here (4.8 nm). The results presented here contrast with previous work in which unphosphorylated VSV P was found to elute from a SEC column as a monomer (19, 24, 25), while phosphorylated forms of VSV P and ST VSV P were reported to form either dimers (24), trimers (25), or tetramers (19). Some of these experiments were carried out at lower protein concentrations (24, 25), which could explain the observation of monomer, but the discrepancies and difficulties to obtain a clear-cut answer about the oligomeric state of *Rhabdoviridae* phosphoproteins arise from several peculiar properties of these proteins. First, P proteins exhibit an anomalous migration on SDS-PAGE, and thus, cross-linking experiments revealed by SDS-PAGE, although they demonstrate the existence of different forms of P, cannot be used to determine the accurate stoichiometry of the oligomers (16, 19, 32). Second, P protein oligomers are elongated and possibly partially disordered, rendering hydrodynamics measurements by SEC and analytical centrifugation difficult to interpret in the absence of an absolute measurement of the molecular mass. Third, VSV P and ST VSV P convert rapidly into soluble large oligomeric species upon incubation at 30 °C. For example, in the His-tag dilution assays used to demonstrate the oligomerization induced by phosphorylation (16, 19) the protein was incubated at 32 °C for 30 min. All experiments that involved a long incubation time above 25 °C should thus be taken with caution. Similar difficulties were encountered for determining the stoichiometry of P oligomers from Sendai and mumps viruses (*Paramyxoviridae*). Sendai virus P protein has been reported to form trimers (11), until it was demonstrated that it forms tetramers (44, 45). With Chandipura virus (CHPV), phosphorylation of a single residue stabilizes the dimers (16), but it has been found recently that different oligomeric forms of unphosphorylated P also coexist in a concentration-dependent equilibrium (28).

The P proteins from VSV and RV exhibit an elongated rodlike shape as suggested by the large Stokes radii measured by SEC and as confirmed by the large frictional ratio obtained from the sedimentation velocity experiments and by the shape of the cross-sectional Guinier plots. The results obtained by sedimentation velocity, dynamic light scattering, and size-exclusion chromatography agree with dimers exhibiting a hydrodynamic radius of 4.8–5.3 nm and a frictional ratio of 1.7–1.9. SANS experiments yielded a radius of gyration of about 5.3–5.5 nm and a cross-sectional radius of gyration of about 1.6–1.8 nm for the three phosphoproteins, which indicated elongated molecules with

one axis 8–7 times longer than the other two. Assuming a cylindrical shape, triaxial parameters, the radius,  $r$ , and the length,  $L$ , can be calculated from the values of  $R_g$  and  $R_c$  (Table 3). The theoretical frictional ratio ( $f/f_0$ ) calculated for a cylinder of such dimensions ranges from 1.3 to 1.9 (46). The partial specific volume of the proteins calculated for such hypothetical cylinders indicates that, for VSV P, the protein occupied 96% of the volume. For RV P, the triaxial parameters, as the molecular mass (see above), were possibly overestimated because of the presence of larger species in the sample. The experimental factor  $\delta = R_g/R_S$  is a sensitive indicator of the compactness of the molecule. The values found here range from 1.0 to 1.2. Its value is 0.775 for a compact sphere, 0.8 for a globular protein (47), about 0.9 for a molten globule (48), and 1.5 for a Gaussian coil (49). The theoretical  $R_g/R_S$  ratio calculated for a cylindrical shape of dimensions similar to those of P proteins ranges from 0.70 to 1.20 (50). Recent studies have concluded that P proteins from *Paramyxoviridae* viruses are intrinsically disordered proteins (IDPs) (51, 52) with structured domains separated by long disordered regions (53, 54). RV and VSV P are sensitive to proteolytic digestion, and analysis of the amino acid sequence suggested that P proteins from *Rhabdoviridae* viruses also belong to this class of proteins (51). Therefore, both an elongated shape and the flexibility inherent to the presence of long disordered regions can explain a  $R_g/R_S$  ratio close to unity. Also, it is noteworthy that a truncated form of Sendai virus P, which contains the central tetramerization domain and the C-terminal N-RNA binding domain (PCT), exhibits similar shape and molecular dimensions (45, 55). The structure of the dimerization domain of VSV, however, reveals that interactions between P molecules in these dimers do not occur through the formation of a long coiled-coil, but rather through central contacts between an  $\alpha$ -helix from each monomer and the formation of two intermolecular  $\beta$ -sheets on the sides (23). For RV P, secondary structure prediction algorithms also predict  $\alpha$ -helices in the central region rather than a coiled-coil, suggesting a structural organization for this domain similar to that of VSV (data not shown).

**Implications for Viral RNA Synthesis.** With both VSV and CHPV, phosphorylation of residues in the N-terminal domain of P proteins or replacements of these residues by acidic residues have been correlated with transcription activity (16, 19). Recent data obtained with VSV suggest that two distinct multimolecular complexes carry out RNA transcription and replication (26, 27) and initiate RNA synthesis at different sites on the genome (56). Multiple results suggest that phosphorylation of P protein from VSV and from CHPV regulates the switch between transcription and replication and that the molecular origin of this change in activity resides in a change in conformation (29, 57) and self-assembly of the P protein (16, 19). Phosphorylation of different residues in the N- and C-terminal domains of VSV P are essential for viral growth (58). Phosphorylation in the C-terminal domain might control replication (59), whereas phosphorylation in the N-terminal domain is necessary for transcription activation (18, 20, 60) but not for replication (60). Replacements of phosphorylation sites in the N-terminal domain by alanine can severely impair transcription activity (21, 60), whereas replacements by glutamate or aspartate residues yield proteins that are constitutively active in transcription



(15, 19, 60). In particular, it has been reported that the double mutant S60D/T62D of VSV P (ST VSV P) mimics phosphorylation by constitutively producing oligomeric species (19). Phosphorylation of one residue in CHPV N-terminal domain has also been shown to control transcription (30). Here, however, we show that unphosphorylated P proteins from VSV and RV as well as the mutant of VSV (S60D/T62D) that is constitutively active in transcription (19, 60) form dimers. Although phosphorylation seems to be required for transcription activation, our results suggest that phosphorylation and transcriptional activation do not need oligomeric structures larger than a dimer, and that self-association of phosphorylated P is not involved in controlling the transcription/replication switch. Alternatively, it is possible that the addition of negative charges in a highly acidic region of P proteins through phosphorylation or mutation might affect the conformation of P (29), its dynamics, or its affinity for another component of the transcription or replication complexes (15). Interactions of P proteins with the leader RNA could also play an important role in the regulation of the replication and transcription activities (28), and phosphorylation of P could also affect these interactions.

Finally, we showed that both VSV P and ST VSV P form filamentous oligomers of larger size (>700 kDa) when incubated at 30 °C. The conversion is cooperative, and no intermediate of small size accumulates during the process. Mutations of the phosphorylation sites have no effect on the kinetics of the oligomerization. In the previous studies (19), phosphorylation has been performed by incubation of P in the presence of casein kinase II for 45 min or more at 32 °C, and thus, it is possible that the appearance of the higher molecular weight forms was attributed wrongly to the formation of tetramers. These filamentous species form under physiological temperatures, and it is now of major interest to understand whether or not they play a relevant role in the replication of the virus and interact with the viral ribonucleoprotein complex (Figure 6C). Additional binding of Sendai virus P protein to N:RNA template is required for stimulating the activity of the RNA polymerase complex (61), and it will thus be important to verify if such filamentous forms of P are involved in the replication processes under physiological conditions.

## ACKNOWLEDGMENT

We thank Danielle Blondel (Unite Mixte de Virologie Moléculaire et Structurale, Gif sur Yvette) for the plasmid encoding the VSV P protein and the sample of VSV RNP; Bertrand Lefèbvre, Denis Daveloos, and Bernard Brasmes (CRSSA, Grenoble) for mass spectrometry measurements; and Christine Ebel (IBS, Grenoble) for the use of the analytical centrifuge and discussion of the results.

## REFERENCES

- Fields, B. N., Knipe, D. M., Howley, P. M., Chanock, R. M., Melnik, T. P., Roizman, B., and Straus, S. E. (1996) *Fields' virology*, 3rd ed., Lippincott-Raven Publishers, New York.
- 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.
- 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.
- Arnheiter, H., Davis, N. L., Wertz, G., Schubert, M., and Lazzarini, R. A. (1985) Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis, *Cell* 41, 259–267.
- 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.
- Pattanaik, A. K., and Wertz, G. W. (1990) Replication and amplification of defective interfering particle RNAs of vesicular stomatitis virus in cells expressing viral proteins from vectors containing cloned cDNAs, *J. Virol.* 64, 2948–2957.
- Gupta, A. K., and Banerjee, A. K. (1997) Expression and purification of vesicular stomatitis virus N-P complex from *Escherichia coli*: role in genome RNA transcription and replication in vitro, *J. Virol.* 71, 4264–4271.
- Masters, P. S., and Banerjee, A. K. (1988) Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA, *J. Virol.* 62, 2658–2664.
- Banerjee, A. K. (1987) Transcription and replication of rhabdoviruses, *Microbiol. Rev.* 51, 66–87.
- Emerson, S. U., and Wagner, R. R. (1973) L protein requirement for in vitro RNA synthesis by vesicular stomatitis virus, *J. Virol.* 12, 1325–1335.
- Curran, J., Boeck, R., Lin-Marq, N., Lupas, A., and Kolakofsky, D. (1995) Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils, *Virology* 214, 139–149.
- 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.
- Barik, S., McLean, T., and Dupuy, L. C. (1995) Phosphorylation of Ser<sup>232</sup> directly regulates the transcriptional activity of the P protein of human respiratory syncytial virus: phosphorylation of Ser<sup>237</sup> may play an accessory role, *Virology* 213, 405–412.
- Asenjo, A., and Villanueva, N. (2000) Regulated but not constitutive human respiratory syncytial virus (HRSV) P protein phosphorylation is essential for oligomerization, *FEBS Lett.* 467, 279–284.
- Gao, Y., and Lenard, J. (1995) Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly, *J. Virol.* 69, 7718–7723.
- Raha, T., Samal, E., Majumdar, A., Basak, S., Chattopadhyay, D., and Chattopadhyay, D. J. (2000) N-terminal region of P protein of Chandipura virus is responsible for phosphorylation-mediated homodimerization, *Protein Eng.* 13, 437–444.
- Chen, M., Ogino, T., and Banerjee, A. K. (2006) Mapping and functional role of the self-association domain of vesicular stomatitis virus phosphoprotein, *J. Virol.* 80, 9511–9518.
- Barik, S., and Banerjee, A. K. (1992) Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P, *Proc. Natl. Acad. Sci. U.S.A.* 89, 6570–6574.
- Gao, Y., and Lenard, J. (1995) Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II, *EMBO J.* 14, 1240–1247.
- Barik, S., and Banerjee, A. K. (1992) Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation, *J. Virol.* 66, 1109–1118.
- Spadafora, D., Canter, D. M., Jackson, R. L., and Perrault, J. (1996) Constitutive phosphorylation of the vesicular stomatitis virus P protein modulates polymerase complex formation but is not essential for transcription or replication, *J. Virol.* 70, 4538–4548.
- Peluso, R. W., and Moyer, S. A. (1988) Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA, *Virology* 162, 369–376.
- 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.
- Das, T., Gupta, A. K., Sims, P. W., Gelfand, C. A., Jentoft, J. E., and Banerjee, A. K. (1995) Role of cellular casein kinase II in the function of the phosphoprotein (P) subunit of RNA polymerase of vesicular stomatitis virus, *J. Biol. Chem.* 270, 24100–24107.
- Gao, Y., Greenfield, N. J., Cleverley, D. Z., and Lenard, J. (1996) The transcriptional form of the phosphoprotein of vesicular stomatitis virus is a trimer: structure and stability, *Biochemistry* 35, 14569–14573.

26. Qanungo, K. R., Shaji, D., Mathur, M., and Banerjee, A. K. (2004) Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA, *Proc. Natl. Acad. Sci. U.S.A.* *101*, 5952–5957.
27. Gupta, A. K., Shaji, D., and Banerjee, A. K. (2003) Identification of a novel tripartite complex involved in replication of vesicular stomatitis virus genome RNA, *J. Virol.* *77*, 732–738.
28. Basak, S., Polley, S., Basu, M., Chattopadhyay, D., and Roy, S. (2004) Monomer and dimer of Chandipura virus unphosphorylated P-protein binds leader RNA differently: implications for viral RNA synthesis, *J. Mol. Biol.* *339*, 1089–1101.
29. Raha, T., Chattopadhyay, D., Chattopadhyay, D., and Roy, S. (1999) A phosphorylation-induced major structural change in the N-terminal domain of the P protein of Chandipura virus, *Biochemistry* *38*, 2110–2116.
30. Chattopadhyay, D., Raha, T., and Chattopadhyay, D. (1997) Single serine phosphorylation within the acidic domain of Chandipura virus P protein regulates the transcription in vitro, *Virology* *239*, 11–19.
31. Chattopadhyay, D., and Chattopadhyay, D. (1994) *Cell. Mol. Biol. Res.* *40*, 693–698.
32. 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.
33. 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.
34. Prehaud, C., Nel, K., and Bishop, D. H. (1992) Baculovirus-expressed rabies virus M1 protein is not phosphorylated: it forms multiple complexes with expressed rabies N protein, *Virology* *189*, 766–770.
35. Uversky, V. N. (1993) Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule, *Biochemistry* *32*, 13288–13298.
36. Wyatt, P. J. (1998) Submicrometer Particle Sizing by Multiangle Light Scattering following Fractionation, *J. Colloid Interface Sci.* *197*, 9–20.
37. Serdyuk, I. N., Zaccai, N. R., and Zaccai, J. (2007) *Methods in Molecular Biophysics: Structure, Dynamics, Function*, Cambridge University Press, Cambridge.
38. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling, *Biophys. J.* *78*, 1606–1619.
39. Svedberg, T. H., and Pedersen, K. O. (1940) *The Ultracentrifuge*, Oxford University Press, Oxford.
40. Jacrot, B., and Zaccai, G. (1981) Determination of molecular weight by neutron scattering, *Biopolymers* *20*, 2413–2426.
41. Glatter, O., and Kratky, O. (1982) *Small Angle X-ray Scattering*, Academic Press, London, U.K.
42. Semenyuk, A. V., and Svergun, D. (1991) GNOM - a program package for small-angle scattering data processing, *J. Appl. Crystallogr.* *24*, 537–540.
43. 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.
44. Tarbouriech, N., Curran, J., Ruigrok, R. W., and Burmeister, W. P. (2000) Tetrameric coiled coil domain of Sendai virus phosphoprotein, *Nat. Struct. Biol.* *7*, 777–781.
45. Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R. W., and Burmeister, W. P. (2000) On the domain structure and the polymerization state of the sendai virus P protein, *Virology* *266*, 99–109.
46. Bloomfield, V., Dalton, W. O., and Van Holde, K. E. (1967) Frictional coefficients of multisubunit structures. I. Theory, *Biopolymers* *5*, 135–148.
47. Damaschun, G., Damaschun, H., Gast, K., Misselwitz, R., Zirwer, D., Guhrs, K. H., Hartmann, M., Schlott, B., Triebel, H., and Behnke, D. (1993) Physical and conformational properties of staphylokinase in solution, *Biochim. Biophys. Acta* *1161*, 244–248.
48. Gast, K., Damaschun, H., Misselwitz, R., Muller-Frohne, M., Zirwer, D., and Damaschun, G. (1994) Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate, *Eur. Biophys. J.* *23*, 297–305.
49. Tanford, C. (1961) *Physical chemistry of macromolecules*, New York.
50. Santos, N. C., and Castanho, M. A. R. B. (1996) Teaching light scattering spectroscopy: The dimension and shape of tobacco mosaic virus, *Biophys. J.* *71*, 1641–1650.
51. Karlin, D., Ferron, F., Canard, B., and Longhi, S. (2003) Structural disorder and modular organization in Paramyxovirinae N and P, *J. Gen. Virol.* *84*, 3239–3252.
52. Bourhis, J. M., Canard, B., and Longhi, S. (2006) Structural disorder within the replicative complex of measles virus: functional implications, *Virology* *344*, 94–110.
53. Radivojac, P., Obradovic, Z., Smith, D. K., Zhu, G., Vucetic, S., Brown, C. J., Lawson, J. D., and Dunker, A. K. (2004) Protein flexibility and intrinsic disorder, *Protein Sci.* *13*, 71–80.
54. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol.* *6*, 197–208.
55. Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W., and Marion, D. (2004) Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution, *Virology* *319*, 201–211.
56. 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.
57. 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.
58. Das, S. C., and Pattnaik, A. K. (2004) Phosphorylation of vesicular stomatitis virus phosphoprotein P is indispensable for virus growth, *J. Virol.* *78*, 6420–6430.
59. Hwang, L. N., Englund, N., Das, T., Banerjee, A. K., and Pattnaik, A. K. (1999) Optimal replication activity of vesicular stomatitis virus RNA polymerase requires phosphorylation of a residue(s) at carboxy-terminal domain II of its accessory subunit, phosphoprotein P, *J. Virol.* *73*, 5613–5620.
60. Pattnaik, A. K., Hwang, L., Li, T., Englund, N., Mathur, M., Das, T., and Banerjee, A. K. (1997) Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication, *J. Virol.* *71*, 8167–8175.
61. Curran, J. (1996) Reexamination of the Sendai virus P protein domains required for RNA synthesis: a possible supplemental role for the P protein, *Virology* *221*, 130–140.

BI7007799

## C – Conclusion

C'est la première démonstration claire que la forme non phosphorylée de la phosphoprotéine de *Rhabdoviridae* est un dimère allongée en solution, que l'on pourrait assimiler à un cylindre:

- les analyses hydrodynamiques en utilisant la chromatographie d'exclusion de taille (SEC) et la diffusion dynamique de la lumière (DLS) indiquent que ces protéines sont asymétriques avec un rayon de Stokes de 4,8 et 5,3 nm, respectivement pour les protéines P VSV et P RV,

- l'analyse par diffusion de lumière (SEC-MALLS-RI) montre clairement que les phosphoprotéines se comportent en dimères en solution,

- la diffusion de neutrons aux petits angles (SANS) et l'ultracentrifugation analytique (AUC) ont confirmé l'état dimérique en solution. Le SANS nous donne un rayon de giration de 5,3 nm en accord avec la chromatographie d'exclusion de taille, mais surtout un rayon de giration de la cross-section de 1,6 – 1,8 nm, indiquant l'asymétrie de la molécule. L'AUC a donné un rapport des coefficients de frictions supérieur à 1,7, confirmant encore l'asymétrie de la molécule,

- la microscopie électronique montre que la protéine P VSV et le mutant P VSV S60D T62D forment des fibres irrégulières lorsqu'incubés à 30°C.

- pour la protéine P VSV, la phosphorylation ne semble pas nécessaire pour l'oligomérisation, puisque le mutant P VSV S60D/T62D se comporte en solution comme la protéine P VSV non phosphorylée.

La technique de SEC-MALLS-RI nous a permis de statuer clairement sur l'homo-dimérisation des formes non phosphorylées de la phosphoprotéine du virus de la rage (P RV) et de la stomatite vésiculaire (P VSV). La caractérisation du mutant P VSV S60D/T62D remet en cause les résultats obtenus concernant les activités de transcription et répliation des formes phosphorylées de la phosphoprotéine de VSV. Il semble que la phosphorylation, tout comme pour la phosphoprotéine du RV, ne soit nécessaire pour l'oligomérisation de la protéine. Aussi, il serait à présent intéressant d'investir à nouveau les rôles de la phosphorylation et de l'oligomérisation de la phosphoprotéine dans les activités de transcription et répliation.

## ***II – La phosphoprotéine du virus de la rage est modulaire***

### **A – Introduction**

Une des premières étapes chez le biochimiste et le biophysicien est d'obtenir des informations sur la protéine d'intérêt sur base de sa séquence en acides aminés (aa). Nous avons effectué dans un premier temps, une analyse bioinformatique détaillée de la séquence primaire des phosphoprotéines, de façon similaire au travail du Dr. Sonia LONGHI (Karlin & *al.*, 2002). Nous avons réalisés des prédictions de structures secondaires, des alignements de séquences en acides aminés de protéines homologues, et des prédictions de régions désordonnées Dans le papier présenté ci-après « *Modular organization of rabies virus phosphoprotein* », nous avons proposé un algorithme simple qui permet d'obtenir un consensus de prédiction. Nous avons utilisé la séquence de la phosphoprotéine du virus Sendai pour laquelle les données structurales sont disponibles, pour valider notre méthode.

Aussi, nous avons caractérisé biochimiquement et biophysiquement la phosphoprotéine du virus de la rage ainsi que des mutants pour mettre en évidence son organisation modulaire :

- la protéine P (**PRV**) (protéine entière, 297 résidus)
- **P<sub>ΔN52</sub>** (protéine tronquée de 52 aa N-terminaux)
- **P<sub>Δ91-131</sub>** (protéine tronquée du domaine central de la phosphoprotéine)
- **P<sub>ΔC120</sub>** (protéine tronquée de la région C-terminale, responsable de l'interaction avec le complexe nucléoprotéine-ARN)
- le domaine C-terminal **P<sub>CTD (186-297)</sub>** (aa 186-297), responsable de l'interaction avec N-ARN
- le domaine d'oligomérisation **P<sub>CED (90-135)</sub>** (aa 90-135), responsable de la dimérisation de la phosphoprotéine.

Nous avons poursuivi cette analyse par des expériences de double hybride, en collaboration avec le Dr. Danielle BLONDEL afin de caractériser les interactions phosphoprotéine-nucléoprotéine et phosphoprotéine-phosphoprotéine.

## B – Article 2: Modular organization of rabies virus phosphoprotein

Auteurs: **Francine C.A. GERARD**, Euripedes de Almeida Ribeiro Jr., Cédric Leyrat, Ivan Ivanov, Danielle Blondel, Sonia Longhi, Noël Tordo, Rob W. H. Ruigrok et Marc Jamin.

En préparation pour soumission à *Journal of Virology*.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

**MODULAR ORGANIZATION OF RABIES VIRUS  
PHOSPHOPROTEIN**

**Francine C. A. Gerard<sup>1\*</sup>, Euripedes de Almeida Ribeiro Jr.<sup>1\*</sup>, Cédric Leyrat<sup>1</sup>,  
Ivan Ivanov<sup>1</sup>, Danielle Blondel<sup>2</sup>, Sonia Longhi<sup>3</sup>, Rob W. H. Ruigrok<sup>1</sup> and Marc Jamin<sup>1\*</sup>**

<sup>1</sup>UJF-EMBL-CNRS UMR 5233 - Unit of Virus Host Cell Interactions,  
6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

<sup>2</sup>Laboratoire de Virologie Moléculaire et Structurale, UMR 2472 CNRS-1157 INRA  
Bât. 14B, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

<sup>3</sup>Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Universités  
Aix-Marseille I et II, Case 932, 163 Avenue de Luminy, 13288 Marseille Cedex 09, France.

\* These authors contributed equally to the work

**Running title:** Organization of rabies virus phosphoprotein

\* **Corresponding author:** Unit of Virus Host Cell Interactions (UVHCI), UMR 5233 UJF-  
EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France  
E-mail : [jamin@embl.fr](mailto:jamin@embl.fr), Phone: + 33 4 76 20 94 62, Fax : + 33 4 76 20 94 00

31 **Abstract**

32 A phosphoprotein (P) is found in all viruses of the *Mononegavirales* order. These proteins  
33 form homo-oligomers, fulfil similar roles in the replication cycles of the various viruses, but  
34 differ in their length and oligomerization state. Sequence alignments reveal no sequence  
35 similarities among proteins from viruses belonging to the same family, and sequence analyses  
36 suggest that phosphoproteins from viruses of the *Paramyxoviridae* contain structured domains  
37 alternating with intrinsically disordered regions. Here, we used predictions of disorder, of  
38 secondary structure and an analysis of sequence conservation to predict the domain  
39 organization of the phosphoprotein from rabies virus (RV). We devised simple voting  
40 procedures for combining the results from multiple prediction methods and locating the  
41 boundaries between disordered regions and structured domains. To validate the proposed  
42 modular organization and to confirm that the putative structured domains correspond to  
43 autonomous folding units, we have used two-hybrid and biochemical approaches to  
44 characterize the properties of several fragments of RV P. We found that both central and C-  
45 terminal domains can fold in isolation, that the central domain constitutes the oligomerization  
46 domain, and that the C-terminal domain binds to nucleocapsids. Our results suggest a  
47 conserved organization of P proteins in the *Rhabdoviridae* family in concatenated functional  
48 domains resembling that of the P proteins in the *Paramyxoviridae* family.

## 49 INTRODUCTION

50 The phosphoprotein (P) is an essential component of the replication machinery of  
51 viruses from the *Rhabdoviridae* family, which includes rabies virus (RV) and vesicular  
52 stomatitis virus (VSV). P has similar functions in other non-segmented negative strand RNA  
53 viruses belonging to the *Paramyxoviridae* (measles virus, Sendai virus, respiratory syncytial  
54 virus, Nipah virus), the *Filoviridae* (Ebola virus, Marburg virus) and the *Bornaviridae* (Borna  
55 disease virus), which are grouped together with the *Rhabdoviridae* in the order of the  
56 *Mononegavirales* (MNV). In all MNV viruses, P is part of a set of five common proteins  
57 encoded in the same order along the negative strand genome from the 3' extremity to the 5'  
58 extremity: the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the  
59 glycoprotein (G) and the RNA dependent RNA-polymerase (L). The nucleoprotein binds to  
60 the viral RNA forming long helical ribonucleoprotein complexes, which serve as templates  
61 for RNA transcription and replication. The P protein plays multiple roles at different stages of  
62 the viral life cycle, forming a complex with the nascent nucleoprotein that prevents the  
63 association of the latter with host-cell RNA and forming a two-subunit viral RNA polymerase  
64 complex with the L protein. Recent studies have shown that P proteins from viruses of the  
65 *Paramyxoviridae* contain long disordered regions that alternate with structured domains<sup>1; 2; 3;</sup>  
66<sup>4; 5</sup>.

67 Various structural and functional data are available for either VSV P or RV P that  
68 suggest the existence of common functional regions, and a previous analysis of the amino  
69 acid sequence suggested that *Rhabdoviridae* P proteins also contain disordered or flexible  
70 sequences<sup>1</sup>. Also, the P proteins of both viruses are highly sensitive to proteolysis; two  
71 domains for which the structure was solved by X-ray crystallography were identified by  
72 limited proteolysis<sup>6; 7; 8</sup>. The N-terminal fragments of both VSV P and RV P interact with N  
73 in the N<sup>0</sup>-P complex, which is sufficient for maintaining N (referred to as N<sup>0</sup>) in a soluble,



74 RNA-free form<sup>9; 10</sup>. A structured domain in the central part of VSV P is responsible for the  
75 dimerization of VSV P<sup>7</sup>, in agreement with our recent demonstration that recombinant, full-  
76 length P proteins from both VSV and RV form dimers in solution<sup>11</sup>. The oligomerization  
77 domain of RV P was also shown to be located in the central region (aa 52 to 189)<sup>12</sup>, although  
78 its precise location and boundaries were not clearly determined<sup>13</sup>. The C-terminal domains of  
79 RV and VSV P bind to N-RNA complexes, and their structures were solved by X-ray  
80 crystallography<sup>6</sup> and NMR<sup>14</sup>, respectively.

81 In infected cells, apart from the full-length protein, four N-terminally truncated forms  
82 of RV P (P2: aa 20-297, P3: aa 53-297, P4: aa 69-297 and P5: aa 83-297) are produced by a  
83 ribosomal scanning mechanism<sup>15</sup>. Full-length P and P2 are located in the cytoplasm, whereas  
84 P3, P4 and P5 are located in the nucleus. The subcellular localization of the different P forms  
85 is regulated by a CRM1-dependent nuclear export signal (NES#1) located between residues  
86 49 and 58 and a NES#2 in close association with the nuclear localization signal (NLS) that are  
87 both located within the globular C-terminal domain (CTD)<sup>16; 17</sup> (aa 211-214 and 260). The C-  
88 terminal region (aa 172-297) also contains binding sites for STAT1 and the promyelocytic  
89 leukaemia protein (PML), two proteins involved in the interferon-induced antiviral response  
90<sup>18; 19; 20</sup>. Finally, a region between residue 139 and 172 contains a binding site for the  
91 cytoplasmic dynein light chain (LC8) that could explain how RV nucleocapsids are  
92 transported in neuronal axons towards the cell body<sup>21; 22</sup>.

93 For many years since the first protein structures were determined<sup>23; 24; 25</sup>, biological  
94 activities of proteins have been associated with the folding of the polypeptide chain into a  
95 well-defined three-dimensional structure. This idea was supported by the discovery that the  
96 biological activities disappear when proteins unfold<sup>26; 27</sup>. However, this structure/function  
97 paradigm was recently challenged with the discovery that many proteins are fully disordered  
98 or contain disordered regions under physiological conditions<sup>28; 29; 30; 31</sup>. Also, several proteins

99 were found to exist in a disordered form in isolation but to fold into a well-defined structure  
100 upon binding to a partner<sup>32; 33; 34; 35; 36</sup>. More surprising was the realization that disorder was  
101 required for, or participated in the biological function of some proteins<sup>28; 31; 33; 37; 38; 39; 40; 41</sup>.  
102 Several approaches were developed for sorting entire proteins between disordered and  
103 structured proteins<sup>29; 38; 42</sup>. Application of such algorithms to whole genomes suggested that  
104 as much as one third of the proteins from eukaryotic organisms contain long disordered  
105 regions<sup>38; 42</sup>. This raised a new challenge in structural biology for identifying disordered  
106 proteins from the amino acid sequence and for locating disordered regions and folded  
107 domains within the sequence. Various algorithms were developed for predicting the location  
108 of disordered regions from the amino acid sequence (for recent reviews, see<sup>5; 43</sup>). Some of  
109 these algorithms rely on the relationship between the amino acid sequence and the disorder  
110 state of the polypeptide chain<sup>44; 45</sup> (SEG, FoldIndex, Globplot, HCA plot, NORSp). For  
111 example, it was early recognized that a protein region rich in charged residues and poor in  
112 hydrophobic residues is likely to be disordered<sup>29; 44</sup>. Other programs rely on trained neural  
113 networks and use sets of known disordered proteins (RONN, DisEMBL, Disopred2), or  
114 combine both approaches (PONDR). However, different methods may yield different  
115 predictions and it is difficult to choose the method appropriate to a given protein.

116 In this paper, we analyzed the amino acid sequence of the phosphoprotein from RV  
117 using algorithms available through WEB servers for predicting the location of disordered  
118 regions and of secondary structure elements, and we analyzed the amino acid sequence  
119 conservation. To avoid choosing among the prediction methods, we devised simple scoring  
120 procedures for integrating the predictions from the different algorithms and for yielding  
121 consensus locations of boundaries between ordered and disordered regions. As a test case, we  
122 applied this approach to the analysis of Sendai virus (SV) P for which the location of several  
123 structured domains and disordered regions are known from previous structural studies<sup>2; 3; 46</sup>.

124 The properties of the predicted structured domains of RV P were tested in a cellular context  
125 using a yeast two-hybrid approach but also *in vitro* by producing and characterizing  
126 recombinant protein fragments. Circular dichroism was used to determine the secondary  
127 structure content and size-exclusion chromatography combined with detection by Multi-  
128 Angle Laser Light Scattering was used to determine the molecular size, molecular mass and  
129 the stoichiometry of the different protein fragments. Native gel shift experiments were used to  
130 test the interaction of the various domains with the nucleoprotein-RNA complex, which  
131 represents the physiological partner. Altogether, these results provided by the *in vivo* and *in*  
132 *vitro* characterization of the various P fragments confirmed the *in silico* analysis and allowed  
133 us to propose a model for the modular organization of RV P.

134

135 **RESULTS**136 **Strategy for predicting the location of disordered and structured regions**

137 Using available programs we attempted to predict the boundaries between structured  
138 and unstructured regions in RV and SV P proteins. The latter was used as a test case since  
139 knowledge about the positions of some of its structured and unstructured domains are  
140 available<sup>2; 3; 47; 48</sup>. The location of disordered regions in the sequence of both P proteins was  
141 predicted from the amino acid sequence by using 16 different algorithms available through  
142 WEB servers (Figure S1 in the Supplementary material). A recently developed meta-server  
143 for disorder prediction, MEADOR<sup>49</sup>, yielded a similar picture (Figure S2 in the  
144 Supplementary material). The predictions showed significant differences in the number,  
145 position and length of the predicted disordered regions. The discrepancies between the results  
146 from different algorithms argue for the existence of different types (“flavors”) of disorder<sup>50; 51</sup>  
147 and raise the problem of choosing the best algorithm and/or strategy for locating the  
148 boundaries of disordered regions in a given protein<sup>43; 51</sup>.

149 Recently, we devised a simple voting procedure that combines multiple predictions of  
150 disordered regions and predicts the boundaries between structured and unstructured regions,  
151 and used it to delimit the Nucleoprotein-RNA (N-RNA) binding C-terminal domain of  
152 vesicular stomatitis virus (VSV) P protein<sup>14</sup>. Here, we improve this approach by including  
153 secondary structure predictions and amino acid sequence conservation, and use it to predict  
154 the modular organization of RV and SV P. This method requires no advised input from the  
155 experimentalist and integrates the results from multiple disorder predictions, as well as from  
156 multiple secondary structure predictions and from amino acid sequence conservation among  
157 homologous proteins. For disorder predictions and secondary structure predictions, we  
158 devised scoring methods based on simple binary voting procedures. For each disorder  
159 prediction, residues were simply ranked in a binary manner as ordered or disordered using the

160 default threshold set for each algorithm in the WEB servers. The statistical weights yielded by  
161 some of these algorithms were not accounted for, since they were not available for each  
162 predicting method and were difficult to compare. For each predicting algorithm, a score of 0  
163 was attributed to each residue predicted to be disordered, whereas a score of 1 was attributed  
164 to each one predicted to be in a structured region. A D-score was calculated by adding the  
165 values for each residue and dividing by the number of used algorithms. We arbitrarily defined  
166 two threshold levels at 0.75 and 0.50; residues with a D-score below 0.50 were assigned as  
167 disordered, while residues with a D-score above 0.75 were assigned as structured. The  
168 structural state of the intermediate zone, with a D-score comprised between 0.50 and 0.75,  
169 was considered as ambiguous. Residues with intermediate D-scores were found to occur at the  
170 boundaries between ordered and disordered regions, but they could also correspond to protein  
171 regions that are mainly stabilized by tertiary or intermolecular interactions.

172 A complementary approach for deciphering the modular organization of a protein  
173 consists in identifying structured domains by predicting the location of secondary structures  
174 <sup>43; 51</sup>. Numerous algorithms, available through WEB servers, predict the location of secondary  
175 structure elements on the basis of the amino acid sequence, and some algorithms combine  
176 multiple methods <sup>52; 53</sup>. Here again, we used a simple voting method for combining  
177 predictions from different algorithms and for defining consensus locations of predicted  
178 secondary structures (SS-score).

179 Finally, as mutation rates tend to be higher in disordered regions than in structured  
180 domains <sup>54</sup>, structured domains are expected to appear as conserved regions in multiple  
181 sequence alignments of homologous proteins. Exceptions could be found in disordered  
182 regions interacting with partners and hence presenting some sequence conservation. To score  
183 the conservation rate in multiple sequence alignments we used the score provided by the  
184 AL2CO software <sup>55</sup>, and we considered regions with a normalized score (C-score) > 0.5 as

185 conserved regions. Finally, we defined a global score (G-score), as a non-weighted, linear  
186 combination of the scores for disorder prediction (D-score), secondary structure prediction  
187 (SS-score) and sequence conservation (C-score). Again, we arbitrarily defined a threshold of  
188 0.5 in order to discriminate between disordered (G-score < 0.5) and structured regions (G-  
189 score > 0.5).

190

### 191 **Amino acid sequence analysis of SV P**

192 To assess the reliability of the above-described scoring methods, we analyzed the  
193 amino-acid sequence of the SV P protein. Figures 1A and 1F show the location of functional  
194 regions and the boundaries of the two domains for which the structure has been determined <sup>2</sup>;  
195 <sup>3</sup>; <sup>46</sup>, respectively. The structure of the central oligomerization domain (aa 320-429) was  
196 solved by X-ray crystallography. It is a tetramer in which each monomer is composed of a  
197 long C-terminal  $\alpha$ -helix (aa 364-429) that forms a tetrameric coiled-coil plus an N-terminal  
198 helical bundle of three short  $\alpha$ -helices <sup>3</sup>. The structure of the C-terminal N-RNA interacting  
199 domain (aa 516-568), solved by NMR spectroscopy, consists of a three-helix bundle <sup>2</sup>. NMR  
200 also showed that the structured C-terminal domain (aa 516-568) is linked to a flexible domain  
201 (aa 474-504) by an even more flexible linker (aa 505-515) <sup>2</sup>; <sup>46</sup>. There is no information on the  
202 structure of residues 1 to 319 nor on residues 430 to 473.

203 Although disorder predictions for SV P displayed significant discrepancies among the  
204 various algorithms (Figure S1A and S2A), the normalized D-score shown in Figure 1B  
205 revealed the existence of three main disordered regions (score  $\leq$  0.5) ranging from residue 1  
206 to residue 321 (with the exception of the 28-49 region), from residue 430 to residue 459 and  
207 from residue 494 to residue 520, which were thus considered as the consensus disordered  
208 regions (Figure 1C). The D-score clearly highlighted the location of the two structured  
209 domains, namely a central region spanning residue 322 to residue 429, which closely

210 corresponds to the tetramerization domain (aa 320-429), and a C-terminal region  
211 encompassing residues 521-568, which corresponds to the structured C-terminal domain (aa  
212 516-568).

213 Predictions of secondary structures ( $\alpha$ -helices and  $\beta$ -strands) supported these  
214 conclusions. Figure 1D shows the normalized SS-score for  $\alpha$ -helix/ $\beta$ -strand prediction and  
215 Figure 1E shows the consensus locations of these secondary structure elements. Within the  
216 known structured domains<sup>2,3</sup> (Figure 1F), secondary structures were correctly predicted, with  
217 the exception of the first  $\alpha$ -helix in the C-terminal domain that was predicted as a  $\beta$ -strand.  
218 This C-terminal domain is rather flexible in isolation and is stabilized upon binding to N<sup>56</sup>. In  
219 particular, residues in the first helix undergo proton exchange and the helix is in  
220 conformational exchange with the unfolded form<sup>48</sup>. The long  $\alpha$ -helix predicted in the central  
221 domain was also predicted to form a coiled-coil (aa 375-420) by the analysis of the HCA plot  
222 (see Figure 2 in<sup>57</sup>), as well as by various coiled-coil predictors (see Material and Methods  
223 section), although with a level of confidence generally lower than that set by default in the  
224 WEB servers (data not shown). The coiled coil of SV P exhibits a different packing pattern  
225 and is more hydrophilic than typical coiled-coils<sup>3</sup>, which may explain the failure of these  
226 algorithms to clearly predict this structure from the sequence.

227 A multiple sequence alignment was obtained for P proteins from four members of the  
228 *Respirovirus* genus in the *Paramyxoviridae* family. The normalized C-score shown in Figure  
229 1G indicated several short conserved regions in the N-terminal half of the protein (aa 27-42,  
230 97-120, 162-174, 186-205, 211-230), and three long and more conserved regions in the C-  
231 terminal half (aa 326-381, 399-481, 500-564) (Figure 1H). The three conserved regions in the  
232 C-terminal half of P cover most of the tetramerization domain and the entire C-terminal  
233 domain involved in binding to N-RNA complexes. However, part of the tetramerization

234 domain is not conserved (aa 382-398) and two regions (aa 430-459 and 500-520) predicted to  
235 be disordered were found to be conserved.

236 Finally, the global score (G-score) shows that the C-terminal half of the protein is  
237 predicted to be more structured than the N-terminal half (Figures 1I and 1J). The two regions  
238 with the highest G-score in the C-terminal half (aa 323-433, 520-566) match the domains of  
239 known three-dimensional structure. In addition, our predictions revealed the presence of three  
240 regions in the N-terminal half (aa 26-36, 105-109 and 208-226) and one region in the C-  
241 terminal half of SV P (aa 458-488) with a G-score  $> 0.5$  suggesting that they are structured.  
242 Two of these regions (aa 26-36 and 458-488) have intermediate D-score, are predicted to  
243 contain secondary structure elements ( $|\text{SS-score}| > 0.5$ ) and are conserved among homologous  
244 proteins (C-score  $> 0.5$ ). The former corresponds to the binding site of N in the N<sup>0</sup>-P complex  
245 while the latter partially overlaps the flexible region (aa 474-504) identified in the NMR  
246 experiments <sup>2</sup>. The two other N-terminal regions (aa 105-109 and 206-226) have D-score  $<$   
247  $0.5$  but that clearly stand out of the background (Figure 1B), are predicted to form secondary  
248 structure elements ( $|\text{SS-score}| > 0.5$ ) and are conserved among homologous proteins (C-score  
249  $> 0.5$ ). Up to now there is no experimental evidence for the presence of structured domains in  
250 these regions. The sequences are short and could correspond to flexible regions in isolation,  
251 which may fold upon interaction with viral or cellular partners.

252 In conclusion, for SV P, our scoring method predicts the location of known boundaries  
253 between disordered and structured regions with an accuracy of a few residues and indicates  
254 the possible existence of other structured regions.

255

## 256 **Amino acid sequence analysis of RV P**

257 A similar analysis was carried out with the amino acid sequence of RV P. Figure 2A  
258 and 2F show the location of functional regions along the sequence of P <sup>12; 13; 21; 22; 58; 59</sup> and the



259 location of the C-terminal N-RNA binding domain whose structure was determined by X-ray  
260 crystallography <sup>6</sup>, respectively. Calculation of the D-score (Figure 2B) revealed the presence  
261 in RV P of two disordered regions (aa 53-86 and 133-194) and three structured regions (aa: 1-  
262 29, 90-130 and 200-297 with the exception of 241-252). Figure 2C shows the location of  
263 consensus disordered regions. Figure 2D shows the SS-score for  $\alpha$ -helix/ $\beta$ -strand prediction,  
264 and Figure 2E shows the consensus location of these secondary structure elements. Most of  
265 the secondary structures experimentally observed in the three-dimensional structure of the C-  
266 terminal domain of RV P (Figure 2F) <sup>6</sup> were correctly predicted, although one  $\beta$ -strand and  
267 three  $\alpha$ -helices were not predicted. A coiled-coil was predicted by four algorithms as well as  
268 by analysis of the HCA plot with a consensus location between residues 9 and 42, but all  
269 these predictions had a level of confidence below the threshold value used in the various  
270 WEB servers and are, therefore, not indicated in Figure 2. Within this region, which binds to  
271 the nucleoprotein in the N<sup>0</sup>-P complex and to L <sup>9; 10; 59</sup>, one  $\beta$ -strand and two  $\alpha$ -helices were  
272 predicted.

273 A multiple alignment of the amino acid sequences of P proteins from 13 members of  
274 the *Lyssavirus* genus revealed an overall 25 % identity and 51 % similarity. The normalized  
275 C-score for sequence conservation calculated from the multiple sequence alignment is shown  
276 in Figure 2G, and boundaries for consensus conserved regions are shown in Figure 2H. In the  
277 sequence of RV P, four conserved regions (C-score > 0.5: aa 1-50, 101-126, 194-244 and  
278 249-292) are located within or corresponding to the otherwise predicted structured regions  
279 (D-score > 0.5, Figure 2A).

280 Finally, the G-score calculated for RV P points out the presence of three structured  
281 domains (aa 1-55, 92-131 and 195-293) separated by two disordered regions (aa 56-91 and  
282 132-194) (Figures 2I and 2J). The predicted N-terminal domain (aa 1-55) encompasses the  
283 binding site of N<sup>0</sup> in the N<sup>0</sup>-P complex (aa 4-40) <sup>9</sup>, whereas the predicted C-terminal

284 structured domain (aa 195-293) corresponds closely to the known N-RNA binding domain (aa  
285 186-297). A small region (aa 243-251) in the middle of the C-terminal domain is predicted  
286 disordered by the different scores. It corresponds to the end of helix  $\alpha_2$ , the entire helix  $\alpha_3$   
287 and the loop connecting the two helices, a part of the C-terminal domain of RV P that is  
288 absent in the homologous C-terminal domain of VSV P. The predicted central domain (aa 91-  
289 132) could correspond to the dimerization domain since it is located within the region  
290 necessary for oligomerization that was previously identified by deletion (aa 52-189)<sup>12</sup>.

291

### 292 **Yeast two-hybrid experiments**

293 In a yeast two-hybrid experiment, P protein associated with itself and with N (Figure  
294 3A). Deletion mutants were used to assess whether the predicted central and C-terminal  
295 domains of RV P were responsible for the dimerization of P<sup>11</sup> or for the interaction with N  
296 (Figure 3A). Truncation of the first 172 N-terminal residues (P <sub>$\Delta$ N172</sub>) or of the central region  
297 ranging from residue 91 to residue 131 (P <sub>$\Delta$ 91-131</sub>) abolished the association of P with itself,  
298 whereas truncation of the 52 N-terminal residues (aa 1-52 - P <sub>$\Delta$ N52</sub>) or of the 125 C-terminal  
299 residues (aa 172-297 - P <sub>$\Delta$ C125</sub>) had no effect on self-association (Figure 3A). The predicted  
300 central domain (aa 92-131 according to the G-score) thus appears critical for the dimerization  
301 of RV P as in VSV P<sup>7</sup>. To characterize the dimerization interface, the central region was  
302 further dissected. The deletion mutant P <sub>$\Delta$ 91-105</sub> still associated with P, whereas the mutants  
303 P <sub>$\Delta$ 106-113</sub>, P <sub>$\Delta$ 114-121</sub> and P <sub>$\Delta$ 122-131</sub> showed no association with P, indicating that residues 106 to  
304 131 are required for dimerization (data not shown). In addition, single amino acid mutations  
305 were introduced in the central region of P and were tested by using the two-hybrid approach.  
306 Five hydrophobic residues were mutated into alanine (F114, W118, I125, Y128 and V129).  
307 Although none of these mutations affected the binding to N, they reduced significantly the  
308 association with P (Figure 3B).

309 Previous data showed that the C-terminal domain of RV P (aa 173-297) binds to N-  
310 RNA complexes<sup>6;60</sup>, while the first 40 residues associate with N in the N<sup>0</sup>-P complex<sup>9</sup>. In the  
311 yeast two-hybrid experiment, truncation of the 125 C-terminal residues of RV P reduced but  
312 did not abolish the interactions with N (Figure 3A). The persisting interaction might reflect  
313 the interaction of the N-terminal domain of P with N in N<sup>0</sup>-P complexes. Truncation of the  
314 172 N-terminal residues or of the central domain (91-131) had no effect on the interactions  
315 with N<sup>58;61</sup>.

316

### 317 **Characterization of recombinant proteins**

318 The full-length P protein and different fragments (P<sub>NTD</sub>, P<sub>CED</sub>, P<sub>CTD</sub>, P<sub>ΔC120</sub>, P<sub>ΔN52</sub> and  
319 P<sub>Δ91-131</sub>) (Figure 4) were prepared as recombinant proteins with a C-terminal His<sub>6</sub>-tag and  
320 purified in a two-step procedure using Ni<sup>2+</sup> affinity and size exclusion chromatography. P<sub>ΔN52</sub>  
321 and P<sub>ΔC120</sub> poorly expressed and could only be obtained in small amounts.

322 The far-UV CD spectrum of full length P, P<sub>CED</sub>, P<sub>CTD</sub> and P<sub>Δ91-131</sub> exhibited two  
323 dichroic bands near 208 and 222 nm indicating the presence of  $\alpha$ -helical structure, while the  
324 spectrum obtained with P<sub>NTD</sub> showed one dichroic band near 222 nm and one below 200 nm  
325 suggesting a low  $\alpha$ -helical content (Figure 5). The proper folding of the central domain, P<sub>CED</sub>,  
326 was confirmed by 1D NMR spectroscopy (data not shown). Because  $\beta$ -strand and random coil  
327 contribute little to the CD signal at 222 nm, the average amount of  $\alpha$ -helix was grossly  
328 estimated from the molar ellipticity value at 222 nm<sup>62</sup>. The values for P<sub>CED</sub> (-21,100  
329 deg.cm<sup>2</sup>.dmol<sup>-1</sup> – 28 aa in  $\alpha$ -helix), P<sub>CTD</sub> (-14,000 deg.cm<sup>2</sup>.dmol<sup>-1</sup> – 45 aa in  $\alpha$ -helix), P<sub>Δ91-131</sub>  
330 (-9,200 deg.cm<sup>2</sup>.dmol<sup>-1</sup> – 67 aa in  $\alpha$ -helix), and full-length P (-11,100 deg.cm<sup>2</sup>.dmol<sup>-1</sup> – 94 aa  
331 in  $\alpha$ -helix) corresponded closely to the fraction of  $\alpha$ -helical structure predicted from the  
332 amino acid sequence of these proteins (P<sub>CED</sub>: 19 aa, P<sub>CTD</sub>: 43 aa, P<sub>Δ91-131</sub>: 73 and P: 92 aa)

333 (Figure 5). The value for  $P_{\text{NTD}}$  ( $-3,600 \text{ deg.cm}^2.\text{dmol}^{-1} -7 \text{ aa}$  in  $\alpha$ -helix) indicated a lower  $\alpha$ -  
334 helical content than that suggested by the secondary structure predictions (Figure 2). The  
335 number of residues in  $\alpha$ -helical conformation in  $P_{\text{CED}}$  and in  $P_{\Delta 91-131}$  adds up to that calculated  
336 for full-length P (28 aa + 67 aa = 95 aa), supporting the hypothesis that  $P_{\text{CED}}$  and  $P_{\text{CTD}}$  are  
337 autonomous folding units. However, the number of  $\alpha$ -helical residues in  $P_{\text{CTD}}$  calculated from  
338 the CD spectrum (45 aa) was significantly lower than that derived from the crystal structure  
339 (67 aa), and summing  $\alpha$ -helical residues in  $P_{\text{NTD}}$ ,  $P_{\text{CED}}$  and  $P_{\text{CTD}}$  yields a number (7 aa + 28 aa  
340 + 45 aa = 80 aa) significantly lower than that measured for full-length protein (94 aa). This  
341 discrepancy could be accounted for by the presence of additional structure in a region of P  
342 other than  $P_{\text{NTD}}$ ,  $P_{\text{CED}}$  or  $P_{\text{CTD}}$ . It is also possible that the predicted  $\alpha$ -helices in the N-terminal  
343 domain only fold in the context of a structure that is larger than the tested fragment from  
344 residue 1 to 68.

345 Analysis by size exclusion chromatography (SEC) using on-line detection by Multi-  
346 Angle Laser Light Scattering (MALLS) and refractometry (RI) indicated that full length P,  
347  $P_{\text{CED}}$ ,  $P_{\text{CTD}}$  and  $P_{\Delta 91-131}$  were monodisperse.  $P_{\text{NTD}}$ ,  $P_{\text{CTD}}$  and  $P_{\Delta 91-131}$  were monomeric with  
348 weight-averaged molecular masses of  $7.6 \pm 0.5 \text{ kDa}$ ,  $15.6 \pm 0.2 \text{ kDa}$  and  $32.5 \pm 0.3 \text{ kDa}$   
349 (Figure 6A), respectively, in agreement with the expected masses calculated from the amino  
350 acid sequence ( $P_{\text{NTD}} = 8696 \text{ Da}$ ,  $P_{\text{CTD}} = 13,989 \text{ Da}$ ;  $P_{\Delta 91-131} = 29,306 \text{ Da}$ ).  $P_{\text{CED}}$  and P form  
351 dimers with average molecular masses of  $17.4 \pm 0.2 \text{ kDa}$  (calculated molecular mass = 6,501  
352 Da ) and  $69 \pm 2 \text{ kDa}$  (calculated molecular mass of 33,616 Da), respectively<sup>11</sup>. The  
353 hydrodynamic radii (Stokes' radii,  $R_s$ ) of full length P,  $P_{\text{NTD}}$ ,  $P_{\text{CED}}$ ,  $P_{\text{CTD}}$ ,  $P_{\Delta C120}$  and  $P_{D91-131}$   
354 were determined by calibrating the SEC column with standard proteins<sup>63</sup>. On a plot of the  
355 hydrodynamic radius as a function of molecular mass (Figure 6B),  $P_{\text{CED}}$  and  $P_{\text{CTD}}$  appeared as  
356 native globular proteins,  $P_{\Delta 91-131}$ ,  $P_{\Delta C120}$  and full-length P were intermediate between native  
357 and unfolded proteins, indicating that these proteins are non-globular, and  $P_{\text{NTD}}$  appeared as

358 an unfolded protein. A previous analysis by Small-Angle Neutron Scattering revealed that  
359 full-length P forms elongated dimeric molecules <sup>11</sup>. These experiments confirmed that the  
360 predicted central domain of P constitutes the dimerization domain and, together with the CD  
361 experiments, that P<sub>CED</sub> folds as an independent domain.

362

### 363 **Binding to recombinant N-RNA rings**

364 The binding of full length P and its deletion mutants to a recombinant circular N-RNA  
365 complex containing 10 nucleoprotein protomers (N<sub>10</sub>-RNA) <sup>64</sup> was tested by native gel  
366 electrophoresis (Figure 7). Lane 1 shows the migration of N<sub>10</sub>-RNA, with a contaminating  
367 band of N<sub>9</sub>-RNA. In the presence of an excess of full-length P, a new band appeared  
368 corresponding to the complex. The band corresponding to N<sub>10</sub>-RNA alone disappeared, while  
369 the excess of P appeared at the bottom of the gel. When N<sub>10</sub>-RNA was incubated with P<sub>Δ91-131</sub>  
370 or with P<sub>CTD</sub>, similar shifts were observed for the complex indicating that, like full-length P,  
371 both protein fragments bound to N<sub>10</sub>-RNA (Lanes 4 and 6). Conversely, when N<sub>10</sub>-RNA was  
372 incubated with P<sub>ΔC120</sub> or P<sub>CED</sub>, the migration of N<sub>10</sub>-RNA was not affected, indicating no  
373 association (Lanes 5 and 7). These results demonstrate that binding of P to N-RNA complexes  
374 only requires the presence of the C-terminal domain and is independent of the presence of  
375 P<sub>CED</sub>, and thus, from the dimerization of P.

376

## 377 **DISCUSSION**

### 378 **Assessing the scoring procedure**

379           Predicting the location of disordered regions in a protein from its amino acid sequence  
380 is a major challenge in structural biology. Many proteins contain disordered regions which  
381 prevent structure determination by X-ray crystallography and which are often found to  
382 participate in their biological activity. Here, we have devised a simple method for combining  
383 different disorder predictions, which also takes into account secondary structure predictions  
384 and sequence conservation for defining consensus boundaries between structured and  
385 disordered regions. Such a voting approach integrates information from a large number of  
386 algorithms and avoids delicate choices between prediction methods. However, it is simplistic  
387 and accounts neither for the specificity of the different algorithms nor for the confidence level  
388 associated with each amino acid by some algorithms and it may underscore some peculiar  
389 types of disorder predicted by some algorithms. Our selection of algorithms is not exhaustive,  
390 and the threshold level at 0.5 for defining disorder was chosen arbitrarily. However, despite  
391 possible objections, this procedure provides a consensus prediction for the location of  
392 disordered regions in the phosphoproteins of SV and RV that are in agreement with many  
393 experimental data. Among the different algorithms used for predicting disordered regions,  
394 GlobPlot, PreLink and the three DisEMBL predictors yielded the most divergent results  
395 (Figure S1). However, removal of these predictions from the score calculation had little effect  
396 on the overall location of disordered regions and modified the location of boundaries only by  
397 a few residues (data not shown).

398

### 399 **A common modular organization within *Rhadoviridae* and *Paramyxoviridae* P**

400           All *MNV* viruses encode a phosphoprotein that plays similar roles in the replication  
401 cycle, but no similarity is found between P sequences from different *MNV* families or even

402 between P sequences within members of the *Rhabdoviridae*. In the *Paramyxoviridae*, the  
403 three-dimensional structure of the C-terminal domain of P is conserved between Sendai and  
404 measles viruses <sup>2; 65</sup>, and the sequence of the oligomerization domain is well-conserved,  
405 suggesting a common homotetrameric coiled-coil <sup>3; 66</sup>. The structure of the C-terminal domain  
406 of P of Sendai and measles viruses, however, is different from the structure of the C-terminal  
407 domains of RV P <sup>6</sup> and VSV P <sup>14</sup>, and the structure of the central dimerization domain of VSV  
408 P, an  $\alpha/\beta$  domain <sup>7</sup>, is different from the tetrameric coiled-coil domain of Sendai virus <sup>3</sup>.

409         Despite this lack of similarity, the available structural and functional information  
410 combined with a computational analysis of the amino acid sequence of RV P suggest a  
411 modular organization similar to that recently proposed for P proteins from members of the  
412 *Paramyxoviridae* <sup>1</sup>, with three structured domains alternating with two disordered regions  
413 (Figure 8). (1) An N-terminal domain with a potential to form  $\alpha$ -helices. The 55 N-terminal  
414 amino acids of RV P constitute an N-terminal hydrophobic region with  $\alpha$ -helical potential,  
415 although the far-UV circular dichroism spectrum obtained with the P<sub>NTD</sub> fragment of RV P  
416 (residues 1 to 68) and the hydrodynamic radius revealed a disordered domain with a low  $\alpha$ -  
417 helical content in solution. This discrepancy between predictions and experiments might  
418 reflect a disordered state in isolation, with a stable structure being induced upon binding to a  
419 physiological partner. In this regard, it is noteworthy that in RV P, the N-terminal domain  
420 contains a binding site for the L protein (aa 1-19) <sup>59</sup>, and for N<sup>0</sup> <sup>9; 10</sup>. In P from SV, a small N-  
421 terminal domain (aa 1-60), is also predicted to adopt an  $\alpha$ -helical conformation, and to be  
422 involved in binding to N<sup>0</sup> <sup>1; 67</sup> supporting a conserved role of this region of the protein in both  
423 MNV families. The N-terminal region was previously proposed to constitute the  
424 oligomerization domain of RV P, but the first 52 residues were found to be dispensable for  
425 dimerization <sup>13</sup>. Here, we confirm that deletion of the first 52 residues of RV P has no effect  
426 on the ability of P to self-associate, thus ruling out a role of this region in P dimerization. (2)

427 *An N-terminal disordered region.* In RV P, the 56-91 region is predicted to be disordered  
428 (Figure 2). A large part of this region (from aa 61) is dispensable for transcription <sup>12</sup>, thus  
429 highlighting another similarity with the *Paramyxoviridae*, in which the corresponding  
430 disordered region (aa 42-319, Figure 1A) is of variable length and dispensable for both  
431 transcription and replication <sup>68</sup>. (3) *A central oligomerization domain.* Recently, we have  
432 shown that RV P and VSV P form dimers in solution <sup>11</sup> and here we demonstrate that the  
433 central domain of RV P is responsible for the dimerization as it is in VSV P. The central  
434 domain of VSV P (aa 107-170) is formed by a kinked  $\alpha$ -helix flanked on each side by two-  
435 stranded  $\beta$ -sheets and assembles into dimers through the packing of the  $\alpha$ -helices thereby  
436 forming two domain-swapped  $\beta$ -sheets <sup>7</sup>. The dimerization domain in RV P is shorter than in  
437 VSV P, although part of this domain could be stabilized by intermolecular interactions and  
438 hence could escape our prediction method. Secondary structure predictions suggest a different  
439 structural organization in RV P with respect to VSV P with three short  $\alpha$ -helices and one  $\beta$ -  
440 strand in each monomer (Figure 2). Experiments with a minireplicon system revealed that the  
441 central region of RV P (aa 61-175) and thus the dimerization of P is dispensable for  
442 transcription and replication <sup>12</sup>, precluding a mechanism where P dimers cartwheel along the  
443 nucleocapsid <sup>68</sup>. In *Paramyxoviridae*, the central folded domain forms tetramers through  
444 coiled-coils <sup>3</sup>, and the tetramerization domain is required for both transcription and replication  
445 <sup>68</sup>. (4) *A C-terminal disordered region.* A second disordered region is predicted in the C-  
446 terminal part of RV P. This region contains the binding site for the cytoplasmic dynein light  
447 chain (LC8) <sup>21; 22</sup>. In P proteins from the *Paramyxoviridae*, this region is considered as a  
448 flexible linker <sup>1; 2</sup>. (5) *A C-terminal N-RNA binding domain.* This domain binds to N-RNA  
449 complexes (<sup>6</sup> and this study), but also to proteins from the interferon-induced cellular  
450 response, including the STAT1 factor <sup>18; 20</sup> and the promyelocytic leukemia protein <sup>19</sup>. The  
451 structure of the C-terminal domain of RV P (aa 186-297), which was solved by X-ray



452 crystallography <sup>6</sup>, is conserved among *Rhabdoviridae*, as shown recently by comparison with  
453 the structure of the C-terminal domain of VSV P <sup>14</sup>. A short structured C-terminal domain was  
454 also found in Sendai and measles viruses that bound to the N-RNA complex <sup>56; 65</sup>.

455 In the modular organization proposed for the P protein from *Rhabdoviridae* and  
456 *Paramyxoviridae* <sup>1</sup> (Figure 8), the structured domains share similar functions, although the  
457 structures of some of the domains are different. The presence of disordered regions, however,  
458 seems well conserved and could provide the flexibility required for proper function of the  
459 proteins, allowing for example the interaction between different partners within the  
460 replication complex. The presence of disordered regions and the function of the structured  
461 domains appear to be conserved during evolution.

## 462 MATERIALS AND METHODS

### 463 Amino acid sequence analysis

464 The location of disordered regions within SV and RV P was predicted by submitting  
465 the amino acid sequence to 16 different algorithms accessible through WEB servers.  
466 Foldindex was run with a window size of 51<sup>44</sup>. With the different PONDR predictors (XL3,  
467 VL-XT, XL1-XT and VSL1) residues with a score higher than 0.5 were considered as  
468 disordered<sup>69; 70</sup>. The NORSp server was used with a window size for disorder set to 25  
469 residues, a structure content cut-off of 12% and with the number of consecutive exposed  
470 residues set to 10<sup>71</sup>. The IUPred predictor for long disordered regions was used and residues  
471 with a score above 0.5 were considered as disordered<sup>72</sup>. The DISOPRED server was run with  
472 the rate threshold for predictions of false positive set at 2%<sup>42</sup>. The DisEMBL server  
473 (Loops/Coils, Hot-loops and Remark465)<sup>73</sup>, RONN server<sup>74</sup> and PreLink server<sup>75</sup> were run  
474 using the default parameters. The SEG program was used for analyzing sequence complexity  
475<sup>76</sup>. HCA plots were drawn with DRAWHCA<sup>77</sup> and disordered regions of more than 20  
476 residues were identified by visual inspection. A simple scoring procedure was used to define  
477 consensus disordered regions. For each prediction, residues predicted to be part of a  
478 disordered region were assigned a score of 0 while other residues were assigned a score of 1.  
479 A D-score for each amino-acid was calculated by adding the values from all predictions and  
480 normalizing the sum. Consensus disordered regions were defined as regions with a  
481 normalized score  $\leq 0.50$ , and structured domains were defined as regions with a normalized  
482 score  $\geq 0.75$ .

483 Similar scoring procedures were used for analyzing predictions of secondary  
484 structures. Secondary structure predictions were obtained from 15 different algorithms freely  
485 available through WEB servers: DPM<sup>78</sup>, DSC<sup>79</sup>, GOR IV<sup>80</sup>, HHNC<sup>81</sup>, PHD<sup>82</sup>, PREDATOR  
486<sup>83</sup>, SIMPA96<sup>84</sup>, SOPM<sup>78</sup>, SOPMA<sup>78</sup>, NNIPREDICT<sup>85</sup>, PORTER<sup>86</sup>, JPRED<sup>87</sup>, PSIPRED<sup>88</sup>,

487 SS-PRO<sup>89</sup> and PROF<sup>90</sup>. All algorithms were run using the default parameters. For each  
 488 prediction, residues predicted to be in  $\alpha$ -helix conformation were assigned a score of 1 while  
 489 residues predicted to be in extended conformation were assigned a score of -1. A SS-score  
 490 was calculated for each amino acid by adding values from each prediction and normalizing  
 491 the sum.  $\alpha$ -helices were defined as regions with a SS-score  $> 0.5$  and  $\beta$ -strands as regions  
 492 with a SS-score  $< -0.5$ . Predicted locations of coiled-coils were obtained from 4 different  
 493 WEB servers: COILS<sup>91</sup>, Paircoil2<sup>92</sup>, Multicoil<sup>93</sup> and Marcoil<sup>94</sup>.

494 Multiple sequence alignments were obtained for P sequences from 4 different  
 495 *Respiroviruses*: Sendai virus (BAC79134), human parainfluenza virus 1 (NP\_604435), human  
 496 parainfluenza virus 3 (BAA00031), bovine parainfluenza virus 1 (NP\_037642), and from 13  
 497 different *Lyssaviruses*: rabies virus CVS-11 strain (P22363), rabies virus Pasteur strain  
 498 (P06747), Australian bat lyssavirus (Q9QSP3), Australian bat lyssavirus (Q8JTH2), European  
 499 bat lyssavirus 1 (A4UHP9), European bat lyssavirus 2 (A4UHQ4), Duvenhage virus  
 500 (O56774), Lagos bat virus (O56773), Mokola virus (P0C569), Aravan virus (Q6X1D7), Irkut  
 501 virus (Q5VKP5), Khujand virus (Q6X1D3), West Caucasian bat virus (Q5VKP1). The  
 502 multiple alignments were performed with ClustalW using the BLOSUM-62 matrix, a gap  
 503 opening penalty of 10 and a gap extension penalty of 0.2<sup>95</sup>. The C-score characterizing the  
 504 sequence conservation was obtained for each residue with the AL2CO software<sup>55</sup> using the  
 505 Henikoff-Henikoff frequency estimation method, a sum of pairs conservation measure, an  
 506 averaging window of 20 residues and a gap fraction of 0.5. The C-score corresponding to the  
 507 standard deviation from the mean was normalized into values ranging from 0 to 1 according  
 508 to:

$$509 \text{ for } Z\text{-score} \geq 0, \quad 1 - 0.5 \cdot \exp(-Z \text{ score}) \quad (\text{Eq. 1})$$

$$510 \text{ for } Z\text{-score} < 0, \quad 0.5 \cdot \exp(Z \text{ score}) \quad (\text{Eq. 2})$$

511 A global G-score for locating boundaries between disordered and structured regions  
 512 was calculated as a linear combination of the D-score for disorder prediction, the absolute  
 513 value of the SS-score for secondary structure prediction and the C-score for sequence  
 514 conservation according to:

$$515 \quad G\text{-score} = \frac{D\text{-Score} + |SS\text{-Score}| + C\text{-score}}{3} \quad (\text{Eq. 3})$$

516

### 517 **Plasmid construction**

518 The cDNAs encoding full-length P, P deleted of the first N-terminal 52 residues  
 519 ( $P_{\Delta N52}$ ), P deleted of the first N-terminal 172 residues ( $P_{\Delta N172}$ ), and P deleted of the last C-  
 520 terminal 125 residues ( $P_{\Delta C125}$ ) were fused to the sequence encoding the DNA binding domain  
 521 (BD) of Lex A and cloned into pLex (Clontech) as described previously<sup>21</sup>. The constructs  
 522 pLex encoding deleted P proteins ( $P_{\Delta 91-105}$ ), ( $P_{\Delta 106-113}$ ), ( $P_{\Delta 114-121}$ ), ( $P_{\Delta 122-131}$ ) and ( $P_{\Delta 91-131}$ )  
 523 were created by deletion in the full-length P gene by using PCR gene fusion. Amino-acid  
 524 substitutions in pLex P ( $P_{F114A}$ ,  $P_{W118A}$ ,  $P_{I125A}$ ,  $P_{Y128A}$  and  $P_{V129A}$ ) were generated by a two-  
 525 step PCR-based site-directed mutagenesis approach as described previously<sup>16</sup>. The plasmid  
 526 pGAD-P was obtained by fusing the full-length P gene with the GAL4 activation domain  
 527 (AD) of pGAD (Clontech). The plasmid pGAD-N was described previously<sup>21</sup>.

528 Plasmids for expressing full length RV P (CVS strain) and several fragments ( $P_{NTD}$  (aa  
 529 1-68),  $P_{CED}$  (aa 91-134),  $P_{CTD}$  (aa 186-297),  $P_{\Delta 91-131}$ ,  $P_{\Delta N52}$  and  $P_{\Delta C120}$ ) in bacteria were  
 530 constructed by cloning the gene or the corresponding fragment into pET22b(+) using PCR.  
 531 For each construct, a His6-tag and a two amino-acid linker (Glu-Leu) were introduced at the  
 532 C-terminal extremity. All plasmids were verified by standard dideoxy sequencing.

533

534

**535 Two-hybrid system**

536 Yeast L40 cells were co-transformed with the plasmids encoding wild-type P or N  
537 fused to the AD of GAL4 and the P mutants fused to the DB of Lex A. The interaction was  
538 assayed by the appearance of blue colonies in the presence of X-Gal. An X-Gal mixture  
539 containing 0.5% agar, 0.1% SDS, 6% dimethylformamide and 0.04% X-Gal (5-bromo-4-  
540 chloro-3-indolyl- $\beta$ -galactosidase) was overlaid on freshly transformed cells grown on Trp<sup>-</sup>  
541 Leu<sup>-</sup> dishes and blue clones were detected after 60 min to 18 hours at 30°C.

542

**543 Production and purification of RV P and its different fragments**

544 The plasmids for expressing full length P or its fragments were transformed into *E.*  
545 *coli* strain BL21(DE3) or *E.coli* BL21(DE3)-RIL for expressing P<sub>CED</sub>. Cells were grown in  
546 LB medium containing 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin at 37°C until O.D<sub>600</sub> reached 0.6-0.8 A.U. At  
547 this point, IPTG at a final concentration of 0.5 mM was added and the temperature was  
548 shifted to 24°C. After an incubation of 5 hours, cells were harvested by centrifugation and  
549 suspended in 20 mM Tris-HCl pH 7.5, 150 mM NaCl (Buffer A) supplemented with anti-  
550 proteases (Complete<sup>TM</sup> Protease Inhibitor Cocktail Tablets, Roche), DNaseI (Sigma) and 20  
551 mM MgSO<sub>4</sub>, and cells were disrupted by sonication. The extract was centrifuged at 20,000 g  
552 during 1h at 4°C. The supernatant was filtered (0.45  $\mu\text{m}$ ) and loaded onto a Ni<sup>2+</sup> resin column,  
553 pre-equilibrated with Buffer A. The resin was washed with three bed volumes of Buffer A,  
554 then with three bed volumes of 20 mM Tris-HCl, 1.5 M NaCl, 10 mM imidazole at pH 7.5,  
555 and the protein was eluted using Buffer A supplemented with 400 mM imidazole. The  
556 recombinant protein was concentrated (Centricon, Amicon-10,000 MWCO) and loaded onto a  
557 size exclusion chromatography column (Superdex 75 HR 10/30, GE Healthcare) equilibrated  
558 with Buffer A. For P<sub>CED</sub>, the size exclusion chromatography was performed in 20 mM Tris-  
559 HCl, 500 mM NaCl at pH 8.5. Separations were performed at a flow rate of 0.5 mL $\cdot\text{min}^{-1}$ .

560 The protein solution was dialyzed against water and stored at -80°C. The purity of the protein  
 561 samples were assessed by SDS-PAGE. The identity and integrity of the proteins were  
 562 confirmed by electrospray mass spectrometry. The experiments were performed on a Quattro  
 563 II mass spectrometer (Micromass, Altricham, U.K.) by continuously injecting the sample with  
 564 a Type 22 pump from Harvard Apparatus at a flow rate of 5  $\mu\text{L}\cdot\text{min}^{-1}$ . Protein concentrations  
 565 were measured by spectrophotometry using extinction coefficients calculated from the amino  
 566 acid sequence <sup>96</sup>.

567

### 568 **Circular dichroism spectroscopy**

569 A J-810 CD spectropolarimeter from JASCO equipped with a temperature controller  
 570 (Peltier system) was used to record the far-UV CD spectra at 20°C, which were accumulated  
 571 at least 5 times. Full-length RV P or fragments thereof were diluted to final concentrations  
 572 ranging from 10  $\mu\text{M}$  to 40  $\mu\text{M}$  in 20 mM Tris-HCl pH 7.5 containing 150 mM NaF using a  
 573 cuvette with a path length of 1 mm.

574 After subtracting the blank signal, the CD signal (in millidegrees) was converted to  
 575 mean molar residue ellipticity (in  $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ ) using the following equation:

$$576 \quad [\theta] = \frac{0.1 (CD \text{ signal})}{l C N_{aa}} \quad (\text{Eq. 4})$$

577 where  $l$  is the pathlength of the cuvette in centimeters,  $C$  is the protein concentration in molar  
 578 and  $N_{aa}$  is the number of amino acids. The ellipticity at 222 nm was converted to fraction  
 579 helix values using:

$$580 \quad f_{\alpha} = \frac{[\theta]}{[\theta]_{helix}} \quad (\text{Eq. 5})$$

581 where  $[\theta]_{helix} = -35000 \text{ deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$  represents the mean molar residue ellipticity for 100%  
 582 helix calculated from the following expression derived for peptides <sup>97</sup>:

$$583 \quad [\theta]_{helix} = (-44000 + 250T) \left(1 - \frac{n}{N_r}\right) \quad (\text{Eq. 6})$$

584 where  $T$  is the temperature in °C,  $n$  is the number of non-H-bonded CO groups in the peptide  
 585 and  $N_r$  is the chain length in residues. For proteins, we used  $n = 1$ , assuming most CO groups  
 586 at the C-terminal extremity of  $\alpha$ -helices are capped and  $N_r = 10$  as an average helix length.

587

### 588 SEC-MALLS-RI

589 Size-exclusion chromatography (SEC) was performed with an S75 Superdex column  
 590 (GE Healthcare) equilibrated with buffer A. Separations were performed at 20°C with a flow  
 591 rate of 0.5 mL.min<sup>-1</sup>. Typically, 50  $\mu$ L of a protein solution at a concentration of 5-10 mg.mL<sup>-1</sup>  
 592 were injected. On-line Multi-Angle Laser Light Scattering (MALLS) detection was  
 593 performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using  
 594 a laser emitting at 690 nm. Data were analyzed and weight-averaged molar masses ( $M_w$ ) were  
 595 calculated using the ASTRA software (Wyatt Technology Corp., Santa Barbara, CA) as  
 596 described previously <sup>11</sup>.

597 The excluded ( $V_0$ ) and total volumes ( $V_t$ ) were measured with blue dextran and  
 598 thymidine, respectively. The partition coefficients ( $K_{av}$ ) were calculated by using

$$599 \quad K_{av} = \frac{V_e - V_0}{V_t - V_0} \quad (\text{Eq. 7})$$

600 where  $V_e$  is the elution volume of the protein. The column was calibrated using proteins of  
 601 known Stokes' radii ( $R_S$ ) and molecular masses ( $M_w$ ) (35): bovine serum albumine ( $R_S = 3.4$   
 602 nm,  $M_w = 67.0$  kDa), RNase A ( $R_S = 1.9$  nm,  $M_w = 13.7$  kDa), ovalbumin ( $R_S = 3.0$  nm,  $M_w$   
 603 = 43.5 kDa),  $\beta$ -lactoglobulin ( $R_S = 2.7$  nm,  $M_w = 36.8$  kDa), and chymotrypsinogen ( $R_S = 2.3$   
 604 nm,  $M_w = 25$  kDa) (35).

605

606

**607 Nucleocapsid binding assay**

608           The interaction between circular N-RNA complexes and the various P proteins was  
609 investigated by native gel electrophoresis. Circular N-RNA complexes containing 10  
610 nucleoprotein protomers (N<sub>10</sub>-RNA) were purified as described previously<sup>64</sup>. The N<sub>10</sub>-RNA  
611 complex was incubated in the presence of full length RV P, P<sub>Δ91-131</sub>, P<sub>ΔN52</sub>, P<sub>ΔC120</sub>, P<sub>CTD</sub> or  
612 P<sub>CTD</sub> at 20°C during 10 min. A native gel of 4% acrylamide and with a 19:1  
613 acrylamide/bisacrylamide ratio was pre-run for 30 min at 200 V in the cold room. Typically,  
614 10 μl of protein mixture at about 1 mg/ml (where P fragments were in molar excess with  
615 respect to N-RNA) were loaded, and the gel was run for 3 hours at 200 V. Complexes were  
616 identified by Coomassie blue staining.

617



618 **ACKNOWLEDGMENTS**

619           This work was supported by the interdisciplinary program “Maladies Infectieuses  
620 Emergentes” from the CNRS and by a grant from the French ANR (ANR-07-001-01  
621 (ANRAGE)) and Lyonbiopôle. We thank Bertrand Lefèbvre, Denis Daveloos and Bernard  
622 Brasmes (CRSSA, Grenoble) for mass spectrometry measurements and Baptiste Fouquet and  
623 Hélène Raux for the amino acid substitutions in P and the test in the two-hybrid system. We  
624 thank Noël Tordo for extensive discussions and critical reading of the manuscript. E.A.R. was  
625 supported by postdoctoral fellowships from the University Joseph Fourier and from the ANR  
626 program (ANR-06-JCJC-0126-01). F.C.A.G. and C.L. were supported by MENRT  
627 fellowships from the French government.

## 628 REFERENCES

- 629 1. Karlin, D., Ferron, F., Canard, B. & Longhi, S. (2003). Structural disorder and  
630 modular organization in Paramyxovirinae N and P. *J Gen Virol* 84, 3239-52.
- 631 2. Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P.,  
632 Ruigrok, R. W. & Marion, D. (2004). Structure and dynamics of the  
633 nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution.  
634 *Virology* 319, 201-11.
- 635 3. Tarbouriech, N., Curran, J., Ruigrok, R. W. & Burmeister, W. P. (2000).  
636 Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat Struct Biol* 7,  
637 777-81.
- 638 4. Llorente, M. T., Garcia-Barreno, B., Calero, M., Camafeita, E., Lopez, J. A.,  
639 Longhi, S., Ferron, F., Varela, P. F. & Melero, J. A. (2006). Structural analysis of  
640 the human respiratory syncytial virus phosphoprotein: characterization of an  
641 alpha-helical domain involved in oligomerization. *J Gen Virol* 87, 159-69.
- 642 5. Bourhis, J. M., Canard, B. & Longhi, S. (2006). Structural disorder within the  
643 replicative complex of measles virus: functional implications. *Virology* 344, 94-  
644 110.
- 645 6. Mavrakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. (2004).  
646 Structure and function of the C-terminal domain of the polymerase cofactor of  
647 rabies virus. *J Mol Biol* 343, 819-31.
- 648 7. Ding, H., Green, T. J., Lu, S. & Luo, M. (2006). Crystal structure of the  
649 oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J*  
650 *Virol* 80, 2808-14.
- 651 8. Ding, H., Green, T. J. & Luo, M. (2004). Crystallization and preliminary X-ray  
652 analysis of a proteinase-K-resistant domain within the phosphoprotein of  
653 vesicular stomatitis virus (Indiana). *Acta Crystallogr D Biol Crystallogr* 60, 2087-  
654 90.
- 655 9. Mavrakis, M., Mehoulas, S., Real, E., Iseni, F., Blondel, D., Tordo, N. & Ruigrok,  
656 R. W. (2006). Rabies virus chaperone: identification of the phosphoprotein  
657 peptide that keeps nucleoprotein soluble and free from non-specific RNA.  
658 *Virology* 349, 422-9.
- 659 10. Chen, M., Ogino, T. & Banerjee, A. K. (2007). Interaction of vesicular stomatitis  
660 virus P and N proteins: Identification of two overlapping domains at the N-  
661 terminus of P that are involved in N0-P complex formation and encapsidation of  
662 viral genome RNA. *J Virol* 81, 13478-13485.
- 663 11. Gerard, F., Ribeiro, E., Albertini, A., Zaccari, G., Ebel, C., Ruigrok, R. & Jamin,  
664 M. (2007). Unphosphorylated Rhabdoviridae phosphoproteins form elongated  
665 dimers in solution. *Biochemistry* 46, 10328-10338.
- 666 12. Jacob, Y., Real, E. & Tordo, N. (2001). Functional interaction map of lyssavirus  
667 phosphoprotein: identification of the minimal transcription domains. *J Virol* 75,  
668 9613-22.
- 669 13. Gigant, B., Iseni, F., Gaudin, Y., Knossow, M. & Blondel, D. (2000). Neither  
670 phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is  
671 required for its oligomerization. *J Gen Virol* 81, 1757-61.
- 672 14. Ribeiro, E. A., Jr., Favier, A., Gerard, F. C., Leyrat, C., Brutscher, B., Blondel,  
673 D., Ruigrok, R. W., Blackledge, M. & Jamin, M. (2008). Solution Structure of the  
674 C-Terminal Nucleoprotein-RNA Binding Domain of the Vesicular Stomatitis  
675 Virus Phosphoprotein. *J Mol Biol* 382, 525-538.

- 676 15. Chenik, M., Chebli, K. & Blondel, D. (1995). Translation initiation at alternate  
677 in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a  
678 ribosomal leaky scanning mechanism. *J Virol* 69, 707-12.
- 679 16. Padeloup, D., Poisson, N., Raux, H., Gaudin, Y., Ruigrok, R. W. & Blondel, D.  
680 (2005). Nucleocytoplasmic shuttling of the rabies virus P protein requires a  
681 nuclear localization signal and a CRM1-dependent nuclear export signal.  
682 *Virology* 334, 284-93.
- 683 17. Moseley, G. W., Filmer, R. P., DeJesus, M. A. & Jans, D. A. (2007).  
684 Nucleocytoplasmic distribution of rabies virus P-protein is regulated by  
685 phosphorylation adjacent to C-terminal nuclear import and export signals.  
686 *Biochemistry* 46, 12053-61.
- 687 18. Vidy, A., Bougrini, J., Cheldi-Alix, M. & Blondel, D. (2007). The  
688 nucleocytoplasmic rabies P protein counteracts interferons signaling by  
689 inhibiting both nuclear accumulation and DNA binding of STAT1. *J Virol.* 81,  
690 4255-4263.
- 691 19. Blondel, D., Regad, T., Poisson, N., Pavie, B., Harper, F., Pandolfi, P. P., De The,  
692 H. & Chelbi-Alix, M. K. (2002). Rabies virus P and small P products interact  
693 directly with PML and reorganize PML nuclear bodies. *Oncogene* 21, 7957-70.
- 694 20. Vidy, A., Chelbi-Alix, M. & Blondel, D. (2005). Rabies virus P protein interacts  
695 with STAT1 and inhibits interferon signal transduction pathways. *J Virol* 79,  
696 14411-20.
- 697 21. Raux, H., Flamand, A. & Blondel, D. (2000). Interaction of the rabies virus P  
698 protein with the LC8 dynein light chain. *J Virol* 74, 10212-6.
- 699 22. Jacob, Y., Badrane, H., Ceccaldi, P. E. & Tordo, N. (2000). Cytoplasmic dynein  
700 LC8 interacts with lyssavirus phosphoprotein. *J Virol* 74, 10217-22.
- 701 23. Perutz, M. F. (1963). X-ray analysis of hemoglobin. *Science* 140, 863-9.
- 702 24. Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C. & Sarma, V.  
703 R. (1965). Structure of hen egg-white lysozyme. A three-dimensional Fourier  
704 synthesis at 2 Angstrom resolution. *Nature* 206, 757-61.
- 705 25. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips,  
706 D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by  
707 x-ray analysis. *Nature* 181, 662-6.
- 708 26. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains.  
709 *Science* 181, 223-30.
- 710 27. Tanford, C. (1968). Protein denaturation. *Adv Protein Chem* 23, 121-282.
- 711 28. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-  
712 assessing the protein structure-function paradigm. *J Mol Biol* 293, 321-31.
- 713 29. Uversky, V. N., Gillespie, J. R. & Fink, A. L. (2000). Why are "natively  
714 unfolded" proteins unstructured under physiological conditions? *Proteins* 41, 415-  
715 27.
- 716 30. Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur J Biochem*  
717 269, 2-12.
- 718 31. Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem Sci* 27,  
719 527-33.
- 720 32. Spolar, R. S. & Record, M. T., Jr. (1994). Coupling of local folding to site-specific  
721 binding of proteins to DNA. *Science* 263, 777-84.
- 722 33. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for  
723 unstructured proteins. *Curr Opin Struct Biol* 12, 54-60.
- 724 34. Sugase, K., Dyson, H. J. & Wright, P. E. (2007). Mechanism of coupled folding  
725 and binding of an intrinsically disordered protein. *Nature* 447, 920-921.

- 726 35. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J.,  
727 Evans, R. M. & Wright, P. E. (2002). Mutual synergistic folding in recruitment of  
728 CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415, 549-53.
- 729 36. Uversky, V. N., Segel, D. J., Doniach, S. & Fink, A. L. (1998). Association-  
730 induced folding of globular proteins. *Proc Natl Acad Sci U S A* 95, 5480-3.
- 731 37. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N. & Dunker, A. K.  
732 (2006). Intrinsic disorder in transcription factors. *Biochemistry* 45, 6873-88.
- 733 38. Bourhis, J. M., Johansson, K., Receveur-Brechot, V., Oldfield, C. J., Dunker, K.  
734 A., Canard, B. & Longhi, S. (2004). The C-terminal domain of measles virus  
735 nucleoprotein belongs to the class of intrinsically disordered proteins that fold  
736 upon binding to their physiological partner. *Virus Res* 99, 157-67.
- 737 39. Romero, P., Obradovic, Z. & Dunker, A. K. (2004). Natively disordered proteins:  
738 functions and predictions. *Appl Bioinformatics* 3, 105-13.
- 739 40. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G.,  
740 Obradovic, Z. & Dunker, A. K. (2004). The importance of intrinsic disorder for  
741 protein phosphorylation. *Nucleic Acids Res* 32, 1037-49.
- 742 41. Tompa, P. & Csermely, P. (2004). The role of structural disorder in the function  
743 of RNA and protein chaperones. *Faseb J* 18, 1169-75.
- 744 42. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. (2004). The  
745 DISOPRED server for the prediction of protein disorder. *Bioinformatics* 20,  
746 2138-9.
- 747 43. Bourhis, J. M., Canard, B. & Longhi, S. (2007). Predicting protein disorder and  
748 induced folding: from theoretical principles to practical applications. *Curr*  
749 *Protein Pept Sci* 8, 135-49.
- 750 44. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O.,  
751 Beckmann, J. S., Silman, I. & Sussman, J. L. (2005). FoldIndex: a simple tool to  
752 predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*  
753 21, 3435-8.
- 754 45. Promponas, V. J., Enright, A. J., Tsoka, S., Kreil, D. P., Leroy, C., Hamodrakas,  
755 S., Sander, C. & Ouzounis, C. A. (2000). CAST: an iterative algorithm for the  
756 complexity analysis of sequence tracts. Complexity analysis of sequence tracts.  
757 *Bioinformatics* 16, 915-22.
- 758 46. Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. &  
759 Blackledge, M. (2005). A structural model for unfolded proteins from residual  
760 dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci U S A*.
- 761 47. Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R. W. & Burmeister, W. P.  
762 (2000). On the domain structure and the polymerization state of the sendai virus  
763 P protein. *Virology* 266, 99-109.
- 764 48. Houben, K., Blanchard, L., Blackledge, M. & Marion, D. (2007). Intrinsic  
765 dynamics of the partly unstructured PX domain from the Sendai virus RNA  
766 polymerase co-factor P. *Biophys J* 93, 2930-2940.
- 767 49. Lieutaud, P., Canard, B. & Longhi, S. (2008). MeDor: a metasever for  
768 predicting protein disorder. *BMC Genomics* 9(Suppl. 2), S25.
- 769 50. Vucetic, S., Brown, C. J., Dunker, A. K. & Obradovic, Z. (2003). Flavors of  
770 protein disorder. *Proteins* 52, 573-84.
- 771 51. Ferron, F., Longhi, S., Canard, B. & Karlin, D. (2006). A practical overview of  
772 protein disorder prediction methods. *Proteins* 65, 1-14.
- 773 52. King, R. D., Ouali, M., Strong, A. T., Aly, A., Elmaghraby, A., Kantardzic, M. &  
774 Page, D. (2000). Is it better to combine predictions? *Protein Eng* 13, 15-9.

- 775 53. Schulz, G. E., Barry, C. D., Friedman, J., Chou, P. Y., Fasman, G. D., Finkelstein,  
776 A. V., Lim, V. I., Pititsyn, O. B., Kabat, E. A., Wu, T. T., Levitt, M., Robson, B. &  
777 Nagano, K. (1974). Comparison of predicted and experimentally determined  
778 secondary structure of adenyl kinase. *Nature* 250, 140-2.
- 779 54. Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield,  
780 C. J., Williams, C. J. & Dunker, A. K. (2002). Evolutionary rate heterogeneity in  
781 proteins with long disordered regions. *J Mol Evol* 55, 104-10.
- 782 55. Pei, J. & Grishin, N. V. (2001). AL2CO: calculation of positional conservation in  
783 a protein sequence alignment. *Bioinformatics* 17, 700-12.
- 784 56. Kingston, R. L., Hamel, D. J., Gay, L. S., Dahlquist, F. W. & Matthews, B. W.  
785 (2004). Structural basis for the attachment of a paramyxoviral polymerase to its  
786 template. *Proc Natl Acad Sci U S A* 101, 8301-6.
- 787 57. Ferron, F., Rancurel, C., Longhi, S., Cambillau, C., Henrissat, B. & Canard, B.  
788 (2005). VaZyMolO: a tool to define and classify modularity in viral proteins. *J*  
789 *Gen Virol* 86, 743-9.
- 790 58. Chenik, M., Chebli, K., Gaudin, Y. & Blondel, D. (1994). In vivo interaction of  
791 rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-  
792 binding sites on P protein. *J Gen Virol* 75 2889-96.
- 793 59. Chenik, M., Schnell, M., Conzelmann, K. K. & Blondel, D. (1998). Mapping the  
794 interacting domains between the rabies virus polymerase and phosphoprotein. *J*  
795 *Virol* 72, 1925-30.
- 796 60. Green, T. J., Macpherson, S., Qiu, S., Lebowitz, J., Wertz, G. W. & Luo, M.  
797 (2000). Study of the assembly of vesicular stomatitis virus N protein: role of the P  
798 protein. *J Virol* 74, 9515-24.
- 799 61. Fu, Z. F., Zheng, Y., Wunner, W. W., Koprowski, H. & Dietzschold, B. (1994).  
800 Both the N- and C-terminal domains of the nominal phosphoprotein of rabies  
801 virus are involved in binding to the nucleoprotein. *Virology* 200, 590-597.
- 802 62. Scholtz, J. M., Qian, H., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991).  
803 Parameters of helix-coil transition theory for alanine-based peptides of varying  
804 chain lengths in water. *Biopolymers* 31, 1463-70.
- 805 63. Uversky, V. N. (1993). Use of fast protein size-exclusion liquid chromatography  
806 to study the unfolding of proteins which denature through the molten globule.  
807 *Biochemistry* 32, 13288-98.
- 808 64. Albertini, A. A., Clapier, C. R., Wernimont, A. K., Schoehn, G., Weissenhorn, W.  
809 & Ruigrok, R. W. (2006). Isolation and crystallization of a unique size category of  
810 recombinant Rabies virus Nucleoprotein-RNA rings. *J Struct Biol* 158, 129-133.
- 811 65. Johansson, K., Bourhis, J. M., Campanacci, V., Cambillau, C., Canard, B. &  
812 Longhi, S. (2003). Crystal structure of the measles virus phosphoprotein domain  
813 responsible for the induced folding of the C-terminal domain of the  
814 nucleoprotein. *J Biol Chem* 278, 44567-73.
- 815 66. Curran, J., Boeck, R., Lin-Marq, N., Lupas, A. & Kolakofsky, D. (1995).  
816 Paramyxovirus phosphoproteins form homotrimers as determined by an epitope  
817 dilution assay, via predicted coiled coils. *Virology* 214, 139-49.
- 818 67. Curran, J., Marq, J. B. & Kolakofsky, D. (1995). An N-terminal domain of the  
819 Sendai paramyxovirus P protein acts as a chaperone for the NP protein during  
820 the nascent chain assembly step of genome replication. *J Virol* 69, 849-55.
- 821 68. Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. (2004). Viral DNA  
822 polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*  
823 318, 463-73.

- 824 69. Li, X., Romero, P., Rani, M., Dunker, A. K. & Obradovic, Z. (1999). Predicting  
825 Protein Disorder for N-, C-, and Internal Regions. *Genome Inform Ser Workshop*  
826 *Genome Inform* 10, 30-40.
- 827 70. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K.  
828 (2001). Sequence complexity of disordered protein. *Proteins* 42, 38-48.
- 829 71. Liu, J., Tan, H. & Rost, B. (2002). Loopy proteins appear conserved in evolution.  
830 *J Mol Biol* 322, 53-64.
- 831 72. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. (2005). IUPred: web server for  
832 the prediction of intrinsically unstructured regions of proteins based on  
833 estimated energy content. *Bioinformatics* 21, 3433-4.
- 834 73. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B.  
835 (2003). Protein disorder prediction: implications for structural proteomics.  
836 *Structure* 11, 1453-9.
- 837 74. Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. (2005). RONN: the bio-  
838 basis function neural network technique applied to the detection of natively  
839 disordered regions in proteins. *Bioinformatics* 21, 3369-76.
- 840 75. Coeytaux, K. & Poupon, A. (2005). Prediction of unfolded segments in a protein  
841 sequence based on amino acid composition. *Bioinformatics* 21, 1891-900.
- 842 76. Wootton, J. (1994). Non-globular domains in protein sequences: automated  
843 segmentation using complexity measures. *Comput. Chem.* 18, 269-285.
- 844 77. Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J.,  
845 Henrissat, B. & Mornon, J. P. (1997). Deciphering protein sequence information  
846 through hydrophobic cluster analysis (HCA): current status and perspectives.  
847 *Cell Mol Life Sci* 53, 621-45.
- 848 78. Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. (2000). NPS@: network  
849 protein sequence analysis. *Trends Biochem Sci* 25, 147-50.
- 850 79. King, R. D., Saqi, M., Sayle, R. & Sternberg, M. J. (1997). DSC: public domain  
851 protein secondary structure prediction. *Comput Appl Biosci* 13, 473-4.
- 852 80. Garnier, J., Gibrat, J. F. & Robson, B. (1996). GOR method for predicting  
853 protein secondary structure from amino acid sequence. *Methods Enzymol* 266,  
854 540-53.
- 855 81. Guermeur, Y., Geourjon, C., Gallinari, P. & Deleage, G. (1999). Improved  
856 performance in protein secondary structure prediction by inhomogeneous score  
857 combination. *Bioinformatics* 15, 413-21.
- 858 82. Rost, B., Sander, C. & Schneider, R. (1994). PHD--an automatic mail server for  
859 protein secondary structure prediction. *Comput Appl Biosci* 10, 53-60.
- 860 83. Frishman, D. & Argos, P. (1997). Seventy-five percent accuracy in protein  
861 secondary structure prediction. *Proteins* 27, 329-35.
- 862 84. Levin, J. M. (1997). Exploring the limits of nearest neighbour secondary  
863 structure prediction. *Protein Eng* 10, 771-6.
- 864 85. Kneller, D. G., Cohen, F. E. & Langridge, R. (1990). Improvements in protein  
865 secondary structure prediction by an enhanced neural network. *J Mol Biol* 214,  
866 171-82.
- 867 86. Pollastri, G. & McLysaght, A. (2005). Porter: a new, accurate server for protein  
868 secondary structure prediction. *Bioinformatics* 21, 1719-20.
- 869 87. Rost, B. & Liu, J. (2003). The PredictProtein server. *Nucleic Acids Res* 31, 3300-4.
- 870 88. Jones, D. T. (1999). Protein secondary structure prediction based on position-  
871 specific scoring matrices. *J. Mol. Biol.* 292, 195-202.

- 872 **89. Pollastri, G., Przybylski, D., Rost, B. & Baldi, P. (2002). Improving the prediction**  
873 **of protein secondary structure in three and eight classes using recurrent neural**  
874 **networks and profiles. *Proteins* 47, 228-35.**
- 875 **90. Ouali, M. & King, R. D. (2000). Cascaded multiple classifiers for secondary**  
876 **structure prediction. *Protein Sci* 9, 1162-76.**
- 877 **91. Lupas, A., Van Dyke, M. & Stock, J. (1991). Predicting coiled coils from protein**  
878 **sequences. *Science* 252, 1162-4.**
- 879 **92. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M. & Kim, P. S. (1995).**  
880 **Predicting coiled coils by use of pairwise residue correlations. *Proc Natl Acad Sci***  
881 ***U S A* 92, 8259-63.**
- 882 **93. Wolf, E., Kim, P. S. & Berger, B. (1997). MultiCoil: a program for predicting**  
883 **two- and three-stranded coiled coils. *Protein Sci* 6, 1179-89.**
- 884 **94. Delorenzi, M. & Speed, T. (2002). An HMM model for coiled-coil domains and a**  
885 **comparison with PSSM-based predictions. *Bioinformatics* 18, 617-25.**
- 886 **95. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving**  
887 **the sensitivity of progressive multiple sequence alignment through sequence**  
888 **weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids***  
889 ***Res* 22, 4673-80.**
- 890 **96. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in**  
891 **proteins. *Biochemistry* 6, 1948-54.**
- 892 **97. Luo, P. & Baldwin, R. (1997). Mechanism of helix induction by trifluoroethanol:**  
893 **a framework for extrapolating the helix-forming properties of peptides from**  
894 **trifluoroethanol/water mixtures back to water. *Biochemistry* 36, 8413-8421.**  
895

896  
897

## FIGURES LEGENDS

898 **Figure 1. Predicting boundaries between structured and disordered regions in Sendai**  
899 **virus (SV) P.** (A) Location of regions of known function<sup>47; 66; 67</sup>. The conventions used in this  
900 figure are shown in the inset below. (B) Score for disorder (D-score) as a function of residue  
901 number calculated as described in the text. (C) Location of consensus disordered regions with  
902 a D-score < 0.50. Regions with a score ranging between 0.50 and 0.75 are shown as striped  
903 boxes. Numbers above the boxes indicate the boundaries of the consensus disordered regions.  
904 (D) Score for secondary structure prediction calculated from 15 different predictions (see  
905 text). (E) Consensus location of predicted secondary structure as obtained from the scoring  
906 procedure described in the text (SS-score > 0.5 or < -0.5, respectively). (F) Location of  
907 secondary structure elements in the known three-dimensional structures of two domains<sup>2; 3</sup>  
908 with the pdb filenames being provided above the drawing. (G) Score for sequence  
909 conservation calculated with the AL2CO software from a multiple sequence alignment  
910 obtained with ClustalW. Conserved residues are defined as regions with a C-score > 0.5. (H)  
911 Consensus location of conserved regions as defined by the C-score. (I) Global score  
912 calculated as a linear combination of the D-score (Panel B), the absolute value of the SS-score  
913 (Panel D) and the C-score for sequence conservation (Panel G). Structured regions have a  
914 score > 0.5 and disordered regions have a score < 0.5. (J) Consensus location of predicted  
915 disordered and structured domains as defined by the G-score.

916

917 **Figure 2. Predicting boundaries between structured and disordered regions in rabies**  
918 **virus (RV) P.** (A) Location of regions of known function<sup>6; 9; 12; 16; 18; 19; 20; 21; 22; 59</sup>. The  
919 conventions used in this figure are shown in the inset below. (B) to (J), see legend of Figure 1.

920



921 **Figure 3. Binding of RV P to P or N as assessed by a yeast two-hybrid approach. (A)**  
922 **Deletion mutants. (B) Point mutations in the central domain.** The interaction between  
923 mutants of P fused to the DNA binding domain (BD) of Lex A (rows in the figure) and P or N  
924 genes fused to GAL4 AD (columns in the figure) were assayed by the appearance of blue  
925 colonies in the presence of X-Gal.

926

927 **Figure 4. Schematic representation of the different constructs used for producing**  
928 **recombinant proteins.** The predicted disordered regions are labeled DR1 and DR2. NTD:  
929 amino-terminal domain, CED: central domain, CTD: carboxy-terminal domain.

930

931 **Figure 5. Circular dichroism spectroscopy of deletion mutants of RV P.** CD spectra of  
932 different constructs of RV P. Protein spectra were recorded at 20°C in 20 mM Tri-HCl pH 7.5  
933 and 150 mM NaF.

934

935 **Figure 6. Molecular mass and size of deletion mutants of RV P. (A) Molecular mass**  
936 **measured by SEC/MALLS/RI.** The line shows the SEC elution profile as monitored by  
937 refractometry. The crosses show the molecular mass calculated from light scattering and  
938 refractometry data. **(B) Plot of Stokes' radius measured by SEC against the molecular**  
939 **mass measured by MALLS/RI.** The filled and open circles show data taken from <sup>63</sup> for  
940 standard globular proteins in native or unfolded forms, respectively.

941

942 **Figure 7. Binding of deletion mutants of P to nucleoprotein-RNA complexes (N-RNA)**  
943 **using native gel electrophoresis.** Recombinant N-RNA rings containing 10 N protomers  
944 purified from insect cells were incubated with the different constructs of RV P and loaded on  
945 a native gel. Lane 1: N<sub>10</sub>-RNA circular complex containing a small contamination of N<sub>9</sub>-RNA

946 complex. Lanes 2-7. Interaction of N<sub>10</sub>-RNA with different deletion mutants of RV P as  
947 indicated above the lanes. Proteins were visualized by Coomassie staining.

948

949 **Figure 8. Modular organization for *Rhabdoviridae* and *Paramyxoviridae***  
950 **phosphoproteins.** Structured regions are represented by wide boxes. The model for  
951 *Paramyxoviridae* P protein is reproduced from Karlin et al. <sup>1</sup>.

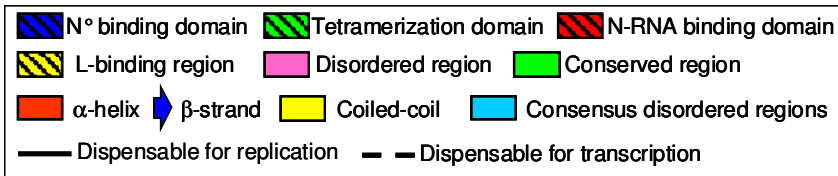
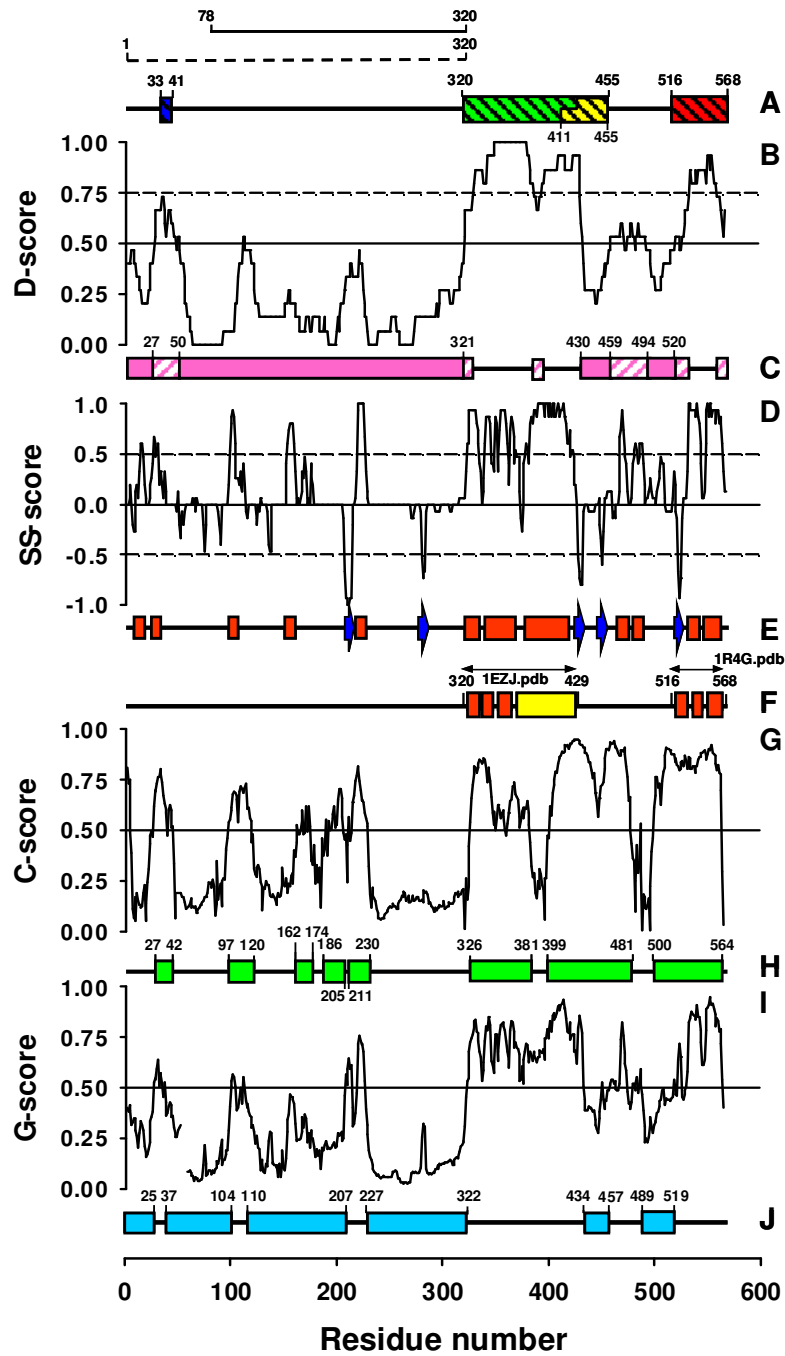
952

953 **Supplementary Material**

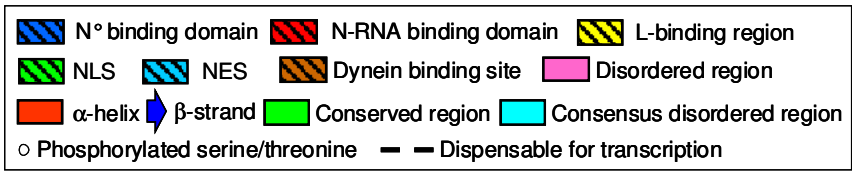
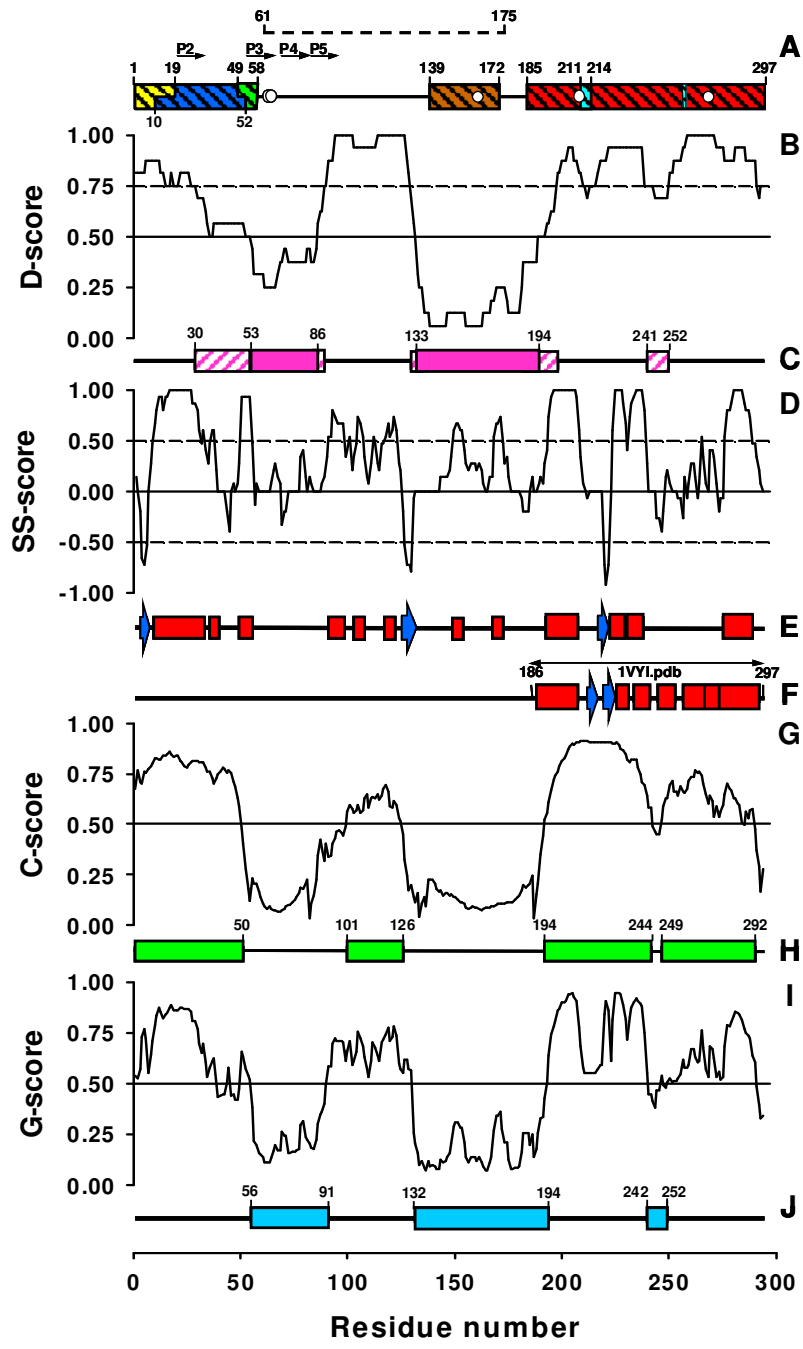
954 **Figure S1. Predictions of structural disorder in SV P (A) and RV P (B) using 16**  
955 **different algorithms.** The different prediction algorithms were used with the default  
956 parameters set in the WEB servers. The filled boxes show the location of predicted disordered  
957 regions.

958

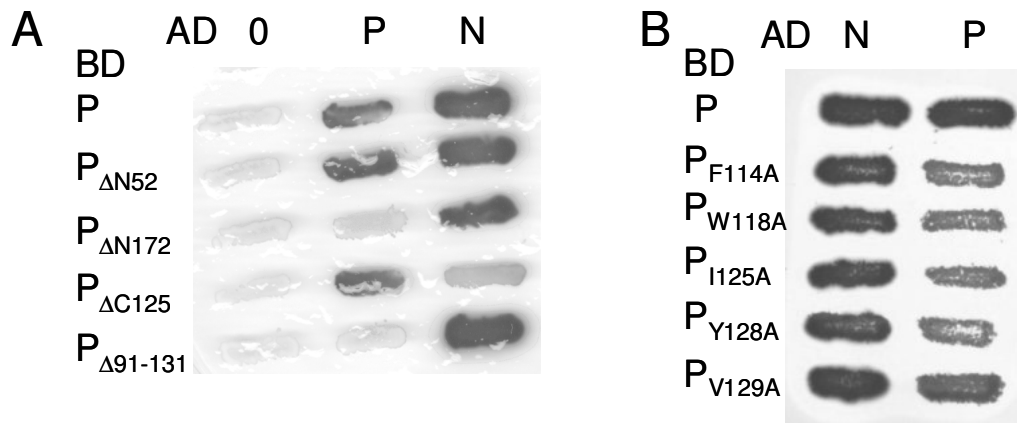
959 **Figure S2. Predictions of structural disorder in SV P (A) and RV P (B) as provided by**  
960 **the Metaserver of Disorder MeDor<sup>49</sup>.** MeDor collects disorder and secondary structure  
961 predictions from servers available on the web and generates a graphical output. Specifically, it  
962 uses predictions from 12 disorder predictors, namely IUPred, Prelink, RONN, FoldUnfold,  
963 DisEMBL, Foldindex, Globplot2, Disprot VL3, Disprot VL3H, Disprot VSL2B, and performs  
964 secondary structure prediction using the pred2ary algorithm using the default parameters (see  
965 <sup>49</sup> and references cited therein). It also incorporates hydrophobic cluster analysis (HCA) and  
966 generates a HCA plot, where regions depleted in hydrophobic clusters correspond to  
967 disordered regions.



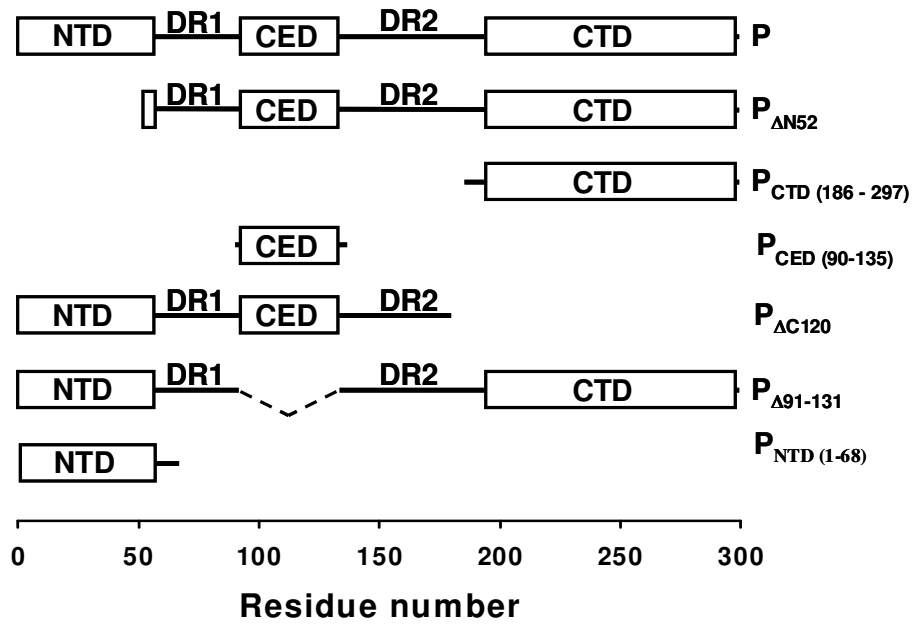
**FIGURE 1**



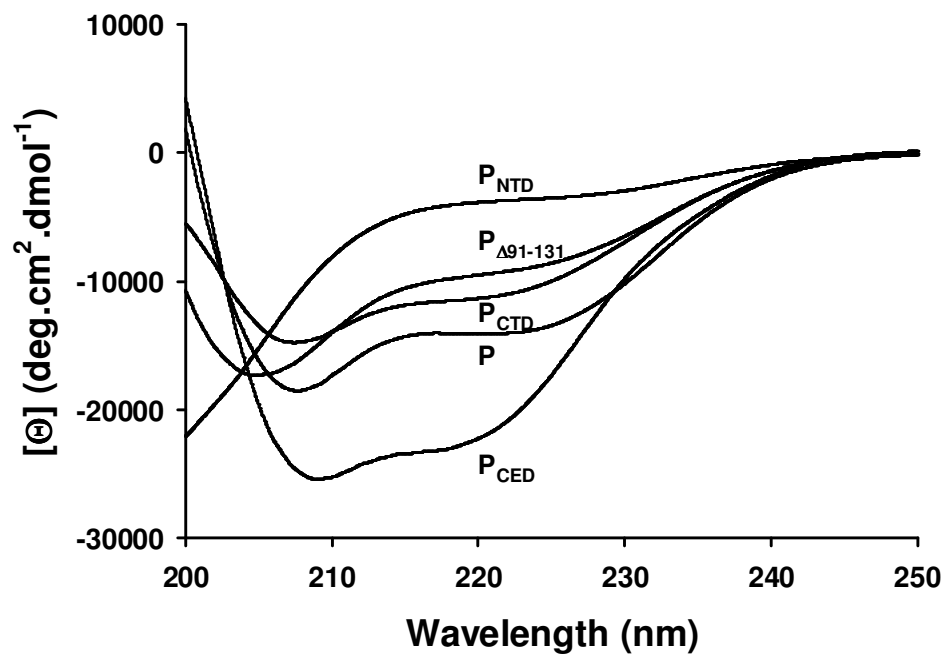
**FIGURE 2**



**FIGURE 3**

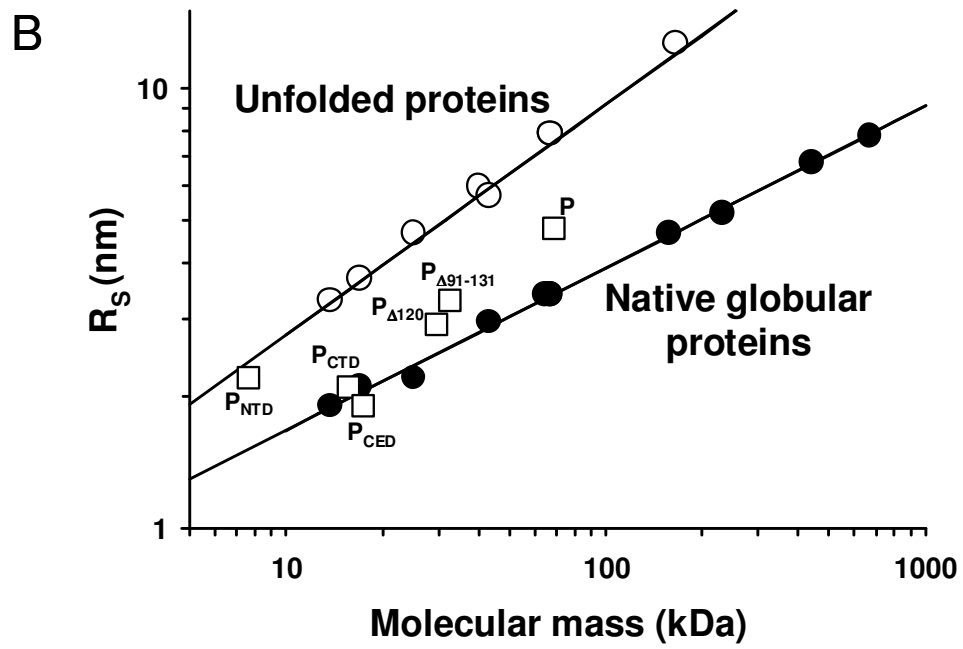
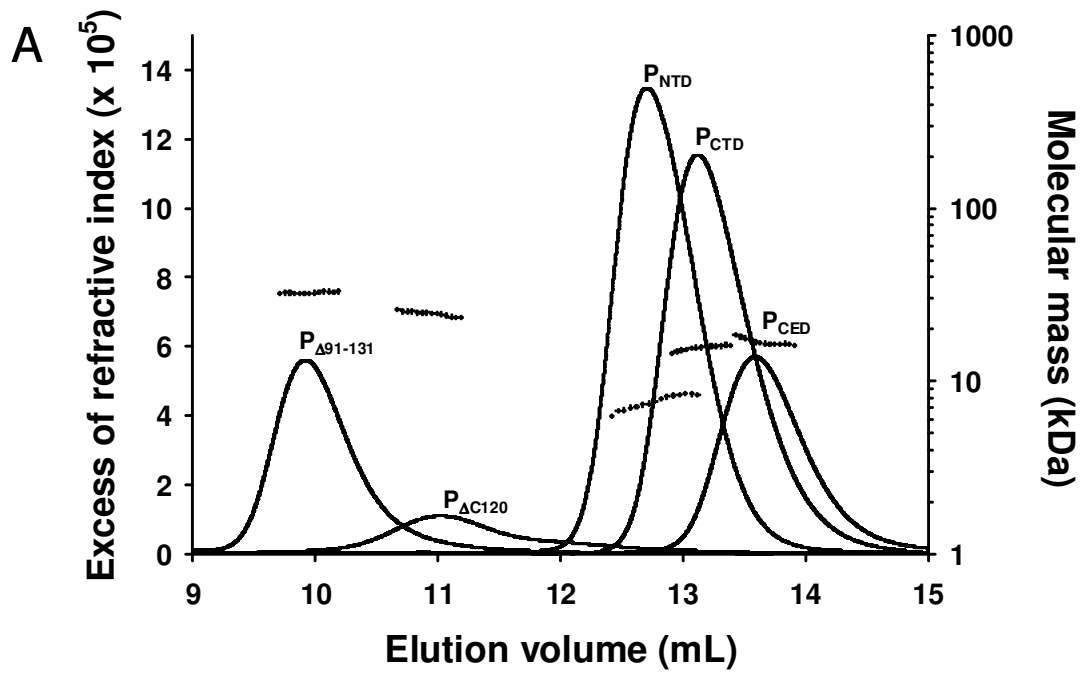


**FIGURE 4**

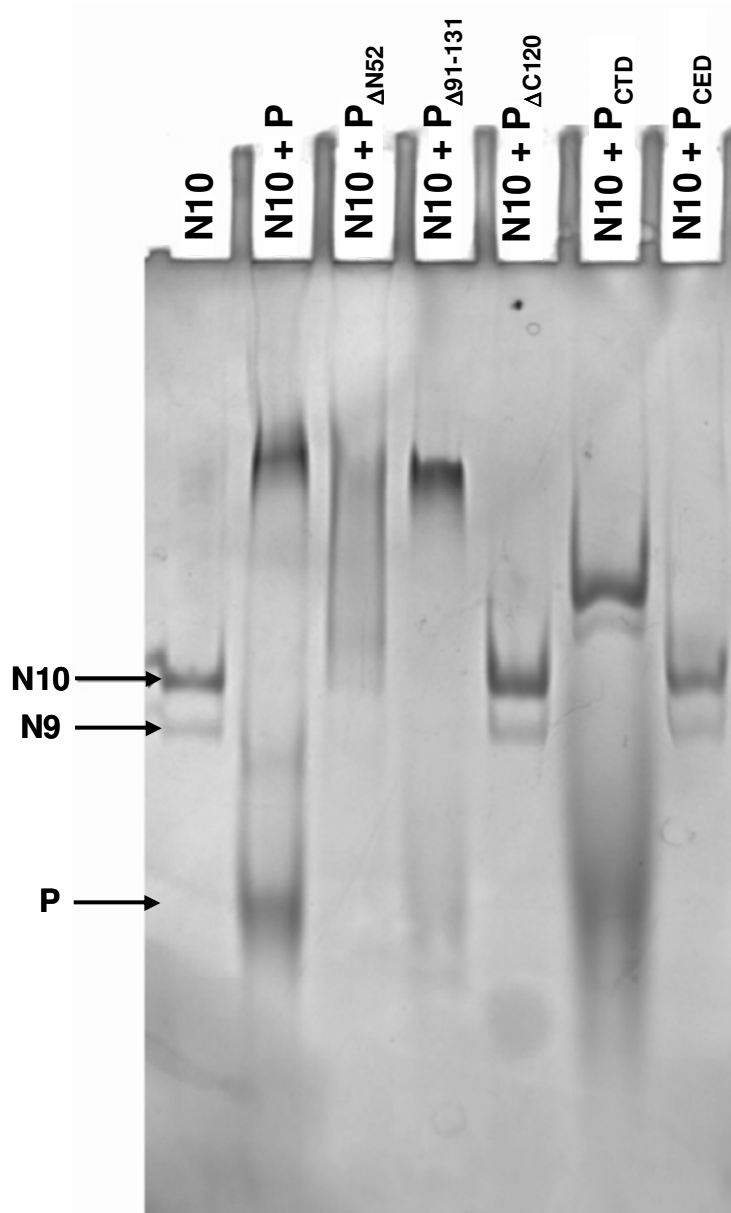


**FIGURE 5**

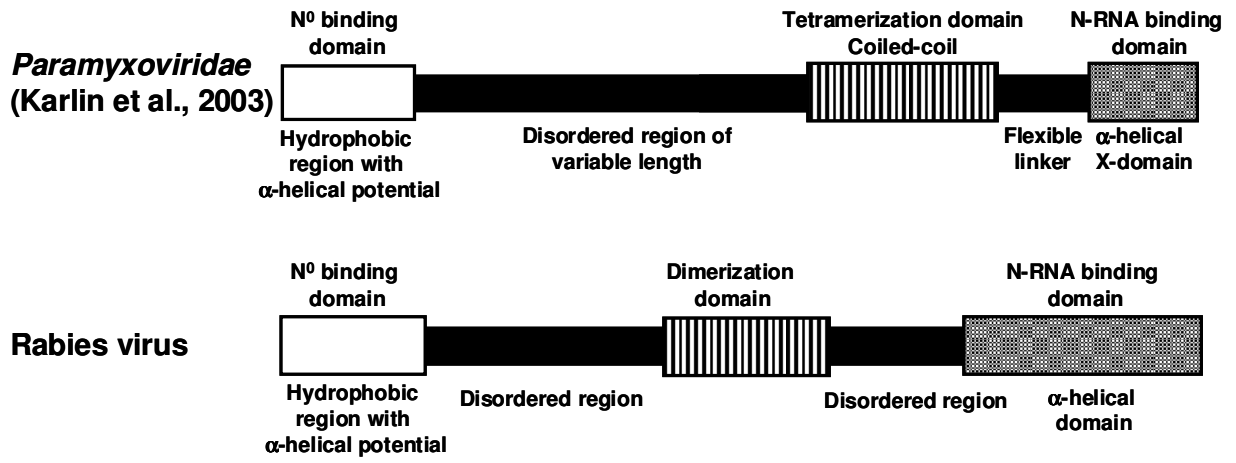




**FIGURE 6**

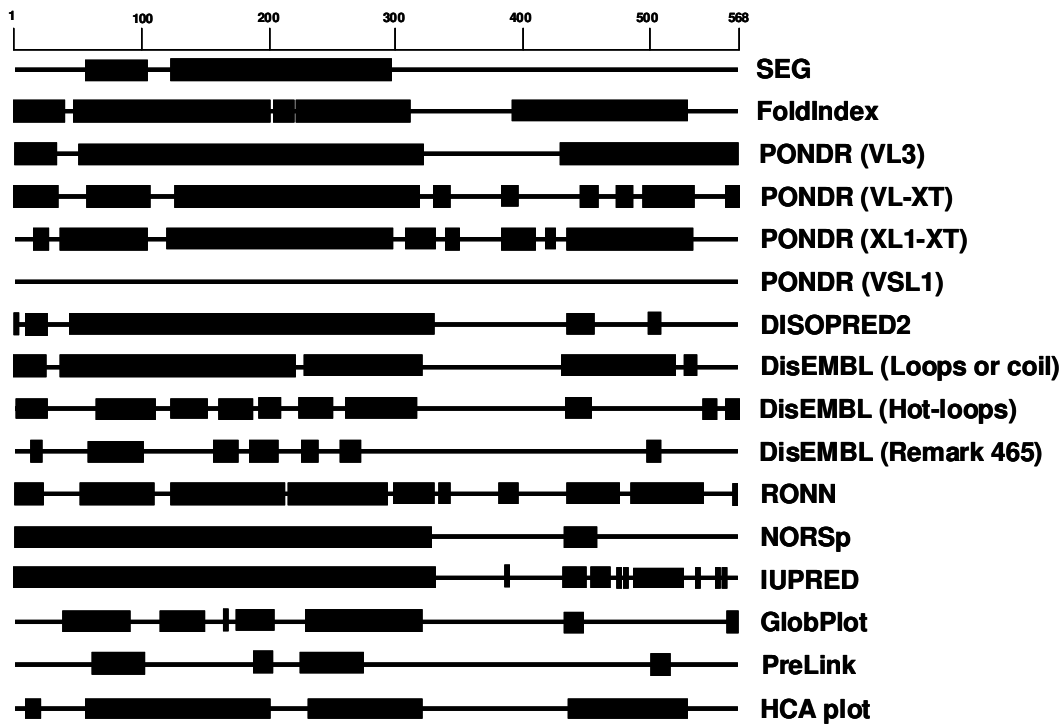


**FIGURE 7**

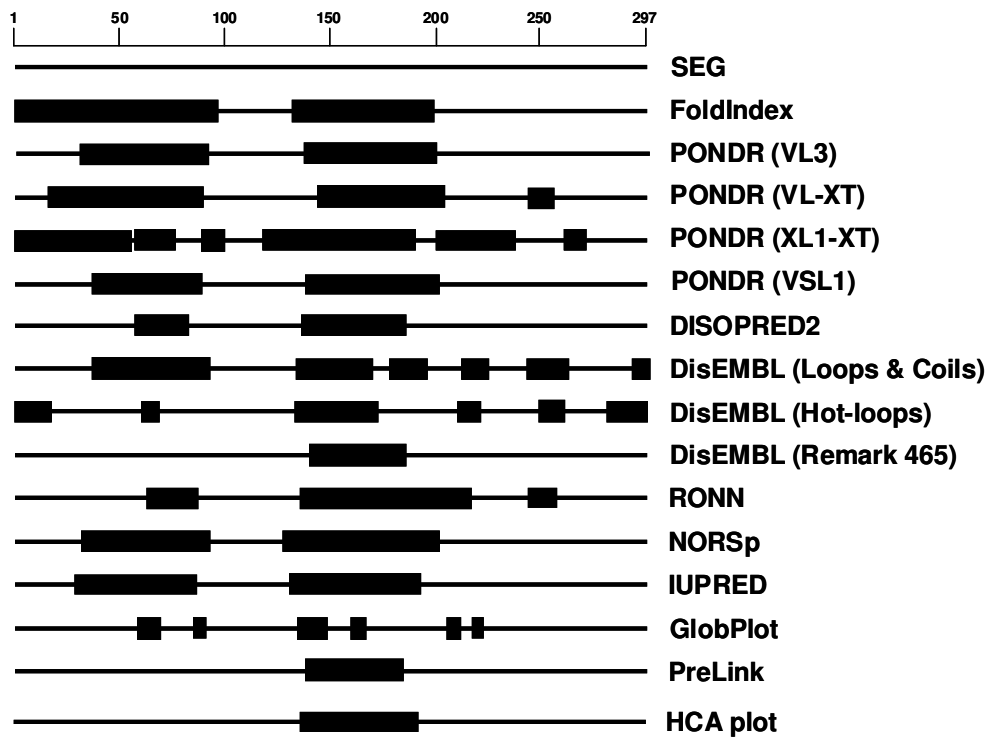


**FIGURE 8**

## A SV



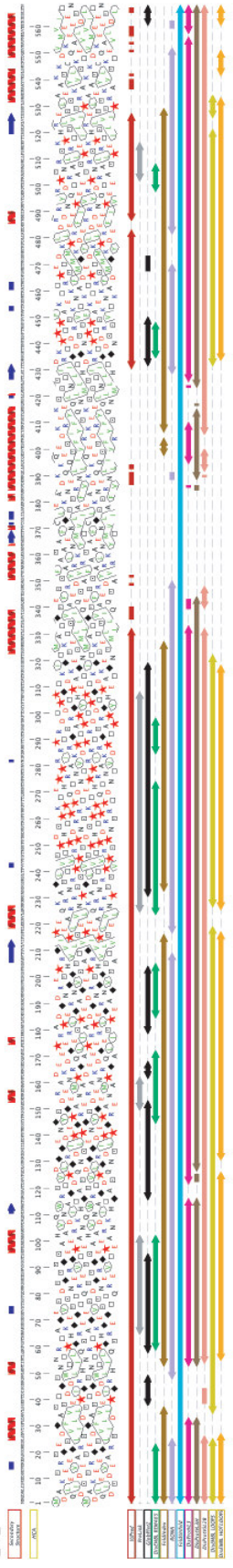
## B RV



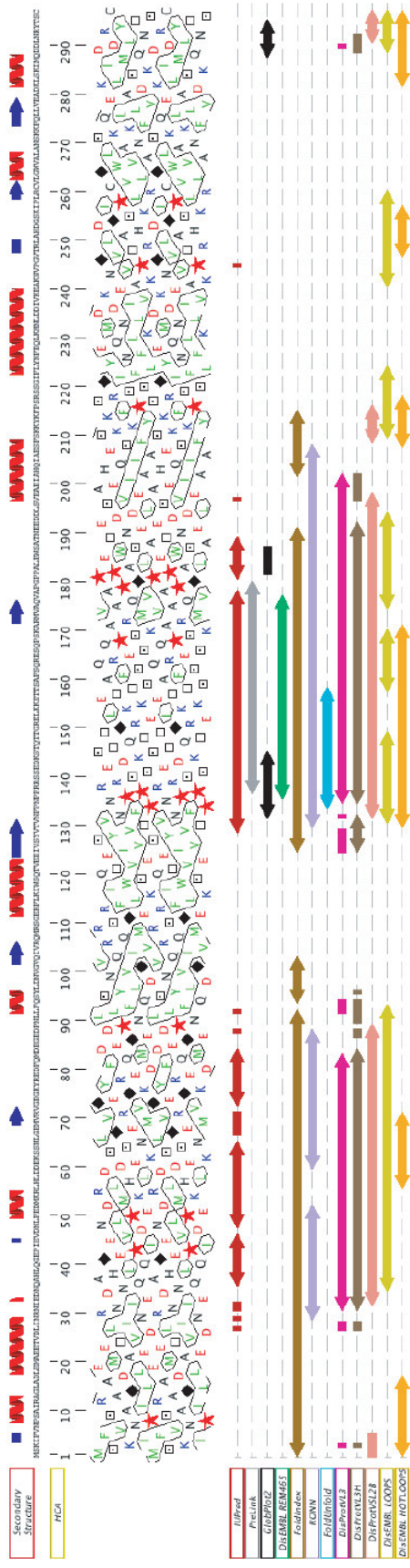
**FIGURE S1**

**FIGURE S2**

**A SV**



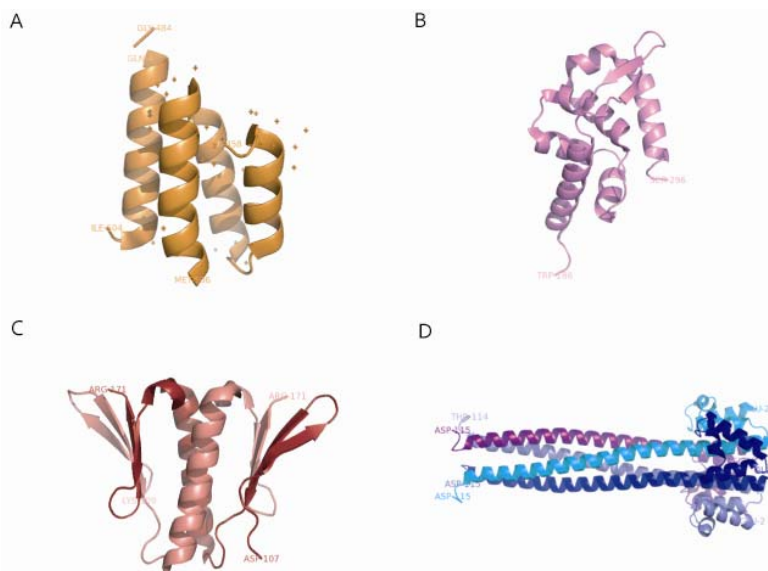
**B RV**



## C – Conclusion

Dans cette publication, nous avons utilisé une méthode simple pour combiner différentes prédictions sur base de la séquence primaire, en prenant en compte les prédictions de structures secondaires ainsi que les conservations de résidus à partir d'alignement de séquences de phosphoprotéines de la même famille. Nous avons mis en évidence la présence de trois régions structurées séparées de deux régions flexibles.

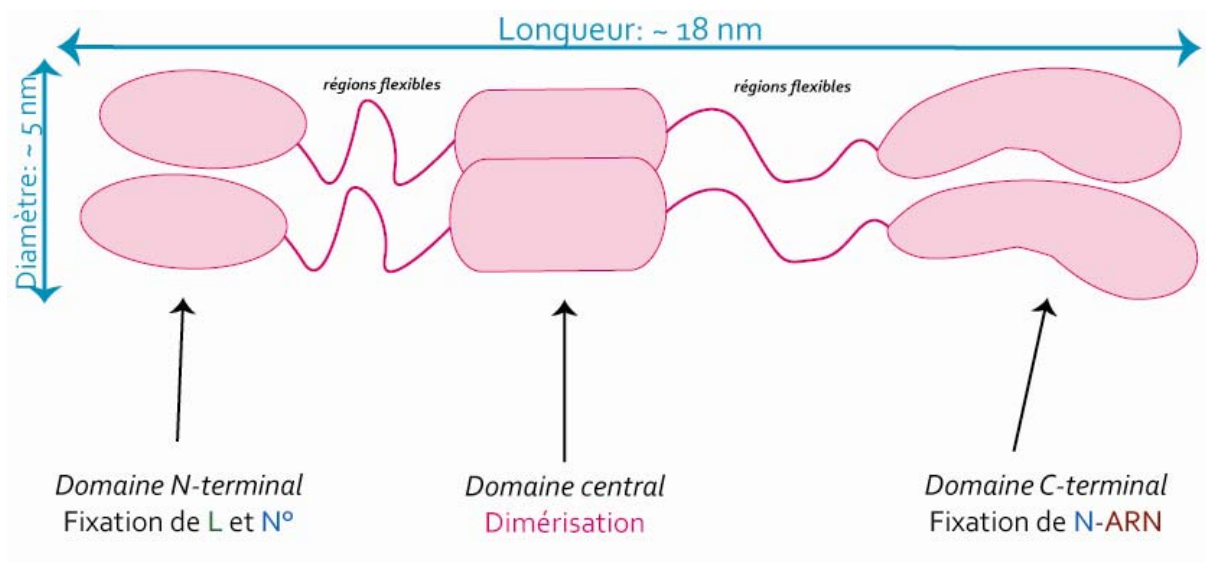
Tous les *Mononegavirales* codent pour une phosphoprotéine qui joue des rôles similaires pour la réplication virale, cependant, peu de similarité de séquences entre ces phosphoprotéines existent au sein de l'ordre des *Mononegavirales* ou d'une même famille. Chez les *Paramyxoviridae*, la structure tridimensionnelle du domaine C-terminal de la phosphoprotéine est conservée entre les virus de la rougeole et Sendai (figure 35, A) (Blanchard & al., 2004) (Johansson & al., 2003), et la séquence du domaine d'oligomérisation est conservé, ce qui suggère un *coiled coil* homotétramérique commun (figure 35,D) (Tarbouriech & al., 2000) (Curran, Boeck & al., 1995). La structure du domaine C-terminal de la phosphoprotéine du virus de Sendai et de la rougeole (figure 35, A) est différente de celle du domaine C-terminal de la phosphoprotéine du virus de la rage (figure 35, B) (Mavrakis & al., 2004), et la structure du domaine d'oligomérisation de la phosphoprotéine du virus de la stomatite vésiculaire (figure 35, C) (Ding & al., 2006) est différente du domaine *coiled coil* tétramérique du virus de Sendai (figure 35, D) (Tarbouriech & al., 2000).



**Figure 35** - Comparaison des domaines C-terminaux et d'oligomérisation des phosphoprotéines de *Paramyxoviridae* et *Rhabdoviridae*. (A) Structure du domaine C-terminal de la phosphoprotéine du virus de la rougeole responsable de la fixation du complexe N-ARN (code pdb : 1T6O). (B) Structure du domaine C-terminal de la phosphoprotéine du virus de la rage (code pdb : 1VYI). (C) Structure du domaine central responsable de l'oligomérisation de la phosphoprotéine du virus de la stomatite vésiculaire (code pdb : 2FQM). (D). Structure du domaine d'oligomérisation de la phosphoprotéine du virus de Sendai, qui forme un tétramère via un *coiled coil* (code pdb : 1EZJ).

Malgré ces faibles homologues, l'information structurale et fonctionnelle, couplée à l'analyse des séquences que nous avons réalisées suggèrent une organisation modulaire de la phosphoprotéine du virus de la rage similaire à celle proposée pour les phosphoprotéines des *Paramyxoviridae*, avec trois domaines structurés séparés de deux régions désordonnées (figure 36) :

- une région N-terminale potentiellement formée d'hélices  $\alpha$ ,
- une région N-terminale désordonnée,
- un domaine central d'oligomérisation,
- une région C-terminale désordonnée,
- un domaine C-terminal responsable de la fixation du complexe nucléoprotéine-ARN.



**Figure 36** - Schématisation de l'organisation modulaire de la phosphoprotéine de *Rhabdoviridae* en solution. Les trois domaines : N-terminal, central et C-terminal sont séparés par deux régions flexibles.

### **III – Structure par Résonance Magnétique Nucléaire (RMN) du domaine C-terminal de la phosphoprotéine du virus de la stomatite vésiculaire (VSV)**

#### **A – Introduction**

L'homologie structurale entre la nucléoprotéine du virus de la stomatite vésiculaire et celle du virus de la rage a été mise en évidence (Albertini & *al.*, 2008) (Luo & *al.*, 2007), bien que les similarités de séquences sont très faibles. L'organisation modulaire de la phosphoprotéine semble conservée au sein de *Rhabdoviridae* et notamment entre les *Vesiculovirus* et les *Lyssavirus*. Il semble donc intéressant de comparer les structures des domaines de la phosphoprotéine. Le Dr. Euripedes RIBEIRO, stagiaire postdoctoral dans le groupe, en collaboration avec Adrien FAVIER du laboratoire de Résonance Magnétique Nucléaire de l'Institut de Biologie Structurale (LRMN, IBS, Grenoble) ont déterminé par RMN la structure du domaine C-terminal de la phosphoprotéine du virus de la stomatite vésiculaire (VSV). Ma contribution à ce travail fut la construction de clones du domaine C-terminal de la phosphoprotéine de VSV.

Dans cet article, nous avons utilisé des outils biochimiques et biophysiques (SEC-MALLS-RI, dichroïsme circulaire) pour caractériser ce domaine. Ce travail a consisté en l'obtention de la structure atomique du domaine C-terminal de la phosphoprotéine du virus de la stomatite vésiculaire en utilisant la Résonance Magnétique Nucléaire (RMN). Pour la résolution de la structure, notre équipe a collaboré avec laboratoire de RMN de l'IBS.



B – Article 3: Solution structure of the C-terminal N-RNA binding domain of the vesicular stomatitis virus phosphoprotein

Auteurs: , Euripedes de Almeida Ribeiro Jr., Adrien Favier, **Francine C.A. GERARD**, Cédric Leyrat, Bernhard Brutscher, Danielle Blondel, Rob W. H. Ruigrok, Martin Blackledge et Marc Jamin.

Publié en 2008 dans le *Journal of Molecular Biology* , volume 382, pages : 525-538.

C – Données supplémentaires

# Solution Structure of the C-Terminal Nucleoprotein–RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein

Euripedes A. Ribeiro Jr<sup>1</sup>†, Adrien Favier<sup>2</sup>†, Francine C. A. Gerard<sup>1</sup>, Cédric Leyrat<sup>1</sup>, Bernhard Brutscher<sup>2</sup>, Danielle Blondel<sup>3</sup>, Rob W. H. Ruigrok<sup>1</sup>, Martin Blackledge<sup>2\*</sup> and Marc Jamin<sup>1\*</sup>

<sup>1</sup>UJF-EMBL-CNRS—UMR 5233—Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

<sup>2</sup>Institut de Biologie Structurale Jean-Pierre Ebel, CNRS, CEA, UJF, UMR 5075, 41 rue Jules Horowitz, F-38027 Grenoble, Cedex, France

<sup>3</sup>Laboratoire de Virologie Moléculaire et Structurale, UMR 2472 CNRS-1157 INRA Bât. 14B, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

Received 10 June 2008;  
accepted 7 July 2008  
Available online  
16 July 2008

Beyond common features in their genome organization and replication mechanisms, the evolutionary relationships among viruses of the *Rhabdoviridae* family are difficult to decipher because of the great variability in the amino acid sequence of their proteins. The phosphoprotein (P) of vesicular stomatitis virus (VSV) is an essential component of the RNA transcription and replication machinery; in particular, it contains binding sites for the RNA-dependent RNA polymerase and for the nucleoprotein. Here, we devised a new method for defining boundaries of structured domains from multiple disorder prediction algorithms, and we identified an autonomous folding C-terminal domain in VSV P ( $P_{CTD}$ ). We show that, like the C-terminal domain of rabies virus (RV) P, VSV  $P_{CTD}$  binds to the viral nucleocapsid (nucleoprotein–RNA complex). We solved the three-dimensional structure of VSV  $P_{CTD}$  by NMR spectroscopy and found that the topology of its polypeptide chain resembles that of RV  $P_{CTD}$ . The common part of both proteins could be superimposed with a backbone RMSD from mean atomic coordinates of 2.6 Å. VSV  $P_{CTD}$  has a shorter N-terminal helix ( $\alpha_1$ ) than RV  $P_{CTD}$ ; it lacks two  $\alpha$ -helices (helices  $\alpha_3$  and  $\alpha_6$  of RV P), and the loop between strands  $\beta_1$  and  $\beta_2$  is longer than that in RV. Dynamical properties measured by NMR relaxation revealed the presence of fast motions (below the nanosecond timescale) in loop regions (amino acids 209–214) and slower conformational exchange in the N- and C-terminal helices. Characterization of a longer construct indicated that  $P_{CTD}$  is preceded by a flexible linker. The results presented here support a modular organization of VSV P, with independent folded domains separated by flexible linkers, which is conserved among different genera of *Rhabdoviridae* and is similar to that proposed for the P proteins of the *Paramyxoviridae*.

© 2008 Elsevier Ltd. All rights reserved.

**Keywords:** phosphoprotein; vesicular stomatitis virus; rabies virus; NMR structure; virus replication

Edited by M. F. Summers

\*Corresponding authors. E-mail addresses: martin.blackledge@ibs.fr; jamin@embl.fr.

† E.A.R. and A.F. contributed equally to this work.

Abbreviations used: P, phosphoprotein; VSV, vesicular stomatitis virus; RV, rabies virus; MNV, *Mononegavirales*; N, nucleoprotein; L, RNA polymerase; NC, nucleocapsid; IND serotype, Indiana serotype; NJ serotype, New Jersey serotype; SEC, size-exclusion chromatography; MALLS, multiangle laser light scattering; RI, refractometry; HSQC, heteronuclear single quantum coherence; PDB, Protein Data Bank; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; RDC, residual dipolar coupling.

## Introduction

Vesicular stomatitis virus (VSV) and rabies virus (RV) belong to the *Rhabdoviridae* [order: *Mononegavirales* (MNV)] and have a nonsegmented negative-strand RNA molecule as genome. VSV causes infections in cattle and horses, rarely in humans, but it bears a high economical impact on the farm industry, in particular in the Americas where the virus is endemic.<sup>1</sup> For many years, VSV has served as a model system for studying the mechanism of replication of negative-strand RNA viruses. Indeed, MNV viruses vary widely in morphology and

cellular interactions, but they share a similar organization of their genomes as well as similar modes of RNA replication and transcription.<sup>2,3</sup> The viral genome (9 to 18 kb) comprises five common genes [successively coding for the nucleoprotein (N), the phosphoprotein (P), the matrix protein, the glycoprotein, and the RNA polymerase (L)] that fulfill similar roles in the life cycles of these viruses. Both VSV and RV use only three of these viral proteins, N, P, and L, to synthesize viral RNA in an efficient and regulated manner.<sup>4,5</sup> The genomic RNA is always encapsidated by N, forming a helical nucleocapsid (NC) that serves as a template for replication and transcription.<sup>6</sup> P is an essential component in the replication machineries of all *Rhabdoviridae*, and it plays multiple roles during the viral replication cycle. In the early stages of viral infection, P forms a complex with N, named the N<sup>0</sup>-P complex, which maintains N in a soluble and monomeric form, prevents the binding of N to cellular RNA, and, therefore, preserves N for encapsidating the viral RNA. In addition, P acts as a cofactor of the viral RNA-dependent L in both transcription and replication processes.<sup>5,7</sup> L carries out all the enzymatic activities but is unable to bind to NCs. P binds to both NCs and L and allows the association of the polymerase with its template.

Two stable domains of VSV P have been identified by limited proteolysis.<sup>8,9</sup> The isolated central domain (from amino acid 107 to 177) forms a dimer whose structure was solved by X-ray crystallography.<sup>10</sup> Recently, we have shown that full-length P from both VSV and RV also forms dimers in solution.<sup>11</sup> Another stable domain was found at the C-terminal extremity of P (P<sub>CTD</sub>).<sup>9</sup> Mutation and deletion studies with VSV P from both the Indiana (IND) and New Jersey (NJ) serotypes identified several functional regions within P<sub>CTD</sub>. At the N-terminal extremity of P<sub>CTD</sub>, one region (amino acids 201–210) is critical for RNA replication and another one (amino acids 191–200) is involved in virus assembly, although the exact roles of these regions are unknown.<sup>12</sup> The C-terminal part of P<sub>CTD</sub> is necessary for binding to N and in particular to N-RNA templates,<sup>9,13–16</sup> and the central region of P<sub>CTD</sub> is necessary for binding to L and contains phosphorylation sites.<sup>13,15–19</sup> No structural information about this domain of VSV P is available, but the three-dimensional (3D) structure of the C-terminal domain of RV P (RV P<sub>CTD</sub>) was solved by X-ray crystallography.<sup>20</sup> Essential residues for transcription and replication were identified by mutational studies in the homologous protein from Mokola virus and mapped onto two regions of RV P<sub>CTD</sub>, one on each face of the molecule. VSV and RV belong to two different genera within the *Rhabdoviridae* family, the *Vesiculovirus* and *Lyssavirus* genera, respectively. Multiple alignments of P from these genera, however, reveal no homology (0% identity, 4% similarity). In the *Paramyxoviridae*, another family of MNV, the C-terminal domain of P also binds to N-RNA complexes. The 3D structures of the C-terminal domain of P from viruses belonging to two different genera of *Paramyxoviridae*, measles virus

and Sendai virus, share a similar structure,<sup>21,22</sup> which is different from that of the C-terminal domain of RV P. This observation raises the question of structural conservation of P proteins within the *Rhabdoviridae* family.

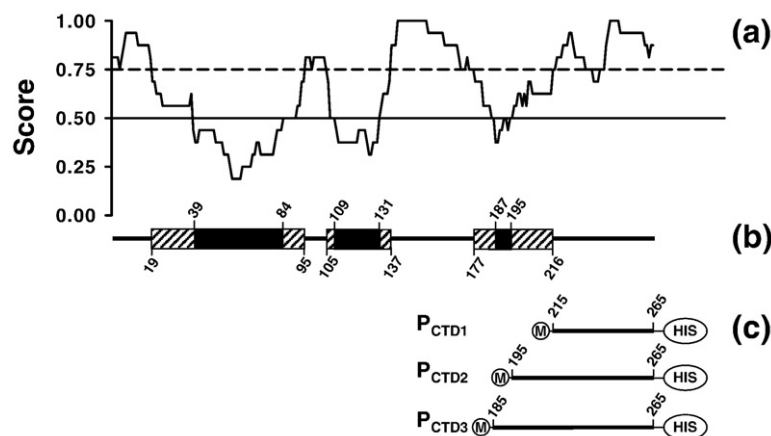
In this work, we analyzed the amino acid sequence of VSV P (IND) to localize the boundary of P<sub>CTD</sub>, and we showed that the identified C-terminal domain of P constitutes an autonomous folding unit. We determined its 3D structure and analyzed its dynamic properties by NMR spectroscopy. The structure revealed homology with the C-terminal domain of RV and is discussed in the light of previous functional data. The results presented here support a common modular organization of P proteins within the *Rhabdoviridae* family.

## Results

### The modular organization of VSV P

P proteins from the *Rhabdoviridae* are intrinsically disordered proteins that consist of structured domains alternating with disordered regions.<sup>23</sup> The location of disordered regions in the sequence of VSV P (IND) was predicted from the amino acid sequence by using 16 different algorithms available through web servers. The predictions showed large differences in the number, position, and length of the predicted disordered regions (data not shown). In order to define consensus locations for the boundaries between structured and disordered regions, we devised a scoring method based on a simple binary voting procedure that integrated the results from all disorder predictions. The calculated score (Fig. 1a) varied from 0 to 1, and, for the sake of simplicity, we arbitrarily defined two threshold levels. Residues with a score equal to or below 0.50 were considered disordered, and residues with a score equal to or above 0.75 were considered structured. Residues with a score in the intermediate zone, between 0.50 and 0.75, were considered ambiguous. For VSV P, the disorder score revealed the presence of three disordered regions (amino acids 39–84, 109–131, and 187–195) and four structured regions (amino acids 1–19, 95–105, 137–177, and 216–265) (Fig. 1b). A similar pattern of disordered and structured regions was found with the amino acid sequence of VSV NJ serotype, which is longer (274 aa) than that of the IND serotype (265 aa) and exhibits two small gaps at positions 45 and 71–73 as well as a large insertion of 14 residues between residues 195 and 196 (numbering according to the IND serotype). The main insertion in the C-terminal half of the NJ serotype was predicted to be disordered, resulting in a longer C-terminal disordered region than in the IND serotype (data not shown).

On the basis of this prediction, three different constructs were generated to locate the boundaries of a stable and structured VSV P<sub>CTD</sub> domain (Fig. 1c). The shortest construct (P<sub>CTD1</sub>: amino acids 215–265)



**Fig. 1.** (a) Modular organization of VSV P. The graph shows the normalized score for disorder as a function of residue number calculated from the results of 16 different predictors. The score was calculated by assigning, for each prediction, a value of 0 to the residues predicted to be disordered and a value of 1 to those predicted as structured, by adding these values for each residue, and by normalizing the results. Residues with a normalized score  $\leq 0.5$  are defined as consensus disordered regions; residues with a normalized score

$\geq 0.75$  are defined as consensus structured regions. (b) Consensus location of disordered regions. The scheme shows the location of consensus disordered regions with a score below 0.50 (black boxes) and with a score ranging between 0.50 and 0.75 (striped boxes). Numbers above the boxes indicate the limits of the consensus disordered regions and numbers below indicate the limits of structured regions. (c) Schematic representation of the constructs of the VSV C-terminal domain of P used in this article. Circled M represents the added N-terminal methionine, and circled HIS represents the His-tag sequence (LEHHHHHH).

was mainly found in inclusion bodies, and after purification on a  $\text{Ni}^{2+}$  column in an unfolded form (6 M urea), the protein could not be properly refolded. The intermediate construct (P<sub>CTD2</sub>: amino acids 195–265) was mainly expressed in soluble form, whereas the longest construct (P<sub>CTD3</sub>: amino acids 185–265) was found in inclusion bodies although a significant fraction was soluble. Therefore, the structural work was carried out mainly with P<sub>CTD2</sub>. It should also be noted that P<sub>CTD3</sub> obtained by refolding from inclusion bodies was undistinguishable from that obtained in the soluble fraction.

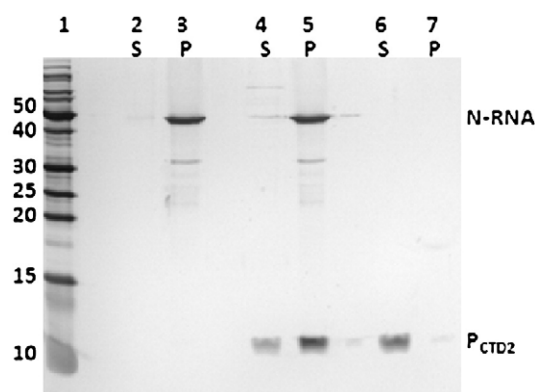
### The C-terminal domain of VSV P (P<sub>CTD</sub>) is an autonomous folding unit

The molecular masses measured by mass spectrometry for P<sub>CTD2</sub> ( $9070 \pm 2$  Da) and its  $^{15}\text{N}$ -labeled form ( $9188 \pm 2$  Da) indicated that it lacked the N-terminal methionine (molecular mass calculated from the gene without the N-terminal Met is 9075 and 9191 Da, respectively). These results were confirmed by N-terminal sequencing. Conversely, P<sub>CTD3</sub> was obtained with an N-terminal methionine, in agreement with the observed preferences of the methionyl aminopeptidase of *Escherichia coli* for the penultimate amino acid.<sup>24</sup> In P<sub>CTD2</sub>, the second amino acid is a serine, one of the most processed penultimate amino acids, whereas in P<sub>CTD3</sub>, it is an asparagine, one of the least processed.<sup>24</sup>

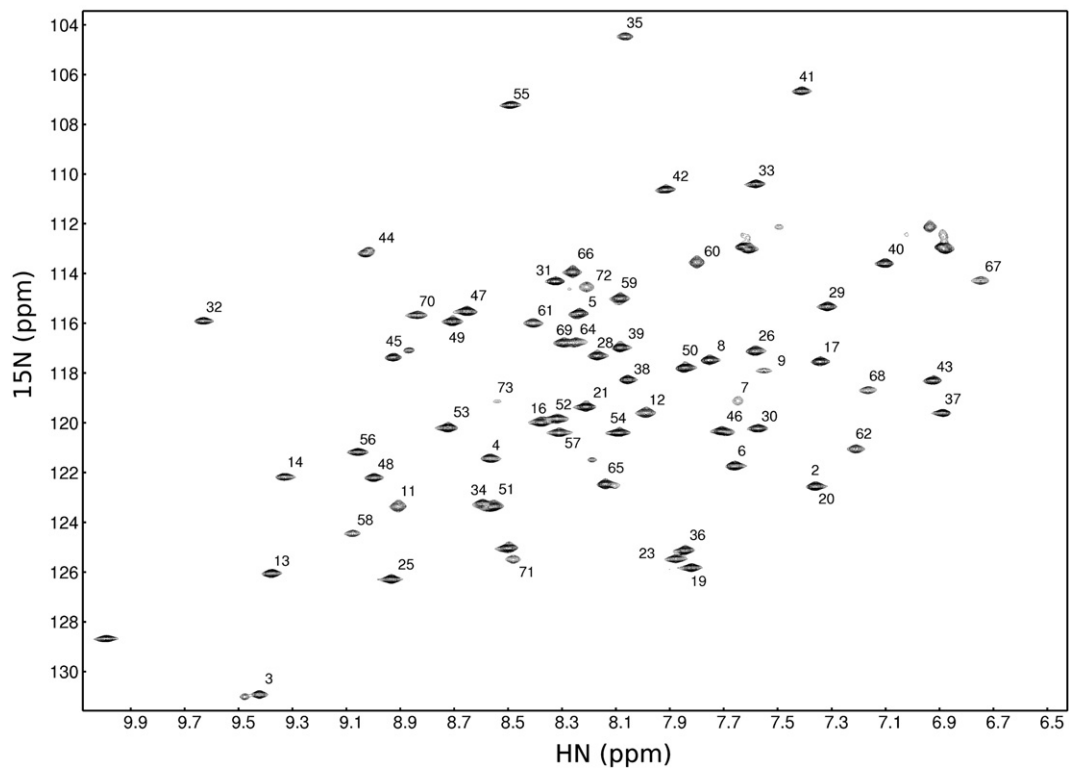
The corresponding domain of RV P binds to RV NCs.<sup>20</sup> The binding of the C-terminal domain of VSV P to viral NCs was tested by centrifugation through a glycerol cushion and SDS-PAGE. NCs isolated from virus-infected cells sedimented through the glycerol and were found in the pellet (lanes 2 and 3, Fig. 2), whereas VSV P<sub>CTD2</sub> remained in the supernatant (lanes 6 and 7, Fig. 2). When P<sub>CTD2</sub> and NCs were previously incubated, both proteins cosedimented, and the C-terminal domain

was found both in the supernatant and in the pellet (lanes 4 and 5, Fig. 2). Similar results were obtained with P<sub>CTD3</sub>, demonstrating that our fragments of VSV P conserved their ability to bind to the viral NCs, as previously shown for another construct (amino acids 192–265).<sup>9</sup>

Analysis by size-exclusion chromatography (SEC) using online detection by multiangle laser light scattering (MALLS) and refractometry (RI) indicated that P<sub>CTD2</sub> and P<sub>CTD3</sub> are monodisperse in solution. Their weight-averaged molecular masses of  $9700 \pm 800$  Da and  $11,800 \pm 300$  Da (Fig. S1a and b) showed that both proteins are monomeric (molecular mass



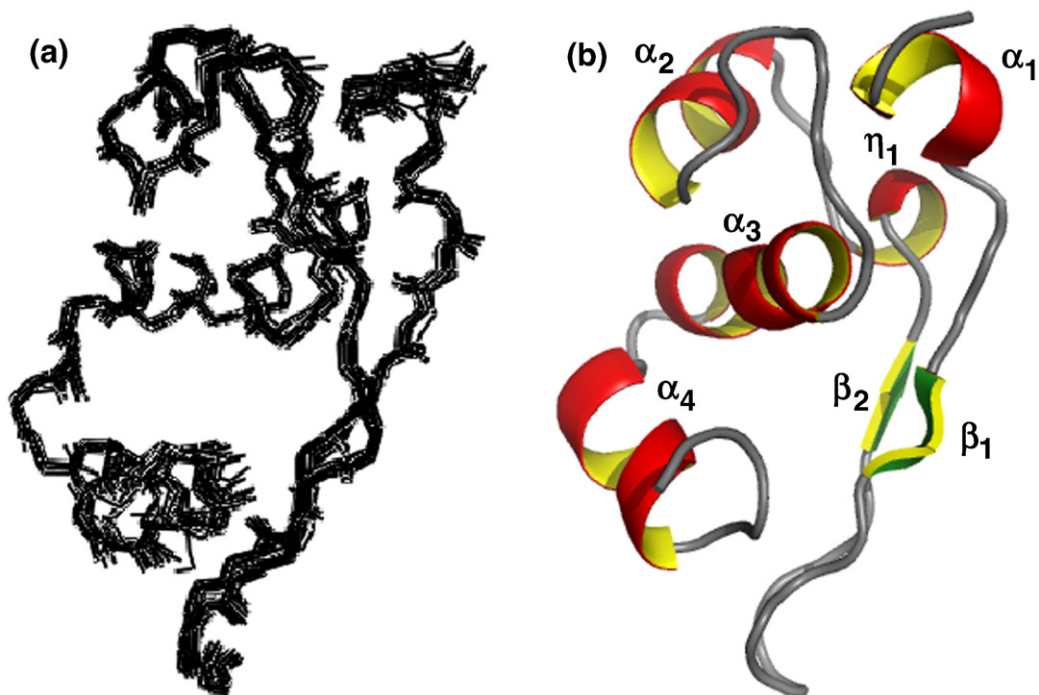
**Fig. 2.** Binding of the C-terminal domain of VSV P<sub>CTD2</sub> to NCs. NCs (N-RNA complex) purified from infected BSR cells by CsCl gradient centrifugation were incubated with P<sub>CTD2</sub> (amino acids 195–265). After centrifugation of the mixture through a 15% (v/v) glycerol cushion, samples of supernatant (S) and pellet (P) were separated by 15% SDS-PAGE and detected by silver staining. NCs were used at a final concentration of 5  $\mu\text{M}$ . Lane 1, protein standards with molecular mass indicated on the left; lanes 2 and 3, N-RNA alone; lanes 4 and 5, N-RNA:P<sub>CTD2</sub> in a 1:1 molar ratio; lanes 6 and 7, P<sub>CTD2</sub> alone.



**Fig. 3.** The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of  $\text{P}_{\text{CTD}2}$ . The spectrum was measured at 800 MHz and at 20 °C. Residues are numbered  $\text{P}_{\text{CTD}2}$  starting at amino acid 195.

calculated from the gene is 9075 and 10,258 Da, respectively) as previously suggested by SEC analysis,<sup>9</sup> in contrast to full-length VSV P that

forms dimers.<sup>11</sup> Also, a value of  $9700 \pm 500$  Da was obtained for  $\text{P}_{\text{CTD}2}$  from small-angle neutron scattering measurements (data not shown).



**Fig. 4.** NMR structure of  $\text{P}_{\text{CTD}2}$ . (a) Ensemble of the 20 best structures calculated from NMR data (PDB accession number: 2k47). (b) Cartoon representation of the lowest-energy NMR structure of VSV  $\text{P}_{\text{CTD}2}$ . Secondary-structure elements are colored and labeled. The figure was prepared with the PyMOL program.

Recombinant P<sub>CTD2</sub> and P<sub>CTD3</sub> were folded as shown by far-UV circular dichroism (CD) (Fig. S2a) and by 1D NMR<sup>25</sup> (data not shown) and exhibited a similar amount of secondary structure although P<sub>CTD3</sub> contains 10 additional residues. Figure 3 shows the <sup>1</sup>H-<sup>15</sup>N heteronuclear single quantum coherence (HSQC) spectrum of P<sub>CTD2</sub> at pH 7.5 and 20 °C. In these conditions, both P<sub>CTD2</sub> and P<sub>CTD3</sub>, prepared as described above, were stable according to HSQC spectra over a period of 6 weeks, and thermal denaturation of both proteins monitored by CD at 222 nm was fully reversible (Fig. S2b). The protein denatured in a single cooperative transition with a *T<sub>m</sub>* near 45 °C for P<sub>CTD2</sub> and near 47 °C for P<sub>CTD3</sub>. The far-UV CD spectra recorded at 25 °C after one cycle of denaturation/renaturation were superimposed to the initial spectra obtained at 25 °C (Fig. 3a).

## NMR structure

### VSV P<sub>CTD</sub> resembles the structure of RV P<sub>CTD</sub>

The 3D structural ensemble defining P<sub>CTD2</sub> is shown in Fig. 4a, and the associated statistical analysis is presented in Table 1. VSV P<sub>CTD2</sub> forms a single compact domain that has the shape of a triangular prism. It is composed of four  $\alpha$ -helices ( $\alpha_1$ - $\alpha_4$ ), one 3<sup>10</sup> helix ( $\eta_1$ ), and two short  $\beta$ -strands ( $\beta_1$ - $\beta_2$ ) that form an antiparallel  $\beta$ -sheet (Fig. 4b). A search of the Protein Data Bank (PDB) using the programs DALI<sup>26</sup> or SSM (secondary-structure matching)<sup>27</sup> revealed no hit with a Z score higher than 2.5 and did not detect the P<sub>CTD</sub> of RV.<sup>20</sup> However, structural alignment with RV P<sub>CTD</sub> using the SSM program<sup>27</sup> gave a Z score of 2.3 and an RMSD value of 2.6 Å over the 44 aligned residues. Figure 5a and b show the superposition of both structures. The short  $\alpha_1$  helix, the two  $\beta$ -strands  $\beta_1$  and  $\beta_2$ , the 3<sup>10</sup> helix  $\eta_1$ , and the helices  $\alpha_2$ ,  $\alpha_3$ , and  $\alpha_4$  of VSV P<sub>CTD</sub> match with corresponding secondary-structure elements in RV P<sub>CTD</sub> ( $\alpha_1$ ,  $\beta_1$ ,  $\beta_2$ ,  $\eta_1$ ,  $\alpha_2$ ,  $\alpha_4$ , and  $\alpha_5$ , respectively). Comparison of the topology of the peptide chain between VSV and RV P<sub>CTD</sub> indicated a similar organization in both proteins (Fig. 5c and d). Helices  $\alpha_3$  and  $\alpha_6$  of RV P are missing in VSV P, helices  $\alpha_1$  and  $\alpha_2$  are shorter in VSV than in RV P, while the loop connecting  $\beta_1$  and  $\beta_2$  is longer in VSV than in RV P (Fig. 5e).

### NMR spectrum of P<sub>CTD3</sub>

The HSQC spectrum obtained for P<sub>CTD3</sub> was similar to that obtained for P<sub>CTD2</sub> (Fig. S3a). Only three additional resonances were visible in the amide backbone region, with <sup>1</sup>H chemical shifts in the range associated with disordered structure. The seven other additional residues did not give rise to observable amide backbone resonances, indicating the presence of intermediate conformational exchange, while N-terminal sequencing performed on the sample used for the NMR experiments confirmed the integrity of the protein from its N-terminal methionine residue. Some resonances in the spectrum of P<sub>CTD2</sub> corresponding to residues 197 and

**Table 1.** NMR structure parameters, data statistics,<sup>a</sup> and details of structure determination

<i>Number of experimental distance constraints</i>	
All	1091
Intraresidual	227
Sequential ( $ i-j =1$ )	313
Short range ( $1 <  i-j  \leq 5$ )	231
Long range ( $ i-j  > 5$ )	320
<i>RMSDs from experimental distance constraints (Å)</i>	
NOEs	0.14±0.03
<i>Number of NOE violations</i>	
Violations >0.5 Å	14±4
Violations >0.3 Å	32±8
Violations >0.1 Å	75±17
<i>RMSDs from experimental dihedral constraints (°)<sup>b</sup></i>	
$\phi/\psi$	(68) 0.19±0.05
<i>Reduced chi-square (<math>\chi^2/N</math>) from experimental RDC restraints<sup>b</sup></i>	
C <sup><math>\alpha</math></sup> -H <sup><math>\alpha</math></sup>	(30) 2.1±0.2
N-H <sup>N</sup>	(40) 1.0±0.1
C <sup><math>\alpha</math></sup> -C'	(36) 0.7±0.1
<i>Alignment tensor eigenvalues</i>	
D <sub>a</sub>	13.0±0.2×10 <sup>-4</sup>
D <sub>r</sub>	3.3±0.1×10 <sup>-4</sup>
<i>Deviation from the idealized covalent geometry</i>	
Bonds (Å)	0.0039±0.0001
Angles (°)	0.48±0.006
Improper (°)	0.42±0.01
<i>Energies (kcal/mol)</i>	
Total	-2430±85
van der Waals	-691±8
Electrostatic	-2217±82
<i>Coordinate precision (Å)</i>	
Backbone secondary structure <sup>c</sup>	0.27±0.06
Heavy atoms secondary structure	0.88±0.09
Backbone	0.43±0.09
Heavy atoms	0.93±0.11
<i>Percentage of non-Gly and non-Pro residues in Ramachandran regions<sup>d</sup></i>	
Core	77.7
Allowed	21.2
Generous	1.2
Disallowed	0.0

<sup>a</sup> The statistics were obtained from a subset of the 20 best structures following the protocol described in Materials and Methods. The number of the various constraints is given in parentheses.

<sup>b</sup> The dihedral angle constraints are inferred from the program TALOS (reference).

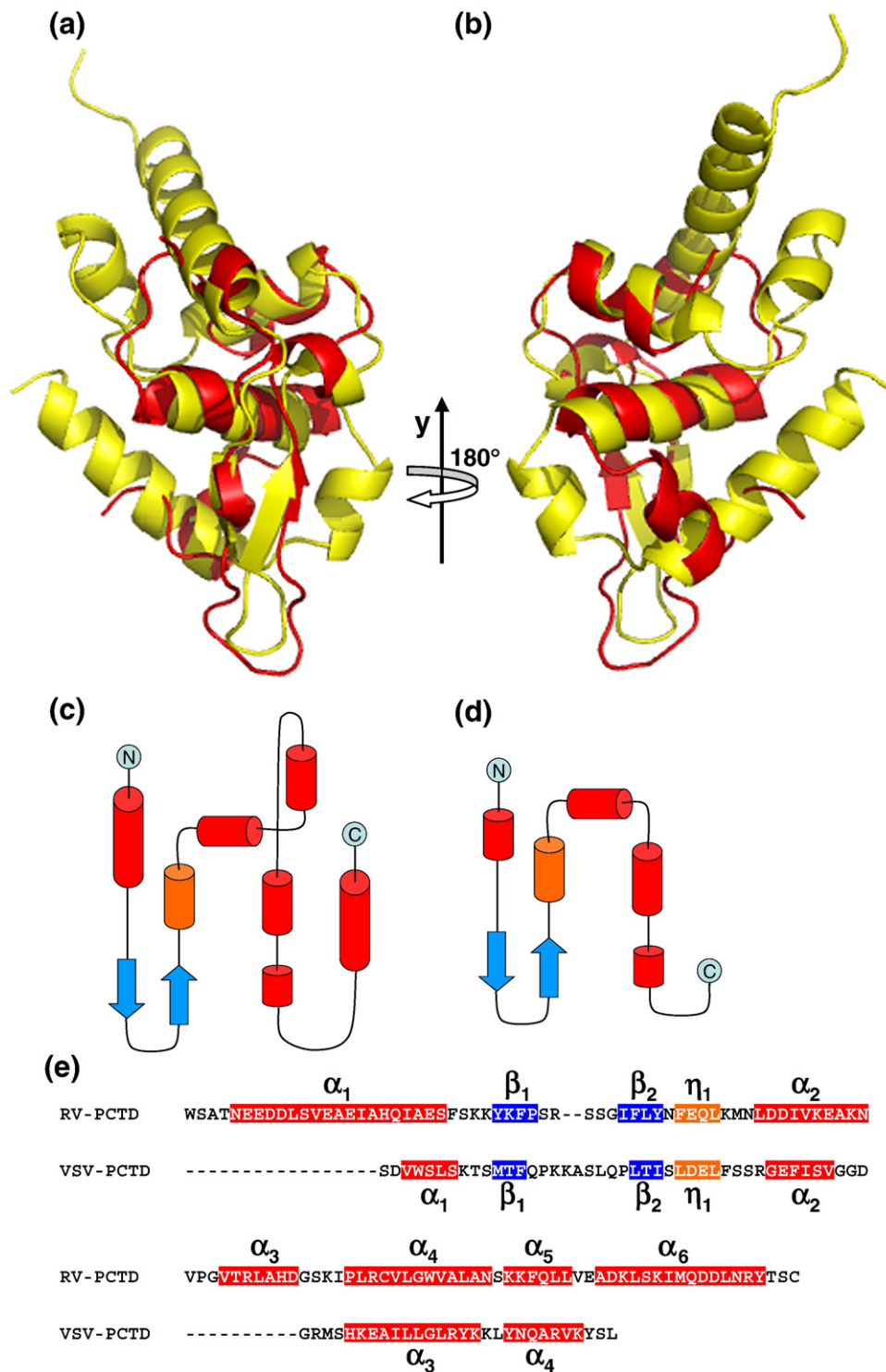
<sup>c</sup> Secondary structural elements defined as (10–13), (23–26), (28–30), (34–41), (49–59), and (65–68).

<sup>d</sup> These values are calculated with the program PROCHECK-NMR.

198 in helix  $\alpha_1$  and to residues 237 to 240 in the loop connecting helices  $\alpha_2$  and  $\alpha_3$  were slightly shifted in P<sub>CTD3</sub> spectrum. These residues are the closest in space to the N-terminal extension in P<sub>CTD3</sub> and the most susceptible to be affected by the presence of an additional flexible N-terminal extension (Fig. S3b).

### Conformational dynamics

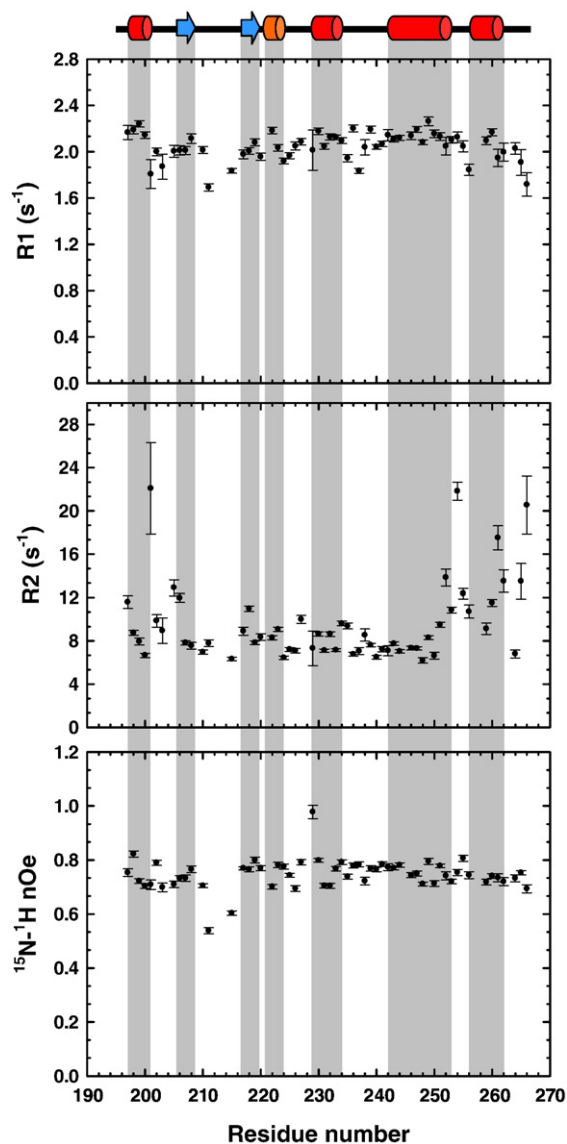
NMR spin relaxation is sensitive to motions of the associated nuclei and can be used to probe local and



**Fig. 5.** Comparison of VSV P<sub>CTD</sub> and RV P<sub>CTD</sub>. (a and b) Structural alignments. In yellow, ribbon representation of RV P<sub>CTD</sub> (PDB accession number: 1VYI); in red, ribbon representation of VSV P<sub>CTD2</sub>. The RMSD is 2.6 Å for the 44 superposed residues. (b) shows the structure rotated by 180° around the *y*-axis. (c and d) Topology of RV (c) and VSV (d) P<sub>CTD</sub>. (e) Alignment of the amino acid sequences of RV and VSV P<sub>CTD</sub> based on structural alignment.

global protein dynamics. The relaxation rates of longitudinal and transverse magnetization of backbone <sup>15</sup>N nuclei (R1 and R2), as well as <sup>1</sup>H-<sup>15</sup>N steady-state heteronuclear nuclear Overhauser enhancement (NOE), are sensitive to motions on a picosecond to nanosecond timescale. We present

experimental relaxation rates rather than a Lipari-Szabo analysis<sup>28</sup> because the apparently significant exchange contributions in the terminal helices, combined with restricted angular sampling in the secondary-structure elements of the molecule, preclude an accurate analysis of the rotational diffusion



**Fig. 6.** Dynamics of  $P_{CTD2}$ . Relaxation rates of longitudinal,  $R1$  (a), and transverse,  $R2$  (b), magnetization of backbone  $^{15}N$  nuclei and  $^1H$ - $^{15}N$  heteronuclear NOE (c). Secondary-structure elements are represented above the plots and by gray shades within the plots.

tensor. The  $R1$  and  $R2$  relaxation rates and heteronuclear NOEs are shown in Fig. 6 for each backbone amide in VSV  $P_{CTD2}$ . In addition to enhanced flexibility in the region of residues 209 to 215 as indicated from the heteronuclear NOE experiment, transverse relaxation rates were most probably affected by conformational exchange on the microsecond to millisecond timescale in the N-terminal (amino acids 195–206) and in particular in the C-terminal (amino acids 251–265) helical region. The protein used for NMR spectroscopy contained the His-tag, but no resonances were observed for this part of the protein, also suggesting a conformational exchange on the microsecond to millisecond timescale. In a similar study of the dynamic behavior of VSV  $P_{CTD3}$ , essentially identical heteronuclear NOE

values were measured throughout the protein for the common regions, indicating that the additional 10 amino acids have a negligible effect on the stability of the protein on the observed timescales. For the three additional amide backbone resonances observed in the NMR spectrum of  $P_{CTD3}$ , one had a heteronuclear NOE value near zero and the other two exhibited negative values, indicating an important flexibility in this region of the protein.

## Discussion

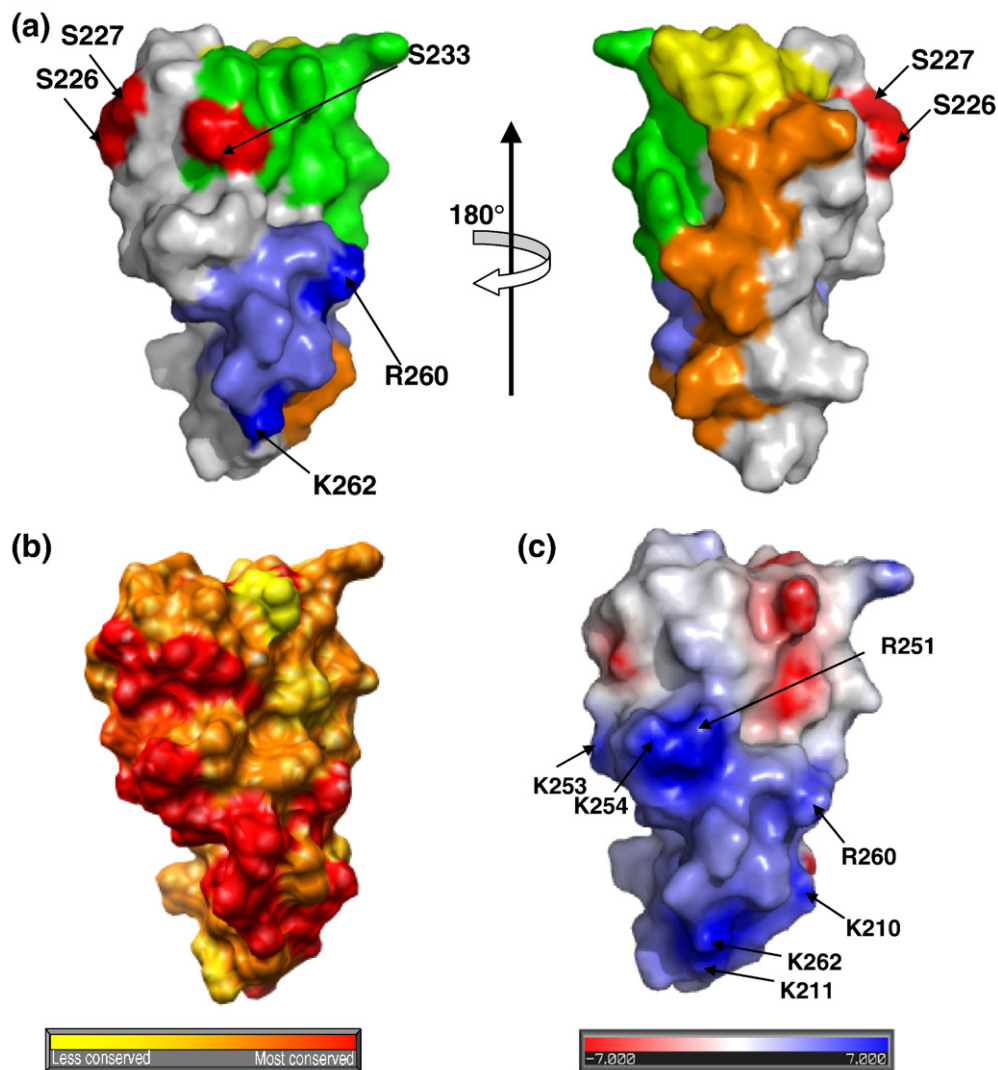
### Modular organization of P

Two observations reported here support a modular organization of VSV P. Firstly, in a modular protein, independent domains are expected to fold in an autonomous process, and we found that recombinant VSV  $P_{CTD}$  folds in *E. coli* and that its unfolding is fully reversible. The shortest construct ( $P_{CTD1}$ ), lacking helix  $\alpha_1$  and strand  $\beta_1$ , could not be expressed as a soluble protein and could not be refolded. Also, the central dimerization domain folds independently of the rest of the protein.<sup>10</sup> Secondly, these two autonomous folding units seem to be attached by a flexible linker. The short region preceding helix  $\alpha_1$  in  $P_{CTD3}$  was predicted to be disordered by our scoring method (Fig. 1a), and NMR spectroscopy confirmed that this region of P is flexible. Three additional amide backbone resonances were visible in the 2D HSQC NMR spectrum obtained with  $P_{CTD3}$  as compared to the spectrum obtained with  $P_{CTD2}$ . Heteronuclear NOEs for these three residues had null (one residue) or negative values (two residues), indicating a high degree of flexibility. The absence of peaks for the other residues probably resulted from conformational exchange on a timescale similar to that of the NMR experiment (microseconds to milliseconds). In P proteins from members of the *Paramyxoviridae* family, the C-terminal N–RNA binding domain pX is also separated from the central tetramerization domain by a flexible linker,<sup>21,29</sup> indicating that the mobility of  $P_{CTD}$  is conserved between different MNV families and suggesting that it is required for its biological function. Also, it is noteworthy to mention that a longer disordered region, separating the central dimerization domain from the N–RNA binding C-terminal domain, is predicted in VSV P from the NJ serotype.

### Mapping of the functional regions within VSV $P_{CTD}$

VSV P is a multifunctional protein and its C-terminal domain ( $P_{CTD}$ ) participates in different processes during viral replication. Mutation and deletion studies have been used to identify amino acids or regions that are involved in the different activities of  $P_{CTD}$  (Fig. 7a). This domain binds to N in the NC and to L<sup>14,16,17,30</sup> but was also shown to participate in virus assembly, possibly by interacting with other viral or cellular proteins.<sup>12</sup> The last 21 C-terminal residues of P (amino acids 235–265) are





**Fig. 7.** Mapping of functional residues, electrostatic surface potential, and amino acid conservation on the surface of VSV P<sub>CTD2</sub>. (a) Functional mapping. Surface representations of P<sub>CTD2</sub> showing the residues involved in virus assembly in yellow (amino acids 195–200), those involved in RNA replication in orange (amino acids 201–210), those involved in binding to L in green (amino acids 231–244), those involved in binding to N in blue (amino acids 255–265 with Arg260 and Lys262 in dark blue), and the phosphorylation sites in red (Ser226, Ser227, and Ser233). (b) Amino acid sequence conservation among VSV strains. The conservation score calculated with the AL2CO program is mapped on the surface of P<sub>CTD2</sub> using the Chimera program, with red denoting the most conserved residues and yellow indicating the least conserved residues. (c) Electrostatic surface potential. The surface potential was calculated with the Delphi program and is color-coded on the surface from red (negatively charged residues;  $-12 k_B T$ ) to blue (positively charged residues;  $+12 k_B T$ ).

required for transcriptional activity and for the binding to NC templates.<sup>16,17,30</sup> This interacting region was narrowed to the last 10 or 11 C-terminal residues.<sup>16,30</sup> Deletion of the last 11 C-terminal residues of VSV P almost completely abrogated viral transcription and replication and significantly reduced the binding to N-RNA complexes, to soluble N, and to L<sup>16</sup> (amino acids 255 to 265 are shown in blue in Fig. 7a). Two critical basic residues, Arg260 and Lys262, were identified in this C-terminal region. Mutation of Arg260, a residue strictly conserved among all *Vesiculoviruses*, is sufficient for significantly reducing the binding of P to N-RNA, while mutation of Lys262 has almost

no effect. However, mutations of Arg260 or of Lys262 have a strong effect on *in vitro* transcription and on binding to L but has no effect on RNA replication or on binding to soluble N.<sup>16</sup> These results underline the importance of C-terminal residues in binding NCs and L. Another region, ranging from residues 221 to 234, is phosphorylated by an L protein-associated kinase<sup>31</sup> (phosphorylation sites are shown in red in Fig. 7a) and is involved in the binding of P to the L protein.<sup>13,17</sup> Two regions of P were identified as binding sites for the polymerase, one in the N-terminal half of P<sup>15</sup> and another in the C-terminal domain.<sup>13,17</sup> The L binding site in P<sub>CTD</sub> was assigned to a region of 14 amino

acids ranging from residue 231 to residue 244<sup>17</sup> (shown in green in Fig. 7a). Phosphorylation of VSV P is carried out by two protein kinases acting sequentially.<sup>32</sup> The casein kinase II regulates transcriptional activity by phosphorylating residues in the N-terminal half of P, whereas an L-polymerase-associated kinase phosphorylates additional sites in P<sub>CTD</sub>. In the IND serotype, important phosphorylation sites in P<sub>CTD</sub> were mapped to Ser226, Ser227, and Ser233<sup>18,19</sup> (shown in red in Fig. 7a), but the role of these secondary phosphorylations is unclear. It has been proposed that hyperphosphorylation of P could lead to the arrest of RNA replication. These residues precede or are part of the L binding site and could thus directly affect the binding to L. Finally, deletions in the N-terminal part of P<sub>CTD</sub> affect virus assembly (amino acids 191 to 200, shown in yellow in Fig. 7a) and suppress RNA replication (amino acids 201 to 210, shown in orange in Fig. 7a) without preventing binding to N and L.<sup>12</sup>

The identification of functional regions in P<sub>CTD</sub> relied mainly on deletion studies. P<sub>CTD</sub> is a small domain of 71 residues, and thus, truncation of large fragments might have major impacts on its structure. Mutants with large deletions in P<sub>CTD</sub>, however, conserve their ability to bind to their partners, suggesting that binding occurs through recognition of sequence patterns in P<sub>CTD</sub> rather than through recognition of 3D motifs. Different parts of VSV P<sub>CTD</sub>, including the C-terminal positively charged region, are highly flexible as seen by NMR spectroscopy. This feature of P<sub>CTD</sub> might explain our failure to find crystallization conditions through the use of different crystallization screens (see Materials and Methods) and might be important for the binding to its partners.

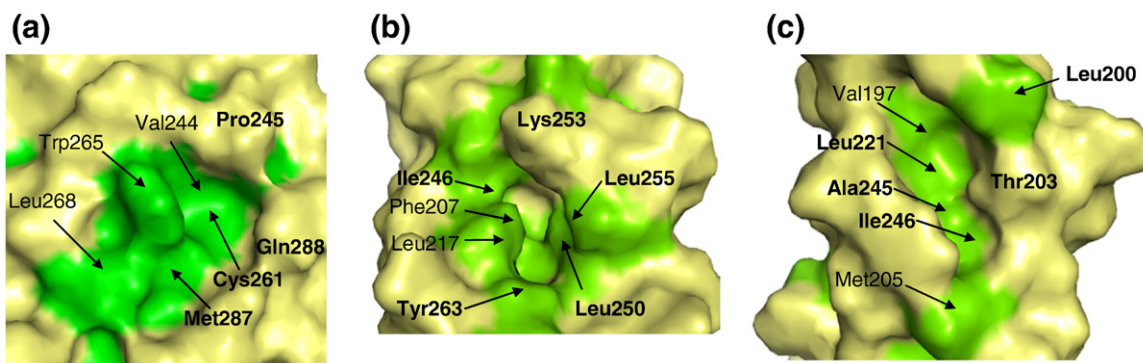
### Structure conservation within the *Rhabdoviridae*

VSV and RV belong to two different genera of *Rhabdoviridae*, the *Vesiculoviruses* and the *Lyssaviruses*, but multiple sequence alignments indicate almost no sequence similarity between these two P proteins. However, P proteins from these two viruses fulfill similar functions and seem to share a

common modular organization. The C-terminal domain in VSV P is shorter than the C-terminal domain of RV P, but the structure of VSV P<sub>CTD</sub> determined here by NMR spectroscopy is related to that of RV P. The topology of the backbone is conserved, but VSV P<sub>CTD</sub> lacks helices  $\alpha_3$  and  $\alpha_6$  of RV P<sub>CTD</sub> (Fig. 5). This new structure suggests that the C-terminal domain and its associated biological functions are conserved among the *Rhabdoviridae*.

The most conserved residues identified through multiple amino acid sequence alignments of all members of the *Vesiculoviruses*, including Isfahan, Piry, and Chandipura viruses, were found in the main hydrophobic core of P<sub>CTD</sub>, with the largest number of conserved residues located in helix  $\alpha_3$ . Restricting sequence comparison to the different VSV strains of both IND and NJ serotypes revealed additional conserved residues mainly in helices  $\alpha_2$  and  $\alpha_4$ . On the surface of the protein, the conserved residues are located in the C-terminal part of helix  $\alpha_3$ , in helix  $\alpha_4$ , and in the short C-terminal loop. They form an elongated patch along one ridge of the molecule and line two large cavities (Fig. 7b). In RV P<sub>CTD</sub>, most of the residues conserved among *Lyssaviruses* map on the rounded face of the domain (Fig. S4a).

Residues of P<sub>CTD</sub> from Mokola virus involved in the binding to N were identified by mutagenesis using a two-hybrid system.<sup>33</sup> P protein from Mokola virus is homologous to RV P, and mapping of these residues on the surface of RV P<sub>CTD</sub> revealed the presence of two major clusters on opposite faces of the domain.<sup>20</sup> The distribution of electrostatic potentials on the surface of RV P<sub>CTD</sub> revealed a bipolar distribution of charges with a positive pole and a negative pole<sup>20</sup> (Fig. S4b). The first cluster is located in the positive pole of the rounded face and comprises conserved positively charged residues (K211, K212, and K214) and Leu224. The second cluster, located on the opposite flat face of P<sub>CTD</sub>, forms a hole on the surface of the protein and comprises Cys261, Trp265, and Met287 (Figs. 8a and Fig. S5a). For VSV P<sub>CTD</sub>, analysis of the distribution of electrostatic potentials (Fig. 7c) also shows a bipolar distribution of charges on the surface of VSV



**Fig. 8.** View of the conserved hydrophobic clusters in RV P<sub>CTD</sub> (a) and VSV P<sub>CTD</sub> (b and c). Hydrophobic side chains are shown in green and are labeled. The bold labels highlight the residues conserved in *Lyssaviruses* (a) or in VSV strains (b and c).

$P_{CTD}$  with a pole of positive charges and a pole containing equivalent numbers of positive and negative charges. The positively charged side chains of Lys210, Lys211, Arg251, Lys253, Lys254, Arg260, and Lys262 form a cluster on one face of  $P_{CTD}$ . Among these basic residues, Lys210, Lys253, Lys254, Arg260, and Lys262 are strictly conserved among the different VSV strains, while Lys211 and Arg243 found in IND serotypes are replaced by Arg and Lys residues in NJ serotypes, respectively. Arg260 is the only basic residue that is strictly conserved among all *Vesiculoviruses*.

Two deep hydrophobic cavities are present on the surface of VSV  $P_{CTD}$  and comprise conserved residues. The first cavity is located between helices  $\alpha_3$  and  $\alpha_4$  and the  $\beta$ -hairpin and is lined by numerous hydrophobic residues (Phe207, Leu217, Ile246, Leu250, Tyr263, and Leu265), of which some are conserved among VSV strains (Figs. 8b and Fig. S5b). This large pocket (1.2 nm long and 0.6 nm wide) is located on a different face of  $P_{CTD}$  compared to the cluster of basic residues and might correspond to the hydrophobic cluster found on the surface of RV  $P_{CTD}$  (Figs. 8a and Fig. S5a). The second cavity in VSV  $P_{CTD}$  is a long groove (1.1 nm long and 0.5 nm wide) located between helix  $\alpha_1$  and the loop connecting helices  $\alpha_2$  and  $\alpha_3$ . Its bottom comprises hydrophobic and conserved residues (Figs. 8c and Fig. S5c). These deep cavities at the surface of  $P_{CTD}$  might represent binding sites for viral or cellular partners. Deletion studies revealed that three amino acids at the C-terminal extremity of N, including a valine and a phenylalanine, are critical for the association with P.<sup>14</sup> Alternatively, cavities at the surface of  $P_{CTD}$  might result from the intrinsic flexibility of this protein domain and indicate that binding to partners involves significant conformational changes in  $P_{CTD}$ .

The presence of clusters of conserved basic and hydrophobic residues on opposite faces of VSV and RV  $P_{CTD}$  supports a conservation of functional sites on the surface of the protein. It will now be interesting to compare the structure of the central oligomerization domain of RV with that determined for VSV<sup>10</sup> and also to characterize the putative structure of the N-terminal domain of both proteins alone or in complex with N<sup>0</sup> to find out whether the different domains were conserved through evolution. Understanding the origin and evolution of RNA viruses is complicated by the high rate of mutation and specific adaptation to host cells.<sup>34–37</sup> The comparison of genome or protein sequences can only probe closely related viruses, and a more thorough analysis awaits accumulation of structural information on both viral and cellular proteins.

## Materials and Methods

### Amino acid sequence analysis

The location of disordered regions within VSV P was predicted by submitting its amino acid sequence to 16

different algorithms accessible through web servers. FoldIndex was run with a window size of 51.<sup>38</sup> With the different PONDR predictors (XL3, VL-XT, XL1-XT, and VSL1), residues with a score higher than 0.5 were considered as disordered.<sup>39,40</sup> The NORSp server was used with a window size for disorder set to 25 residues, the structure content cutoff set to 12%, and the number of consecutive exposed residues set to 10.<sup>41</sup> The IUPred predictor for long disordered regions was used and residues with a score above 0.5 were considered as disordered.<sup>42</sup> The DISOPRED server was run with the rate threshold for predictions of false positive set at 2%.<sup>43</sup> The DisEMBL server (Loops/Coils, Hot-loops, and Remark465),<sup>44</sup> RONN server,<sup>45</sup> and PreLink server<sup>46</sup> were run using the default parameters. The SEG program was used for analyzing sequence complexity.<sup>47</sup> Hydrophobic cluster analysis plots were drawn with DRAWHCA,<sup>48</sup> and disordered regions of more than 20 residues were identified by visual inspection. A simple scoring procedure was used to define consensus disordered regions. For each prediction, residues were simply ranked in a binary manner as ordered or disordered using the default threshold set for each algorithm in the web servers. The statistical weights yielded by some of these algorithms are not accounted for, since they are not available for each predicting method and are difficult to compare. For each algorithm, a score of 0 is attributed to each residue predicted to be disordered, whereas a score of 1 is attributed to each one predicted to be in a structured region. The values for each residue are added and divided by the number of used algorithms in order to normalize the score. Consensus disordered regions were defined as regions with a normalized score  $\leq 0.50$ .

### Cloning, production, and purification of VSV P and its C-terminal domain

Three constructs of the C-terminal domain of VSV P,  $P_{CTD1}$  (amino acids 215–265),  $P_{CTD2}$  (amino acids 195–265), and  $P_{CTD3}$  (amino acids 185–265) (Fig. 1c), were obtained by PCR amplification from the full-length gene. For each construct, an NdeI restriction site and a methionine codon were introduced downstream of the coding sequence and an XhoI restriction site was introduced upstream. The PCR products were cloned into pET22b between the NdeI and XhoI sites, thus introducing a short linker of two amino acids (Leu-Glu) and a His<sub>6</sub>-tag at the C-terminal extremity of the coding sequence. The constructions were verified by DNA sequencing. The proteins were expressed in *E. coli* BL21(DE3). <sup>15</sup>N and <sup>13</sup>C protein samples for NMR spectroscopy were produced in M9 minimum medium supplemented with 1.0 g L<sup>-1</sup> of <sup>15</sup>NH<sub>4</sub>Cl, 2.0 g L<sup>-1</sup> of <sup>13</sup>C glucose, and MEM vitamins (Gibco).

Soluble proteins were purified in 20 mM Tris-HCl and 150 mM NaCl, pH 7.5 (Buffer A), in a two-step procedure using a Ni<sup>2+</sup> resin column and a SEC column (Superdex 75 HR 10/30, GE Healthcare) as previously described for full-length VSV P.<sup>11</sup> Inclusion bodies were washed three times by suspending the pellet in Buffer A and centrifuging at 27,200g for 30 min. The pellet was then solubilized with 20 mM Tris-HCl, pH 7.5, containing 8 M urea (Buffer B), and the solution was centrifuged at 48,000g for 30 min to remove insoluble material. The supernatant was loaded onto Ni<sup>2+</sup> resin column, pre-equilibrated with Buffer B. The resin was washed successively with 1 bed volume of Buffer B, with 3 bed volumes of 20 mM Tris-HCl at pH 7.5 to refold the protein, with 3 bed volumes of 20 mM Tris-HCl and 1 M NaCl at pH 7.5 supplemented with 10 mM imidazole,

and, finally, with 3 bed volumes of 20 mM Tris-HCl at pH 7.5. The protein was eluted using 20 mM Tris-HCl and 400 mM imidazole at pH 7.5 and loaded onto a SEC column (Superdex 75 HR 10/30, Amersham Pharmacia) equilibrated with 20 mM Tris-HCl, pH 7.5. The separation was performed at a flow rate of 0.5 mL min<sup>-1</sup>.

The protein preparations were checked by SDS-PAGE, by electrospray mass spectrometry, and by N-terminal sequencing. Mass spectrometry experiments were performed on a Quattro II Mass Spectrometer (Micromass, Altrincham, UK) by continuously injecting the sample with a Type 22 pump from Harvard Apparatus at a flow rate of 5  $\mu$ L min<sup>-1</sup>. Proteins were diluted in H<sub>2</sub>O containing 1% formic acid at a concentration of about 1 mg mL<sup>-1</sup>. N-terminal sequencing was performed by Edman degradation on a Procise Model 492 protein sequencer (Applied Biosystems). All protein concentrations were measured by absorbance spectroscopy using Edelhoch's method<sup>49</sup> and the calculated extinction coefficient ( $\epsilon_{280} = 9540 \text{ M}^{-1} \text{ cm}^{-1}$ ).

Solutions of purified P<sub>CTD2</sub> at different concentrations ranging from 4.0 to 19.0 mg mL<sup>-1</sup> were submitted to extensive crystallization trials with a PixSys4200 robot (Cartesian) using crystal screens I and II and Index (Hampton research) and the vapor diffusion method in drops of 0.1  $\mu$ L protein solution and 0.1  $\mu$ L reservoir solution, but no protein crystals were produced.

### CD spectroscopy

Far-UV CD spectra were recorded at 20 °C on a JASCO model J-810 CD spectropolarimeter equipped with a Peltier temperature controller. VSV P<sub>CTD</sub> was diluted to final concentrations of about 40  $\mu$ M, in 20 mM Tris-HCl, pH 7.5, containing 150 mM NaCl. Spectra were measured in a cuvette with a path length of 1 mm. Mean residue ellipticities were calculated using the number of residues [P<sub>CTD2</sub>: 71 residues + linker (2 aa) + His-tag (6 aa); P<sub>CTD3</sub>: 82 residues + linker (2 aa) + His-tag (6 aa)].

The temperature-induced unfolding and refolding were recorded by CD at 222 nm by varying the temperature between 25 and 60 °C using a scan rate of 60 °C/h. VSV P<sub>CTD</sub> was diluted in 2 mM Tris-HCl, pH 7.5. Protein concentration was 5  $\mu$ M and path length was 10 mm.

### SEC-MALLS-RI

SEC was performed with a S200 Superdex column (GE Healthcare) equilibrated with 20 mM Tris-HCl and 150 mM NaCl, pH 7.5. Separations were performed at 20 °C with a flow rate of 0.5 mL min<sup>-1</sup>. Typically, 20  $\mu$ L of protein solution at concentrations of 5–10 mg mL<sup>-1</sup> was injected. Online MALLS detection and refractive index measurements were performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting at 690 nm and with a RI2000 detector (Schambeck SFD), respectively. Data were analyzed and weight-averaged molar masses ( $M_w$ ) were calculated using the ASTRA software (Wyatt Technology Corp.) as described previously.<sup>11</sup>

### NC binding assay

The interaction between viral NC and P<sub>CTD2</sub> was investigated by centrifugation through a glycerol cushion. VSV NC was purified from infected BSR cells as described previously.<sup>50</sup> In this assay, VSV NC was incubated in the

presence of P<sub>CTD2</sub> at 20 °C for 10 min. The final NC concentration was 5  $\mu$ M in 20 mM Tris-HCl, pH 7.5, containing 50 mM NaCl, and P<sub>CTD2</sub> was added to NC at a 1:1 N:P ratio. These mixtures (40  $\mu$ L) were loaded on top of 400  $\mu$ L of 20 mM Tris-HCl, pH 7.5, containing 15% (v/v) glycerol and centrifuged at 25,000 rpm for 2 h at 20 °C (SW55Ti rotor using Ultra Clear tubes of 0.8 mL). After centrifugation, 40  $\mu$ L of sample present at the top of the glycerol cushion was recovered and the pellet was suspended in 40  $\mu$ L of 20 mM Tris-HCl, pH 7.5. These fractions were analyzed by 15% SDS-PAGE. Five microliters of each sample was loaded on the gel, and proteins were detected by silver staining.

### NMR spectroscopy

Automated sequential backbone resonance assignment of P<sub>CTD2</sub> was performed using the Batch protocol recently developed in our laboratory for time-optimized automated backbone resonance assignment. The Batch protocol is based on longitudinal relaxation-optimized BEST-type correlation experiments.<sup>51,52</sup> Sequential connectivity information is extracted from these spectra using COBRA,<sup>53</sup> a correlation-coefficient-based analysis tool. This connectivity information is complemented by additional amino-acid-type information extracted from a HADAMAC spectrum.<sup>54</sup> This automated approach provided assignment for 77% of the non-proline residues using NMR data recorded in 15 h on a 600-MHz spectrometer equipped with a cold probe.

Due to the low sensitivity of some peaks located in dynamic regions of the protein, additional standard 3D HNCACB and CBCA(CO)NH experiments were recorded in order to complete the backbone assignment. Aliphatic carbon and proton side-chain resonances were assigned using 4D HC(CCO)NH-type correlation spectra reconstructed from a set of four 3D projections with (H,C) tilt angles of 0°,  $\pm 20^\circ$ , and 90°. <sup>1</sup>H and <sup>13</sup>C resonance assignment of aromatic groups was done using standard 2D (HB)CB(CGCDCE)HE and (HB)CB(CGCD)HD spectra. The completeness of the final atom assignment is 83.9% for all atoms and 88.2% for <sup>1</sup>H atoms.

NMR distance restraints were derived from 3D <sup>15</sup>N-edited NOE spectroscopy (NOESY)-HSQC, 3D aliphatic <sup>13</sup>C-edited NOESY-HSQC, 3D aromatic <sup>13</sup>C-edited NOESY-HSQC, and 3D methyl-selective NOESY-HSQC experiments.<sup>55</sup> The following NOE mixing times were used: 125 ms (aliphatic and aromatic <sup>13</sup>C-edited NOESY-HSQC), 100 ms (<sup>15</sup>N-edited NOESY-HSQC), and 160 ms (3D methyl-selective NOESY-HSQC).

Residual dipolar couplings (RDCs) were collected on a uniformly <sup>13</sup>C, <sup>15</sup>N-labeled sample suspended in a liquid-crystalline medium consisting of a mixture of 5% pentaerythritol monododecyl ether (C<sub>12</sub>E<sub>5</sub>) and hexanol with a molar ratio of 0.96.<sup>56</sup> <sup>1</sup>D<sub>NH</sub>, <sup>1</sup>D<sub>C $\alpha$</sub> , and <sup>1</sup>D<sub>C $\alpha$ Ha</sub> couplings were obtained using 3D BEST HNCOC or HNCOCOA-type experiments. Coupling constants were calculated from the difference between line splitting measured in the isotropic and in the anisotropic sample using the NMRPipe nlinLS fitting routine.<sup>57</sup>

R1, R2(CPMG), and {<sup>1</sup>H}-<sup>15</sup>N heteronuclear NOE relaxation experiments were acquired at 600 MHz using standard pulse sequences<sup>58</sup> on P<sub>CTD2</sub>. For P<sub>CTD3</sub>, only the {<sup>1</sup>H}-<sup>15</sup>N heteronuclear NOE relaxation experiment was acquired. The relaxation-caused magnetization decay was sampled at 0, 100, 200, 400, 600, 800, 1100, 1500, and 1900 ms for longitudinal and at 10, 30, 50, 70, 90, 130, 170, 210, and 250 ms for transverse relaxation.

Standard pulse sequences were taken from the Varian BioPack. All NMR spectra were processed and analyzed using NMRPipe and NMRView.<sup>59</sup>

### Structure determination

Structure determination was accomplished with the ATNOS-CANDID NOESY assignment program,<sup>60,61</sup> using the CYANA program to calculate structural ensembles.<sup>62</sup> CYANA was used for seven cycles using only NOEs and dihedral angle restraints derived from the program TALOS.<sup>63</sup> Following this, a second run of ATNOS-CANDID was performed starting from cycle 2 using the structures determined with CYANA. In this case, the program CNS,<sup>64</sup> modified to incorporate the program SCULPTOR to handle RDC constraints, was used.<sup>65</sup> The second run involves a restrained molecular dynamics calculation using the standard CNS force field, with the initial sampling period at 1000 K for 5 ps during which time RDCs are scaled from 0.1% to 100% of their final values. A further sampling period of 0.5 ps precedes a slow cooling to 100 K over a period of 5 ps. Steps of 0.3 fs were used throughout the RDC refinement. All alignment tensor parameters were allowed to float throughout the protocol.<sup>66</sup>

### Structure analysis and comparisons

Sequence alignments were made using the CLUSTALW2 server,<sup>67</sup> and structure alignments were made with the SSM program.<sup>27</sup> Electrostatic maps were calculated using the Delphi program.<sup>68</sup> Conservation score was calculated with the AL2CO program.<sup>69</sup> Figures of the structure were prepared with PyMOL<sup>70</sup> and Chimera.<sup>71</sup>

### PDB accession number

Coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics PDB with the accession number 2k47.

## Acknowledgements

This work was supported by the interdisciplinary program "Maladies Infectieuses Emergentes" from the Centre National de la Recherche Scientifique and by a grant from the French Agence Nationale de la Recherche (ANR) [ANR-07-001-01 (ANRAGE) and ANR-05-JCJC-0077] and Lyonbiopôle. We thank Bertrand Lefèbvre, Denis Daveloos, and Bernard Brasmes [CRSSA (Le Centre de recherche du Service de santé Émile Pardé), Grenoble] for mass spectrometry measurements; Jean-Pierre Andrieu [IBS (Institut de Biologie Structurale), Grenoble] for N-terminal sequencing; Rodolfo Rasia (IBS) for help with RDC analysis; and Ewen Lescop (IBS) for the Batch processing protocols. E.A.R. was supported by postdoctoral fellowships from the University Joseph Fourier and from the ANR program (ANR-06-JCJC-0126-01). F.C.A.G. and C.L. were supported by MENRT (Ministère d'Éducation National, de la Recherche et de la Technologie) fellowship from the French government.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.07.028

## References

- Rodriguez, L. L. (2002). Emergence and re-emergence of vesicular stomatitis in the United States. *Virus Res.* **85**, 211–219.
- Whelan, S. P., Barr, J. N. & Wertz, G. W. (2004). Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr. Top. Microbiol. Immunol.* **283**, 61–119.
- Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. (2004). Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, **318**, 463–473.
- Emerson, S. U. & Wagner, R. R. (1973). L protein requirement for in vitro RNA synthesis by vesicular stomatitis virus. *J. Virol.* **12**, 1325–1335.
- Emerson, S. U. & Yu, Y. (1975). Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. *J. Virol.* **15**, 1348–1356.
- Arnheiter, H., Davis, N. L., Wertz, G., Schubert, M. & Lazzarini, R. A. (1985). Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell*, **41**, 259–267.
- Qanungo, K. R., Shaji, D., Mathur, M. & Banerjee, A. K. (2004). Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. *Proc. Natl. Acad. Sci. USA*, **101**, 5952–5957.
- Ding, H., Green, T. J. & Luo, M. (2004). Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2087–2090.
- Green, T. J., Macpherson, S., Qiu, S., Lebowitz, J., Wertz, G. W. & Luo, M. (2000). Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J. Virol.* **74**, 9515–9524.
- Ding, H., Green, T. J., Lu, S. & Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* **80**, 2808–2814.
- Gerard, F., Ribeiro, E., Albertini, A., Zaccai, G., Ebel, C., Ruigrok, R. & Jamin, M. (2007). Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry*, **46**, 10328–10338.
- Das, S. C. & Pattnaik, A. K. (2005). Role of the hypervariable hinge region of phosphoprotein P of vesicular stomatitis virus in viral RNA synthesis and assembly of infectious virus particles. *J. Virol.* **79**, 8101–8112.
- Gill, D. S., Chattopadhyay, D. & Banerjee, A. K. (1986). Identification of a domain within the phosphoprotein of vesicular stomatitis virus that is essential for transcription in vitro. *Proc. Natl. Acad. Sci. USA*, **83**, 8873–8877.
- Takacs, A. M., Das, T. & Banerjee, A. K. (1993). Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. *Proc. Natl. Acad. Sci. USA*, **90**, 10375–10379.
- Emerson, S. U. & Schubert, M. (1987). Location of the binding domains for the RNA polymerase L and the

- ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA*, **84**, 5655–5659.
16. Das, T., Pattnaik, A. K., Takacs, A. M., Li, T., Hwang, L. N. & Banerjee, A. K. (1997). Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. *Virology*, **238**, 103–114.
  17. Paul, P. R., Chattopadhyay, D. & Banerjee, A. K. (1988). The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype). *Virology*, **166**, 350–357.
  18. Chen, J. L., Das, T. & Banerjee, A. K. (1997). Phosphorylated states of vesicular stomatitis virus P protein in vitro and in vivo. *Virology*, **228**, 200–212.
  19. Gao, Y. & Lenard, J. (1995). Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly. *J. Virol.* **69**, 7718–7723.
  20. Mavrikakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. (2004). Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J. Mol. Biol.* **343**, 819–831.
  21. Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W. & Marion, D. (2004). Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*, **319**, 201–211.
  22. 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.
  23. Karlin, D., Ferron, F., Canard, B. & Longhi, S. (2003). Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* **84**, 3239–3252.
  24. Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G. & Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. USA*, **86**, 8247–8251.
  25. Schanda, P., Forge, V. & Brutscher, B. (2006). HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magn. Reson. Chem.* **44**, S177–S184; Spec No.
  26. Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science*, **273**, 595–603.
  27. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2256–2268.
  28. Lipari, G. & Szabo, A. (1982). A model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559.
  29. Houben, K., Blanchard, L., Blackledge, M. & Marion, D. (2007). Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P. *Biophys. J.* **93**, 2830–2844.
  30. Takacs, A. M. & Banerjee, A. K. (1995). Efficient interaction of the vesicular stomatitis virus P protein with the L protein or the N protein in cells expressing the recombinant proteins. *Virology*, **208**, 821–826.
  31. Chattopadhyay, D. & Banerjee, A. K. (1987). Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro. *Cell*, **49**, 407–414.
  32. Barik, S. & Banerjee, A. K. (1992). Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation. *J. Virol.* **66**, 1109–1118.
  33. Jacob, Y., Real, E. & Tordo, N. (2001). Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.* **75**, 9613–9622.
  34. Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. (2004). Does common architecture reveal a viral lineage spanning all three domains of life? *Mol. Cell*, **16**, 673–685.
  35. Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982). Rapid evolution of RNA genomes. *Science*, **215**, 1577–1585.
  36. Bamford, D. H., Grimes, J. M. & Stuart, D. I. (2005). What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* **15**, 655–663.
  37. Worobey, M. & Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **80**, 2535–2543.
  38. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S. *et al.* (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, **21**, 3435–3438.
  39. Li, X., Romero, P., Rani, M., Dunker, A. K. & Obradovic, Z. (1999). Predicting protein disorder for N-, C-, and internal regions. *Genome Inform. Ser. Workshop Genome Inform.* **10**, 30–40.
  40. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Struct. Funct. Genet.* **42**, 38–48.
  41. Liu, J., Tan, H. & Rost, B. (2002). Loopy proteins appear conserved in evolution. *J. Mol. Biol.* **322**, 53–64.
  42. Dosztanyi, Z., Csizmek, V., Tompa, P. & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
  43. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. (2004). The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, **20**, 2138–2139.
  44. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure*, **11**, 1453–1459.
  45. Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, **21**, 3369–3376.
  46. Coeytaux, K. & Poupon, A. (2005). Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics*, **21**, 1891–1900.
  47. Wootton, J. (1994). Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* **18**, 269–285.
  48. Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, B., Chomilier, J. *et al.* (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell. Mol. Life Sci.* **53**, 621–645.
  49. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, **6**, 1948–1954.
  50. Iseni, F., Baudin, F., Blondel, D. & Ruigrok, R. W. (2000). Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *RNA*, **6**, 270–281.

51. Schanda, P. & Brutscher, B. (2007). Hyperdimensional protein NMR spectroscopy in peptide-sequence space. *J. Am. Chem. Soc.* **129**, 11916–11917.
52. Lescop, E., Schanda, P. & Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187**, 163–169.
53. Schanda, P., Van Melckebeke, H. & Brutscher, B. (2006). Speeding up three-dimensional protein NMR experiments to a few minutes. *J. Am. Chem. Soc.* **128**, 9042–9043.
54. Lescop, E., Rasia, R. & Brutscher, B. (2008). Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *J. Am. Chem. Soc.* **130**, 5014–5015.
55. Van Melckebeke, H., Simorre, J. P. & Brutscher, B. (2004). Amino acid-type edited NMR experiments for methyl–methyl distance measurement in <sup>13</sup>C-labeled proteins. *J. Am. Chem. Soc.* **126**, 9584–9591.
56. Rückert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7797.
57. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
58. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G. *et al.* (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
59. Johnson, B. A. & Blevins, R. A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 606–614.
60. Herrmann, T., Guntert, P. & Wuthrich, K. (2002). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, **24**, 171–189.
61. Herrmann, T., Guntert, P. & Wuthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227.
62. Guntert, P., Mumenthaler, C. & Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
63. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
64. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
65. Hus, J. C., Marion, D. & Blackledge, M. (2000). De novo determination of protein structure by NMR using orientational and long-range order restraints. *J. Mol. Biol.* **298**, 927–936.
66. Sibille, N., Blackledge, M., Brutscher, B., Coves, J. & Bersch, B. (2005). Solution structure of the sulfite reductase flavodoxin-like domain from *Escherichia coli*. *Biochemistry*, **44**, 9086–9095.
67. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
68. Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.
69. Pei, J. & Grishin, N. V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, **17**, 700–712.
70. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific, Palo Alto, CA, USA.
71. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E. (2006). Tools for integrated sequence–structure analysis with UCSF Chimera. *BMC Bioinformatics*, **7**, 339.

SUPPLEMENTARY MATERIAL

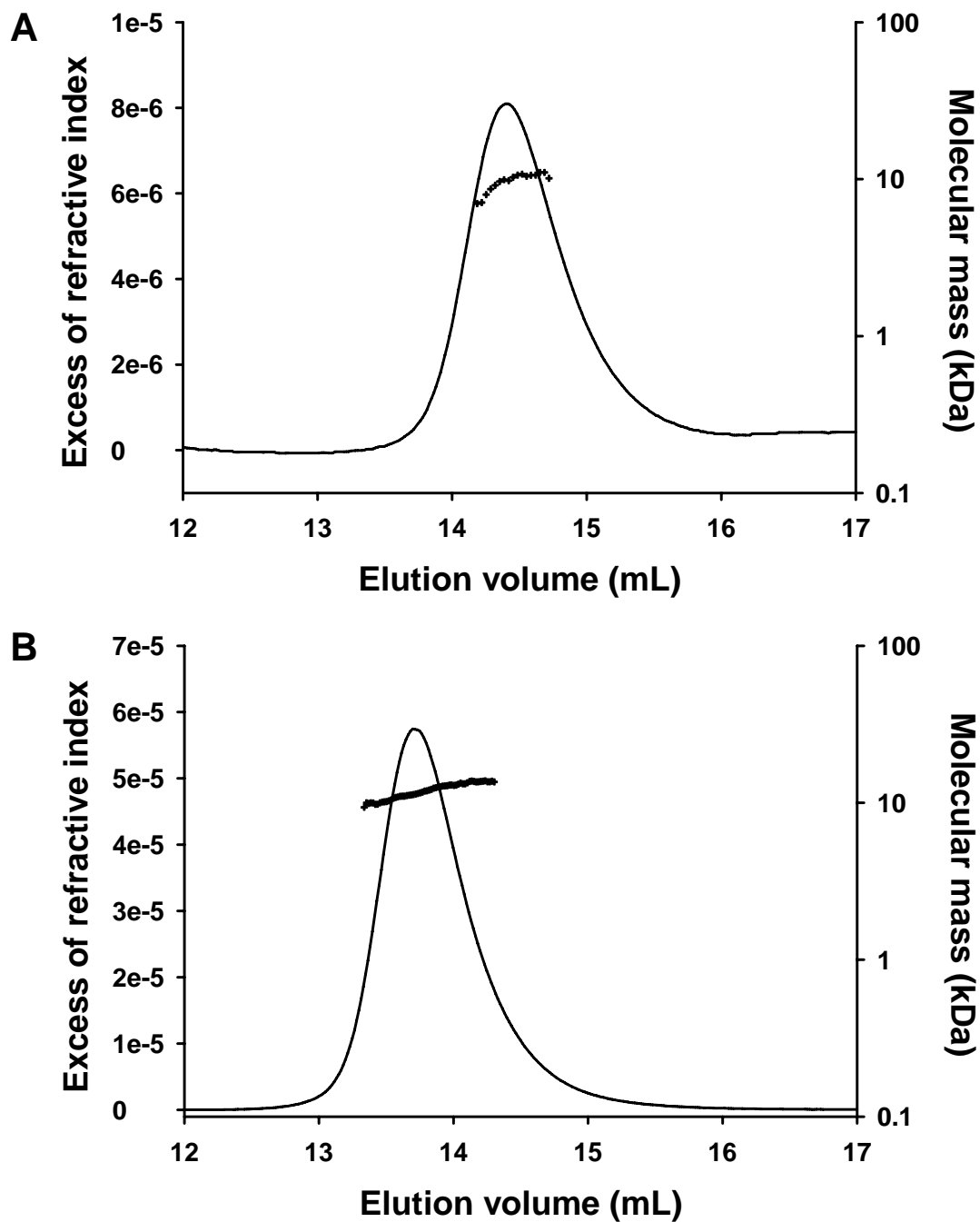
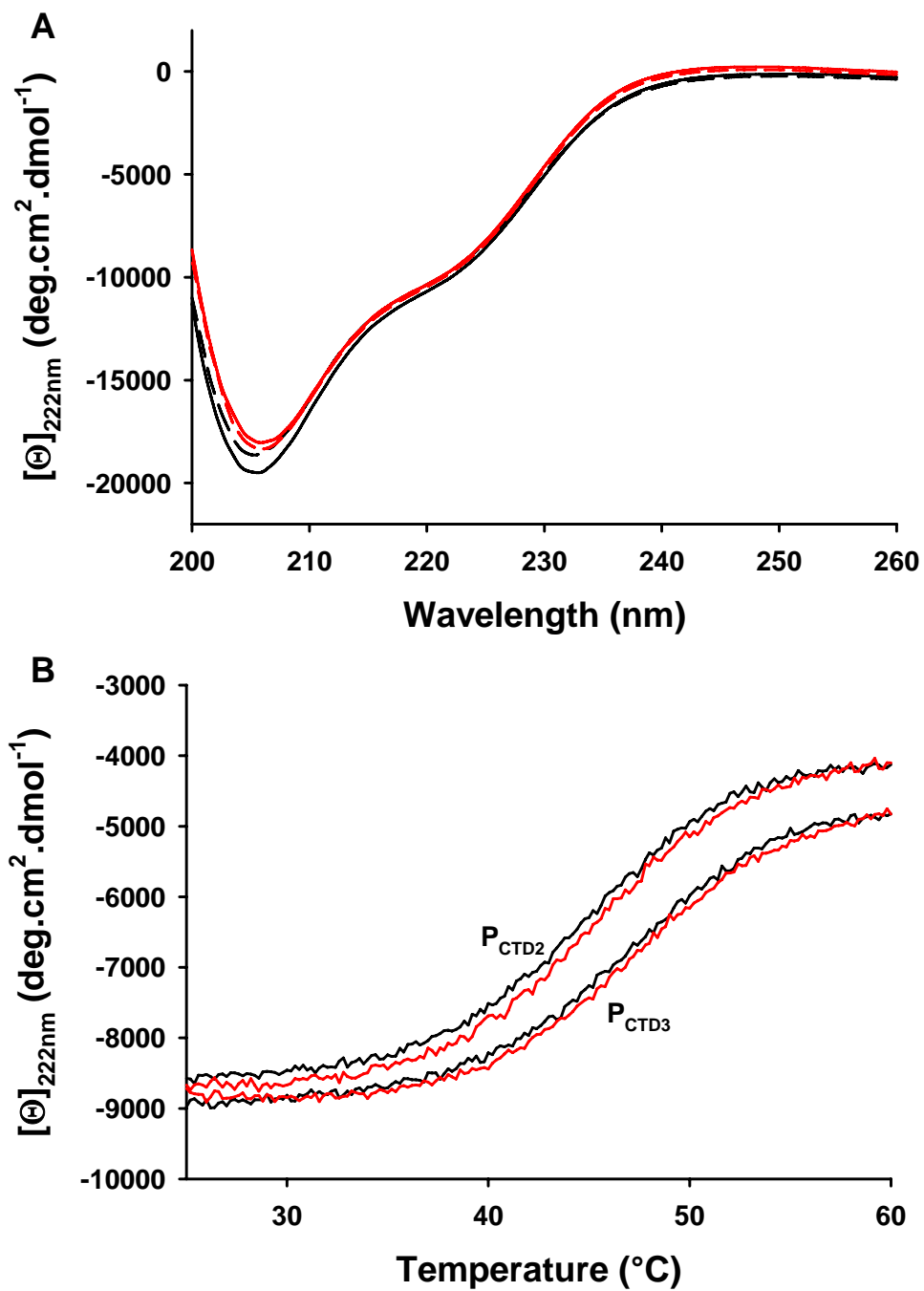
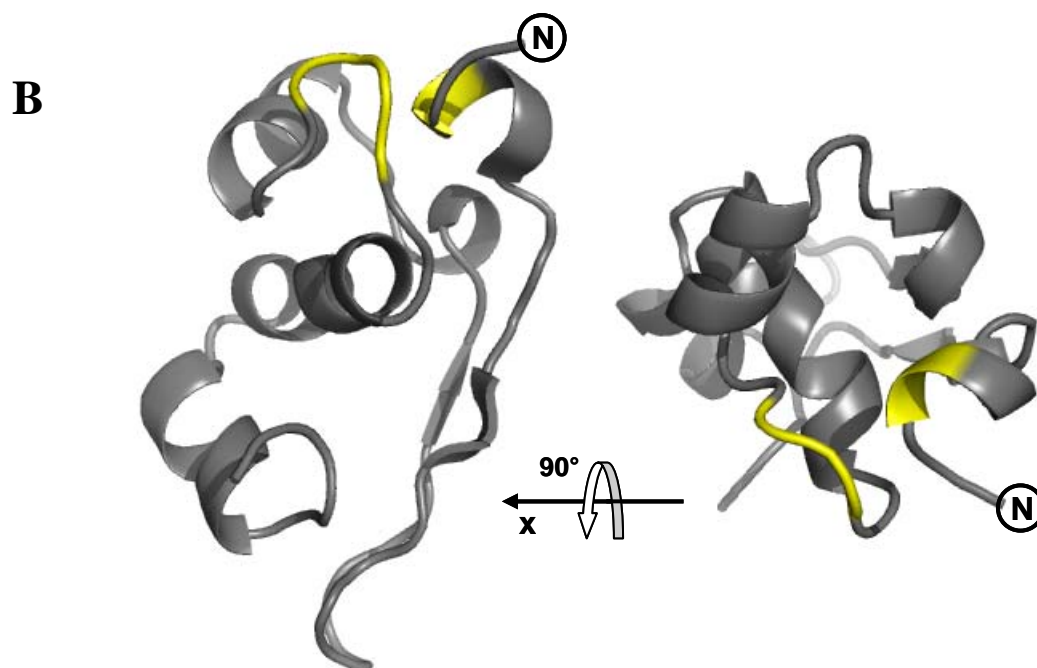
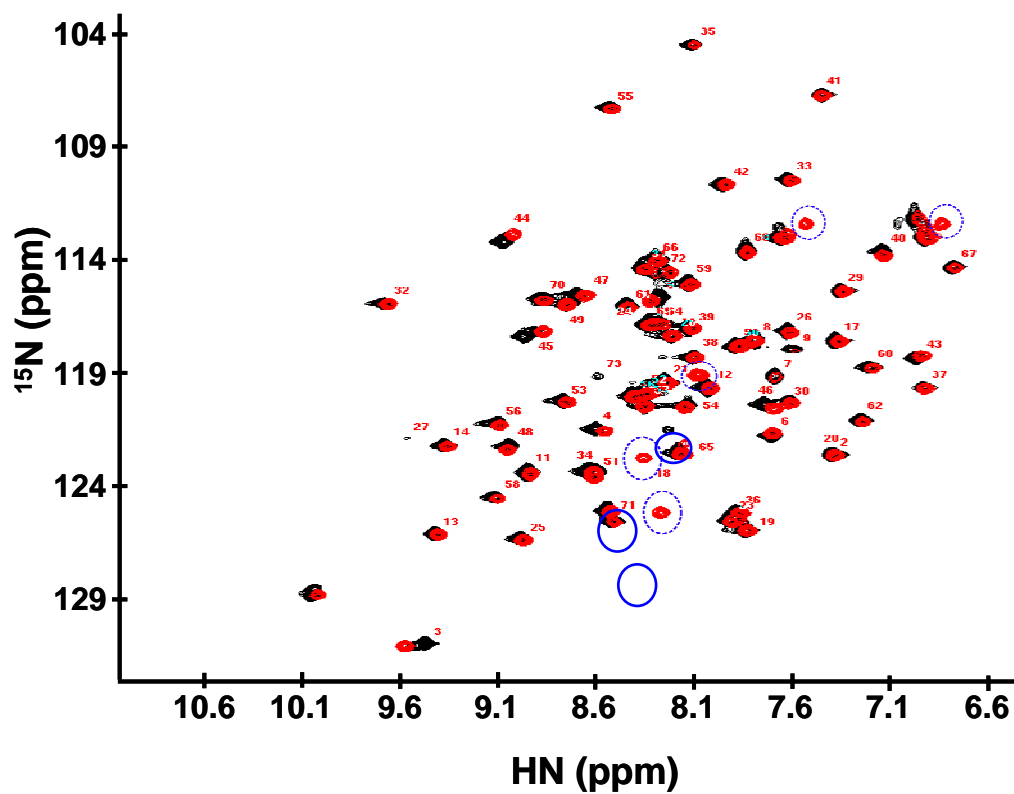


FIGURE S1 – SEC-MALLS-RI analysis of  $P_{\text{CDT2}}$  (A) and  $P_{\text{CDT3}}$  (B).

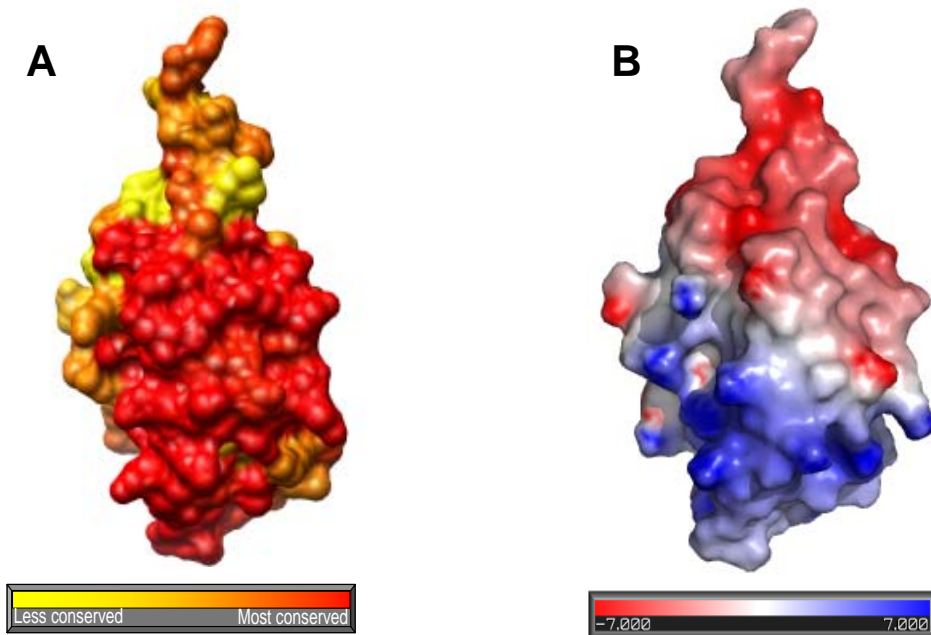




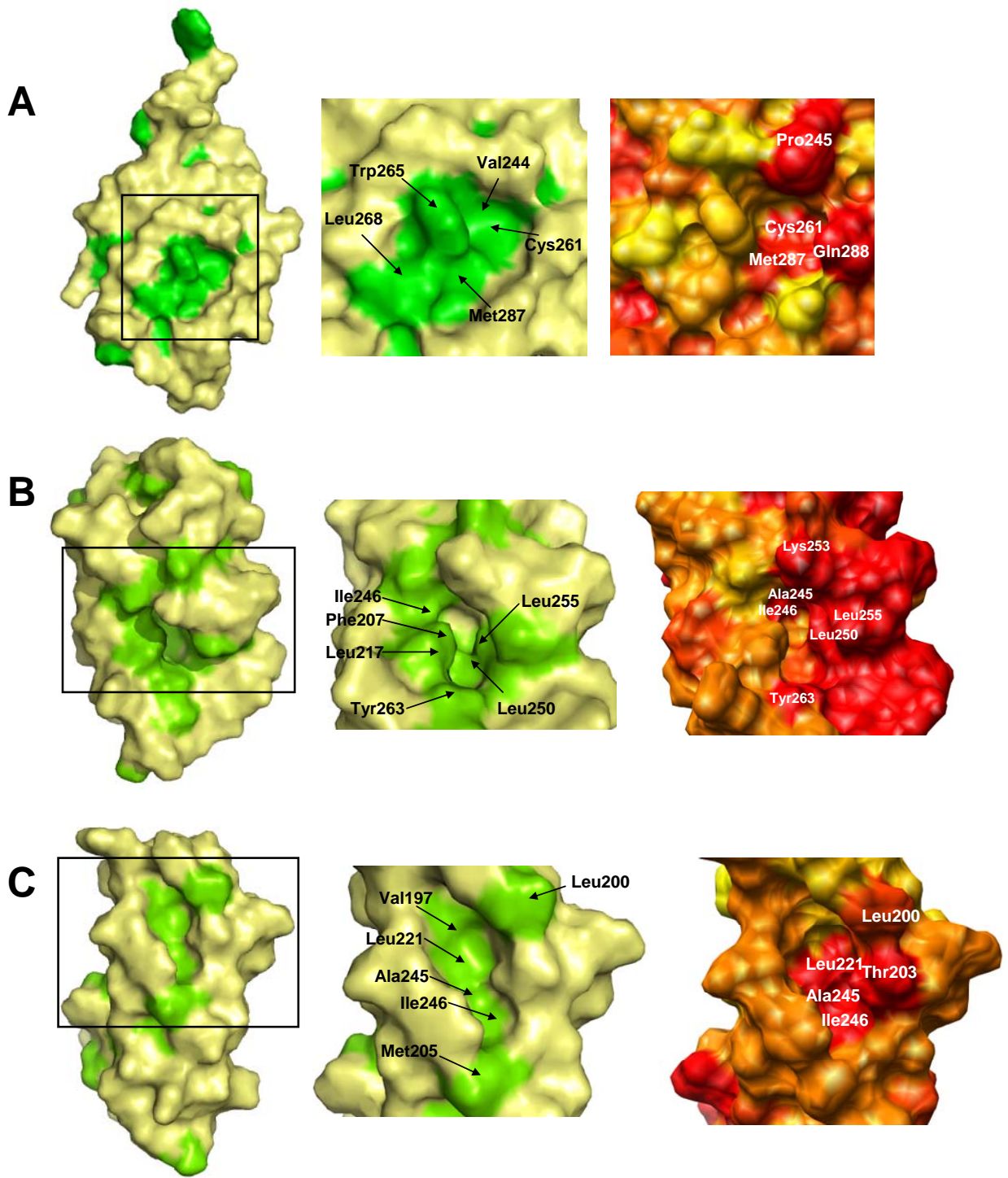
**FIGURE S2 – (A) CD spectrum of PCDT2 after each cycle of thermal denaturation/renaturation. (C) Thermal denaturation of P<sub>CTD2</sub> and P<sub>CTD3</sub>.**



**Figure S3. Comparison of NMR spectra of  $P_{CTD2}$  and  $P_{CTD3}$ .** (A) Superposed HSQC spectra of  $P_{CTD2}$  (Black) and  $P_{CTD3}$  (Red). Additional amide resonances in  $P_{CTD3}$  are circled with solid blue line, while additional side chain resonances are circles with dashed blue line. (B) Cartoon representation of  $P_{CTD2}$  showing the location in yellow of residues for which the amide backbone resonance is slightly shifted in the NMR spectrum of  $P_{CTD3}$ .



**FIGURE S4 – (A) Representation according to residus conservation (yellow, residus less conserved, red, residus most conserved. (B) Electrostatic view of the PCDT2**



**FIGURE S5**

## D - Conclusion

Dans cette publication, nous avons utilisé la méthode d'analyse bioinformatique que nous avons mise au point (Gerard & *al.*, 2008) pour définir les régions structurées de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV). Nous avons identifié un domaine indépendant dans la région C-terminale de la phosphoprotéine du virus de la stomatite vésiculaire (VSV) (figure 1, P<sub>CTD</sub>). En effet, l'analyse par SEC-MALLS-RI (figure S1), mais aussi par dichroïsme circulaire et suivi du dépliement et repliement induit par la température (figure S2), nous a permis de montrer que le domaine P<sub>CTD</sub> est un domaine autonome. L'expérience sur coussin de glycérol nous a permis de montrer, comme le domaine C-terminal de la phosphoprotéine du virus de la rage, que le domaine P<sub>CTD</sub> est capable de fixer les nucléocapsides virales (figure 2).

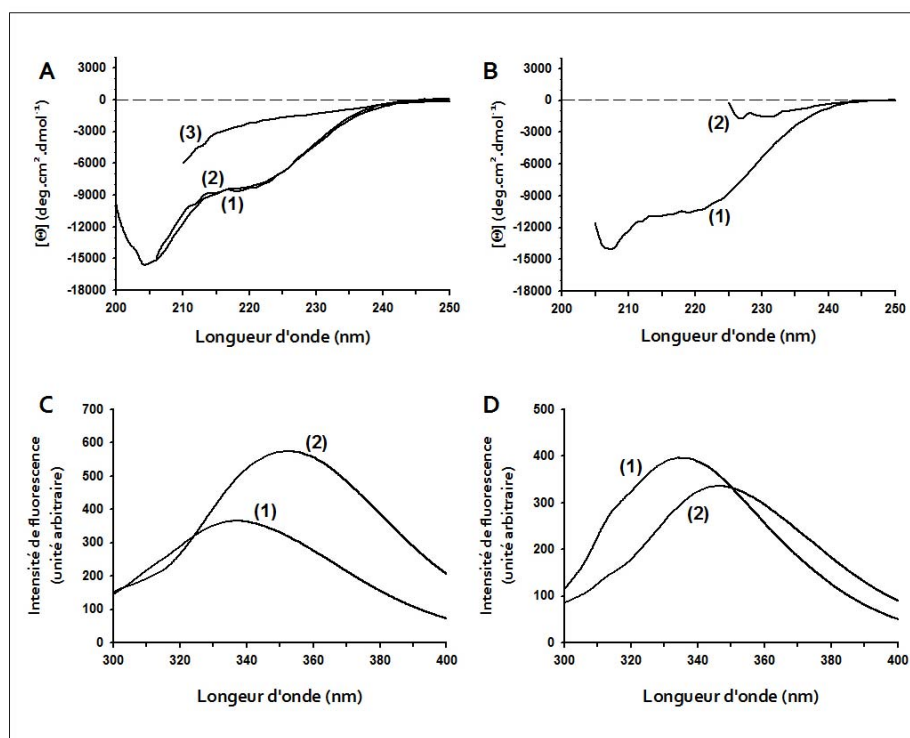
La structure atomique du domaine C-terminal de la phosphoprotéine du virus de la stomatite vésiculaire a été résolue par Résonance Magnétique Nucléaire (RMN). La topologie du domaine C-terminal de la phosphoprotéine de VSV est similaire à celle du domaine C-terminal de la phosphoprotéine du virus de la rage (RV). Le domaine P<sub>CTD</sub> du VSV est composé de 4 hélices  $\alpha$  (1 à 4), une hélice  $3^{10}$   $\eta$ 1 ainsi que deux brins  $\beta$ 1 et  $\beta$ 2 parallèles (figure 4). L'hélice  $\alpha$ 1 en C-terminal du domaine P<sub>CTD</sub> du VSV est plus courte que celle du domaine C-terminal de la phosphoprotéine du RV. Le domaine P<sub>CTD</sub> de VSV ne possède pas les hélices nommées  $\alpha$ 3 et  $\alpha$ 6 chez le RV, et la boucle entre les brins parallèles  $\beta$ 1 et  $\beta$ 2 est plus longue chez le domaine P<sub>CTD</sub> de VSV. Même si il existe peu de similarité de séquence, nous avons montré ici une homologie de structure entre les deux domaines C-terminaux. En effet, les parties communes des domaines ont pu être superposé avec un rmsd de 2,6 Å (figure 5).

Des études dynamiques par relaxation en RMN ont permis de mettre en évidence des mouvements rapides (inférieur au ns) dans la région de la boucle 209-214 et des échanges conformationnels plus lents dans les hélices N et C terminales. La caractérisation de la construction 185-265 (P<sub>CTD3</sub>) indique que le P<sub>CTD</sub> est précédé d'une région flexible. Ces résultats confirment l'organisation modulaire de la phosphoprotéine du VSV, avec des domaines indépendants séparés de régions flexibles.

#### IV – Analyse de la stabilité de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) et du virus de la rage (P RV)

##### A – Spectroscopie de Dichroïsme Circulaire (CD) et de fluorescence

Afin de caractériser la structure de la phosphoprotéine en solution, d'étudier sa flexibilité conformationnelle et de mesurer sa stabilité thermodynamique, nous avons utilisé la spectroscopie CD dans l'UV lointain, une sonde de la présence de structures secondaires, et la spectroscopie de fluorescence des tryptophanes, une sonde de la présence d'un cœur hydrophobe afin de suivre les transitions de dénaturation induites par l'urée.



**Figure 37 - Spectres CD et fluorescence des Trp. (A) spectre CD dans l'UV lointain de la phosphoprotéine du VSV (P VSV).** Courbe (1) P VSV produite en cellule d'insectes, courbe (2) P VSV produite en bactérie et courbe (3) P VSV produite en cellules d'insecte dans 6 M urée. **(B) Spectre CD dans l'UV lointain de la phosphoprotéine du RV (P RV).** Courbe (1) P RV produite en cellules d'insectes et courbe (2) P RV produite en cellules d'insectes et dénaturées dans 6 M urée. **(C) Spectre de fluorescence des Trp de P VSV.** Courbe (1) P VSV produite en bactérie et courbe (2) P VSV produite en bactérie et dénaturée en 6 M urée. **(D) Spectre fluorescence des Trp de P RV.** Courbe (1) P RV produite en cellule d'insecte et courbe (2) P RV produite en cellule d'insectes et dénaturée en 6 M urée.

Les spectres CD des protéines natives indiquent une présence significative d'hélices  $\alpha$ , un peu plus important pour la phosphoprotéine du RV que pour la phosphoprotéine du VSV (figure 37, A & B). En utilisant la valeur de référence de l'ellipticité molaire par résidu à 222 nm pour 100% d'hélices  $\alpha$ ,  $[\Theta]_{222\text{nm}} = -35000 \text{ deg.cm}^2.\text{dmol}^{-1}$  (Scholtz & Baldwin, 1992), le pourcentage d'hélices  $\alpha$  pour la protéine P VSV est de 23% ( $[\Theta]_{222\text{nm}} = -8000 \text{ deg.cm}^2.\text{dmol}^{-1}$ ) et pour la protéine P RV de 28% ( $[\Theta]_{222\text{nm}} = -10000 \text{ deg.cm}^2.\text{dmol}^{-1}$ ).

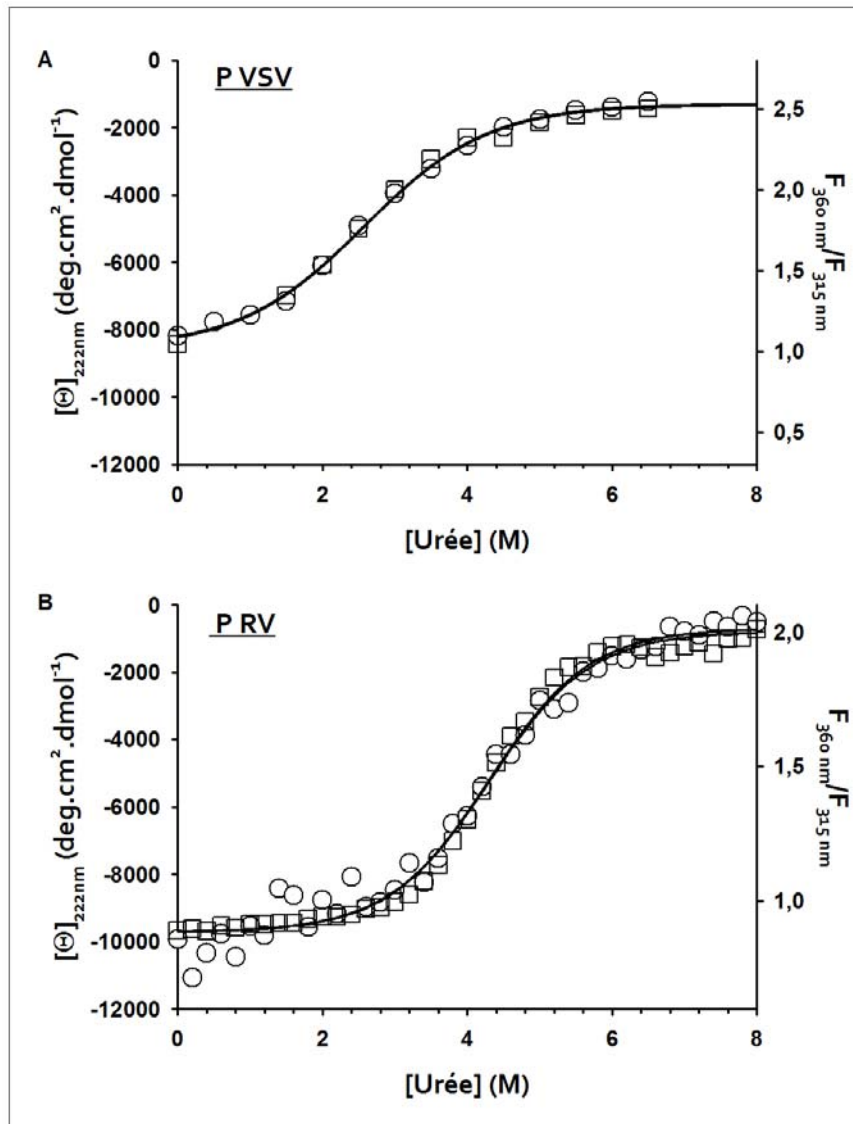
Les spectres de fluorescence intrinsèque pour les deux phosphoprotéines montrent un maximum d'émission de fluorescence à 335 nm, caractéristique de protéines où les résidus Trp sont enfouis dans un cœur hydrophobe (figure 37, C & D). L'addition de l'urée résulte en la disparition des structures secondaires, comme l'indique la disparition des bandes dichroïques caractéristiques entre 200 et 250 nm et à l'exposition des résidus Trp au solvant, comme l'indique le décalage vers le rouge du spectre d'émission des Trp ( $\lambda_{\max} = 350 \text{ nm}$ ) (figure 37, C & D).

## B - Transition de dénaturation induite par l'urée

La stabilité de la phosphoprotéine a été déterminée par analyse de la transition de dénaturation induite par l'urée. Les transitions de dénaturation induite par l'urée pour chaque protéine, suivies par les deux sondes spectroscopiques se superposent, ce qui suggère que le dépliement de la phosphoprotéine suit un processus en deux étapes dans lequel la protéine existe soit dans sa forme native soit dans sa forme dépliée (figure 38). La superposition de transitions mesurées dans les deux sens, vers le dépliement ou vers le repliement (données non montrées) démontre que la réaction atteint l'équilibre en moins de 5 minutes et que le dépliement est complètement réversible. Les données de fluorescence et de CD ont été ajustées sur la base d'un modèle simple à deux états en utilisant une procédure qui prend en compte les lignes de base pour N (forme native) et U (forme dépliée) à l'intérieur et l'extérieur de la région de transition (Santoro & Bolen, 1988). La constante d'équilibre K et la différence d'énergie libre entre N et U,  $\Delta G^\circ$ , ont été déterminées dans la région de transition. Comme pour de nombreuses protéines,  $\Delta G^\circ$  varie de façon linéaire avec la concentration molaire de dénaturant dans la région de transition. La pente du graphe de  $\Delta G^\circ$  fonction de la concentration molaire en urée,  $\left(\frac{\partial \Delta G^\circ}{\partial [\text{Urea}]}\right)$ , nommé la valeur m, nous donne une indication de la surface qui devient accessible au solvant lors du dépliement de la protéine ( $\Delta \text{Accessible Surface Area}$ ) (Myers & al., 1995). L'extrapolation à 0 M urée permet d'obtenir une estimation de la stabilité de la forme native N en absence d'agent dénaturant,  $\Delta G^\circ(\text{H}_2\text{O})$  (tableau 6). Alors que les valeurs de m sont similaires pour les protéines P VSV et P RV, la stabilité de la protéine P RV en absence d'urée est significativement plus élevée que celle de la protéine P VSV.

Paramètres	P VSV (cellules d'insectes)	P VSV (bactérie)	P RV (cellules d'insectes)
$\Delta G^\circ(\text{H}_2\text{O}) (\text{cal} \cdot \text{mol}^{-1})$	$1,500 \pm 150$	$1,780 \pm 80$	$3,660 \pm 200$
$m (\text{cal} \cdot \text{mol}^{-1} \cdot \text{M}^{-1})$	$620 \pm 50$	$690 \pm 30$	$850 \pm 50$

**Tableau 6** - Paramètres thermodynamiques obtenus à partir des transitions de dénaturation induite par l'urée.



**Figure 38** - Transition de dénaturation induite par l'urée de la phosphoprotéine du VSV (A) et de la phosphoprotéine du RV (B). Les transitions ont été suivies par CD à 222 nm (cercles) et par fluorescence des Trp (présenté ici par le ratio d'intensité de fluorescence à 360 nm et à 325 nm) (carrés). Les données ont été superposées pour la présentation en adaptant les échelles des axes. L'ajustement est dessiné en utilisant les équations 5 & 6 dérivant de l'analyse d'un modèle à deux états et les paramètres sont décrits dans le [tableau 6](#). Conditions: 20 mM Tris-HCl, 150 mM NaCl, pH 7,5 à 30°C, concentration en protéine de 3 µM.

### C - Dénaturation thermique

La dénaturation et la renaturation thermique de la phosphoprotéine du VSV ont été induites en augmentant ou en diminuant de façon continue la température entre 4°C et 60°C, et en suivant les changements de conformation par CD dans l'UV lointain. La forme dénaturée par la température de la phosphoprotéine du VSV conserve environ deux tiers de la structure secondaire native, et plus aucun changement n'intervient entre 60 et 95°C ([figure 39, A](#)). Plus de quatre cycles de dénaturation /renaturation sont superposables, ce qui indique que la réaction atteint rapidement l'équilibre et que la dénaturation thermique est réversible. Les transitions thermiques suivent un modèle coopératif de



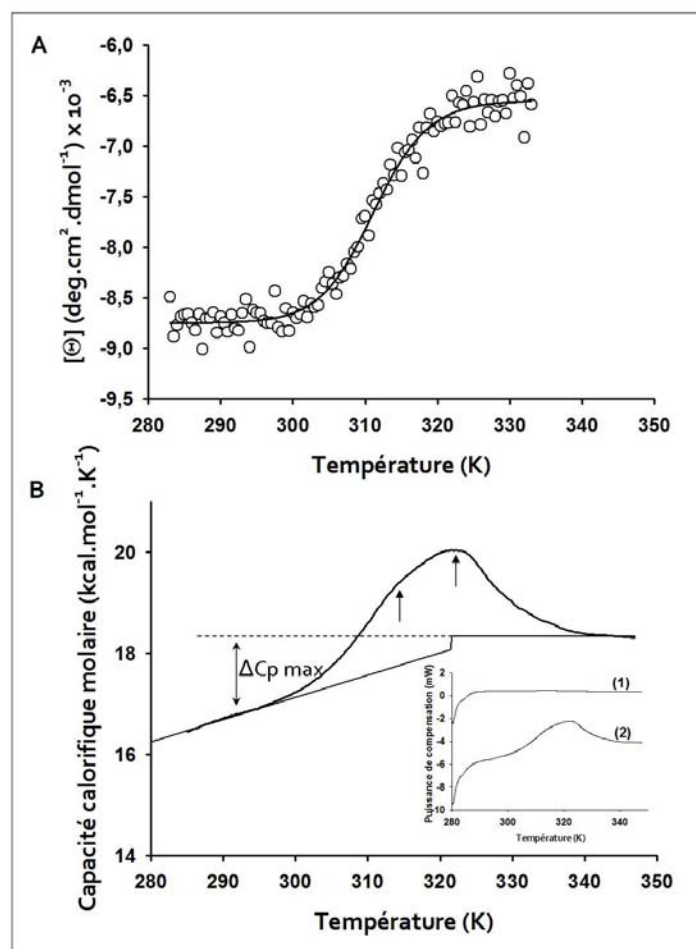
dénaturation à deux états, et peuvent être utilisées pour calculer la différence d'énergie libre entre la forme native et la forme dépliée (par la température) de la protéine P VSV. En supposant un modèle à deux états, l'analyse de la transition thermique donne une valeur  $T_m$  de  $311 \pm 1$  K, et la variation d'enthalpie qui accompagne ce processus de dénaturation est de  $51,0 \pm 3,0$  kcal.mol<sup>-1</sup>. Pour la protéine P RV, la dénaturation thermique dans ces conditions est irréversible.

## D – Calorimétrie différentielle à balayage (DSC)

Ce travail a été réalisé en collaboration avec Jean-Luc GARDEN et le Dr. Hervé GUILLOU du Centre de Recherches sur les Très Basses Températures (CNRS, UJF et INPG) à Grenoble.

La [figure 39, B](#) nous indique la capacité calorifique molaire de la phosphoprotéine du VSV en fonction de la température mesurée par DSC. La dénaturation thermique de la protéine P VSV est accompagnée par un excès de capacité calorifique sur le pic principal de  $321 \pm 1$  K et pour l'épaulement à environ 310 K (indiquer par des flèches sur la [figure 39, B](#)). La température caractéristique de l'épaulement correspond à la température de fusion dans la transition de dépliement suivie par CD. L'origine du pic majeur reste mal comprise : il pourrait résulter d'une agrégation de la protéine, étant donné que les expériences ont été réalisées à une concentration plus importante (35  $\mu$ M contre 2  $\mu$ M en CD), ou de l'existence d'une seconde transition qui n'affecte pas la quantité d'hélices  $\alpha$  de la protéine. Les lignes de bases linéaires pour la capacité calorifique de la forme native et dénaturées par la température ont été utilisées pour analyser la courbe. L'excès de capacité de chaleur entre la température initiale (22°C) et finale (71°C) est réduit après chaque cycle indiquant la présence de processus irréversible qui doivent être analysés. Cependant, des changements similaires de capacité calorifique ( $\Delta C_p$ ) sont mesurés pour les processus de dénaturation successifs. Une limite supérieure pour un  $\Delta C_p$  de  $1,5 \pm 0,5$  kcal.mol<sup>-1</sup>.K<sup>-1</sup> a été obtenue comme la différence de capacité de chaleur entre la température initiale et la température finale.

Par la suite, nous avons obtenu de bons résultats avec le domaine C-terminal de la protéine P VSV qui se comportent comme un domaine indépendant et qui se replie et déplie en variant la température. Ceci a été très largement confirmé par CD et est présenté dans le papier présenté en partie III (Ribeiro & *al.*, 2008).

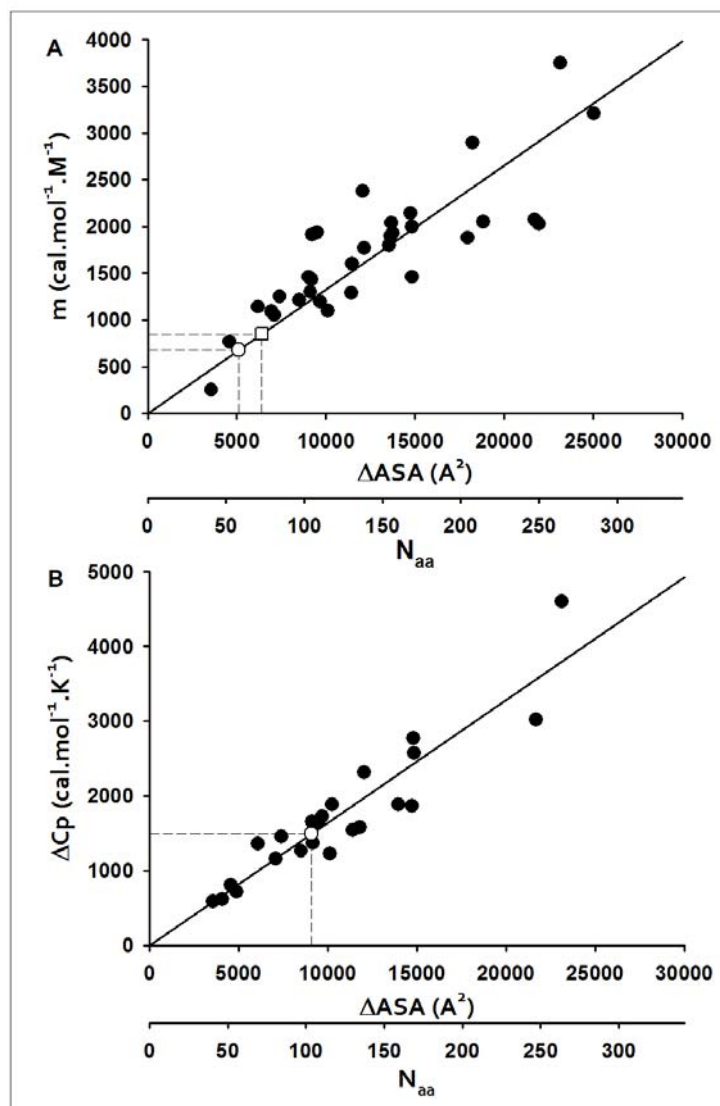


**Figure 39** - Dénaturation thermique de la protéine P VSV. (A) CD de la protéine P VSV en fonction de la température. L'ellipticité à 222 nm a été mesurée alors que la température variait de 4 K.min<sup>-1</sup>. La courbe correspond à la courbe théorique calculé en utilisant l'équation 7, avec les paramètres suivants :  $T_m = 311$  K,  $\Delta H = 51$  kcal.mol<sup>-1</sup>,  $[\Theta]_N = 8750$  deg.cm<sup>2</sup>.dmol<sup>-1</sup> and  $[\Theta]_D = 6550$  deg.cm<sup>2</sup>.dmol<sup>-1</sup>. Conditions: 20 mM Tris-HCl, 150 mM NaCl à pH 7,5, concentration en protéine = 3  $\mu$ M. (B) Capacité calorifique molaire en fonction de la température pour la protéine P VSV par DSC. Le scan calorimétrique a été réalisé dans 5 mM phosphate de sodium à pH 7 avec une concentration en protéine de 35  $\mu$ M avec une vitesse de scan de 1 K.min<sup>-1</sup>. La capacité de chaleur a été analysée des mesures brutes de la puissance de compensation en utilisant les lignes de bases linéaires pour les états natif et déplié (ligne fine). La position des deux pics a été indiquée par des flèches et la limite supérieure pour  $\Delta C_p$  est calculée comme la différence de capacité calorifique au début et à la fin de la transition (double flèche). Le graphe indique les mesures de lbrutes de la puissance de compensation pour le solvant (1) et pour la solution de protéine (2).

## E – Changement de surface accessible à l'agent dénaturant

Pour les petites protéines, la valeur  $m$  obtenue a partir des transitions induites par dénaturation et le changement de capacité calorifique  $\Delta C_p$ , ont été corrélés avec le changement de surface accessible lors de la dénaturation ( $\Delta ASA$ ) et avec la taille de la protéine (Myers & *al.*, 1995). Les valeurs  $m$  pour les protéines P VSV et P RV sont faibles pour des protéines de plus de 250 aa. Sur le graphique (figure 40) représentant la valeur  $m$  en fonction de  $\Delta ASA$  obtenue pour diverses protéines globulaires (figure 40, A), elles correspondent à des valeurs de  $\Delta ASA$  mesurées typiquement pour des protéines globulaires de 60

et 70 aa, respectivement. La limite supérieure calculée pour la valeur  $\Delta C_p$  ( $1,5 \text{ kcal.mol}^{-1}.\text{K}^{-1}$ ) correspond à une valeur de  $\Delta ASA$  typique pour une protéine de 100 aa (figure 40, B). Ces résultats indiquent clairement que la partie structurée de la protéine P implique environ une centaine d'acides aminés.



**Figure 40** - Graphique de la valeur  $m$  (A) et de  $\Delta C_p$  (B) en fonction de  $\Delta ASA$  et du nombre d'acides aminés ( $N_{aa}$ ). Les valeurs  $m$  obtenues des courbes de transitions de dénaturation par l'urée et la valeur de  $\Delta C_p$  obtenus par l'expérience de DSC sont proportionnels au changement de surface accessible à l'urée lors du dépliement. Les données pour des petites protéines globulaires (cercles noirs pleins) sont obtenus de (Myers & *al.*, 1995). Le second axes des abscisses est obtenu en corrélant la valeur  $\Delta ASA$  et le nombre d'acides aminés ( $N_{aa}$ ):  $\Delta ASA (\text{\AA}^2) = 87.80 N_{aa}$  (Myers & *al.*, 1995). Le cercle vide correspond à la protéine P de VSV et le carré à la protéine P de RV. La valeur  $m$  correspond à celle prédite pour des protéines globulaires de 60 aa pour P VSV ( $5100 \text{\AA}^2$ ) et de 70 aa pour P RV ( $6400 \text{\AA}^2$ ). La valeur de  $\Delta C_p$  pour P VSV correspond à une protéine globulaire de 100 aa ( $9100 \text{\AA}^2$ ).

## F - Conclusions

Ces caractérisations biophysiques des phosphoprotéines de RV et VSV nous ont montré que ces protéines étaient peu stables en solution. L'analyse des courbes de dénaturation, par CD et fluorescence, a montré qu'une petite partie (environ un tiers de la protéine) était sensible à l'agent dénaturant

(paragraphe B et E). Aussi, des différences de comportement vis-à-vis de la dénaturation thermique ont pu être notées (paragraphe C et D). En effet, seule la phosphoprotéine du VSV (ou son domaine C-terminal) a un repliement réversible lorsqu'elle est soumise à des changements de température. L'analyse par DSC de la protéine P VSV entière n'est pas certaine, mais confirme bien une faible stabilité de la protéine. Pour la protéine P de RV, le mutant P<sub>Δ91-131</sub> ou le domaine C-terminal de P RV, nous n'avons jamais obtenu, de repliements réversibles.

L'analyse de la stabilité de la phosphoprotéine en solution laisse des questions en suspens. En effet, d'après nos analyses, il semble que nous n'observions le dépliement que d'un tiers de la protéine en une étape, alors que nous nous attendions aux dépliements successifs des domaines indépendants de la phosphoprotéine. Il est possible que nous n'observions qu'un seul chemin de dépliement. En effet, les domaines de la phosphoprotéine c'est-à-dire le domaine central et de domaine C-terminal, puisque nous n'avons aucune preuve que le domaine N-terminal est replié (son repliement est sûrement induit par la présence de la nucléoprotéine), pourrait suivre des chemins de dépliement similaires ou presque similaires. Les domaines individuels chez VSV semble se replie et deplie de façon réversible mais nous n'avons observé cela en DSC. Il semblerait que la formation d'agrégats induits par la température pose problème.

## **V – Etudes structurales des phosphoprotéines du virus de la stomatite vésiculaire (P VSV) et du virus de la rage (P RV)**

La cristallisation de la phosphoprotéine du virus de la rage (P RV) ou du virus de la stomatite vésiculaire (P VSV) entière n'ayant pas donné de résultats, nous nous sommes tournés vers les techniques de diffusions aux petits angles (SAS : *Small Angle Scattering*) comme le SANS (*Small Angle Neutron Scattering*) et le SAXS (*Small Angle X-ray Scattering*). Ces techniques de SAS permettent d'étudier les macromolécules biologiques en solution dans une large gamme de conditions, de physiologiques à dénaturantes.

L'analyse par SANS des phosphoprotéines P RV et P VSV nous a indiqué une asymétrie de l'homodimère, assimilable à un cylindre (Gerard & al., 2007). Nous avons aussi utilisé le SAXS, qui permet d'obtenir des données pour des valeurs plus élevées de Q, et dont l'une des applications majeures en biologie est la modélisation des complexes macromoléculaires. Il est possible, grâce à des algorithmes d'obtenir des reconstructions *ab initio* 3D à basse résolution à partir de la courbe de diffusion 1D. Cependant, les informations contenues dans les données de diffusions sont limitées par rapport aux

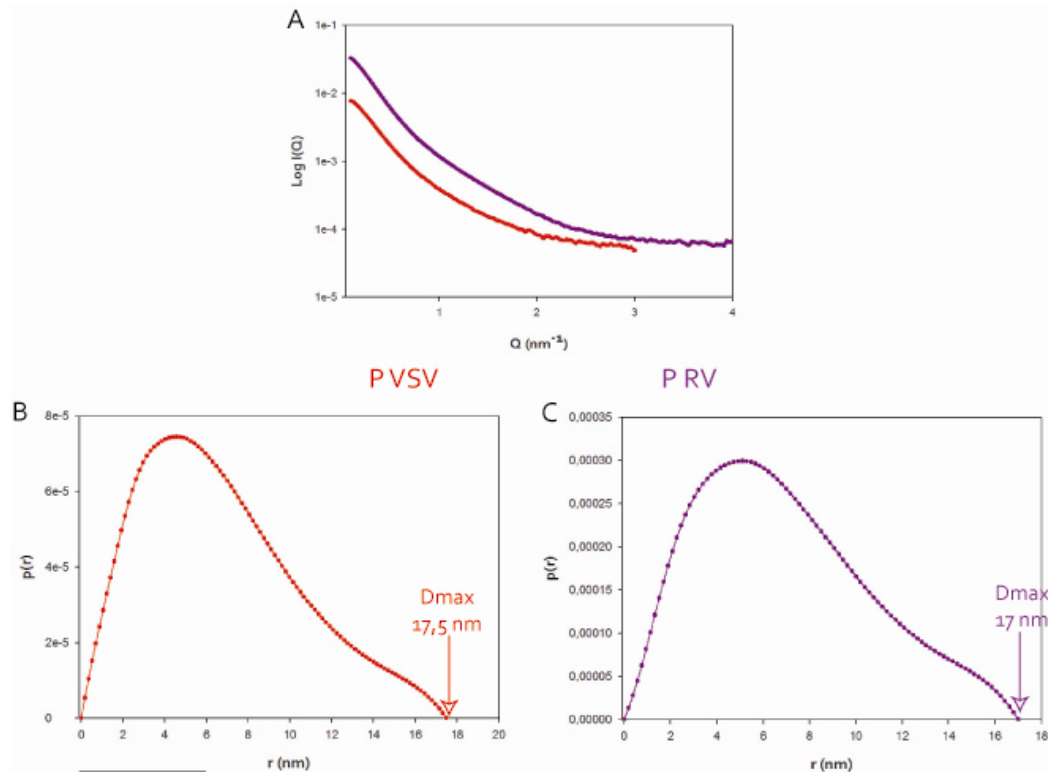
données à haute résolution de cristallographie des rayons X ou de RMN. En effet, les particules en solution diluée et monodisperse, adoptent toutes les orientations dans l'espace de part le mouvement brownien. La seule grandeur accessible expérimentalement est donc la moyenne des intensités de diffusions. Cependant, un des avantages de la diffusion aux petits angles, réside dans la sensibilité de la méthode envers les domaines flexibles et/ou mal définis, qui sont ininterprétables ou indétectables dans les cartes de densités électroniques lors des analyses cristallographiques. La construction d'un modèle est largement facilité par l'existence de structures haute résolution de fragment de la protéine (données de cristallographie des rayons X ou RMN). Pour obtenir une reconstruction 3D à partir d'images de microscopie électronique, cela nécessite un échantillon de plus grand poids moléculaire, et moins flexible. L'analyse par RMN (Résonance Magnétique Nucléaire) serait possible alors pour des protéines de cette taille, mais le plus souvent il faut trouver une condition de stabilité dans le temps, ce qui est difficile pour ces protéines qui possèdent des régions désordonnées. C'est dans l'optique d'obtenir des modélisations de la phosphoprotéine des *Rhabdoviridae* que nous avons utilisé le SAXS.

## A – Extraction des paramètres structuraux

Nous avons mesurés les courbes de diffusions des rayons X aux petits angles pour la phosphoprotéine du virus de la rage (P RV) et de la stomatite vésiculaire (P VSV). La diffusion sur des solutions de protéine à 3 mg/mL pour P RV et 5 mg/mL pour P VSV ont été mesurées pour deux distances échantillons-détecteur de 1 m ( $0,01 \text{ nm}^{-1} < Q < 0,95 \text{ nm}^{-1}$ ) et 5 m ( $0,1 \text{ nm}^{-1} < Q < 4,7 \text{ nm}^{-1}$ ) et les données ont été regroupées afin d'obtenir une seule courbe dans une gamme de vecteur de diffusion de  $0,01 \text{ nm}^{-1} < Q < 4 \text{ nm}^{-1}$  pour P RV et de  $0,01 \text{ nm}^{-1} < Q < 3 \text{ nm}^{-1}$  pour P VSV (figure 41, A). Les paramètres structuraux que nous avons pu extraire de ces courbes, notamment l'analyse par approximation de Guinier (valable aux très petits angles,  $Q.R_g < 1,3$ ) qui permet d'obtenir le rayon de giration de la protéine en solution, sont en accord avec les paramètres obtenus en SANS et publié dans (Gerard & al., 2007). Le rayon de giration, calculé à partir des données de diffusion à l'aide du logiciel PRIMUS (Konarev, 2003), pour la phosphoprotéine du virus de la rage (P RV) est de  $5,7 \pm 0,2 \text{ nm}$ , et de  $5,2 \pm 0,2 \text{ nm}$  pour la phosphoprotéine du virus de la stomatite vésiculaire (P VSV).

La fonction  $p(r)$  correspond à la distribution des paires de distances interatomiques de la molécule et possède une forme en cloche asymétrique (Gerard & al., 2007). La valeur  $p(r) = 0$  correspond à la distance maximale entre deux atomes de la molécule et donc la plus grande dimension de la protéine ( $D_{\text{max}}$ ). La valeur de  $D_{\text{max}}$  calculée à partir du programme GNOM (Svergun, 1992) sur la courbe de diffusion de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) est de 17,5 nm et celle de la phosphoprotéine du virus de la rage (P RV) en solution est de 17 nm. L'allure de cette courbe est

caractéristique d'une protéine asymétrique: la valeur de  $D_{max}/2$  est supérieure à la valeur du rayon de gyration, indiquant une forme allongée de la phosphoprotéine en solution.



**Figure 41** - (A) Courbes de diffusions des rayons X aux petits angles de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) et du virus de la rage (P RV). (B et C) Représentation de la fonction de distribution de paires  $p(r)$  (unité arbitraire) pour la diffusion de P VSV (B) et P RV (C). La valeur  $D_{max}$  est indiquée.

## B – Modélisations *ab initio*

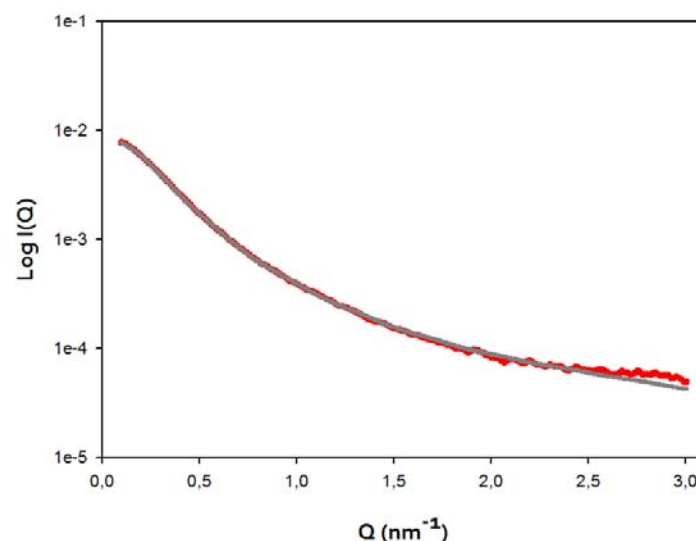
Afin de déterminer une enveloppe à basse résolution de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) et du virus de la rage (P RV), nous avons utilisé dans une première approche le programme DAMMIN (Svergun, 1999). Ce programme représente la protéine comme une suite de boules positionnées selon un réseau hexagonal compact et utilise une méthode de recuit simulé pour construire un modèle compatible avec les données de la courbe expérimentale de diffusion, sous des contraintes assurant la connectivité et la compacité du modèle.

### 1 – Modélisations *ab initio* de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV)

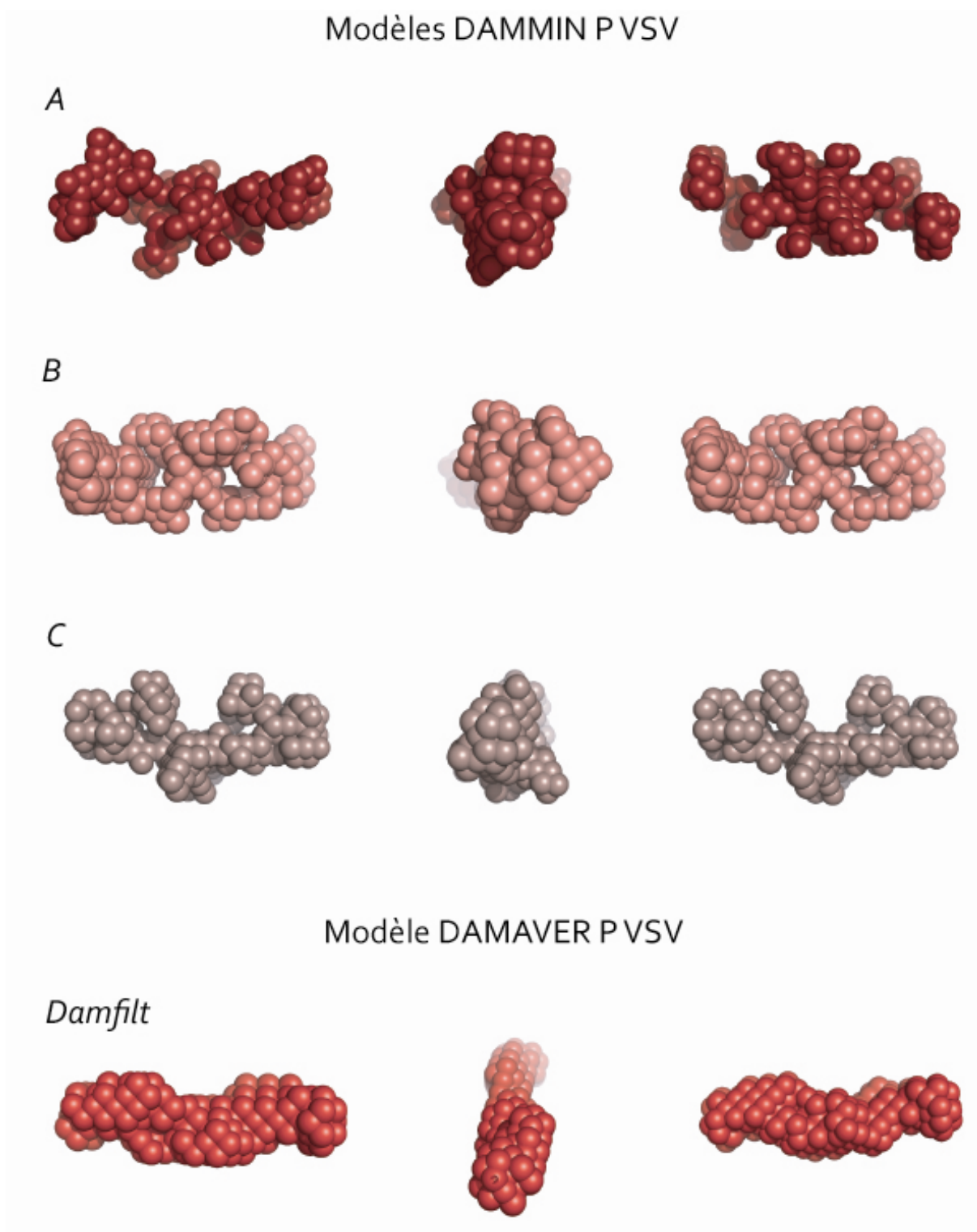
Nous avons réalisé 50 modèles indépendants avec le programme DAMMIN. Ayant démontré au préalable que la phosphoprotéine du VSV forme des dimères en solutions, nous avons imposé une contrainte de symétrie P2 au programme DAMMIN. Les modèles obtenus forment une galerie de formes

(figure 42, A, B et C) qui peuvent être comparées avec le programme DAMAVER (Volkova, 2003). DAMAVER détermine la valeur d'une pseudo-distance dite « *Normalized Spatial Discrepancy* » (NSD) entre les modèles qui est utilisée pour mesurer le degré de similarité entre les modèles. Lorsque les modèles sont similaires, typiquement la valeur de NSD est proche ou inférieure à 1. Pour les modèles de la protéine P VSV générés par DAMMIN, les valeurs de NSD varient entre 0,88 et 0,1 (valeur moyenne du NSD : 0,94 ; écart-type : 0,077) indiquant une excellente corrélation. En effet, chaque modèle de DAMMIN pris individuellement présente des différences, mais l'allure générale de ces modèles, reflétée dans le modèle filtré « *damfilt* », nous indique l'allure générale de la protéine (figure 42, *Damfilt*). La phosphoprotéine du virus de la stomatite vésiculaire a une forme allongée, assimilable à un cylindre en solution.

La courbe de diffusion expérimentale en fonction de  $Q$  est toujours comparée à la courbe de diffusion théorique des modèles issus de DAMMIN. L'accord entre les deux courbes de diffusions est mesuré par un facteur  $\chi$ , qui est proche ou inférieur à 1 dans le cas d'une bonne superposition des courbes. La comparaison de la courbe expérimentale (figure 43, courbe rouge) avec la courbe calculée pour l'un des modèles obtenus avec DAMMIN (figure 43, courbe grise) donne une valeur de  $\chi = 0,07$ , montrant la bonne qualité de l'ajustement.



**Figure 43** - Graphe de l'intensité de diffusion par la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) en fonction de  $Q$ . La courbe expérimentale est représentée en rouge et la courbe calculée à partir d'un modèle issu de DAMMIN est en gris.

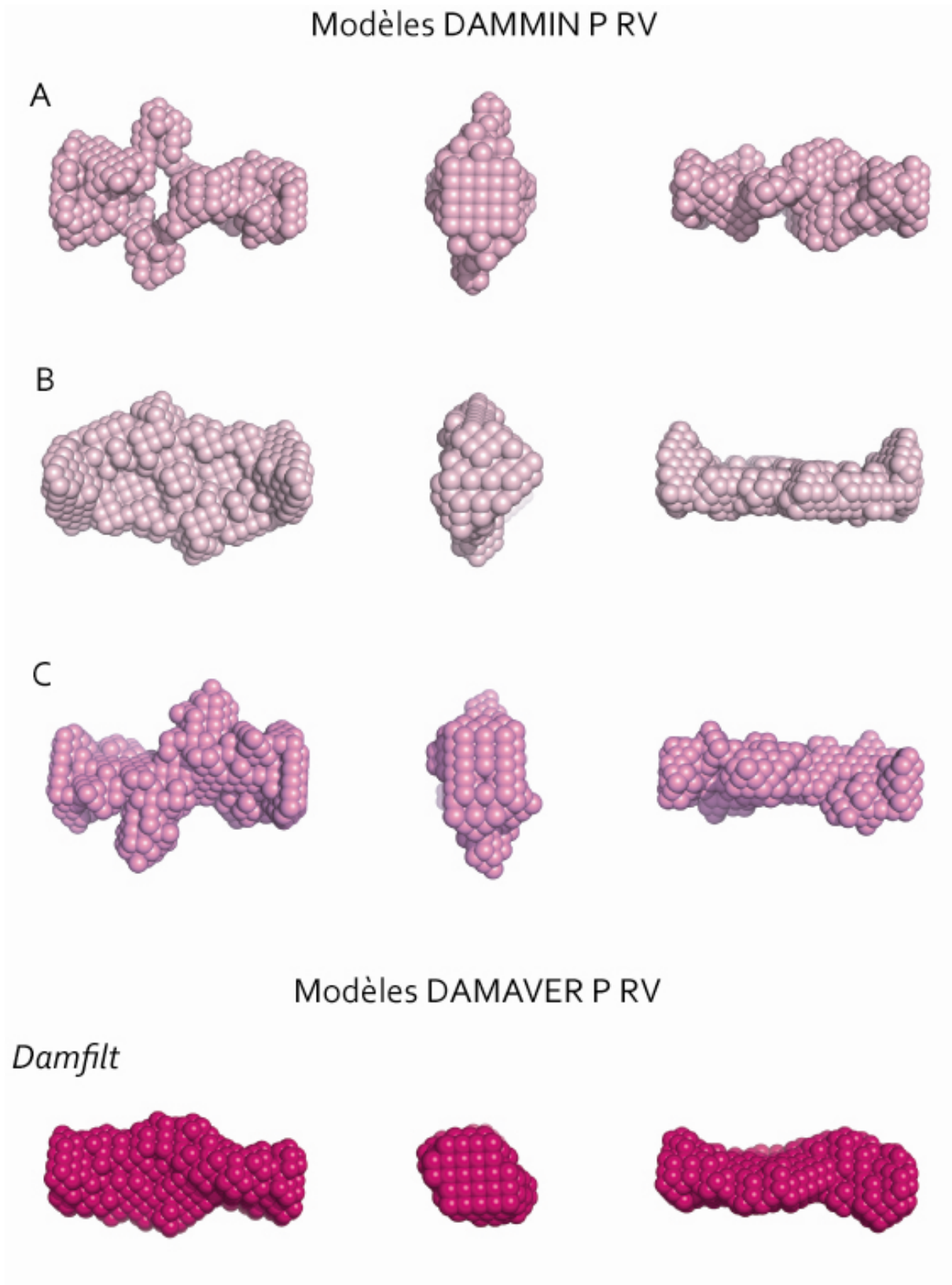


**Figure 42** - Modèles de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) obtenus avec le logiciel **DAMMIN** (modèles A, B et C, sous trois vues différentes). La forme globale de la protéine en solution est indiquée dans le modèle filtré « Damfilt » obtenu avec le logiciel **DAMAVER**. Chaque sphère à un diamètre de 5,0 Å.



## 2 – Modélisations *ab initio* de la phosphoprotéine du virus de la rage (P RV)

Nous avons obtenu 15 modèles indépendants pour la phosphoprotéine du virus de la rage avec le programme DAMMIN en imposant une symétrie P2 pour tenir compte de l'homodimérisation de la protéine en solution (figure 44, A, B et C). Les valeurs de NSD pour les modèles de la phosphoprotéine du virus de la rage varient entre 0,9 et 1,0 (valeur moyenne du NSD : 0,95, écart-type : 0,03) ce qui indique une excellente corrélation. Le modèle filtré « *damfilt* » nous indique une allure générale de la protéine (figure 44, *Damfilt*) comparable à celle de la phosphoprotéine du VSV, avec une forme allongée, assimilable à un cylindre en solution. De la même façon, les courbes théoriques des modèles s'ajustent correctement avec la courbe de diffusion expérimentale.



**Figure 44** - Modèles de la phosphoprotéine du virus de la rage (P RV) obtenus avec le logiciel DAMMIN (modèles A, B et C, sous trois vues différentes). La forme globale de la protéine en solution est indiquée avec le modèle filtré « Damfilt » obtenu avec le logiciel DAMAVER. Chaque sphère a un diamètre de 4,1 Å.

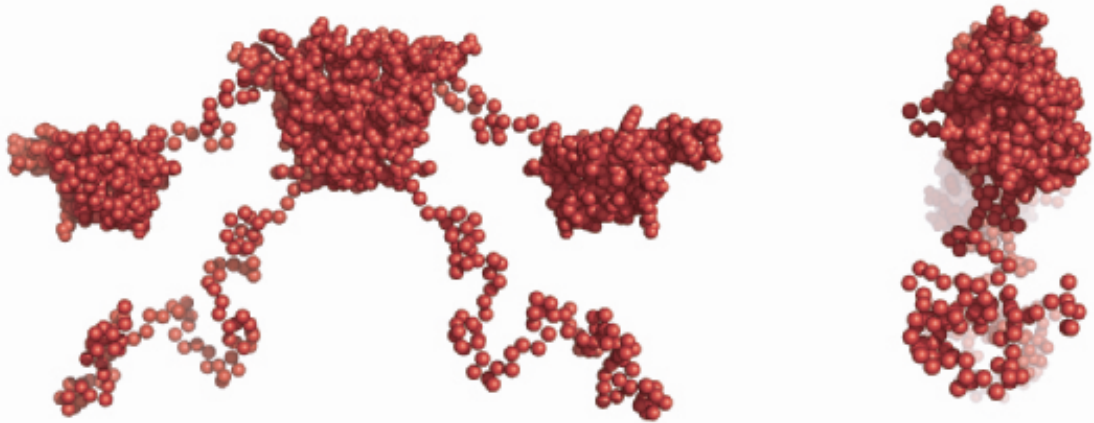
## C – Modélisation de la phosphoprotéine du virus de la stomatite vésiculaire (P VSV) avec le logiciel BUNCH

Cent modèles indépendants ont été obtenus avec le programme BUNCH. Les structures cristallographiques du domaine C-terminal de la phosphoprotéine du VSV (code pdb : 2K47) et du domaine d'oligomérisation de la protéine P VSV (code pdb : 2FQM) ont été utilisées pour la modélisation. Par ailleurs, le programme BUNCH ne peut considérer que des monomères. Nous avons donc créé un fichier de contact (représentant les contacts observés dans la structure 2FQM) et imposé la symétrie P2 aux modèles. Les trois meilleurs modèles indépendants de BUNCH sont présentés ci-après (figure 45), c'est-à-dire les modèles dont la valeur de  $\chi$  était la plus faible (0,18). Ces modèles ont aussi été contraints de façon à respecter les dimensions pour le dimère de phosphoprotéine de VSV que l'on a pu obtenir avec les modélisations *ab initio*. Les modèles ici ont été présentés en positionnant le domaine central de la protéine. On remarque que les modélisations créent deux « bras » de part et d'autre du domaine central, qui sont des zones flexibles et non globulaires, faites d'une succession de pseudo-atomes. Les modèles créent des dimères antiparallèles, imposés par les contacts imposés dans le domaine central d'oligomérisation.

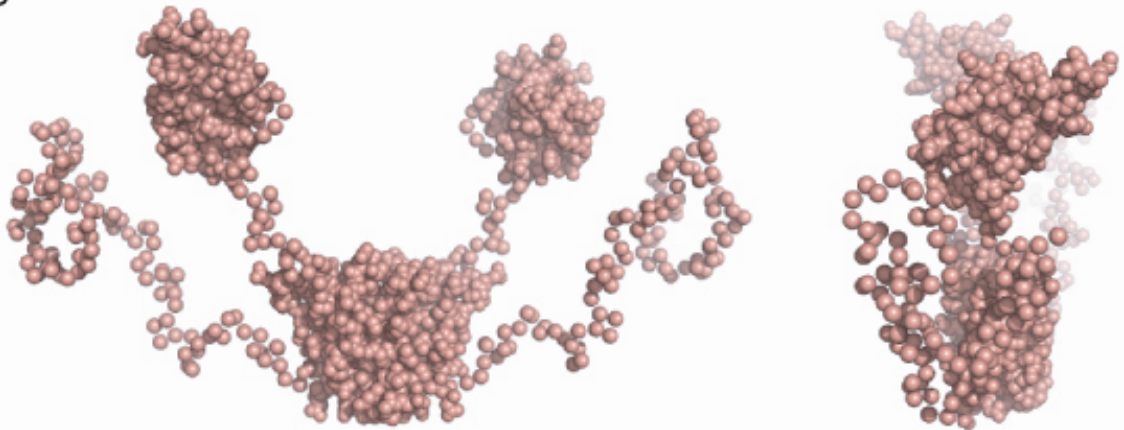
Nous avons initié une collaboration avec le Dr. Martin BLACKLEDGE au laboratoire de RMN à l'Institut de Biologie Structurale à Grenoble afin de la dynamique moléculaire sur les modèles de phosphoprotéine de VSV et d'estimer la flexibilité de la protéine. Aussi ces domaines flexibles pourraient être importants pour la fixation à d'autres partenaires, viraux ou cellulaires, provoquant un repliement induit.

## Modèles BUNCH P VSV

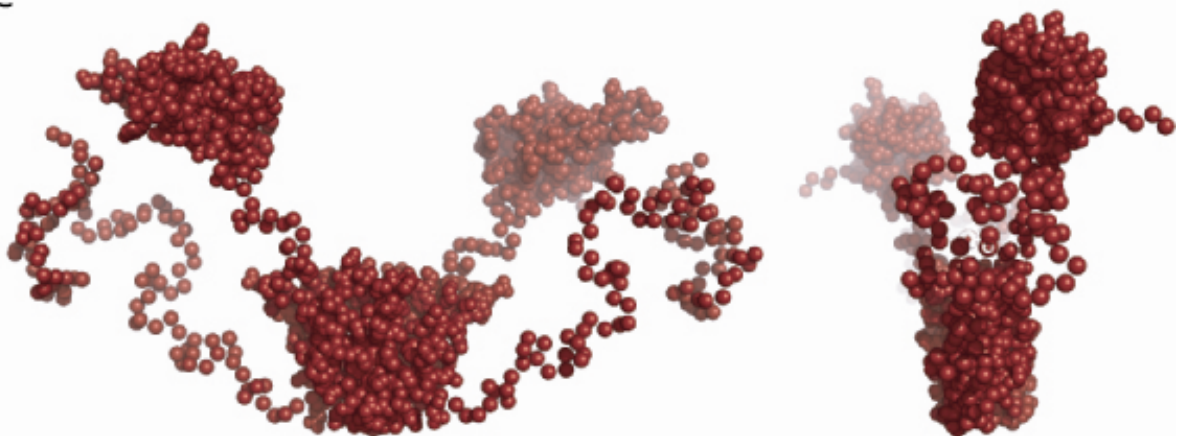
A



B



C



**Figure 45** -Trois modèles de la phosphoprotéine du virus de la stomatite vésiculaire (PVSV) obtenus avec le programme BUNCH. La structure cristallographique du domaine central (domaine responsable de la dimérisation de la phosphoprotéine du VSV, code pdb : 2FQM), et du domaine C-terminal (code pdb : 2K47) ont été imposés pour créer le modèle. Chaque sphère à un diamètre de 1 Å.

## D – Conclusion

L'analyse par SANS/SAXS nous a permis de modéliser l'organisation modulaire de la phosphoprotéine de *Rhabdoviridae*. En effet, la phosphoprotéine est un dimère allongé en solution et l'organisation en domaine est similaire pour la phosphoprotéine du VSV et de RV. Des conservations structurales existent même si il y a peu de conservation de la séquence en acides aminés (Ribeiro & *al.*, 2008) (Albertini & *al.*, 2008) (Albertini & *al.*, 2008). Grâce à la combinaison des données en solution (SAXS) avec les données de structures atomiques (Luo & *al.*, 2007) et structure cristallographique du domaine central de VSV (Ding & *al.*, 2006), il est possible d'obtenir un modèle de la protéine entière. La résolution de la structure du domaine central de la phosphoprotéine du virus de la rage (P<sub>CD</sub>) par cristallographie des rayons X ou par RMN permettra de modéliser de façon similaire la phosphoprotéine du virus de la rage et de confirmer cette organisation modulaire similaire au sein de *Rhabdoviridae*.

La compréhension des rôles fonctionnels et structuraux des domaines de la phosphoprotéine peut aider à la compréhension des rôles de cette protéine dans les mécanismes de transcription et de réplication virale. La caractérisation des complexes nucléoprotéine-ARN- phosphoprotéine va pouvoir aider à la compréhension de la dynamique de réplication des *Rhabdoviridae*.

**CHAPITRE IV : BASES STRUCTURALES DE L'INTERACTION ENTRE LE COMPLEXE  
NUCLEOPROTEINE – ARN ET LA PHOSPHOPROTEINE**

## ***I – Interaction entre les anneaux nucléoprotéine-ARN et la phosphoprotéine***

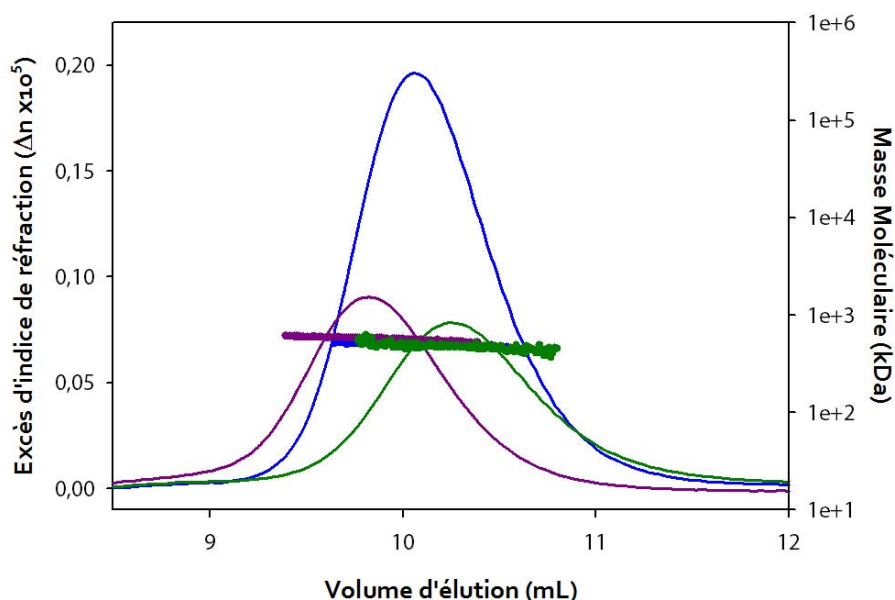
La structure des anneaux à 11 sous unités de nucléoprotéine a été résolue il y a 2 ans dans le laboratoire (Albertini & *al.*, 2006). La mise au point de la purification permet de séparer les anneaux contenant de 9 à 12 sous-unités de nucléoprotéine (N9, N10, N11 et N12), par une technique d'électrophorèse préparative (Albertini & *al.*, 2007). Une caractérisation par SEC-MALLS-RI (*Size Exclusion Chromatography coupled with Multi Angle Laser Light Scattering and refractometry*) nous a permis d'établir la stœchiométrie des complexes anneaux-phosphoprotéine, anneaux-domaine C-terminal de la phosphoprotéine ainsi que anneaux-P<sub>Δ91-131</sub>. L'étude biochimique par électrophorèse en condition native nous a permis d'observer les différentes formes discrètes de complexes formés. Les analyses structurales par cristallographie des rayons X et par SAXS (*Small Angle X-ray Scattering*) se sont avérées non concluantes, mais les essais par microscopie électronique, en utilisant la phosphoprotéine du virus de la rage en fusion avec la *Maltose Binding Protein* (MBP) (P-MBP) nous ont permis de localiser plusieurs densités électroniques supplémentaires sur les anneaux, qui pourrait correspondre à la protéine MBP. Aussi, nous avons cherché à identifier l'affinité de la phosphoprotéine pour la nucléoprotéine par résonance plasmonique de surface.

### **A – Caractérisations des complexes anneaux (nucléoprotéine-ARN) et de la phosphoprotéine par SEC-MALLS-RI**

#### ***1 – Caractérisation des anneaux N9, N10 et N11 par SEC-MALLS-RI***

Dans un premier temps, nous avons analysé les anneaux purifiés N9, N10 et N11 avec une colonne de chromatographie d'exclusion de taille (S200, HR 10/30, GE Healthcare) et nous avons déterminé leurs masses moléculaires par SEC-MALLS-RI. Comme présenté en [figure 46](#), les anneaux sont élués en un seul pic, indiquant leur homogénéité en solution. Les anneaux N9 sont élués à un volume de 10,2 mL, les N10 à 10,1 mL et les N11 à 9,8 mL. La différence de volume d'éluion indique que le rayon hydrodynamique est différent. Par la calibration de la colonne de chromatographie d'exclusion de taille (Gerard & *al.*, 2007) nous avons ainsi déterminé les rayons hydrodynamiques des anneaux N9, N10 et N11 en solution ([tableau 7](#)), qui sont respectivement de 7,4, 7,6 et  $8,0 \pm 0,1$  nm. La valeur de 8,0 nm que nous avons obtenu pour les N11 est en accord avec les distances observées dans la structure cristallographique et notamment avec le diamètre externe de l'anneau de 160 Å (Albertini & *al.*, 2006) (code pdb : 2GTT).

Les masses moléculaires déterminées par SEC-MALLS-RI, sont respectivement de  $470 \pm 40$  kDa,  $500 \pm 20$  kDa et  $560 \pm 30$  kDa pour les N9, N10 et N11. Ces masses moléculaires moyennes mesurées sont en accord avec les masses moléculaires théoriques calculées (457 632 Da pour les N9, 508 480 Da pour les N10 et 559 328 Da pour les N11) sur la base du nombre de monomères de nucléoprotéine et de la taille de l'ARN encapsidé dans ces différents anneaux (tableau 7).



**Figure 46** - Profil d'élution des complexes nucléoprotéine-ARN (N-ARN) sous formes d'anneaux avec 9 sous-unités de nucléoprotéine (N9, courbe verte), 10 sous-unités de N (N10, courbe bleue) et 11 sous-unités de N (N11, courbe violette). La masse moléculaire est calculée à partir des mesures de l'intensité de lumière diffusée et de l'indice de réfraction à différents temps et est représenté par des points sous le chromatogramme.

	Volume d'élution (mL)	Rayon hydrodynamique (nm)	Masse Moléculaire théorique (Da)	Masse Moléculaire mesurée (kDa)
<b>N9</b>	10,2	$7,4 \pm 0,1$	$9 * 50578 + 81 * 30 = 457\ 632$	$470 \pm 40$
<b>N10</b>	10,1	$7,6 \pm 0,1$	$10 * 50578 + 90 * 30 = 508\ 480$	$500 \pm 20$
<b>N11</b>	9,8	$8,0 \pm 0,1$	$11 * 50578 + 99 * 30 = 559\ 328$	$560 \pm 30$

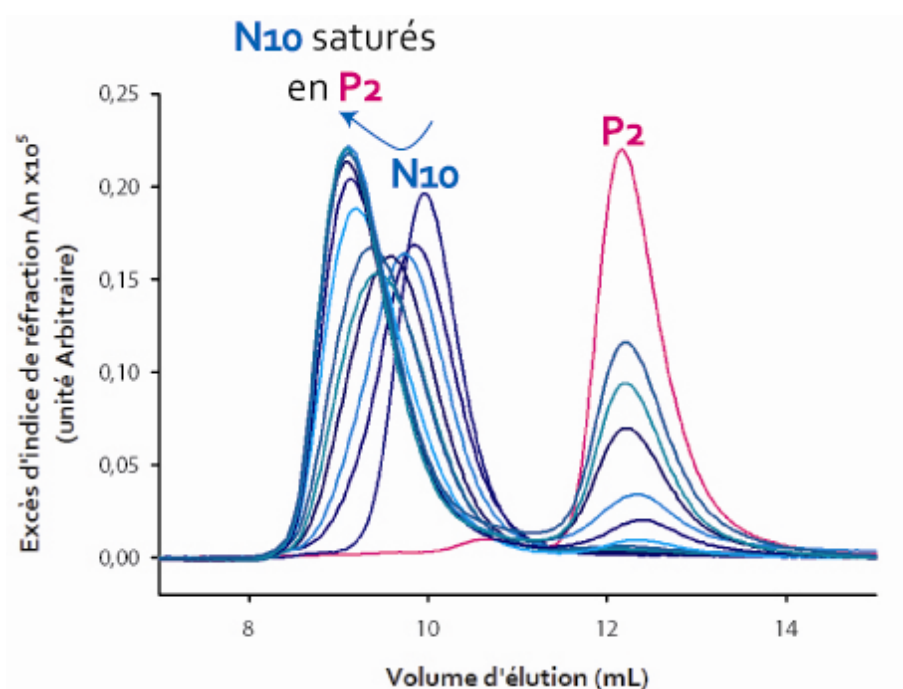
**Tableau 7** – Analyse par SEC-MALLS-RI des anneaux N9, N10 et N11. Le volume d'élution des N9, N10 et N11 est indiqué et le calcul du rayon hydrodynamique a été possible grâce à la calibration de la colonne d'exclusion de taille avec des protéines standards globulaires de rayon de Stokes (Rs) connus (Gerard & al., 2007). La masse théorique de chaque taille d'anneaux a été calculée : la masse moléculaire d'un monomère de nucléoprotéine du virus de la rage est de 50 578 Da et nous avons estimé à 30 Da le poids d'un nucléotide et est comparée à la masse moléculaire mesurée moyenne par SEC-MALLS-RI.

Cette caractérisation biophysique des anneaux de nucléoprotéines valide l'utilisation de cette méthode pour caractériser les différentes formes de complexes formés.



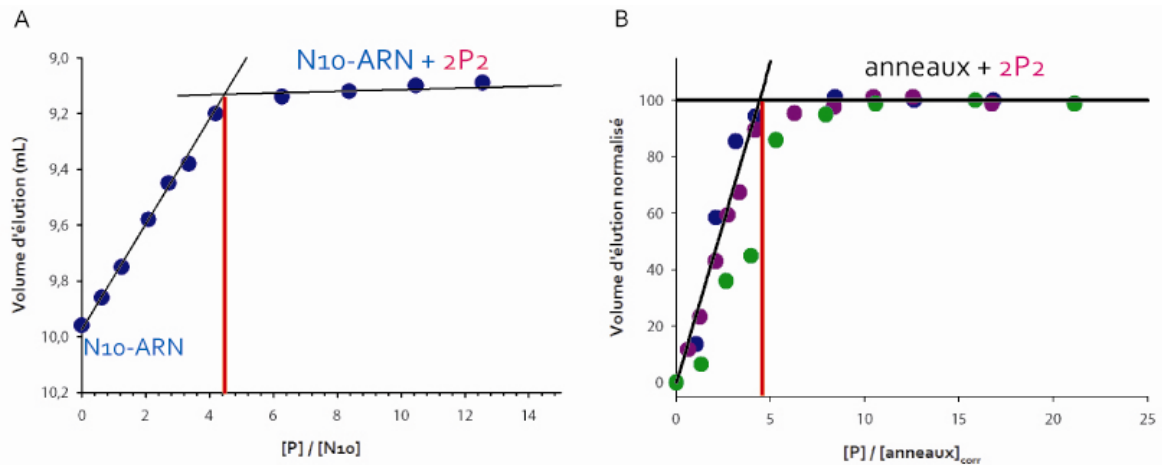
## 2– Stoechiométrie des complexes anneaux-phosphoprotéine par SEC-MALLS-RI

Nous avons titré les anneaux N9, N10 et N11 avec la phosphoprotéine du virus de la rage (P RV). La [figure 47](#) présente la titration de N10 avec la protéine P RV où des quantités croissantes de phosphoprotéine ont été incubées avec un échantillon d'anneaux N10. L'addition de la phosphoprotéine provoque un déplacement du pic de complexes N10, indiquant que la phosphoprotéine se fixe sur les anneaux. Ce déplacement n'évolue plus lorsque les anneaux sont saturés en phosphoprotéine ([figure 47, N10 saturés P<sub>2</sub>](#)). En effet, on peut alors observer l'apparition de dimères libres de phosphoprotéine non associés aux anneaux ([figure 47, P<sub>2</sub>](#)).



**Figure 47** - Profil d'élution des complexes N10 avec différentes quantités de dimères de phosphoprotéine. En rose est indiqué le profil d'élution de la phosphoprotéine du virus de la rage (P RV). Les différents tons de bleu présentent les profils d'élution des N10 avec différentes quantités de protéine P RV.

L'analyse de la position du pic d'élution en fonction du rapport de concentrations [phosphoprotéine] / [anneau] démontre que les anneaux N10 sont saturés par la fixation de deux dimères de phosphoprotéines ([figure 48, A](#)). En effet, la position du pic d'élution reste stable à 9,1 mL lorsqu'on augmente le rapport [P] / [N10], indiquant que les N10 ne peuvent pas fixer plus de deux dimères de phosphoprotéine. De la même façon, les titrations réalisées avec la phosphoprotéine pour les N9 et N11 montrent que les anneaux sont saturés avec deux dimères de phosphoprotéine (2 P<sub>2</sub>) ([figure 48, B](#)).

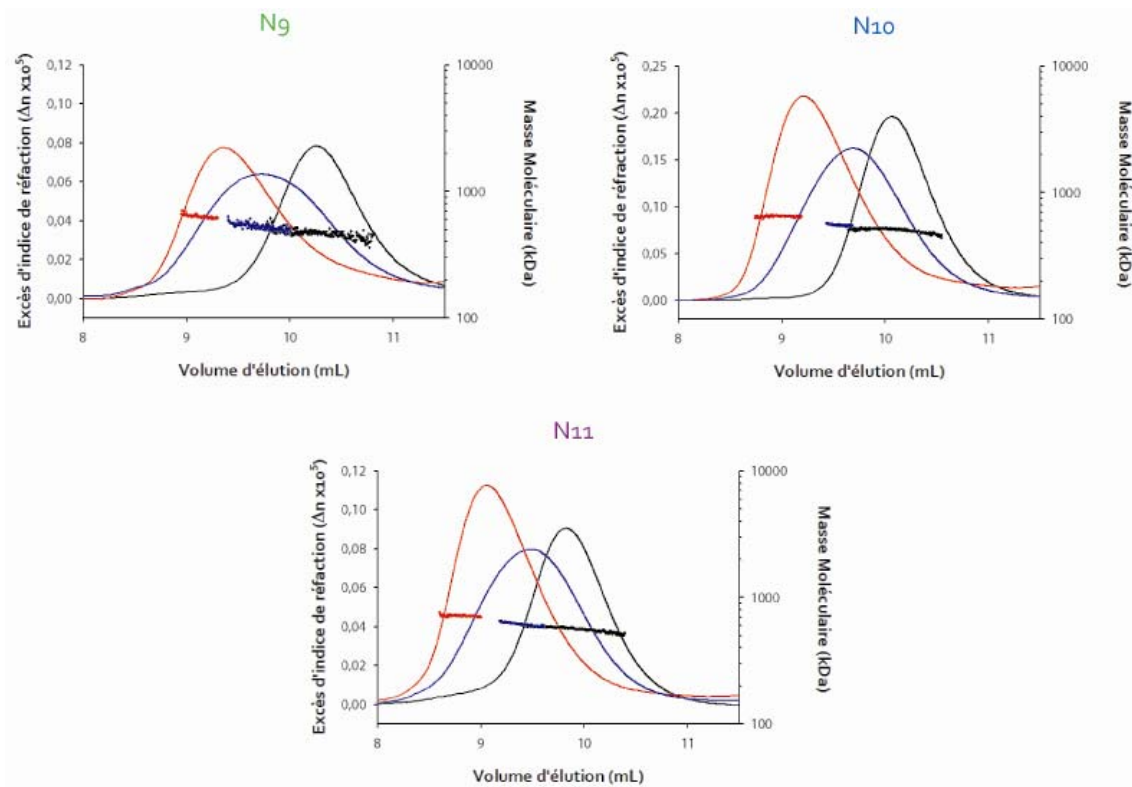


**Figure 48** - (A) Titrage des anneaux N10 avec la phosphoprotéine. (B) Titrage des N9 (vert), N10 (bleu) et N11 (violet) par la phosphoprotéine. Les données expérimentales sont représentées ainsi : position du volume d'éluion en fonction du ratio [phosphoprotéine] / [anneaux].

Deux espèces de complexes anneaux –phosphoprotéine sont formés et peuvent être visualisés sur le SEC-MALLS-RI (figure 49) :

- un complexe transitoire formé d'un anneau et d'un dimère de phosphoprotéine (P<sub>2</sub>),
- un complexe saturé formé par l'association d'un anneau et de deux dimères de phosphoprotéine (2P<sub>2</sub>).

La masse moléculaire de ces différents complexes a été calculée pour les N9, N10 et N11 (figure 49). Dans le tableau 8 sont récapitulés les volumes d'éluions, rayons hydrodynamiques et masses moléculaires mesurées pour les différents complexes anneaux-phosphoprotéine. D'après les expériences de SEC-MALLS-RI, la masse moléculaire du dimère de phosphoprotéine du virus de la rage est de  $69 \pm 2$  kDa. Le calcul des masses confirme que deux espèces de complexes anneaux phosphoprotéines existent (tableau 8).



**Figure 49** - Identification des complexes anneaux (N<sub>9</sub>, N<sub>10</sub> et N<sub>11</sub>), anneaux + un dimère de phosphoprotéine (+ P<sub>2</sub>) et anneaux + deux dimères de phosphoprotéine (+ 2P<sub>2</sub>). Les points correspondent au calcul de la masse moléculaire pour chaque point du chromatogramme.

	Volume d'élution (mL)	Rayon de Stoke (nm)	Masse Moléculaire (kDa)
<b>N<sub>9</sub></b>	10,2	7,4	470 ± 40
<b>N<sub>9</sub> + P<sub>2</sub></b>	9,7	8,1	530 ± 30
<b>N<sub>9</sub> + 2P<sub>2</sub></b>	9,4	8,7	640 ± 20
<b>N<sub>10</sub></b>	10,1	7,6	500 ± 20
<b>N<sub>10</sub> + P<sub>2</sub></b>	9,7	8,2	555 ± 10
<b>N<sub>10</sub> + 2P<sub>2</sub></b>	9,2	9,0	650 ± 10
<b>N<sub>11</sub></b>	9,8	8,0	565 ± 30
<b>N<sub>11</sub> + P<sub>2</sub></b>	9,5	8,5	615 ± 20
<b>N<sub>11</sub> + 2P<sub>2</sub></b>	9,1	9,2	720 ± 10

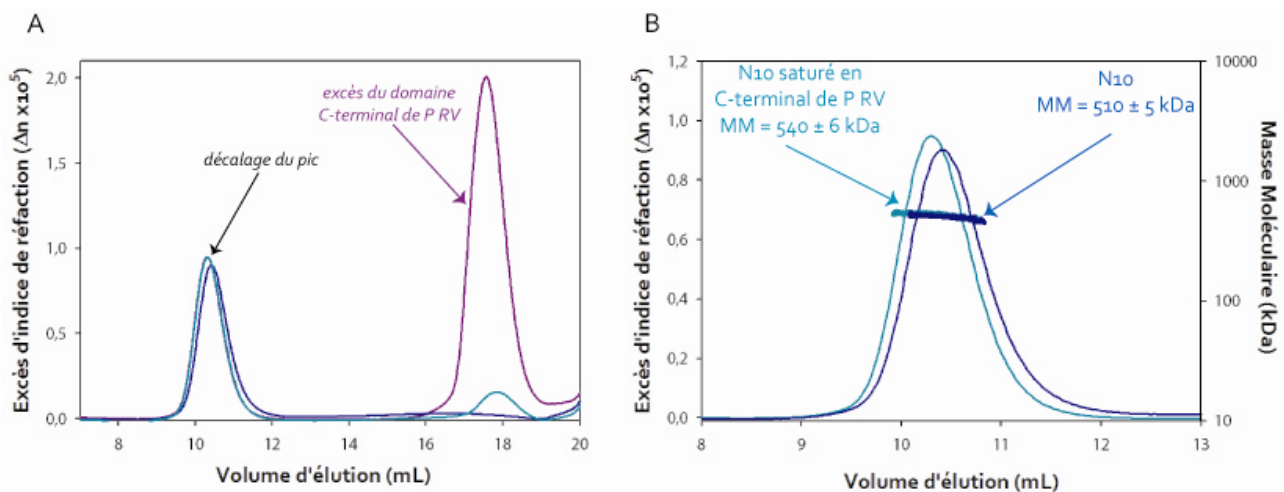
**Tableau 8** - Analyse des profils d'élutions des complexes anneaux –phosphoprotéines par SEC-MALLS-RI.

### 3 – Stoechiométrie des complexes anneaux-CterP RV par SEC-MALLS-RI

Les données de la littérature ont montré que le domaine C-terminal de la phosphoprotéine est responsable de l'interaction avec la nucléoprotéine sous sa forme nucleoprotéine-ARN (Schoehn & al.,

2001). La protéolyse du domaine C-terminal de la nucléoprotéine entraîne la perte de l'interaction avec la phosphoprotéine. Aussi, l'analyse structurale des surfaces électrostatiques du monomère de nucléoprotéine et du domaine C-terminal montre une complémentarité des surfaces électrostatiques (Albertini & *al.*, 2008). D'après cette analyse, il semble possible que chaque domaine C-terminal de P RV puisse interagir avec un monomère de nucléoprotéine avec une stoechiométrie 1 :1. Nous avons donc caractérisé la formation des complexes anneaux-domaine C-terminal de la phosphoprotéine du virus de la rage (CterP RV).

Dans un premier temps, nous avons vérifié par SEC-MALLS-RI que la protéolyse prolongée du domaine C-terminal de la nucléoprotéine provoquait la perte de l'interaction avec la phosphoprotéine (données non montrées). Ensuite, nous avons réalisé les mêmes expériences de titration en ajoutant des quantités croissantes de domaine C-terminal (CterP RV) à des anneaux N10 (Schoehn & *al.*, 2001). Le décalage du pic des anneaux N10 est très faible, même lorsque le complexe est saturé (présence d'excès du domaine C-terminal de P RV). En effet, dans ces conditions le volume du pic d'élution est de 10,3 mL pour les anneaux N10 et de 10,4 mL pour le complexe N10-CterP RV saturé (figure 50, A). La masse moléculaire mesurée par SEC-MALLS-RI pour le domaine C-terminal de la phosphoprotéine RV est de  $15 \pm 1$  kDa (Gerard & *al.*, 2008). La masse moléculaire calculée pour les N10 est de  $510 \pm 5$  kDa et la masse moléculaire calculée pour le complexe anneau-domaine C-terminal de P RV est de  $540 \pm 6$  kDa, ce qui indique que seuls deux domaines C-terminaux se sont fixés aux anneaux (figure 50, B).



**Figure 50** - (A) Profil d'élution des complexes anneaux N10 et anneaux-domaine C-terminal de P RV. (B) Analyse des masses des complexes par SEC-MALLS-RI.

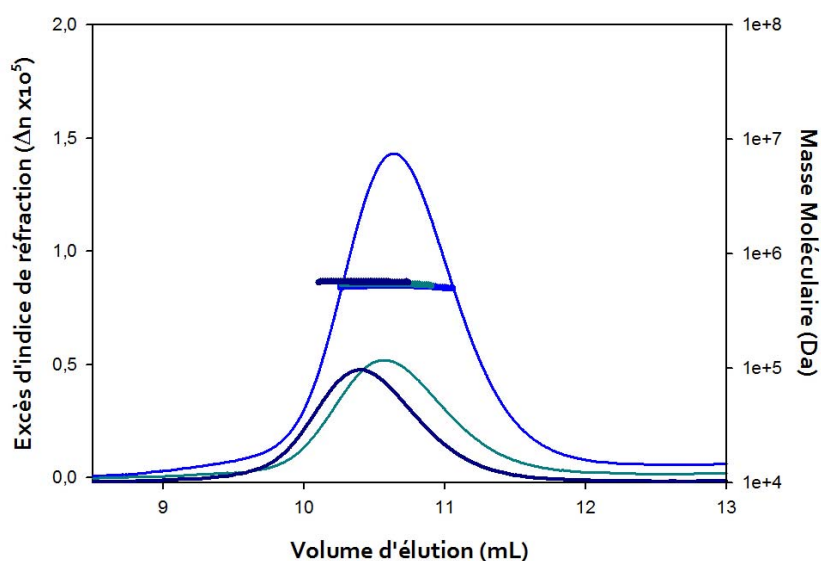
Nous avons mesurée la masse des complexes anneaux-CterP RV pour les N9, N10 et N11. L'écart de masse mesurée entre les anneaux seuls et en complexe avec le CterP RV est de 30 kDa, quel que soit le type d'anneau utilisé, indiquant que la stoechiométrie du complexe anneau-CterP RV est de deux

domaines C-terminaux de P RV. En effet, il semble qu'il existe un encombrement stérique au sein de l'anneau qui permet de fixer uniquement deux domaines C-terminaux de la phosphoprotéine.

Dans la particule virale on compte une phosphoprotéine pour deux nucléoprotéines, soit un dimère de phosphoprotéine pour quatre nucléoprotéines. Il semble que dans le contexte de l'anneau, cette règle soit conservée. Il semblerait que deux dimères de P RV se fixent par l'intermédiaire de deux domaines C-terminaux.

#### 4 – Stoechiométrie des complexes anneaux- $P_{\Delta 91-131}$ par SEC-MALLS-RI

Nous avons aussi caractérisé l'interaction entre les anneaux N10 et les différents mutants de la phosphoprotéine. En effet, il était important de vérifier quels domaines étaient responsables de l'interaction avec le complexe nucléoprotéine-ARN et si nous observions la même stoechiométrie pour le mutant  $P_{\Delta 91-131}$ , monomérique que pour le domaine C-terminal. Nous avons pu constater que seuls les mutants de P RV qui conservaient le domaine C-terminal de la phosphoprotéine (P RV,  $P_{\Delta N52}$ ,  $P_{\Delta 91-131}$  et  $P_{\text{CDT}}$ ) fixaient les anneaux, contrairement à  $P_{\Delta C120}$  et  $P_{\text{CED}}$  (données non montrées). Par ailleurs, nous avons confirmé la stoechiométrie du complexe N10-  $P_{\Delta 91-131}$ , où deux monomères de  $P_{\Delta 91-131}$  se fixent sur les N10 (figure 51). En effet, la masse moléculaire calculée pour le complexe N10-  $P_{\Delta 91-13}$  est de  $565 \pm 5$  kDa, (le mutant  $P_{\Delta 91-13}$  a une masse moléculaire calculée par SEC-MALLS-RI de  $30 \pm 2$  kDa (Gerard, 2008 #340)) et la masse moléculaire moyenne mesurée pour les N10 est de  $516 \pm 2$  kDa.



**Figure 51** - Profil d'élution des anneaux N10 (courbe bleue), en complexe avec le domaine C-terminal de la phosphoprotéine (courbe verte) ainsi qu'avec le mutant  $P_{\Delta 91-131}$  (courbe bleue foncée). La masse moléculaire moyenne mesurée pour les N10 est  $516 \pm 5$  kDa, de  $540 \pm 4$  kDa pour le complexe N10-CterP RV et de  $565 \pm 5$  kDa pour le complexe N10- $P_{\Delta 91-131}$ . La masse moléculaire mesurée pour le domaine C-terminal de la phosphoprotéine du virus de la rage (CterP RV) est de  $15 \pm 1$  kDa et celle pour le mutant  $P_{\Delta 91-131}$  est de  $30 \pm 2$  kDa, d'après (Gerard & al., 2008).

## 5 – Conclusions

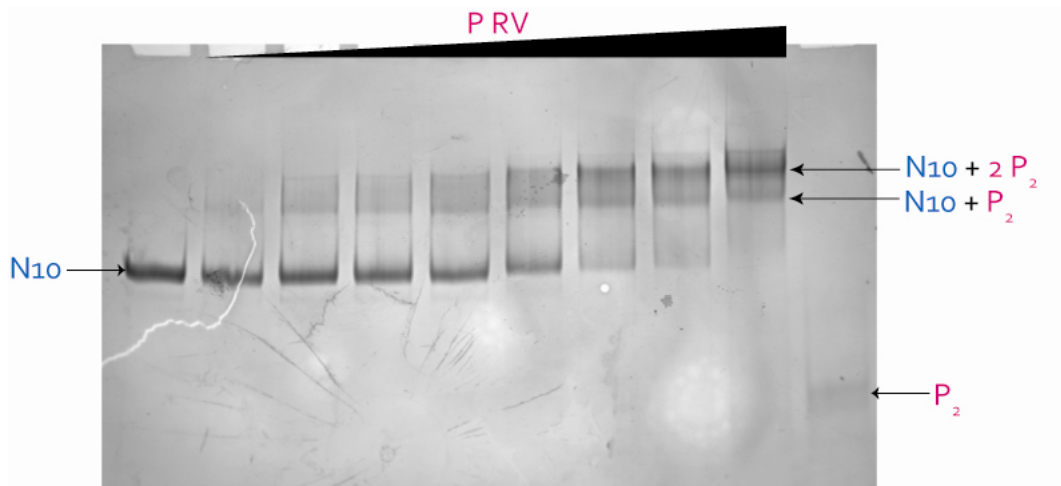
L'analyse par SEC-MALLS-RI de la masse des complexes anneaux-phosphoprotéine, anneaux-CterP RV et anneaux- P<sub>Δ91-131</sub> nous a montré que deux dimères de phosphoprotéine peuvent se fixer sur les anneaux (N9, N10 et N11) par l'intermédiaire de deux domaines C-terminaux. Nous avons pu isoler deux types de complexes anneaux-phosphoprotéine : les anneaux avec un dimère de P RV et les anneaux saturés avec deux dimères de P RV. Cependant, il n'a pas été possible de caractériser les deux formes de complexes anneaux-Cter RV par cette technique.

### B – Analyse des complexes anneaux N-ARN et phosphoprotéine par électrophorèse en conditions natives

Les anneaux avec un nombre de 9 à 12 sous-unités de nucléoprotéine ont été séparés par électrophorèse préparative (Albertini & *al.*, 2007). La même approche nous a permis de mettre en évidence les complexes formés avec la phosphoprotéine (P RV) (N-ARN-P), le mutant P<sub>Δ91-131</sub> (N-ARN-P<sub>Δ91-131</sub>).et le domaine C-terminal de la phosphoprotéine (CterP RV) (N-ARN-CterP) par l'observation d'un retard sur le gel d'électrophorèse en conditions natives.

#### 1 – Complexes N10-phosphoprotéine

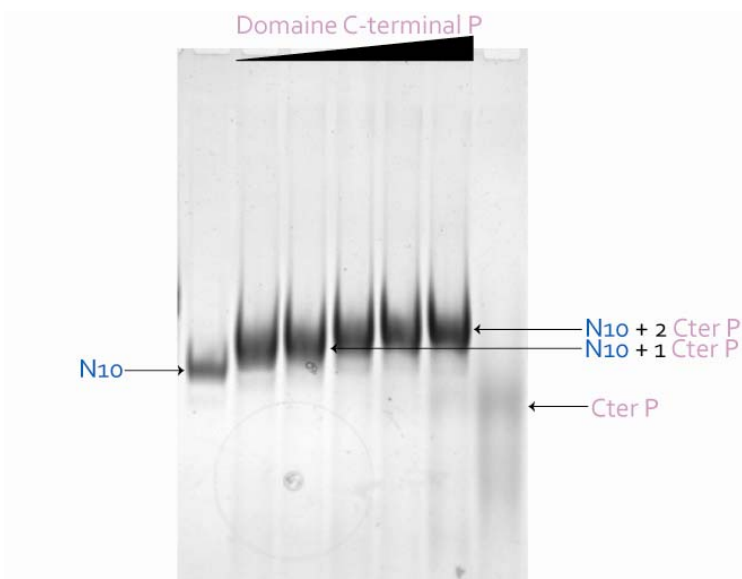
Le plus souvent, une partie des échantillons préparés pour la titration suivie par SEC-MALLS-RI étaient déposés sur gel d'électrophorèse en condition native. On observe un retard sur gel pour les complexes N10-phosphoprotéine formés (figure 52), ce qui confirme que la phosphoprotéine interagit avec la nucléoprotéine. Cette interaction est suffisamment forte pour persister lors de la migration à travers la matrice du gel natif, mais aussi sur chromatographie d'exclusion de taille (*partie A*). En augmentant la quantité de phosphoprotéine (figure 52), plusieurs formes discrètes du complexes N10-ARN-P peuvent être isolées. Le premier est formé par l'association entre un anneau N10 et un dimère de phosphoprotéine (N10 + P<sub>2</sub>), et le second correspond à la forme saturé du complexe, l'association d'un anneau N10 et de deux dimères de phosphoprotéine (N10 + 2 P<sub>2</sub>). L'observation de ces différentes formes est en accord avec l'analyse des masses des différents complexes obtenues par SEC-MALLS-RI.



**Figure 52** - Gel natif 4% de polyacrylamide mettant en évidence l'interaction entre des anneaux N10 et la phosphoprotéine. Le gel a été coloré au Bleu de Coomassie. On observe les formes discrètes du complexe N10-ARN-P, notées  $N10 + P_2$  et  $N10 + 2P_2$  (état saturé).

## 2 – Complexes N10-domaine C-terminal de la phosphoprotéine

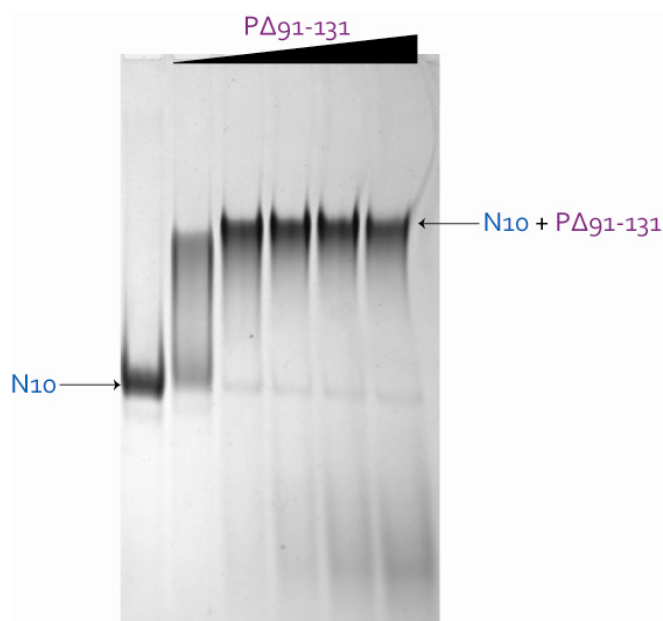
Le retard sur gel observé pour l'interaction anneaux (N10) et domaine C-terminal de la phosphoprotéine du virus de la rage (CterP RV) est plus petit (figure 53). Toutefois, deux formes de complexes N10-ARN-CterP sont visibles : une forme correspondant à l'association des anneaux de N10 et d'un domaine C-terminal de la protéine P ( $N10 + 1 \text{ CterP}$ ), et une forme saturée, correspondant à l'association des anneaux de N10 et deux domaines C-terminaux de la protéine P, en accord avec la masse moléculaire mesurée par SEC-MALLS-RI. Dans ces conditions natives, nous avons pu observer les deux types de complexes formés avec le domaine C-terminal de la phosphoprotéine du virus de la rage, même si le retard sur gel observé est faible comparé à la migration des N10.



**Figure 53** - Gel natif 4% de polyacrylamide mettant en évidence l'interaction entre des anneaux N10 et le domaine C-terminal de la phosphoprotéine (186-297) du virus de la rage. Le gel a été coloré au Bleu de Coomassie. On observe les formes discrètes du complexe N10-ARN-CterP, notées  $N10 + 1 \text{ CterP}$  et  $N10 + 2 \text{ CterP}$  (état saturé).

### 3 – Complexes N10-P<sub>Δ91-131</sub>

Nous avons aussi caractérisé en conditions natives l'interaction entre les anneaux N10 et le mutant P<sub>Δ91-131</sub> de la phosphoprotéine du virus de la rage. Ce mutant, délété de la région centrale de la phosphoprotéine, responsable de la dimérisation de la molécule, interagit toujours avec les anneaux mais avec une stoechiométrie différente (figure 54). Nous n'observons par électrophorèse en gel natif qu'une seule forme pour le complexe N10-ARN- P<sub>Δ91-131</sub>. Un anneau N10 saturé par 2 protéines mutantes P<sub>Δ91-131</sub> (figure 54), d'après nos études par SEC-MALLS-RI. Cependant, nous n'observons pas d'état intermédiaire de saturation, comme observé avec la phosphoprotéine entière, et avec le domaine C-terminal de P RV.



**Figure 54** - Gel natif 4% de polyacrylamide mettant en évidence l'interaction entre des anneaux N10 et le mutant P<sub>Δ91-131</sub> de la phosphoprotéine du virus de la rage. Le gel a été coloré au Bleu de Coomassie. On observe une forme saturée du complexe N10-ARN-P<sub>Δ91-131</sub>, ainsi que l'excès de P<sub>Δ91-131</sub>.

### 4 – Conclusions

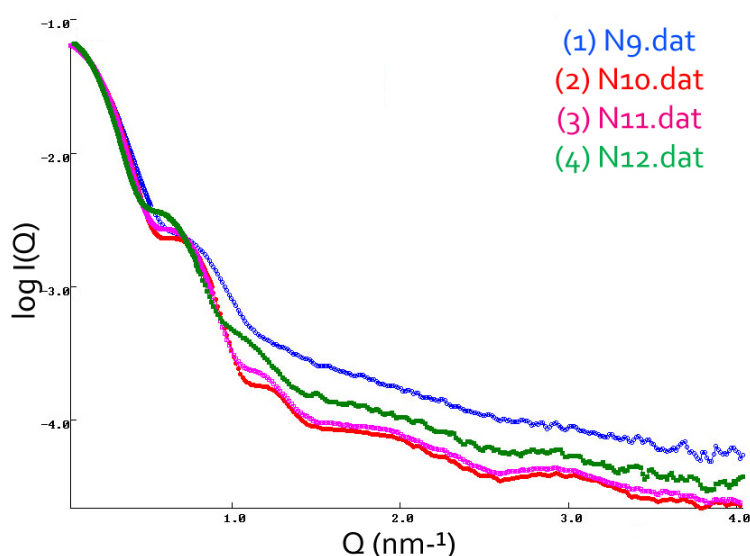
L'analyse en conditions natives des complexes anneaux-phosphoprotéine nous a permis de confirmer les formes d'anneaux seuls et en complexes observés et analysés par SEC-MALLS-RI : anneaux, anneaux + P<sub>2</sub> et anneaux + 2P<sub>2</sub>. Cette analyse en conditions natives nous a permis de voir deux formes de complexes des anneaux avec le domaine C-terminal de P RV, que nous n'avions pu observer en SEC-MALLS-RI, et dont la forme saturée correspond bien à l'association de deux domaines C-terminal de P RV sur un anneau. Aussi, l'analyse des complexes formés avec le mutant P<sub>Δ91-131</sub> nous permet de confirmer que la fixation de la phosphoprotéine s'effectue par l'intermédiaire de deux domaines C-terminaux, vraisemblablement, un par dimère de phosphoprotéine, même si seule la forme saturée du complexe anneaux- P<sub>Δ91-131</sub> a été observée.



## C – Etudes des complexes N-ARN-P et N-ARN-CterP par SAXS

### 1 – Données de diffusions des rayons X

La courbe de diffusion des rayons X obtenues pour les anneaux à 9, 10, 11 et 12 sous-unités montre un profil particulier (figure 55), qui met en évidence la présence d'une cavité à l'intérieur de la molécule étudiée (la structure des anneaux peut être assimilée à celle de beignets).

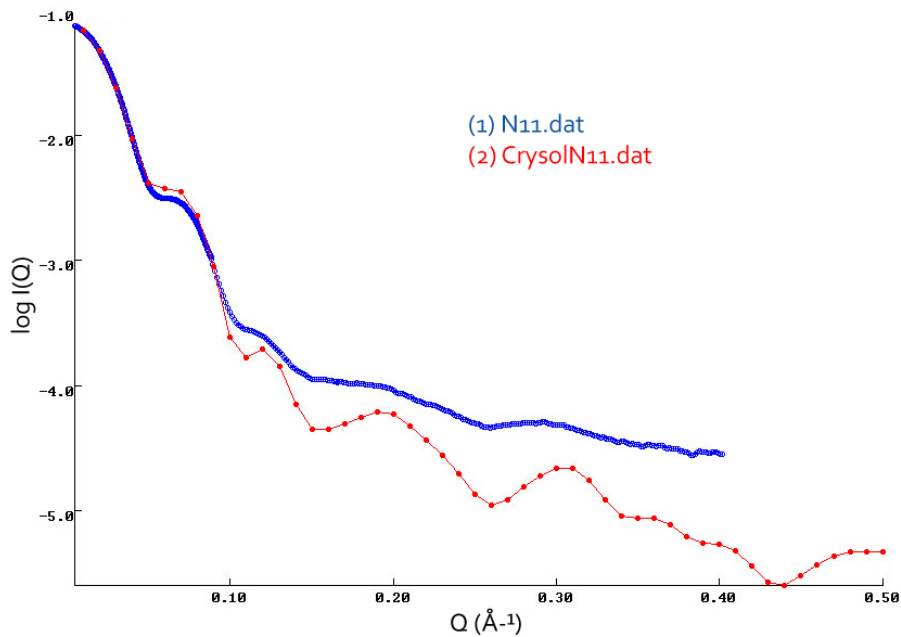


**Figure 55 - Courbes de diffusions des rayons X aux petits angles des différents anneaux.** Le graphique représente le log de l'intensité de lumière diffusée en fonction de  $s$  ( $\text{nm}^{-1}$ ). En vert, courbe de diffusion des anneaux N9, en bleu, courbe de diffusion des N10, en violet, courbe de diffusion des N11 et en rouge courbe de diffusion des N12. Les données ont été enregistrées pour des valeurs  $0,01 \text{ nm}^{-1} < s < 4 \text{ nm}^{-1}$ .

Le plot de Guinier des données obtenues aux très petits angles est linéaire, et le rayon de giration des différents anneaux a été calculé (tableau 9). Le rayon de giration obtenu pour les N9 est de  $6,2 \pm 0,1$  nm, de  $6,0 \pm 0,1$  nm pour les N10, de  $6,1 \pm 0,1$  nm pour les N11 et de  $6,8 \pm 0,1$  nm pour les N12. Cependant, les données que nous avons enregistré pour les N9 et les N12 ont été mesurées sur des échantillons à faible concentration ( $0,5 \text{ mg/mL}$ ), montre un bruit de fond plus important même après avoir groupé les mesures à deux distances échantillon-détecteur. Par la suite, nous nous sommes principalement intéressé aux anneaux N10 et N11.

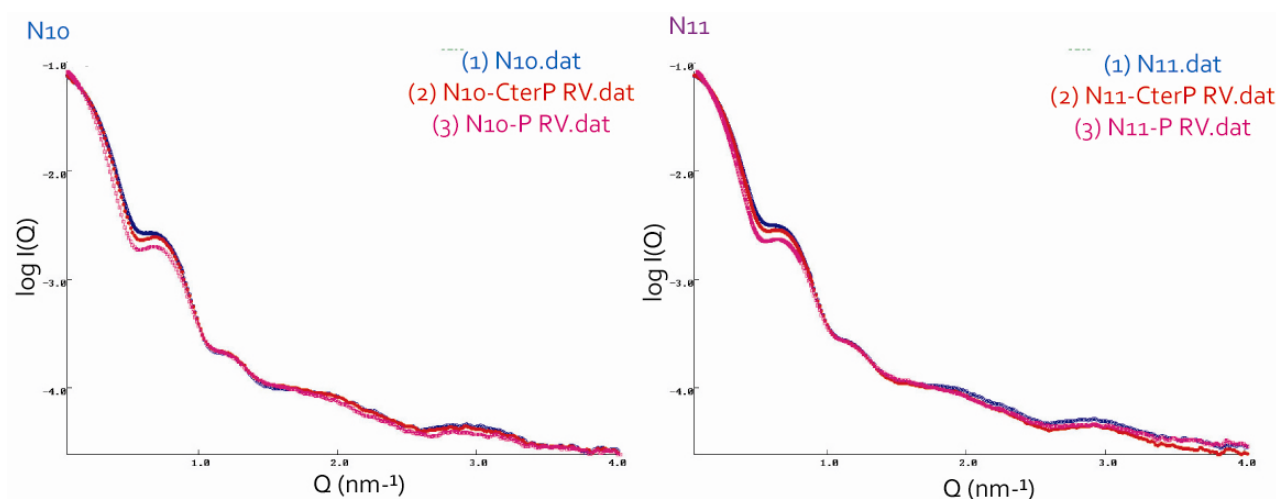
Pour vérifier que les données expérimentales s'accordaient avec les données atomiques, le programme CRY SOL (Svergun 1995) a été utilisé. Ce programme permet de calculer une courbe de diffusion théorique à partir d'une structure atomique connue. Nous avons calculé cette courbe pour un

anneau N11 (code pdb : 2GTT), présenté en [figure 56](#). Comme on peut le voir les courbes sont similaires pour des valeurs de  $s$  comprises entre  $0,01$  et  $0,1 \text{ \AA}^{-1}$ . En effet, pour les valeurs de  $s$  supérieures à  $0,15 \text{ \AA}^{-1}$ , nos données expérimentales ne sont pas d'assez bonnes qualités. Cependant, les informations sur la forme de l'anneau en solution indique que la conformation de l'anneau est la même en solution et dans le cristal. Le rayon de gyration de  $6,1 \text{ nm}$  obtenu à partir de la mesure expérimentale est en bon accord avec le rayon de gyration calculé à partir de la courbe théorique de SAXS obtenue à partir de la structure cristallographique (programme CRY SOL).



**Figure 56** - Comparaison des courbes de diffusion expérimentales et théorique (généralisé grâce au logiciel CRY SOL) pour les anneaux N11.

Les courbes de diffusions des rayons X aux petits angles ont été enregistrées pour les complexes N10 et N11 saturés en phosphoprotéine (N10-P RV et N11-P RV) et en domaine C-terminal de phosphoprotéine (N10-CterP RV et N11-CterP RV) ([figure 57](#)). Le plot de Guinier des données obtenues aux très petits angles est linéaire, et le rayon de gyration des différents complexes a été calculé. Dans ce [tableau 9](#) est récapitulée l'analyse des valeurs de  $R_g$  obtenus pour chacun des types d'anneaux et pour les complexes. L'allure de ces courbes de diffusion est similaire, mais le rayon de gyration calculé par l'approximation de Guinier aux très petits angles, indiquent que les complexes ont un  $R_g$  plus important. La différence de  $R_g$  est significative lorsque le complexe anneaux-phosphoprotéine est formé, puisque le  $R_g$  augmente de  $1,2$  et  $1,3 \text{ nm}$  pour les N10 et N11 respectivement. Pour les complexes avec le CterP RV, le rayon est sensiblement différent.



**Figure 57** - Comparaison des courbes de diffusions des rayons X aux petits angles des complexes **N10**, **N10-P** et **N10-CterP** et **N11**, **N11-P** **N11-CterP**. Le graphe représente le log de l'intensité de lumière diffusée en fonction de  $s$  ( $\text{nm}^{-1}$ ). (N10) En bleu, courbe de diffusion des N10, en rouge, courbe de diffusion de N10-CterP et en violet courbe de diffusion N10-P. (N11) En bleu, courbe de diffusion des N11, en rouge courbe de diffusion des N11-CterP et en violet courbe de diffusion de N11-P. Les données ont été enregistrées pour des valeurs  $0,01 \text{ nm}^{-1} < s < 4 \text{ nm}^{-1}$ .

Anneaux	Rayon de gyration (Rg) en nm
<b>N9</b>	$6,2 \pm 0,1$
<b>N10</b>	$6,0 \pm 0,1$
<b>N10CterP</b>	$6,1 \pm 0,1$
<b>N10P</b>	$7,2 \pm 0,1$
<b>N11</b>	$6,1 \pm 0,1$
<b>N11CterP</b>	$6,2 \pm 0,1$
<b>N11P</b>	$7,4 \pm 0,2$
<b>N12</b>	$6,8 \pm 0,2$

**Tableau 9** - Valeurs des rayons de gyrations des anneaux et complexes anneaux-phosphoprotéine, anneaux-domaine C-terminal de P RV obtenus à partir de l'approximation de Guinier sur les courbes de diffusions de SAXS.

Ces données sont en accord avec les formes de complexes anneaux-phosphoprotéine et anneaux-domaine C-terminal de la phosphoprotéine que nous avons pu observer en SEC-MALLS-RI et électrophorèse en conditions natives. En effet, le rayon de gyration des complexes anneaux-phosphoprotéine est plus grand ce qui est en accord avec une élution à des volumes plus petits sur SEC-MALLS-RI.

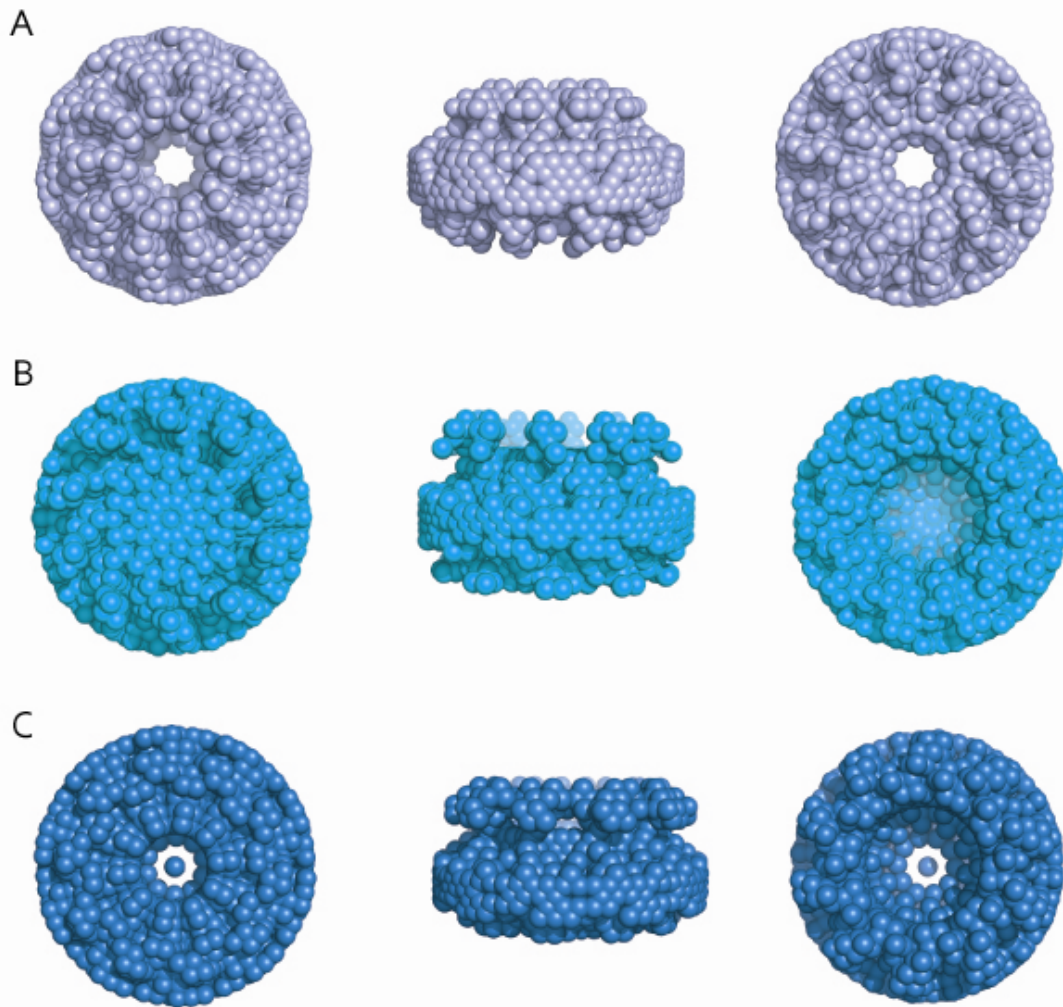
## 2 – Modélisations *ab initio* des anneaux

Les modélisations *ab initio* permettent d'obtenir une idée de l'enveloppe tridimensionnelle d'un objet calculée à partir de la courbe de diffusion des rayons X. Les données que nous avons obtenues pour les N9 et N12 ont été mesurées pour des concentrations faibles (0,5 mg/mL) et les modèles *ab initio* que nous avons obtenus ne sont pas de bonnes qualités. Les valeurs de NSD que nous avons obtenu pour ces modèles étaient supérieures à 1, et les ajustements de courbes de diffusions expérimentales avec les courbes calculées à partir du modèle étaient de mauvaises qualités. Les modèles indépendants créés avec le logiciel DAMMIN pour les anneaux N12 montrent une hétérogénéité. Nous avons donc principalement travaillé avec les données concernant les anneaux N10 et N11, seuls ou en complexe avec le domaine C-terminal de la phosphoprotéine (CterP RV) ou avec la phosphoprotéine (P RV).

Dans un premier temps, nous avons réalisé 15 modèles indépendants en suivant la même procédure avec le logiciel DAMMIN. La seule contrainte imposée a été la symétrie P10 pour la modélisation *ab initio* des anneaux N10 et la symétrie P11 pour la modélisation *ab initio* des anneaux N11. Pour chaque type d'anneaux, nous présentons trois modèles indépendants de DAMMIN ainsi que le modèle filtré par le programme DAMAVER (figure 58 pour les N10 et figure 59 pour les N11). Comme on peut le voir sur les modèles individuels de DAMMIN, le programme tend à trouver la cavité à l'intérieur de l'anneau. Le modèle filtré créé par le programme DAMAVER « *damfilt* » nous permet d'avoir une idée de la forme globale de la molécule, en forme de beignet.

Les dimensions des modèles obtenus respectent les paramètres obtenus expérimentalement par diffusion des rayons X aux petits angles.

### Modèles DAMMIN des anneaux N10



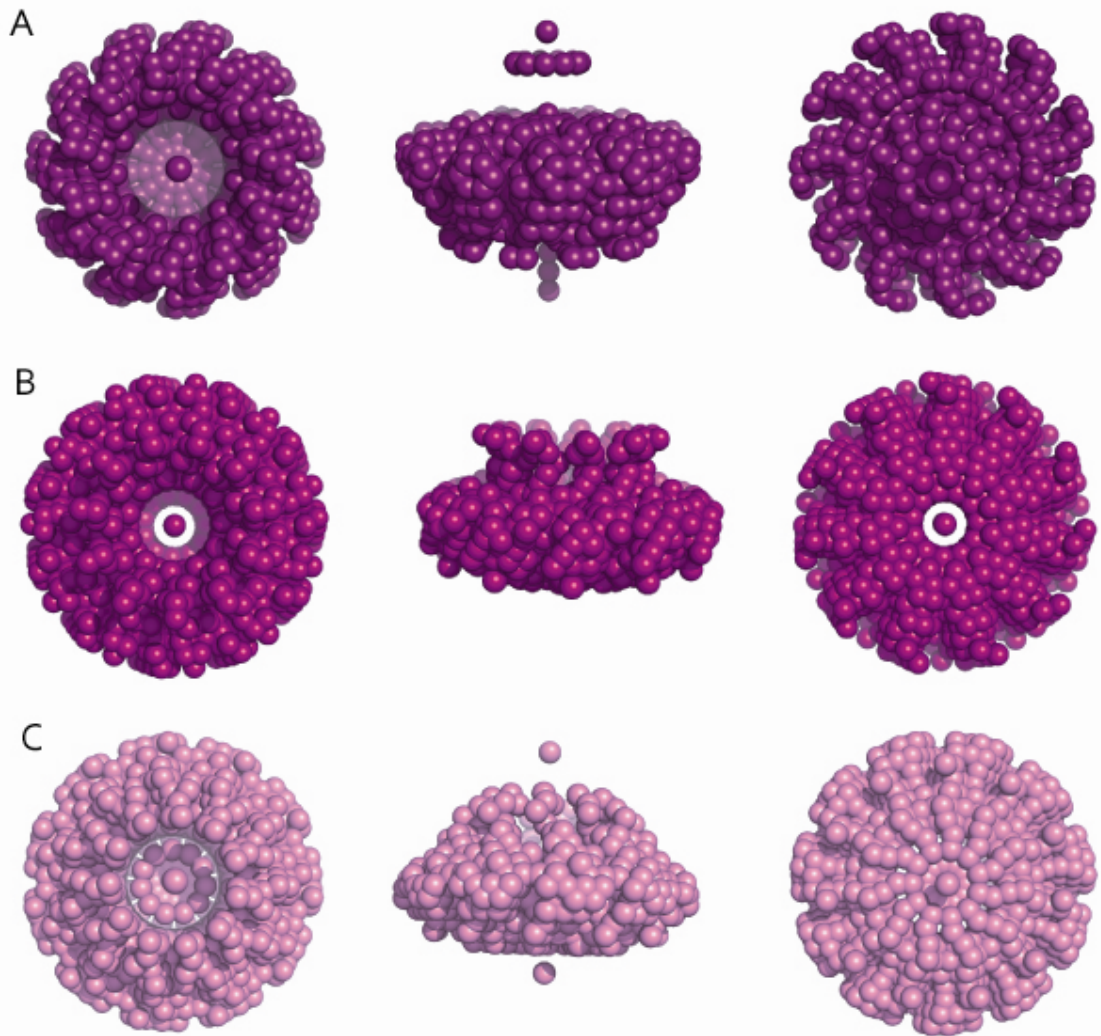
### Modèles DAMAVER des anneaux N10

*Damfilt*



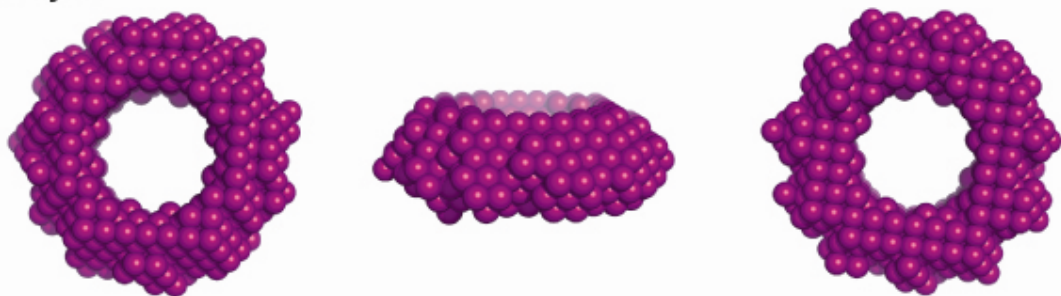
**Figure 58** - Modèles des anneaux de nucléoprotéine-ARN N10 du virus de la rage obtenus avec le logiciel DAMMIN (modèles A, B et C, sous trois vues différentes). La forme globale du complexe protéine-ARN en solution est indiquée avec le modèle *Damfilt* obtenu avec le logiciel DAMAVER. Chaque sphère a un diamètre de 5,0 Å.

### Modèles DAMMIN des anneaux N11



### Modèles DAMAVER des anneaux N11

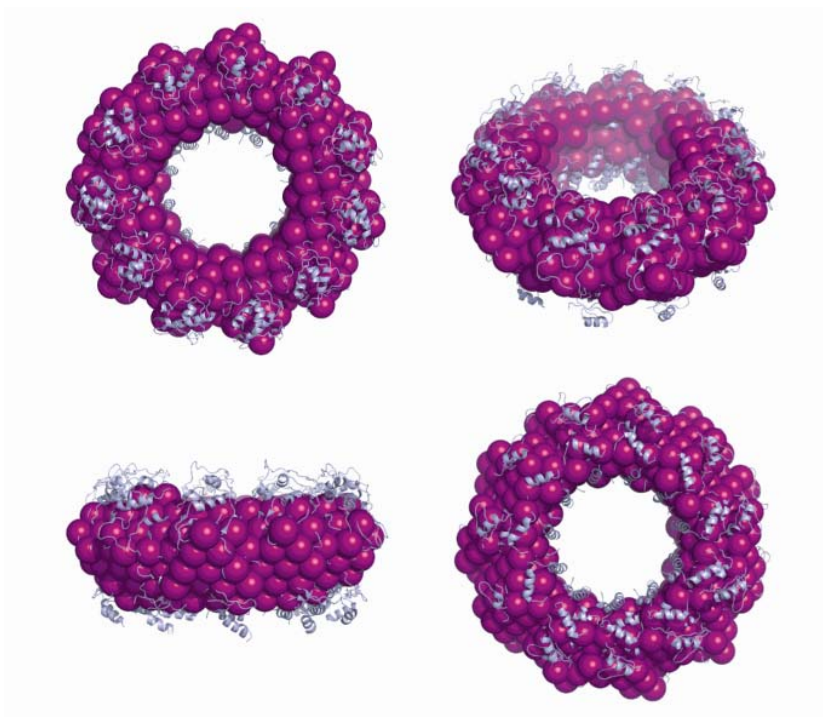
*Damfilt*



**Figure 59** - Modèles des anneaux de nucléoprotéine-ARN N11 du virus de la rage obtenus avec le logiciel DAMMIN (modèles A, B et C, sous trois vues différentes). La forme globale du complexe protéine-ARN en solution est indiquée avec le modèle Damfilt obtenu avec le logiciel DAMAVER. Chaque sphère a un diamètre de 5,0 Å.

### 3 – Superposition de la structure cristallographique des N11 avec le modèle filtré des N11 (damfilt)

Le modèle filtré obtenu avec le programme DAMAVER a été superposé à la structure atomique des N11 en utilisant le logiciel SUPCOMB (figure 6o). Lorsque nous superposons la structure cristallographique d'un anneau N11 avec le modèle issu de cette modélisation *ab initio*, nous pouvons remarquer que la structure à haute résolution est capable d'occuper convenablement le volume à basse résolution déterminé. Ce résultat nous permet d'apprécier la qualité de la méthode car le volume déterminé à partir de la diffusion de lumière de la protéine en solution correspond au volume occupé par la structure. Ainsi on peut estimer qu'il n'y a pas de modification majeure dans l'arrangement tridimensionnel global entre la structure cristalline et la structure en solution.

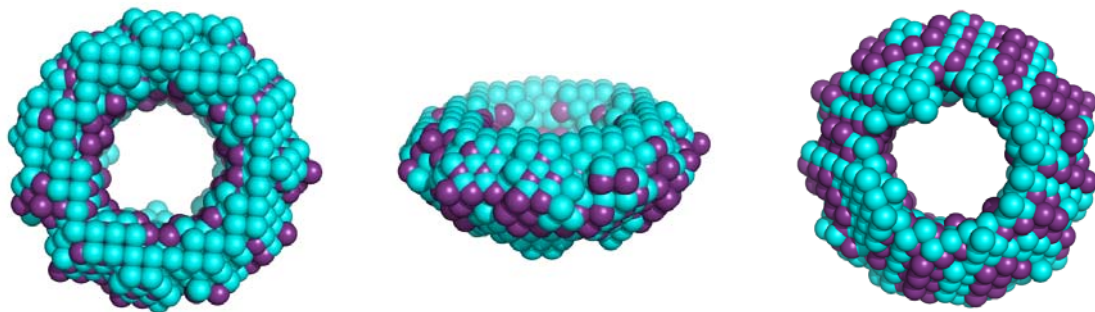


**Figure 6o** - Superposition de la structure cristallographique des N11 avec le modèle « damfilt » (DAMAVÉR). Cette superposition a été obtenue grâce au programme SUPCOMB. La structure des N11 (code pdb : 2GTT) est représentée en bleu clair et le modèle filtré « damfilt » en violet.

### 4 – Modélisations *ab initio* des complexes N11, N11-phosphoprotéine et N11-domaine C-terminal de la phosphoprotéine

Les modélisations *ab initio* des complexes anneaux-phosphoprotéine et anneaux-domaine C-terminal nous ont posé problème. En effet, même si nous connaissons la stoechiométrie des complexes, celle-ci ne peut être imposée au programme DAMMIN. Cependant, la symétrie P11 des anneaux reste

conservée. Pour chacun des complexes, nous avons réalisé 15 cycles indépendants avec le programme DAMMIN, en imposant une symétrie P11 ou P1 et nous avons ensuite comparés les modèles indépendants avec le programme DAMAVER. Les valeurs de NSD que nous avons obtenu pour les modèles N11-CterP avec la symétrie P11 imposée variaient de 0,73 à 0,82 ( $0,785 \pm 0,05$ ), et avec une symétrie P1 imposée de 0,76 à 0,86 ( $0,8 \pm 0,05$ ). Cependant, les modèles obtenus en imposant ou non la symétrie étaient sensiblement les mêmes. Nous avons voulu comparer le modèle filtré obtenu pour la modélisation *ab initio* des N11 avec celui obtenu pour le complexe N11-CterP RV. Le logiciel SUPCOMB a été utilisé afin de superposer ces deux modèles filtrés (figure 61) : à cette basse résolution, il n'est pas possible d'identifier une densité supplémentaire pouvant correspondre à la fixation du domaine C-terminal de la phosphoprotéine sur les N11. Aussi, l'analyse est réalisée en solution et il est possible que le signal fournit par la fixation du CterP RV soit moyenné.

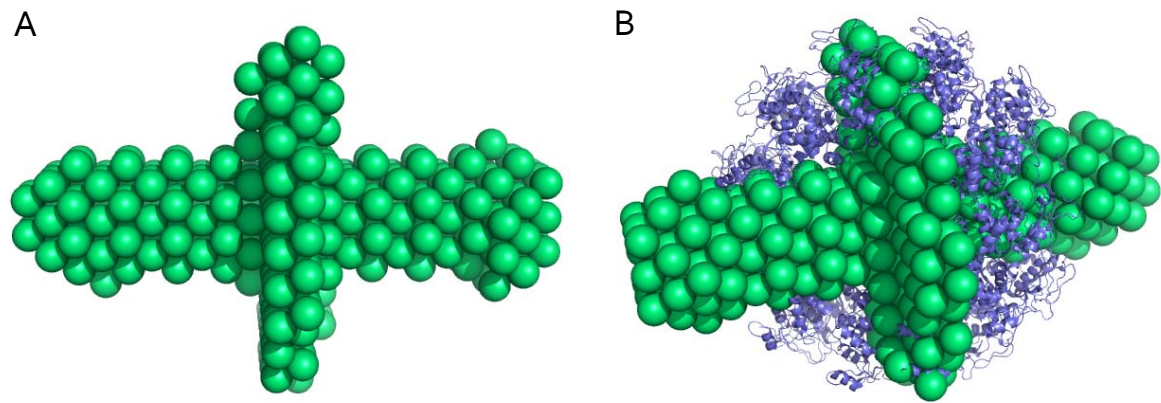


**Figure 61** – Superposition avec le logiciel SUPCOMB des modèles filtrés (damfilt) pour les modélisations *ab initio* des N11 (violet) et des complexes N11-Cter P RV (cyan). Trois vues sont présentées. Le diamètre des sphères est de 4,2 Å.

De la même façon, 15 cycles indépendants DAMMIN en imposant la symétrie ou non ont été réalisés pour le complexe N11-P RV. Les valeurs de NSD pour les modèles N11-P RV sont plus élevées, de 0,87 à 0,92 ( $0,89 \pm 0,02$ ) où aucune symétrie n'a été imposée et de 0,94 à 1,4 ( $1,1 \pm 0,2$ ) où la symétrie P11 a été imposée. Les modèles que nous avons obtenus respectent les données expérimentales, mais les modèles obtenus possèdent une forme particulière en « croix », où il semblerait que la phosphoprotéine se fixe à l'intérieur de l'anneau, et la densité cylindrique de la phosphoprotéine est modélisée de part et d'autres de l'anneau (figure 62, A). Cependant, lorsqu'on utilise le programme SUPCOMB, la superposition des anneaux ne semble pas adéquate (figure 62, B).

Nous essayons aussi de réaliser des « docking » du domaine C-terminal de la phosphoprotéine sur les anneaux N11, de façon à obtenir un modèle (type pdb) pour lequel nous pouvons créer la courbe théorique de diffusions de rayons X aux petits angles (logiciel CRY SOL) et ainsi comparer les données expérimentales et du modèle.





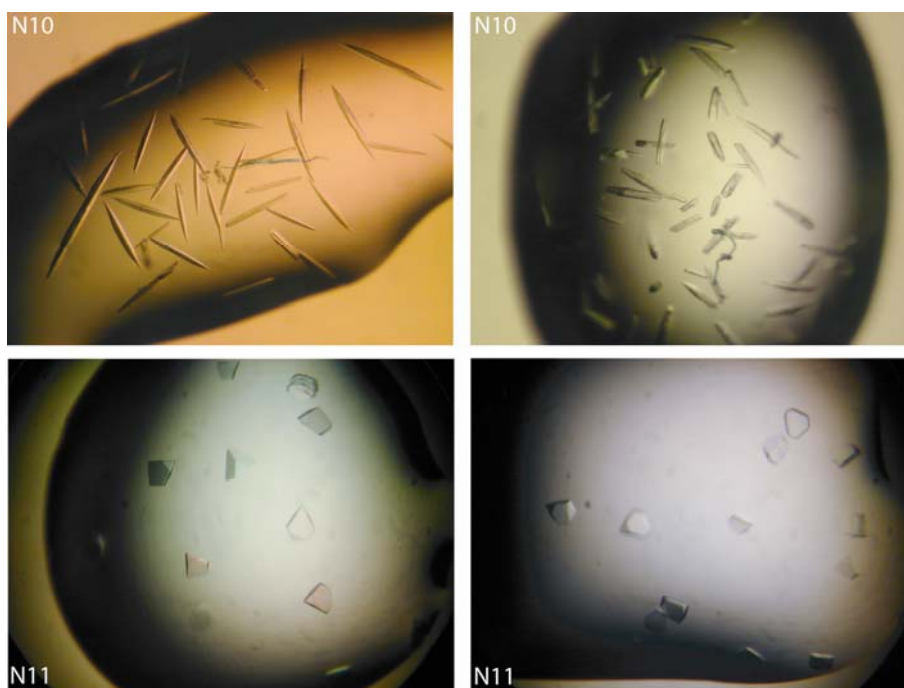
**Figure 62** – (A) Modèle filtré pour le complexe N<sub>11</sub>-P RV obtenu avec le programme DAMAVER à partir de 15 cycles indépendants DAMMIN. (B) Superposition du modèle filtré N<sub>11</sub>-P RV avec la structure atomique (2GTT, en bleu) avec le programme SUBCOMB. Chaque sphère est représentée avec un diamètre de 4,2 Å.

## D – Essai de co-cristallisation

Les anneaux à 9, 10 et 11 sous-unités de nucléoprotéine cristallisent dans de nombreuses conditions, avec toutefois un pouvoir de diffraction limité. La résolution de diffraction des cristaux de N11 a été optimisée grâce à une procédure de déshydratation (Albertini & *al.*, 2007) permettant ainsi de résoudre la structure des anneaux à 11 sous-unités de nucléoprotéine (Albertini & *al.*, 2006). Le complexe formé entre les anneaux et la phosphoprotéine ou son domaine C-terminal se comportant de façon stable, nous avons donc envisagé plusieurs stratégies en vue d’obtenir sa structure.

### 1 – Trempage de cristaux de N11 et N10

Nous avons répété de façon manuelle les cristaux d’anneaux N10 (condition : 9,5% PEG 6K, 1,3M LiCl et 0,1M MES pH 6) (figure 63, N10) et d’anneaux N11 (condition : 60% tascimate et 10 mM L-cystéine) (figure 63, N11). Ces cristaux ont été ainsi aisément reproduits pour des concentrations d’anneaux variant de 4 à 8 mg/ml. Les cristaux pour les N11 avaient une taille plus importante que ceux obtenus auparavant, avec des dimensions de 300 par 450 µm. Nous avons ensuite pêcher ses cristaux et nous les avons trempé dans une goutte contenant de la phosphoprotéine du virus de la rage (P RV) ou son domaine C-terminal (CterP RV) (équilibrée dans la même condition de cristallisation des N10 ou N11), dans des temps variant de 30 min à plusieurs heures. Avant la congélation des cristaux dans le flux d’azote liquide (100°K), nous avons ensuite procéder à la déshydratation des cristaux dans des temps variant de 6 à 20 min.



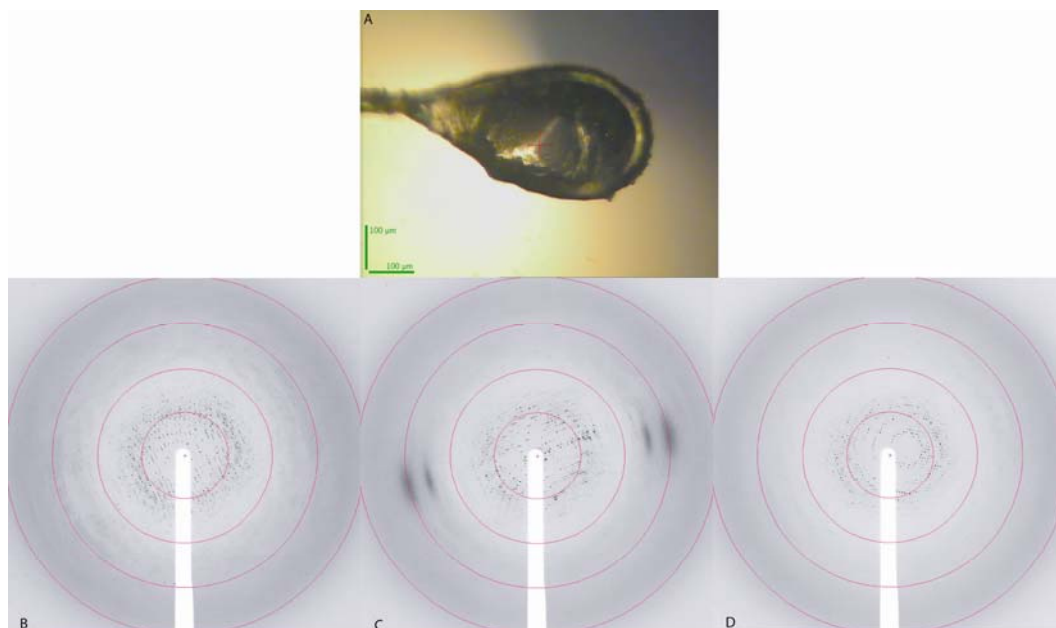
**Figure 63** - Cristaux obtenus en gouttes suspendues (1  $\mu$ l + 1  $\mu$ l) de N10 (condition : 9,5% PEG 6K, 1,3M LiCl et 0,1M MES pH 6) et N11 (condition : 60% tascimate et 10 mM L-cystéine) reproduit de (Albertini, 2007 #400).

Nous avons ensuite mesurés ces cristaux sur la ligne de rayonnement synchrotron de l'ESRF (*European Radiation Synchrotron Facility, Grenoble*) ID23eh1 dédiée à la cristallographie des rayons X. Les cristaux de N10 diffractant très mal (entre 10 et 15 Å de résolution) nous avons collecté quatre jeux de données natifs pour les cristaux de N11 trempés :

- Jeu 1 : N11 trempé dans le domaine C-terminal de la phosphoprotéine (CterP RV) avec 10 min déshydratation,
- Jeu 2 : N11 trempé dans le CterP RV avec 6 min déshydratation,
- Jeu 3 : N11 trempé dans le CterP RV avec 11min déshydratation,
- Jeu 4 : N11 trempé dans la phosphoprotéine du virus de la rage (P RV) avec 7 min déshydratation.

Le jeu de données n° 1 à une résolution raisonnable collecté sur un cristal de N11 trempé dans du domaine C-terminal (figure 64, A) comportait des taches de diffraction à 3,5 Å sur les premiers clichés (figure 64, B). Après une collecte complète et l'analyse des statistiques avec le logiciel MOSFLM (Leslie 2006), la résolution globale s'est avérée être à 6 Å à cause des dommages liés aux radiations (figure 64, C & D). En effet, après la collecte de quelques images, les points de diffraction se limitaient à 6 Å. Le groupe d'espace a été déterminé avec le programme MOSFLM comme étant orthorhombique de type P222. En effet, après l'analyse des données après l'indexation et l'intégration (MOSFLM), les extinctions systématiques selon h et k ont permis de préciser le groupe d'espace en P2<sub>1</sub>2<sub>1</sub>2, identique à celui obtenu pour les N11. Pour les 3 autres jeux de données, de la même façon des taches de diffraction apparaissaient

aux alentours de 3,5 / 4 Å de résolution sur les premières images, mais la résolution globale est aux alentours de 6 Å. Nous avons déterminé le groupe d'espace des cristaux en utilisant le programme MOSFLM. Dans le [tableau 10](#) est récapitulé les informations relatives aux collectes de données.



**Figure 64** - Clichés de diffraction collecté pour le cristal de N11 trempé plusieurs heures dans le domaine C-terminal de P RV (jeu de données n°1). (A) Cliché du cristal dans la boucle. (B) Image 1. (C) Image 45. (D) Image 90. Cercles roses indiquant la résolution (respectivement 2,9, 3,9, 5,8 et 11,7 Å).

Les jeux de données sont de bonnes qualités, quasi complets mais la résolution reste limitée à environ 6 Å. Pour tous les cristaux trempés dans le CterP RV, le groupe d'espace ne change pas ( $P2_12_12$ ) alors que le cristal trempé dans la P RV a un groupe d'espace toujours orthorhombique mais de type  $C222_1$ . Nous avons alors vérifié que l'empilement des anneaux dans la maille n'était pas complètement aberrant. Nous avons reproduit avec le programme Pymol (Delano scientific LLC) l'unité asymétrique du jeu de données n°4 et nous avons pu constater que l'empilement était très similaire à celui présent dans le groupe d'espace  $P2_12_12$ . Il semblerait donc que le trempage du cristal de N11 dans la protéine P RV ait affecté le groupe d'espace, l'empilement restant similaire aux cristaux appartenant au groupe d'espace  $P2_12_12$ .

Par la suite, nous avons résolu la structure par remplacement moléculaire en utilisant un anneau N11 comme modèle ceci dans le but de localiser la densité électronique de la phosphoprotéine ou du domaine C-terminal dans les cartes expérimentales. Malheureusement, après plusieurs essais, nous n'avons pu obtenir des cartes de qualité suffisante. Il n'a pas été possible de localiser une densité supplémentaire, quel que soit le jeu de donnée traité. Plusieurs facteurs peuvent expliquer ceci : les cartes obtenues sont à une résolution trop basse (6 Å) pour pouvoir visualiser la densité électronique de la

phosphoprotéine ou de son domaine C-terminal. Il est également possible que le trempage du cristal dans la protéine P RV ou du CterP RV n'ait pas entraîné 100% d'occupations des sites sur la nucléoprotéine.

Jeu de données	<u>Jeu 1</u> (n11ctrvp_10min)	<u>Jeu 2</u> (n11ctrvp_6min)	<u>Jeu 3</u> (n11ctrvp_11min)	<u>Jeu 4</u> (n11rvp_7min)
Nombre d'images	210	300	300	285
Oscillation	0,5	0,5	0,5	0,5
Groupe d'espace	P <sub>2</sub> <sub>1</sub> <sub>2</sub> <sub>1</sub> 2 (n° 18)	P <sub>2</sub> <sub>1</sub> <sub>2</sub> <sub>1</sub> 2	P <sub>2</sub> <sub>1</sub> <sub>2</sub> <sub>1</sub> 2	C <sub>2</sub> <sub>2</sub> <sub>2</sub> <sub>1</sub>
Dimensions de la maille élémentaire				
a (Å)	271,8802	271,8773	269,1033	235,5631
b (Å)	284,4086	283,5133	281,1390	281,1581
c (Å)	237,2169	238,2221	237,3862	260,3087
Longueur d'onde (Å)	0,87260	0,9537	0,9537	0,9537
Limites de résolution (Å)	94,916 – 6 (0,00011 – 0,002778)	136,083 – 6 (0,00005 – 0,02778)	68,199 – 5,7 (0,00021 – 0,03078)	130,189 – 5,5 (0,00006 – 0,03306)
Rsym (%)	0,118 (0,492)	0,210 (0,707)	0,099 (0,535)	0,074 (0,431)
I / $\sigma$ I	4,6 (1,6)	2,0 (1,0)	11 (3,3)	6,8 (1,8)
Complétude (%)	93,7 (94,5)	95,1 (95,1)	100 (100)	100 (100)
Multiplicité	4,6 (4,6)	5,0 (5,2)	5,8 (5,9)	5,5 (5,7)

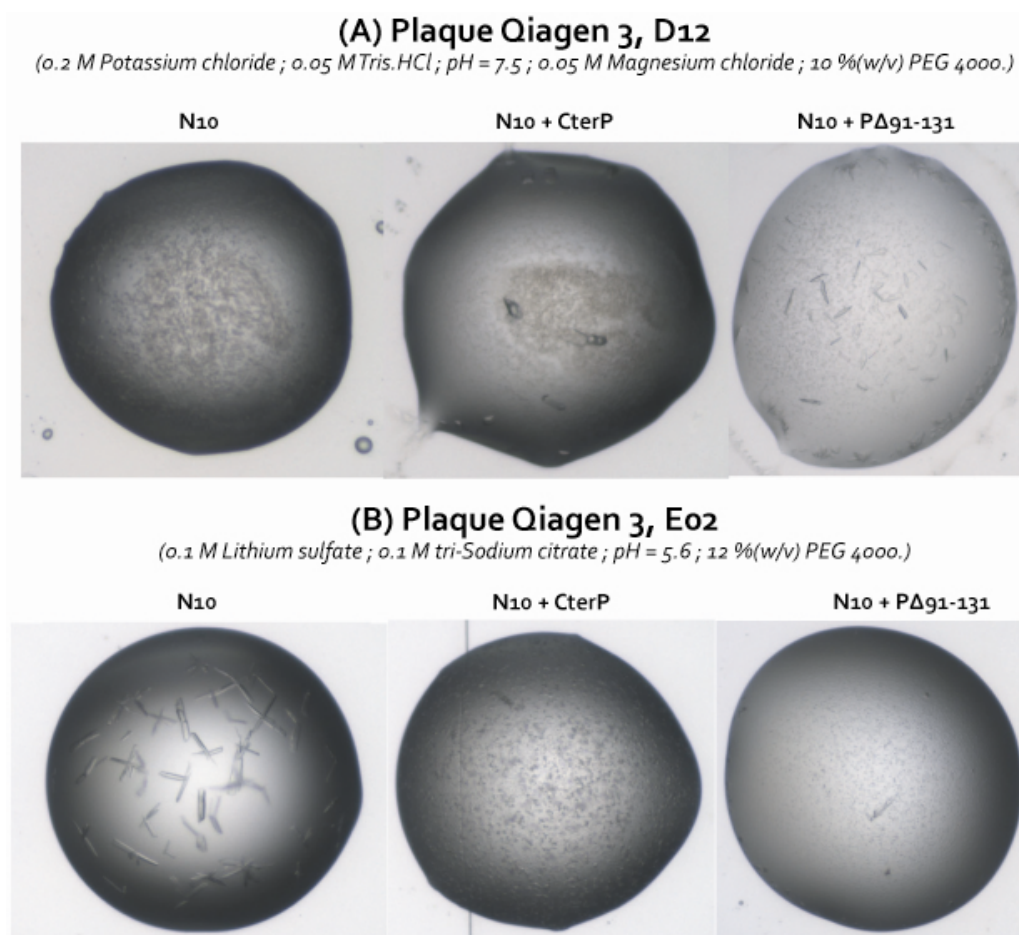
**Tableau 10** - Statistiques relatives aux collectes de données.

Nous avons répété ces trempages à plusieurs reprises, mais nous n'avons pu collecter des données que peu de fois. Ces données restent basse en résolution et leur traitement n'indique pas clairement si le domaine est fixé ou pas. Même si les cristaux de N11 obtenus étaient plus gros que précédemment, cela n'a pas amélioré le pouvoir de diffraction du cristal trempé dans le partenaire viral. Il est aussi possible aujourd'hui d'utiliser des marqueurs fluorescents pour la protéine ligand et ainsi, réalisé des trempages et vérifiés si le ligand a bien diffusé dans le cristal (Kettenberger and Cramer 2006). La résolution de la structure de ce complexe reste un challenge, mais de nombreux problèmes y font obstacles, en effet, il réside toujours ce problème de stœchiométrie du complexe, et comment identifier 2 domaines C-terminaux sur les anneaux si la fixation n'est pas similaire sur chaque anneau.

Nous essayons aujourd'hui encore d'obtenir des données de ce complexe à plus haute résolution. Il semble clair que ce travail de caractérisation structurale doit être réalisé avec le domaine C-terminal de phosphoprotéine du virus de la rage puisque c'est le domaine responsable de l'interaction avec le complexe N-ARN. Aussi, la flexibilité de la protéine P RV entière rend difficile son étude par cristallographie des rayons X.

## 2 – Co-cristallisation des anneaux N10 et N11 avec le domaine C-terminal de la phosphoprotéine et le mutant $P_{\Delta 91-131}$

Aussi, nous avons décidé d'utiliser le robot de cristallisation de la plateforme de haut débit de l'EMBL (<https://htxlab.embl.fr/>), pour passer au crible un maximum de 600 conditions de cristallogénèse. En comparant la cristallisation des anneaux seuls (N10 et N11) et des anneaux complexés avec le domaine C-terminal de P RV et le mutant  $P_{\Delta 91-131}$ , nous espérons ainsi pouvoir comparer la cristallisation des anneaux N-ARN et des anneaux en complexe et identifier dans quelles conditions des cristaux apparaissent de façon spécifique au complexe. Parmi toutes ces conditions testées, peu de conditions négatives pour les anneaux étaient positives pour les anneaux en complexe (figure 65, A). En effet, pour de nombreuses conditions, les cristaux présents pour les anneaux seuls ne le sont plus pour les anneaux en complexe (figure 65, B). J'ai tout de même essayé de reproduire quelques conditions intéressantes en gouttes suspendues de façon manuelle, mais malheureusement, aucune d'entre elle n'a permis d'obtenir de cristaux.

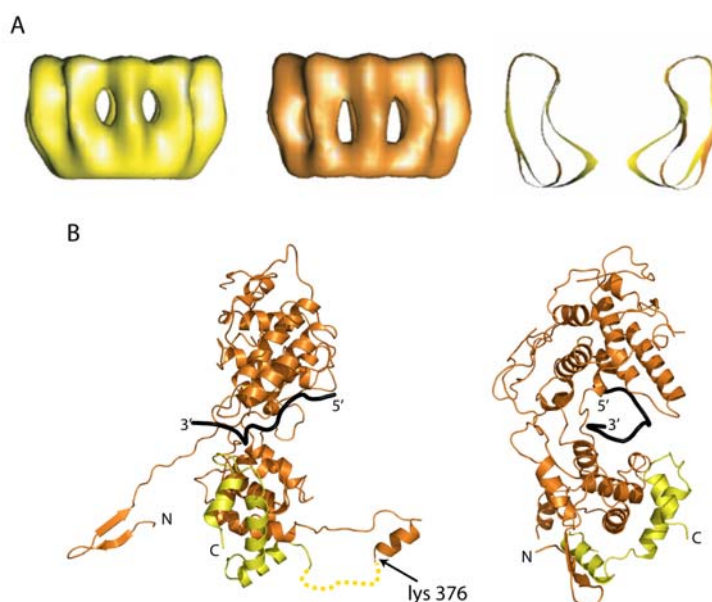


**Figure 65** - Exemple de nanogouttes obtenus avec des anneaux N10-ARN. (A) Plaque Qiagen 3, puit D12 (condition : 200 mM M KCl, 50 mM Tris-HCl pH 7,5, 50 mM MgCl<sub>2</sub>, 10% PEG 4000) : des cristaux sont présents pour les complexes N10-ARN-CterP et N10-ARN-P $\Delta 91-131$ . (B) Plaque Qiagen 3, puit E02 (condition : 100 mM LiSO<sub>4</sub>, 100 mM Na citrate, pH 5,6, 12% PEG 4000). La cristallisation est perturbée lorsque les N10 sont en complexe avec CterP RV et P $\Delta 91-131$ .

Le problème majeur que j'ai rencontré ici est la concentration des anneaux et des complexes qui restait faible. En effet, pour former le complexe une étape de purification sur chromatographie d'exclusion de taille (S200, GE Healthcare), ainsi qu'une étape de concentration supplémentaire sont nécessaires. Ceci pose problème car à chaque étape de concentration, on perd de l'échantillon. Aussi, tous les essais de reproduction manuelle des gouttes intéressantes n'ont pas aboutis. Il semble donc tout intéressant d'essayer à une concentration plus élevée du complexe.

## E – Analyse en microscopie électronique du complexe anneau de nucléoprotéine-ARN et la phosphoprotéine

En 2001, le Dr. Guy SCHOEHN a réalisé deux reconstructions tridimensionnelles (3D) des anneaux à 10 sous-unités de nucléoprotéines du virus de la rage à partir d'images de microscopie électronique (Schoehn & *al.*, 2001). La première reconstruction a été réalisée sur des anneaux non protéolysés (figure 66, jaune), et la seconde sur des anneaux protéolysés (figure 66, orange) par action de la trypsine. La trypsine réalise une coupure au niveau de la lysine 376 (Kouznetzoff & *al.*, 1998), et lors d'une protéolyse prolongée, le domaine C-terminal de la nucléoprotéine est totalement éliminé. Grâce à ces reconstructions avec et sans domaine C-terminal de la nucléoprotéine le site de fixation de la phosphoprotéine sur l'anneau a été identifié (figure 66).

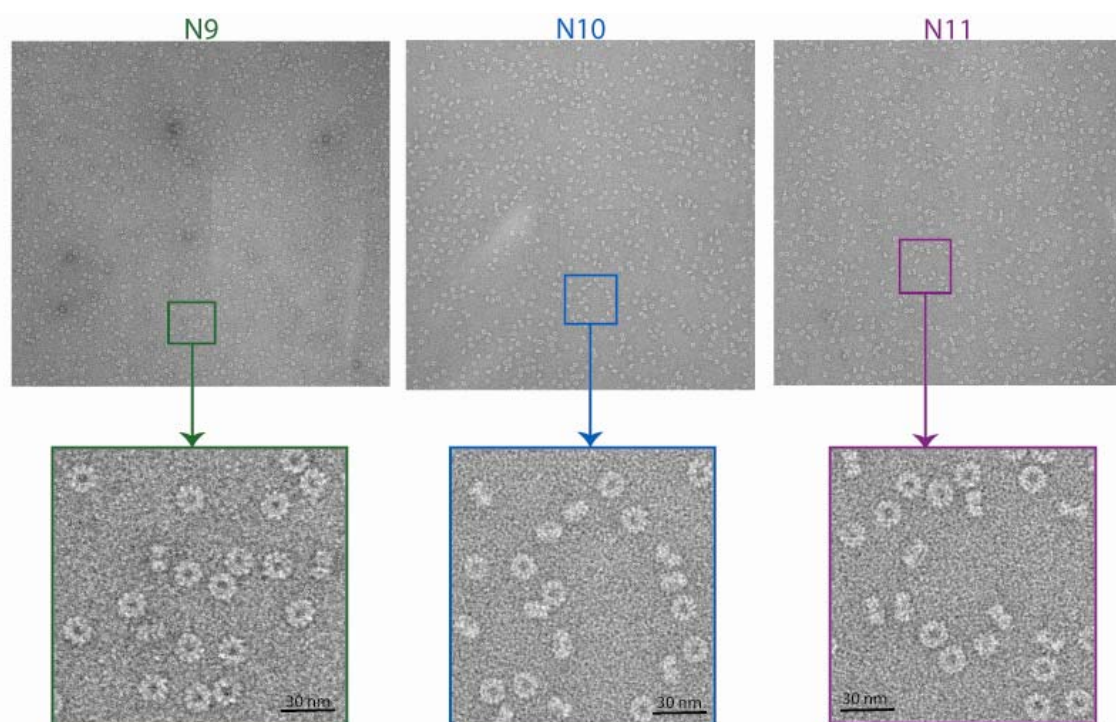


**Figure 66 – (A) Reconstruction par microscopie électronique des anneaux à 10 sous-unités de nucléoprotéines, N10 entiers (jaune) et protéolysés (orange).** Identification de la région C-terminal de la nucléoprotéine responsable de l'interaction avec la phosphoprotéine. **(B) Visualisation en mode « cartoon » du monomère de nucléoprotéine, d'après la structure cristallographique (Albertini & *al.*, 2006).** La zone représentée en orange correspond à la forme protéolysée de la nucléoprotéine. On peut ainsi observer la région C-terminal de la nucléoprotéine responsable de l'interaction avec la nucléoprotéine. D'après (Albertini & *al.*, 2008).

Aussi, suite à la résolution de la structure des anneaux à 11 protomères de nucléoprotéines, nous avons entrepris un travail de caractérisation structural des complexe N-ARN-P et N-ARN-CterP afin d'identifier plus clairement le site de fixation de la phosphoprotéine sur les anneaux. Ceci est un travail en étroite collaboration avec le Dr. Irina GUTSCHE, qui a observé de nombreux échantillons de complexes N-ARN-P et N-ARN-CterP.

### 1 – Microscopie électronique sur des N9, N10 et N11 et reconstruction d'images

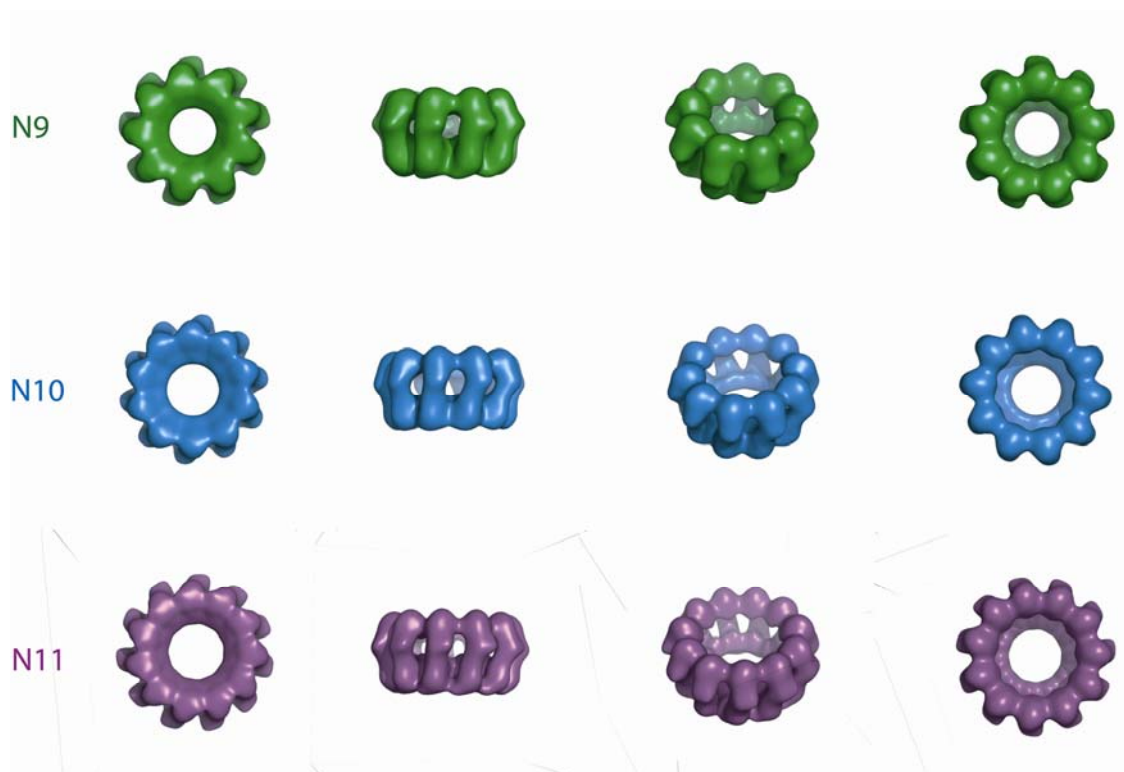
Dans un premier temps, les échantillons d'anneaux purifiés ont été observés en microscopie électronique en coloration négative en acétate d'uranyle (figure 67). Les images sélectionnées ont permis de réaliser des reconstructions tridimensionnelles des différents anneaux par rétroprojection. Nous avons obtenus des modèles 3D des anneaux pour les N9, N10 et N11 à une résolution entre 16 et 18 Å (figure 68). Ainsi, nous pourrons comparer par la suite les reconstructions obtenues pour le complexe N-ARN-P ou N-ARN-CterP et identifier la densité correspondant à P/CterP par soustraction.



**Figure 67** - Images de microscopie électronique en coloration négatives (acétate d'uranyle) des N9, N10 et N11 ainsi que un zoom. Il est possible de distinguer les sous-unités des anneaux, et de réaliser de façon simple une classification de ces anneaux. Diamètre des anneaux N11 : 180 Å.

La conformation du monomère de nucléoprotéine est similaire dans les différentes reconstructions obtenues (figure 68). En effet, il a été possible d'ajuster le monomère de nucléoprotéine, sans les domaines échangeables par symétrie dans ces reconstructions. Ceci indique que la monomère

de nucléoprotéine possède la même conformation au niveau de ces domaines NTD et CTD dans les anneaux N9, N10 et N11. Aussi, la boucle flexible non visualisable dans la structure cristallographique des anneaux N11, n'est pas visible dans ces reconstructions.



**Figure 68** - Reconstructions 3D des anneaux de différentes tailles : N9, N10 et N11. Différentes vues sont présentées ici (représentation de surface avec le programme Pymol). La première vue du dessus nous montre le domaine C-terminal de la nucléoprotéine, responsable d'une grande surface d'interactions entre les monomères, alors que la dernière vue du dessous, nous montre la partie N-terminale de la nucléoprotéine. Les reconstructions obtenues sont à une résolution de 18 Å.

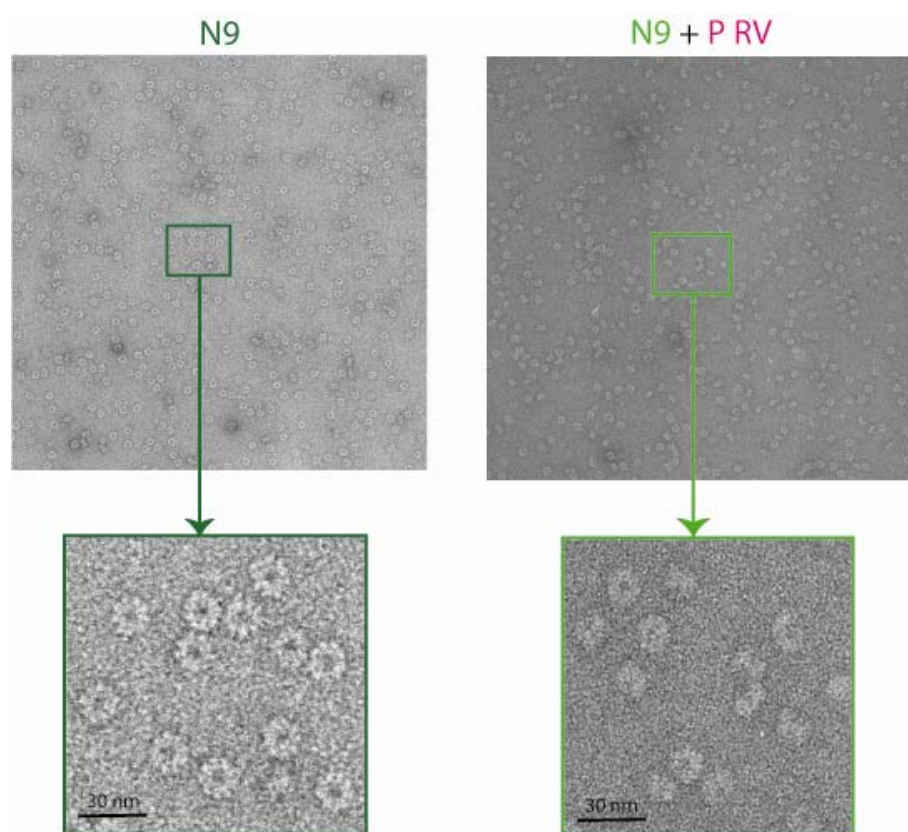
## 2 – Microscopie électronique sur le complexes N9-P RV et reconstruction d'images

Les premiers échantillons de complexes N-ARN-P et N-ARN-CterP ont été produits en mélangeant de façon stœchiométrique des anneaux et de la phosphoprotéine et du domaine C-terminal de la protéine P. Les images obtenues de ces complexes n'étaient pas satisfaisantes. En effet, le bruit de fond dû à la présence de la protéine P seule (ou du domaine C-terminal de la protéine P) rend l'analyse d'image difficile. Par la suite, j'ai préparé des échantillons de complexes N-ARN-P ou N-ARN-CterP purifiés sur chromatographie d'exclusion de taille sur une colonne S200 (HR 10/30, GE Healthcare). Ainsi, le complexe en conditions saturantes de P ou CterP était injecté, et seule la fraction du pic correspondant au complexe était observée en ME, après vérification par SDS-PAGE et gel natif. La [figure 69](#) nous présente les images brutes en ME en coloration négative des anneaux à 9 sous-unités seuls et en complexe avec la phosphoprotéine entière. Si l'on observe ces clichés attentivement ([figure 69, zooms](#)), il est impossible de différencier les anneaux seuls des anneaux complexés. Ces expériences ont été réalisées pour tous les



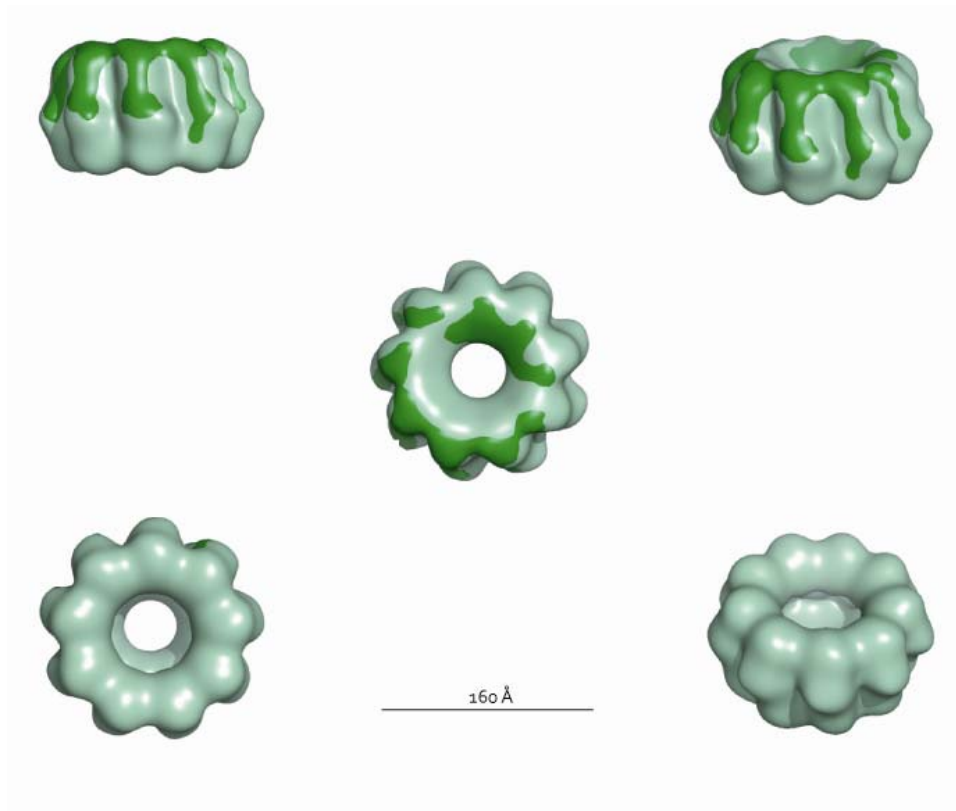
types d'anneaux, des N9 aux N11. Pour aucun d'entre eux, il n'a été possible d'identifier la phosphoprotéine fixée par une visualisation directe des images. Pour les images des anneaux N9, on identifie clairement sur le cliché chaque sous-unité de nucléoprotéine de l'anneau. Pour les anneaux complexés avec la protéine P, il semble plus difficile de les distinguer (figure 69).

Les échantillons de N9 sont les plus homogènes, en effet, ce sont les premiers types d'anneaux à être élués lors de l'électrophorèse préparative. Nous avons obtenu une reconstruction du complexe purifié à partir d'images de coloration négative (figure 70). L'utilisation du même type de colorant pour l'observation des N9 seuls et en complexe avec la phosphoprotéine, ainsi que le traitement similaire des images nous a permis de comparer les reconstructions tridimensionnelles obtenues (figure 70).



**Figure 69** - Images de ME en coloration négative des N9 seuls et saturés avec la protéine P entière, et zooms.

Lorsque nous avons superposé les reconstructions 3D des N9 seuls et en complexe, il n'a pas été possible d'identifier une extra densité correspondant la protéine P, comme si la protéine, dans ces conditions d'analyses en microscopie électronique, n'était pas fixée sur l'anneau (figure 70).



**Figure 70-** Comparaison des reconstructions tridimensionnelles des anneaux N9 (vert foncé) et des N9P (vert clair).

### 3 – Protéines fusions à la Maltose Binding Protein (MBP)

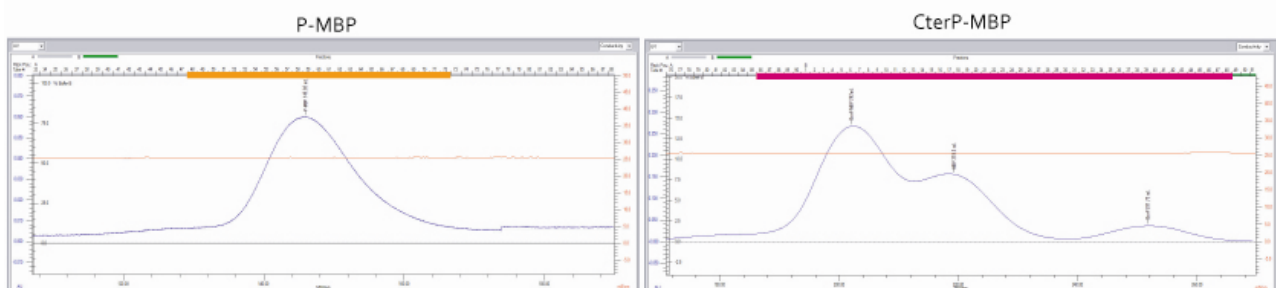
Aucun résultat concluant sur la localisation de la protéine P ou du CterP sur le complexe N-ARN n'étant obtenu avec les complexes « natifs », nous avons décidé de produire la protéine P du RV ainsi que son domaine C-terminal en fusion avec la MBP (*Maltose Binding Protein*). En effet, la protéine MBP étant globulaire et de masse moléculaire d'environ 43 kDa, la protéine fusion fixée sur les anneaux de nucléoprotéine-ARN pourrait être plus facilement localisable en microscopie électronique. L'analyse en biochimie montre clairement que le complexe N-ARN-P ou N-ARN-CterP se forme et est stable. Nous avons défini la stœchiométrie de ces complexes grâce au SEC-MALLS-RI. La fixation de deux dimères de protéine P sur les anneaux (450 à 700 kDa), ajoutent une masse d'environ 140 kDa, la fixation de 2 domaines C-terminaux ajoutent une masse d'environ 30 kDa sur les anneaux (450 à 700 kDa). Il est possible de ne pas visualiser directement sur les images brutes, 2 domaines de 14 kDa sur les anneaux. Aussi, la flexibilité de la phosphoprotéine entière pourrait être à l'origine de ce problème. Il est possible que les parties flexibles de la protéine qui ne fixe pas les anneaux soient mobiles et que nous ne les observions pas par moyenne sur les images et reconstructions de microscopie électronique. Aussi, le

complexe N-ARN-P/CterP peut être déstabilisé lors de la préparation de l'échantillon pour la coloration négative.

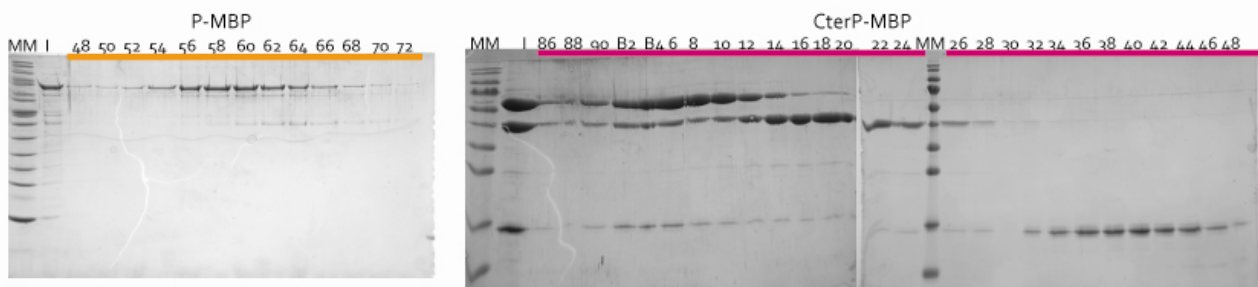
Nous avons donc produits et purifiés les protéines fusions P-MBP et CterP-MBP afin de réaliser à nouveau le complexe anneaux-phosphoprotéine en fusion MBP. Les deux premières étapes de chromatographie d'affinité nous ont permis d'isoler les protéines P-MBP et CterP-MBP des autres contaminants.

Afin d'éliminer d'éventuels agrégats des protéines fusions, nous avons utilisé une chromatographie d'exclusion de taille (S200, HR 10/30, GE Healthcare). La protéine fusion P-MBP sur la colonne S200 est éluée en un seul pic (figure 71, A), symétrique et l'analyse par gel SDS-PAGE montre que le début du pic est homogène en présence de P-MBP mais que la fin du pic renferme des formes protéolysées de P-MBP (figure 71, B) (à partir de la fraction 60). En ce qui concerne l'échantillon de CterP-MBP (figure 71, A), on voit clairement 3 pics, que l'on peut facilement attribuer grâce au gel SDS-PAGE (figure 71, B) : le premier correspond à un mélange de CterP-MBP et MBP (volume d'éluion de 202 mL), le second correspond uniquement à la MBP (volume d'éluion de 219 mL) et enfin le dernier, au domaine C-terminal de P seul (volume d'éluion de 257 mL).

(A) Chromatographie d'exclusion de taille (S200 Hiloal 26/60, GE Healthcare) des protéines fusionnées à la MBP



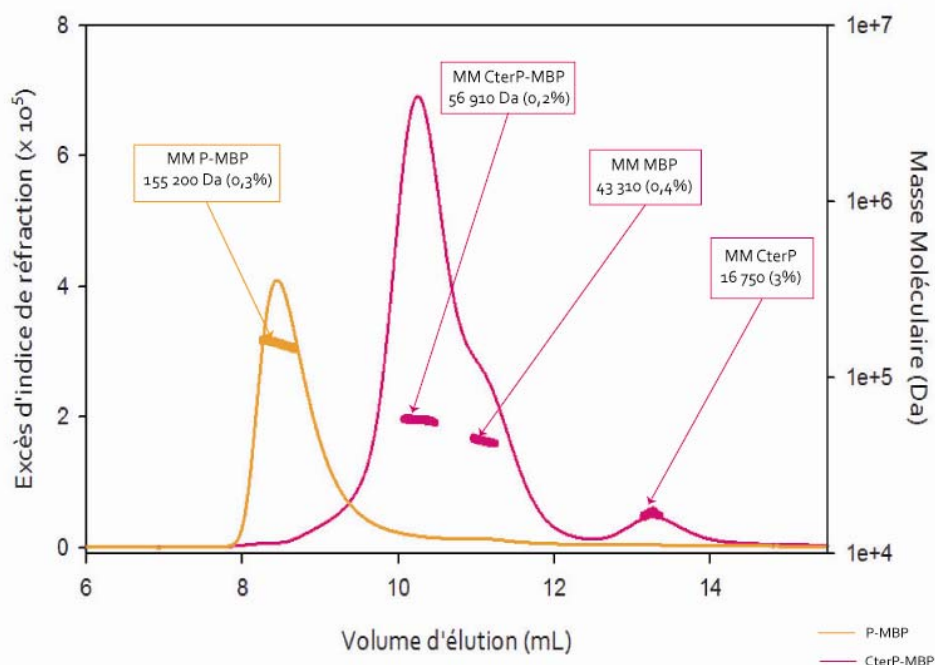
(B) Analyse par gel SDS PAGE 15% des chromatogrammes



**Figure 71** - Profils d'éluion de P-MBP et CterP-MBP sur colonne d'exclusion de taille de type S200. (A) Chromatogrammes obtenus pour P-MBP et CterP-MBP. (B) Analyse en gel polyacrylamide en conditions dénaturantes 15% des fractions d'éluions indiquées (en orange pour P-MBP et rose pour CterP-MBP).

Afin de vérifier la qualité des échantillons de phosphoprotéines fusions produites, mais surtout vérifier que l'ajout de la MBP en N-terminal de la phosphoprotéine et de son domaine C-terminal ne modifie pas l'état d'oligomérisation, nous avons injecté les échantillons de P-MBP et CterP-MBP sur SEC-MALLS-RI. La protéine fusion P-MBP (figure 72, courbe orange) s'élué en un seul pic sur la chromatographie d'exclusion de taille (S200, HR 10/30, GE Healthcare). L'analyse de la masse par SEC-MALLS-RI nous donne une masse moléculaire moyenne de 155 kDa, ce qui montre que la protéine P-MBP est un dimère en solution (tableau 11). L'ajout de la protéine MBP en N-terminal ne perturbe donc pas l'état d'oligomérisation de la phosphoprotéine. La protéine fusion CterP-MBP (figure 72, courbe fuschia) s'élué quant à elle en deux pics : le premier pic (volume d'élué de 9 à 12 mL) possède un épaulement, montrant l'existence de deux espèces et le second pic (volume d'élué 13 à 14 mL) correspond au domaine C-terminal de la phosphoprotéine. La mesure des masses moléculaires moyennes de ces différentes espèces en solution (tableau 11) nous montrent clairement que l'on trouve la protéine fusion CterP-MBP (56,9 kDa), la protéine MBP (43,3 kDa) et le domaine C-terminal de la phosphoprotéine (16,8 kDa). Aussi, l'état d'oligomérisation de la protéine fusion CterP-MBP est le même que pour le domaine C-terminal de la phosphoprotéine seul, à savoir monomérique.

Ces échantillons de protéine fusion, où l'on a ajouté en N-terminal de la protéine (phosphoprotéine ou domaine C-terminal) la protéine MBP, possède le même état d'oligomérisation en solution que les protéines sauvage : P-MBP est un dimère et CterP-MBP un monomère.



**Figure 72** - Chromatographie d'exclusion de taille couplée à la détection par diffusion statique de lumière et à la réfractométrie. Profil d'élution des protéines fusion P-MBP (orange) et CterP-MBP (rose) sur colonne S200 (HR 10/30 GE Healthcare). Les crochets indiquent la masse moléculaire déterminée à partir des mesures de diffusion de lumière et de réfractométrie à différents temps d'élution en utilisant  $dn/dc = 0,185$ . Conditions : 20 mM Tris-HCl pH 7,5 et 150 mM NaCl.

	<b>P-MBP</b>	<b>CterP-MBP</b>	<b>MBP</b>	<b>CterP</b>
<b>Masse théorique (Da)</b>	77 361	57 136	40 340	13 989
<b>Masse mesurée (Da)</b>	155 000 ± 500	56 900 ± 120	43 300 ± 200	16 800 ± 500
<b>Etat d'oligomérisation en solution</b>	dimère	monomère	monomère	monomère

**Tableau 11** - Analyse des masses théoriques et mesurées des différentes protéines fusions et détermination de l'état d'oligomérisation en solution.

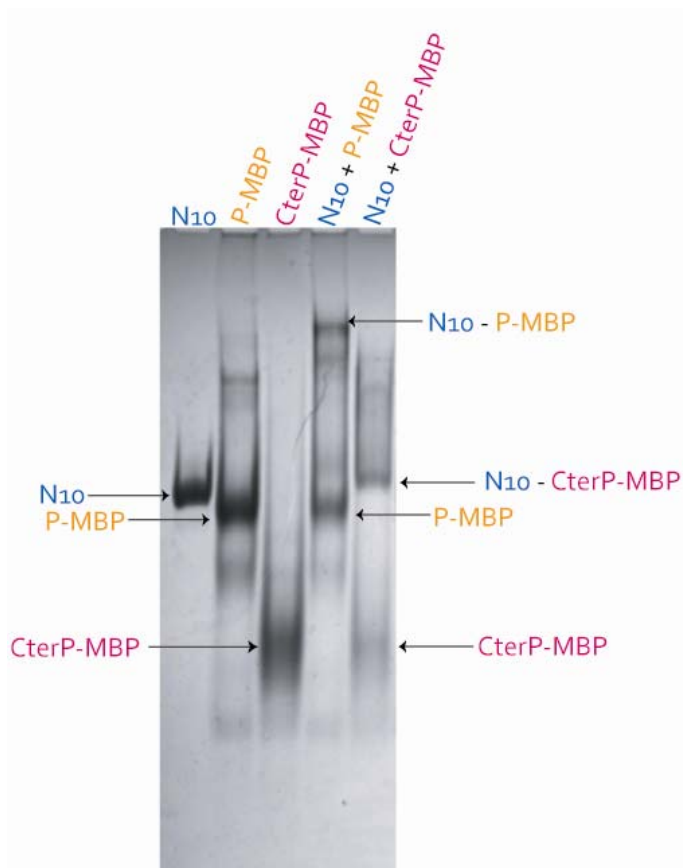
D'après les résultats d'homogénéité d'échantillons obtenus, il semblait clair d'utiliser la protéine fusion P-MBP pour former le complexe avec les anneaux et les analyser en microscopie électronique. En effet, la protéine fusion CterP-MBP ne semble pas être stable au cours de la purification (cette protéolyse s'étant produite pour plusieurs purifications).

#### *4 – Electrophorèse en conditions natives des complexes N10-P-MBP et N10-CterP-MBP*

Afin de vérifier que les anneaux fixaient bien les phosphoprotéines fusions, les complexes N10-P-MNBP et N10-CterP-MBP, où les anneaux N10 ont été incubées avec un excès de phosphoprotéines fusions ont été analysé sur électrophorèse en conditions natives. Nous avons pu observer un retard sur gel des anneaux N10 en complexe avec la phosphoprotéine fusion MBP et avec le domaine C-terminal de PRv en fusion avec la MBP (figure 73). Le retard sur gel pour le complexe N10-P-MBP est important, et la migration de la P-MBP seule dans ces conditions se situe dans la même zone que la migration des N10. Le retard sur gel du complexe N10-CterP-MBP est faible, mais on observe la présence d'excès de CterP-MBP.

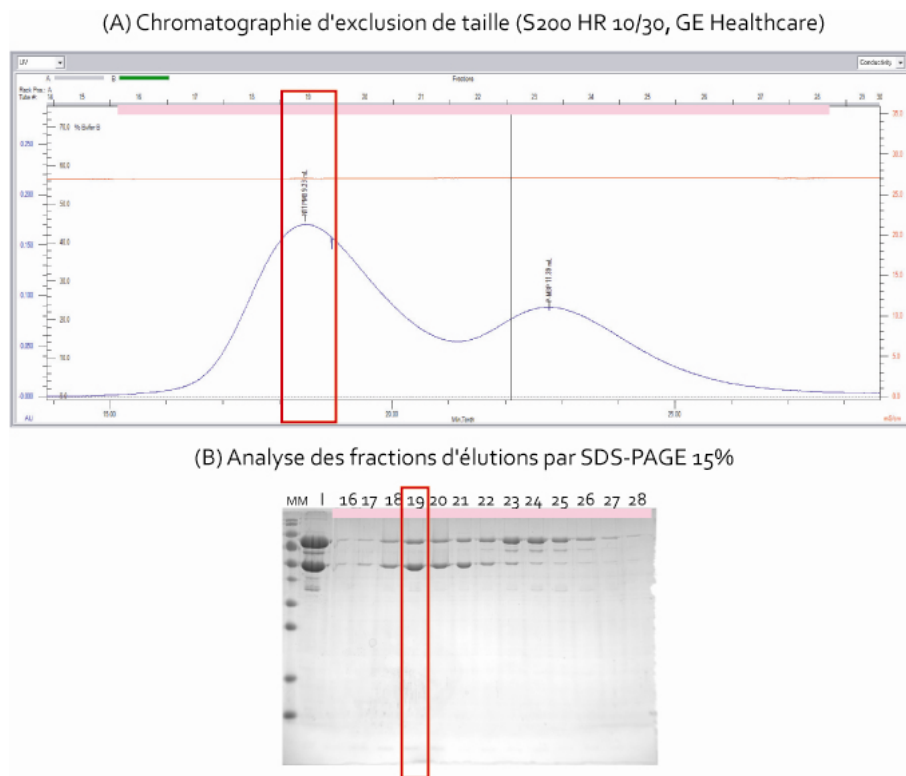
Cette étude biochimique en conditions natives nous a permis de vérifier que l'ajout de la protéine MBP ne perturbe pas l'interaction entre la nucléoprotéine (N-ARN) et le domaine C-terminal de la phosphoprotéine. La masse moléculaire de ces complexes mesurées par SEC-MALLS-RI nous a permis de mettre en évidence que la stoechiométrie des complexes anneaux-P-MBP et anneaux-CterP-MBP est la même que pour les complexes anneaux-P RV et anneaux-CterP RV (données non montrées).

**Figure 73** - Analyse sur électrophorèse en condition native des protéines fusions P-MBP et CterP-MBP seules et en complexe avec les anneaux (N10).



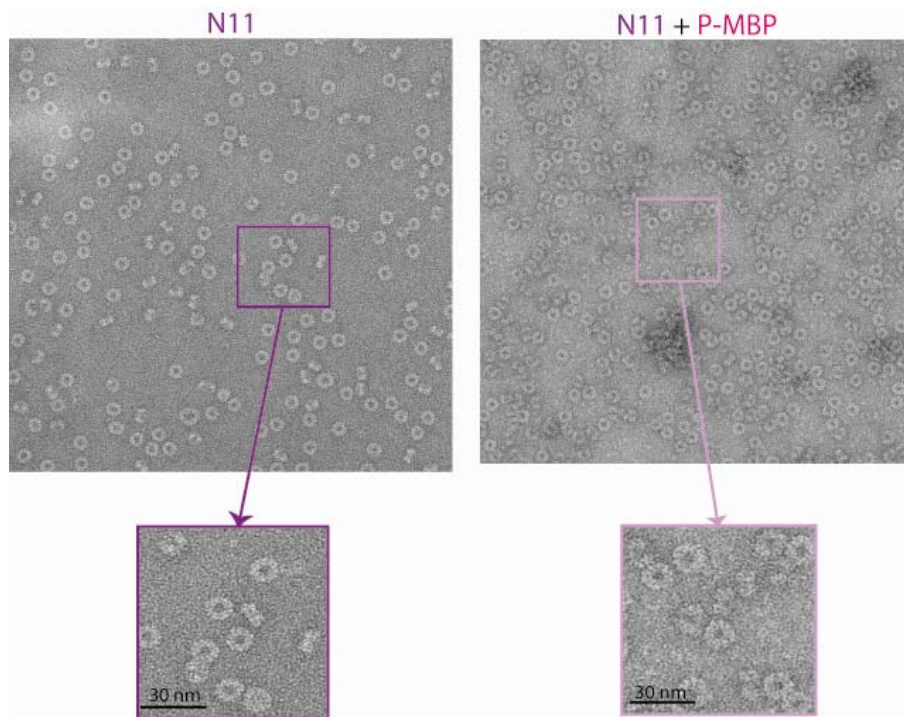
### 5 – Microscopie électronique du complexe N11-P-MBP et reconstruction d’images

Les analyses structurales en microscopie électronique ont été réalisées sur le complexe N11-P-MBP. Le complexe N11-P-MBP a été purifié sur chromatographie d’exclusion de taille avec une colonne S200 (HR 10/30, GE Healthcare). Des anneaux N11 à une concentration de 4 mg/ml ont été complexés avec la phosphoprotéine fusion P-MBP, en conditions saturantes de P-MBP (figure 74, Input). Le mélange a été injecté sur la colonne d’exclusion de taille (figure 74, A) et les fractions correspondant aux pics présents sur le chromatogramme ont été analysé par gel de polyacrylamide en conditions dénaturantes (figure 74, B). Comme on peut l’observer sur le chromatogramme, le mélange est élué en deux pics (figure 74, A) : le premier correspondant au complexe saturé N11-P-MBP, et le second correspondant à l’excès de phosphoprotéine fusion. Ceci a été confirmé par l’analyse sur gel SDS-PAGE. Nous avons par la suite, observé l’échantillon correspondant à la fraction 19 (encadré rouge) en microscopie électronique. Cet échantillon étant le plus concentré ( $DO_{280nm}$  de 1,7 sur le chromatogramme en figure 74, A) et le plus homogène (figure 74, B, fraction 19).

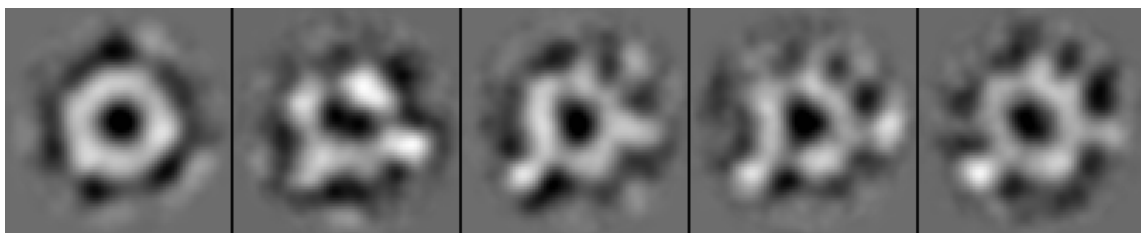


**Figure 74 - Profils d'élution du complexe N11-P-MBP en conditions saturantes de P-MBP sur colonne d'exclusion de taille de type S200.** (A) Chromatogrammes obtenus pour le complexe N11-P-MBP. On peut observer le second pic qui correspond à la protéine P-MBP (B) Analyse en gel polyacrylamide en conditions dénaturantes 15% des fractions d'élutions indiquées en rose (16 à 28). MM : marqueur de poids moléculaire. I : input (mélange chargé sur la colonne).

Nous avons observé tout d'abord l'échantillon du complexe N11-P-MBP par microscopie électronique en coloration négative. Là encore, les images de coloration négative ne nous ont pas permis d'identifier la protéine fusion (figure 75). Comparativement aux anneaux seuls, sur les images de microscopie électronique en coloration négative, le complexe N11-P-MBP ne semble pas être différent. Toutefois, l'analyse du complexe N11-P-MBP en cryomicroscopie électronique s'est révélé plus prometteuse (figure 75). Le contraste de l'image étant faible, nous avons procédé aux traitements de images afin d'obtenir une première classification des particules sélectionnées (figure 76). Sur ces images, il apparaît clairement une sensité à l'extérieur de l'anneau. Sur les 3 dernières images de classification, on visualise très facilement 4 densités à l'extérieure de l'anneau, avec une répartition inhomogène sur l'anneau. Cette première classification nous donne bon espoir quand à obtenir une reconstruction tridimensionnelle du complexe N11-P-MBP. La densité que nous observons peut être attribuée à la protéine fusion MBP. En effet, les complexes N11-P-MBP ont la même stoechiométrie que les complexes N11-P RV et d'après les données de la littérature, le domaine C-terminal de la phosphoprotéine va interagir avec le domaine C-terminal de la nucléoprotéine, sur l'anneau. L'observation de quatre extra-densités à l'extérieur de l'anneau confirme par ailleurs que la protéine P-MBP est un dimère, chaque densité pouvant correspondre à la MBP. Cependant, nous n'avons encore pas de données avec une résolution plus haute pour identifier clairement comment la phosphoprotéine se fixe sur les anneaux.



**Figure 75** - (A) images de microscopie électronique en coloration négative (acétate d'uranyl) de N11, et de N11 saturé en phosphoprotéine fusion P-MBP.



**Figure 76** - Première classification des images de cryomicroscopie électronique des complexes N11-P-MBP. Il est possible d'identifier une densité à l'extérieur de l'anneau, globulaire.

## 6- Conclusions

Ces résultats sont prometteurs concernant l'obtention d'une reconstruction par microscopie électronique du complexe. Nous sommes actuellement en train d'optimiser la qualité des échantillons tout comme la qualité de la congélation pour l'observation en cryomicroscopie électronique. En effet, ici il n'est plus possible de distinguer les sous unités de nucléoprotéine sur les images de la première classification (figure 76), ce qui va affecter la qualité de la carte de densité de microscopie électronique.



## F – Détermination de l'affinité du complexe N-ARN-P

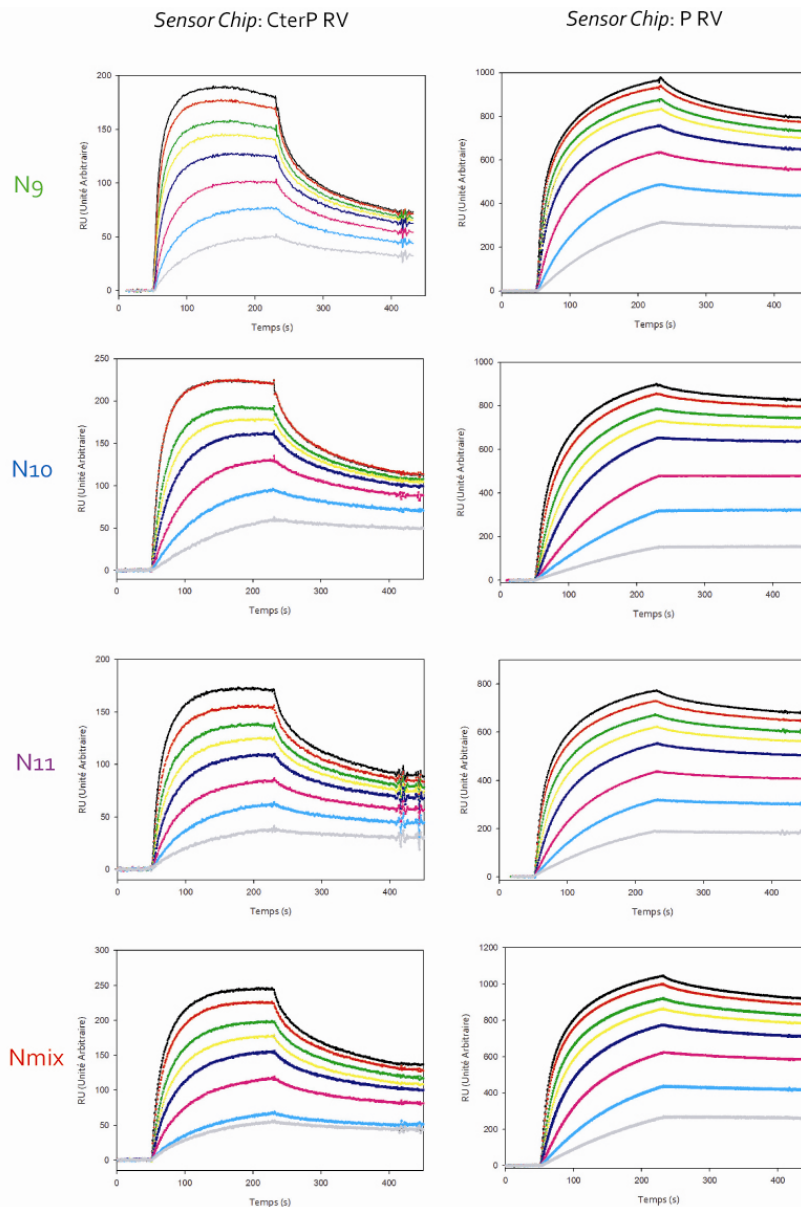
Le PSB (*Partnership for Structural Biology*) qui se situe dans le bâtiment dans lequel nous travaillons, permet d'avoir accès à un grand nombre de techniques différentes afin de caractériser les molécules d'intérêts. Dans un premier temps, j'ai essayé d'utiliser la calorimétrie isotherme à titration (ITC : *Isothermal Titration Calorimetry*). L'ITC permet de mettre en évidence une interaction et de déterminer la constante d'affinité  $K_a$  ( $10^3$  à  $10^8$   $M^{-1}$ ) et la stœchiométrie du système. Cette technique est largement utilisée pour décrire des interactions protéines/protéines, mais aussi protéines/peptides. Cependant, malgré mes efforts, je n'ai pu enregistrer aucune variation de signal lorsque le complexe N-ARN avec la protéine P ou son domaine C-terminal se formait. Cette technique demande de grandes quantités d'échantillons, facteur limitant lorsque l'on travaille avec les anneaux N-ARN. Aussi, des essais de dialyses à l'équilibre, où l'on peut quantifier les échantillons après échange entre les deux compartiments, ne se sont pas révélés concluants. Nous avons donc essayé la technique de résonance plasmonique de surface (SPR) ou BIAcore pour caractériser l'affinité de ce complexe.

### *1 – Expériences de résonance plasmonique de surface sur le complexe anneaux-phosphoprotéine*

Nous avons réalisés plusieurs cinétiques d'associations en injectant différents types d'anneaux sur les deux *sensor chips* présentant la phosphoprotéine et le domaine C-terminal. Nous avons injecté une solution d'anneaux nommée Nmix, qui correspond à des anneaux de différentes tailles (de 9 à 12 sous-unités de nucléoprotéine). Nous avons aussi injecté des anneaux purifiés : N9, N10 et N11 à différentes concentrations. La *sensor chip* a été régénérée après chaque injection en injectant une solution de NaCl à 1 M. Les résultats sont présentés par des graphes qui montrent l'association et la dissociation du complexe (figure 77). L'axe des ordonnées à pour unité le RU (événements de fixation de biomolécules à la surface qui provoque des changements dans un signal de résonance plasmonique de surface, qui sont exprimés en unités de résonance (une RU est équivalente à un picogramme par  $mm^2$  de surface sensible)) en fonction du temps.

Nous avons injecté les échantillons de nucléoprotéine à différentes concentrations : 400, 300, 200, 150, 100, 50, 25 et 10 nM et nous avons obtenu du signal significatif de l'interaction (de l'ordre de la centaine de RU). Ceci nous confirme bien que l'interaction a lieu, dans ces conditions. L'expérimentation a été simple, dans le sens où pour chacune des conditions, nous avons obtenus un signal significatif et

nous avons pu régénérer la puce. Ce fut l'analyse des données qui nous a posé problème quand à la détermination des paramètres  $k_a$ ,  $k_d$  et  $k_D$  de ces courbes.



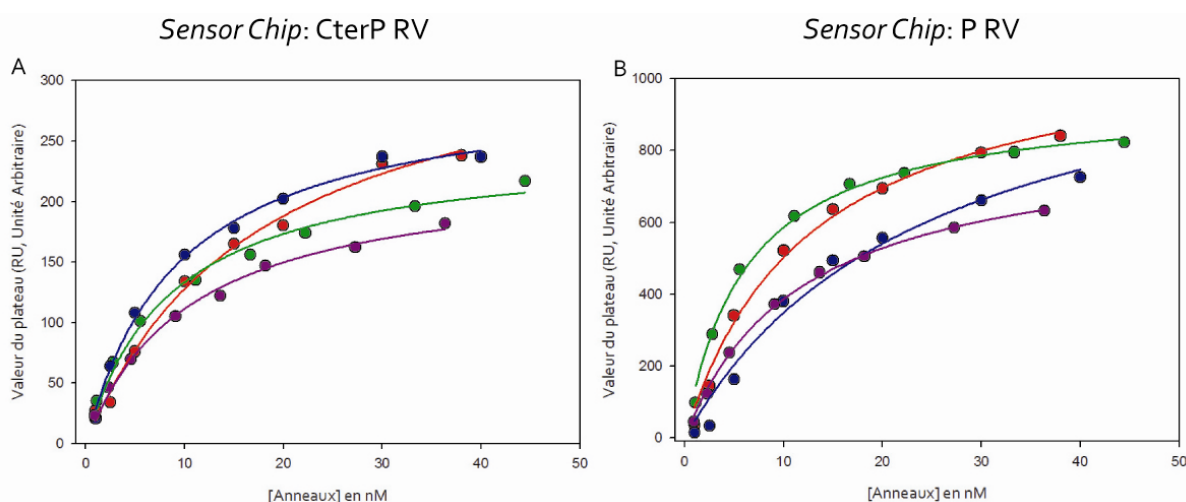
**Figure 77** – Cinétiques des expériences de résonance plasmonique de surface sur un appareil BIAcore X (GE Healthcare). Chaque courbe correspond à une mesure au cours du temps pour une injection de 60  $\mu$ l d'anneaux (N9, N10, N11 ou Nmix) à une concentration donnée (10, 25, 50, 100, 150, 200, 300 et 400 nM).

## 2 – Analyse des données

Les courbes cinétiques étant obtenues, nous avons utilisé dans un premier temps le logiciel BIAevaluation afin d'ajuster les courbes dans leur globalité, en utilisant différents modèles. Cependant, aucun des modèles disponibles dans ce logiciel ne nous a permis d'ajuster les courbes dans leur globalité et d'obtenir des paramètres pour l'association et la dissociation. Nous avons donc utilisé l'analyse à l'équilibre des données. En effet, pour un certain nombre de ces cinétiques, l'association du complexe

arrive à une valeur RU plateau, qui peut être utilisée. Nous avons donc représenté sur un graphique les valeurs au plateau en fonction de la concentration de ligand injecté (figure 78), pour chacun des complexes formés.

Comme on peut le constater, l'ajustement des valeurs expérimentales du plateau atteint lors de l'association du complexe est de bonne qualité. Les paramètres extraits de ces ajustements sont récapitulés dans le tableau 12. On remarque que la valeur de la constante d'association reste sensiblement la même quelle que soit les conditions de complexes qu'on utilise. En effet, celle-ci est aux alentours de  $10 \text{ nM}^{-1}$ , que ce soit pour l'interaction anneaux-phosphoprotéine ou anneaux-domaine C-terminal de la phosphoprotéine.



**Figure 78 – Ajustement à l'équilibre des cinétiques de résonance plasmonique de surface.** Les points représentent les valeurs expérimentales obtenues et les courbes correspondent à l'ajustement de ces courbes suivant le modèle  $f = \frac{RU \max \times X}{k_D + X}$ . En vert sont présentées les données pour les N9, en bleu pour les N10, en violet pour les N11 et en rouge pour les Nmix.

	N9 vs CterP	N10 vs CterP	N11 vs CterP	Nmix vs CterP
<b>RU max</b>	247 ± 9	299 ± 10	228 ± 9	356 ± 22
<b>k<sub>D</sub> (nM)</b>	8,5 ± 1	9,4 ± 0,9	10,5 ± 1,1	17,8 ± 2,5
	N9 vs P	N10 vs P	N11 vs P	Nmix vs P
<b>RU max</b>	949 ± 25	1211 ± 175	842 ± 20	1134 ± 62
<b>k<sub>D</sub> (nM)</b>	6,2 ± 0,6	24,8 ± 7	11,8 ± 0,7	12,6 ± 1,8

**Tableau 12 - Paramètres RUmax et ka obtenus à partir de l'ajustement à l'équilibre des courbes de cinétiques de résonance plasmonique de surface (BIAcore).**

### 3 - Conclusion

La technique de résonance plasmonique de surface (BIAcore) nous a permis d'obtenir un signal significatif de l'interaction entre les anneaux et la phosphoprotéine ou son domaine C-terminal. Nous n'avons pu ajuster les courbes dans leur globalité, et l'ajustement à l'équilibre de ces cinétiques nous indique que la constante de dissociation est du même ordre de grandeur, c'est-à-dire  $k_D \approx 10$  nM pour l'interaction anneaux-phosphoprotéine ou anneaux-domaine C-terminal de P RV. Ceci nous confirme encore que seul le domaine C-terminal de la phosphoprotéine est responsable de l'interaction avec le complexe nucléoprotéine-ARN.

## G – Conclusions

Les résultats que nous avons obtenus montrent clairement que la phosphoprotéine se lie aux anneaux nucléoprotéine-ARN via son domaine C-terminal, et uniquement ce domaine. Les complexes que nous avons pu observer en SEC-MALLS-RI et gel natif, montre clairement que la phosphoprotéine est fixée. Toutefois, les études structurales de ces complexes par microscopie électronique ou encore SAXS ne nous permettent pas d'avoir des modèles 3D qui indiqueraient la localisation de la phosphoprotéine sur les anneaux. Il semble donc que l'ajout de la phosphoprotéine sur les anneaux ne modifie pas la conformation de la nucléoprotéine. Ceci remet en question notre vision du complexe de réplication des Rhabdoviridae, puisque nous pensions que la phosphoprotéine allait jouer le rôle d'interrupteur et allait modifier la conformation de la nucléoprotéine pour permettre à la polymérase virale d'accéder au génome. Il semble clair que le modèle des anneaux ne permettent pas de conclure sur ce phénomène. Il semble important de caractériser aussi les états de phosphorylations de la phosphoprotéine.

Des expériences complémentaires de caractérisation du complexe par SANS, avec la phosphoprotéine du virus de la rage deutérée, dans différent pourcentage de D<sub>2</sub>O vont permettre de mieux caractériser les différents composants du complexe. En effet, à 0%, 40%, 70% ou 100% de D<sub>2</sub>O, la diffusion des neutrons aux petits angles va nous permettre d'obtenir de l'information structurale pour chacun des composants du complexe : protéine native, protéine deutérée mais aussi l'ARN, et nous pourrons comparer quels composants du complexe subissent des éventuels changements de conformation.

Il semble aussi que ce modèle, facile d'étude en biochimie et biophysique ne soit pas complètement adéquate pour caractériser la dynamique du complexe de réplication du virus de la rage. Aussi, l'état de phosphorylation de la phosphoprotéine et de la nucléoprotéine pourrait jouer le recrutement de partenaires cellulaires et viraux. Il semble le modèle des anneaux est pratique pour comprendre au niveau moléculaire les interactions nucléoprotéine-phosphoprotéine mais il semble plus pertinent de regarder au niveau des nucléocapsides, quel est l'effet de la phosphoprotéine.

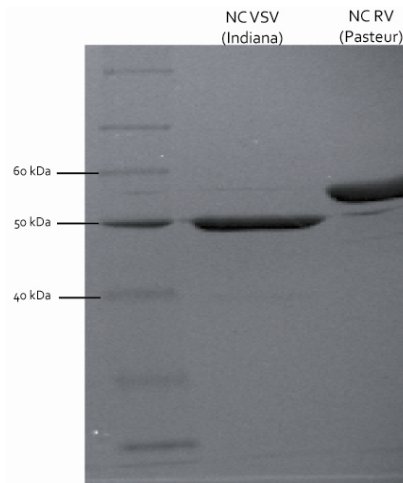
## ***II – Interaction entre les nucléocapsides virales (NC) et la phosphoprotéine (P)***

L'analyse des changements structuraux avec les complexes de type anneaux N-ARN-P ou N-ARN-CterP s'est avéré non concluant. En effet, nous n'avons pu observer directement des changements de conformation de la nucléoprotéine ou de la phosphoprotéine. Nous avons caractérisé l'interaction entre la phosphoprotéine et son domaine C-terminal sur les nucléocapsides virales (NC), matrice pour la transcription et réplication virale, en visualisant par microscopie électronique l'effet de la phosphoprotéine et notamment de son domaine C-terminal sur ces structures hélicoïdales. Par ailleurs, nous avons également utilisé l'anisotropie de fluorescence et la résonance plasmonique de surface pour déterminer l'affinité de la phosphoprotéine et de son domaine C-terminal pour les nucléocapsides.

### **A – Caractérisation structurale des nucléocapsides virales (NC VSV et NC RV)**

La collaboration avec le Dr. Danielle Blondel nous permet de recevoir des nucléocapsides virales du virus de la stomatite vésiculaire (souche Indiana Mudd-Summers) (NC VSV) et du virus de la rage (souche Pasteur) (NC RV), provenant de cellules BSR infectées. La purification des NC sur gradient de chlorure de césium élimine la majeure partie des protéines associées (polymérase virale, phosphoprotéine et protéine de matrice ([figure 79](#))).

Les nucléocapsides sont préparées dans un tampon qui contient 150 mM de NaCl. Cependant, nous avons observé en microscopie électronique en coloration négative le comportement de ces NC dans l'eau, dans 150 mM NaCl et dans des hautes concentrations salines (4 M NaCl) ([figure 80](#)). Les nucléocapsides du virus de la rage ont une structure rigidifiée en absence de sel, et présentent une structure de forme hélicoïdale lâche en présence de 150 ou 4 M NaCl.

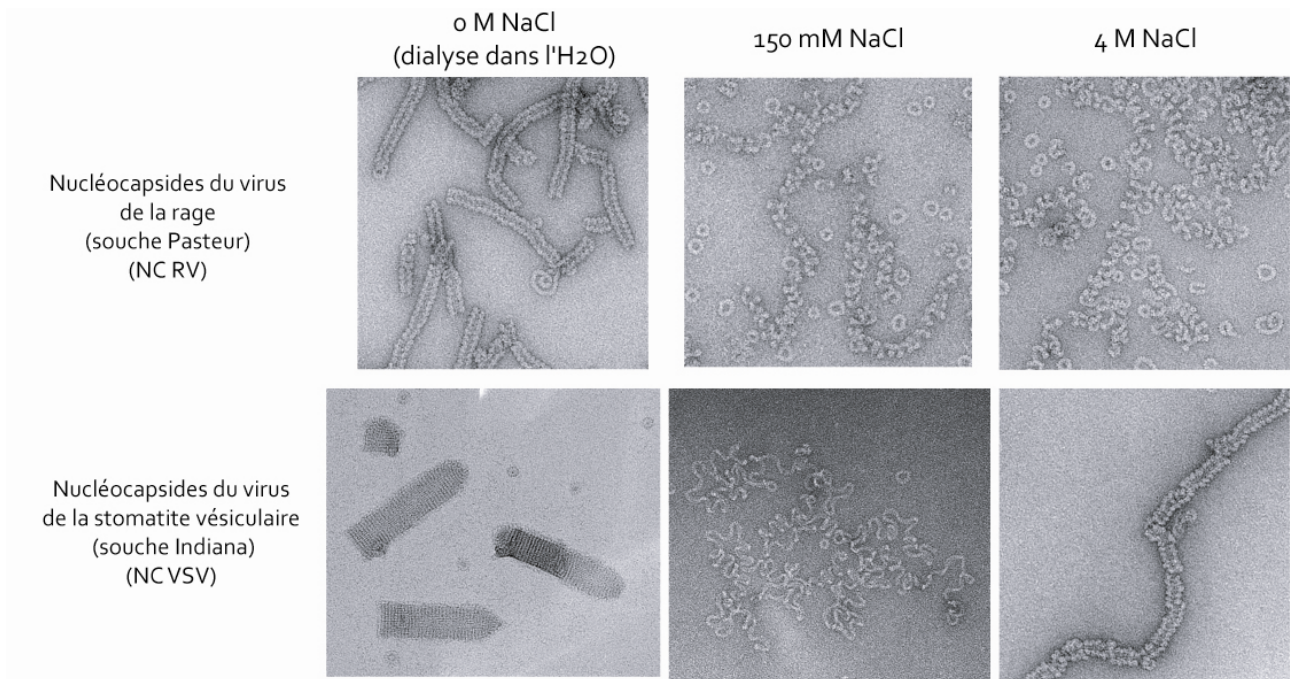


**Figure 79** - Gel SDS-PAGE 12% montrant pureté des nucléocapsides de VSV (NC VSV) et de RV (NC RV) après purification sur gradient 20/40% w/w de CsCl (Dr. Danielle BLONDEL).

Les NC de RV forment des structures beaucoup plus rigides dans l'eau que dans les conditions normales de préparation (figure 80), et pourraient alors être utilisées pour effectuer des reconstructions hélicoïdales. Nous pourrions par la suite comparer ses structures et les paramètres décrivant l'hélice (pas de l'hélice, nombre de sous-unité par tour), avec les reconstructions réalisées pour les NC du virus de la rougeole (MeV), un *Paramyxovirus* (Schoehn & al., 2004).

En ce qui concerne les nucléocapsides du virus de la stomatite vésiculaire (NC VSV), différents états de compactations sont observés : les NC VSV forment des structures en forme de balle de fusil lorsqu'elles sont dialysées dans l'eau, d'un diamètre d'environ 36 nm. Leur structure hélicoïdale est plus lâche que pour les NC de RV à 150 mM NaCl et celle-ci se compacte en présence de 4 M de NaCl (figure 80).

Nous n'avons pu reproduire l'état «en balle de fusil» sur toutes les préparations de nucléocapsides de VSV que nous avons reçu. Les informations que nous avons pu obtenir avec ses NC dialysées contre l'eau ne s'avèrent pas concluante et nous sommes toujours en train de travailler sur ces structures, et de mieux contrôler les conditions dans lesquelles nous travaillons pour caractériser plus rigoureusement les changements morphologiques que nous observons.

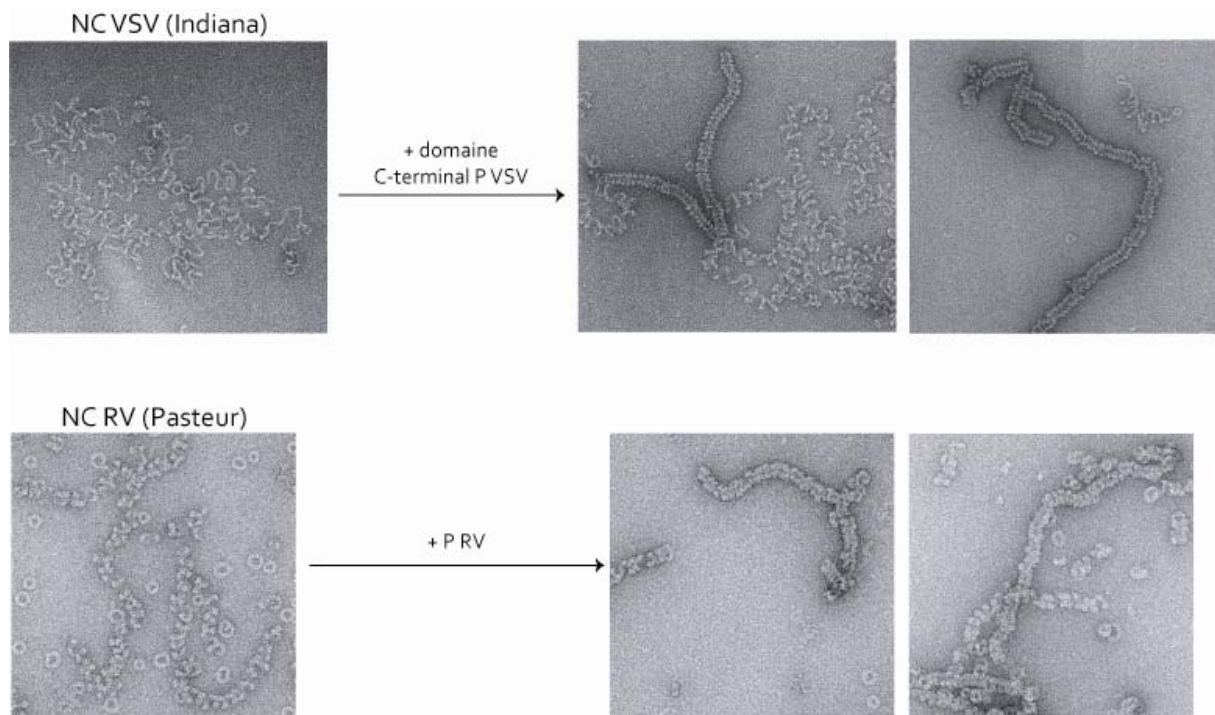


**Figure 8o** – Caractérisations structurale par microscopie électronique en coloration négative des nucléocapsides du virus de la rage (NC RV) et des nucléocapsides du virus de la stomatite vésiculaire (NC VSV) dans différentes conditions salines. Les nucléocapsides ont été observées pour des 0, 150 mM et 4 M NaCl.

## B – Compaction des nucléocapsides virales induite par la phosphoprotéine

L'analyse par microscopie électronique de ces longues structures est adéquate, la biochimie difficile. Dans les mêmes conditions de tampons, nous avons ajouté, à des NC virales de VSV et RV des quantités croissantes de phosphoprotéine, ou de son domaine C-terminal, et nous avons réalisés des grilles afin d'observer ces structures hélicoïdales par microscopie électronique en coloration négative. Lorsque l'on ajoute la phosphoprotéine ou son domaine C-terminal, les nucléocapsides subissent un changement de conformation flagrant et deviennent beaucoup plus compacte, que ce soit pour les NC de VSV que les NC de RV. Les nucléocapsides compactées ont un diamètre estimé à 20 nm.

Néanmoins, concernant les NC de VSV, l'effet de compaction semble plus important, car leur état d'enroulement initial est plus faible que pour les NC de RV. Aussi, l'ajout uniquement du domaine C-terminal et non de la protéine entière réduit le bruit de fond de l'image. Sur la première image de compaction pour les NC de VSV, on peut observer, sur le même cliché, les états non compacté et compacté de la NC. Ceci pourrait indiquer un effet coopératif de l'interaction du domaine C-terminal sur la matrice N-ARN des NC.



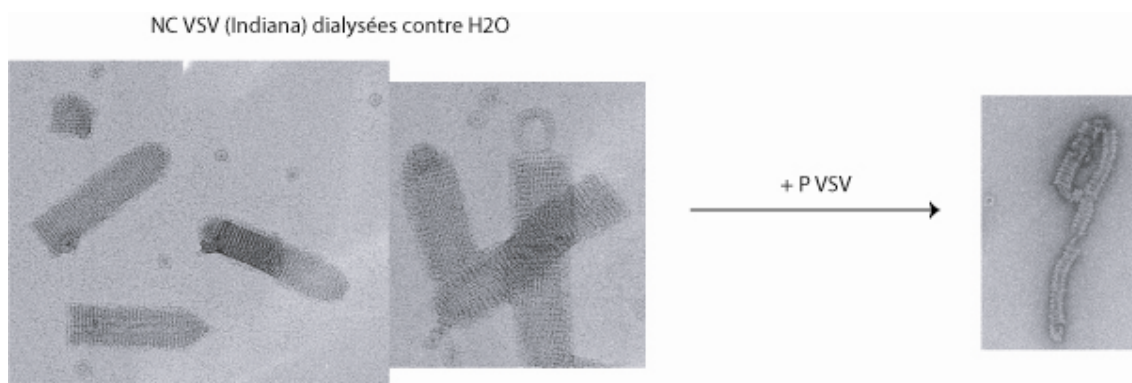
**Figure 81** - Effet de compaction du domaine C-terminal de P VSV sur les NC VSV et de la phosphoprotéine du virus de la rage sur les NC RV virales. Images de microscopie électronique à coloration négative du changement de d'état de compaction des NC de VSV et RV sous l'effet de la phosphoprotéine ou de son domaine C-terminal.

Ce résultat étant très intéressant d'un point de vue morphologie des nucléocapsides. Nous avons reproduits à plusieurs reprises ces compactations en ajoutant la phosphoprotéine, son domaine C-terminal sur les NC de VSV et RV, mais aussi avec le mutant P VSV S60D/T62D sur les NC de VSV. Ces protéines induisent le même changement de conformation, c'est-à-dire le même état de compaction. Ceci confirme que seul le domaine C-terminal est responsable de l'interaction avec le complexe nucléoprotéine-ARN et est responsable de la compaction des NC. Toutefois, la force ionique ayant été décrite comme modifiant l'enroulement et la flexibilité des NC (Heggeness & *al.*, 1980), nous avons réalisé plusieurs tests contrôles afin de vérifier que l'effet observé n'était pas du à un effet de force ionique. Nous avons aussi testé que des peptides poly-lysines ne modifiait pas la conformation des ces NC (données non montrées).

Les NC de VSV dialysées contre de l' $H_2O$  forment des structures en balle de fusil (*bullet shape*) qui pourraient être confondue avec des particules virales (figure 82). Si on ajoute de la phosphoprotéine aux NC VSV, cela perturbe et rompt ces structures en balle de fusil et les NC se retrouvent sous la même forme de compaction que décrite auparavant (figure 82). Malheureusement, ces résultats n'ont pas pu être reproduits, ce qui pose la question de la présence de la protéine M lors de la purification des NC virales, protéine décrite essentielle pour le bourgeonnement de la particule virale mais aussi pour la condensation du génome viral. Nous n'avons pu tester sur toutes les préparations de NC quelles quantités résiduelles de



protéine M et de phosphoprotéine étaient présentes. Nous sommes toujours en train de travailler sur ces structures et nous aimerions très rapidement tester l'effet de la protéine M sur le NC seules, mais aussi sur les NC en complexe avec la phosphoprotéine. Ce travail nous permettra de mieux comprendre les mécanismes d'assemblage et de condensation du génome viral lors du bourgeonnement, en identifiant plus précisément les rôles respectifs de la phosphoprotéine et de la protéine matrice dans la compaction de ces nucléocapsides.



**Figure 82 :** Nucléocapsides de VSV dialysés contre de l'eau, formant des structures en balle de fusil. L'ajout de phosphoprotéine détruit ses structures mais les NC gardent leur compaction observée auparavant.

## C – Affinité de la phosphoprotéine et de son domaine C-terminal pour les nucléocapsides (NC)

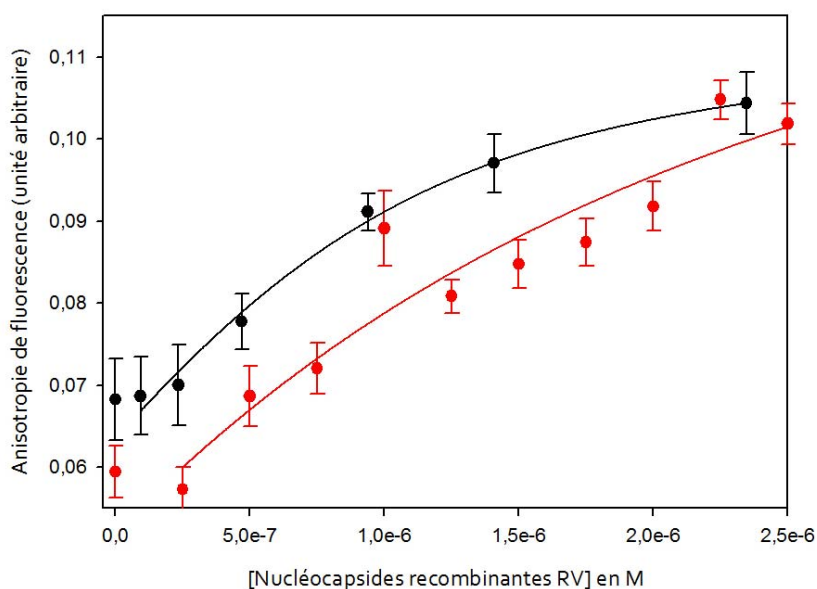
Nous avons essayé de caractériser l'affinité de la phosphoprotéine et de son domaine C-terminal pour les nucléocapsides (NC) recombinantes du virus de la rage (RV), obtenues lors de la production et purification de la nucléoprotéine du RV en cellules d'insectes en utilisant 2 techniques : l'anisotropie de fluorescence et la résonance plasmonique de surface (BIAcore).

### 1 – Anisotropie de fluorescence

L'anisotropie de fluorescence est une grandeur sans dimension liée au temps de corrélation de rotation des fluorophores qui fournit un signal indépendant de la concentration qui augmente avec le volume de la particule. La technique peut être utilisée pour suivre la formation de complexes engendrant une augmentation de la taille de la protéine marquée (Lundblad & *al.*, 1996). Dans notre cas, la mesure est utilisée pour suivre la fixation de la protéine P RV marquée avec le fluorophore IAEDANS, qui se fixe de façon covalente aux cystéines de la protéine ( $P_{2IAEDANS}$ ), ou du domaine C-terminal de P RV ( $CterP_{IAEDANS}$ ) sur des nucléocapsides recombinantes. La [figure 83](#) représente l'anisotropie moyenne mesurée ( $n=3$ ) en fonction de la concentration de nucléoprotéine ajoutée (NC recombinantes du RV) à une quantité

constante de phosphoprotéine marquée P<sub>ZIAEDANS</sub> (1 μM) ou du domaine C-terminal de P RV marqué CterP<sub>IAEDANS</sub> (0,5 μM). Les données expérimentales ne peuvent être analysées et ajustées avec un modèle simple d'association. Celles-ci ont été ajustées avec une équation quadratique par régression non linéaire.

On observe un changement de la valeur d'anisotropie lorsqu'on ajoute de façon croissante des nucléocapsides (figure 83, A & B). Cependant, lorsque l'on ajuste les courbes, dans une gamme de concentration en nucléocapsides pour lesquelles il n'y a pas de bruit parasite, les paramètres que l'on obtient n'ont pas exactement le même ordre de grandeur. En effet, les concentrations en phosphoprotéine ou domaine C-terminal de P et en nucléoprotéine étant du même ordre de grandeur, il est nécessaire de prendre en compte la déplétion des protéines ( $[P_2] \neq [P_2]_t$ ,  $[N] \neq [N]_t$ ) se produisant lorsque celle-ci interagissent. L'ajustement fournit une estimation de la constante de dissociation du complexe, et des paramètres  $A_f$  et  $A_b$  correspondant respectivement à l'anisotropie de P<sub>ZIAEDANS</sub> (ou CterP<sub>IAEDANS</sub>) et de P<sub>ZIAEDANS</sub>-NC (ou CterP<sub>IAEDANS</sub>-NC). Le  $K_D$  obtenu pour l'interaction entre les nucléocapsides et la protéine P<sub>ZIAEDANS</sub> est de  $4 \pm 2 \cdot 10^{-7} \text{ M}$  et le  $K_D$  obtenu pour l'interaction entre les nucléocapsides et le domaine C-terminal de la protéine P, CterP<sub>IAEDANS</sub> est de  $2 \pm 3 \cdot 10^{-6} \text{ M}$ . L'affinité de la phosphoprotéine ou de son domaine C-terminal est relativement forte, de l'ordre du μM. Il y a un facteur 5 de différence entre ces deux constantes calculées, ce qui ne semble pas significatif. En effet, nous avons confirmé que seul le domaine C-terminal était responsable de l'interaction avec les NC, et ceci est confirmé par le calcul de ces  $K_D$  qui sont très proches.

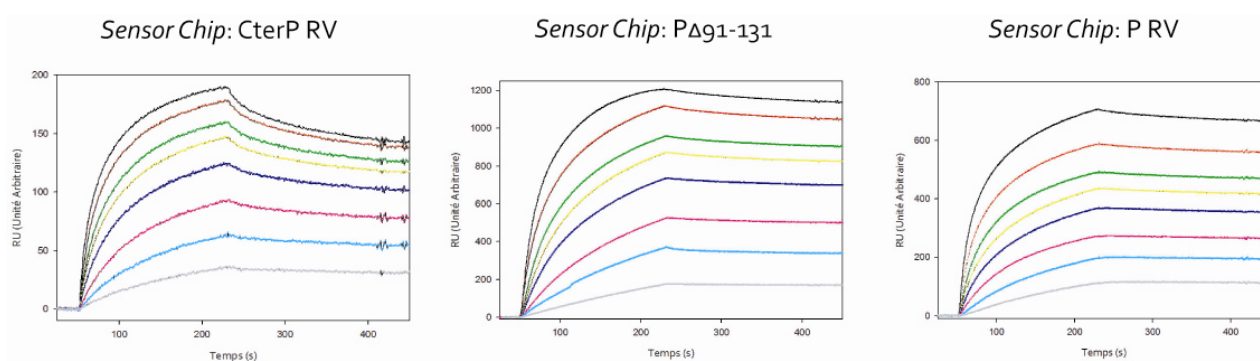


**Figure 83** - Mesure par anisotropie de fluorescence de l'affinité de la phosphoprotéine (P RV) ou de son domaine C-terminal (CterP RV) pour les nucléocapsides recombinants du virus de la rage. Les cercles noirs (titration NC par la protéine P) et rouges (titration NC par le domaine C-terminal de P) assortis de barres d'erreurs représentent une moyenne de l'anisotropie mesurée sur trois expériences indépendantes. La courbe représente une régression non linéaire des données selon le modèle quadratique décrit ci-dessus ( $R = 0,99$ ). L'intensité totale de fluorescence pour chaque point expérimental est constante (données non montrées)

Il semble clair que l'interaction entre le domaine C-terminal de la phosphoprotéine et les nucléocapsides ne suit pas un modèle simple. Il est possible que les changements de conformations (observé en microscopie électronique) modifient la valeur d'anisotropie mesurée et rend l'analyse des données difficile. Aussi, il se peut que le fluorophore ne soit pas fixé de façon homogène sur la protéine : la phosphoprotéine de la rage possède 2 cystéines en position 261 et 297 (souche Pasteur), c'est pourquoi nous sommes actuellement en train de réaliser des mutants de P RV :  $P_{\Delta C261}$ ,  $P_{\Delta C297}$  et  $P_{\Delta C261\Delta 297}$ . Ainsi, le marquage fluorophore sera unique et nous pourrons comparer le signal obtenu.

## 2 – Résonance plasmonique de surface

Sur les mêmes puces que nous avons préparées avec la protéine P du virus de la rage,  $P_{\Delta 91-131}$  et le domaine C-terminal de la protéine P (RV), nous avons aussi injecté des nucléocapsides recombinantes de la rage. De la même façon que pour les anneaux de nucléoprotéine-ARN, nous avons obtenu un signal significatif mais il a été impossible d'ajuster de façon globale les courbes que nous avons obtenues. Les données expérimentales brutes d'association et de dissociation du complexe sont présentées en [figure 84, A](#). Comme on peut le remarquer, pour les injections où en nucléocapsides et même pour les courbes où la concentration en NC est la plus haute, le plateau de saturation n'est pas atteint, ce qui rend aussi impossible l'analyse de ces données à l'équilibre. Nous avons alors essayé d'ajuster les courbes pour l'association et la dissociation de façon indépendante, et nous avons observé une dépendance des valeurs de constantes trouvées en fonction de la concentration en nucléocapsides injectées (données non montrées). Ceci nous indique clairement que l'association et la dissociation ne peuvent être ajustées de façon globale avec des paramètres simples, indiquant une certaine complexité lors de l'interaction.



**Figure 84** - Données brutes obtenus par résonance plasmonique de surface avec les *sensor chip* *CM5* où la protéine P (RV),  $P_{\Delta 91-131}$  et le domaine CterP RV sont fixés. Chaque courbe correspond à une mesure au cours du temps pour une injection de 60  $\mu$ l de NC à une concentration donnée (10, 25, 50, 100, 150, 200, 300 et 400 nM).

## D – Conclusions

Nous avons initié ce travail sur les nucléocapsides virales grâce à une collaboration avec le Dr. Danielle BLONDEL, afin d'observer l'effet de la phosphoprotéine et de son domaine C-terminal sur les NC. Clairement, la protéine P, et notamment son domaine C-terminal, responsable de l'interaction N-ARN, induit un changement de conformation de la nucléocapside. Il reste à confirmer si cette compaction garde l'ARN protégé au sein de la nucléocapside, ou est-ce que cette compaction, modifiant la forme de la NC va permettre à l'ARN d'être accessible. En effet, nous pensons que la phosphoprotéine joue le rôle d'un intermédiaire, mais surtout elle permet à l'ARN polymérase ARN-dépendante d'avoir accès à l'information du génome viral. Il semble donc essentiel de regarder l'accessibilité de l'ARN dans les différentes conformations des nucléocapsides (seules, en présence de la protéine P ou de son domaine C-terminal).

L'affinité de la protéine P ou de son domaine C-terminal pour la nucléoprotéine a été étudiée par deux méthodes : l'anisotropie de fluorescence et la résonance plasmonique de surface. Pour ces deux techniques nous obtenons un signal significatif, confirmant l'interaction NC-ARN-P, cependant, il nous a été impossible d'ajuster ces données de façon simple, suggérant une interaction plus complexe, qui met en évidence changement de conformation des NC. Nous essayons toujours de trouver un modèle adéquat pour l'analyse de ces données afin d'obtenir les paramètres pour l'association et la dissociation du complexe.

## **CHAPITRE V : DISCUSSION**

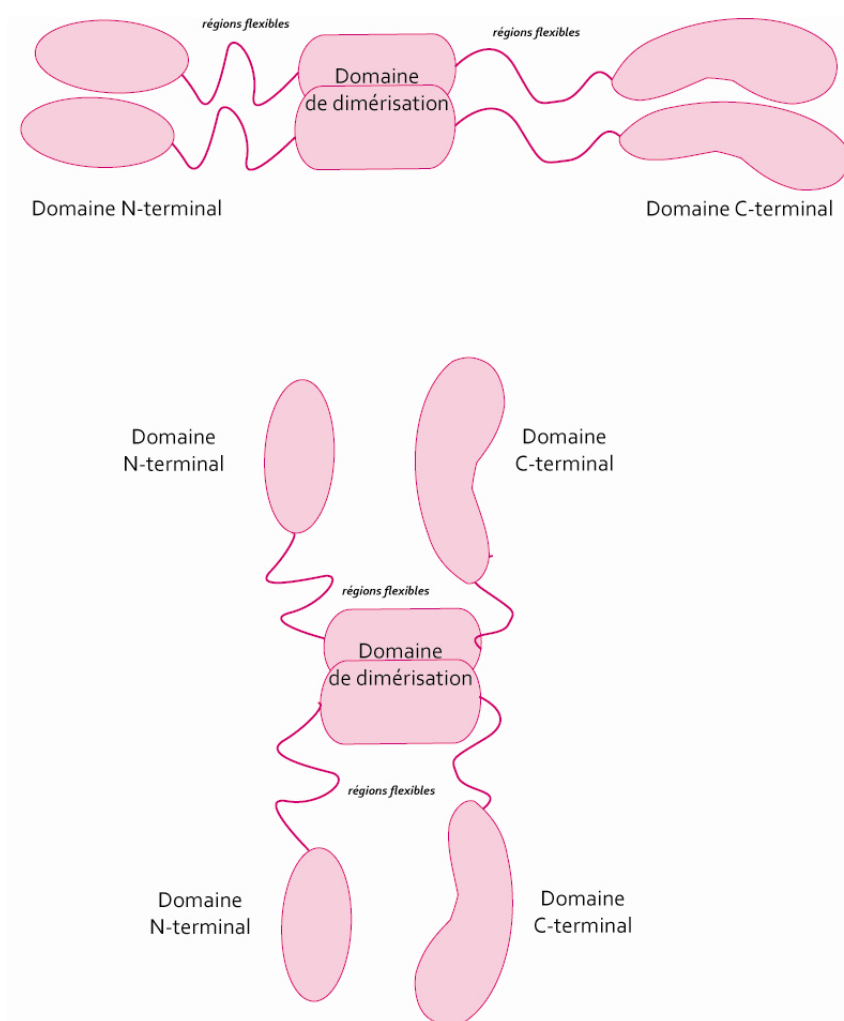
## ***I – Organisation modulaire de la phosphoprotéine des Rhabdoviridae***

Les formes non phosphorylées des phosphoprotéines du virus de la rage et de la stomatite vésiculaire sont dimériques en solution. Leur caractérisation biophysique par ultracentrifugation analytique et diffusion des neutrons aux petits angles nous a permis de mettre en évidence une asymétrie de la phosphoprotéine. La forme de la phosphoprotéine du virus de la rage et de la stomatite vésiculaire peut être assimilée à un cylindre d'environ 18 nm de longueur et 5 nm de diamètre. Aussi, la caractérisation du mutant de P VSV S60D/T62D, mimant la phosphorylation N-terminale de la phosphoprotéine du VSV nous indique que la phosphorylation n'est pas nécessaire pour l'oligomérisation de la protéine, tout comme chez la phosphoprotéine du virus de la rage. Pour la phosphoprotéine du VSV, une incubation à 30°C induit la formation de structure en forme de filament (Gerard & al., 2007).

L'analyse bioinformatique des séquences primaires des phosphoprotéines de RV et VSV a révélé trois régions structurées séparées par deux régions flexibles. La caractérisation biophysique et biochimique des domaines de la phosphoprotéine du virus de la rage montre une organisation modulaire qui peut être comparable à celle proposée pour la phosphoprotéine des *Paramyxovirus* (Gerard 2008). Même si les structures de ces domaines entre les différentes familles sont différentes, il semble que l'organisation en module de la phosphoprotéine soit conservée au sein des *Mononegavirales*. Chaque module de la phosphoprotéine joue un rôle dans l'interaction avec ses partenaires cellulaires et viraux, et de permet d'assurer les différentes fonctions au cours de la multiplication virale.

La structure du domaine C-terminal de la phosphoprotéine du VSV est homologue à la structure du domaine C-terminal de la phosphoprotéine du virus de la rage. Même si les conservations de séquences en acides aminés est faible au sein d'une même famille de l'ordre des *Mononegavirales*, il semble que la structure tridimensionnelle soit conservée (Ribeiro & al., 2008). En effet, les homologues de structures chez les virus de la rage et de la stomatite vésiculaire ont été mises en évidence pour la nucléoprotéine, le domaine C-terminal de la phosphoprotéine et aussi pour la protéine matrice (Ribeiro & al., 2008) (Assenberg & al., 2008) (Albertini & al., 2008). Leur organisation structurale semble conservée tout comme leur états d'oligomérisation. Au sein d'une même famille, l'organisation structurale est conservée, et au sein de l'ordre des *Mononegavirales*, l'organisation modulaire de la phosphoprotéine est également conservée. Chez les *Paramyxovirus*, l'oligomérisation de la phosphoprotéine en *coiled coil* est au niveau de leur domaine d'oligomérisation est conservée, notamment concernant la phosphoprotéine du virus Sendai, de la rougeole et aussi du RSV (Tarbouriech & al., 2000) (Llorente & al., 2008). Chez les *Rhabdoviridae*, il semble que l'état dimérique de la phosphoprotéine soit aussi conservé, chez RV, VSV mais aussi chez le virus Chandipura.

Les données de diffusions des rayons X aux petits angles ont permis de modéliser la phosphoprotéine du VSV en incluant les données structurales du domaine central et du domaine C-terminal. Ainsi, nous avons pu confirmer l'organisation modulaire de la phosphoprotéine dans le contexte d'un dimère allongé. Cependant, les modèles que nous avons réalisés imposent que le dimère de phosphoprotéine soit parallèle, c'est-à-dire que les domaines C-terminaux ou N-terminaux de la phosphoprotéine se retrouvent du même côté, comme la structure du domaine d'oligomérisation le laisse supposer. Les modules de la phosphoprotéine sont séparés par des régions flexibles et il est possible que la conformation de la phosphoprotéine soit alors antiparallèle (figure 85). En effet, la forme générale en forme de cylindre la phosphoprotéine serait respectée. Aussi, nous ne pouvons exclure le fait qu'il existe dans la particule virale ces formes.



**Figure 85** – Hypothèse sur l'association du dimère de phosphoprotéine chez le virus de la rage et de la stomatite vésiculaire.

Sur la base des données obtenues pour la phosphoprotéine du virus Mokola, il a été montré que la phosphoprotéine monomérique permettait la transcription et la réplication virale (Jacob & *al.*, 2001). Ceci indique que l'oligomérisation de la phosphoprotéine n'est pas essentielle pour les activités de transcription et de réplication. Il serait à présent nécessaire de réévaluer les effets de la phosphorylation

en N-terminal et C-terminal de la phosphoprotéine sur l'état d'oligomérisation ainsi que sur les activités de transcription et de réplication. L'oligomérisation de la phosphoprotéine pourrait être nécessaire pour d'autres processus lors de la multiplication virale, comme l'encapsidation du génome viral par la nucléoprotéine. L'organisation en module permet à la phosphoprotéine de se lier au complexe nucléoprotéine-ARN et d'ajouter à l'ARN naissant une nucléoprotéine pour l'encapsidation.

Il serait intéressant de comparer les activités de transcription et de réplication *in vitro* de la phosphoprotéine non phosphorylée avec celle de la phosphoprotéine extraite de la particule virale. Il semble clair que la phosphorylation des résidus N-terminaux de la protéine P ait un rôle clé pour les activités de transcription et de réplication. Il serait possible dans les deux complexes transcriptases et répliques trouvé par Qanungo & *al.* que la phosphoprotéine soit dans un état de phosphorylation différent (Qanungo & *al.*, 2004), ce qui lui permettrait d'interagir avec d'autres partenaires cellulaires, et d'avoir des effets sur des voies de signalisation de la cellule hôte. Il serait intéressant de revoir si la phosphorylation de la phosphoprotéine induit un changement de conformation de la protéine ou des protéines virales en complexe.

## **II – Interaction de la phosphoprotéine et des anneaux de nucléoprotéines-ARN**

Les analyses par SEC-MALLS-RI ont permis de conclure sur la stoechiométrie des complexes anneaux-phosphoprotéine, anneaux-domaine C-terminal de la phosphoprotéine et anneaux-P<sub>Δ91-131</sub> du virus de la rage. En effet, les anneaux N9, N10 et N11 peuvent fixer deux dimères de phosphoprotéine. Le dimère de phosphoprotéine interagit avec les anneaux par l'intermédiaire d'un de ces domaines C-terminaux. Il existe au sein de l'anneau de nucléoprotéine-ARN un encombrement stérique qui ne permet une stoechiométrie 1:1 nucléoprotéine : domaine C-terminal de la phosphoprotéine. Cependant, la stoechiométrie présente sur l'anneau est représentative de la stoechiométrie nucléoprotéine/phosphoprotéine que l'on retrouve dans la particule virale (figure 86).

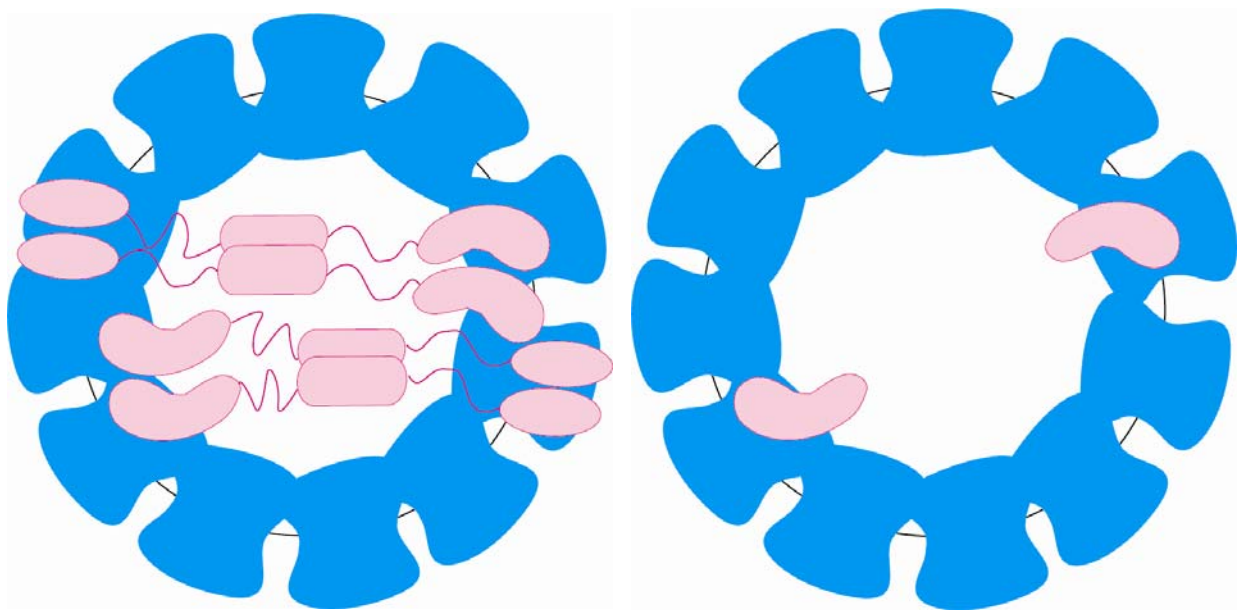
Les études biochimiques par électrophorèse en conditions natives ont permis d'observer les différentes formes de complexes anneaux-phosphoprotéine, anneaux-domaine C-terminal de la phosphoprotéine et anneaux-P<sub>Δ91-131</sub>, et notamment de préciser les formes intermédiaires de complexes avant la saturation.

Les analyses structurales par cristallographie des rayons X et diffusion des rayons X aux petits angles ne permettent pas de conclure quant à un changement éventuel de conformation de la nucléoprotéine. Les conditions d'analyse en microscopie électronique en coloration négative semblent



dissocier le complexe anneaux-P ou anneaux-CterP. En effet, nous n'avons pas testé l'effet du pH sur ce complexe et il est possible que lors du traitement avec l'acétate d'uranyle, la variation de pH dissocie le complexe. Les analyses en cryo-microscopie électronique du complexe N11-P-MBP ont permis d'identifier sur la première classification d'images quatre densités électroniques pouvant correspondre à la protéine MBP. Cependant, l'organisation de ces densités à l'extérieur de l'anneau laisse supposer que la phosphoprotéine peut exister sous la conformation parallèle et antiparallèle en solution. Ceci laisse supposer que la phosphoprotéine est une protéine dynamique et qu'elle peut adopter plusieurs conformations dans la nucléocapside.

Nous n'avons pas pu clairement identifier les deux sites de fixation du domaine C-terminal de la phosphoprotéine sur les anneaux, que ce soit les anneaux N9, N10 ou N11. Cependant, la stoechiométrie des complexes anneaux-P et anneaux-CterP étant la même, cela suggère qu'il existe un site particulier sur l'anneau pour la fixation du CterP, par exemple au niveau des extrémités 5' ou 3' de l'ARN, où les nucléoprotéines formeraient un site préférentiel pour le domaine C-terminal de la phosphoprotéine.



**Figure 86** – Schématisation de la fixation de la phosphoprotéine (à gauche) et du domaine C-terminal de la phosphoprotéine (à droite) sur les anneaux N11. Le dimère de phosphoprotéine est schématisé parallèle, mais la conformation antiparallèle peut être envisagée.

Il est possible que la conformation fermée de l'anneau empêche d'observer des changements de conformation dus à la phosphoprotéine. En effet, il est possible que dans ces conditions la nucléoprotéine puisse toujours interagir par l'intermédiaire de son domaine C-terminal avec le domaine C-terminal de la phosphoprotéine mais que le changement de conformation ne puisse avoir lieu puisque la nucléoprotéine séquestre totalement l'ARN et possède une conformation complètement fermée.

Cependant, toutes ces données appuient le modèle pour la réplication et la transcription proposé auparavant dans le laboratoire (Albertini & *al.*, 2008).

### **III – Interaction de la phosphoprotéine et des nucléocapsides**

La conformation des nucléocapsides virales est variable suivant les conditions dans lesquelles elles se trouvent. En effet, nous avons pu observer que les nucléocapsides du virus de la stomatite vésiculaire forment des structures en balle de fusil, similaire à celle retrouvée dans la particule virale lorsqu'elles sont dans l'eau. L'ajout de phosphoprotéine dans des conditions salines normales (150 mM NaCl) engendre un changement de conformation important des nucléocapsides, montrant une compaction de la structure hélicoïdale, que ce soit pour les nucléocapsides du virus de la stomatite vésiculaire ou pour celle du virus de la rage. Les essais réalisés ont confirmés que seul le domaine C-terminal est responsable de l'interaction avec le complexe nucléoprotéine-ARN, et car il permet à lui seul de compacter les nucléocapsides.

Cependant, les analyses en microscopie électronique en coloration négative des anneaux suggèrent que dans ces conditions la phosphoprotéine et le domaine C-terminal se dissocie des anneaux. Les seules données structurales ont été obtenues pour le complexe N11-P-MBP en conditions natives. Pour les nucléocapsides, il semblerait que la phosphoprotéine soit toujours fixée, l'effet étant clairement observé. Il semblerait que la phosphoprotéine fixe mieux les nucléocapsides, ou que la réorganisation conformationnelle de la nucléocapside protège la phosphoprotéine du changement de pH opéré lors de la préparation en microscopie électronique en coloration négative, empêchant la dissociation du complexe. Nous n'avons pas testé si la phosphoprotéine était présente dans les nucléocapsides, il serait intéressant de réaliser un marquage anticorps afin de confirmer la présence de la phosphoprotéine dans les nucléocapsides compactées. Aussi, notre hypothèse de travail est que la phosphoprotéine va permettre à la polymérase virale d'accéder au génome. Les conditions d'expérimentation montrent que la phosphoprotéine compacte les nucléocapsides mais l'accessibilité de l'ARN doit être testée. Il est possible que ce changement de conformation permette à l'ARN d'être plus accessible au solvant et donc à la polymérase.

## **REFERENCES BIBLIOGRAPHIQUES**

- (1994). "The CCP4 suite: programs for protein crystallography." *Acta Crystallogr D Biol Crystallogr* **50**(Pt 5): 760-3.
- Abraham, G., Rhodes, D.P. and Banerjee, A.K. (1975). "The 5' terminal structure of the methylated mRNA synthesized in vitro by vesicular stomatitis virus." *Cell* **5**(1): 51-8.
- Adrian, M., Dubochet, J., Lepault, J. and McDowell, A.W. (1984). "Cryo-electron microscopy of viruses." *Nature* **308**(5954): 32-6.
- Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K. and Miyano, M. (1999). "Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus." *Structure* **7**(11): 1417-26.
- Albertini, A.A., Clapier, C.R., Wernimont, A.K., Schoehn, G., Weissenhorn, W. and Ruigrok, R.W. (2007). "Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings." *J Struct Biol* **158**(1): 129-33.
- Albertini, A.A., Schoehn, G., Weissenhorn, W. and Ruigrok, R.W. (2008). "Structural aspects of rabies virus replication." *Cell Mol Life Sci* **65**(2): 282-94.
- 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**(5785): 360-3.
- 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**(1): 308-13.
- Arnheiter, H., Davis, N.L., Wertz, G., Schubert, M. and Lazzarini, R.A. (1985). "Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis." *Cell* **41**(1): 259-67.
- Asenjo, A. and Villanueva, N. (2000). "Regulated but not constitutive human respiratory syncytial virus (HRSV) P protein phosphorylation is essential for oligomerization." *FEBS Lett* **467**(2-3): 279-84.
- Assenberg, R., Delmas, O., Graham, S.C., Verma, A., Berrow, N., Stuart, D.I., Owens, R.J., Bourhy, H. and Grimes, J.M. (2008). "Expression, purification and crystallization of a lyssavirus matrix (M) protein." *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**(Pt 4): 258-62.
- Badrane, H., Bahloul, C., Perrin, P. and Tordo, N. (2001). "Evidence of two Lyssavirus phylogroups with distinct pathogenicity and immunogenicity." *J Virol* **75**(7): 3268-76.
- Baer, G.M., Shaddock, J.H., Quirion, R., Dam, T.V. and Lentz, T.L. (1990). "Rabies susceptibility and acetylcholine receptor." *Lancet* **335**(8690): 664-5.
- 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**(2): 572-6.
- Banerjee, A.K. (1987). "Transcription and replication of rhabdoviruses." *Microbiol Rev* **51**(1): 66-87.
- Banerjee, A.K. and Barik, S. (1992). "Gene expression of vesicular stomatitis virus genome RNA." *Virology* **188**(2): 417-28.
- 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**(12): 7246-53.
- Barik, S. and Banerjee, A.K. (1992). "Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation." *J Virol* **66**(2): 1109-18.
- Barik, S., McLean, T. and Dupuy, L.C. (1995). "Phosphorylation of Ser232 directly regulates the transcriptional activity of the P protein of human respiratory syncytial virus: phosphorylation of Ser237 may play an accessory role." *Virology* **213**(2): 405-12.

- Barr, J.N., Whelan, S.P. and Wertz, G.W. (2002). "Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus." *Biochim Biophys Acta* **1577**(2): 337-53.
- Basak, S., Polley, S., Basu, M., Chattopadhyay, D. and Roy, S. (2004). "Monomer and dimer of Chandipura virus unphosphorylated P-protein binds leader RNA differently: implications for viral RNA synthesis." *J Mol Biol* **339**(5): 1089-101.
- 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**(13): 3158-65.
- Bearzotti, M., Delmas, B., Lamoureux, A., Loustau, A.M., Chilmonczyk, S. and Bremont, M. (1999). "Fish rhabdovirus cell entry is mediated by fibronectin." *J Virol* **73**(9): 7703-9.
- Bhella, D., Ralph, A., Murphy, L.B. and Yeo, R.P. (2002). "Significant differences in nucleocapsid morphology within the Paramyxoviridae." *J Gen Virol* **83**(Pt 8): 1831-9.
- Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W.P., Ruigrok, R.W. and Marion, D. (2004). "Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution." *Virology* **319**(2): 201-11.
- 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**(52): 7957-70.
- 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): 769-79.
- Bourhy, H., Kissi, B. and Tordo, N. (1993). "Molecular diversity of the Lyssavirus genus." *Virology* **194**(1): 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): 2063-74.
- Broughan, J.H. and Wunner, W.H. (1995). "Characterization of protein involvement in rabies virus binding to BHK-21 cells." *Arch Virol* **140**(1): 75-93.
- 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**(12): 7673-81.
- 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**(10): 5803-12.
- Bukreyev, A., Skiadopoulos, M.H., Murphy, B.R. and Collins, P.L. (2006). "Nonsegmented negative-strand viruses as vaccine vectors." *J Virol* **80**(21): 10293-306.
- Burrage, T.G., Tignor, G.H. and Smith, A.L. (1985). "Rabies virus binding at neuromuscular junctions." *Virus Res* **2**(3): 273-89.
- 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**(8): 4822-30.
- Cann, A.J. (1997). "Principles of Molecular Virology".
- Carneiro, F.A., Stauffer, F., Lima, C.S., Juliano, M.A., Juliano, L. and Da Poian, A.T. (2003). "Membrane fusion induced by vesicular stomatitis virus depends on histidine protonation." *J Biol Chem* **278**(16): 13789-94.
- 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**(3): 407-14.

- Chattopadhyay, D. and Chattopadhyay, D. (1994). "Cloning of the chandipura virus phosphoprotein encoding gene and its expression in *Escherichia coli*." *Cell Mol Biol Res* **40**(7-8): 693-8.
- Chattopadhyay, D., Raha, T. and Chattopadhyay, D. (1997). "Single serine phosphorylation within the acidic domain of Chandipura virus P protein regulates the transcription in vitro." *Virology* **239**(1): 11-9.
- 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**(2): 200-12.
- Chen, M., Ogino, T. and Banerjee, A.K. (2006). "Mapping and functional role of the self-association domain of vesicular stomatitis virus phosphoprotein." *J Virol* **80**(19): 9511-8.
- 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* **12**(2): 227-35.
- 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**(2): 707-12.
- 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): 2889-96.
- 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**(3): 1925-30.
- Chong, L.D. and Rose, J.K. (1993). "Membrane association of functional vesicular stomatitis virus matrix protein in vivo." *J Virol* **67**(1): 407-14.
- Chow, T.L., Chow, F.H. and Hanson, R.P. (1954). "Morphology of vesicular stomatitis virus." *J Bacteriol* **68**(6): 724-6.
- 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**(5470): 1432-5.
- Clague, M.J., Schoch, C., Zech, L. and Blumenthal, R. (1990). "Gating kinetics of pH-activated membrane fusion of vesicular stomatitis virus with cells: stopped-flow measurements by dequenching of octadecylrhodamine fluorescence." *Biochemistry* **29**(5): 1303-8.
- Claverie, J.M., Ogata, H., Audic, S., Abergel, C., Suhre, K. and Fournier, P.E. (2006). "Mimivirus and the emerging concept of "giant" virus." *Virus Res* **117**(1): 133-44.
- 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**(2): 197-204.
- Conti, C., Hauttecoeur, B., Morelec, M.J., Bizzini, B., Orsi, N. and Tsiang, H. (1988). "Inhibition of rabies virus infection by a soluble membrane fraction from the rat central nervous system." *Arch Virol* **98**(1-2): 73-86.
- Cotton, W.E. (1927). "Vesicular stomatitis." *Vet. Med.*
- Crimmins, D.L., Mehard, W.B. and Schlesinger, S. (1983). "Physical properties of a soluble form of the glycoprotein of vesicular stomatitis virus at neutral and acidic pH." *Biochemistry* **22**(25): 5790-6.
- Curran, J., Boeck, R., Lin-Marq, N., Lupas, A. and Kolakofsky, D. (1995). "Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils." *Virology* **214**(1): 139-49.
- Curran, J., Marq, J.B. and Kolakofsky, D. (1995). "An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication." *J Virol* **69**(2): 849-55.

- Das, T., Gupta, A.K., Sims, P.W., Gelfand, C.A., Jentoft, J.E. and Banerjee, A.K. (1995). "Role of cellular casein kinase II in the function of the phosphoprotein (P) subunit of RNA polymerase of vesicular stomatitis virus." *J Biol Chem* **270**(41): 24100-7.
- David M Knipe, P.M.H. "Fields Virology".
- De Brito, T. and Araujo Mde, F. (1973). "Ultrastructure of the Negri body in human rabies." *J Neurol Sci* **20**(4): 363-72.
- Delmas, O., Holmes, E.C., Talbi, C., Larrous, F., Dacheux, L., Bouchier, C. and Bourhy, H. (2008). "Genomic diversity and evolution of the lyssaviruses." *PLoS ONE* **3**(4): e2057.
- 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**(16): 4228-36.
- 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**(6): 2808-14.
- 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**(Pt 11): 2087-90.
- Doms, R.W., Keller, D.S., Helenius, A. and Balch, W.E. (1987). "Role for adenosine triphosphate in regulating the assembly and transport of vesicular stomatitis virus G protein trimers." *J Cell Biol* **105**(5): 1957-69.
- Durrer, P., Gaudin, Y., Ruigrok, R.W., Graf, R. and Brunner, J. (1995). "Photolabeling identifies a putative fusion domain in the envelope glycoprotein of rabies and vesicular stomatitis viruses." *J Biol Chem* **270**(29): 17575-81.
- 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**(5): 2233-43.
- Emerson, S.U. (1982). "Reconstitution studies detect a single polymerase entry site on the vesicular stomatitis virus genome." *Cell* **31**(3 Pt 2): 635-42.
- 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**(2): 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**(6): 1348-56.
- Emsley, P. and Cowtan, K. (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-32.
- 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**(22): 12074-82.
- 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**(Pt 6): 1613-21.
- Flamand, A., Raux, H., Gaudin, Y. and Ruigrok, R.W. (1993). "Mechanisms of rabies virus neutralization." *Virology* **194**(1): 302-13.
- Fooks, A.R., Brookes, S.M., Johnson, N., McElhinney, L.M. and Hutson, A.M. (2003). "European bat lyssaviruses: an emerging zoonosis." *Epidemiol Infect* **131**(3): 1029-39.
- Fooks, A.R., McElhinney, L.M., Pounder, D.J., Finnegan, C.J., Mansfield, K., Johnson, N., Brookes, S.M., Parsons, G., White, K., McIntyre, P.G. and Nathwani, D. (2003). "Case report: isolation of a European bat lyssavirus type 2a from a fatal human case of rabies encephalitis." *J Med Virol* **71**(2): 281-9.
- Gao, Y., Greenfield, N.J., Cleverley, D.Z. and Lenard, J. (1996). "The transcriptional form of the phosphoprotein of vesicular stomatitis virus is a trimer: structure and stability." *Biochemistry* **35**(46): 14569-73.

- Gao, Y. and Lenard, J. (1995). "Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly." *J Virol* **69**(12): 7718-23.
- Gao, Y. and Lenard, J. (1995). "Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II." *Embo J* **14**(6): 1240-7.
- 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)**: 2437-40.
- Gaudier, M., Gaudin, Y. and Knossow, M. (2001). "Cleavage of vesicular stomatitis virus matrix protein prevents self-association and leads to crystallization." *Virology* **288**(2): 308-14.
- Gaudier, M., Gaudin, Y. and Knossow, M. (2002). "Crystal structure of vesicular stomatitis virus matrix protein." *Embo J* **21**(12): 2886-92.
- Gaudin, Y. (2000). "Reversibility in fusion protein conformational changes. The intriguing case of rhabdovirus-induced membrane fusion." *Subcell Biochem* **34**: 379-408.
- Gaudin, Y., Barge, A., Ebel, C. and Ruigrok, R.W. (1995). "Aggregation of VSV M protein is reversible and mediated by nucleation sites: implications for viral assembly." *Virology* **206**(1): 28-37.
- Gaudin, Y., de Kinkelin, P. and Benmansour, A. (1999). "Mutations in the glycoprotein of viral haemorrhagic septicaemia virus that affect virulence for fish and the pH threshold for membrane fusion." *J Gen Virol* **80 ( Pt 5)**: 1221-9.
- Gaudin, Y., Ruigrok, R.W., Knossow, M. and Flamand, A. (1993). "Low-pH conformational changes of rabies virus glycoprotein and their role in membrane fusion." *J Virol* **67**(3): 1365-72.
- Gaudin, Y., Ruigrok, R.W., Tuffereau, C., Knossow, M. and Flamand, A. (1992). "Rabies virus glycoprotein is a trimer." *Virology* **187**(2): 627-32.
- Gaudin, Y., Tuffereau, C., Benmansour, A. and Flamand, A. (1991). "Fatty acylation of rabies virus proteins." *Virology* **184**(1): 441-4.
- Gaudin, Y., Tuffereau, C., Durrer, P., Flamand, A. and Ruigrok, R.W. (1995). "Biological function of the low-pH, fusion-inactive conformation of rabies virus glycoprotein (G): G is transported in a fusion-inactive state-like conformation." *J Virol* **69**(9): 5528-34.
- Gaudin, Y., Tuffereau, C., Segretain, D., Knossow, M. and Flamand, A. (1991). "Reversible conformational changes and fusion activity of rabies virus glycoprotein." *J Virol* **65**(9): 4853-9.
- Gerard, F.C., Jr, E.D., Albertini, A.A., Gutsche, I., Zaccari, G., Ruigrok, R.W. and Jamin, M. (2007). "Unphosphorylated Rhabdoviridae Phosphoproteins Form Elongated Dimers in Solution." *Biochemistry*.
- Gerard, F.C.A., Ribeiro Jr., Euripedes de Almeida, Leyrat, Cédric, Ivanov, Ivan, Blondel, Danielle, Longhi, Sonia, Ruigrok, Rob W. H. & Jamin, Marc. (2008). "Modular Organization of rabies virus phosphoprotein."
- 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**(Pt 7): 1757-61.
- 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**(49): 46864-70.
- Green, T.J., Zhang, X., Wertz, G.W. and Luo, M. (2006). "Structure of the vesicular stomatitis virus nucleoprotein-RNA complex." *Science* **313**(5785): 357-60.
- Gupta, A.K. and Banerjee, A.K. (1997). "Expression and purification of vesicular stomatitis virus N-P complex from *Escherichia coli*: role in genome RNA transcription and replication in vitro." *J Virol* **71**(6): 4264-71.



- 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**(1): 91-8.
- Gupta, A.K., Das, T. and Banerjee, A.K. (1995). "Casein kinase II is the P protein phosphorylating cellular kinase associated with the ribonucleoprotein complex of purified vesicular stomatitis virus." *J Gen Virol* **76** (Pt 2): 365-72.
- Gupta, A.K., Shaji, D. and Banerjee, A.K. (2003). "Identification of a novel tripartite complex involved in replication of vesicular stomatitis virus genome RNA." *J Virol* **77**(1): 732-8.
- Hanson, R.P. (1981). "Virus diseases of food animals".
- Harty, R.N., Brown, M.E., McGettigan, J.P., Wang, G., Jayakar, H.R., Huibregtse, J.M., Whitt, M.A. and Schnell, M.J. (2001). "Rhabdoviruses and the cellular ubiquitin-proteasome system: a budding interaction." *J Virol* **75**(22): 10623-9.
- Harty, R.N., Brown, M.E., Wang, G., Huibregtse, J. and Hayes, F.P. (2000). "A PPxY motif within the VP40 protein of Ebola virus interacts physically and functionally with a ubiquitin ligase: implications for filovirus budding." *Proc Natl Acad Sci U S A* **97**(25): 13871-6.
- He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., Ge, R., Rao, Z. and Liu, Y. (2008). "Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus." *Nature* **454**(7208): 1123-6.
- 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**(5): 2631-5.
- 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**(5784): 217-20.
- 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**(5320): 1845-8.
- Herman, R.C. (1986). "Internal initiation of translation on the vesicular stomatitis virus phosphoprotein mRNA yields a second protein." *J Virol* **58**(3): 797-804.
- Hogenhout, S.A., Redinbaugh, M.G. and Ammar el, D. (2003). "Plant and animal rhabdovirus host range: a bug's view." *Trends Microbiol* **11**(6): 264-71.
- 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**(6): 3078-83.
- 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**(8): 4901-8.
- Horne, R.W. and Wildy, P. (1963). "Virus Structure Revealed by Negative Staining." *Adv Virus Res* **10**: 101-70.
- Irie, T., Licata, J.M., Jayakar, H.R., Whitt, M.A., Bell, P. and Harty, R.N. (2004). "Functional analysis of late-budding domain activity associated with the PSAP motif within the vesicular stomatitis virus M protein." *J Virol* **78**(14): 7823-7.
- 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): 2909-19.
- Iseni, F., Baudin, F., Blondel, D. and Ruigrok, R.W. (2000). "Structure of the RNA inside the vesicular stomatitis virus nucleocapsid." *Rna* **6**(2): 270-81.
- Iseni, F., Baudin, F., Garcin, D., Marq, J.B., Ruigrok, R.W. and Kolakofsky, D. (2002). "Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids." *Rna* **8**(8): 1056-67.

- Iverson, L.E. and Rose, J.K. (1981). "Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription." *Cell* **23**(2): 477-84.
- Jacob, Y., Badrane, H., Ceccaldi, P.E. and Tordo, N. (2000). "Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein." *J Virol* **74**(21): 10217-22.
- Jacob, Y., Real, E. and Tordo, N. (2001). "Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains." *J Virol* **75**(20): 9613-22.
- Johansson, K., Bourhis, J.M., Campanacci, V., Cambillau, C., Canard, B. and 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**(45): 44567-73.
- Johnson, K.M., Tesh, R.B. and Peralta, P.H. (1969). "Epidemiology of vesicular stomatitis virus: some new data and a hypothesis for transmission of the Indian serotype." *J Am Vet Med Assoc* **155**(12): 2133-40.
- Kaptur, P.E., McKenzie, M.O., Wertz, G.W. and Lyles, D.S. (1995). "Assembly functions of vesicular stomatitis virus matrix protein are not disrupted by mutations at major sites of phosphorylation." *Virology* **206**(2): 894-903.
- Karlin, D., Longhi, S., Receveur, V. and Canard, B. (2002). "The N-terminal domain of the phosphoprotein of Morbilliviruses belongs to the natively unfolded class of proteins." *Virology* **296**(2): 251-62.
- 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**(2): 395-407.
- Kettenberger, H. and Cramer, P. (2006). "Fluorescence detection of nucleic acids and proteins in multi-component crystals." *Acta Crystallogr D Biol Crystallogr* **62**(Pt 2): 146-50.
- 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**(2): 463-73.
- Konarev, P.V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J and Svergun, D. I. (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." *Journal of Applied Crystallography* **36**: 1277-1282
- 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): 1005-13.
- Kozak, M. (1989). "The scanning model for translation: an update." *J Cell Biol* **108**(2): 229-41.
- Kozin, M.B.a.S., D. I. (2001). "Automated matching of high- and low-resolution structural models." *Journal of Applied Crystallography* **34**: 33-41.
- Kurath, G., Ahern, K.G., Pearson, G.D. and Leong, J.C. (1985). "Molecular cloning of the six mRNA species of infectious hematopoietic necrosis virus, a fish rhabdovirus, and gene order determination by R-loop mapping." *J Virol* **53**(2): 469-76.
- 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**(3): 773-8.
- 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**(17): 5240-4.
- Kuzmin, I.V., Wu, X., Tordo, N. and Rupprecht, C.E. (2008). "Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions." *Virus Res* **136**(1-2): 81-90.
- Laggner, P., Glatter, O., Muller, K., Kratky, O., Kostner, G. and Holasek, A. (1977). "The lipid bilayer structure of the abnormal human plasma lipoprotein X. An X-ray small-angle-scattering study." *Eur J Biochem* **77**(1): 165-71.

- 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**(4): 298-307.
- Lesburg, C.A., Cable, M.B., Ferrari, E., Hong, Z., Mannarino, A.F. and Weber, P.C. (1999). "Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site." *Nat Struct Biol* **6**(10): 937-43.
- Leslie, A.G. (2006). "The integration of macromolecular diffraction data." *Acta Crystallogr D Biol Crystallogr* **62**(Pt 1): 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**(6): 857-865.
- Lichty, B.D., Power, A.T., Stojdl, D.F. and Bell, J.C. (2004). "Vesicular stomatitis virus: re-inventing the bullet." *Trends Mol Med* **10**(5): 210-6.
- Llorente, M.T., Taylor, I.A., Lopez-Vinas, E., Gomez-Puertas, P., Calder, L.J., Garcia-Barreno, B. and Melero, J.A. (2008). "Structural properties of the human respiratory syncytial virus P protein: evidence for an elongated homotetrameric molecule that is the smallest orthologue within the family of paramyxovirus polymerase cofactors." *Proteins* **72**(3): 946-58.
- 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**(20): 18638-48.
- Lundblad, J.R., Laurance, M. and Goodman, R.H. (1996). "Fluorescence polarization analysis of protein-DNA and protein-protein interactions." *Mol Endocrinol* **10**(6): 607-12.
- Luo, M., Green, T.J., Zhang, X., Tsao, J. and Qiu, S. (2007). "Conserved characteristics of the rhabdovirus nucleoprotein." *Virus Res*.
- Luo, M., Green, T.J., Zhang, X., Tsao, J. and Qiu, S. (2007). "Structural comparisons of the nucleoprotein from three negative strand RNA virus families." *Virology* **4**(1): 72.
- Malet, H., Egloff, M.P., Selisko, B., Butcher, R.E., Wright, P.J., Roberts, M., Gruez, A., Sulzenbacher, G., Vornrhein, C., Bricogne, G., Mackenzie, J.M., Khromykh, A.A., Davidson, A.D. and Canard, B. (2007). "Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5." *J Biol Chem* **282**(14): 10678-89.
- Masters, P.S. and Banerjee, A.K. (1988). "Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA." *J Virol* **62**(8): 2658-64.
- Masters, P.S. and Banerjee, A.K. (1988). "Resolution of multiple complexes of phosphoprotein NS with nucleocapsid protein N of vesicular stomatitis virus." *J Virol* **62**(8): 2651-7.
- Mavrikakis, 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**(2): 406-14.
- Mavrikakis, M., Kolesnikova, L., Schoehn, G., Becker, S. and Ruigrok, R.W. (2002). "Morphology of Marburg virus NP-RNA." *Virology* **296**(2): 300-7.
- Mavrikakis, 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**(4): 819-31.
- Mavrikakis, 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*.
- McColl, K.A., Tordo, N. and Aguilar Setien, A.A. (2000). "Bat lyssavirus infections." *Rev Sci Tech* **19**(1): 177-96.
- McCreedy, 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**(2): 902-6.

- McKinley, E.B. (1932). "A Concept of the Ultramicroscopic Virus Diseases and a Classification." *Science* **76**(1977): 449-454.
- McPherson, A. (2001). "A comparison of salts for the crystallization of macromolecules." *Protein Sci* **10**(2): 418-22.
- Mebatsion, T. (2001). "Extensive attenuation of rabies virus by simultaneously modifying the dynein light chain binding site in the P protein and replacing Arg333 in the G protein." *J Virol* **75**(23): 11496-502.
- Miyamoto, K. and Matsumoto, S. (1965). "The nature of the Negri body." *J Cell Biol* **27**(3): 677-82.
- 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**(1): 59-67.
- Myers, J.K., Pace, C.N. and Scholtz, J.M. (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." *Protein Sci* **4**(10): 2138-48.
- 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**(3): 1055-62.
- Obayashi, E., Yoshida, H., Kawai, F., Shibayama, N., Kawaguchi, A., Nagata, K., Tame, J.R. and Park, S.Y. (2008). "The structural basis for an essential subunit interaction in influenza virus RNA polymerase." *Nature* **454**(7208): 1127-31.
- Ogino, T. and Banerjee, A.K. (2007). "Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus." *Mol Cell* **25**(1): 85-97.
- Ogino, T. and Banerjee, A.K. (2008). "Formation of guanosine(5')tetraphospho(5')adenosine cap structure by an unconventional mRNA capping enzyme of vesicular stomatitis virus." *J Virol* **82**(15): 7729-34.
- Pace, C.N., Shirley, B. A., and Thompson, J. A. (1989). "Measuring the conformational stability of a protein. " Oxford, IRL Press.
- Pace, C.N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995). "How to measure and predict the molar absorption coefficient of a protein." *Protein Sci* **4**(11): 2411-23.
- Pak, C.C., Puri, A. and Blumenthal, R. (1997). "Conformational changes and fusion activity of vesicular stomatitis virus glycoprotein: [125]iodonaphthyl azide photolabeling studies in biological membranes." *Biochemistry* **36**(29): 8890-6.
- Pattnaik, A.K. and Wertz, G.W. (1990). "Replication and amplification of defective interfering particle RNAs of vesicular stomatitis virus in cells expressing viral proteins from vectors containing cloned cDNAs." *J Virol* **64**(6): 2948-57.
- Peluso, R.W. and Moyer, S.A. (1988). "Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA." *Virology* **162**(2): 369-76.
- Petersen, J.M., Her, L.S., Varvel, V., Lund, E. and Dahlberg, J.E. (2000). "The matrix protein of vesicular stomatitis virus inhibits nucleocytoplasmic transport when it is in the nucleus and associated with nuclear pore complexes." *Mol Cell Biol* **20**(22): 8590-601.
- Petoukhov, M.V., Eady, N.A., Brown, K.A. and Svergun, D.I. (2002). "Addition of missing loops and domains to protein models by x-ray solution scattering." *Biophys J* **83**(6): 3113-25.
- Petoukhov, M.V. and Svergun, D.I. (2005). "Global rigid body modeling of macromolecular complexes against small-angle scattering data." *Biophys J* **89**(2): 1237-50.
- 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**(12): 3867-74.

- 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**(Pt 11): 2691-6.
- Prehaud, C., Harris, R.D., Fulop, V., Koh, C.L., Wong, J., Flamand, A. and Bishop, D.H. (1990). "Expression, characterization, and purification of a phosphorylated rabies nucleoprotein synthesized in insect cells by baculovirus vectors." *Virology* **178**(2): 486-97.
- Prehaud, C., Nel, K. and Bishop, D.H. (1992). "Baculovirus-expressed rabies virus M1 protein is not phosphorylated: it forms multiple complexes with expressed rabies N protein." *Virology* **189**(2): 766-70.
- Qanungo, K.R., Shaji, D., Mathur, M. and Banerjee, A.K. (2004). "Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA." *Proc Natl Acad Sci U S A* **101**(16): 5952-7.
- Quiroz, E., Moreno, N., Peralta, P.H. and Tesh, R.B. (1988). "A human case of encephalitis associated with vesicular stomatitis virus (Indiana serotype) infection." *Am J Trop Med Hyg* **39**(3): 312-4.
- Raha, T., Chattopadhyay, D., Chattopadhyay, D. and Roy, S. (1999). "A phosphorylation-induced major structural change in the N-terminal domain of the P protein of Chandipura virus." *Biochemistry* **38**(7): 2110-6.
- Raha, T., Samal, E., Majumdar, A., Basak, S., Chattopadhyay, D. and Chattopadhyay, D.J. (2000). "N-terminal region of P protein of Chandipura virus is responsible for phosphorylation-mediated homodimerization." *Protein Eng* **13**(6): 437-44.
- Raux, H., Flamand, A. and Blondel, D. (2000). "Interaction of the rabies virus P protein with the LC8 dynein light chain." *J Virol* **74**(21): 10212-6.
- Ribeiro, E.A., Jr., Favier, A., Gerard, F.C., Leyrat, C., Brutscher, B., Blondel, D., Ruigrok, R.W., Blackledge, M. and Jamin, M. (2008). "Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein." *J Mol Biol* **382**(2): 525-38.
- Roche, S., Albertini, A.A., Lepault, J., Bressanelli, S. and Gaudin, Y. (2008). "Structures of vesicular stomatitis virus glycoprotein: membrane fusion revisited." *Cell Mol Life Sci* **65**(11): 1716-28.
- 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**(5784): 187-91.
- Roche, S. and Gaudin, Y. (2002). "Characterization of the equilibrium between the native and fusion-inactive conformation of rabies virus glycoprotein indicates that the fusion complex is made of several trimers." *Virology* **297**(1): 128-35.
- Roche, S., Rey, F.A., Gaudin, Y. and Bressanelli, S. (2007). "Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G." *Science* **315**(5813): 843-8.
- Rolls, M.M., Webster, P., Balba, N.H. and Rose, J.K. (1994). "Novel infectious particles generated by expression of the vesicular stomatitis virus glycoprotein from a self-replicating RNA." *Cell* **79**(3): 497-506.
- Rose, J.K., Adams, G.A. and Gallione, C.J. (1984). "The presence of cysteine in the cytoplasmic domain of the vesicular stomatitis virus glycoprotein is required for palmitate addition." *Proc Natl Acad Sci U S A* **81**(7): 2050-4.
- Rose, J.K. and Iverson, L. (1979). "Nucleotide sequences from the 3'-ends of vesicular stomatitis virus mRNA's as determined from cloned DNA." *J Virol* **32**(2): 404-11.
- 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**(10): 1219-26.
- Santoro, M.M. and Bolen, D.W. (1988). "Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants." *Biochemistry* **27**(21): 8063-8.

- Schlegel, R., Tralka, T.S., Willingham, M.C. and Pastan, I. (1983). "Inhibition of VSV binding and infectivity by phosphatidylserine: is phosphatidylserine a VSV-binding site?" *Cell* **32**(2): 639-46.
- Schmidt, M.F. and Schlesinger, M.J. (1979). "Fatty acid binding to vesicular stomatitis virus glycoprotein: a new type of post-translational modification of the viral glycoprotein." *Cell* **17**(4): 813-9.
- 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**(1): 490-8.
- 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**(2): 301-12.
- Schofield, S.A., Wolynes, P.G. and Wyatt, R.E. (1995). "Computational study of many-dimensional quantum energy flow: From action diffusion to localization." *Phys Rev Lett* **74**(19): 3720-3723.
- Scholtz, J.M. and Baldwin, R.L. (1992). "The mechanism of alpha-helix formation by peptides." *Annu Rev Biophys Biomol Struct* **21**: 95-118.
- Sellers, R.F. (1984). "Vesicular viruses", London: Edward Arnold.
- Sellers, R.F. and Maarouf, A.R. (1990). "Trajectory analysis of winds and vesicular stomatitis in North America, 1982-5." *Epidemiol Infect* **104**(2): 313-28.
- Sha, B. and Luo, M. (1997). "Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1." *Nat Struct Biol* **4**(3): 239-44.
- Spadafora, D., Canter, D.M., Jackson, R.L. and Perrault, J. (1996). "Constitutive phosphorylation of the vesicular stomatitis virus P protein modulates polymerase complex formation but is not essential for transcription or replication." *J Virol* **70**(7): 4538-48.
- Spiropoulou, C.F. and Nichol, S.T. (1993). "A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus." *J Virol* **67**(6): 3103-10.
- 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)**: 1031-42.
- Superti, F., Hauttecoeur, B., Morelec, M.J., Goldoni, P., Bizzini, B. and Tsiang, H. (1986). "Involvement of gangliosides in rabies virus infection." *J Gen Virol* **67 ( Pt 1)**: 47-56.
- Superti, F., Seganti, L., Tsiang, H. and Orsi, N. (1984). "Role of phospholipids in rhabdovirus attachment to CER cells. Brief report." *Arch Virol* **81**(3-4): 321-8.
- Suzan-Monti, M., La Scola, B. and Raoult, D. (2006). "Genomic and evolutionary aspects of Mimivirus." *Virus Res* **117**(1): 145-55.
- Svergun, D.I. (1991). "Mathematical methods in small-angle scattering data analysis." *Journal of Applied Crystallography* **24**: 485-492.
- Svergun, D.I. (1992). "Determination of the regularization parameter in indirect-transform methods using perceptual criteria." *Journal of Applied Crystallography* **25**: 495-503.
- Svergun, D.I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." *Biophys J* **76**(6): 2879-86.
- Svergun, D.I., Barberato, C. & Koch, M.H.J. (1995). "CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates." *Journal of Applied Crystallography* **28**: 768-773.
- Tanford, C. (1970). "Protein denaturation. C. Theoretical models for the mechanism of denaturation." *Adv Protein Chem* **24**: 1-95.

- Tarbouriech, N., Curran, J., Ruigrok, R.W. and Burmeister, W.P. (2000). "Tetrameric coiled coil domain of Sendai virus phosphoprotein." *Nat Struct Biol* **7**(9): 777-81.
- Tesh, R., Saidi, S., Javadian, E., Loh, P. and Nadim, A. (1977). "Isfahan virus, a new vesiculovirus infecting humans, gerbils, and sandflies in Iran." *Am J Trop Med Hyg* **26**(2): 299-306.
- Tesh, R.B., Chaniotis, B.N. and Johnson, K.M. (1972). "Vesicular stomatitis virus (Indiana serotype): transovarial transmission by phlebotomine sandflies." *Science* **175**(29): 1477-9.
- 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**(2): 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**(8): 4098-107.
- 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**(9): 7181-90.
- Tordo, N. and Kouknetzoff, A. (1993). "The rabies virus genome: an overview." *Onderstepoort J Vet Res* **60**(4): 263-9.
- Tordo, N., Poch, O., Ermine, A. and Keith, G. (1986). "Primary structure of leader RNA and nucleoprotein genes of the rabies genome: segmented homology with VSV." *Nucleic Acids Res* **14**(6): 2671-83.
- Tordo, N., Poch, O., Ermine, A., Keith, G. and Rougeon, F. (1986). "Walking along the rabies genome: is the large G-L intergenic region a remnant gene?" *Proc Natl Acad Sci U S A* **83**(11): 3914-8.
- Tordo, N., Poch, O., Ermine, A., Keith, G. and Rougeon, F. (1988). "Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses." *Virology* **165**(2): 565-76.
- Torreira, E., Schoehn, G., Fernandez, Y., Jorba, N., Ruigrok, R.W., Cusack, S., Ortin, J. and Llorca, O. (2007). "Three-dimensional model for the isolated recombinant influenza virus polymerase heterotrimer." *Nucleic Acids Res* **35**(11): 3774-83.
- Tuffereau, C., Benejean, J., Blondel, D., Kieffer, B. and Flamand, A. (1998). "Low-affinity nerve-growth factor receptor (P75NTR) can serve as a receptor for rabies virus." *Embo J* **17**(24): 7250-9.
- Tuffereau, C., Desmezieres, E., Benejean, J., Jallet, C., Flamand, A., Tordo, N. and Perrin, P. (2001). "Interaction of lyssaviruses with the low-affinity nerve-growth factor receptor p75NTR." *J Gen Virol* **82**(Pt 12): 2861-7.
- Uversky, V.N. (1993). "Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule." *Biochemistry* **32**(48): 13288-98.
- Vidal, S. and Kolakofsky, D. (1989). "Modified model for the switch from Sendai virus transcription to replication." *J Virol* **63**(5): 1951-8.
- 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**(22): 14411-20.
- Volkova, V.V.S., D. I. (2003). "Uniqueness of ab initio shape determination in small-angle scattering". *Journal of Applied Crystallography*. **36**: 860-864.
- Vulliamoz, D. and Roux, L. (2001). "'Rule of six': how does the Sendai virus RNA polymerase keep count?" *J Virol* **75**(10): 4506-18.
- 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**(7): 3501-6.

- Whelan, S.P., Barr, J.N. and Wertz, G.W. (2004). "Transcription and replication of nonsegmented negative-strand RNA viruses." *Curr Top Microbiol Immunol* **283**: 61-119.
- Whelan, S.P. and Wertz, G.W. (1999). "Regulation of RNA synthesis by the genomic termini of vesicular stomatitis virus: identification of distinct sequences essential for transcription but not replication." *J Virol* **73**(1): 297-306.
- 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**(14): 9178-83.
- Willoughby, R.E., Jr. (2007). "A cure for a rabies?" *Sci Am* **296**(4): 88-95.
- Willoughby, R.E., Jr., Tieves, K.S., Hoffman, G.M., Ghanayem, N.S., Amlie-Lefond, C.M., Schwabe, M.J., Chusid, M.J. and Rupprecht, C.E. (2005). "Survival after treatment of rabies with induction of coma." *N Engl J Med* **352**(24): 2508-14.
- Wunner, W.H., Reagan, K.J. and Koprowski, H. (1984). "Characterization of saturable binding sites for rabies virus." *J Virol* **50**(3): 691-7.
- Wyatt, P. (1993). "Light scattering and the absolute characterization of macromolecules." *Analytica Chimica Acta* **272** (1): 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**(1): 107-17.
- 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**(2): 1661-4.
- Yap, T.L., Xu, T., Chen, Y.L., Malet, H., Egloff, M.P., Canard, B., Vasudevan, S.G. and Lescar, J. (2007). "Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution." *J Virol* **81**(9): 4753-65.
- Ye, Q., Krug, R.M. and Tao, Y.J. (2006). "The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA." *Nature* **444**(7122): 1078-82.

## REFERENCES ELECTRONIQUES

Institut Pasteur	<a href="http://www.pasteur.fr/ip/index.jsp">http://www.pasteur.fr/ip/index.jsp</a>
ICTV ( <i>International Taxonomy on Viruses</i> )	<a href="http://www.ncbi.nlm.nih.gov/ICTVdb/">http://www.ncbi.nlm.nih.gov/ICTVdb/</a>
Taxonomy Pubmed	<a href="http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy">http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy</a>
Organisation Mondiale Santé (Rabnet)	<a href="http://www.who.int/GlobalAtlas/">http://www.who.int/GlobalAtlas/</a>
ILL	<a href="http://www.ill.fr/Computing/prop">http://www.ill.fr/Computing/prop</a>
CCP4	<a href="http://www.ccp4.ac.uk/index.php">http://www.ccp4.ac.uk/index.php</a>
SAS Softwares Svergun Hamburg	<a href="http://www.embl-hamburg.de/ExternalInfo/Research/Sax/">http://www.embl-hamburg.de/ExternalInfo/Research/Sax/</a>
ESRF SAXSutilities	<a href="http://www.esrf.eu/UsersAndScience/Experiments/SCMatter/ID02">http://www.esrf.eu/UsersAndScience/Experiments/SCMatter/ID02</a>
EMBL HTx Lab	<a href="https://htxlab.embl.fr/">https://htxlab.embl.fr/</a>
IMAGIC	<a href="http://www.imageScience.de">www.imageScience.de</a>
EMAN	<a href="http://blake.bcm.edu/EMAN/">http://blake.bcm.edu/EMAN/</a>
SPIDER	<a href="http://www.wadsworth.org/spider_doc/spider/docs/spider.html">http://www.wadsworth.org/spider_doc/spider/docs/spider.html</a>
PyMOL	<a href="http://pymol.sourceforge.net/">http://pymol.sourceforge.net/</a>



## **ARTICLES SUPPLEMENTAIRES**

## Cooperative Sub-Millisecond Folding Kinetics of Apomyoglobin pH 4 Intermediate

Sebastien Weisbuch, **Francine Gérard**, Marielle Padeloup, Jérémy Cappadoro,  
Yves Dupont et Marc Jamin

*Biochemistry*, 2005, 44, 7013-7023.

Ce papier fut l'objet du travail de thèse de Sébastien Weisbuch, et j'ai participé durant les stages de Licence et Master 1, que j'ai réalisé au Laboratoire de Biophysique Moléculaire (UJF, CEA Grenoble) avec Marc Jamin, à la confirmation des résultats obtenus par stopped flow.

Pour les petites protéines, la formation de la conformation native (N) à partir d'une forme complètement dépliée (U) ou à partir d'un intermédiaire de repliement (I) se produit généralement dans un processus hautement coopératif qui peut être décrit avec un modèle à deux états. Cependant, il n'est pas clair comment la coopérativité apparaît au cours de la réaction de repliement et comment les intermédiaires de repliement sont formés lors de ces processus. Ici, nous présentons que chaque étape précédemment identifiée qui conduisent l'apomyoglobine (ApoMb) à sa forme dépliée, c'est-à-dire, les réactions  $U \leftrightarrow I_a$ ,  $I_a \leftrightarrow I_b$ , et  $I_b \leftrightarrow N$ , sont des réactions à deux états. Tout d'abord, les cinétiques de repliement et dépliement de la réaction  $U \leftrightarrow I_a$  sont mesurables à pH 4,2 en utilisant l'urée (Jamin, 1996 #447) (Jamin, 1998 #446). Dans ce travail, la cinétique en sub-millisecondes mesurées par CD dans l'UV lointain, sonde de la structure secondaire des protéines, sont similaires à celles mesurées par fluorescence intrinsèque des Trp, sonde de la formation d'un cœur hydrophobe et de collapse des chaînes (Jamin #446). Ces résultats confirment que le premier état intermédiaire  $I_a$  se forme dans un processus coopératif dans lequel collapse hydrophobique et la formation de structure secondaires se passent de façon concomitante avec le cœur A(B) GH. Ensuite, lorsque le repliement de N est mesuré à pH basique, en commençant avec l'ensemble acide déplié, la formation de  $I_a$  se passe durant le temps de mélange du stopped flow en sub-millisecondes. Mais les étapes suivantes, les réactions  $I_a \leftrightarrow I_b$ , et  $I_b \leftrightarrow N$ , montrent des cinétiques similaires par spectroscopie CD dans l'UV lointain et fluorescence intrinsèque des Trp. Ceci indique que ces états tardifs lors du repliement de l'ApoMb se passent aussi dans des conditions de coopérativité, avec un modèle à deux états.

Cooperative Sub-Millisecond Folding Kinetics of Apomyoglobin pH 4 Intermediate<sup>†</sup>

Sebastien Weisbuch, Francine Gérard, Marielle Padeloup, Jérémy Cappadoro, Yves Dupont, and Marc Jamin\*

*Laboratoire de Biophysique Moléculaire et Cellulaire, Université Joseph Fourier, BMC/DRDC, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France**Received December 15, 2004; Revised Manuscript Received March 8, 2005*

**ABSTRACT:** For small single-domain proteins, formation of the native conformation (N) from a fully unfolded form (U) or from a partially folded intermediate (I) occurs typically in a highly cooperative process that can be described by a two-state model. However, it is not clear whether cooperativity arises early along the folding reaction and whether folding intermediates are also formed in highly cooperative processes. Here, we show that each previously identified step leading apomyoglobin from its unfolded form to its native form, namely, the  $U \rightleftharpoons I_a$ , the  $I_a \rightleftharpoons I_b$ , and the  $I_b \rightleftharpoons N$  reactions, exhibits typical features of a two-state reaction. First, refolding and unfolding kinetics of the earliest  $U \rightleftharpoons I_a$  reaction are measurable at pH 4.2 within the urea-induced unfolding transition [Jamin, M., and Baldwin, R. L. (1996) *Nat. Struct. Biol.* 3, 613–618; Jamin, M., and Baldwin, R. L. (1998) *J. Mol. Biol.* 276, 491–504], and we report here that sub-millisecond kinetics measured by far-UV circular dichroism (CD), a probe of secondary structure, are similar to those measured by Trp fluorescence, a probe of hydrophobic core formation and chain collapse. These results confirm that folding of the earliest intermediate,  $I_a$ , occurs in a highly cooperative process, in which hydrophobic collapse and secondary structure formation occur concomitantly in the A(B)GH core. Second, when the refolding of N is measured at high pH, starting from the acid-unfolded ensemble, the formation of  $I_a$  occurs in the mixing time of the sub-millisecond stopped-flow, but the subsequent steps, the  $I_a \rightleftharpoons I_b$  and  $I_b \rightleftharpoons N$  reactions, exhibit similar kinetics by far-UV CD and Trp fluorescence, indicating that these two late stages of the apoMb folding process also occur in highly cooperative, two-state reactions.

Many small globular proteins fold by forming partly structured intermediates in the earliest times of their folding process. The mechanisms by which these early intermediates fold from an unfolded ensemble remain poorly understood, essentially because these reactions are too fast to be seen even with the fastest mixers available. With some proteins, including apomyoglobin, similar partially folded forms are observed at equilibrium under partially destabilizing conditions. Equilibrium characterization of these stable intermediates is easier and directly informs on the properties and behaviors of the more relevant kinetic intermediates. However, equilibrium studies provide information neither about the mechanism of the reaction nor about the properties of the transition state. Gaining this information relies only on kinetic studies.

Apomyoglobin (apoMb)<sup>11</sup> is particularly interesting to investigate the role of folding intermediates and the mechanism of their formation because this protein populates equilibrium intermediates near pH 4.0 under low-salt condi-

tions that resemble kinetic intermediates observed when the protein refolds to its native conformation at pH 6.0 (2–4). In native conditions, all available data indicate that the native form N appears with a rate constant of about  $2.5 \text{ s}^{-1}$  at  $5^\circ\text{C}$  (2, 3), but at least two different intermediates have been identified that form on very different time scales along this process. A prominent folding intermediate, named here  $I_a$ , forms during the dead-time of the fastest continuous flow mixing devices ( $<200 \mu\text{s}$ ) (2–3, 5–7).  $I_a$  has molten globule structural features, and hydrogen/deuterium (H/D) exchange experiments revealed that it contains a structured core formed by the A, H, and G helices of Mb as well as by a part of the B helix (3). The central part is more loosely packed and contains fluctuating D and E helices (8). A second intermediate, named  $I_b$ , forms more slowly with a rate constant of  $35 \text{ s}^{-1}$  at  $5^\circ\text{C}$ .  $I_b$  was first identified by stopped-flow Trp fluorescence (2), but its occurrence was since confirmed by other probes (5–7). Both intermediates,  $I_a$  and  $I_b$ , coexist near pH 4.0, in an equilibrium that depends on pH, the presence or absence of denaturant, and the presence or absence of stabilizing anions (2). Detailed structural characterization of the equilibrium mixture of  $I_a$  and  $I_b$  by NMR spectroscopy and by H/D exchange experiments reveals a similar structural organization to that of the dead-time intermediate at pH 6, with a stable and compact A(B)GH core and a central part containing fluctuating native and non-native helices (4, 9–10).

Although at pH 6 the formation of  $I_a$  is too fast to be seen, it was shown recently that folding and unfolding kinetics of

<sup>†</sup> This work was supported by the Emergence program (Région Rhône-Alpes, France).

\* To whom correspondence should be addressed: Laboratoire de Biophysique Moléculaire et Cellulaire, Université Joseph Fourier, BMC/DRDC, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France. Telephone: +(32) 4 38 78 42 12. Fax: + 32 4 38 78 54 87. E-mail: Marc.Jamin@cea.fr.

<sup>1</sup> Abbreviations: apoMb, apomyoglobin; Mb, myoglobin; CD, circular dichroism; N, native form; I, intermediate form; U, unfolded form; NATA, *N*-acetyl-tryptophanamide; NBS, *N*-bromosuccinimide; Trp, tryptophan.

$I_a$  can be measured by stopped-flow Trp fluorescence within its urea-induced unfolding transition at pH 4.2 (1, 11). When monitored by Trp fluorescence, refolding and unfolding kinetics of the  $U \rightleftharpoons I_a$  reaction follow exponential time courses of increasing and decreasing fluorescence intensity, respectively, in agreement with equilibrium fluorescence values (1). Unfolding kinetics at pH 4.2, however, deviates from a two-state model because of the presence of  $I_b$  and of aggregated forms of I. At pH 4.2,  $I_b$  is less stable than  $I_a$  and converts into  $I_a$  in the presence of moderate concentrations of urea (2). The unfolding transition observed above 1 M urea is that of  $I_a$ , but unfolding of  $I_b$  in this urea range gives rise to a burst phase in the unfolding kinetics (1, 11). A second source of deviation from a strict two-state model comes from the presence of multimeric species that give rise to a second slow decreasing phase in unfolding kinetics (1, 11). Besides these complications, the data are in agreement with predictions for a cooperative two-state  $U \rightleftharpoons I_a$  reaction (12, 13). Plotting the apparent rate constant for folding and the fast apparent rate constant for unfolding as a function of urea concentration yields a V-shaped graph, typical of a two-state kinetic process. More importantly, the  $\Delta G_{H_2O}$  and  $m$  values derived from these kinetic data are similar to those obtained from the analysis of the  $U \rightleftharpoons I_a$  equilibrium transition (1, 11).

Additional evidence confirming that the  $U \rightleftharpoons I_a$  reaction is highly cooperative would be to compare the kinetics monitored by Trp fluorescence and by far-UV circular dichroism (CD) and find out whether secondary structure formation is concomitant with hydrophobic core formation and chain collapse. To perform this test, we use here a newly designed stopped-flow mixer and microcuvette assembly that operates with a dead-time of about 400  $\mu$ s and monitor the folding and unfolding kinetics of  $I_a$  at pH 4.2 both by Trp fluorescence and by far-UV CD at 225 nm. We also extend the test to the other steps of the formation of native apoMb and measure refolding kinetics of N under native conditions by far-UV CD and Trp fluorescence.

## MATERIALS AND METHODS

Urea, glucuronolactone, *N*-acetyl-tryptophanamide, *N*-bromo-succinimide, and lysozyme were purchased from Sigma. Concentration of urea solutions was determined from refractive index measurements (14). All experiments were performed in 2 mM citrate buffer containing 30 mM NaCl.

**Protein Expression and Purification.** Horse myoglobin was purchased from Sigma and apoMb was prepared by acid-acetone precipitation (15). Sperm whale apoMb was produced by heterologous expression in *Escherichia coli*. The synthetic gene for sperm whale myoglobin (16) was cloned between NdeI and BamHI sites in a pET17b expression vector (Novagen) of *E. coli*. The sequence of the cloned gene was verified by DNA sequencing. Protein was expressed in *E. coli* strain BL21 pLysS and purified according to the procedure described in ref 17. Briefly, freshly transformed *E. coli* cells were grown in LB medium. Production was induced by addition of 1 mM IPTG when the O.D. at 600 nm reached 0.8. Cells were harvested 4 h after induction and were suspended in a lysis buffer (50 mM Tris buffer, 100 mM NaCl, 10 mM EDTA, pH 8.0). Crude extract was prepared by adding lysozyme and DNase. ApoMb was

solubilized from the pellet by iterative extractions with a 0.1% trifluoroacetic solution. Fractions containing apoMb were identified by SDS PAGE and loaded on a radial-compression reversed-phase HPLC column (Delta-Pak C4, 15  $\mu$ m, Waters) equilibrated in 0.1% TFA–25% acetonitrile and operating at 12 mL/min. The protein was eluted by a 25–60% acetonitrile linear gradient over 30 min. Collected fractions containing apoMb were pooled and lyophilized.

Protein solutions were prepared in distilled water, filtered, and dialyzed extensively against distilled water before use. Protein concentration was determined by absorbance in 6.0 M GdmCl (20 mM Na phosphate, pH 6.5) as described (18), using  $\epsilon_{280 \text{ nm}} = 15\,200 \text{ M}^{-1} \text{ cm}^{-1}$  and  $\epsilon_{288 \text{ nm}} = 10\,800 \text{ M}^{-1} \text{ cm}^{-1}$ .

**Circular Dichroism Spectroscopy.** Equilibrium folding and unfolding CD experiments were performed with a Jasco J810 spectropolarimeter using a thermostated 1 cm quartz cell.

**Stopped-Flow Measurements.** Stopped-flow measurements were made by using a SFM-4 stopped-flow instrument (Bio-Logic, Claix, France). Kinetics were monitored simultaneously by Trp fluorescence and by far-UV CD with the same apparatus and the same observation chamber and settings. When both traces were measured simultaneously, a single wavelength, set at 225 nm, was used for both measuring the CD signal and for exciting Trp fluorescence. Fluorescence emission was detected at 90°, using an optical cutoff filter (50% transmittance at 305 nm). An excitation of Trp fluorescence at 288 nm yielded similar kinetics. The path length of the observation chamber was 0.8 mm. For CD measurements, a quartz blade was used to allow for corrections of the signal due to constraints on the quartz walls of the observation cell. Control experiments were performed with glucuronolactone to calibrate the CD signal and with lysozyme. Refolding of lysozyme exhibited typical measurable biphasic kinetics with an overshoot of CD signal corresponding to the formation of an intermediate.

Mixing efficiency and dead-time were measured by using the bimolecular quenching of *N*-acetyl-tryptophanamide (NATA) fluorescence by *N*-bromo-succinamide (NBS) (19). The reaction was measured in 10 mM sodium phosphate pH 7 at 20 °C. Final NATA concentration was 5  $\mu$ M, and the kinetics were measured for NBS concentrations ranging from 0 to 10 mM. Kinetic traces were fitted to a one-exponential equation with the software provided by Bio-Logic. The dead-time of the apparatus was determined from a plot of  $\ln(\Delta F_{\text{Tot}}/\Delta F_{\text{Obs}})$  versus  $k$  according to eq 1 (20), where  $t_0$  is the dead-time and  $k$  is the apparent rate constant:

$$\ln(\Delta F_{\text{Tot}}/\Delta F_{\text{Obs}}) = t_0 k \quad (1)$$

For stopped-flow CD experiments performed at pH values ranging from 6.0 to 11.0, the unfolded protein at pH 2.2 was diluted in a 1:10 ratio into the refolding buffer. Final protein concentration was 10  $\mu$ M, and the kinetic traces show averages of 15 individual measurements.

The refolding kinetics for the  $U \rightleftharpoons I_a$  reaction of sperm whale apoMb were measured at increasing temperature and analyzed with a one-exponential equation. The temperature dependence of the apparent rate constant ( $1/\tau$ ) was modeled with the Arrhenius equation:

$$\ln\left(\frac{1}{\tau}\right) = \ln A - \frac{E_a}{RT} \quad (2)$$

where  $A$  is the preexponential factor,  $E_a$  is the activation energy,  $R$  is the gas constant, and  $T$  is the absolute temperature.

*Fit to a Two-State Model.* Folding and unfolding of  $I_a$  can be schematized with a two-state model:



for which the equilibrium constant is given by

$$K = \frac{[U]}{[I_a]} = \frac{k_{21}}{k_{12}} \quad (4)$$

and the apparent rate constant is given by

$$\frac{1}{\tau} = k_{12} + k_{21} \quad (5)$$

The equilibrium constant and the rate constants for folding ( $k_{12}$ ) and unfolding ( $k_{21}$ ) are expressed as functions of urea molarity ( $[\text{urea}]$ ) using the standard relations (12–13):

$$K = K(\text{H}_2\text{O})e^{(m[\text{urea}]/RT)} \quad (6a)$$

$$k_{12} = k_{12}(\text{H}_2\text{O})e^{(m_{12}^\ddagger[\text{urea}]/RT)} \quad (6b)$$

$$k_{21} = k_{21}(\text{H}_2\text{O})e^{(m_{21}^\ddagger[\text{urea}]/RT)} \quad (6c)$$

$K(\text{H}_2\text{O})$ ,  $k_{12}(\text{H}_2\text{O})$ , and  $k_{21}(\text{H}_2\text{O})$  represent the equilibrium constant and the rate constants for folding and unfolding at 0 M urea, respectively. The equilibrium parameters  $\Delta G^\circ(\text{H}_2\text{O})$  and  $m$ , are related to the kinetics parameters  $k_{12}(\text{H}_2\text{O})$ ,  $k_{21}(\text{H}_2\text{O})$ ,  $m_{12}^\ddagger$  and  $m_{21}^\ddagger$  according to the following relations:

$$\Delta G^\circ(\text{H}_2\text{O}) = -RT \ln \frac{k_{21}(\text{H}_2\text{O})}{k_{12}(\text{H}_2\text{O})} \quad (7a)$$

$$m = m_{21}^\ddagger - m_{12}^\ddagger \quad (7b)$$

From those equations, eq 8 can be derived according to the procedure described in ref 21 to represent the equilibrium urea-induced unfolding transition monitored by CD:

$$\text{CD} = \left[ (\text{CD}_{I_a} + b_{I_a}[\text{urea}]) + \frac{(\text{CD}_U + b_U[\text{urea}]) \exp\left(-\frac{\Delta G^\circ(\text{H}_2\text{O})}{RT} + \frac{m[\text{urea}]}{RT}\right)}{\left[1 + \exp\left(-\frac{\Delta G^\circ(\text{H}_2\text{O})}{RT} + \frac{m[\text{urea}]}{RT}\right)\right]} \right] \quad (8)$$

where  $\text{CD}_U$  and  $\text{CD}_{I_a}$  are the CD values at 222 nm of U and  $I_a$  at 0 M urea, and  $b_U$  and  $b_{I_a}$  are the slopes of the linear baselines for the variation of the CD signal of U and  $I_a$  with urea concentration, respectively. Equation 9 can be derived to represent the dependence of the apparent rate constant  $1/\tau$  versus urea molarity:

$$\frac{1}{\tau} = k_{12}(\text{H}_2\text{O}) \exp\left(\frac{m_{12}^\ddagger[\text{urea}]}{RT}\right) + k_{21}(\text{H}_2\text{O}) \exp\left(\frac{m_{21}^\ddagger[\text{urea}]}{RT}\right) \quad (9)$$

The kinetic parameters were determined in a combined fit of these two data sets (Figures 3 and 4) using SIGMAPLOT (Jandel, San Rafael, CA).

*Modeling the pH Dependence of the Rate Constants.* The rate constants for folding ( $k_{12}$ ) and unfolding ( $k_{21}$ ) are expressed as functions of pH by assuming that the proton binding sites are equivalent and independent, using the following equations:

$$k_{12} = k_{12}^{\text{ref}} \frac{\prod_i [1 + 10^{(pK_{ai}^\ddagger - \text{pH})}]}{\prod_i [1 + 10^{(pK_{ai}^U - \text{pH})}] } \quad (10a)$$

$$k_{21} = k_{21}^{\text{ref}} \frac{\prod_j [1 + 10^{(pK_{aj}^\ddagger - \text{pH})}]}{\prod_j [1 + 10^{(pK_{aj}^{I_a} - \text{pH})}] } \quad (10b)$$

$$K = \frac{k_{12}^{\text{ref}}}{k_{21}^{\text{ref}}} \quad (10c)$$

where  $k_{12}^{\text{ref}}$  and  $k_{21}^{\text{ref}}$  are the reference rate constants for the  $U \rightarrow I_a$  and  $I_a \rightarrow U$  reactions at high pH, respectively,  $K$  is the folding equilibrium constant at high pH, and the  $pK_{ai}$  and  $pK_{aj}$  values refer to residues  $i$  and  $j$  in the U,  $I_a$ , and  $I^\ddagger$  forms. Simulations were performed with SIGMAPLOT (Jandel, San Rafael, CA).

## RESULTS

*A Sub-Millisecond Stopped-Flow Apparatus.* Conventional stopped-flow instruments have dead-times on the order of 1 ms in the fluorescence detection mode and on the order of 5 ms in the CD (CD) mode. Here, we use a newly designed micromixer and microcuvette with reduced volumes (Bio-Logic, Claix, France) that operates with a similar dead-time in both fluorescence and CD modes. In the fluorescence mode, optimal settings of the instrument bring the dead-time down to 250  $\mu\text{s}$ . These performances were obtained by reducing the mixer and cuvette volumes while keeping the solution flow rate as fast as possible. The cuvette has a light path of 0.8 mm and a volume of 4  $\mu\text{L}$ . The maximum flow rate attainable with the SFM-400 system is 18–20  $\text{mL s}^{-1}$ . However, transient mechanical stress due to pressure changes on the cuvette quartz wall produces a dynamic linear dichroism artifact that prevents measurements in CD mode at such high flow rate values. By using glucuronolactone and measuring the CD signal at 225 nm, we found that a maximum flow rate of 10–12  $\text{mL s}^{-1}$  is usable in CD experiments. Static linear dichroism was also observed that originates from the vertical strain exerted by the cuvette mount. This linear dichroism signal was compensated by inserting a quartz blade before the observation cuvette that was set under an adjustable static strain oriented at a right angle from that exerted by the flow. The strain on the blade

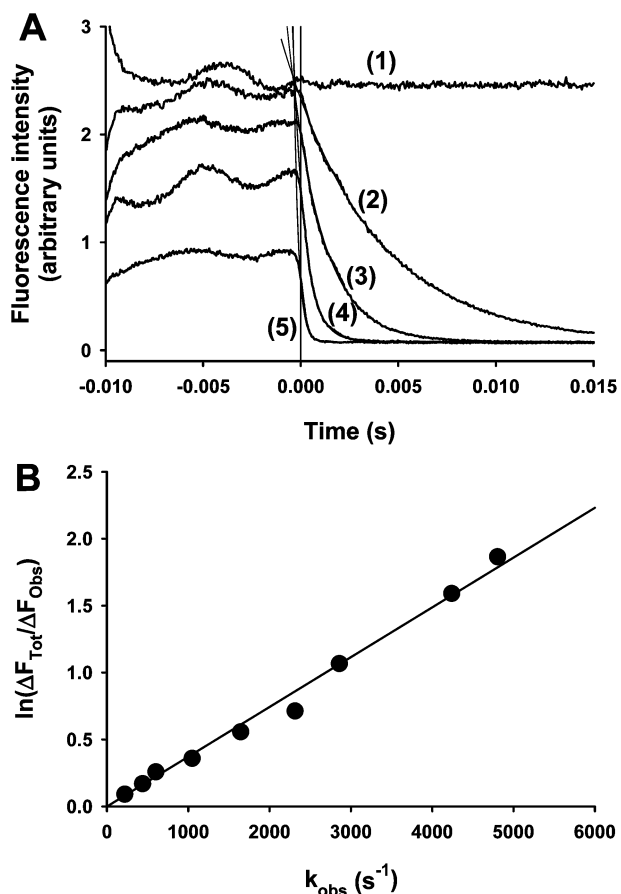


FIGURE 1: Tests of the sub-millisecond stopped-flow. (A) Bimolecular test reaction. Kinetics of the reaction between NATA at a final concentration of 5  $\mu\text{M}$  and NBS at varying concentrations: 0 mM (curve 1), 0.3 mM (curve 2), 0.9 mM (curve 3), 2.4 mM (curve 4), and 6.0 mM (curve 5). Conditions: NATA and NBS are mixed in a 1:1 ratio in 10 mM phosphate buffer pH 7, 20 °C. The excitation was at 295 nm, and a cutoff filter at 320 nm was used for the emission. Curves 2 and 3 are fitted to a one-exponential equation. (B) Plot of  $\ln(\Delta F_{\text{Tot}}/\Delta F_{\text{Obs}})$  versus  $k$ .  $\Delta F_{\text{Tot}}$  is the difference at time 0 between curve 1 and the final fluorescence measured. The slope of this graph provides a dead-time value of  $370 \pm 15 \mu\text{s}$ .

was adjusted by measuring the static CD value of a calibrated solution of glucuronolactone at 225 nm.

The efficiency of mixing and the dead-time of the apparatus in the settings used in CD mode are estimated by using a fluorescence quenching reaction between NATA and NBS (14). Mixing efficiency is estimated by monitoring fluorescence intensity when the liquid flows through the observation cell before the stop (Figure 1A). In our experiment, the optical signal was recorded during the flow for 10 ms before the stop, and time zero refers to the time at which liquid flow stops. Accordingly, the flowing period preceding the stop appears as negative time. Figure 1A shows that, during this time, fluorescence emission increases and reaches a plateau about 5 ms before the stop, indicating that mixing is complete. The kinetic traces from time zero are well fitted by a single-exponential equation and extrapolate to the initial fluorescence about 400  $\mu\text{s}$  before time 0. A plot of the observed rate constant as a function of NBS concentration is linear, and its slope yields a value of  $670\,000 \pm 11\,000 \text{ M}^{-1} \text{ s}^{-1}$  for the second-order rate constant of the reaction. The dead-time is determined from a plot of  $\ln(\Delta F_{\text{Tot}}/\Delta F_{\text{Obs}})$  versus  $k$  (Figure 1B), where  $\Delta F_{\text{Tot}}$ , the total fluorescence change, is calculated for each curve as the difference between

the fluorescence intensity of NATA and the fluorescence intensity at the end of the kinetics, and  $\Delta F_{\text{Obs}}$  is the fluorescence amplitude of each kinetic curve measured from time 0 (20). The plot is linear, and its slope directly yields a dead-time of  $370 \pm 15 \mu\text{s}$  for the mixer. This dead-time is similar to that calculated by dividing the volume that separates the mixing chamber from the observation chamber by the flow rate. Whereas in these conditions this mixer only provides a two-fold improvement of the dead-time in fluorescence mode, as compared with classical mixer and cuvette, it provides a 10-fold improvement in CD mode since CD kinetics can be measured with the same mixing chamber and settings and thus with the same dead-time.

*Folding and Unfolding Kinetics of  $I_a$  at pH 4.2.* The refolding and unfolding kinetics of apoMb  $I_a$  are measured here within its urea-induced unfolding transition (above 1 M urea) both by Trp fluorescence and by far-UV CD at 225 nm. In the native protein, Trp fluorescence is usually considered as a probe of tertiary structure, but strictly it reports on the burial of Trp residues in a nonpolar environment. Upon formation of a folding intermediate from an unfolded ensemble, Trp fluorescence more likely reports on partial or global collapse of the polypeptide chain, leading to the burial of Trp side chains inside a hydrophobic core. Both horse and sperm whale apoMb have two Trp residues located in A helix, and both Trp residues become buried, at least partially, in the A(B)GH core in the folding intermediate. The changes in Trp fluorescence upon formation of  $I_a$  are thus used here as a probe of the collapse of this core, whereas changes in far-UV CD are used as a probe of secondary structure formation.

Figure 2 shows typical folding and unfolding kinetics for horse and sperm whale apomyoglobin. At equilibrium,  $I_a$  has a higher fluorescence intensity and a lower ellipticity at 225 nm than U. Accordingly, in refolding kinetics, fluorescence intensity increases and molar ellipticity at 225 nm decreases in single-exponential time courses (Figure 2A,B). No significant change occurs during the mixing time, and up to 90% of the reaction is observed. At the same final urea concentration, similar values of the observed rate constant are found for kinetics monitored either by Trp fluorescence or by far-UV CD. In unfolding kinetics, fluorescence intensity decreases and molar ellipticity at 225 nm increases (Figure 2C–F). The same complications are observed here as previously (1, 11). The burst phase in Trp fluorescence, assigned to the conversion of  $I_b$  into  $I_a$ , is clearly visible in Figure 2C,E. By far-UV CD, the signal-to-noise ratio is too small to reveal the presence of a corresponding burst phase (Figure 2D,F). If, however, it occurs during the mixing time, the  $I_b \rightleftharpoons I_a$  reaction involves only a small variation in CD signal, in agreement with other observations (5). With both probes two kinetic phases are measurable. The fast phase observed by Trp fluorescence was previously assigned to the  $U \rightleftharpoons I_a$  reaction, and the slow phase, which is detected as a baseline change on the time scale shown in Figure 2, was assigned to the presence of aggregated species in the initial conditions (1). For the fast measurable phase, similar rate constants are obtained with both probes.

At pH 4.2, horse apoMb is less stable than sperm whale apoMb, and it unfolds at lower urea concentrations as shown when equilibrium urea-induced unfolding transitions are monitored by far-UV CD (Figure 3 and Table 1). As

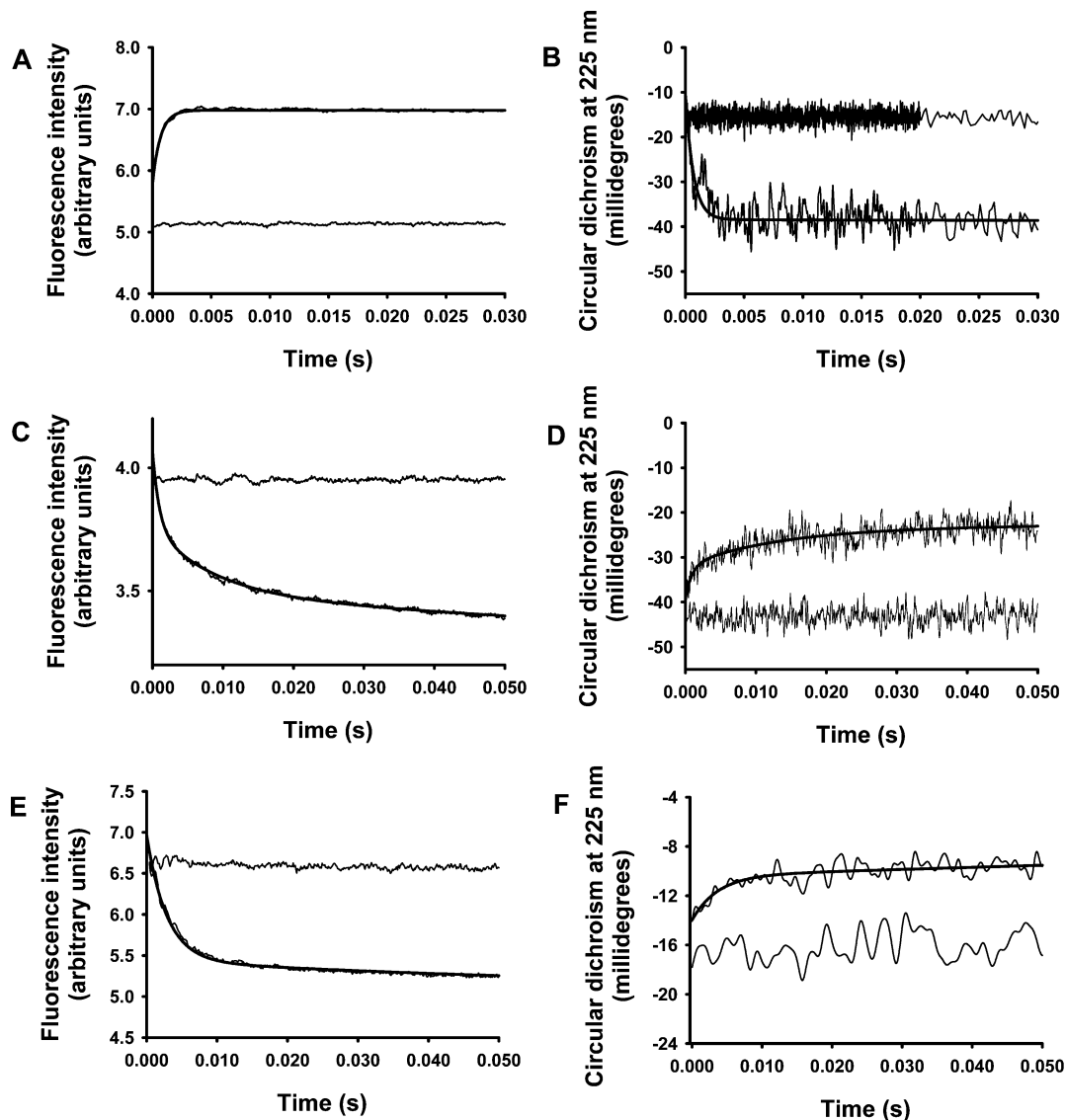


FIGURE 2: Urea-induced refolding (A, B) and unfolding (C–F) kinetics of the pH 4 intermediate measured by Trp fluorescence (A, C, E) and by far-UV CD (B, D, F) for horse (A–D) and sperm whale (E–F) apomyoglobin. The starting material for refolding experiments is U at pH 4.2 in 4 M urea and for unfolding it is a mixture of  $I_a$  and  $I_b$  at pH 4.2 in 0 M urea. The final urea concentration is 1.1 M (A, B) or 1.9 M urea (C–F). Stopped-flow fluorescence was monitored with an excitation at 225 nm and a cutoff filter at 305 nm for the emission. Stopped-flow CD was monitored at 225 nm. Fluorescence kinetics show averages of three shots, whereas CD kinetics show averages of about 20 shots. Kinetics are fitted to a one-exponential equation for refolding and to a two-exponential equation for unfolding. Conditions: 2 mM Na citrate, 30 mM NaCl, pH 4.2, 5 °C; final protein concentration is 50  $\mu$ M (A, B), 100  $\mu$ M (C, D), or 24  $\mu$ M (E, F).

previously observed (1), final values in the kinetic experiments monitored by Trp fluorescence and far-UV CD reproduce the equilibrium unfolding transitions (data not shown). For both horse (Figure 4A) and sperm whale (Figure 4B) apoMb, a single V-shaped dependence is observed when the apparent rate constant for the  $U \rightleftharpoons I_a$  reaction is plotted as a function of urea concentration. For sperm whale apoMb (Figure 4B), the data obtained in this study are in agreement with previous kinetic data obtained by measuring Trp fluorescence with a stopped-flow apparatus operating with a dead-time of 1 ms (small circles in Figure 4B show data taken from ref 1). Both equilibrium and kinetic data are modeled by assuming a two-state reaction (see Materials and Methods). The rate constants for folding ( $k_{12}$ ) and unfolding ( $k_{21}$ ) and their dependences on urea concentration ( $m_{12}^\ddagger$  and  $m_{21}^\ddagger$  values) have been obtained in a combined fitting procedure of the equilibrium constant ( $K$ ) and of the apparent rate constant ( $1/\tau$ ) (Table 1). According to a two-state model,

the midpoint of the equilibrium transition occurs at the urea concentration ( $C_m$ ) given by the following equation:

$$C_m = -RT \frac{\ln\left(\frac{k_{21}(\text{H}_2\text{O})}{k_{12}(\text{H}_2\text{O})}\right)}{m_{21}^\ddagger - m_{12}^\ddagger} \quad (11)$$

whereas the lowest value of  $1/\tau$  is found at a urea concentration ( $C_{(1/\tau)\min}$ ), where the first-order derivative with respect to urea concentration of the equation giving the urea-dependence of the apparent rate constant (eq 9 in Materials and Methods) is equal to zero:

$$C_{(1/\tau)\min} = -RT \frac{\ln\left(-\frac{k_{21}(\text{H}_2\text{O})m_{21}^\ddagger}{k_{12}(\text{H}_2\text{O})m_{12}^\ddagger}\right)}{m_{21}^\ddagger - m_{12}^\ddagger} \quad (2)$$

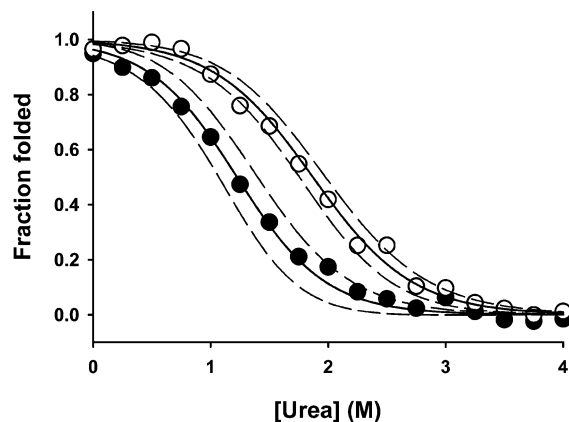


FIGURE 3: Equilibrium urea-induced unfolding transition for horse (●) and sperm whale (○). Urea-induced unfolding transitions were monitored by far-UV CD at 222 nm. The full lines are drawn using eq 8, and the parameters shown in Table 1. The dashed lines show the 95% confidence intervals for the combined fit of equilibrium and kinetic data. Conditions: 2 mM Na citrate, 30 mM NaCl, pH 4.2, 5 °C; final protein concentration is between 1 and 5  $\mu$ M.

Table 1: Parameters of the Urea-Induced Unfolding Transition at pH 4.2

parameters	sperm whale	horse
$k_{12(\text{H}_2\text{O})}$ ( $\text{s}^{-1}$ )	$3980 \pm 800$	$6000 \pm 1,500$
$k_{21(\text{H}_2\text{O})}$ ( $\text{s}^{-1}$ )	$47 \pm 8$	$230 \pm 50$
$m_{12}$ (cal/mol/M)	$-900 \pm 75$	$-1050 \pm 150$
$m_{21}$ (cal/mol/M)	$410 \pm 30$	$420 \pm 60$
$C_{(1/\tau)_{\text{min}}}$ (M)	$2.2 \pm 0.2$	$1.6 \pm 0.2$
$\Delta G_{(\text{H}_2\text{O})}$ (cal/mol)	$2450 \pm 140$	$1800 \pm 200$
$m$ (cal/mol/M)	$1310 \pm 80$	$1470 \pm 160$
$C_m$ (M)	$1.9 \pm 0.2$	$1.2 \pm 0.2$

If the absolute value of  $m_{12}^{\ddagger}$  is different from that of  $m_{21}^{\ddagger}$ , the value of  $C_m$  is close to that of  $C_{(1/\tau)_{\text{min}}}$  but not equal. According to the difference in equilibrium stability between the two proteins, the midpoint of the V-shaped plot ( $C_{(1/\tau)_{\text{min}}}$ ) is found at a lower urea concentration for horse apoMb than for sperm whale apoMb, but because the value of  $m_{12}^{\ddagger}$  is larger than that of  $m_{21}^{\ddagger}$ ,  $C_{(1/\tau)_{\text{min}}}$  values are larger than  $C_m$  values (Table 1). The lower stability of horse apoMb is mainly due to an increase of the unfolding rate.

**Protein Concentration Dependence of the Folding Kinetics of  $I_a$ .** Refolding kinetics for sperm whale apoMb at pH 4.2 were measured by stopped-flow Trp fluorescence at varying final protein concentrations, ranging from 1 to 8.8  $\mu$ M. A solution of unfolded apoMb in 4 M urea was diluted 3 times in a refolding buffer at a final urea concentration of 1.9 M. Each kinetic curve could be fitted to a one-exponential equation. The amplitude of the kinetic curves exhibits a linear dependence on protein concentration (Figure 5A). The rate constant is similar for each protein concentration (Figure 5B) and matches that measured, at the same final urea concentration, at higher protein concentrations (Figure 2), showing that, in these conditions, the folding kinetics of the pH 4 intermediate of apoMb are not complicated by an aggregation process.

**Temperature Dependence of the Folding Kinetics of  $I_a$ .** Refolding kinetics for sperm whale apoMb at pH 4.2 were measured by stopped-flow Trp fluorescence at two different final urea concentrations and at various temperatures, ranging from 5 to 15 °C. Folding kinetics at the same final urea concentrations had been obtained previously at 20 °C by

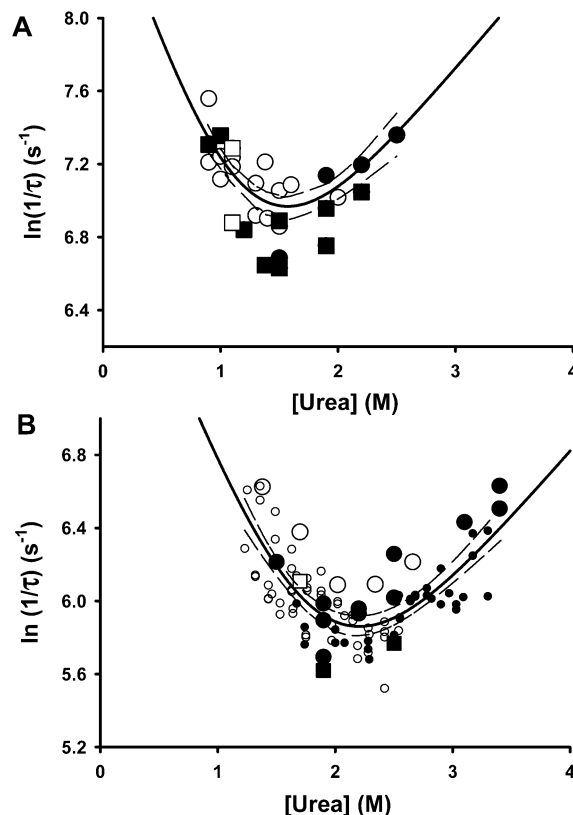


FIGURE 4: Apparent rate constants ( $1/\tau$ ) versus [urea] for refolding from 4 M urea and for unfolding from 0 M urea. (A) Horse apomyoglobin. (B) Sperm whale apomyoglobin. Open symbols are for refolding kinetics, and closed symbols are for unfolding kinetics. Circles are for kinetics measured by Trp fluorescence, and squares are for kinetics measured by far-UV CD. Small open and closed circles in panel B show data taken from ref 1 that were measured by stopped-flow fluorescence in identical conditions. The full lines are drawn using eq 9 and the parameters shown in Table 1, and the dashed lines show the 95% confidence interval for the combined fit of equilibrium and kinetic data.

continuous-flow Trp fluorescence (11). Figure 6 shows the Arrhenius plots of these rate constants. Activation energies of  $23.6 \pm 0.4$  and  $22.3 \pm 1.3$  kcal/mol are obtained at final urea concentrations of 1.5 and 1.8 M, respectively, indicating that the  $U \rightleftharpoons I_a$  reaction involves a high energy barrier.

**pH Dependence of the Folding Kinetics of  $I_a$ .** A kinetic phase observed by monitoring Trp fluorescence in pH-jump experiments, starting from U at pH 2.0 and ending at final pH values below 3.5, has been previously assigned to the  $U \rightleftharpoons I_a$  reaction (11). The rate constant of this kinetic phase increases with pH and becomes too fast to be measured above pH 3.5. At pH 4.2, an estimate of the rate for the  $U \rightleftharpoons I_a$  reaction has been obtained by extrapolating to zero urea concentration, the rate constants measured within the urea-induced unfolding transition (1).

A simple model can be used to predict the pH dependence of the apparent rate constant for the  $U \rightleftharpoons I_a$  reaction. This model is adapted from the Monod-Wyman-Changeux model (MWC model) (22) to describe the effect of preferential binding of  $n$  protons on folding and unfolding kinetics. This model has been applied previously to simulate the pH dependence of kinetics for the  $I \rightleftharpoons N$  reaction of apomyoglobin at higher pH values (23). For kinetics induced by pH change, the preferential binding of protons results in lowered  $pK_a$  values of the contributing residues in both  $I_a$  and in the



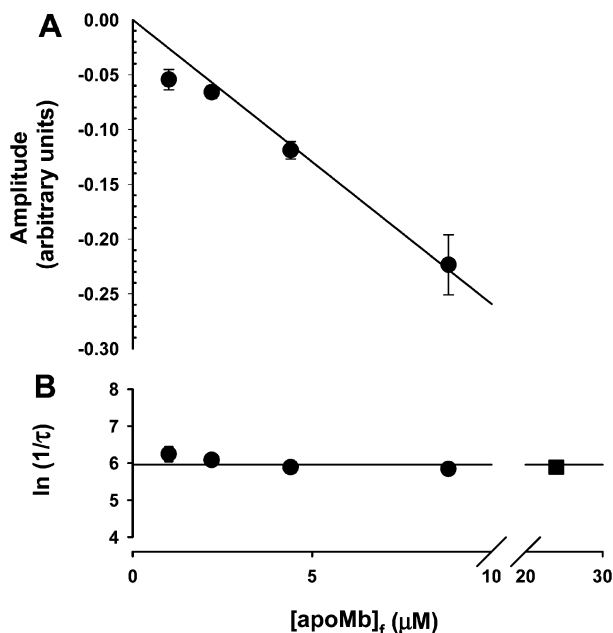


FIGURE 5: Protein concentration dependence of the folding kinetics of  $I_a$ . Refolding kinetics were measured by Trp fluorescence following a urea-jump from 4 M urea to 1.9 M urea. The kinetics were fitted with a one-exponential equation. Error bars represent the standard deviation. (A) Plot of the amplitude versus final protein concentration. (B) Plot of the apparent rate constant versus final protein concentration. The filled square shows the observed rate at 1.9 M urea obtained by measuring Trp fluorescence in the conditions used in Figure 2 (excitation wavelength = 225 nm). Conditions: 2 mM Na citrate, 30 mM NaCl, pH 4.2, 5 °C; mixing ratio 1:3; excitation wavelength was set at 288 nm.

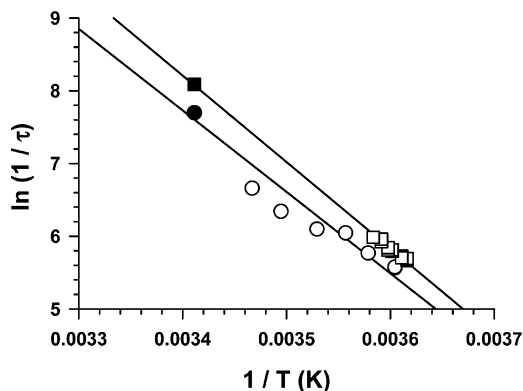


FIGURE 6: Arrhenius plots for the apparent rate constant of the  $U \rightleftharpoons I_a$  reaction for sperm whale apoMb. Refolding kinetics were measured at two different final urea concentrations and at different temperatures by stopped-flow fluorescence: (○) [urea] = 1.5 M, (□) [urea] = 1.8 M. Kinetics were fitted to a one-exponential equation, and the plot shows the logarithm of the apparent rate constant ( $1/\tau$ ) as a function  $1/T$ . The closed symbols shows the apparent rate constant for unfolding obtained at 20 °C at the same final urea concentrations, using a continuous-flow fluorescence apparatus (data taken from ref 11). The lines are drawn using eq 2.

transition of the  $U \rightleftharpoons I_a$  reaction. The model relies on the assumptions that all groups responsible for the transition are equivalent and independent and thus do not vary with titration of other ionizable groups.

The A(B)GH core of apoMb contains 13 acidic groups. If we assume that these groups have model compound  $pK_a$  in U (a  $pK_a = 4.1$  is taken as an average between Glu and Asp  $pK_a$  values), a lower  $pK_a$  in  $I_a$  ( $pK_a = 3.6$ ) and an intermediate  $pK_a$  in the transition state, closer to that of  $I_a$  than to that of

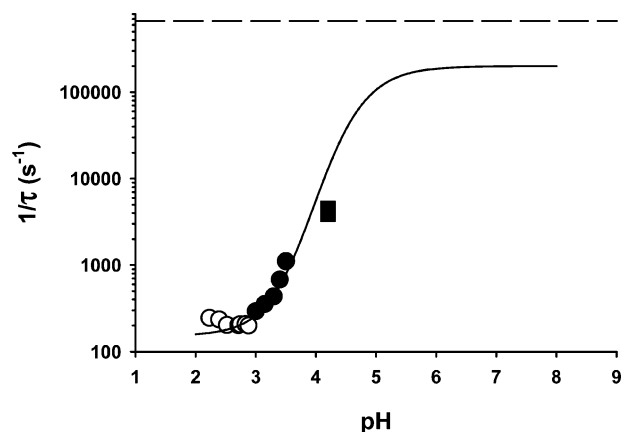


FIGURE 7: pH dependence of the time constant for the  $U \rightleftharpoons I_a$  reaction. Folding and unfolding kinetics for sperm whale apoMb at 4.5 °C were induced by pH-jump, starting from pH 2.2 for refolding or from pH 6.0 for unfolding. The symbols show the apparent rate constant for folding (●) and for unfolding (○). The filled squares (■) show the apparent rate for folding at pH 4.2 obtained by extrapolating data within the urea-induced unfolding transition for both sperm whale and horse apoMb. The line shows a prediction of the apparent rate constant for the  $U \rightleftharpoons I_a$  reaction, calculated using a simple model (eq. where acidic groups in the A(B)GH core have different  $pK_a$  values in U ( $pK_a = 4.1$ ),  $I_a$  ( $pK_a = 3.6$ ) and in the transition state separating U from  $I_a$  ( $pK_a = 3.8$ ) and using  $k_{12}^{ref} = 200\,000\text{ s}^{-1}$ ,  $k_{21}^{ref} = 20\text{ s}^{-1}$ ,  $i = 11$  and  $j = 2$ . The dashed line shows the fastest rate predicted for collapse of apoMb (53).

U ( $pK_a = 3.8$ ), then the model predicts a large increase of the apparent rate constant with pH in fair agreement with available data (Figure 7). Although the choice of  $pK_a$  values is arbitrary, this simulation provides a rationale for the variation of the rate constant with pH. It should be noted also that the Linderstrøm-Lang smeared charge model has also been used for simulating equilibrium acid-induced unfolding transition (24) and that similar pH dependences are obtained with this model.

**Folding of N under Native Conditions.** Refolding kinetics of N, starting from U at pH 2.2, were measured here at various final pH values ranging from pH 6.0 to pH 11.0, both by Trp fluorescence and by far-UV CD at 225 nm. In this pH range, kinetics monitored by Trp fluorescence are similar. They clearly show a burst phase followed three measurable exponential phases, corresponding to the previously assigned  $I_a \rightleftharpoons I_b$ ,  $I_b \rightleftharpoons N$ , and  $I_{agg} \rightleftharpoons N$  reactions (2). The rate constants for each measurable phase do not change significantly (data not shown), but the ratio of the amplitude for the fast ( $I_a \rightleftharpoons I_b$  reaction) over that for the middle phase ( $I_b \rightleftharpoons N$  reaction) changes from 0.7 at pH 6.0 to 1.4 at pH 11.0, indicating that  $I_b$  is more populated at pH 11.0 than at pH 6.0.

By far-UV CD, a major kinetic phase is detected whose rate constant matches exactly that measured by Trp fluorescence for the  $I_b \rightleftharpoons N$  reaction. At pH 6.0, the  $I_a \rightleftharpoons I_b$  reaction is not clearly detectable by far-UV CD but, when refolding kinetics are monitored at pH 11.0, a fast phase corresponding to the  $I_a \rightleftharpoons I_b$  reaction seen by fluorescence is also apparent in the kinetics monitored by far-UV CD (Figure 8A). The kinetic trace by far-UV CD is better fitted with a two-exponential equation than with a one-exponential equation (Figure 8B–D), and the rate constants for the two phases match those measured by Trp fluorescence. These results

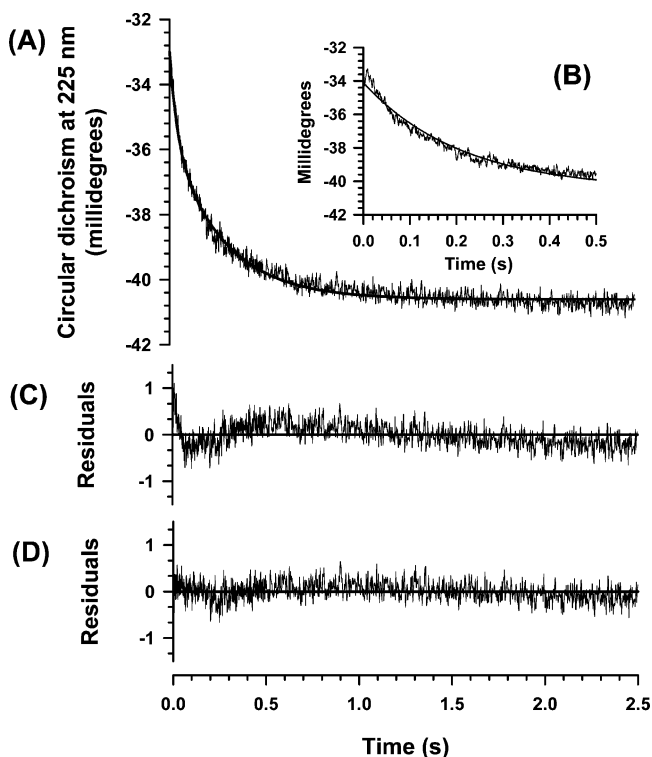


FIGURE 8: Refolding kinetics at pH 11 measured by far-UV CD. (A) Refolding kinetics following a pH jump from pH 2.2 to pH 11.0. The fit to a two-exponential equation is shown as a line and yields the following parameters:  $A_1 = 2.60 \pm 0.06$  millidegrees,  $A_2 = 5.05 \pm 0.03$  millidegrees,  $k_1 = 25.0 \pm 1 \text{ s}^{-1}$ ,  $k_2 = 3.55 \pm 0.04 \text{ s}^{-1}$  and  $\text{CD}_{\text{final}} = -40.60 \pm 0.03$  millidegrees. (B) The insert shows the same data fitted to a one-exponential equation. (C) Residuals for fitting to a one-exponential equation. (D) Residuals for fitting to a two-exponential equation. Conditions: 2 mM Na citrate, 30 mM NaCl, 5 °C; final protein concentration is 10  $\mu\text{M}$ .

suggest that the  $I_a \rightleftharpoons I_b$  reaction also gives rise to a change in CD signal, but one that is more difficult to detect because of the lower signal-to-noise ratio in CD than in the fluorescence signal. The slowest phase corresponding to the formation of N from an aggregated fraction of I is not resolved in the kinetics monitored by far-UV CD.

## DISCUSSION

**Folding Pathway of apoMb.** The folding process of apoMb has been extensively characterized in both equilibrium and kinetic studies (25–26). Similar intermediates are found both as stable species at equilibrium near pH 4 and as transient species during refolding kinetics at pH 6 (3–4). Both forms are compact, with a radius of gyration of the order of 23 Å (5, 27). They contain a hydrophobic core in which both Trp residues are buried, and they contain between 45 and 60% of the helical Mb structure (3–5), but they lack the tight packing of the side chains (7, 28). However, both equilibrium and kinetic intermediates are heterogeneous in nature and consist of ensembles of conformations. In addition, two different intermediate forms,  $I_a$  and  $I_b$ , separated by a sizable kinetic barrier, have been identified by Trp fluorescence (2). The existence of a second intermediate was already suggested in the original pulse labeling H/D exchange experiments in which residues in the B helix were found to acquire protection against H/D exchange more slowly than those forming the AGH core but more rapidly than those forming the rest of the protein (3). These pulse labeling experiments

were performed in the presence of 0.8 M urea, and  $I_b$  is sensitive to the presence of urea. In the presence of 0.8 M urea, the relative amplitude of the  $I_a \rightleftharpoons I_b$  reaction in refolding kinetics monitored by Trp fluorescence at pH 6.0 is reduced as compared to that measured in the absence of denaturant (data not shown). More recent H/D exchange experiments confirm the heterogeneous nature of the burst phase intermediate by revealing the presence of labile secondary structures in B helix, at the junction between D and E helices, and inside E helix (8). The occurrence of the  $I_a \rightleftharpoons I_b$  reaction in the refolding process of N has also been detected by different probes with both horse and sperm whale apoMb (5–7). A kinetic phase with a rate constant similar to that measured for the  $I_a \rightleftharpoons I_b$  reaction by Trp fluorescence was measured when folding of native horse apoMb was monitored with a sub-millisecond continuous-flow apparatus by far-UV CD (5) and by resonance Raman spectroscopy (6) in pH-jump experiments, starting from U at pH 2. Here, we show that, with sperm whale apoMb, the  $I_a \rightleftharpoons I_b$  reaction is not resolved by far-UV CD at pH 6.0 but is resolved at pH 11.0, certainly because  $I_b$  is more populated at pH 11.0 than at pH 6.0. In native conditions, the  $I_a \rightleftharpoons I_b$  reaction was not detected in kinetics monitored by SAXS (5), but pulsed thiol-disulfide exchange experiments, carried out with sperm whale apoMb variants, indicate that side chain packing at the BG interface occurs on the same time scale (7). These data suggest that formation of  $I_b$  involves no major change in the compaction of the chain but rather corresponds to the strengthening of the helical structure already present in  $I_a$  and to the stabilization of the A(B)GH core. This hypothesis is supported by the observations that some of the marginally stable helices present in the intermediate are recruited in the structured core upon single mutation (29, 30) or by addition of a stabilizing anion (28).

Two major arguments point toward a linear  $U \rightleftharpoons I_a \rightleftharpoons I_b \rightleftharpoons N$  model where  $I_a$  and  $I_b$  are compulsory on-pathway intermediates. First, all observed intermediate forms of apoMb contain essentially nativelike structure (3–10) and can be arranged in a series where structure and compaction are progressively acquired (17). Second, two kinetic studies provide supporting evidence for the linear model and exclude simple models either where N is formed by parallel fast and slow pathways or where I is a dead-end complex (2, 31).

ApoMb from two species are commonly used in folding studies, those from sperm whale and horse. They have usually been used in different conditions, and no formal comparison of their folding process has been done. Here, we found that both proteins exhibit similar folding kinetics at pH 4.2, even if faster unfolding is measured for horse apoMb than for sperm whale apoMb, suggesting that the folding intermediate is not as stable in horse as in sperm whale apoMb.

**Cooperativity in apoMb Folding Process.** Cooperativity in proteins arises because stabilizing interactions cooperate to stabilize the structure and provide an overall stabilizing effect larger than the sum of each individual contribution (32). Practically, a highly cooperative folding behavior means that no intermediate is detected during a folding/unfolding transition, and thus this transition can be described by a simple two-state model in which the initial and final states are separated by a unique free energy barrier. Native

conformations bear hallmarks of highly cooperative structures and are separated from the denatured ensemble by a large kinetic barrier. However, many proteins fold through folding intermediates, which are distinct from the unfolded ensemble and more strictly consist of ensembles of collapsed and fluctuating conformations. Whether the formation of these intermediates from the unfolded ensemble is highly cooperative or not is still extensively debated.

Different tests have been applied for determining whether the formation of apoMb pH 4 intermediate from the unfolded ensemble is highly cooperative. A major test for cooperativity is the calorimetric test: in a two-state process, the enthalpy change measured by calorimetry should be equivalent to the effective van't Hoff enthalpy change calculated from the equilibrium thermal unfolding transition (12, 33). Although the apoMb pH 4 intermediate exhibits a broad but measurable thermal unfolding transition, no significant enthalpy change is detected either by differential scanning calorimetry (34) or by isothermal titration calorimetry at 20 °C upon acid unfolding (35). The failure of the calorimetric test can result from the presence of more than two species in the mixture (36), and, at pH 4.2, we found that apoMb exists in at least two different forms,  $I_a$  and  $I_b$  (see above). At pH 2.2, where apoMb forms a similar intermediate in the presence of anions, a small but measurable excess of heat is released upon thermal unfolding (37).

Conversely, several other tests, including the kinetic experiments described here, support a highly cooperative mechanism for the formation of  $I_a$ . First, urea-induced equilibrium unfolding transitions measured at pH 4.2 by far-UV CD and by Trp fluorescence can be superimposed (1, 38). Second, the introduction by site-directed mutagenesis of a glycine or of a proline residue in the middle of A or G helix affects the overall stability and the structure of the intermediate (39), as expected for a highly cooperative system. Third, a kinetic test, that is more sensitive to the presence of intermediates than equilibrium measurements (13) has confirmed the two-state behavior of the  $U \rightleftharpoons I_a$  reaction above 1 M urea (1, 11).

Data reported here show that the compaction of the A(B)-GH core and the formation of secondary structure occur concomitantly, in single-exponential kinetics with similar apparent rate constants. In its urea-unfolded form, apoMb is expanded and devoid of almost any secondary structure, and thus formation of  $I_a$  at pH 4.2 from the urea-unfolded ensemble requires the collapse of the polypeptide chain, the formation of  $\alpha$ -helices, and the packing of these helices into the A(B)GH core. The large temperature dependence of the apparent rate for refolding of  $I_a$  supports the presence of a large free-energy barrier separating U and  $I_a$  and the two-state character of the reaction. If single-exponential kinetics can be observed with multiple-state systems (40, 41), the existence of a large free-energy barrier and the observation of similar kinetics by far-UV CD and by Trp fluorescence are more difficult to explain with a gradual model than with a two-state model. It is not clear however whether the signal change in Trp fluorescence reports on the global collapse of the polypeptide chain into a compact hydrophobic core, or if it reports only on the compaction of a previously collapsed chain, restricting the accessibility of the hydrophobic interior to solvent molecules. Several studies show that a rapid collapse of the chain could precede the rate-limiting step

leading to the formation of the first stable intermediate (42–44). On another hand, different results suggest that protein collapse can be highly cooperative, exhibiting features of a barrier-limited process, such as single-exponential kinetics and large Arrhenius temperature dependences (45). Using a laser T-jump apparatus, Hagen and Eaton monitored the collapse of cytochrome C and found that these kinetics display typical features of a two-state process (45). For this reaction, they measured activation energy of the order of 28 kJ mol<sup>-1</sup>, the same order of magnitude as for the activation energy measured here for the formation of  $I_a$ . With lysozyme, a prominent folding intermediate is formed during the dead-time of stopped-flow mixers. Application of the superposition test to curves where burst phase changes measured by Trp fluorescence and by far-UV CD are plotted against final guanidinium chloride concentrations reveals that the collapse of lysozyme is also a cooperative process (46). With  $\alpha$ -lactalbumin, formation of the A state appears noncooperative according to various relevant tests (47–49), but a variant, lacking all four disulfide bridges, exhibits a more cooperative behavior as judged by the superposition test of equilibrium unfolding curves (50, 51). With apoMb, the cooperativity can also be modulated by mutations or by the addition of anions (39, 52). Cooperativity has been linked to stability (39), and the formation of  $I_a$  could require a concerted process because a minimal stability must be reached to maintain the nascent structure.

It should also be noted that unfolding kinetics of a low pH form of apoMb, named E, also exhibits features of a cooperative two-state process since one of the two kinetics phases detected by infrared absorbance (IR) spectroscopy also gives rise to a change in Trp fluorescence (53, 54). The second phase seen by IR spectroscopy is not detected by Trp fluorescence.

*Why is the Formation of  $I_a$  So Slow near pH 4?* Recent studies have been carried out to estimate the upper limit of folding rate. Although it is still debated which theory should be used to predict folding rates, it seems that a consensus between experiments and theoretical predictions for the maximum rate constant attainable for an overall collapse and folding reaction is given by  $1/\tau = 10^8/N \text{ s}^{-1}$ , where  $N$  is the number of residues of the protein (55). This estimate would set a limit at 670 000 s<sup>-1</sup> ( $\tau = 1.5 \mu\text{s}$ ) for apomyoglobin folding reaction. In our studies, the formation of the first stable intermediate, containing secondary structure and hydrophobic packing is too fast to be measured at pH 6.0 with our mixing apparatus but occurs much more slowly at lower pH values, with a rate constant of 4000 s<sup>-1</sup> ( $\tau = 250 \mu\text{s}$ ) at pH 4.2 and a rate constant of 300 s<sup>-1</sup> at pH 3.0 ( $\tau = 3300 \mu\text{s}$ ) (11).

Why would the rate decrease at lower pH? Decreasing the pH reduces the stability of the intermediate, very likely because of the build-up of a net positive charge on the protein (24). Because it is more favorable to have multiple positive charges on an expanded form of the protein (U) than on a compact one ( $I_a$ ), electrostatic repulsions lead to the progressive destabilization of  $I_a$  when the pH is decreased. If the transition state for the  $U \rightleftharpoons I_a$  reaction has a compaction level close to that of  $I_a$ , it is likely that its stability is also reduced when the pH is decreased, and thus that the rate constant for folding decreases.

## ACKNOWLEDGMENT

We thank Prof. Buzz Baldwin for critical reading of the manuscript.

## REFERENCES

- Jamin, M., and Baldwin, R. L. (1996) Refolding and unfolding kinetics of the equilibrium folding intermediate of apomyoglobin, *Nat. Struct. Biol.* 3, 613–618.
- Jamin, M., and Baldwin, R. L. (1998) Two forms of the pH 4 folding intermediate of apomyoglobin, *J. Mol. Biol.* 276, 491–504.
- Jennings, P. A., and Wright, P. E. (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin, *Science* 262, 892–896.
- Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) Structural characterization of a partly folded apomyoglobin intermediate, *Science* 249, 1544–1548.
- Uzawa, T., Akiyama, S., Kimura, T., Takahashi, S., Ishimori, K., Morishima, I., and Fujisawa, T. (2004) Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of alpha-helical content and compactness, *Proc. Natl. Acad. Sci. U.S.A.* 101, 1171–1176.
- Haruta, N., and Kitagawa, T. (2002) Time-resolved UV resonance Raman investigation of protein folding using a rapid mixer: characterization of kinetic folding intermediates of apomyoglobin, *Biochemistry* 41, 6595–6604.
- Ha, J. H., and Loh, S. N. (1998) Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange, *Nat. Struct. Biol.* 5, 730–737.
- Nishimura, C., Dyson, H. J., and Wright, P. E. (2002) The apomyoglobin folding pathway revisited: structural heterogeneity in the kinetic burst phase intermediate, *J. Mol. Biol.* 322, 483–489.
- Eliezer, D., Yao, J., Dyson, H. J. and Wright, P. E. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding, *Nat. Struct. Biol.* 5, 148–155.
- Eliezer, D., Chung, J., Dyson, H. J., and Wright, P. E. (2000) Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin, *Biochemistry* 39, 2894–2901.
- Jamin, M., Yeh, S. R., Rousseau, D. L., and Baldwin, R. L. (1999) Submillisecond unfolding kinetics of apomyoglobin and its pH 4 intermediate, *J. Mol. Biol.* 292, 731–740.
- Tanford, C. (1970) Protein denaturation. Part C. Theoretical models for the mechanism of denaturation, *Adv. Prot. Chem.* 24, 2–95.
- Jackson, S. E., and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, *Biochemistry* 30, 10428–10435.
- Pace, C. N., Shirley, B. A. and Thompson, J. A. (1989) Measuring the conformational stability of a protein, in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 311–330, IRL Press, Oxford, UK.
- Fanelli, A. R., Antonini, E. and Caputo, A. (1958) Studies on the structure of hemoglobin. I. Physicochemical properties of human globin, *Biochim. Biophys. Acta* 30, 608–615.
- Springer, B. A. and Sligar, S. G. (1987) High-level expression of sperm whale myoglobin in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* 84, 8961–8965.
- Jennings, P. A., Stone, M. J., and Wright, P. E. (1995) Overexpression of myoglobin and assignment of its amide, C alpha and C beta resonances, *J. Biomol. NMR* 6, 271–276.
- Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins, *Biochemistry* 6, 1948–1954.
- Peterman, B. F. (1979) Measurement of the dead time of a fluorescence stopped-flow instrument, *Anal. Biochem.* 93, 442–444.
- Hiroimi, K. (1979) *Kinetics of Fast Enzyme Reactions: Theory & Practice*, John Wiley & Sons, New York.
- Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants, *Biochemistry* 27, 8063–8068.
- Monod, J., Wyman, J. and Changeux, J. P. (1965) On the nature of allosteric transitions: A plausible model, *J. Mol. Biol.* 12, 88–118.
- Jamin, M., Geierstanger, B. and Baldwin, R. L. (2001) The pK<sub>a</sub> of His24 in the folding transition state of apomyoglobin, *Proc. Natl. Acad. Sci. U.S.A.* 98, 6127–6131.
- Kay, M. S., and Baldwin, R. L. (1998) Alternative models for describing the acid unfolding of the apomyoglobin folding intermediate, *Biochemistry* 37, 7859–7868.
- Wright, P. E., and Baldwin, R. L. (2000) Case study 1: The folding process of apomyoglobin, in *Mechanisms of Protein Folding* (Pain, R. H., Ed.) pp 309–329, Oxford University Press, Oxford, UK.
- Jamin, M. (2005) Apomyoglobin folding process, *Prot. Pept. Lett.*, in press.
- Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F., and Goto, Y. (1995) Structural characterization of the molten globule and native states of apomyoglobin by solution X-ray scattering, *J. Mol. Biol.* 249, 215–228.
- Loh, S. N., Kay, M. S., and Baldwin, R. L. (1995) Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway, *Proc. Natl. Acad. Sci. U.S.A.* 92, 5446–5450.
- Garcia, C., Nishimura, C., Cavagnero, S., Dyson, H. J., and Wright, P. E. (2000) Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue, *Biochemistry* 39, 11227–11237.
- Kiefhaber, T., and Baldwin, R. L. (1995) Intrinsic stability of individual alpha helices modulates structure and stability of the apomyoglobin molten globule form, *J. Mol. Biol.* 252, 122–132.
- Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., and Wright, P. E. (1999) Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate, *Protein Sci.* 8, 45–49.
- Creighton, T. E. (1995) Protein folding. An unfolding story, *Curr. Biol.* 5, 353–356.
- Privalov, P. L. (1979) Stability of proteins: small globular proteins, *Adv. Prot. Chem.* 33, 167–241.
- Griko, Y. V., and Privalov, P. L. (1994) Thermodynamic puzzle of apomyoglobin unfolding, *J. Mol. Biol.* 235, 1318–1325.
- Jamin, M., Antalík, M., Loh, S. N., Bolen, D. W., and Baldwin, R. L. (2000) The unfolding enthalpy of the pH 4 molten globule of apomyoglobin measured by isothermal titration calorimetry, *Protein. Sci.* 9, 1340–1346.
- Freire, E. (1995) Thermodynamics of partly folded intermediates in proteins, *Annu. Rev. Biophys. Biomol. Struct.* 24, 141–165.
- Nishii, I., Kataoka, M., Tokunaga, F., and Goto, Y. (1994) Cold denaturation of the molten globule states of apomyoglobin and a profile for protein folding, *Biochemistry* 33, 4903–4909.
- Kay, M. S., and Baldwin, R. L. (1996) Packing interactions in the apomyoglobin folding intermediate, *Nat. Struct. Biol.* 3, 439–445.
- Luo, Y., Kay, M. S., and Baldwin, R. L. (1997) Cooperativity of folding of the apomyoglobin pH 4 intermediate studied by glycine and proline mutations, *Nat. Struct. Biol.* 4, 925–930.
- Tsong, T. Y., and Baldwin, R. L. (1972) A sequential model of nucleation-dependent protein folding: kinetic studies of ribonuclease A, *J. Mol. Biol.* 63, 453–469.
- Parker, M. J., and Marqusee, S. (2000) A statistical appraisal of native state hydrogen exchange data: evidence for a burst phase continuum?, *J. Mol. Biol.* 300, 1361–1375.
- Agashe, V. R., Shastry, M. C. R., and Udgaonkar, J. B. (1995) Initial hydrophobic collapse in the folding of barstar, *Nature* 377, 754–757.
- Sadqi, M., Lapidus, L. J., and Muñoz (2003) How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci. U.S.A.* 100, 12117–12122.
- Magg, C., and Schmid, F. X. (2004) Rapid collapse precedes the fast two-state folding of the cold shock protein, *J. Mol. Biol.* 335, 1309–1323.
- Hagen, S. J., and Eaton, W. A. (2000) Two-state expansion and collapse of a polypeptide, *J. Mol. Biol.* 301, 1019–1027.
- Bachmann, A., Segel, D., and Kiefhaber, T. (2002) Test for cooperativity in the early kinetic intermediate in lysozyme folding, *Biophys. Chem.* 96, 141–151.
- Xie, D., Bhakuni, V., and Freire, E. (1991) Calorimetric determination of the energetics of the molten globule intermediate in protein folding: apo-alpha-lactalbumin, *Biochemistry* 30, 10673–10678.

48. Schulman, B. A., and Kim, P. S. (1996) Proline scanning mutagenesis of a molten globule reveals non-cooperative formation of a protein's overall topology, *Nat. Struct. Biol.* 3, 682–687.
49. Schulman, B. A., Kim, P. S., Dobson, C. M. and Redfield, C. (1997) A residue-specific NMR view of the non-cooperative unfolding of a molten globule, *Nat. Struct. Biol.* 4, 630–634.
50. Redfield, C., Schulman, B. A., Milhollen, M. A., Kim, P. S., & Dobson, C. M. (1999)  $\alpha$ -lactalbumin forms a compact molten globule in the absence of disulfide bonds, *Nat. Struct. Biol.* 6, 948–952.
51. Luo, Y., & Baldwin, R. L. (1999) The 28–111 disulfide bond constrains the alpha-lactalbumin molten globule and weakens its cooperativity of folding, *Proc. Natl. Acad. Sci. U.S.A.* 96, 11283–11287.
52. Kay, M. S., Ramos, C. H., and Baldwin, R. L. (1999) Specificity of native-like interhelical hydrophobic contacts in the apomyoglobin intermediate, *Proc. Natl. Acad. Sci. U.S.A.* 96, 2007–2012.
53. Gilmanshin, R., Callender, R. H., and Dyer, R. B. (1998) The core of apomyoglobin E-form folds at the diffusion limit, *Nat. Struct. Biol.* 5, 363–365.
54. Gulotta, M., Gilmanshin, R., Buscher, T. C., Callender, R. H., and Dyer, R. B. (2001) Core formation in apomyoglobin: probing the upper reaches of the folding energy landscape, *Biochemistry* 40, 5137–5143.
55. Kubelka, J., Hofrichter, J. and Eaton, W. A. (2004) The protein folding 'speed limit', *Curr. Opin. Struct. Biol.* 14, 76–88.

BI047372V

## **Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins**

Cécile Hamès, Denis Ptchelkine, Clemens Grimm, Emmanuel Thevenon,  
Edwige Moyroud, **Francine Gérard**, Jean-Louis Martiel, Reyes Benlloch,  
François Parcy et Christoph W. Müller

*EMBO Journal*, 2008

L'appareillage de diffusion statique de lumière couplé à la chromatographie d'exclusion de taille et réfractométrie (SEC-MALLS-RI) dont nous disposons au laboratoire, équipement obtenu par mon codirecteur de thèse, le Pr. Marc JAMIN, a permis de caractériser protéines et complexes macromoléculaires en solution. Ainsi, nous avons collaboré avec des chercheurs et étudiants sur des projets très différents (Lata, 2008 #448) (Hodak, 2008 #449) (Macheboeuf, 2008 #450) (Erlich, 2008 #453) (Dumestre-Perard, 2007 #451) (Buisson, 2006 #454). Cet appareillage permet de répondre à des questions d'état d'oligomérisation d'une protéine mais aussi de complexe ainsi que sur la monodispersité des échantillons.

La protéine LEAFY (LFY) est un régulateur clé dans le développement des angiospermes. Ici, Cécile Hamès et ses collaborateurs ont obtenus la structure cristallographique du domaine de liaison à l'ADN de LFY d'*Arabidopsis thaliana*, en liaison avec 2 éléments promoteurs cibles. LFY adopte un nouveau repliement en 7 hélices, et lie l'ADN de façon coopérative en formant un dimère. Le dimère forme des contacts spécifiques avec les bases de l'ADN dans les cavités majeures et mineures. Ici, nous avons avec Cécile Hamès injecter ses échantillons de protéine LFY seule et en complexe avec l'ADN, et l'analyse par SEC-MALLS-RI nous a permis de conclure sur la dimérisation de la protéine lors de la fixation de l'ADN ([figure 1, A](#)).

# Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins

Cécile Hamès<sup>1,6</sup>, Denis Ptchelkine<sup>2,3,6</sup>,  
Clemens Grimm<sup>2,7</sup>, Emmanuel Thevenon<sup>1</sup>,  
Edwige Moyroud<sup>1</sup>, Francine Gérard<sup>4</sup>,  
Jean-Louis Martiel<sup>5</sup>, Reyes Benlloch<sup>1,8</sup>,  
François Parcy<sup>1,\*</sup> and Christoph W Müller<sup>2,3,\*</sup>

<sup>1</sup>Laboratoire Physiologie Cellulaire Végétale, UMR5168, Centre National de la Recherche Scientifique, Commissariat à l'énergie atomique, Institut National de la Recherche Agronomique, Université Joseph Fourier, Grenoble, France, <sup>2</sup>European Molecular Biology Laboratory, Grenoble, France, <sup>3</sup>European Molecular Biology Laboratory, Structural and Computational Biology Unit, Heidelberg, Germany, <sup>4</sup>Unit of Virus Host Cell Interactions, UMR5233 Université Joseph Fourier—European Molecular Biology Laboratory—Centre National de la Recherche Scientifique, Grenoble, France and <sup>5</sup>TIMC-IMAG Laboratory, Université Joseph Fourier, CNRS UMR5525, INSERM, Grenoble, France

**The LEAFY (LFY) protein is a key regulator of flower development in angiosperms. Its gradually increased expression governs the sharp floral transition, and LFY subsequently controls the patterning of flower meristems by inducing the expression of floral homeotic genes. Despite a wealth of genetic data, how LFY functions at the molecular level is poorly understood. Here, we report crystal structures for the DNA-binding domain of *Arabidopsis thaliana* LFY bound to two target promoter elements. LFY adopts a novel seven-helix fold that binds DNA as a cooperative dimer, forming base-specific contacts in both the major and minor grooves. Cooperativity is mediated by two basic residues and plausibly accounts for LFY's effectiveness in triggering sharp developmental transitions. Our structure reveals an unexpected similarity between LFY and helix-turn-helix proteins, including homeodomain proteins known to regulate morphogenesis in higher eukaryotes. The appearance of flowering plants has been linked to the molecular evolution of LFY. Our study provides a unique framework to elucidate the molecular mechanisms underlying floral development and the evolutionary history of flowering plants.**

*The EMBO Journal* advance online publication, 11 September 2008; doi:10.1038/emboj.2008.184

**Subject Categories:** plant biology; structural biology

\*Corresponding authors. F Parcy, Laboratoire Physiologie Cellulaire Végétale, CNRS, UMR5168, CEA, 17 av. des Martyrs, bât. C2, 38054 Grenoble, France. Tel.: +33 438 784 978; Fax: +33 438 784 091; E-mail: francois.parcy@cea.fr or CW Müller, Structural and Computational Biology Unit, EMBL Meyerhofstrasse 1, 69012 Heidelberg, Germany. Tel.: +49 6221 387 8320; Fax: +49 6221 387 519; E-mail: christoph.mueller@embl.de

<sup>6</sup>These authors contributed equally to this work

<sup>7</sup>Present address: Institut für Biochemie, Biozentrum der Universität, Am Hubland, 97074 Würzburg, Germany

<sup>8</sup>Present address: Department of Forest Genetics and Plant Physiology, Umeå Plant Science Centre, Swedish University of Agricultural Sciences, 90183 Umeå, Sweden

**Keywords:** crystal structure; flower development; homeotic genes; LEAFY; transcriptional regulation

## Introduction

Homeotic genes control developmental patterns and organ morphogenesis. In animals, they encode transcription factors of the homeodomain family, such as Hox and paired proteins, which contact DNA through one or several helix-turn-helix (HTH) motifs (Gehring *et al*, 1994; Underhill, 2000). In plants, most homeotic genes determining the identity of floral organs encode MADS-box transcription factors, suggesting that plants and animals have adopted distinct types of homeotic regulators (Meyerowitz, 1997; Ng and Yanofsky, 2001). In addition to organ identity genes, plants also use another class of regulators named 'meristem identity genes', which control floral meristem versus shoot/inflorescence fate. In *Arabidopsis thaliana*, the meristem identity genes *LEAFY* (*LFY*) and *APETALA1* (*API*) induce flower development, whereas *TERMINAL FLOWER1* (*TFL1*) promotes inflorescence development (Blazquez *et al*, 2006). Mutations or ectopic expression of these genes result in complete or partial interconversions between flower and inflorescence meristems.

The *LFY* gene encodes a plant-specific transcription factor, which has a cardinal function in this process, regulating both the transition to flowering and the subsequent patterning of young floral meristems. During the plant vegetative growth, *LFY* expression increases in newly formed leaves until a certain threshold is reached. *LFY* then induces the expression of *API* and *CAULIFLOWER* (*CAL*) genes and triggers the abrupt floral transition (Blazquez *et al*, 2006). Once the floral meristem is established, *LFY* governs its spatial patterning by inducing the expression of the floral homeotic ABC genes, such as *API*, *AP3* or *AGAMOUS* (*AG*), which control the identity of stereotypically arranged floral organs (Coen and Meyerowitz, 1991; Lohmann and Weigel, 2002).

*LFY* is found in all terrestrial plants from moss to angiosperms; its sequence shows a high level of conservation throughout the plant kingdom but no apparent similarity to other proteins (Maizel *et al*, 2005). Unlike many plant transcription factors that evolved by gene duplication to form a multigene family (Riechmann and Ratcliffe, 2000; Shiu *et al*, 2005), *LFY* is present in single copy in most angiosperms and *lfy* mutants available from several species such as snapdragon, petunia, tomato or maize show, as in *Arabidopsis*, partial or complete flower-to-shoot conversions (Coen *et al*, 1990; Souer *et al*, 1998; Molinero-Rosales *et al*, 1999; Bomblies *et al*, 2003). In gymnosperms, a paralogous *NEEDLY* (*NLY*) clade of genes exists. No mutant is available in these species, but *LFY* and *NLY* expression patterns are also consistent with a role in reproductive organ development

Received: 11 April 2008; accepted: 22 August 2008

(reviewed in Frohlich and Chase, 2007). Because of its central role in determining floral meristem identity, and considering that *NLY* disappeared concomitantly with the appearance of flowers, *LFY* has been put at the centre of different evolutionary scenarios that rationalize the appearance of the successful angiosperm group (Albert *et al*, 2002; Frohlich, 2003; Frohlich and Chase, 2007; Theissen and Melzer, 2007).

*LFY* activates gene expression by recognizing pseudo-palindromic sequence elements (CCANTGT/G) in the promoters of its target genes, including *API* (one site) and *AG* (four sites; *AG-I* to *AG-IV*) (Parcy *et al*, 1998; Busch *et al*, 1999; Lohmann *et al*, 2001; Lamb *et al*, 2002; Hong *et al*, 2003). *LFY* has two domains, a partially conserved N-terminal domain that is thought to contribute to transcriptional activation and a highly conserved C-terminal domain responsible for DNA binding (*LFY-C*) (Coen *et al*, 1990; Maizel *et al*, 2005). *LFY* functions synergistically with coregulators such as the *WUSCHEL* (*WUS*) homeodomain protein (Lenhard *et al*, 2001; Lohmann *et al*, 2001) or the *UFO* F-Box protein (Lee *et al*, 1997; Parcy *et al*, 1998; Chae *et al*, 2008).

In this study, we show that *LFY* binds DNA cooperatively as a dimer, a property shown to be essential to trigger developmental switches. The crystal structure of *LFY-C* bound to DNA reveals the molecular basis for sequence-specific recognition and cooperative binding as well as an unexpected similarity of *LFY* with HTH proteins such as homeodomain transcription factors. Our findings enable to formulate new hypotheses on the appearance of angiosperms in evolution.

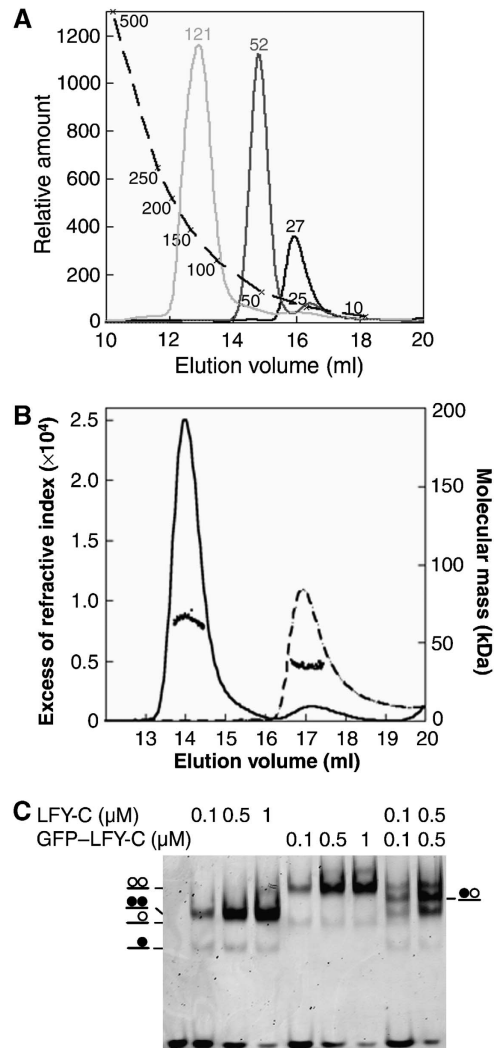
## Results and discussion

### *LFY-C* dimerizes on DNA binding

We produced the recombinant *LFY* DNA-binding domain (*LFY-C*, residues 223–424) and showed by size-exclusion chromatography (SEC) that it is monomeric in the absence of DNA (Figure 1A and B). In electrophoretic mobility shift assays (EMSAs), *LFY-C* recognized a DNA probe bearing an *API* site as two distinct species: a major protein–DNA complex and a minor one of higher mobility (Figure 1C). Multi-angle laser light scattering (MALLS) coupled to SEC demonstrated that the major complex contained two *LFY-C* molecules per DNA duplex (Figure 1B). The homodimeric nature of *LFY* in this complex was confirmed by mixing untagged and GFP-tagged *LFY-C* and observing a single new species attributable to the formation of an *LFY-C*/GFP-*LFY-C*/DNA complex (Figure 1C). Using probes mutated in one half-site of the palindrome, we confirmed that the minor, high-mobility species corresponds to a single *LFY-C* monomer bound to DNA (Supplementary Figure 2).

### Structure of the *LFY* DNA-binding domain bound to its DNA recognition site

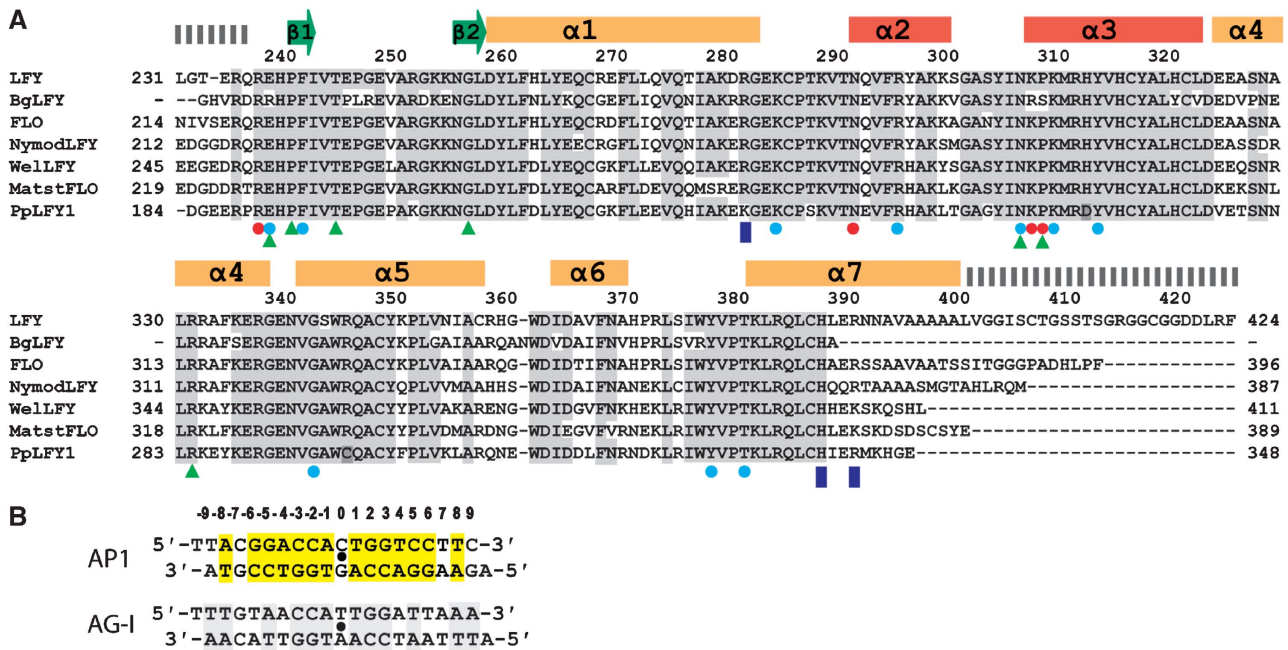
To understand how *LFY* specifically recognizes its DNA target sequences, we crystallized *LFY-C* in complex with DNA. We solved the structure of *LFY-C* bound to two different *LFY*-binding sites, *API* and *AG-I* at 2.1- and 2.3-Å resolution, respectively (Figures 2, and 3A and B; Table 1). The overall structure shows an *LFY-C* dimer bound to a pseudo-palindromic DNA duplex, where the *LFY-C* monomers are related by a crystallographic dyad. The DNA duplexes used for co-crystallization deviate from strict two-fold symmetry at the 5' ends and at base pairs (bp)  $\pm 9$ ,  $\pm 7$  and  $\pm 0$  in the *API* site



**Figure 1** DNA-dependent dimerization of *LFY-C*. (A) Size-exclusion chromatography. *LFY-C* (40  $\mu$ M, black curve), *API* DNA (10  $\mu$ M, dark grey curve), *LFY-C* (40  $\mu$ M) + *API* DNA (10  $\mu$ M, light grey curve) were analysed. *LFY-C* elution at a volume corresponding to 28 kDa is consistent with the monomer size (25.7 kDa), the DNA duplex elutes earlier than expected at a volume corresponding to 52 kDa because of its elongated shape. The *LFY-C*/DNA complex elutes at a volume corresponding to 121 kDa. Molecular weights estimated from the calibration curve (dashed line) are indicated. (B) Molecular mass of *LFY-C* alone (dashed line) or in combination with *API* DNA (solid line) determined by multi-angle laser light scattering and refractometry combined with size-exclusion chromatography. Elution profiles were monitored by excess refractive index (left ordinate axis). Dots show the molecular mass distribution (right ordinate axis). Average molecular mass is  $64 \pm 2$  kDa for the *LFY-C*/DNA complex (65 kDa theoretical size for a dimeric complex) and  $35 \pm 1$  kDa for *LFY-C* alone (26 kDa theoretical size for *LFY-C* monomer). (C) Electrophoretic mobility shift assay (EMSA) with 10 nM *API* DNA and various *LFY-C* or GFP-*LFY-C* concentrations. Schematic complexes with *LFY-C* (filled circle) and GFP-*LFY-C* (open circle) are depicted.

(5' end, bp  $\pm 7$ ,  $\pm 6$ ,  $\pm 4$  and  $\pm 0$  in the *AG-I* site). Nevertheless, the pseudo-dyads of the DNA duplexes coincide with the crystallographic dyad, probably as a result of the random bimodal orientation of the DNA duplex around the dyad (see Materials and methods). The resulting molecular averaging does not impair our interpretation of the protein–DNA interface. In the final  $2F_0 - F_c$  elec-





**Figure 2** Sequence alignments. (A) Aligned C-terminal amino-acid sequences of LFY (*Arabidopsis thaliana*, AAA32826), BgLFY (*Brownia grandiceps*, AAS79888), FLO (*Antirrhinum majus*, P23915), NymodLFY (*Nymphaea odorata*, AAF77609), WelLFY (*Welwitschia mirabilis*, AAF23870), MatstLFY (*Matteuccia struthiopteris*, AAF77608) and PpLFY1 (*Physcomitrella patens*, BAD91043). Identical and conservatively substituted residues are depicted on a grey background. Secondary structure elements are indicated. Residues involved in interactions with DNA bases and backbone are labelled with red and blue circles, respectively. Dashed bars indicate disordered regions in the crystal, blue rectangles indicate the residues involved in dimerization. Green triangles indicate the position of *Arabidopsis* mutations and residues divergent in PpLFY1 are highlighted in pink. (B) Two DNA duplexes containing the LEAFY-binding sites from *AP1* and *AG-I* promoters present in the LEAFY-DNA complex crystals are depicted. Base pairs related by a dyad (indicated by a black dot) are highlighted in yellow.

tron density map, but also in the initial solvent-flattened single isomorphous replacement with anomalous scattering (SIRAS) electron density map (Supplementary Figure 1), the sugar-phosphate backbone of the DNA is well defined and shows no evidence of conformational averaging. The density for palindromic DNA bases is clearly defined, whereas the density at non-palindromic positions is consistent with the superposition of two different base pairs. Furthermore, all residues close to the DNA are clearly defined and we do not observe any diffuse density, which suggests that each monomer undergoes only minor changes to adapt to the slightly different half-sites. Despite the differences between the *AP1* and *AG-I* binding sites (Figure 2B), both complex structures are very similar and can be superimposed with an r.m.s. distance of 0.55 Å for 163 C<sub>α</sub> and 19 phosphate atoms.

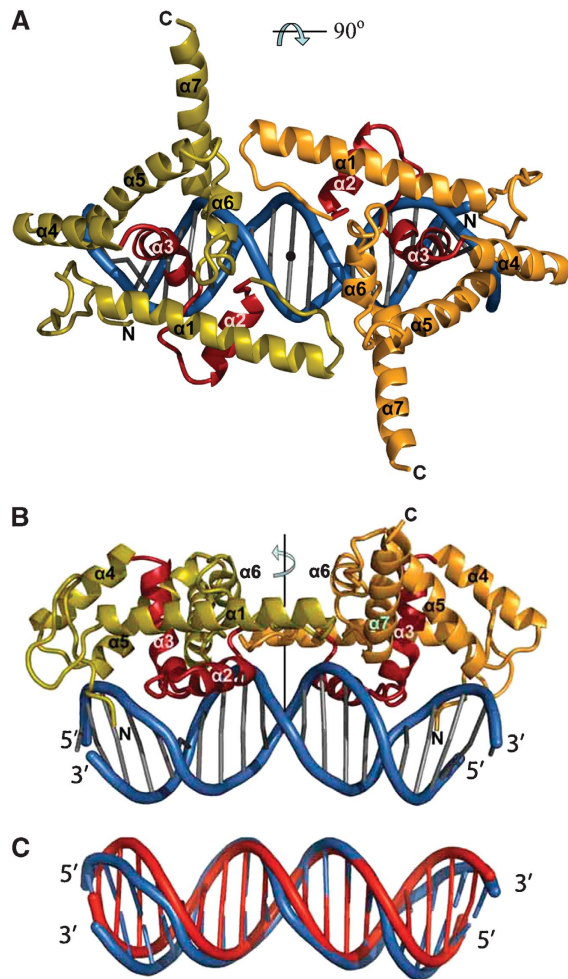
LFY-C (with residues 237–399 ordered in the crystal structure) adopts a compact fold that interacts principally with a single DNA half-site (Figures 3A and B, and 4A and B). The fold is defined by two short β-strands followed by seven helices connected by short loops (Figure 3A and B). The absence of any extended hydrophobic patches at its surface suggests that LFY-C represents an autonomous DNA-binding domain without a large interface to its N-terminal domain. Helices α2 and α3 define a HTH motif (Aravind *et al*, 2005), with helix α3 occupying the major groove and mediating most of the DNA contacts. The DNA in the complex adopts a B-DNA-like conformation exhibiting an overall bend of about 20° (Figure 3C), which can be localized to two kinks of about 10° at base pairs ±2/±3. Both ends of the DNA duplex are AT rich and the minor grooves are narrower compared with

classical B-DNA. Narrowing of the minor groove is slightly more pronounced in the *AG-I* duplex than in *AP1*.

### DNA recognition in the major and minor grooves

Sequence-specific contacts between LFY and the DNA involve both the minor and major grooves. Base-specific contacts in the major groove are formed by Asn291 and Lys307 in helices α2 and α3, which together specify the two invariant guanines at positions ±2 and ±3 (Figure 4A and B). Mutating either of these residues into alanine resulted in considerably lower DNA-binding affinity (Figure 4C), whereas previous studies showed loss of binding when the corresponding base pairs were mutated (Parcy *et al*, 1998; Busch *et al*, 1999). The *Arabidopsis lfy-20* mutation (N306D) adjacent to Lys307 also leads to a reduced DNA-binding affinity (Supplementary Figure 3) and a weak *lfy* phenotype *in planta* (Weigel *et al*, 1992; Maizel *et al*, 2005), presumably because the negatively charged aspartate interacts unfavourably with the DNA backbone (Figure 4A and B).

Base-specific recognition in the minor groove is mediated by Arg237, which is the first ordered N-terminal residue in the crystal structure. At the *AP1* site, its side chain points towards A:T base pairs ±8 and contacts the exocyclic O2 of thymine 8 and also the O2 of cytosine 7 in one half-site, or the O2 of cytosine-9 in the other half-site (Figure 4A and B; Supplementary Figure 4). In the *AG-I* site, T:A base pair 8 is replaced by A:T, and in the LFY/*AG-I* complex, the Arg237 side chain adopts a different conformation, which allows it to recognize the thymine of the opposite strand (Supplementary Figure 4). The importance of this interaction is underscored by



**Figure 3** Structure of the LFY-C dimer bound to DNA. (A, B) Two orthogonal views of the LFY-C dimer (residues 237–399) bound to DNA. Monomers are coloured in olive and orange with the helix-turn-helix (HTH, helices  $\alpha 2$  and  $\alpha 3$ ) motif in red. The DNA duplex is depicted in blue. Figures 3, 4B, 5A and 6 were produced with program Pymol (Delano, 2002). (C) Superposition of the DNA duplex found in the LEAFY–DNA complex (blue) with regular B-form DNA (red).

the presence of A:T or T:A base pairs at position 8 in all 12 confirmed half-sites (Parcy *et al*, 1998; Busch *et al*, 1999; Lohmann *et al*, 2001; Lamb *et al*, 2002; Hong *et al*, 2003). The consensus LFY-binding site is therefore more accurately defined as T/ANNNNCCANTGT/GNNNNT/A (with the centre of the pseudo-palindrome underlined). The Arg237 side chain is inserted into an AT-rich narrow minor groove (Figure 4A and B), similar to that observed in the Hox homeodomain–Exd–DNA complex, where the narrow minor groove was shown to enhance the electrostatic interaction between DNA backbone and arginine side chain (Joshi *et al*, 2007). The R237A mutation led to a strongly reduced affinity of LFY-C for *AP1* (Figure 4C). In contrast, changing the adenine 8 into a cytosine in *AP1* reduced only moderately the LFY-C-binding affinity (Figure 4C; *AP1* m5), presumably because the arginine side chain can contact the adjacent base (Supplementary Figure 4). Finally, next to Arg237, the two *lfy* mutations (*lfy-4* (E238K) and *lfy-5* (P240L)) result in decreased *in vitro* binding affinities (Supplementary Figure 3) and lead to a mutant phenotype *in planta* (Weigel *et al*, 1992).

An unusual contact with DNA is mediated by Pro308 that points between the guanines in base pairs  $\pm 5$  and  $\pm 6$ , which results in a pronounced propeller twist for base pair  $\pm 5$  and local bending of the DNA at this position (Figure 4B). The mutant *lfy-28* (P308L) is impaired in DNA binding and gives rise to an intermediate to strong phenotype *in planta* (Figure 4C–F), as a likely consequence of a steric clash of the leucine side chain with the guanine bases. In contrast, a small side chain such as alanine perfectly fits in this protein–DNA interface and, indeed, the mutant protein P308A showed a wild-type-binding affinity (Figure 4C). Pro308 is not strictly conserved and is substituted by serine in some *Brownea* species (Figure 2A). This substitution probably modifies DNA binding, because serine can form direct hydrogen bonds to DNA bases at positions  $\pm 4$  and  $+ 5$  and it replaces P308, which locally distorts DNA. However, the conformational flexibility of serine and its ability to function as hydrogen bond donor or acceptor makes it difficult to predict the preferred binding specificity. Moreover, P308S is systematically associated with the K307R substitution, affecting the base-contacting residue Lys307 (Figure 2A). LFY proteins from *Brownea* species might therefore recognize significantly different DNA target sites.

Similar to most protein–DNA co-crystal structures, not all bases in the consensus LFY site T/ANNNNCCANTGT/GNNNNT/A are specified through direct interactions with the protein. Additional specificity presumably arises from sequence-dependent deformability of the DNA, sometimes referred to as ‘indirect readout’. Dinucleotide steps CA/TG at bp  $\pm 1$ /bp  $\pm 2$  are part of the consensus LFY site and are particularly flexible, which might facilitate the observed kink of the DNA at base pairs  $\pm 2/\pm 3$ . However, these particular sequences are not critically required as they are not conserved in the *AP3-I* binding site (Lamb *et al*, 2002).

Not all LFY mutations directly affect DNA contacts. Mutations *lfy-3* (T244M) and *lfy-9* (R331K) (Weigel *et al*, 1992) disturb two interacting amino acids, both of which contribute to a polar network that connects N-terminal residues with helices  $\alpha 1$  and  $\alpha 4$ . Similarly, two other residues (His312 and Arg345) interact in a typical planar stacking. His312 and Arg345 are conserved except in the LFY protein from *Physcomitrella patens* (PpLFY1) where they are substituted by aspartate and cysteine, respectively (Maizel *et al*, 2005). His312 forms part of helix  $\alpha 3$  and is located just one helical turn above Pro308 at the N-terminal end of helix  $\alpha 3$ . In addition, the preceding residue Lys307 directly contacts the guanine in base pairs  $\pm 2$  (Figure 4A). The loss of the His312/Arg345 stacking interaction in the moss PpLFY1 likely affects the orientation of helix  $\alpha 3$ , explaining the altered DNA-binding properties of this orthologue, whereas reverting the aspartate into histidine restores the binding activity of PpLFY1 to canonical LFY-binding sites (Maizel *et al*, 2005).

#### Structural basis for cooperative DNA binding

The structure of the LFY-C/DNA complex also reveals important monomer–monomer interactions governing its DNA-binding mode (Figure 5A). Our EMSA analysis shows that LFY-C binds DNA in a cooperative manner: the monomeric complex is present only in minor amounts as compared with the dimeric complex, even at low LFY-C concentrations (Figure 5B) and binding of the second monomer occurs with a 90-fold higher affinity than binding of the first one

**Table I** Structure determination of the LEAFY–DNA complex

<i>Data statistics</i>					
Data set	Resolution (Å) <sup>a</sup>	Reflections measured/unique	$R_{\text{meas}}$ (%) <sup>b</sup>	$I/\sigma$	Completeness (%)
<i>API/LFY</i> : space group P6 <sub>5</sub> 22, unit cell dimensions a = b = 98.8 Å, c = 177.4 Å					
Native	20–2.1 (2.2–2.1)	503 443/29 859	4.8 (54.3)	38.5 (5.0)	97.4 (83.9)
EMTS	20–2.4 (2.5–2.4)	283 416/20 463	11.1 (50.2)	15.5 (4.7)	99.6 (99.3)
<i>AG-I/LFY</i> : space group P6 <sub>5</sub> 22, unit cell dimensions a = b = 98.4 Å, c = 176.4 Å					
Native	20–2.3 (2.4–2.3)	328 932/23 456	7.9 (75.1)	26.7 (3.4)	99.8 (100.0)
<i>Phasing statistics for the EMTS derivative (SIRAS)</i>					
Wavelength (Å)	0.934				
Phasing power <sup>c</sup>	1.68				
Figure of merit	0.468				
$R_{\text{cullis}}$ <sup>d</sup>	0.674				
Number of mercury sites	4				
<i>Refinement statistics</i>					
		LFY/API	LFY/AG-I		
Resolution <sup>a</sup>		19.1–2.1 (2.2–2.1)	20.0–2.3 (2.4–2.3)		
Total number of non-hydrogen protein atoms		1352	1332		
Total number of non-hydrogen DNA atoms		526	567		
Number of water molecules		148	105		
R-factor (%) <sup>a</sup>		21.0 (25.0) for 26 896 reflections	22.1 (29.0) for 22 559 reflections		
$R_{\text{free}}$ (%) <sup>a,e</sup>		23.7 (26.4) for 1521 reflections	24.9 (30.7) for 1183 reflections		
<i>r.m.s. deviations</i>					
Bond lengths (Å)		0.009	0.008		
Bond angles (deg)		1.22	1.21		
<i>Average temperature factors (Å<sup>2</sup>)</i>					
Protein		41.1	39.8		
DNA		48.1	46.4		
Solvent		42.8	48.5		
r.m.s.d. of covalently linked atoms (Å <sup>2</sup> )		2.55	3.45		
<i>Residues in Ramachandran plot<sup>f</sup></i>					
Most favoured regions (%)		93.2 (136)	95.2 (139)		
Additionally allowed regions (%)		6.8 (10)	4.8 (7)		
Generously allowed regions (%)		0 (0)	0 (0)		
Disallowed region (%)		0 (0)	0 (0)		

<sup>a</sup>Values for the highest resolution range are given in parenthesis.

<sup>b</sup> $R_{\text{meas}}$  is a redundancy independent R-factor as defined in Diederichs and Karplus (1997).

<sup>c</sup>Phasing power is the mean value of the heavy-atom structure factor amplitudes divided by the mean lack of closure.

<sup>d</sup> $R_{\text{cullis}}$  is the mean lack of closure divided by the mean isomorphous difference.

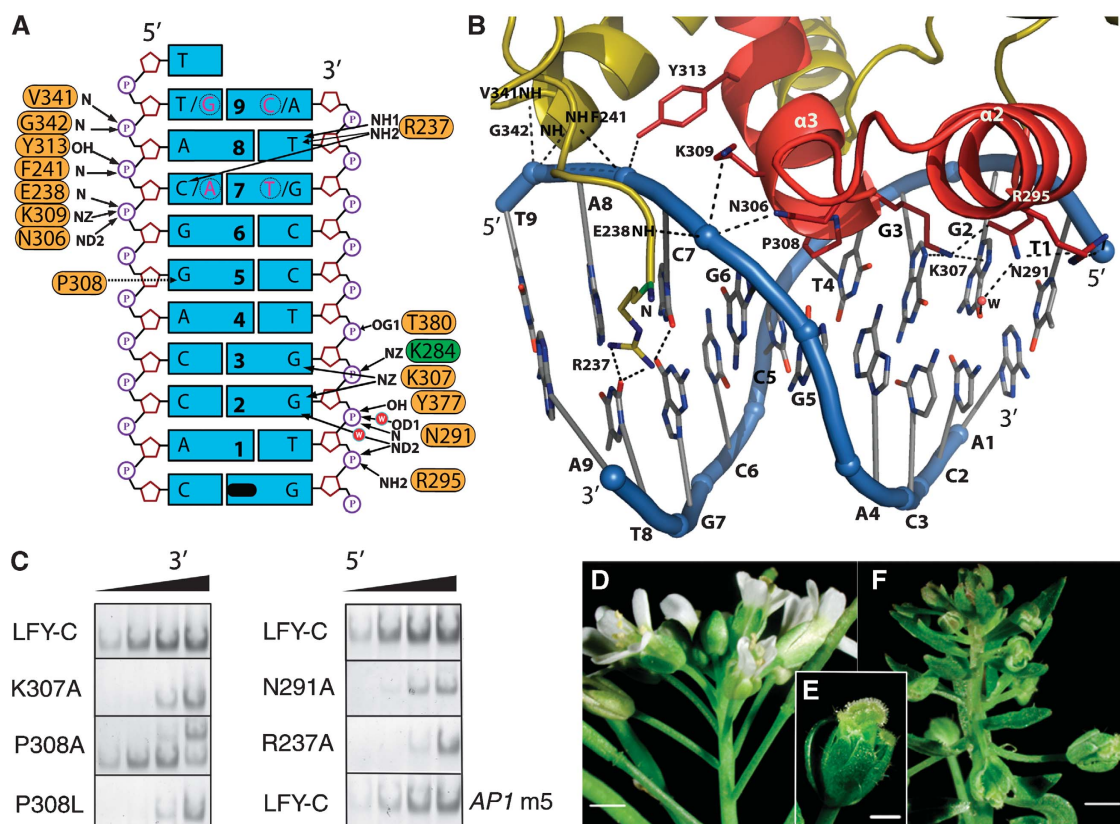
<sup>e</sup> $R_{\text{free}}$  was calculated from a subset of 5% of the data.

<sup>f</sup>Number of residues are given in parentheses.

(Figure 5C; Supplementary Figure 5). This type of cooperative binding can result either from DNA conformability, where binding of one monomer favours the binding of the second monomer, or from protein–protein interactions between DNA-bound monomers (Senear *et al*, 1998; Schumacher *et al*, 2002; Panne *et al*, 2004). Our structure suggests the latter. The LFY dimer comprises a small interface of 420 Å<sup>2</sup> buried surface area formed by loop  $\alpha$ 12 and helix  $\alpha$ 7 in which the two residues His387 and Arg390 form hydrogen bonds with the backbone carbonyl of Asp280 (Figure 5A). We validated the importance of these contacts by mutagenesis: cooperativity of binding is moderately affected in the H387A or R390A single mutants but more strongly reduced in a H387A/R390A double mutant (Figure 5B and C; Supplementary Figure 5). Therefore, the small monomer–monomer interface (with a major contribution of His387 and

Arg390) rather than DNA conformability is responsible for the cooperative binding. Whether the N-terminal domain of LFY also participates in dimerization, in the presence or absence of DNA, will require additional experiments.

A better understanding of LFY's DNA-binding mode also provides insight into its molecular switch function. DNA-binding cooperativity, as well as dimerization, allows transcription factors to work at lower concentrations and to enhance the sigmoidality of their response curves. When combined with feedback loops, it has been shown essential for threshold-dependent genetic switches (Burz *et al*, 1998; Cherry and Adler, 2000). LFY is involved in a positive autoregulation loop through activation of the homologous *API* and *CAL* genes, that in turn activate *LFY* expression (Bowman *et al*, 1993; Liljegren *et al*, 1999). LFY-binding cooperativity combined with the *API*/*CAL* feedback loop



**Figure 4** DNA recognition by LEAFY. **(A)** Protein–DNA interactions in one *AP1* half-site. Dyad-related base pairs 7 and 9 from the other half-site are shown in pink and encircled. Polar and hydrophobic interactions are shown with solid and dashed arrows, respectively. K284 belongs to the other monomer and is depicted in green. The pseudo-dyad coinciding with the crystallographic dyad is depicted in black. **(B)** Ribbon diagram of one LEAFY monomer bound to its *AP1* half-site. The protein is coloured in olive except for the HTH motif shown in red. Polar interactions are indicated by dashed lines. For clarity, only side chains in contact with DNA are shown. **(C)** Effect of selected mutations on LFY–C DNA-binding affinity to *AP1* DNA. EMSAs were performed with wild-type and mutant LFY–C (100–250–750–2000 nM from left to right). Only dimeric complexes are shown except for P308A that gave rise to an unknown higher complex. *AP1* m5 mutant DNA contains base pair C:G instead of A:T at position  $\pm 8$  (see Supplementary Table 3 for full DNA sequences). Phenotype of the wild-type *Arabidopsis* inflorescence **(D)** and *lfy-28* (P308L) mutant inflorescence **(F)** and flower **(E)**. Scale bar is 1 mm on **(D, F)** and 0.5 mm on **(E)**.

therefore provides a plausible explanation for the threshold-dependent floral switch triggered by LFY.

Many transcription factors bind DNA as homodimers but also form heterodimers, thereby extending their spectrum of recognized DNA target sequences (Klemm *et al*, 1998; Garvie and Wolberger, 2001). LFY has been shown to activate the *AG* organ identity gene synergistically with the homeodomain protein WUS (Lohmann *et al*, 2001). As adjacent WUS- and LFY-binding sites are present on the *AG* regulatory sequence, it has been suggested that LFY and WUS could bind simultaneously (Lohmann *et al*, 2001; Hong *et al*, 2003). Preliminary model building indicates that LFY homodimers cannot be accommodated with WUS at adjacent LFY- and WUS-binding sites. This observation raises the intriguing possibility that they might either compete for the same binding sites or more likely could form LFY–WUS heterodimers.

#### LFY shows similarities with HTH proteins

The nature and origin of LFY had so far remained elusive: LFY–C's primary sequence shows unusually strong sequence conservation within its family but has no apparent similarity to any described transcription factor. The crystal structure of LFY–C bound to DNA reveals a seven-helix domain with many residues involved in protein–DNA interactions, tightly

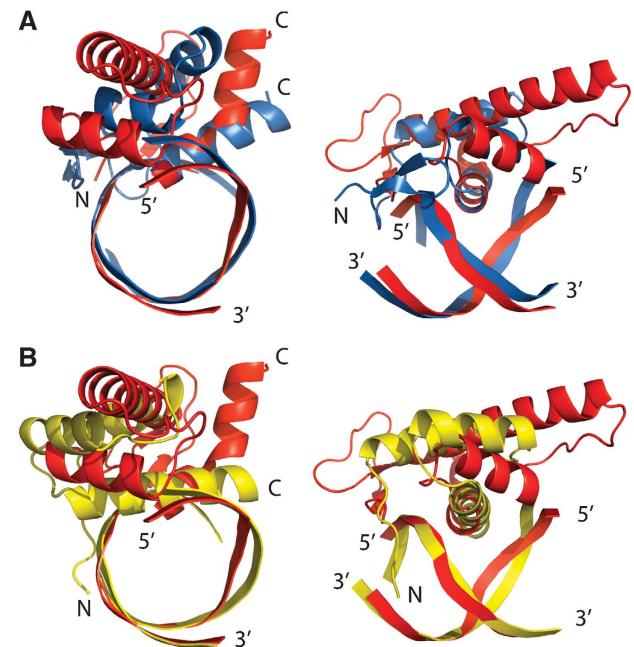
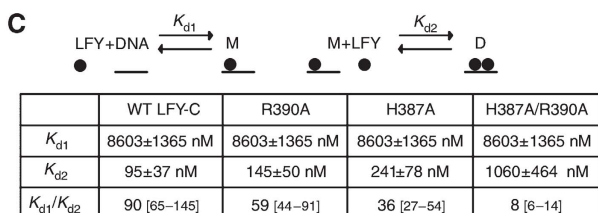
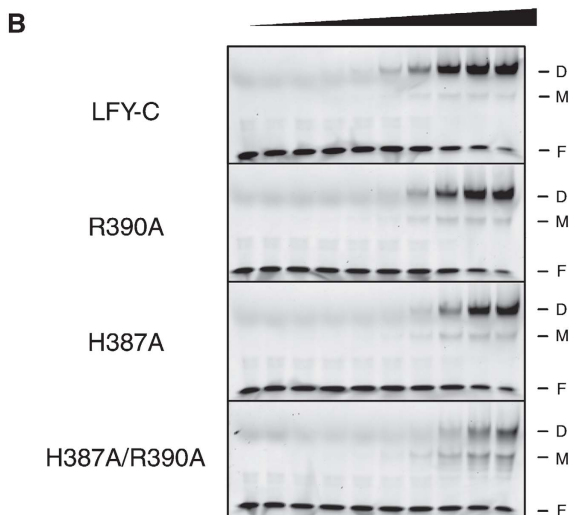
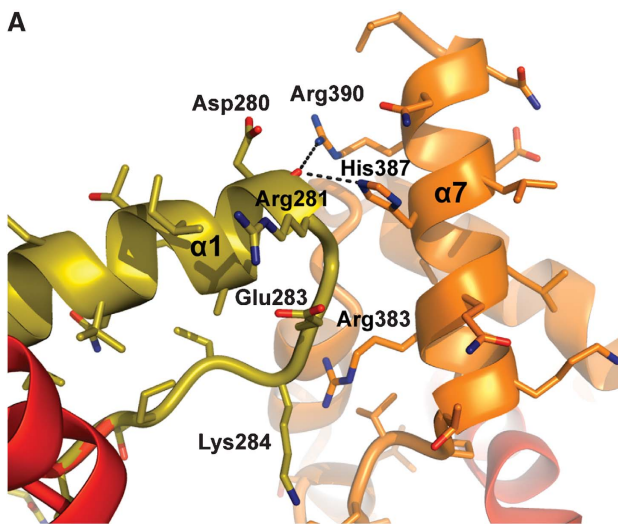
constrained packing interactions in the hydrophobic core and protein–protein interactions with the other monomer. Presumably, these observed tight structural and functional constraints on many residues spread over the entire DNA-binding domain explain the high level of sequence conservation within LFY–C.

The LFY–C structure contains an unpredicted HTH motif formed by helices  $\alpha 2$  and  $\alpha 3$  as part of the overall fold. HTH motifs are present in a wide variety of DNA-binding proteins throughout the three kingdoms of life. They are typically found in a bundle of 3–6  $\alpha$ -helices or combined with  $\beta$ -sheets (winged HTH/fork head domain), which provide a stabilizing hydrophobic core (Weigel and Jackle, 1990; Aravind *et al*, 2005). Comparison of LFY–C against the Protein Data Bank using program DALI (Holm and Sander, 1993) detects similarity of relative short  $\alpha$ -helical segments ( $\sim 60$  amino-acid residues) with different  $\alpha$ -helical proteins including HTH proteins (maximal Dali Z-score 3.0, pairs with  $Z < 2.0$  are structurally dissimilar). A search comprising only the first three N-terminal helices, including the HTH motif, mainly showed similarity to different HTH proteins with slightly higher scores (maximal Dali Z-score: 4.5). When considering just the three helices  $\alpha 1$ ,  $\alpha 2$  and  $\alpha 3$ , LFY aligns well with other three-helix bundle HTH proteins, including the homeodo-

main protein engrailed (r.m.s.d.<sub>40C $\alpha$</sub>  = 2.9 Å), the paired domain (r.m.s.d.<sub>44C $\alpha$</sub>  = 3.5 Å) and the Tc3 transposase (r.m.s.d.<sub>30C $\alpha$</sub>  = 2.4 Å). LFY and partitioning protein KorB (r.m.s.d.<sub>71C $\alpha$</sub>  = 3.7 Å) share some similarity beyond the typical DNA/RNA-binding three-helical bundle core (Russell and Barton, 1992; Khare *et al*, 2004), where five of the seven LFY-C helices, including the HTH motif, roughly superimpose with KorB helices. However, LFY cannot be easily assigned to any of the described classes of HTH proteins (Aravind *et al*, 2005) and it therefore represents a new variant of multi-helical bundle proteins.

The DNA recognition mode of LFY is similar to those observed for the paired domain, Tc3A transposase, Hin recombinase and  $\lambda$  repressor (van Pouderooyen *et al*, 1997; Xu *et al*, 1999). The axis of the recognition helix  $\alpha_3$  in the

HTH of these proteins is oriented parallel to the edges of the nucleotide bases. Only the N terminus of the recognition helix is inserted into the major groove of the DNA, whereas the short helix  $\alpha_2$  has a supporting function. In contrast, in homeodomain proteins, the long probe helix  $\alpha_3$  runs more parallel to the neighbouring DNA phosphate backbone, and mainly the central part of helix  $\alpha_3$  contacts the DNA



**Figure 6** Comparison of LFY-C with paired and homeodomain DNA binding. (A) Two orthogonal views of LFY-C helices  $\alpha_1$ – $\alpha_3$  bound to their DNA target site (red) superimposed with the three-helical bundle core of the N-terminal subdomain of the paired domain of *Drosophila* Prd (blue, PDB-id: 1pdn). (B) Superposition with the homeodomain of *Drosophila* engrailed bound to DNA (yellow, PDB-id: 1hdd), where the centre of recognition helix  $\alpha_3$  inserts into the major groove.

**Figure 5** The LFY-C dimer interface mediates cooperative binding. (A) The dimer interface is viewed perpendicular to the DNA axis. Polar contacts between the two monomers (in orange and olive) are shown with dashed lines. (B) EMSA with increasing concentrations (0, 10, 20, 50, 100, 200, 500, 1000, 2000 and 3000 nM from left to right) of LFY-C wild-type, R390A mutant, H387A mutant, and H387A/R390A double mutant and 50  $\mu$ M *API* DNA. Free DNA (F), monomeric (M) and dimeric (D) complexes are indicated. (C) Estimation of dissociation constants for wild-type LFY-C and three mutant versions (H387A, R390A and H387A/R390A). Binding of LFY-C to *API* DNA was modelled as two equilibrium reactions as detailed in Supplementary data: (1) Binding of a first LFY-C monomer to *API* DNA, leading to the formation of the monomeric complex (M) and characterized by the  $K_{d1}$  dissociation constant; (2) binding of a second LFY-C monomer to M, leading to the formation of the dimeric complex (D) and characterized by  $K_{d2}$ . EMSA signals from (B) were quantified and the corresponding experimental values were fitted with theoretical equations describing the two equilibria. The errors and intervals between square brackets indicated correspond to the 95% confidence interval. An elevated  $K_{d1}/K_{d2}$  ratio reflects a high level of cooperativity, whereas a ratio of 1 would indicate an absence of binding cooperativity. The single mutations resulted in a weak decrease of cooperativity, whereas the H387A/R390A double mutation strongly decreased the cooperativity.

(Figure 6). Similarity between LFY and the paired domain also includes a small two-stranded  $\beta$ -sheet, which precedes the three-helix bundle, and N-terminal residues, which are inserted into the minor groove. However, the minor groove contacting residues are located at the most N-terminal end of LFY-C, whereas in the paired domain they protrude from the loop connecting the two short N-terminal  $\beta$  strands.

Sequence similarities are too weak to suggest a precise evolutionary origin for LFY, although structural resemblances indicate that it might derive from ancestral HTH proteins, including paired and homeodomain proteins (Rosinski and Atchley, 1999; Breitling and Gerber, 2000; Aravind *et al*, 2005). Until now, most plant homeotic genes were found to encode MADS box transcription factors, whereas plant homeodomain proteins rather control meristem homeostasis and cell division (Meyerowitz, 1997; Ng and Yanofsky, 2001). Our study reveals that the LFY master regulator, which determines flower meristem fate and controls the expression of floral organ identity genes, shares structural similarity with other HTH proteins, indicating that this universal DNA-binding motif has also been adopted in plants to trigger major developmental switches.

### Prospects regarding the appearance of angiosperms

The LFY-C structure combined with more than 200 LFY sequences from all types of terrestrial plants offers a unique opportunity to detect key residues in evolution. Some charged LFY-C surface residues (such as Lys253 or Lys254) are strictly conserved, suggesting that they might participate in interactions with other proteins. Other residues are conserved in all angiosperms but not in the non-flowering plants. For example, R390, identified as one of the residues mediating interaction between monomers and cooperative binding, has been conserved in angiosperm LFY proteins, whereas most LFY from non-flowering plants, such as gymnosperms and ferns, show a lysine at this position. This amino-acid change presumably weakens the interaction between monomers and thereby reduces the DNA-binding affinity. The acquisition of R390 might therefore have been important for flower evolution. Because LFY stands at the very centre of the network regulating flower development, it has been proposed that modifications of the LFY gene contributed to the appearance of floral structures in evolution (Albert *et al*, 2002; Frohlich, 2003; Frohlich and Chase, 2007; Theissen and Melzer, 2007). The availability of the LFY-C crystal structure provides a unique framework for generating plausible hypotheses that relate the appearance of angiosperms to specific events during the molecular evolution of LFY. The ‘functional synthesis’ approach that combines phylogeny, biochemical and structural analyses with functional assays *in vivo* (Dean and Thornton, 2007) can now be applied to LFY to try to solve one of the most puzzling enigmas of plant biology: the origin of flowers.

## Materials and methods

### Plant material

The *lfy-28* mutant allele of *A. thaliana* (accession Landsberg *erecta*) was kindly provided by D Weigel (Max Planck Institute, Tübingen, Germany) and originally isolated by J Fletcher (PGENC, Albany). *lfy-28* mutant had been back-crossed twice with the wild type, and individuals showing a mutant phenotype were selected from segregating populations. Plants were grown at 25°C in long days (16 h light).

### Plasmid constructions

**Expression plasmids.** LFY-C (residues 223–424 from *A. thaliana* LFY cDNA) was amplified from pIL-8 (obtained from D Weigel) with Pfu Turbo Polymerase (Stratagene, France) and primers oFP1242 (5′CTCTCGAGCCCGGGCTAGAAACGCAAGTCGTCGCC3′) and oFP1244 (5′CTCTCGAGCCCGGGCTATCCGGTACAGCTAATACCGCC3′), subcloned into pCR-TOPO-BluntII (Invitrogen, Cergy Pontoise, France) and shuttled to pETM-11 (Dummler *et al*, 2005) as *NcoI/XhoI* fragment to yield the pCH28 expression vector. pETM-11 contains an N-terminal 6 × His tag followed by a tobacco etch virus (TEV) cleavage site.

**LFY–GFP plasmid.** A GFP fragment was amplified from pBS–GLFY plasmid obtained from X Wu (Wu *et al*, 2003) using primers oETH1001 5′CCCACTACTGAGAATCTTTATTTTCAGGGCCAGTTCAG TAAAGGAGAAGAAC3′ and oETH1002 5′CCCCAAACCACTACTCCG TTGCCGTTATCTGTTTGTATAGTTTCATCCAT3′. The amplified fragment was subsequently used as a megaprimer to amplify plasmid pCH28 and yield pETH8 (6His–TEV–GFP–LFY-C).

**Expression plasmids for mutant LFY-C.** pCH45 (K307A), pCH46 (N291A), pCH47 (R237A), pCH48 (P308A), pCH49 (D280K), pCH50 (H387A/R390A), pCH54 (H387A), pEDW127 (R390A), pCH55 (*lfy-28*, P308L), pETH21 (*lfy-4*, E238K), pETH23 (*lfy-20*, N306D) and pCH56 (*lfy-5*, P240L) were derived from pCH28 using the megaprimer strategy with appropriate primers (Kirsch and Joly, 1998). All plasmids were verified by sequencing.

### Protein expression, purification and crystallization

Wild-type and mutant LFY-C domains were expressed using *Escherichia coli* strain RosettaBlue(DE3)pLysS (Novagen, Strasbourg, France). After induction by 0.5 mM IPTG, cells were grown overnight at 22°C. For cell lysis, the pellet of 1 l culture was sonicated in 30 ml lysis buffer A (500 mM NaCl, 20 mM Tris–HCl pH 8, 5 mM imidazole, 5% glycerol, 5 mM Tris(2-carboxyethyl)phosphine hydrochloride), one protease inhibitor cocktail tablet Complete EDTA-free (Roche, Meylan, France) and centrifuged for 40 min at 30 000 g. The clear supernatant was incubated for about 1 h with 1 ml Ni-NTA resin (Qiagen, Courtaboeuf, France). The resin was transferred into a column, washed with 20 column volumes (CVs) of buffer A, buffer A + 50 mM imidazole (10 CV) and eluted with buffer A + 380 mM imidazole. The fractions containing the protein were pooled and applied to a Hi-load Superdex-200 16/60 prep grade column (GE Healthcare, Orsay, France) equilibrated with 200 mM NaCl, 20 mM Tris–HCl pH 8, 5 mM dithiothreitol (DTT) to eliminate aggregated proteins by SEC. Protein concentration was estimated using the Bradford assay (Bradford, 1976).

For crystallographic experiments, after elution on the metal-affinity column, the histidine tag was cleaved at 4°C overnight with TEV protease (0.01% w/w, 16 h, 4°C) during the dialysis step against buffer B (500 mM NaCl, 20 mM Tris pH 7.5, 5 mM DTT). The TEV protease, the histidine tag and the uncleaved protein were removed by re-passing the dialysed sample over the Ni-affinity column. The protein was separated from the remaining DNA contamination using the anion-exchange column MonoQ HR10/10 (GE Healthcare) pre-equilibrated in buffer B. Pure protein was recovered in the flow-through, whereas DNA remained bound to the resin. Aggregated protein was removed by SEC with Superdex S75GL column (GE Healthcare) in 200 mM NaCl, 10 mM Tris pH 7.5 and 5 mM DTT. The protein concentration was adjusted to 7.5 mg/ml. DNA oligonucleotides were chemically synthesized and purified by anion-exchange chromatography following established procedures (Cramer and Muller, 1997).

### EMSAs

Single-stranded oligonucleotides, 5′-labelled with tetra-methylcarboxy-rhodamine (Sigma, Saint Quentin Fallavier, France), were annealed to non-fluorescent complementary oligonucleotides in annealing buffer (10 mM Tris pH 7.5, 150 mM NaCl and 1 mM EDTA). The sequences of oligonucleotides used are indicated in Supplementary Table 1. Binding reactions were performed in 20  $\mu$ l binding buffer (150 mM NaCl, 20 mM Tris–HCl pH 7.5, 1% glycerol, 0.25 mM EDTA, 2 mM MgCl<sub>2</sub> and 1 mM DTT) supplemented with 28 ng/ $\mu$ l fish sperm DNA (Roche) and 10 nM double-stranded DNA probe or 140 ng/ $\mu$ l fish sperm DNA for 50 nM DNA probe (Figure 5). Binding reactions were loaded onto native 6% polyacrylamide gels

0.5 × TBE (45 mM Tris, 45 mM boric acid and 1 mM EDTA pH 8) and electrophoresed at 90 V for 80 min at 4°C. Gels were scanned on a Typhoon 9400 scanner (Molecular Dynamics, Sunnyvale, CA; excitation light 532 nm, emission filter 580 BP 30) and signals were quantified using ImageQuant software (Molecular Dynamics). Estimations of  $K_{d1}$  and  $K_{d2}$  (Figure 5; Supplementary Figure 5) were based on the quantifications of binding experiments shown in Figure 5B. The binding model equations used to calculate these  $K_d$  values are explained in detail in Supplementary data.

### SEC

The molecular size of LFY-C/*API* complex was determined using a Superdex-200 10/300GL column (GE Healthcare), equilibrated with buffer containing 150 mM NaCl, 16 mM Tris–HCl pH 7.5, 0.6 mM EDTA and 1 mM DTT, and calibrated with low and high molecular weight protein standards (gel filtration calibration kit; GE Healthcare). Our samples (LFY-C 40 µM, *API* WT 10 µM and LFY-C 40 µM + *API* WT 10 µM) were analysed in the same buffer as protein standards, and molecular size is deduced from the standard curve.

### Analytical SEC and MALLS-SEC

Separation by SEC was carried out with a S200 Superdex column (GE Healthcare). The column was equilibrated in 20 mM Tris–HCl, 150 mM NaCl buffer at pH 7.5. Separations were performed at 20°C with a flow rate of 0.6 ml min<sup>-1</sup>. Protein solution (50 µl) at a concentration of 5 mg ml<sup>-1</sup> was injected. The elution was monitored by using a DAWN-EOS detector with a laser emitting at 690 nm for online MALLS measurement (Wyatt Technology Corp., Santa Barbara, CA), and with a RI2000 detector for online refractive index measurements (Schambeck SFD). Molecular mass calculation was performed as described using the ASTRA software (Gerard *et al*, 2007).

### Crystallization

For co-crystallization using the hanging drop method, protein and DNA duplexes were mixed in a molar ratio of 2:1. The best crystals were obtained at 4°C with 20-mer oligonucleotides bearing complementary A:T overhangs and with 10% PEG 400, 100 mM KCl, 10 mM CaCl<sub>2</sub>, 50 mM HEPES (NaOH) pH 7.0 as reservoir solution. Single crystals grew to a maximal size of 300 × 300 × 500 µm<sup>3</sup> and were stepwise transferred to reservoir solution containing 30% (v/v) glycerol for cryo-protection. For preparation of the mercury derivative, the crystals were soaked in the reservoir solution supplemented with 0.1 mM ethylmercury thiosalicylate (EMTS) for 2 h.

### X-ray structure determination

The crystals of the LFY-C/*API*/DNA complex belong to space group P6<sub>5</sub>22 ( $a = b = 98.8 \text{ \AA}$ ,  $c = 177.4 \text{ \AA}$ ), diffracted up to 2.1 Å resolution and contain half a complex per asymmetric unit. Crystals of the LFY-C/*AG-I*/DNA complex are isomorphous but diffracted slightly weaker (Table I). Diffraction data collected at ESRF beamlines ID14-1, ID29 and ID23-2 were processed using program XDS (Kabsch, 1993). The structure of the LFY-C/*API*/DNA complex was solved using the SIRAS method with EMTS as derivative. The quality of native and derivative data sets is summarized in Table I. Mercury sites were located using program SOLVE (Terwilliger and Berendzen, 1999) and phases were calculated with program SHARP (de la Fortelle and Bricogne, 1997). The experimental electron density map (Supplementary Figure 1) allowed us to automatically build the initial model using program ARP/wARP (Perrakis *et al*, 2001) followed by manually adjusting some side chain conformations

## References

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**: 1948–1954

Albert VA, Oppenheimer DG, Lindqvist C (2002) Pleiotropy, redundancy and the evolution of flowers. *Trends Plant Sci* **7**: 297–301

Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM (2005) The many faces of the helix–turn–helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**: 231–262

Blazquez MA, Ferrandiz C, Madueno F, Parcy F (2006) How floral meristems are built. *Plant Mol Biol* **60**: 855–870

with program COOT (Emsley and Cowtan, 2004) and refinement with program Refmac5 including a TLS refinement with seven groups (Murshudov *et al*, 1997) and later with program Phenix (Adams *et al*, 2002). In space group P6<sub>5</sub>22, the two monomers bound to the pseudo-palindromic DNA duplex are related by a crystallographic dyad. In the crystal, the pseudo-dyad of the DNA coincides with the crystallographic dyad, although the DNA duplexes deviate from strict two-fold symmetry at base pairs 0, ±7, ±9 and the overhanging 5'-end, in the *API* site and at base pairs 0, ±4, ±6 and ±7 and the overhanging 5'-end in the *AG-I* site. To confirm our space group assignment and the underlying assumption that the DNA duplexes used for co-crystallization are randomly distributed in two orientations, the data were reprocessed in the lower symmetry space group P6<sub>5</sub> lacking the dyad, which did not significantly change the  $R_{\text{meas}}$  values. Subsequently, models of the LFY-C dimer bound to the 20-mer DNA duplex were built in space group P6<sub>5</sub> for the *API* and *AG-I* sites and refined in two independent orientations yielding very similar final  $R_{\text{cryst}}$  and  $R_{\text{free}}$  values compared with the refinement in space group P6<sub>5</sub>22. In both orientations (and for both target sites), the final  $F_o - F_c$  electron density maps showed pairs of difference Fourier peaks ( $\sim 7\sigma$ ) of similar height at the non-palindromic bases, indicating that a unique orientation of the DNA duplexes does not correctly describe the situation in the crystals. Finally, simulated-annealing omit maps in space group P6<sub>5</sub> where the non-palindromic bases were omitted showed averaged densities for the omitted bases in both complexes, further confirming the assigned space group P6<sub>5</sub>22.

To account for the two orientations of the DNA in the crystal during the refinement, two nucleotides with 50% occupancy were introduced at the non-palindromic positions. The final model of the LFY-C/*API* complex at 2.1-Å resolution ( $R_{\text{cryst}} = 21.0\%$ ;  $R_{\text{free}} = 23.7\%$ ) comprises residues 237–399 of the LFY DNA-binding domain, whereas the poorly conserved 25 C-terminal residues are disordered. For the refinement of the LFY/*AG-I* complex, the *API* DNA sequence in the LFY/*API* complex was replaced with the *AG-I* sequence. Multiple rounds of refinement (including TLS refinement with seven groups) using program Refmac5 (Murshudov *et al*, 1997) and Phenix (Adams *et al*, 2002) yielded a model with  $R_{\text{cryst}}$  of 22.1% and  $R_{\text{free}}$  of 24.9% using data between 20 and 2.3 Å resolution. The atomic coordinates and structure factors for the LFY/*API* and LFY/*AG-I* complexes have been deposited with the Protein Data Bank under accession codes 2vy1, r2vy1sf and 2vy2, r2vy2sf, respectively.

### Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

## Acknowledgements

We thank D Weigel for providing materials and advice, X Wu for material, L Blanchoin, C Guérin, R Dumas, M Jamin, C Ebel and G Schoehn for help with protein expression and characterization, L Blanchoin, A Maizel, M Blazquez, E Dorcey and C Petosa for critical reading of the paper, R Russell for structure comparisons, the EMBL/ESRF Joint Structural Biology Group for access and support at the ESRF beamlines and the crystallization facility of the Partnership for Structural Biology for support. Funding was provided by ATIP (CNRS) to FP and RB, ATIP + (CNRS) and ANR BLAN-0211 to FP, Region Rhône-Alpes/Cluster 9 to CH, Programme Emergence of the Region Rhône-Alpes to DP.

Bombly K, Wang RL, Ambrose BA, Schmidt RJ, Meeley RB, Doebley J (2003) Duplicate *FLORICAULA/LEAFY* homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. *Development* **130**: 2385–2395

Bowman JL, Alvarez J, Weigel D, Meyerowitz EM, Smyth DR (1993) Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* **119**: 721–743

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* **72**: 248–254

Breitling R, Gerber JK (2000) Origin of the paired domain. *Dev Genes Evol* **210**: 644–650

- Burz DS, Rivera-Pomar R, Jackle H, Hanes SD (1998) Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo. *EMBO J* **17**: 5998–6009
- Busch MA, Bomblies K, Weigel D (1999) Activation of a floral homeotic gene in *Arabidopsis*. *Science* **285**: 585–587
- Chae E, Tan QK, Hill TA, Irish VF (2008) An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development* **135**: 1235–1245
- Cherry JL, Adler FR (2000) How to make a biological switch. *J Theor Biol* **203**: 117–133
- Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* **353**: 31–37
- Coen ES, Romero JM, Doyle S, Elliot R, Murphy G, Carpenter R (1990) *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**: 1311–1322
- Cramer P, Muller CW (1997) Engineering of diffraction-quality crystals of the NF-kappaB P52 homodimer:DNA complex. *FEBS Lett* **405**: 373–377
- de la Fortelle E, Bricogne G (1997) Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* **276**: 472–494
- Dean AM, Thornton JW (2007) Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* **8**: 675–688
- Delano WL (2002) *The PyMOL Molecular Graphics System*. Palo Alto, CA: DeLano Scientific
- Diederichs K, Karplus P (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**: 269–275
- Dummler A, Lawrence AM, de Marco A (2005) Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb Cell Fact* **4**: 34
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126–2132
- Frohlich MW (2003) An evolutionary scenario for the origin of flowers. *Nat Rev Genet* **4**: 559–566
- Frohlich MW, Chase MW (2007) After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* **450**: 1184–1189
- Garvie CW, Wolberger C (2001) Recognition of specific DNA sequences. *Mol Cell* **8**: 937–946
- Gehring WJ, Affolter M, Burglin T (1994) Homeodomain proteins. *Annu Rev Biochem* **63**: 487–526
- Gerard FC, Ribeiro Ede Jr A, Albertini AA, Gutsche I, Zaccai G, Ruigrok RW, Jamin M (2007) Unphosphorylated rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry* **46**: 10328–10338
- Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**: 123–138
- Hong RL, Hamaguchi L, Busch MA, Weigel D (2003) Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell* **15**: 1296–1309
- Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, Crickmore MA, Jacob V, Aggarwal AK, Honig B, Mann RS (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **131**: 530–543
- Kabsch WJ (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* **26**: 795–800
- Khare D, Ziegelin G, Lanka E, Heinemann U (2004) Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. *Nat Struct Mol Biol* **11**: 656–663
- Kirsch RD, Joly E (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res* **26**: 1848–1850
- Klemm JD, Schreiber SL, Crabtree GR (1998) Dimerization as a regulatory mechanism in signal transduction. *Annu Rev Immunol* **16**: 569–592
- Lamb RS, Hill TA, Tan QK, Irish VF (2002) Regulation of *APETALA3* floral homeotic gene expression by meristem identity genes. *Development* **129**: 2079–2086
- Lee I, Wolfe DS, Nilsson O, Weigel D (1997) A LEAFY co-regulator encoded by UNUSUAL FLORAL ORGANS. *Curr Biol* **7**: 95–104
- Lenhard M, Bohnert A, Jurgens G, Laux T (2001) Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between *WUSCHEL* and *AGAMOUS*. *Cell* **105**: 805–814
- Liljegrén SJ, Gustafson-Brown C, Pinyopich A, Ditta GS, Yanofsky MF (1999) Interactions among *APETALA1*, *LEAFY*, and *TERMINAL FLOWER1* specify meristem fate. *Plant Cell* **11**: 1007–1018
- Lohmann JU, Hong RL, Hobe M, Busch MA, Parcy F, Simon R, Weigel D (2001) A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. *Cell* **105**: 793–803
- Lohmann JU, Weigel D (2002) Building beauty: the genetic control of floral patterning. *Dev Cell* **2**: 135–142
- Maizel A, Busch MA, Tanahashi T, Perkovic J, Kato M, Hasebe M, Weigel D (2005) The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* **308**: 260–263
- Meyerowitz EM (1997) Plants and the logic of development. *Genetics* **145**: 5–9
- Molinero-Rosales N, Jamilena M, Zurita S, Gomez P, Capel J, Lozano JU (1999) *FALSIFLORA*, the tomato orthologue of *FLORICAULA* and *LEAFY*, controls flowering time and floral meristem identity. *Plant J* **20**: 685–693
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240–255
- Ng M, Yanofsky MF (2001) Function and evolution of the plant MADS-box gene family. *Nat Rev Genet* **2**: 186–195
- Panne D, Maniatis T, Harrison SC (2004) Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *EMBO J* **23**: 4384–4393
- Parcy F, Nilsson O, Bush MA, Lee I, Weigel D (1998) A genetic framework for floral patterning. *Nature* **395**: 561–566
- Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS (2001) ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr* **57**: 1445–1450
- Riechmann JL, Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. *Curr Opin Plant Biol* **3**: 423–434
- Rosinski JA, Atchley WR (1999) Molecular evolution of helix-turn-helix proteins. *J Mol Evol* **49**: 301–309
- Russell RB, Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* **14**: 309–323
- Schumacher MA, Miller MC, Grkovic S, Brown MH, Skurray RA, Brennan RG (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J* **21**: 1210–1218
- Senear DF, Ross JB, Laue TM (1998) Analysis of protein and DNA-mediated contributions to cooperative assembly of protein-DNA complexes. *Methods* **16**: 3–20
- Shiu SH, Shih MC, Li WH (2005) Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol* **139**: 18–26
- Souer E, van der Krol A, Kloos D, Spelt C, Bliker M, Mol J, Koes R (1998) Genetic control of branching pattern and floral identity during *Petunia* inflorescence development. *Development* **125**: 733–742
- Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55**: 849–861
- Theissen G, Melzer R (2007) Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann Bot (London)* **100**: 603–619
- Underhill DA (2000) Genetic and biochemical diversity in the Pax gene family. *Biochem Cell Biol* **78**: 629–638
- van Pouderooyen G, Ketting RF, Perrakis A, Plasterk RH, Sixma TK (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. *EMBO J* **16**: 6044–6054
- Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM (1992) LEAFY controls floral meristem identity in *Arabidopsis*. *Cell* **69**: 843–859
- Weigel D, Jackle H (1990) The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell* **63**: 455–456
- Wu X, Dinneny JR, Crawford KM, Rhee Y, Citovsky V, Zambryski PC, Weigel D (2003) Modes of intercellular transcription factor movement in the *Arabidopsis* apex. *Development* **130**: 3735–3745
- Xu HE, Rould MA, Xu W, Epstein JA, Maas RL, Pabo CO (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev* **13**: 1263–1275



## Résumé

Le virus de la stomatite vésiculaire (*Vesiculovirus*, VSV) et le virus de la rage (*Lyssavirus*, RV) appartiennent à la famille des *Rhabdoviridae*. Le VSV sert de modèle pour l'étude de la multiplication des virus appartenant à l'ordre des *Mononegavirales* alors que la rage reste un sérieux problème de santé publique causant plus de 50 000 décès par an dans le monde. Le génome de ces virus est composé d'une seule molécule d'ARN simple brin de polarité négative. Pour VSV et RV, cet ARN code cinq protéines virales dont la nucléoprotéine (N) et la phosphoprotéine (P). N s'associe étroitement à l'ARN viral, pour former un complexe N-ARN qui sert de matrice pour la réplication et la transcription virale. P est le cofacteur de la polymérase virale et chaperonne la nucléoprotéine. En interagissant avec les complexes N-ARNs par son domaine C-terminal et avec l'ARN polymérase ARN-dépendante virale (L) par son domaine N-terminal, la phosphoprotéine assure le lien physique entre l'ARN viral et la polymérase. La stœchiométrie de l'oligomère de P, sa structure et son rôle exact dans les processus de transcription et de réplication ont fait l'objet de rapports contradictoires ou ne sont pas connus. Mon travail de thèse a eu pour objectif d'obtenir une caractérisation biophysique et structurale de la phosphoprotéine et des complexes que cette protéine forme avec les complexes N-ARN afin de mieux comprendre la dynamique du complexe de réplication de ces virus.

La mesure de la masse moléculaire et du rayon hydrodynamique par SEC-MALLS-RI (Size Exclusion Chromatography coupled with Multi Angle Laser Light Scattering and refractometry) des phosphoprotéines du RV et du VSV (P RV et P VSV) montre que ces protéines existent exclusivement sous forme de dimère en solution et qu'elles ont une forme allongée. L'analyse bioinformatique des séquences en acides aminés a révélé une organisation modulaire de la phosphoprotéine, similaire à celle observée pour la P des *Paramyxovirus*. Des études biochimiques et biophysiques de mutants de la phosphoprotéine du RV ont confirmé ces résultats. La structure du domaine C-terminal de la P du VSV a été résolue par RMN et montre une homologie avec la structure du domaine C-terminal de la P de RV. Enfin, les modélisations *ab initio* réalisées à partir des données de diffusion des rayons X aux petits angles (SAXS) nous ont permis de confirmer que la phosphoprotéine est un dimère allongé en solution.

Lorsqu'elle est exprimée seule, la nucléoprotéine (N) encapside des ARNs cellulaires pour former des structures hélicoïdales de type nucléocapside, ainsi que des formes annulaires comportant un nombre variable de monomères de nucléoprotéines (9 à 15). La caractérisation de l'interaction entre la P et les anneaux de N-ARN par SEC-MALLS-RI ainsi que leur étude par électrophorèse en conditions natives, a révélé l'existence de deux types de complexes N-ARN-P, contenant respectivement un et deux dimères de P par anneau. La fixation des deux dimères de phosphoprotéine est assurée par leurs domaines C-terminaux. L'analyse par microscopie électronique des complexes N-ARN-P nous a permis d'identifier une densité pouvant correspondre à la fixation de la phosphoprotéine. L'étude par microscopie électronique des complexes nucléocapsides-phosphoprotéine a permis de mettre en évidence un changement de conformation important.

Pour devenir accessible à la polymérase virale, l'ARN génomique du virus de la rage doit se dissocier localement de la 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. L'interaction entre la P et le complexe N-ARN représente potentiellement une nouvelle cible pour le développement d'une thérapie antivirale puisqu'elle n'existe qu'au sein des virus.

**Mots clés :** virus de la stomatite vésiculaire, virus de la rage, virus à ARN négatifs, phosphoprotéine, nucléoprotéine, diffusion statique de la lumière couplée à la chromatographie d'exclusion de taille et à la réfractométrie (MALLS-SEC-RI: Multi Angle Laser Light Scattering coupled with Size Exclusion Chromatography and refractometry), diffusion aux petits angles, modélisations *ab initio*, microscopie électronique, anisotropie de fluorescence, résonance plasmonique de surface.