

Etude biophysique et structurale du complexe de réplication des virus à ARN négatif

Ivan Yavorov Ivanov

▶ To cite this version:

Ivan Yavorov Ivanov. Etude biophysique et structurale du complexe de réplication des virus à ARN négatif. Sciences agricoles. Université de Grenoble, 2011. Français. NNT: 2011GRENV083. tel-01674154

HAL Id: tel-01674154 https://theses.hal.science/tel-01674154

Submitted on 2 Jan 2018 $\,$

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.

UNIVERSITÉ DE GRENOBLE

THÈSE

Soumise pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : BIOCHIMIE ET BIOLOGIE STRUCTURALE

Arrêté ministériel : du 7 Août 2006

Présentée par

Ivan IVANOV

Thèse co-dirigée par **Dr. Marc JAMIN** et par **Dr. Rob RUIGROK**

Réalisée à Unit of Virus Host Cell Interactions et à l'ÉCOLE DOCTORALE CHIMIE ET SCIENCES DU VIVANT

Etude biophysique et structurale du complexe de réplication des virus à ARN négatif

Thèse soutenue publiquement le **02 Décembre 2011**, devant le jury composé de :

Dr. Yves GAUDIN

Laboratoire de Virologie Moléculaire et Structurale, Gif sur Yvette, Rapporteur **Dr. Laurent ROUX** Université de Genève, Faculté de Médecine, Genève, Rapporteur **Dr. Denis GERLIER** INSERM U758 CERVI, Lyon, Examinateur **Dr. Joseph ZACCAI** Institut Laue Langevin, Grenoble, Président **Dr. Marc JAMIN** Unit of Virus Host Cell Interactions, Grenoble, Co-Directeur de thèse **Dr. Rob RUIGROK** Unit of Virus Host Cell Interactions, Grenoble, Co-Directeur de thèse



Résumé :

Les rhabdovirus, dont les virus de la stomatite vésiculaire (VSV) et de la rage (RAV) constituent des prototypes, sont des virus enveloppés dont le génome est constitué d'une seule molécule d'ARN simple brin de polarité négative qui font partie de l'ordre des *Mononegavirales* (MNV). La machinerie de transcription/réplication de ces virus est constituée de l'ARN génomique et de trois protéines qui sont communes à tous les virus de l'ordre des MNV, la (N) qui encapside le génome viral, la grande sous-unité de l'ARN polymérase ARN dépendante (L) et la phosphoprotéine (P) qui est un cofacteur non-catalytique de la L et sert de chaperonne à la N.

Le premier objectif de mon travail de thèse consistait à déterminer la structure cristallographique du domaine de dimérisation de la phosphoprotéine du virus de la rage. La P des rhabdovirus est une protéine modulaire qui contient deux régions intrinsèquement désordonnée, un domaine central responsable de la dimérisation et un domaine C-terminal responsable de la fixation sur la matrice N-ARN. Le modèle atomique obtenu à une résolution de 1.5 Å montre que la structure est très différente de celle du domaine correspondant chez VSV.

Le second objectif de mon travail était la caractérisation structurale de la grande sous-unité L de la polymérase du virus de la stomatite vésiculaire. Cette enzyme de 2109 aa, possède six régions conservées. Le domaine conservé III comprend les régions impliquées dans l'activité de polymérisation et les domaines V et VI sont responsables de la formation de la coiffe des ARNm. Plusieurs stratégies ont été envisagées successivement. (1) Sur la base de prédictions de structures secondaires et de prédictions de désordre, nous avons essayé d'exprimer différents fragments en système d'expression bactérien. Les constructions testées se sont avérées insolubles et certaines d'entre elles fixaient GroEL, indiquant un problème de repliement. (2) Nous avons alors essayé d'exprimer la L seule ou en complexe avec la P en système d'expression eucaryote. La purification s'est avérée impossible, la protéine L restant toujours associée à des protéines cellulaires visibles par coloration au bleu de Coomassie. (3) Finalement nous avons réussi à purifier la polymérase à partir de virus entier. La préparation de la polymérase était très homogène et a permis d'entreprendre une caractérisation par microscopie électronique. Une classification d'images a permis de construire un premier modèle à basse résolution. Le modèle révèle la présence d'un domaine annulaire avec plusieurs domaines structurés attachés au coeur de la polymérase. La cryo-microscopie électronique et la tomographie permettront d'obtenir plus de détails sur cette protéine.

Mots clefs : Virus de la stomatite vésiculaire, virus de la rage, virus à ARN négatif, nucléoprotéine, phosphoprotéine, polymérase, microscopie électronique.

Abstract :

Rhabdoviruses, including vesicular stomatitis virus (VSV) and rabies virus (RAV), are enveloped viruses which genome is made of a single molecule of negative-sense RNA and are classified in the order *Mononegavirales* (MNV). The transcription/replication machinery of these viruses consists of the genomic RNA and of three proteins, which are common to all other viruses of the order MNV, a nucleoprotein (N) that encapsidates the viral genome, a large subunit of the RNA-dependent RNA polymerase (L) and a phosphoprotein (P) that acts as a non-catalytic cofactor of L and a chaperone of N.

The first goal of my research project was to determine the crystallographic structure of the dimerization domain of the rabies virus phosphoprotein. The P protein of the rhabdoviruses is a modular protein, which contains two intrinsically disordered regions, a central dimerization domain and a C-terminal domain involved in binding to the N-RNA template. The atomic model obtained at a resolution of 1.5 Å showed that the structure is different from that of the corresponding domain of VSV.

The second goal was the structural characterization of the large subunit L of VSV polymerase. The enzyme of 2109 aa has six conserved regions. Conserved region III includes the residues involved in the RNA synthesis activity, whereas domains V and VI are involved in mRNA capping formation. Three strategies were successively developed : (1) On the basis of secondary structure and disorder predictions, we tried to express different fragments in bacterial expression systems. These constructions appeared to be insoluble and some of them bound GroEL suggesting a folding problem; (2) We tried to express L alone or co-express it with P in eukaryotic expression system. The purification appeared to be impossible, the L protein always remaining associated with host-cell proteins in amounts detectable by Coomassie staining; (3) We succeeded in purifying the L protein from the virus. The L samples were homogeneous and allowed a characterization by electron microscopy. Image classes allowed the reconstruction of a first low-resolution model. This model revealed the presence of a large ring-like domain and several globular domains. Cryo-electron microscopy and tomography should lead to a more detailed description of this protein.

Keywords : Vesicular stomatitis virus, rabies virus, negative sense RNA virus, nucleoprotein, phosphoprotein, polymerase, electron microscopy.

Remerciements...

L'histoire de la vie d'un chercheur peut parfois se résumer à une rencontre de quelques minutes avec une personne quelconque ou à des opportunités...

Comment je me suis retrouvé dans ce laboratoire est une très, très longue histoire. Mon petit parcours dans le domaine scientifique a commencé avec un petit stage d'un mois en Licence à l'Institut de Biochimie et Génétique Cellulaire à Bordeaux dans le groupe de Dr. Jean Velours. J'ai tout de suite pris connaissance avec des thésards des autres groupes dans ce même institut.

A l'époque je me posais des questions, des questions qui préoccupent à cet âge, beaucoup d'étudiants à l'Université. Que faire après une Licence où on apprend tellement de choses mais l'aspect pratique, on ne le voit pas trop? A la différence de la plupart des gens inscrits à l'Université, moi j'avais en tête que la science me pationne et d'autant plus les virus et leurs fonctionnements. Un des étudiants en thèse, Bertrant, dans l'équipe de Pr. Guy Lauquin m'a dit la phrase suivante : si tu veux faire une thèse en virologie, va à Grenoble dans le groupe de Pr. Rob RUIGROK!!! Après plusieurs discussions avec des scientifiques de Bordeaux, je me suis rendu compte que c'est une très bonne opportunité dans la vie d'un jeune scientifique.

En Master 1 à l'Université Joseph Fourier, dans un module proposé par Pr. Rob RUIGROK et Pr. Patrice MORAND, j'ai découvert le monde de la virologie. Ce module m'a encore plus enthousiasmé à continuer dans ce domaine.

Ma chance a été de pouvoir rencontrer et effectuer mon stage de Master 2 et par la suite ma thèse sous la direction des Prs. Rob RUIGROK et Marc JAMIN, où j'ai pu combiner les approches biophysiques de la recherche avec la virologie. Je tiens à les remercier d'avoir diriger mes travaux avec patience et beaucoup d'attention, pour leur énergie et aussi leur transmission de leur savoir-faire.

Marc, merci de m'avoir toujours soutenu pendant tout mon parcours, de m'avoir donné le courage de poursuivre. Désolé pour toutes les nuits blanches que tu as passées à te soucier de mes travaux sur la grande sous-unité de l'ARN polymérase ARN dépendante et de mon avenir après ma thèse. Un GRAND MERCI pour tout ce que tu m'as appris !!!

Rob, merci de ton soutien moral, l'encadrement et l'intérêt que tu prêtes à mes travaux.

Je voudrais également remercier mon référent de l'Institut Laue Langevin (ILL), Dr. Joseph ZACCAI. Merci Joe pour toutes les discussions scientifiques et ton encadrement pendant mes trois ans de thèse. Merci d'avoir eu la patience à venir avec moi sur les lignes de l'ILL (D22, IN13, IN16 et IN6). Tu es la personne unique dans le bâtiment que je n'ai jamais tutoyée, mais je pense que c'est à cause de l'âge...

Je remercie les Rapporteurs de ce manuscript le Dr. Yves GAUDIN et le Dr. Laurent ROUX d'avoir accepté de juger mon travail scientifique. Je remercie également Denis GERLIER et Joseph ZACCAI d'avoir accepté à faire partie de mon jury de thèse.

Je remercie de directeur de l'EMBL Grenoble et directeur de l'UVHCI, le Dr. Stephen CUSACK de m'avoir accueilli au sein de son institut où règne un environnement scientifique unique grâce au PSB regroupant l'EMBL, l'UVHCI, l'ILL, l'ESRF et l'IBS.

Je suis aussi extrêmement reconnaissant envers Dr. Danielle BLONDEL et Dr. Aurélie ALBERTINI. Danielle, merci pour ta gentillesse, merci de m'avoir transmis ton savoir-faire de Biologie Cellulaire, merci de m'avoir accueilli à Gif. Aurélie, un grand MERCI à toi aussi de m'avoir donné de "trésor de guerre" qui m'a servi de purifier la polymérase de VSV. Merci également pour ton humour, ta gentillesse et pour tous les moments agréables qu'on a passés ensemble. Je dois remercier quatre personnes du fond de mon coeur :

- Dr. Thibaut CREPIN de m'avoir plongé dans le monde de la cristallographie. Merci de m'avoir amené sur les lignes de lumière de l'ESRF, de m'avoir appris à traiter mes données cristallographiques avec XDS, MOSFLM de m'avoir montré les logiciels de collectes de données sur ID14-4. Un GRAND MERCI POUR TOUT.
- Dr. Julien PERARD, de m'avoir plongé dans le monde de la Biologie Moléculaire dès mon arrivée dans le laboratoire. Merci pour tous les conseils que tu m'as donnés pour l'expression de protéines, leurs purifications, et les petits astuces qui sont toujours utiles. Merci de m'avoir formé à faire des expériences en radioactivité, de t'avoir toujours intéressé à mes travaux. Merci pour tous les conseils et tes propositions pour nos expériences difficiles qui préoccupent toujours mon esprit. Un GRAND MERCI POUR TOUT.
- Delphine GUILLIGAY, de m'avoir aidé à faire des clonages pour les expressions des différentes constructions de la polymérase en bactéries et cellules d'insectes. UN GRAND MERCI pour ton support technique.
- Dr. Euripedes RIBEIRO, Orip, tu étais et tu es toujours mon idol de la science. Merci de m'avoir encadré pendant mon Master 2 et ma première année de thèse. Merci pour tous les conseils de Biochimie et Biologie. Pour tout le laboratoire tu étais un support indispensable, tu as toujours la réponse à toute les questions. UN GRAND MERCI ORIP!!! Je pense à toi et à Bettina.

Je voudrais aussi remercier très chaleureusement le microscopiste, Guy SCHOEHN, qui a toujours était très gentil et très patient avec moi. (tu sais que desfois tu fais peur aux gens et au début de mon arrivé au laboratoire, tu étais l'un des premiers à m'avoir fait beaucoup de peur, MAIS je sais que tu es quelqu'un de très gentil...). MERCI GUY, d'avoir toujours prêté beaucoup d'attention à mes échantillons de polymérase, MERCI de pour ton professionnalisme.

Sincères remerciements aussi à Léandro ESTROZI d'avoir toujours été intéressé par mes travaux sur la polymérase de VSV, de m'avoir plongé dans la microscopie et le traitement de signal, merci pour toutes les conversations scientifiques et les jolies cartes de densités électroniques à 4 Åque tu m'as montrées avec une grande patience.

Je tiens à remercier tous les membres du laboratoire 010 : Thibaut, Sylvie, Delphine, Guillaume et Alex.

Je ne manquerai pas de remercier les autres membres de l'EMBL et de l'UVHCI : Fred, Andres, Thomas, Juan, Yan, Simon, Christophe, Max, Filip, David, Pauline, Aurélien, Emilie, Nicolas, Nicolas (le Martinellours), Nicolas TARBOURIECH, Laurence, Marlyse, Sandrine, Elodie, Pascal, Habib, Particia, Céleste, Eric (le petit), Charlotte, Audrey, Chaterine... et bien d'autres...

UN GRAND MERCI aux ex-doctorants et post-doctorants : Euripedes, Francine, Majida, Lucy, Alex, Ambroise et Cédric... et bien d'autres...

Je tiens également à remercier la plateforme EEF de Dr. Imre BERGER, ainsi que la plateforme de microscopie électronique et plus particulièrement Daphna FENEL. Egalement j'aurais souhaité remercier Adrien FAVIER, Dr. Malène JENSEN et Dr. Martin BLACKLEDGE au laboratoire de RMN à l'Institut de Biologie Structurale à Grenoble.

Sincères remerciements à tous les développeurs de LATEX qui m'ont permis d'écrire ce manuscript.

Je voudrais dire aussi un GRAND MERCI à ma famille, mes parents Siyka et Yavor, mon frère Yordan et son épouse Lili, à mes amis et mes proches qui m'ont toujours soutenus dans des moments difficiles.

Je voudrais faire un bisous à mon neveu, Yassen, qui a neuf mois. Le seul conseil que j'ai à te donner, "ne fais pas d'aussi longues études comme ton oncle". Je vous aime TOUS et je vous remercie encore une fois de tout mon coeur.

Je tiens également à remercier mon unique grand-père, 87 ans, qui est toujours vivant, papi tu resteras toujours un MODÈLE de FIERTÉ pour moi.

Enfin un GRAND MERCI du plus profond de mon coeur à ma chérie, Vélina, qui me soutient depuis presque trois ans. Merci de l'amour que tu m'as donné. Malgré la distance qui nous sépare tu trouves toujours les mots qui réchauffent mon coeur !!!

A mes parents, Siyka et Yavor, A mon grand-père, Ivan, A mes grands-mères, Yordanka et Anna, A mon petit canard, Vélina.

Table des matières

1	Intr	roduction	1
	1.1	Introduction générale	2
		1.1.1 Classification des virus	2
		1.1.2 Les virus à ARN négatif	5
	1.2	La transcription et la réplication chez les <i>Mononegavirales</i>	16
		1.2.1 Génome des <i>Rhabdoviridae</i>	16
		1.2.2 La transcription virale \ldots	17
		1.2.3 La réplication virale	20
		1.2.4 Structures des protéines virales	21
	1.3	Objectifs du travail de thèse	42
2	Car	actérisation structurale et fonctionnelle de la phosphoprotéine du virus de la	
	rage	e	43
	2.1	Introduction des phosphoprotéines de Mononegavirales	44
	PAF	PER I : Modular Organization of the Rabies Virus Phosphoprotein	47
	2.2	Conclusion	66
	2.3	Approches envisagées pour la résolution de la structure du domaine central de la	
		phosphoprotéine du virus de la rage	67
		2.3.1 Détermination du groupe d'espace et collecte des jeux de données	69
	2.4	Stratégies utilisées pour obtention des phases	70
		2.4.1 Construction du modèle	72
	PAF	PER II : Structure of the Dimerization Domain of the Rabies Virus Phosphoprotein	76
	2.5	Conclusion	80
3	Exp	$ {\rm pression, purification \ et \ recherche \ de \ domaines \ solubles \ de \ la \ polymérase \ de \ VSV } $	83
	3.1	Introduction	84
	3.2	Expression de fragments en bactéries et cellules d'insectes	85
		3.2.1 Sélection des fragments	85
		3.2.2 Vecteurs d'expressions, milieux de culture et tests à petite échelle	86
		3.2.3 Expression et solubilité à petite échelle	87
	3.3	Purification du fragment N-terminal en fusion avec la MBP	88
		3.3.1 Purification sur chromatographie d'affinité (amylose)	88
		3.3.2 Purification par chromatographie d'exclusion de taille	90
		3.3.3 Purification par colonne Bleu sepharose	90
		3.3.4 Recherche de tampon de stabilisation du domaine N-terminal après clivage à la	
		TEV	91
	3.4	Expression du complexe L-P en cellules d'insectes	93
		3.4.1 Introduction	93
		3.4.2 Clonage du complexe L-P et tests d'expression	93

		$3.4.4 \\ 3.4.5$	Purification du complexe L-P par gel filtration et analyse par Western Blot Caractérisation structurale et fonctionnelle du complexe.	96 97
	3.5	Concl	usions	98
4	Stru	ucture	of the vesicular stomatitis virus L subunit of the RNA-dependent RNA	. 99
	PAF	PER III RNA	I : Structure of the Vesicular Stomatitis Virus L subunit of the RNA-dependent polymerase by Electron Microscopy	101
5	CO	NCLU	USIONS & PERSPECTIVES	113
	$5.1 \\ 5.2$	Organ La gra	ande sous-unité (L) du complexe polymérasique (L-P)	$\begin{array}{c} 114\\ 115 \end{array}$
	5.3	Fixati	ion de la phosphoprotéine à la matrice N-ARN et le mécanisme de transcription	116
	5.4	Fixati	ion du complexe N°P sur la matrice N-ARN et le mécanisme de réplication proposé	110
Bi	ibliog	graphi	e	121
\mathbf{A}	AN	NEXE		141
La	a diff	usion	de neutrons pour l'étude des macromolécules	143
	A.1	Neutr	ons en Biologie	144
		A.1.1	La diffusion des neutrons	144
в	AN	NEXE		155
Pı	rodu	ction o	l'anticorps monoclonaux de VSV	157
С	AN	NEXE		161
	REV	VIEW I	${\tt I}: {\tt Structural insights into the rhabdovirus transcription/replication complex}$	161
	PAF	PER IV	: Structural Disorder in Proteins of the Rhabdoviridae Replication Complex	174
	PAF	'ER V	: Structure, interactions with host cell and functions of rhabdovirus phosphoprotein	183

Table des figures

1.1	Classification des virus.	4
1.2	Image en microscopie électronique des particules virales de virus appartenant	
	à l'ordre des Mononegavirales.	6
1.3	Représentation de la particule virale en forme de balle de révolver des	
	Rhabdoviridae.	7
1.4	Architecture du virion de VSV.	8
1.5	Répartition mondiale du virus de la rage.(Schnell et al. 2010)	10
1.6	Les trois phases du cycle viral des <i>Rhabdoviridae</i> (Schnell et al. 2010)	14
1.7	Organisation schématique du génome des Rhabdoviridae, du virus de la rage	
	(RAV) (Lyssavirus) et du virus de la stomatite vésiculaire (VSV) (Vésiculovirus)	16
1.8	Cinétique du cycle viral du VSV.	19
1.9	Structure de la phosphoprotéine.	22
1.10	Images de microscopie électronique en coloration négative de certaines	
	nucléocapsides de <i>Mononegavirales</i> (Albertini et al. 2005)	24
1.11	Différents complexes de nucléoprotéines-ARN du virus de la rage observés	
	en microscopie électronique en coloration négative (Iseni et al. 1998, Schoehn	
	et al. 2001, Mavrakis et al. 2003, Albertini et al. 2007, 2008)	26
1.12	Images de microscopie électronique en coloration négative de certains	
	nucléocapsides de Mononegavirales (Schoehn et al. 2001)	26
1.13	Structure cristallographique des anneaux N_{11} -ARN du virus de la rage	
	(Albertini et al. 2006)	27
1.14	Comparaison des structures et conservation de séquence des nucléoprotéines	
	du virus de la rage et du VSV (Luo et al. 2007, Ivanov et al. 2011)	28
1.15	Comparaison des structures de nucléoprotéines de différents virus à ARN	
	négatif (Ruigrok et al. 2011)	29
1.16	Structure cristallographique du complexe N°P.	30
1.17	Schéma détaillé des régions conservées de la polymérase (L) du VSV et les	
	structures correspondantes à basse résolution.	32
1.18	Alignement de séquence de la région conservée III des polymérases de	
	Rhabdovirus.	33
1.19	Mécanisme conventionnel et non conventionnel de formation de coiffe des	
	ARNm. (Ogino & Banerjee 2011)	35
1.20	Alignement de séquence de la région conservée V des polymérases de	
	Rhabdovirus.	36
1.21	La tétrade impliquée dans la formation de coiffe (Li et al. 2008)	37
1.22	Alignement de séquence de la région conservée VI des polymérases de	
	Rhabdovirus.	37
1.23	Comparaison des domaines Méthyltransférase de <i>E.coli</i> et du VSV	38
1.24	Le complexe de transcriptase et réplicase du VSV (Qanungo et al. 2004)	41

2.1	Organisation modulaire de la phosphoprotéine de <i>Rhabdoviridae</i> (Takacs et al. 1992).	44
$2.2 \\ 2.3$	Alignements de séquences de phosphoprotéines de Lyssavirus et Vesiculovirus. Caractérisation du domaine de dimérisation de la phosphoprotéine du virus	66
	de la rage.	67
$2.4 \\ 2.5$	Spectre RMN HSQC ¹⁵ N-H du domaine de dimérisation	68
-	manuellement.	69
2.6	Carte de densité électronique expérimentale.	74
2.7	Structure des phosphoprotéines des Rhabdovirus	80
$3.1 \\ 3.2$	Recherche de domaines solubles	85
	F 410	88
3.3	Profil de gel filtration et analyse par gel SDS-PAGE du fragment MBP-F14-	
3.4	F410	90
	MBP-F14-F410.	91
3.5	Vecteur de clonage et suivi de l'expression du complexe L-P par mesure de	
	YFP	93
$3.6 \\ 3.7$	Purification du complexe L-P par chromatographie d'affinité Ni-NTA Chromatographie d'exclusion de taille et analyse du complexe L-P par	95
	Western Blot.	96
5.1	Modèle de transcription/réplication des Rhabdovirus	118
A.1	Diffusion cohérente de certains atomes.	145
A.2	Longueurs de diffusion de différents composants en fonction du pourcentage	
	de D_2O	146
A.3	Données expérimentales de diffusion de neutrons aux petits angles du complexe N^0P	1/18
A 4	Différents types de modèles utilisés lors de la modélisation du complexe N^0 -P	148
A 5	Section efficace <i>incohérente</i> des atomes constituant les protéines	150
A.6	Principe du spectromètre à rétrodiffusion IN13 de l'ILL.	151
A.7	Représentation schématique de la phosphoprotéine et son domaine C-terminal.	152
A.8	Courbes de diffusion élastique incohérente sur IN13 du domaine C-terminal	
	et la phosphoprotéine entière.	153
A.9	Evolution des déplacements carrés moyens en fonction de la température 1	153

x

Liste des tableaux

1.1	Tableau des virus d'ARN négatif non-ségmenté de l'ordre de MNV infectant				
	l'homme et leurs caracteristiques				
2.1	Statistiques relatives aux collectes de données				
3.1	Tests de solubilisation des constructions polymérasiques en bactéries et				
	cellules d'insectes				
3.2	Matrice de stabilisation de la construction F14-F410				
A.1	Sections efficaces et longueurs de diffusion des principaux éléments composant les protéines. 145				

Liste des Abréviations

βME :	β Mercapto Ethanol						
3D :	3 Dimension						
ARNm :	\mathbf{A} cide \mathbf{R} ibo \mathbf{N} ucléique messager						
CK-II :	Casein Kinase II						
CTD :	C Terminal Domaine						
CV:	Column Volume						
FIP :	French beamline for Investigation of Proteins						
HSQC :	Heteronuclear Single Quantum Coherence						
IDR_{CT} :	Intrinsically Disordered Region C-Terminal						
MBP :	Maltose Binding Protein						
MoRE :	Molecular Recognition Element						
MoRE :	Molecular Recognition Element						
NC :	NucléoCapside						
NTD :	N Terminal Domaine						
\mathbf{P}_{NTR} :	P N-Terminal Region						
PH :	Polyhedrin						
PKC :	Protein Kinase C						
RE :	\mathbf{R} éticulum Endoplasmique						
RMN :	Résonance Magnétique Nucléaire						
rmsd :	root mean square deviation						
RVPK :	Rabies Virus Protein Kinase						
SAD :	Single Anomalous Dispersion						
SAM :	\mathbf{S} -Adenosine Methionine						
SAXS :	Small Angle X-Ray Scattering						
SEC-MALLS-R	\mathbf{RI} : Size Exclusion Chromatography-Multi Angle Laser Light Scattering Refractometry						
SV5:	Simien Virus 5						
YFP :	Yellow Fluorescent Protein						

Alanine	А	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	Κ	Lys
Asparagine	Ν	Asn	Methionine	Μ	Met
Aspartic acid	D	Asp	Phenylalanine	F	Phe
Cysteine	С	Cys	Proline	Р	Pro
Glutamic acid	Ε	Glu	Serine	\mathbf{S}	Ser
Glutamine	Q	Gln	Threonine	Т	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	Η	His	Tyrosine	Υ	Tyr
Isoleucine	Ι	Ile	Valine	V	Val

Code à une et trois lettres pour les acides aminés :

Chapitre 1 Introduction

"Ayez le culte de l'esprit critique" (Louis Pasteur, Discours d'inauguration de l'Institut Pasteur, 14 november 1888)

Sommaire

1.1 Intr	oduction générale
1.1.1	Classification des virus
1.1.2	Les virus à ARN négatif
1.2 La t	ranscription et la réplication chez les Mononegavirales 16
1.2.1	Génome des Rhabdoviridae 16
1.2.2	La transcription virale
1.2.3	La réplication virale
1.2.4	Structures des protéines virales
1.3 Obj	ectifs du travail de thèse

1.1 Introduction générale

1.1.1 Classification des virus

De nos jours on dénombre plus de 4000 virus, se distinguant les uns des autres par leur forme, leur taille, la nature de leur matériel génétique qui peut être de l'ADN monocaténaire ou bicaténaire (ciruclaires ou linaires), ou de l'ARN monocaténaire (de polarité positive ou négative) ou bicaténaire (linéaire ou ségmenté). Ils se distinguent également par la symétrie de la capside ou de la nucléocapside (icosaédrique par exemple pour l'adénovirus, ou hélicoïdale pour les nucléocapsides des virus à ARN négatif), et la présence ou l'absence d'enveloppe.

Il existe deux types de systèmes de classifications : la classification hiérarchique proposée par Lwoff qui se base sur des critères morphologiques et la classification de Baltimore qui repose sur la nature de l'acide nucléique utilisée comme génome http://www.nlv.ch/Virologytutorials/Classification. htm.

Dans la classification par hiérarchie proposée en 1962 par André Lwoff, R.W. Horne and P. Tournier (LWOFF et al. 1962*a*, Lwoff et al. 1962*b*), les virus sont classés dans l'ordre suivant : phylum - classe - ordre - famille - sous-famille - genre - espèce - souche/type. Cette classification virale se base sur la similarité des propriétés virales et non pas sur les propriétés des organismes qu'ils infectent. Quatre caractéristiques sont utilisées :

- la nature de l'acide nucléique : ADN ou ARN
- la symétrie de la capside
- la présence ou absence d'enveloppe
- les dimensions du virion et de la capside

Pour le moment les quatre points cités plus haut sont appliqués pour classer les virus en genres et familles au bas de l'échelle de classification. Toutes les familles portent le suffixe « -viridae » e.g. *Calciviridae*, *Picornaviridae*, *Rhabdoviridae*. Le genre porte le suffixe « -virus ».

Le système de classification virale de Baltimore est basé sur les mécanismes que les virus utilisent pour se répliquer dans la cellule hôte. Le thème central de cette classification est que tous les virus afin de produire les protéines virales doivent fabriquer à partir de leurs génomes des ARN messagers (mARN) de polarité positive. Les mécanismes différents utilisés par les virus permettent de les classer en sept groupes chacun utilisant différentes stratégies de réplication.

• Les virus à **ADN double brin** (Adénovirus, Herpesvirus, Poxvirus, etc.) Au sein de ces différentes familles, les modes de réplications sont aussi différents : Les adénovirus se répliquent dans le noyau. Les Poxvirus se répliquent dans le cytoplasme.

• Les virus à ADN simple brin de polarité positive (Parvovirus)

La réplication se fait dans le noyau impliquant la formation de brins de polarité négative qui servent par la suite pour la fabrication d'un ARN de polarité positive et de nouveaux ADN génomiques.

- Les virus à ARN double brin (Réovirus, Birnavirus)
 Ces virus ont des génomes ségmentés. Chaque segment est transcrit séparément pour produire un ARNm monocistronique.
- Les virus à **ARN simple brin de polarité positive** (Picornavirus, Togavirus, etc.) Les picornavirus ou le virus de l'Hépatite A ont des ARNm polycistroniques. Leur génome viral est l'ARNm. La traduction directe résulte dans la formation d'une polyprotéine qui est clivée par la suite pour former des protéines matures.
- Les virus à **ARN simple brin de polarité négative** (Orthomyxovirus, Rhabdovirus etc.) Ces virus doivent posséder une particule virale avec une ARN polymérase ARN dépendante. Les orthomyxovirus ségmentés ou bien les rhabdovirus non ségmentés doivent transcrire l'ARN de polarité négative pour produire des ARNm monocistroniques.
- Les virus à **ARN simple brin de polarité positive avec un intermédiaire ADN** (Rétrovirus)

Les virus possèdent une transcriptase inverse qui permet de transformer l'ARN de polarité positive en matrice d'ADN.

• Les virus à **ADN double brin avec un intermédiaire d'ARN** (Hepadnavirus) Les virus de ce groupe possèdent également une transcriptase inverse, mais chez les hepadnavirus la réplication du génome se produit à l'intérieur de la capside virale.



FIGURE 1.1 – Classification des virus.

A : Classification virale d'après « Lwoff » selon leurs propriétés génomiques, leurs formes, leurs hôtes et leurs processus de réplication et de transcription.

B: Classification des virus d'après Baltimore. Les images sont issues de http://www.nlv.ch/Virologytutorials/Replication.htm

1.1.2 Les virus à ARN négatif

1.1.2.1 Généralités sur les virus à ARN négatif

Le génome des virus à ARN négatif est constitué d'un brin d'ARN de polarité négative complémentaire à l'ARN messager. Le génome des ces virus est constitué d'une seule molécule d'ARN (*Mononegavirales*) ou de plusieurs molécules séparées (virus à ARN ségmenté). Les familles des *Rhabdoviridae* (virus de la rage), *Paramyxoviridae* (virus de la rougeole, virus respiratoire syncytial, virus Hendra et Nipah), *Filoviridae* (virus Ebola) et *Bornaviridae* (Bornavirus) appartiennent à l'ordre des *Mononegavirales*. Ces virus présentent des similarités d'organisation de leur génome et leurs mécanismes de transcription et de réplication sont similaires (Pringle 1997). Le génome des virus des familles des *Arenaviridae* (2 segments d'ARN, virus Lassa), *Bunyaviridae* (3 segments, virus de la fièvre de la vallée du rift) et *Orthomyxoviridae* (7-8 segments, virus de la grippe) sont constitués de plusieurs molécules d'ARN simple brin de polarité négative.

Les virus à ARN négatif sont des virus enveloppés qui ont des morphologies variables, infectent un large spectre d'hôtes (plantes, invertébrés, mammifères) et sont à l'origine de nombreuses pathologies humaines (tableau 1.1, page 5, figure 1.2, page 6)

Dans le tableau ci-dessous sont présentés différents virus à ARN négatif avec leurs caractéristiques correspondantes :

	CLASSIFICATION	VI	RAL PROPERTIES		
groupe	virus disease	shape	size (nm)	size (kb or kpb)	genes
Filoviridae	Ebola virus, Marburg virus	filamentous, pleomorphic	790-970*90	18-19	7
Paramyxoviridae	Nipah, Hendra, Mumps	pleomorphic	150-300	13-16	8
Rhabdoviridae	Rabies virus	bullet shaped	180 x 75	11-15	5
Bornaviridae	Bornavirus	spherical	70-130	8-9	6

TABLE 1.1 – Tableau des virus d'ARN négatif non-ségmenté de l'ordre de MNV infectant l'homme et leurs caractéristiques.

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

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

Les morphologies des virus appartenant à l'ordre des *Mononegavirales* sont représentées sur la figure 1.2.





Les images proviennent de la galerie de l'ICTVdB (International Committee on Taxonomy of Viruses database), (Buchen-Osmond 2003)

(a) : (virus Ebola *Filoviridae*) (b) : (virus Parainfluenza *Paramyxoviridae*) (c) : (virus de la stomatite vésiculaire *Rhabdoviridae*) (d) : (virus de Borna *Bornaviridae*)

D'un point de vue structural, tous les virus de l'ordre des *Mononegavirales* ont cinq gènes codant pour cinq protéines structurales communes à tous les virus appartenant à cet ordre. L'ARN viral n'est jamais nu, ni dans les particules virales ni dans les cellules infectées mais toujours sous forme de complexe ribonucléoprotéique. La protéine majoritaire de ce complexe est la nucléoprotéine (N) qui recouvre entièrement l'ARN viral de polarité négative et l'ARN complémentaire de polarité positive. Chez les *Mononegavirales*, deux autres protéines sont également associées à la nucléocapside viral : l'ARN polymérase ARN dépendante (L) et son cofacteur (P) (Curran et al. 1994).

1.1.2.2 Les Rhabdoviridae

1. Classification

La famille des *Rhabdoviridae* comporte plus de 200 virus infectant des mammifères, des poissons, des crustacés, des insectes, des reptiles et des plantes (Fields 1996). La famille des *Rhabdoviridae* contient six genres :

- Lyssavirus : virus de la rage
- Vesiculovirus : virus de la stomatite vésiculaire
- Ephemerovirus : fièvre éphémère bovine
- Novirhabdovirus : nécrose infectieuse hématopoïétique
- Cytorhabdovirus : necrose jaune de la laitue
- Nucleorhabdovirus : mosaïque du maïs

Parmis les six genres, les genres Vesiculovirus, Lyssavirus et Ephemerovirus infectent des animaux, les genres Cytorhabdovirus et Nucleorhabdovirus infectent des végétaux alors que les Novirhabdovirus infectent de nombreux poissons. Il semblerait qu'un grand nombre d'entre eux soient véhiculés par des insectes. (Ammar & Hogenhout 2008, Ammar et al. 2009, Hogenhout et al. 2003). Le prototype du genre Vesiculovirus est le virus de la stomatite vésiculaire qui est le sujet principal de mon étude. Il existe deux sérotypes majeurs : Indiana et New Jersey. Ce virus infecte le bétail et provoque des lésions vésiculaires bénignes qui ont néanmoins des conséquences économiques importantes dans les élevages. Il existe d'autres Vesiculovirus capables de provoquer des lésions similaires chez les animaux et d'infecter l'homme : le virus Piry (Brun et al. 1995), le virus Chandipura (initialement isolé chez l'homme en Inde) (Rao et al. 2004).

Le genre Lyssavirus est divisé en sept sérotypes sur la base d'analyse de séquence de la nucléoprotéine (Bourhy et al. 1993, Badrane & Tordo 2001) mais aussi plus récemment sur la base d'analyse du génome entier (Delmas et al. 2008). Le sérotype 1 englobe toutes les souches de rage alors que les autres sérotypes correspondent à des virus apparentés à la rage (*Lagos Bat* : sérotype 2); (*Mokola* : sérotype 3), (*Duvenhage* : sérotype 4); (EBVL1 *European Bat LyssaVirus* : sérotype 5); (EBLV2 : sérotype 6); (ABLV *Australian Bat Lyssavirus* : sérotype 7). Tous les sérotype, à l'exception du sérotype 3, pour lequel l'espèce hôte reste inconnue, ont des chiroptères comme réservoir naturel (Sabeta et al. 2007). Récemment, quatre sérotypes additionnels, qui infectent des chauves-souris en Asie centrale et du sud-est, ont été ajoutés : les virus Aravan, Khujand, Irkut et West Caucacian Bat (Kuzmin et al. 2008).

2. Morphologie

Les virus appartenant à la famille des rhabdovirus partagent la même organisation structurale et génomique ainsi que des modes similaires de réplication et transcription de leur ARN (Knipe 2006). Ce sont des virus enveloppés dont la membrane lipidique dérive de la membrane de la cellule hôte infectée. La particule virale des *Rhabdoviridae* possède une forme caractéristique évoquant une forme d'obus (« bullet shape ») (figure 1.3 et figure 1.4 page 8) (Ge et al. 2010, Libersou S et al. 2010).



FIGURE 1.3 – Représentation de la particule virale en forme de balle de révolver des *Rhabdoviridae*.

A : Représentation schématique de la particule virale des *Rhabdoviridae* (Schnell et al. 2010). Les protéines P et L, qui sont fixées sur la face interne de la nucléocapside virale, sont néanmoins représentées sur la partie externe pour plus de clarté. B : Image des particules virales du virus de la stomatite vésiculaire prises en microscopie électronique par Pr. Rob Ruigrok. La barre d'échelle correspond à une taille de 100 nm.

L'une des extrémités du virus est plate alors que l'autre est de forme conique. Les rhabdovirus de

plante sont bacilliformes. Le diamètre des particules virales est aux alentours de 80 nm et la longueur entre 180-300 nm (Howatson & Whitmore 1962). Ces virus sont composés de 74% de protéines, 20% de lipides, 3% de carbohydrates et 3% d'ARN (Thomas et al. 1985).



FIGURE 1.4 – Architecture du virion de VSV.

A : Représente une moyennation de classes 2 D du virion entier.

B : Représente un modèle de l'extrémité conique et du tronc obtenu par reconstruction en cryo-microscopie électronique. La protéine N est représentée en vert, M en bleu, et les faces internes (2) et externes (1) de la membrane sont en violet et en rose respectivement. En insert, illustration de la base du virion. Le "X" marque l'absence d'un tour d'hélice de M sous le dernier tour de l'hélice de N.

C : Représente une coupe transversale du tronc de la particule de VSV. 10 tours d'hélice ont été reconstruit à partir de la carte de densité électronique déposée dans la databank de microscopie électronique sous le nom (EMD-1663). Le fit de la nucléoprotéine représentée en vert et l'ARN en bleu dans la densité électronique a été effectué avec VEDA (interface graphique d'UROX (Navaza et al. 2002)). La carte de microscopie électronique est représentée avec un σ de 2.9.

Les cinq protéines communes à tous les membres de cette famille jouent les rôles clefs dans l'organisation du virion, dans les étapes d'entrée et de sortie de la cellule hôte et dans la réplication proprement dite du virion. La structure interne de la particule virale est constituée de trois protéines virales qui s'associent à l'ARN génomique afin de former un complexe hélicoïdal nommé nucléocapside (NC) (Newcomb et al. 1982) qui constitue l'unité infectieuse du virus. L'ARN n'est jamais nu, ni dans les particules virales ni dans les cellules infectées, mais il est étroitement associé à la nucléoprotéine (N) (Albertini et al. 2006, Green et al. 2006) et c'est ce complexe N-ARN qui sert de matrice pour la réplication et la transcription (Arnheiter et al. 1985). Une fois le virus entré dans la cellule, la synthèse des ARNm se fait à partir de la matrice N-ARN viral. Deux protéines sont nécessaires pour cette réaction : l'ARN polymérase ARN dépendante (L) qui possède les activités enzymatiques de réplication et de transcription (Emerson & Yu 1975, Pattnaik & Wertz 1990) et son cofacteur la phosphoprotéine

(P). Ces deux protéines L-P constituent un complexe polymérasique à deux sous-unités.

La glycoprotéine (G) est enchâssée dans la membrane lipidique (David 1973) et forme des spicules trimériques que l'on peut observer en microscopie électronique (Doms et al. 1987, Gaudin et al. 1992). Elle reconnaît le récepteur cellulaire, induit l'endocytose du virion et permet la fusion des membranes endosomales et virales. La cinquième protéine est la protéine matrice (M) qui a le rôle de stabiliser la nucléoprotéine sous forme condensée (Newcomb et al. 1982).

Pour VSV, chaque particule virale est constituée d'environ 50 copies de L, 460 copies de P, 1250 copies de N, 1820 de M et 1200 monomères de G (Thomas et al. 1985, Banerjee 1987).

Récemment, la structure quasi-atomique des virions de VSV a été résolue par cryo-microscopie électronique (Ge et al. 2010), révélant l'architecture du virion et l'agencement relatif des protéines N et M au sein de la particule virale (figure 1.4 B page 8). La reconstruction tridimensionnelle révèle la présence de 37.5 sous-unités de nucléoprotéine par tour d'hélice au sein du tronc de la particule virale. Malheureusement, les protéines G présentes à la surface du virion, ainsi que les protéines P et L associées aux nucléocapsides n'ont pas pu être reconstruites car elles ne respectent pas la même symétrie hélicoïdale. La reconstruction de la pointe n'a été faite non plus.

La particule virale avec une longueur aux alentours de 175 nm et un diamètre interne de 31 nm (figure 1.4 B,C page 8) est suffisamment grande pour accueillir 50 copies de polymérase, 460 copies de P et approximativement 500 molécules de M qui ne font pas partie de la couche intermédiaire dans la reconstruction par cryo-microscopie électronique (Ge et al. 2010).

1.1.2.3 Le virus de la rage

La rage est une maladie ancienne, décrite depuis l'Antiquité, dont l'origine infectieuse a été proposée au XIX^{ième} siècle. Le virus de la rage est un représentant du genre Lyssavirus et sert de modèle d'étude pour les autres membres de la famille des *Rhabdoviridae*. La rage est une zoonose virale affectant les mammifères terrestres qui est répandue dans le monde (plus de 2.5 milliards de personnes vivent dans des régions endémiques de rage) (figure 1.5 page 10) (Haupt 1999).

Le virus de la rage est un virus neurotrope provoquant des infections du cerveau toujours mortelles une fois déclarées. La rage est transmise à l'homme principalement par le chien qui constitue le vecteur principal du virus dans le monde. Chez l'homme, des contaminations rabiques sont le plus souvent dues à un contact (morsure) avec la salive de l'animal enragé, toutefois, le virus peut pénétrer dans l'organisme par griffure, excoriations cutanées ou léchage d'une plaie ouverte. D'autres contaminations exceptionnelles ont été décrites, par greffe d'organe, provenant de donneurs contaminés (Gode & Bhide 1988, Srinivasan et al. 2005).

Pendant la période d'incubation, le virus se propage dans l'organisme en cheminant le long des neurones par voie axonale centripète (Tirawatnpong et al. 1989). Cette période a une durée variable, dépendante de la quantité de l'inoculum et de la distance entre le site de morsure et le SNC. Le virus remonte le long des nerfs périphériques puis de la moëlle épinière pour finalement aboutir au cerveau. Dans une deuxième phase, le virus se propage par voie axonale centrifuge vers divers organes préférentiels tels que les reins, la cornée et les glandes salivaires. Chez les personnes décédées à la suite d'une infection par le virus de la rage, de hauts titres viraux sont retrouvés dans ces organes mais aussi dans les muqueuses. Selon la zone du cerveau infectée, les dégâts causés par le virus peuvent induire un comportement agressif chez l'animal ou l'homme qui permet la transmission du virus. L'animal enragé va disséminer le virus à un autre hôte par morsure.





A gauche, la distribution mondiale du virus classique (sérotype 1) en 2007 est représentée en orange. Les régions comprenant uniquement des Lyssavirus appartenant aux sérotypes 2 à 7 sont colorées en jaune, et les régions indemne de lyssavirus sont en vert. A droite répartition du nombre de morts imputables à la rage en 2004 en Asie.

Deux formes cliniques de la rage existent chez l'homme (Hemachudha et al. 2002). La forme spastique dite rage furieuse représente 70% des cas. Les symptômes incluent des troubles du comportement, une hyperactivité, des hallucinations et une hydrophobie - impliquant une aérophobie, une dysphagie et une salivation abondante. La phase aigüe dure de 2 à 7 jours puis le patient entre dans un coma et meurt par paralysie du système cardiovasculaire en environ une semaine. La forme paralytique dite rage muette représente 30% des cas. De ces 30% seulement 50% des personnes infectées présentent les symptômes d'hydrophobie et d'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.

Dans la nature, le réservoir du virus sont les chauve-souris. (McColl et al. 2000). La rage est très rarement transmise aux humains directement à partir des chauve-souris, sauf lors de contacts directs avec les animaux ou de contacts avec des aérosols dans les grottes où vivent les animaux qui sont porteurs sains du virus. Les espèces insectivores sont des vecteurs de la 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 des chauves-souris hématophages transmettent la rage au bétail (Kobayashi et al. 2008). Les chiroptères pouvant excréter le virus dans leur salive pendant de longues périodes sont des agents de contamination potentiellement redoutables et impossibles à vacciner.

Divers animaux sauvages constituent un réservoir secondaire de la maladie. A leur contact, les animaux domestiques, les chiens voire l'homme peuvent se contaminer. Selon les régions, de nombreux vecteurs sont retrouvés, tels que le chacal, le raton-laveur, la mouflette, le renard, le loup et autres. Le virus est présent dans la salive 3 à 5 jours avant l'apparition des symptômes neurologiques. L'animal meurt dans les 15 jours qui suivent la déclaration de la maladie. Il est probable que des Lyssavirus infectant les chauve-souris soient entrés en contact avec la population de renards et aient établi un cycle réplicatif dans ce nouvel hôte par une série de mutations, comme le suggèrent des analyses phylogénétiques réalisées sur la glycoprotéines de surface responsable de l'entrée du virus dans les cellules hôtes (Badrane & Tordo 2001). Une adaptation similaire pourrait avoir eu lieu aux Etats-Unis, où la rage affecte les ratons laveurs et les mouflettes (Rupprecht et al. 1995).

Diagnostic de la rage

Aldechi Negri fut le premier à identifier en 1903 la présence de corps d'inclusion visibles dans le cytoplasme des cellules infectées, correspondant à des accumulations de protéines virales que l'on a appelé des corps de Negri (Miyamoto 1965). Cependant la détection de ces structures n'est pas fiable à 100% et d'autres techniques de détections comme la PCR et des tests immunoenzymatiques sont actuellement utilisés.

Prévention, vaccination et traitement de la rage

A l'heure actuelle il n'existe pas de traitement curatif contre la rage. La vaccination reste la seule protection efficace. Le vaccin contre la rage fut développé par Louis Pasteur à partir de moëlle épinière de lapin. Aujourd'hui des progrès considérables ont été réalisés dans la préparation des vaccins antirabiques. Ils sont réalisés en culture cellulaire ou sur des oeufs embryonnés de poulet. Ensuite le virus est concentré et inactivé. Le traitement post-exposition consiste à une injection d'immunoglobuline antirabique. Ce traitement reste cher mais efficace lorsque la maladie est traitée à temps.

Le virus de la rage comme outil pour l'étude du système nerveux central

Le caractère neurotropique du virus rabique, ses capacités de transport rétrograde trans-synaptique, la relative lenteur de sa progression au sein du système nerveux central combinés aux possibilités d'études histologiques basées sur l'immunomarquage fluorescent (Ugolini 1995) en font un outil de choix pour diverses études neuroanatomiques, permettant notamment de mettre en évidence l'architecture des connexions entre chaînes neuronales (Kelly & Strick 2000). Ces propriétés ont été utilisées avantageusement dans l'étude de certains processus neurologiques (Astic et al. 1993, Tang et al. 1999, Ruigrok et al. 2008).

1.1.2.4 Le virus de la stomatite vésiculaire (VSV)

Cette maladie du bétail a été décrite aux Etats-Unis aux XIX^{ième} siècle (Hanson 1952). Elle se localise principalement en Amérique du Nord et du Sud, toutefois, des épidémies de plus faible ampleur ont été décrites chez les chevaux en France et en Amérique du Sud à la fin du XIX^{ième} et au début du XX^{ième} siècle. Le virus a été isolé par Cotton (Cotton 1927) et la morphologie virale a été décrite par Chow (Chow 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, la bouche et les sabots des animaux. Les animaux éprouvent des difficultés à se nourrir et se mouvoir ce qui entraîne une perte de poids. 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 et se propagent rapidement dans le troupeau, mais pas nécessairement aux fermes adjacentes (Sellers & Maarouf 1990). Les tentatives de développement de vaccin à base d'ADN ou utilisant la glycoprotéine du virus comme agent immunogène ont donné des résultats médiocres (Yilma et al. 1985, Cantlon & Gordy 2000). Récemment un vaccin constitué de virus atténué, obtenu par altération du gradient d'expression génique naturel du virus a été développé et testé sur des porcs avec des résultats satisfaisants (Flanagan et al. 2001).

L'infection est transmise par voie cutanée ou par contact avec des muqueuses infectées et le virus est retrouvé dans les fluides des vésicules et dans l'épithélium. Il ne persiste pas après la guérison. Le virus n'est excrété ni dans les fèces, ni dans l'urine, ni dans le lait. La maladie chez les animaux est brève avec une virémie minimale. De nombreux Vesiculovirus ont été isolés chez les moustiques, les mouches, les moucherons, les mites et les tiques (Johnson et al. 1969, Tesh et al. 1972, 1987, Comer et al. 1990). 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 par des piqures d'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 (Johnson et al. 1966), mais le plus souvent l'infection est asymptômatique. Un cas d'encéphalite causée par VSV a été documenté (Quiroz et al. 1988). Après exposition, la période d'incubation est d'environ 9 jours et les symptômes peuvent persister 3 à 4 jours. Des infections ont été observées chez des personnes en contact avec un haut titre viral (inhalation d'aérosols, ou contamination accidentelle des yeux) comme le personnel travaillant dans les fermes en contact avec les animaux infectés. Cependant, dans certaines zones de l'Amérique centrale où le VSV est endémique, la majorité des résidents ont développé des anticorps anti-VSV (Hanson 1952, Tesh et al. 1987).

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.

VSV comme outil thérapeutique

Ce virus se multiplie facilement dans le cytoplasme de nombreuses souches cellulaires, et des tests de transcription et de 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. Depuis quelques années VSV semble être un bon outil de vaccination (Lichty et al. 2004, Ramsburg et al. 2004, Palin et al. 2007), mais peut aussi être utilisé comme virus oncolytique (Ebert et al. 2003). Toutefois, peu d'informations sont disponibles quant à l'effet de l'administration de ce virus à l'homme, et les tissus infectés ne sont pas clairement établis.

Entre autre, il semble que VSV constitue un système adapté pour l'expression d'antigènes de surface étrangers (Schnell et al. 1996), une propriété qui a été utilisée avantageusement pour la conception de nouveaux vaccins (Haglund et al. 2002).

1.1.2.5 Le cycle viral

Le cycle viral du VSV est exclusivement cytoplasmique comme pour tous les Rhabdovirus à l'exception des Rhabdovirus de plantes dont une partie du cycle viral se déroule dans le noyau (Martins et al. 1998, Goodin et al. 2001, 2005, Deng et al. 2007, Bandyopadhyay et al. 2010). Dans le cas de RAV et VSV, certaines protéines virales pénètrent dans le noyau pour interagir avec les protéines nucléaires, et notamment inhiber la réponse immunitaire de la cellule hôte (Blondel et al. 2002, Pasdeloup et al. 2005, Chelbi-Alix et al. 2006, Vidy et al. 2007). Le cycle viral se décompose en trois étapes principales (figure 1.6) : l'entrée du virus (adhésion, endocytose, fusion membranaire et libération des nucléocapsides dans le cytoplasme), la multiplication du virus (transcription, traduction, réplication), et enfin l'assemblage et le bourgeonnement des nouveaux virions.

Entrée du virus

Le cycle viral débute par l'adhésion du virus à la membrane plasmique cellulaire qui est médiée par les spicules de glycoprotéines présentes à la surface de la particule virale (Matlin et al. 1982).

Pour VSV il a été montré que l'endocytose est un processus clathrine-dynéine dépendant (Cureton et al. 2009). Des expériences d'internalisation de virus marqué avec de la ³⁵S-méthionine démontrent que l'entrée du virus s'effectue en moins de trois minutes (Johannsdottir et al. 2009, Sun et al. 2005), et est médiée par l'interaction de la protéine G avec des lipides membranaires cellulaires (Carneiro et al. 2002, 2006). Le processus de fusion membranaire est déclenché par l'acidification de l'endosome (White et al. 1981, Gaudin 2000, Carneiro et al. 2001, Roche & Gaudin 2004) et libère la NC dans le cytoplasme de la cellule hôte. Chez VSV, les données les plus récentes montrent que la libération des NC s'accompagne de l'import nucléaire d'une fraction des protéines de matrice M présentes au sein du virion, alors que la majeure partie de ces protéines reste liée à la membrane endosomale (Mire

et al. 2009, 2010). Par ailleurs, sur la base d'analyses cinétiques *in vivo* de virions de VSV par imagerie de fluorescence, il a été estimé que la libération des NC virales dans le cytoplasme est un processus relativement rapide de l'ordre de la demi-heure (Das et al. 2009) (figure 1.8 page 19).



FIGURE 1.6 – Les trois phases du cycle viral des *Rhabdoviridae* (Schnell et al. 2010).

La première phase comporte la fixation au(x) récepteur(s) et l'internalisation du virion par la cellule hôte (1), suivie de la fusion des membranes virales et endosomales et de la libération dans le cytoplasme du matériel viral (2). Les étapes (1) et (2) constituent la phase 1 du cycle viral. Dans la seconde phase, les composants du virion sont produits (transcription, réplication et synthèse protéique (3)). La troisième phase du cycle viral consiste en l'assemblage des composants du virion et leur bourgeonnement hors de la cellule hôte (4). Les virions peuvent débuter une nouvelle infection.

Multiplication du virus

L'étape suivante est la transcription primaire. L'ARN viral est transcrit en premier lieu en ARNm par le complexe ARN polymérase ARN dépendante-phosphoprotéine (L-P) (étape 3) (Emerson & Yu 1975, Abraham & Banerjee 1976, Banerjee et al. 1977). Ces ARNm sont coiffés et polyadénylés par l'ARN polymérase virale puis traduits en protéines virales par la machinerie de traduction cellulaire (étape 3).

Lorsque la quantité de protéines virales néo-synthétisées atteint un niveau suffisant, la polymérase virale commence à répliquer le génome viral. Chez VSV, des complexes protéiques différents semblent être impliqués dans la réplication et la transcription (Qanungo et al. 2004). Cependant le mécanisme qui gouverne le passage de la transcription à la réplication n'est pas connu (Perlman & Huang 1973, Perrault et al. 1983, Barik & Banerjee 1992*a,b*, Wertz et al. 1994, Das et al. 1997, Gupta & Banerjee 1997, Li & Pattnaik 1999, Whelan & Wertz 1999*b*). En mode réplicatif, le complexe réplicase ignore les signaux de début et fin de gène et synthétise une molécule d'ARN positif simple brin, l'antigénome. Ceux-ci peuvent alors servir soit de matrice pour la synthèse des ARNm viraux (transcription secondaire) soit pour la synthèse de nouveaux génomes viraux (réplication) (Banerjee 1987). Dès le début de leur synthèse, les génomes et antigénomes sont encapsidés par la nucléoprotéine néo-synthétisée (Patton JT

1984, Moyer et al. 1991).

Assemblage et bourgeonnement des particules virales

Ces dernières étapes du cycle viral se font au niveau de la membrane cellulaire. Une étude menée par Swinteck et Lyles (Swinteck & Lyles 2008) sur l'organisation des composants viraux au niveau de la membrane cellulaire a permis une meilleure compréhension des processus d'assemblage et de bourgeonnement. Pour l'assemblage des nouvelles particules virales, les ARN viraux néo-formés associés à la nucléoprotéine, à la phosphoprotéine et à la polymérase virale sont exportés vers la membrane plasmique par l'intermédiaire des microtubules (David 1973, Hunt et al. 1976, Das et al. 2006). La synthèse de G se fait dans le RE et transite par l'appareil de Golgi. Ensuite G forme des micro-domaines dans la membrane cellulaire (Bergmann JE 1981, Strous et al. 1983, Lotti et al. 1992). La protéine M a une double localisation, cytoplasmique et membranaire. Chez le virus de la rage, elle se place sur la face interne de la membrane plasmique où elle interagit avec les queues cytoplasmiques des spicules de glycoprotéine (Mebatsion et al. 1999). Pour le VSV, il a été montré que la queue cytoplasmique de la protéine G était nécessaire pour le bourgeonnement (Whitt MA 1989), mais il était possible de modifier extensivement sa séquence, suggérant l'absence d'interaction spécifique (Schnell et al. 1998). Par ailleurs, M interagit également avec les nucléocapsides dont elle stabilise la condensation en hélice (Newcomb & Brown 1981, Mebatsion et al. 1999, Ge et al. 2010). De nouvelles particules virales compactes bourgeonnent hors de la cellule hôte en détournant des protéines cellulaires (Harty et al. 1999, Irie et al. 2004).

1.2 La transcription et la réplication chez les Mononegavirales

L'ARN polymérase ARN dépendante virale possède des activités réplicase et transcriptase responsables des processus de synthèse d'ARN durant le cycle viral. Une fois libérée dans le cytoplasme de la cellule hôte, l'ARN viral subit en premier la transcription puis la réplication, ceci s'appliquant à tous les virus de l'ordre des *Mononegavirales*. La plupart des connaissances actuelles sur le sujet dérivent d'études sur des virus modèles : le virus de la stomatite vésiculaire (Rhabdovirus) et le virus Sendai (Paramyxovirus).

1.2.1 Génome des Rhabdoviridae

Le génome des *Rhabdoviridae* est constitué d'un brin d'ARN de polarité négative (figure 1.7). Cet ARN comporte de 5 à 12 gènes selon les virus. Cinq de ces gènes sont communs à tous les *Rhabdoviridae*. Ils sont organisés dans le même ordre le long de la molécule d'ARN et codent successivement de l'extrémité 3' vers l'extrémité 5' pour : la nucléoprotéine (N), la phosphoprotéine (P), la protéine matrice (M), la glycoprotéine (G) et l'ARN polymérase ARN dépendante (L) (Tordo et al. 1986). La taille du génome est comprise entre 11 et 15 kb : pour les *Rhabdoviridae* - 11 161 nucléotides (nt) pour le VSV et 11 932 nt pour RAV. Il existe des séquences spécifiques à l'extrémité 3' du génome (région leader) et à l'extrémité 5' (région trailer) (Blumberg et al. 1981, Keene et al. 1981, Blumberg et al. 1983).





Les gènes viraux sont séparés par des régions intergéniques de tailles variables pour RAV, mais réduite à 2 nt pour VSV (Rose & Iverson 1979). Pour RAV, 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/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 ψ (Tordo et al. 1986). Il est possible que cette longue zone intergénique soit la conséquence de la perte d'un gène au cours de l'évolution des virus. Chez VSV cette région se limite seulement à 2 nucléotides (figure 1.7). En revanche chez le virus IHNV (virus de la Nécrose Hématopoïétique Infectieuse), Rhabdovirus tochant les poissons, cette région constitue un sixième gène qui code pour la protéine NV (No-virion protein). Cette protéine de 12 kDa est exprimée dans les cellules infectées (Kurath 1985), et constitue un facteur de virulence (Thoulouze et al. 2004).

1.2.2 La transcription virale

Pour tous les virus à ARN négatif, la transcription constitue la première étape après l'entrée du virion dans la cellule hôte. Cette étape précède la réplication virale et s'effectue de manière séquentielle (Abraham & Banerjee 1976, Ball & White 1976, Banerjee et al. 1977). La transcription du génome viral est initiée au niveau d'un site unique d'entrée de la polymérase virale situé à l'extrémité 3' du génome et l'accès au gène suivant dépend du signal de terminaison du gène précédent (Emerson 1982, Whelan & Wertz 2002). La transcription débute par la synthèse d'un ARN court de 48 nt pour VSV et 58 nt pour RAV, appelé l'ARN leader (Colonno et al. 1976, Colonno & Banerjee 1977, 1978, Tordo et al. 1986, Tordo & Kouknetzoff 1993), qui n'est ni coiffé, ni polyadénylé et n'est donc pas traduit. Certaines régions du leader sont importantes pour la transcription virale (Whelan & Wertz 1999a). L'addition de la coiffe nécessite que l'ARNm atteigne une longueur minimale de 31 nucléotides qui pourrait correspondre à la distance séparant les sites actifs indiqués dans la synthèse d'ARN et dans la formation de la coiffe au sein de la protéine L (Tekes et al. 2011). L'addition de la coiffe sur les ARNm viraux s'effectue par un mécanisme propre aux Rhabdoviridae, qui a été étudié en détail chez VSV (Abraham et al. 1975, Moyer et al. 1975, Testa & Banerjee 1977, Horikami & Moyer 1982, Barr et al. 1997a, Stillman & Whitt 1999, Gupta et al. 2002, Li et al. 2005, 2006, 2007, 2009, Rahmeh et al. 2009, Ogino et al. 2010). A la fin du gène, la polymérase atteint une séquence conservée qui est le signal de fin de gène. Par un processus itératif la polymérase ajoute une queue polyA d'une taille variable allant de 150 à 300 bases (Barr et al. 1997b). La polymérase passe au gène suivant et initie la transcription au niveau de la séquence conservée UUGURRNGA (Iverson & Rose 1981). Ce même processus de transcription, d'addition de coiffe et de polyadénylation est poursuivi pour les 5 gènes de VSV (Abraham & Banerjee 1976). L'atténuation de la transcription qui se produit à chaque région intergénique crée ainsi un gradient de concentration des ARNm (Iverson & Rose 1981, Finke et al. 2000) décroissant en fonction de leurs positions dans le génome viral (figure 1.8 page 19). La position des gènes régule la synthèse des protéines virales. Le gène de la nucléoprotéine est le plus proche du promoteur, assurant ainsi la synthèse massive de N nécessaire pour l'encapsidation du génome viral. A l'inverse, le gène codant pour l'ARN polymérase ARN dépendante (L), dont la concentration requise est inférieure, est située en position distale (Abraham & Banerjee 1976).

Dans les années 80, trois modèles plausibles de transcription du génome viral ont été comparés : <u>le modèle par clivage</u> présentant une entrée unique pour la polymérase à l'extrémité 3' du génome avec un seul site d'initiation de la transcription. Ce modèle impliquait la fabrication d'un ARN unique qui serait coupé par la suite à des endroits spécifiques afin de former les ARN messagers correspondant à chaque gène (Colonno et al. 1976). Le deuxième modèle appelé le modèle du <u>STOP-START</u> avec un seul site d'entrée à l'extrémité 3' du génome mais avec des sites d'initiation multiples. A chaque fin de gène un ARN messager est libéré (Banerjee et al. 1977). Le troisième modèle appelé <u>le modèle d'initiation multiple</u> avec plusieurs sites d'entrée de la polymérase et plusieurs sites d'initiations (Testa et al. 1980). Afin de différencier ces modèles des expériences d'irradiation aux rayons ultraviolet (UV-mapping) ainsi que des expériences de mutagénèse des régions intergéniques ou d'insertion de gènes étrangers ont été menées (Emerson 1982, Banerjee 1987, Barr et al. 2008).

Le modèle le plus probable semble être le modèle STOP-START (modèle de transcription séquentielle) avec un seul site d'entrée, c'est-à-dire, un seul promoteur. En début de chaque gène la polymérase fabrique la coiffe des ARN messagers comme décrit précédemment et en fin de gène par un mécanisme de dérapage rajoute la queue polyA et continue la transcription du second gène. La transcription des gènes est polaire, c'est-à-dire les ARN messagers ne sont pas produits en quantité équimolaire - plus les gènes sont éloignés des promoteurs, moindre est leur expression (figure 1.8 page 19) (Villarreal et al. 1976). Des mesures cinétiques montrent que la polymérase procède à une vitesse de 3-4 nucléotides/sec et la transcription est un phénomène discontinu avec des pauses dans les régions intergéniques allant de 2.5 à 5.7 minutes (Iverson & Rose 1981). Toutes ces données expérimentales de cinétique enzymatique suggèrent que la terminaison de la transcription et la fabrication de la coiffe sont des processus lents comparés à celui de synthèse d'ARN (Iverson & Rose 1981). Ce processus de régulation est propre aux *Mononegavirales*.



La cinétique du cycle viral est divisée en trois phases : phase 1 - entrée endosomale et libération des RNP dans le cytoplasme, la phase 2 est constituée de deux cycles de transcription séparés par un cycle de réplication (Simonsen et al. 1979) et la phase 3 consiste à l'assemblage des particules virales nouvellement formées. Les temps approximatifs nécessaires pour chaque processus sont marqués sur la figure.

La formation de gradient des ARNm est indiquée avec des flèches noires sous chaque gène en vert pour la N, rouge pour la P, gris-clair pour la M, gris-foncé pour la G et jaune pour la L. Les pourcentages d'atténuation de transcription et les pauses intergéniques sont indiqués pour chaque transcription (primaire et secondaire) (Bishop & Roy 1972, Iverson & Rose 1981, Villarreal et al. 1976, Rose 1978, Ball et al. 1999).

La réplication du génome viral est montrée, ARN(-) est répliqué en ARN(+) qui sert de matrice pour la formation de nouveau génome viral. La vitesse de polymérisation est indiquée également.
1.2.3 La réplication virale

La réplication de l'ARN négatif nécessite la production d'une copie complémentaire du génome, appelée antigénome ou ARN complémentaire viral, qui est distincte des ARNm. L'ARN génomique, tout comme l'ARN antigénomique, est produit sous la forme d'un complexe ribonucléoprotéique. La polymérisation de l'ARN est associée à l'encapsidation. Au cours de l'encapsidation, la séquence leader est encapsidée en premier (Blumberg et al. 1983, Tordo et al. 1986), et certaines données suggèrent que l'ARN leader contiendrait des signaux d'encapsidation (Blumberg et al. 1983, Yang et al. 1998, 1999). Il a été montré que chez VSV, l'ARN leader s'accumule dans les noyaux des cellules infectées, inhibant la transcription des gènes de la cellule hôte (Kurilla et al. 1982). Cette propriété de l'ARN leader pourrait jouer un rôle dans le basculement entre transcription et réplication du virus, dont le mécanisme précis demeure à l'heure actuelle mal compris (Perlman & Huang 1973, Perrault et al. 1983, Wertz et al. 1994, Das et al. 1997, Gupta & Banerjee 1997, Li & Pattnaik 1999, Whelan & Wertz 1999a, b). Cependant, les différentes études indiquent que le basculement entre les deux activités pourrait être régulé notamment par la quantité de nucléoprotéine déjà produite au cours du cycle virale (Arnheiter et al. 1985) et la quantité de complexe N°P présente (Blumberg et al. 1981, Gupta & Banerjee 1997), par des phosphorylations au niveau de la protéine P (Barik & Banerjee 1992a, b, Takacs et al. 1992), ou encore par la présence de signaux spécifiques dans les régions leader (Li & Pattnaik 1999) et trailer (Whelan & Wertz 1999a).

Dans le mode réplicatif, 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 (Wertz et al. 1994), la polymérase virale peut amorcer la synthèse au niveau de plusieurs promoteurs (voir figure 1.8 page 19) :

- 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 et de 64 respectivement pour RAV et VSV, constitue à son tour un signal d'encapsidation par la nucléoprotéine (Bourhy et al. 1989) lorsque la polymérase virale copie l'antigénome de façon processive.

La réplication se fait de manière asymétrique, et produit 10 fois plus de génomes que d'antigénomes (Whelan & Wertz 2002). En effet, la région leader est un promoteur moins fort que le promoteur situé sur la région anti-trailer (Whelan & Wertz 1999a,b). L'ARN antigénomique peut alors être répliqué par la polymérase virale pour produire plus de copies génomiques.

Durant un processus de réplication le taux d'erreur d'incorporation de nucléotides varie de 10^{-3} à 10^{-4} par copie (Steinhauer & Holland 1986). Ce taux d'erreur couplé à une absence de mécanisme de correction de copie résulte dans l'existence d'une population de virus différente, qui a conduit à l'idée que le virus ne peut être considéré comme une espèce unique mais doit au contraire être considéré comme une quasi-espèce (Domingo & Holland 1997).

1.2.4 Structures des protéines virales

Afin de mieux comprendre le processus moléculaire de transcription et réplication, il est indispensable de déterminer la structure atomique des différents composants viraux, de comprendre comment ils s'assemblent et comment ils changent de conformation ou se réarrangent.

En 2006, la résolution de la structure cristallographique de la nucléoprotéine de VSV et celle du RAV ont permis de comprendre comment N encapside l'ARN viral. L'ARN se retrouve complètement séquestré à l'intérieur du polymère de nucléoprotéines et est ainsi protégé contre les ribonucléases et le système de l'immunité innée (Albertini et al. 2006, Green et al. 2006).

Des expériences menées dans notre laboratoire démontrent que la P des Rhabdovirus contient des domaines globulaires interconnectés par des régions désordonnées flexibles (Gerard et al. 2007, 2009, Leyrat et al. 2010). Cette organisation modulaire de la phosphoprotéine ne rend pas envisageable la cristallisation de la protéine entière (figure 1.9 page 22).

Aucune structure cristallographique à l'heure actuelle n'est connue pour l'ARN polymérase ARN dépendante. Des données récentes de microscopie électronique suggèrent que la polymérase du virus de la stomatite vésiculaire possède un domaine circulaire et trois domaines globulaires (Rahmeh et al. 2010). Cependant une structure à plus haute résolution est nécessaire afin de comprendre le mécanisme moléculaire d'addition de la coiffe ainsi que le mode de polymérisation de l'ARN viral et les éventuelles régulations du basculement entre transcription et réplication.

1.2.4.1 Protéines impliquées dans la transcription et la réplication du virus

• La phosphoprotéine (P)

Tous les virus appartenant à l'ordre des *Mononegavirales* possèdent un cofacteur essentiel de la machinerie de transcription et de réplication virale équivalent aux phosphoprotéines des virus de Sendai, VSV et RAV (Emerson & Yu 1975). Les séquences, les tailles et les niveaux de régulation (oligomérisations/phosphorylation(s)) des phosphoprotéines de ces différents virus sont très variables (virus de Sendai 568 résidus de P, virus de la rage 297 résidus et pour le virus de la stomatite vésiculaire 265 résidus), mais les phosphoprotéines partagent des caractéristiques communes :

- activité de chaperonne : P forme un complexe avec la nucléoprotéine empêchant ainsi sa liaison aux ARN cellulaires cette forme est dite soluble, et le complexe est nommé N°P, « ° ≫ signalant l'absence d'ARN (Masters & Banerjee 1988, Takacs et al. 1991, 1993, Curran et al. 1995, Takacs & Banerjee 1995, Gupta & Banerjee 1997, Mavrakis et al. 2003, 2006, Chen et al. 2007).
- activité transcriptionnelle : P se fixe sur le complexe nucléoprotéine-ARN (NC) : (Paul et al. 1988, Takacs et al. 1993, Takacs & Banerjee 1995, Schoehn et al. 2001, Ribeiro et al. 2008) mais également sur l'ARN polymérase ARN dépendante formant ainsi le complexe L-P et jouant le rôle de « connecteur »entre la L et la NC virale (Emerson & Yu 1975, Mellon & Emerson 1978,

De & Banerjee 1984, 1985, Gill et al. 1986, Emerson & Schubert 1987, Howard & Wertz 1989).
- auto-oligomérisation : P forme des homo-oligomères (Tarbouriech, Curran, Ruigrok & Burmeister 2000, Tarbouriech, Curran, Ebel, Ruigrok & Burmeister 2000, Choudhary et al. 2002, Ding et al. 2006, Llorente et al. 2006, Gerard et al. 2007, Ivanov et al. 2010).



FIGURE 1.9 – Structure de la phosphoprotéine.

A : Représente une organisation schématique de la phosphoprotéine en général.

Elle est constituée de trois régions, une région Nterminale dépliée (P_{NTR}), un domaine central responsable à l'oligomérisation (P_{CED}) et une région structurée située en C-terminal appelée le domaine à liaison à N-ARN (P_{CTD}). La région entre le domaine central et le domaine Cterminal est une région intrinséquement dépliée appelée IDR_{CT}. Le cercle en pointillés représente la région MoRE qui adopte une structure en hélice α après liaison à N°.

B : Représente un schéma structural de la phosphoprotéine des *Rhabdoviridae* avec les deux structures à haute résolution : le P_{CTD} de RAV avec pdb code : 2k47 (Mavrakis et al. 2004) et le domaine central de VSV avec le pdb code : 2fqm (Ding et al. 2006). Les pointillées représentent les régions dépliées. Le rond en N-terminal représente la région MoRE.

La P du VSV (figure 1.9) et celle du virus de la rage ont fait l'objet de nombreuses études. Les deux phosphoprotéines présentent des organisations moléculaires similaires. Elles sont toutes les deux dimériques en solution et renferment des régions intrinséquement dépliées (Gerard et al. 2007, Leyrat et al. 2010, Leyrat, Jensen, Ribeiro, Gérard, Ruigrok, Blackledge & Jamin 2011). Elles ont deux sites d'interaction avec la nucléoprotéine. Le site d'interaction en N-terminal de la phosphoprotéine du virus de la rage a été déterminé par protéolyse limitée étant le peptide 11-30 interagissant avec la nucléoprotéine pour former le complexe N°P (Chen et al. 2007, Mavrakis et al. 2006, Castel et al. 2009). Cette même région a été prédite structurée par des serveurs de méta-prédictions (Gerard et al. 2009). Les résidus 4-40 pour la phosphoprotéine du virus de la rage sont nécessaires et suffisants à garder soluble le complexe N°P. Le peptide correspondant pour VSV se trouve dans la région de 1 à 60. Récemment la structure cristallographique de N°P a été obtenue dans le laboratoire montrant l'interaction du peptide P_{1-60} avec la nucléoprotéine et comment celui-ci maintient la nucléoprotéine dans sa forme monomérique-soluble (figure 1.16 page 30) (Leyrat, Yabukarski, Tarbouriech, Ribeiro, Jensen, Blackledge, Ruigrok & Jamin 2011). Un second site de fixation de nucléoprotéine existe sur la phosphoprotéine du virus de la rage et celle de VSV, il s'agit de l'extrémité C-terminale (P_{CTD}). Le domaine pour VSV étant $P_{195-265}$ et celui de la rage (RAV) $P_{185-297}$ sont des domaines autonomes, capables de se replier

indépendemment sans la présence nécessaire des autres régions de la phosphoprotéine (Paul et al. 1988, Takacs et al. 1993, Takacs & Banerjee 1995, Ribeiro et al. 2008). Ces deux régions structurées ont été également déterminées par des serveurs de méta-prédiction comme étant des régions structurées (Gerard et al. 2009). Les structures des deux domaines montrent clairement qu'ils sont homologues (Mavrakis et al. 2004, Ribeiro et al. 2008).

Le domaine d'oligomérisation de la phosphoprotéine de VSV (P_{71-177}) a été cristallisé et la structure suggère la présence d'un dimère parallèle avec une hélice α de chaque monomère formant des interactions dans la forme dimérique. Egalement deux brins β de chaque monomère interagissent ensemble pour former un feuillet β (figure 1.9 page 22). Au moment où nous avons débuté ce travail la structure de la région correspondante chez le virus de la rage était inconnue, mais le domaine d'oligomérisation de la région centrale de la protéine avait été identifiée par délétion (Jacob et al. 2001) et une analyse par SEC-MALLS-RI avait montré que la protéine entière formait des dimères comme VSV (Gerard et al. 2007).

Chez VSV, la jonction entre le domaine central et le domaine C-terminal ainsi que la partie N-terminale sont intrinséquement dépliées (Ribeiro et al. 2008, Leyrat, Jensen, Ribeiro, Gérard, Ruigrok, Blackledge & Jamin 2011). Des données de RMN montrent que la région N-terminale est dépliée en solution avec une faible propension à former des hélices α . Dans le complexe N°P, cette région se structure en hélices α , démontrant qu'elle constitue un MoRE c'est-à-dire une région qui est désordonnée dans sa forme isolée mais qui se structure lors de sa liaison à son partenaire. La phosphoprotéine est présente dans la cellule infectée ainsi qu'au sein du virion avec plusieurs degrés de phosphorylation (Das et al. 1995, Chen et al. 1997, Mathur et al. 1997, Pattnaik et al. 1997, Gupta et al. 2000, Das & Pattnaik 2004). La phosphoprotéine du virus de la rage est modifiée par deux kinases cellulaires : la PKC et une autre kinase spécifique - la RVPK (Gupta et al. 2000). La phosphoprotéine du VSV est quant à elle phosphorylée au niveau de la région N-terminale par la CK-II et au niveau de sa région C-terminale par une protéine kinase associée à l'ARN polymérase virale (L) (Chattopadhyay & Banerjee 1987, Chen et al. 1997). Cette kinase associée à la polymérase virale n'a pas pu être identifiée et n'est pas présente en quantité stoechiométrique dans les complexes de transcription et de réplication du VSV (Qanungo et al. 2004).

Le rôle de la phosphorylation et les mécanismes de la régulation restent à l'heure actuelle mal compris.

Les phosphoprotéines également

- interfèrent avec le mécanisme de l'immunité innée de la cellule hôte et dérégule la fonction cellulaire : (Blondel et al. 2002, Brzozka et al. 2005, Vidy et al. 2005, Brzozka et al. 2006, Chelbi-Alix et al. 2006, Vidy et al. 2007, Ito et al. 2010).
- se fixent à des protéines du cytosquelette suggérant une implication dans le traffic intracellulaire du virus : (Fujinami & Oldstone 1983, Zafrullah et al. 1997, Jacob et al. 2000, Raux et al. 2000, Poisson et al. 2001, Moseley et al. 2009).

– La nucléoprotéine (N)

La nucléoprotéine est la protéine la plus abondante dans la particule virale. Sa fonction principale est d'encapsider le génome et l'antigénome viral, maintenant ainsi l'ARN sous une forme simple brin et lui conférant ainsi une grande stabilité et une haute résistance aux ribonucléases (Simpson & Hauser 1966, Wagner et al. 1969, Huang AS 1970, Kiley & Wagner 1972, Soria et al. 1974, Condra & Lazzarini 1980). Elle protège certainement aussi le génome contre les mécanismes de l'immunité innée.

Dans les cellules infectées on retrouve la nucléoprotéine sous deux formes :

A. Sous sa forme soluble, formant un complexe N°P.

B. Sous la forme oligomérique, liée à l'ARN et formant une matrice ribonucléoprotéique (N-ARN) utilisée pour la transcription et la réplication virale.

La nucléoprotéine du virus de la rage est phosphorylée au niveau de la sérine 389 par la caséine kinase II (Wu et al. 2003), lorsqu'elle est présente dans la cellule sous forme N-ARN (Kawai et al. 1999). Ce phénomène de régulation par phosphorylation de la nucléoprotéine N, module l'affinité du complexe N-ARN pour la phosphoprotéine (Toriumi et al. 2002, Toriumi & Kawai 2004) et ne semble pas se retrouver chez VSV.

A. Les complexes Nucléoprotéines-ARN :

- La nucléocapside

Les nucléocapsides des virus à ARN négatif ont une symétrie hélicoïdale (Compans & Choppin 1967, Sokol et al. 1969, Egelman et al. 1989, Bhella et al. 2002, 2004) et forment des longues hélices flexibles, mais les paramètres hélicoïdaux, ainsi que la régularité des structures varient en fonction des conditions expérimentales et du virus considéré, comme le montrent les clichés de microscopie électronique présentés dans la figure 1.10.



FIGURE 1.10 – Images de microscopie électronique en coloration négative de certaines nucléocapsides de *Mononegavirales* (Albertini et al. 2005).

(a) : Nucléocapsides recombinantes du virus de la rougeole (*Paramyxoviridae*).

(b) : Nucléocapsides virales du virus de la rage - (*Rhabdoviridae*).

(c) : Nucléo capsides recombinantes du virus de Marburg (Filoviridae).

(d) : Nucléocapsides recombinantes du virus de Nipah (*Henipahviridae*).

L'affinité de l'ARN pour la nucléoprotéine est très forte. Il a été montré que les nucléocapsides de *Paramyxoviridae* et de *Rhabdoviridae* sont très stables et peuvent résister à de hautes concentrations salines et à une force de gravité importantes (Simpson & Hauser 1966, Soria et al. 1974, Condra & Lazzarini 1980, Heggeness et al. 1980, Blumberg et al. 1984). Ainsi, les nucléocapsides du virus de la rage résistent à des concentrations drastiques d'agent dénaturant (Iseni et al. 1998). Les changements importants de ces nucléocapsides en réponse à la force ionique indiquent la plasticité de cette structure hélicoïdale (Heggeness et al. 1980). 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 pour adopter la structure hautement organisée présente à l'intérieur du virion (figure 1.4 B page 8) (Ge et al. 2010).

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 le virus de Sendai (*Paramyxoviridae*), le nombre de nucléoprotéines par tour d'hélice est de 13 (Egelman et al. 1989, Schoehn et al. 2004). Pour RSV et le SV5, ce nombre varie de 10 à 14 (Bhella et al. 2002). Pour les nucléocapsides de Rhabdovirus le nombre de nucléoprotéine par tour d'hélice dans la particule virale est de 37,5. L'hétérogénéité et la plasticité de ces nucléocapsides rendent impossible 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 1.10 page 24).

<u>Fixation à l'ARN</u>

La nucléoprotéine des virus à ARN négatif se fixe sur un nombre entier de nucléotides. Des études de microscopie électronique ont permis de déterminer le nombre de nucléotides fixés par nucléoprotéine. Ce nombre est spécifique pour chaque famille de virus au sein de l'ordre des *Mononegavirales* : la nucléoprotéine des *Paramyxoviridae* fixe six nucléotides (Egelman et al. 1989) à l'exception de RSV qui en fixe 7 nucléotides (Tawar et al. 2009), celle des *Filoviridae* fixe entre douze et quinze nucléotides (Mavrakis et al. 2002) et celle des *Rhabdoviridae* fixe neuf (Thomas et al. 1985).

Pour les Paramyxoviridae, à l'exception des Pneumovirus (Samal & Collins 1996), il existe une règle appelée « règle de six »(Calain & Roux 1993, Pelet et al. 1996, Kolakofsky et al. 1998, Phillips et al. 1998, Peeters et al. 2000, Halpin et al. 2004, Kolakofsky et al. 2005), qui indique que le génome viral doit contenir un multiple entier de six nucléotides pour que la polymérase virale réplique efficacement le génome. En revanche, aucune « règle de neuf » n'a été mise en évidence chez les *Rhabdoviridae* (Iseni et al. 2000). La longueur du génome viral n'est pas un multiple de 9 nucléotides suggérant que quelques-uns pourraient être exposés au solvant et accessibles à la polymérase pour le processus de transcription/réplication.

Les anneaux de nucléoprotéines

Lorsque la nucléoprotéine du virus de la rage ou du VSV est exprimée sous forme recombinante dans des cellules d'insectes ou des bactéries, celle-ci encapside des ARN de façon non-spécifique pour former de longues structures hélicoïdales de type nucléocapside, ainsi que des structures circulaires comportant un nombre variable de nucléoprotéines (9 à 15), fixant des ARN courts (figure 1.11).



FIGURE 1.11 – Différents complexes de nucléoprotéines-ARN du virus de la rage observés en microscopie électronique en coloration négative (Iseni et al. 1998, Schoehn et al. 2001, Mavrakis et al. 2003, Albertini et al. 2007, 2008).

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.

Une reconstruction 3D des anneaux formés de dix nucléoprotéines a été obtenue à partir d'images de cryo-microscopie électronique (Schoehn et al. 2001) (figure 1.12) et le site de fixation de la phosphoprotéine a été localisé dans le domaine C-terminal de N sur la base d'expériences de protéolyse limitée. Le traitement de ces anneaux par la trypsine entraîne un clivage au niveau du résidu 376 de la nucléoprotéine, éliminant ainsi une partie du domaine C-terminal et inhibant la fixation de la phosphoprotéine (Kouznetzoff et al. 1998, Schoehn et al. 2001).



FIGURE 1.12 – Images de microscopie électronique en coloration négative de certains nucléocapsides de *Mononegavirales* (Schoehn et al. 2001).

trypsine sont représentés en orange.

(i-l) : Superpositions des anneaux N_{10} intacts et digérés. Les deux structures sont très similaires. Une portion significative de la densité est absente au niveau d'une des faces des anneaux digérés. Dans les deux cas, l'angle mesuré entre les deux sous-unités est de 144°.

En 2006, la structure cristallographique des anneaux N_{11} -ARN du virus de la rage a été obtenue à 3.5 Å de résolution (Albertini et al. 2006), et la structure cristallographique des anneaux N_{10} -ARN du virus de la stomatite vésiculaire à 2.9 Å (Green et al. 2006). Ce sont les premières structures de nucléoprotéines de virus à ARN négatifs, où la nucléoprotéine est associée à l'ARN. Ces complexes ont tous les deux été cristallisés dans le groupe d'espace $P2_12_12$ et les structures ont été résolues par SAD. Pour le virus de la rage, les anneaux N₁₁ (environ 550 kDa) sont positionnés face à face dans l'unité asymétrique et chaque anneau possède un diamètre externe de 160 Å, un diamètre interne de 60 Å et une hauteur de 74 Å.

Le protomère de nucléoprotéine consiste en deux domaines, NTD et CTD, qui sont en contact avec neuf nucléotides d'ARN simple brin (figure 1.13 D). Le NTD constitué de résidus 32 à 233 est replié en 6 hélices reliées par des longues boucles. Le CTD constitué de résidus 236 à 356 est composé de 11 hélices reliées par des boucles plus courtes (figure 1.13 C page 27). Les deux domaines NTD et CTD agissent comme des mâchoires en séquestrant la molécule d'ARN (figure 1.13 A,B). Deux régions sont flexibles, et absentes dans la structure cristallographique (NTD : 105-118 et CTD : 376-397). Il existe également deux autres sous-domaines appelés le bras N-terminal et la boucle C-terminale (figure 1.13 C) qui interviennent dans l'assemblage des polymères de N. Le bras N-terminal vient ainsi interagir avec la boucle C-terminale du voisin N+2.



FIGURE 1.13 – Structure cristallographique des anneaux N_{11} -ARN du virus de la rage (Albertini et al. 2006).

A : Vue du dessus de l'anneau N_{11} : chaque nucléoprotéine est de couleur différente et l'ARN est représenté en noir.

B : Vue de côté de l'anneau montrant l'intérieur de six protomères de nucléoprotéines. L'ARN est séquestré dans la cavité centrale située entre les deux domaines de la nucléoprotéine.

C : Protomère de nucléoprotéine en vue de face indiquant la localisation des différents éléments de structure secondaire y compris la boucle Cterminale sensible à la digestion à la trypsine nécessaire à la liaison de la phosphoprotéine et le bras N-terminal.

D : Représentation schématique de l'interaction d'un protomère de nucléoprotéine avec les neuf bases d'ARN.

Les interactions entre la nucléoprotéine et l'ARN sont principalement d'ordre électrostatique et font intervenir de nombreux résidus basiques de la cavité de N et les phosphates de l'ARN (Albertini et al. 2006, Green et al. 2006, Luo et al. 2007, Albertini et al. 2008). Ceci est en accord avec les résultats d'Iseni (Iseni et al. 2000) qui montraient que la nucléoprotéine interagit principalement avec les sucres ou les phosphates et laisse les bases accessibles à l'ARN polymérase ARN dépendante pour la transcription et la réplication. La structure cristallographique révèle que chaque nucléotide est engagé dans des contacts polaires *via* ses groupements phosphates (7 des 9 nucléotides) ou son ribose (2 des 9 nucléotides). Les nucléotides 4,6 et 7 sont empilés, et leurs bases pointent vers l'intérieur de la cavité. La base du nucléotide 5 pointe vers le solvant, tout comme les bases 1,2,3,8 et 9 (voir figure 1.13 D page 27).

Les nucléoprotéines du virus de la rage et du VSV possèdent des structures très similaires (figure 1.14 A) (Luo et al. 2007, Albertini et al. 2008), et ce en dépit d'une faible similarité de séquence (15 % d'identité). Les résidus basiques responsables de l'interaction avec l'ARN sont conservés entre les nucléoprotéines du virus de la rage et de la stomatite vésiculaire, de même pour l'agencement tertiaire des hélices α composant les domaines N et C terminaux (figure 1.14 B). La différence principale entre ces deux structures réside dans la longueur des boucles C-terminales désordonnées, impliquées dans la fixation de la phosphoprotéine (Schoehn et al. 2001), qui contiennent par ailleurs le site de phosphorylation Ser389 chez le virus de la rage (Wu et al. 2003), absent chez le VSV.



FIGURE 1.14 – Comparaison des structures et conservation de séquence des nucléoprotéines du virus de la rage et du VSV (Luo et al. 2007, Ivanov et al. 2011)

A : Les structures des nucléoprotéines du virus de la rage et du VSV sont superposées (*rmsd* 2,18 Å) et représentées en cartoon. Le domaine N-terminal de VSV-N est coloré en vert et le domaine C-terminal est en jaune. La nucléoprotéine du virus de la rage est représentée en gris et les régions présentant des différences structurales importantes sont en rouge. Les résidus N et C-terminaux, ainsi que les éléments de structures secondaires sont indiqués d'après la numérotation de la protéine de VSV-N.

B : Les résidus conservés sont montrés en rouge, les résidus similaires sont montrés en jaune et en orange. L'alignement de séquence a été fait avec clustalW2.

A l'heure actuelle, sept structures de nucléoprotéines sont disponibles : deux nucléoprotéines de *Rhabdoviridae* (figure 1.15 B,C page 29), dans lesquelles chaque protomère est en interaction avec 9 nucléotides d'ARN simple brin (Albertini et al. 2006, Green et al. 2006), la nucléoprotéine du virus

respiratoire syncytial de *Paramyxoviridae* (figure 1.15 D page 29) (Tawar et al. 2009) qui révèle un mode d'interaction avec l'ARN différent de ce qui est observé chez les *Rhabdoviridae*, la nucléoprotéine du virus de la maladie de Borna (figure 1.15 A) (*Bornaviridae*, cristallisée en l'absence d'ARN) sous la forme d'homo-tétramère (Rudolph et al. 2003), la nucléoprotéine du virus de la grippe (figure 1.15 E) (*Orthomyxoviridae*), également cristallisée sans ARN, sous la forme d'homo-trimère (Ye et al. 2006) et deux nucléoprotéines de *Bunyaviridae* celle du virus de la fièvre de la vallée du Rift (figure 1.15 F) (Raymond et al. 2010) et celle du virus Lassa cristallisée en absence d'ARN (figure 1.15 G) (Qi et al. 2010).



FIGURE 1.15 – Comparaison des structures de nucléoprotéines de différents virus à ARN négatif (Ruigrok et al. 2011).

A : Nucléoprotéine du virus de la maladie de Borna (code PDB : 1N93)

B : Nucléoprotéine du virus de la rage (code PDB : 2GTT)

C : Nucléoprotéine du virus de la stomatite vésiculaire (code PDB : 2GIC)

D : Nucléoprotéine du virus respiratoire syncytial (code PDB : 2WJ8)

E : Nucléoprotéine du virus de la grippe (code PDB : 2QO6). Cette protéine ne possède pas de domaines N ou C-terminaux. L'épingle est impliquée dans l'interaction entre monomère.

F: Nucléoprotéines du virus de la fièvre de la vallée du Rift (code PDB : 30V9). La figure de gauche représente une forme fermée et la figure de droite une forme ouverte avec un bras impliqué dans l'oligomérisation.

G : Nucléoprotéine du virus Lassa (code PDB : 3MWP).

En haut les structures sont représentées en cartoon, les couleurs varient de pourpre au rouge allant du N-terminal vers le C-terminal.

En bas les protéines sont représentées en surfaces électrostatiques calculés avec DelPhi colorées du bleu (+5kT/e) au rouge (-5kT/e) (Rocchia et al. 2002) et tournées de façon à montrer les résidus chargés positivement. Toutes les nucléoprotéines sont représentées sans ARN pour plus de clarté.

Même s'il est difficile de superposer les structures des nucléoprotéines des *Rhabdoviridae* avec celles des autres virus à ARN négatif, on observe la présence de deux domaines (NTD et CTD), principalement composés d'hélices α , et responsables de l'interaction avec l'ARN (figure 1.15). Pour la plupart des nucléoprotéines on retrouve également les deux sous-domaines impliqués dans la polymérisation de la nucléoprotéine. Cependant on ne retrouve pas cette séparation en deux domaines pour le virus de la grippe dont le génome est ségmenté (figure 1.15 E) et le virus de la fièvre de la vallée du Rift (figure 1.15 F). La nucléoprotéine du virus de la grippe fixe 24 nucléotides, et l'ARN est facilement

éliminé par des traitements à haute molarité en sel. Aussi, chez le virus de la grippe, les bases sont plus exposées au solvant (Baudin et al. 1994) que chez les *Rhabdoviridae* (Iseni et al. 2000), et *Paramyxoviridae* (Iseni et al. 2002).

B. Le complexe Nucléoprotéine-Phosphoprotéine N°P :

Le complexe N°P se forme de manière transitoire au cours du cycle viral, et joue un rôle critique lors de la transcription et la réplication. La formation du complexe permet d'empêcher l'encapsidation non-spécifique d'ARN cellulaire, tandis que ce complexe doit se dissocier afin que la N encapside le génome viral néo-formé. Le complexe N°P de la rage a été exprimé en cellule d'insecte et purifié sous forme soluble et monomérique (Mavrakis et al. 2003). L'ultracentrifugation analytique, la spéctrométrie de masse et la microscopie électronique suggéraient que le complexe est composé d'une nucléoprotéine pour un dimère de phosphoprotéine (Mavrakis et al. 2003). La nucléoprotéine est présente sous forme non phosphorylée (Kawai et al. 1999).

Il semblerait que cette nucléoprotéine non phosphorylée fixe plus efficacement l'ARN (Yang et al. 1998, 1999), et que sa phosphorylation ait lieu pendant ou après l'encapsidation de l'ARN viral (Toriumi et al. 2002). Pour le VSV, le complexe N°P inhibe la transcription *in vitro* (Gupta & Banerjee 1997). Cependant l'addition du complexe N°P provenant de cellules infectées à un système de transcription *in vitro* résulte en la synthèse du génome, laissant supposer que le complexe N°P joue un rôle dans la régulation entre les activités de transcription et de réplication (Gupta & Banerjee 1997).

Récemment dans le laboratoire la structure cristallographique du complexe N°P₁₋₆₀ de VSV a été résolue. Cette structure montre comment la partie N-terminale de la phosphoprotéine empêche l'encapsidation d'ARN cellulaire par la nucléoprotéine. En particulier, le site de fixation de P sur N° recouvre partiellement le site de fixation de l'ARN et empêche l'association stable des protomères de nucléoprotéine (voir figure 1.16).



FIGURE 1.16 – Structure cristallographique du complexe N°P.

A : Représente un protomère de VSV N-ARN avec le pdb code : (2gtt). Le protomère de nucléoprotéine est présenté en surface en vert et l'ARN en bleu.

B : Représente un protomère de $N_{\Delta 21}$ °P₁₋₆₀ avec le code pdb : (3pmk). La nucléoprotéine est représentée également en vert et le peptide visible dans la structure cristallographique est représenté sous forme d'hélice α en rouge.

• L'ARN polymérase ARN dépendante (L)

Les premières ARN polymérases qui ont été découvertes sont l'ARN polymérase ADN dépendante du bactériophage T7 (Jones & Berg 1966, Summers & Siegel 1970, Chamberlin & Ring 1973), l'ARN polymérase ADN dépendante du virus de la vaccine (Munyon et al. 1967) et l'ARN polymérase ARN dépendante des virus de la famille des *Reoviridae* (Shatkin & Sipe 1968, Kalmakoff et al. 1969, Lewandowski et al. 1969). L'activité ARN polymérase ARN dépendante (L) du virus de la stomatite vésiculaire (VSV) fut mise en évidence peu de temps après à partir des virions purifiés (Baltimore et al. 1970).

La protéine L constitue la sous-unité principale de l'ARN polymérase ARN dépendante chez les *Mononegavirales*. Associée à son cofacteur, la phosphoprotéine (P), la polymérase assure à la fois les activités de transcription et de réplication du génome viral, en utilisant le complexe N-ARN viral comme matrice (Bishop & Roy 1972, Emerson & Wagner 1972, Emerson & Yu 1975). Elle est également responsable de la synthèse et de la méthylation de la coiffe et de la polyadénylation des ARNm viraux (Banerjee et al. 1974, Abraham et al. 1975, Moyer et al. 1975, Testa & Banerjee 1977, Horikami & Moyer 1982, Barr et al. 1997*a*, Stillman & Whitt 1999, Gupta et al. 2002, Grdzelishvili et al. 2005, Li et al. 2005, 2006, 2007, Ogino & Banerjee 2007, Li et al. 2009, Rahmeh et al. 2009, Ogino et al. 2010).

Chez les *Rhabdoviridae*, le nombre de molécules de protéine L par virion est faible, estimé à environ 50 molécules (Thomas et al. 1985, Banerjee 1987), qui sont associées au complexe N-ARN (Bishop & Roy 1972, Emerson & Wagner 1972, Emerson & Yu 1975). La polymérase est une protéine de 2142 aa chez le virus de la rage (Tordo et al. 1988), et de 2109 aa chez le VSV (Schubert et al. 1984). Son gène représente ainsi 60% du génome entier de VSV. Cette protéine est la plus fortement conservée au sein des *Rhabdoviridae*, et au sein de l'ordre des *Mononegavirales* (Tordo et al. 1988).

Aujourd'hui, peu d'informations structurales sont disponibles sur les polymérases des virus à ARN négatif. Cependant, ces polymérases possèdent une similarité de séquence avec les polymérases des virus à ARN positif, pour lesquelles plusieurs structures sont connues, notamment celle du virus de l'hépatite C (Ago et al. 1999, Lesburg et al. 1999), du virus de la Dengue (Yap et al. 2007) ou du virus West Nile (Yap et al. 2007). La partie commune de toutes ces polymérases d'ARN négatif ou positif est l'anneau polymérasique.

A. Analyse des domaines conservés de l'ARN polymérase ARN dépendante (L) :

Six domaines conservés ont été identifiés (Tordo et al. 1988, Poch et al. 1989, 1990), ce qui a permis de repérer les sites probables d'interaction avec la matrice N-ARN (Chenik et al. 1998) et les nucléotides (Sleat & Banerjee 1993). Des données de mutagénèse obtenues avec la protéine L de VSV suggèrent que l'ARN polymérase se fixe dans la région N-terminale de la phosphoprotéine (P) VSV (Emerson & Schubert 1987, Castel et al. 2009). Pour le virus RAV la région N-terminale de P a été aussi déterminée comme responsable de la liaison avec L (Chenik et al. 1998). On sait également que la région C-terminale de la polymérase de VSV est responsable de la liaison avec la P (Canter & Perrault 1996), alors que pour le virus Sendai, la phosphoprotéine se fixe à la polymérase par sa partie N-terminale (Holmes &



FIGURE 1.17 – Schéma détaillé des régions conservées de la polymérase (L) du VSV et les structures correspondantes à basse résolution.

Les domaines conservés I-VI sont représentés en jaune et leurs séquences sont indiquées d'après (Poch et al. 1990). Les images de coloration négative des quatre fragments issus de protéolyse limitée d'après (Rahmeh et al. 2010) sont représentées.

Les micrographies sont au même grossissement représenté sur les images du fragment 1-860. Le trait correspond à 20 nm. La polymérase entière (1-2109aa) en coloration négative est représentée en haut de la figure ainsi que les projections à 2D correspondantes à chaque micrographie. Une représentation de la polymérase est dessinée en haut à droite avec l'anneau polymérasique et les trois domaines additionnels responsables de l'ajout et la méthylation de la coiffe (Rahmeh et al. 2010).

La séquence primaire ainsi que les régions conservées sont dessinées à l'échelle indiquée sur la figure.

Aucune structure cristallographique de polymérase de *Mononegavirales* n'a été, même que partiellement, résolue. Cependant l'organisation modulaire de la polymérase a été observée par microscopie électronique. L'analyse de fragments purifiés de polymérase obtenus par digestion ménagée à la trypsine a permis de placer les régions conservées dans les structures visualisées par microscopie électronique (figure 1.17). La polymérase est composée d'un anneau (figure 1.17 fragment 1-1114) auquel sont associées trois structures globulaires. Ainsi, l'anneau qui est composé des domaines I-IV porte l'activité de synthèse d'ARN. La structure globulaire proche de l'anneau contient le domaine V impliqué dans la formation de la coiffe des ARNm. Enfin, les deux structures globulaires restantes sont

composées du domaine VI et de régions charnières, elles interviennent dans la méthylation de la coiffe (Rahmeh et al. 2010).

• Le domaine polymérasique.

Le motif « QGDNQ » (position 712-716 aa chez VSV, région conservée III, figure 1.18) impliqué dans la réaction de synthèse d'ARN au moment de la transcription et de la réplication du génome viral est l'un des motifs les mieux conservés parmis les ARN polymérases ARN dépendantes dans les règnes végétal et animal, chez les bactériophages (Kamer & Argos 1984) ainsi que dans l'ordre des *Mononegavirales*. Pour cet ordre, il est situé dans une boucle se trouvant dans la région conservée III (Poch et al. 1989, 1990, Blumberg et al. 1988).



FIGURE 1.18 – Alignement de séquence de la région conservée III des polymérases de Rhabdovirus.

Les codes comme suit P03523, Q5K2K3, Q8B6R7, Q9E784, Q8B6J5, Q9Q5P0, P0C568 correspondent à VSV Indiana virus (VSNJV), Isfahan virus (ISFV), Flanders virus, Bovine ephemeral fever virus (BEFV), Rabies virus (RABV), Australian bat lyssavirus (ABLV) et Mokola virus (MOKV).

Les acides aminés de couleur blanche représentés en fond rouge sont identiques et ceux en fond blanc et de couleur rouge sont similaires.

L'alignement de séquence a été fait avec UniProt Knowledgebase (UniProtKB) disponible depuis Expasy (Apweiler et al. 2004) et les figures ont été générées avec ESPript 2.2 (Gouet et al. 1999).

Un alignement de séquence de Rhabdovirus montre que le motif « $QGDNQ \gg$ est conservé parmis ces polymérases (Barik et al. 1990, Choi et al. 1992, Feldhaus & Lesnaw 1988, Galinski et al. 1988, Kawano et al. 1991, Parks et al. 1992, Schubert et al. 1984, Shioda et al. 1986, Tordo et al. 1988, Yusoff et al. 1987).

Une comparaison entre les ARN polymérases ARN dépendantes et les ADN polymérases révèle certaines similarités de séquences (Kamer & Argos 1984, Poch et al. 1989). Chez les virus à ARN de polarité positive le motif conservé se présente sous la forme « GDD » et est entouré de régions hydrophobes responsables de la reconnaissance de l'ARN viral.

Des mutagénèses ponctuelles du motif « QGDNQ » révèlent l'importance de l'Asp (D) en position 714. Par ailleurs, lorsque l'Asn (N) en position 715 est mutée en Asp (D), formant un motif « GDD », la polymérase garde son activité transcriptionnelle, confirmant la correspondance entre le motif « GDN » des virus à ARN négatif et le motif « GDD » des virus à ARN positif (Sleat & Banerjee 1993). La glutamine en position 716 est également nécessaire à l'activité de transcription.

Il est important de noter que ce même motif « GDD » se trouve aussi dans le fragment de Klenow de la polymérase I d'*E.coli* mais également aussi dans l'ADN polymérase des bactériophages T7 et $Q\beta$. Ce motif pourrait jouer le rôle du site actif de la synthèse d'ARN en fixant les métaux divalents (Kamer & Argos 1984, Poch et al. 1989, Delarue et al. 1990).

• Maturation des ARNm et les domaines responsables à la fabrication.

Divers stratégies sont utilisées par différents virus qui produisent des ARN coiffés et polyadénylés semblables aux ARN cellulaires. Ces stratégies s'appliquent aux virus à ADN, à ARN positif mais aussi aux virus à ARN négatif dont le génome est ségmenté.

- Certains virus dont le virus de la vaccine (Poxvirus), le virus de la Dengue (Flavivirus), ou le virus Bluetongue possèdent leurs propres enzymes responsables à la formation de coiffe des ARNm (Barbosa & Moss 1978, Egloff et al. 2007, Sutton et al. 2007, Zhou et al. 2007, Dong et al. 2008, 2010, Yap et al. 2010).
- Le virus de la grippe appartenant à la famille des Orthomyxovirus, utilise un mécanisme de vol de coiffe (cap snatching) (Bouloy et al. 1978, Plotch et al. 1981, Li et al. 2001, Dias et al. 2009).
- Le virus de la stomatite vésiculaire possède une polymérase possédant toutes les activités enzymatiques nécessaires à la fabrication d'ARNm matures à partir de l'ARN viral (Baltimore et al. 1970). Pour arriver au même composer final elle utilise un mécanisme différent de celui des cellules eucaryotes(voir figure 1.19B) (Ogino & Banerjee 2007, Ogino et al. 2010).

Chez les eucaryotes et chez certains virus à ADN (virus de la vaccine), ou à ARN double brin (réovirus), la coiffe (Gp_{γ} - $p_{\beta}p_{\alpha}N$) est synthétisée en deux étapes impliquant une ARN 5'-triphosphatase (RTPase) [1] et une GTP/ARN guanyltransférase (GTase) [1,2] (voir figure 1.19A page 35) (Li et al. 2008, Ogino & Banerjee 2007, Ogino et al. 2010).

La première étape est une étape de déphosphorylation en 5' de l'ARN triphosphate produisant un ARN diphosphate [1] qui sert de récepteur pour le groupe guanyl. Les étapes [2,3] conduisent à la formation de l'ARN coiffé. L'enzyme attaque le $p_{\gamma}p_{\beta}p_{\alpha}$ -G en position α créant une liaison covalente E- p_{α} G et libérant un pyrophosphate (PPi) [2]. La troisième étape implique le transfert du groupe p_{α} -G à partir de l'intermédiaire (E- p_{α} G [2]) vers l'ARN diphosphate ([1]), formant ainsi l'ARN mature (G- $p_{\alpha}p_{\beta}p_{\alpha}$ ARN [3]).

L'ARN polymérase ARN dépendante initie la synthèse de l'ARN messager au premier gène (Whelan & Wertz 2002, Tekes et al. 2011), transcrit une trentaine de bases, marque une pause de quelques secondes, ajoute la coiffe (Stillman & Whitt 1999, Whelan & Wertz 1999b, Wang et al. 2007, Tekes et al. 2011) et procède au contrôle de qualité. A ce stade, deux issues sont possibles : si la coiffe est présente la polymérase continue la transcription de l'ARN viral, si la coiffe n'est pas ajoutée la polymérase arrête la transcription et l'ARNm est dégradé (Stillman & Whitt 1999).

Afin d'avoir un ARNm mature (5'-G-N7_{CH3}-ppp-A-2'OH_{CH3}-ARN-3') plusieurs activités

enzymatiques sont nécessaires : une activité de phosphatase - hydrolyse du GTP et de l'ARN, polyribonucléotidyltransférase (PRNTase), une guanyltransférase (GTase) qui servira à coiffer l'ARN et une dernière méthyltransférase (MTase) qui servira à la méthylation en 7N du guanyl et le 2'OH du ribose de l'adénine.



FIGURE 1.19 – Mécanisme conventionnel et non conventionnel de formation de coiffe des ARNm. (Ogino & Banerjee 2011)

A : Le mécanisme conventionnel est effectué par un domaine GTase qui hydrolyse un GTP (en rouge) en GMP et le transfère sur pp-ARN (en bleu) via la formation d'une liaison covalente (lisyl-N^{ε})-GMP ou encore E-pG. Le résidu K dans le motif conservé KxDG agit comme un nucléophile.

B : Chez VSV la PRNTase transfère p-ARN (en bleu) sur un GDP (en rouge) via un intermédiaire de liaison covalente avec (histidyl- N^{ϵ_2})-pARN ou encore L-pARN. Le résidu histidine est conservé dans le motif HR est joue le rôle de nucléophile pour la formation de l'intermédiaire réactionnel.

Les étapes réactionnelles sont une guanyltriphosphatase (GTPase [4]) qui produit du GDP et du Pi et une polyribonucléotidyltransférase (PRNTase [5,6]) (Ogino & Banerjee 2007, Ogino et al. 2010). L'intermédiaire réactionnel dans l'étape [5] de la synthèse de la coiffe est L- p_{α} ARN tandis que pour les autres polymérases citées plus haut il s'agit d'E- p_{α} G. Egalement avec la polymérase de VSV [6] c'est le GDP qui est transféré sur le complexe L- p_{α} ARN alors que pour les autres enzymes c'est le GMP qui est transféré sur un ARN diphosphate. Toutes ces étapes de formation de la coiffe chez le virus de la stomatite vésiculaire ont été confirmées par spéctrométrie de masse prouvant que l'histidine en position 1227 (voir figure 1.19 B motif HR page 35) est impliquée dans la formation d'une liaison covalente entre l'enzyme et l'ARN monophosphate [5] (Ogino et al. 2010).

La fabrication de la coiffe chez le virus de la stomatite vésiculaire est séquence spécifique (Ogino & Banerjee 2007, Tekes et al. 2011). Dans le génome viral chaque gène commence avec la même séquence de 5 bases (3'-AACAG-5') qui est nécessaire et suffisante pour être coiffée par la polymérase (Ogino & Banerjee 2007). En revanche l'ARN leader possédant une séquence (3'-ACGAA-5') n'est pas coiffé. La mutation d'une des bases 1, 3 ou 5 de la séquence (3'-AACAG-5') conduit à une absence de formation

de la coiffe.

Des alignements de séquence et des études de mutagénèse ponctuelle ont permis de localiser les régions de la protéine nécessaires aux activités de PRNTase, GTase et MTase.



FIGURE 1.20 – Alignement de séquence de la région conservée V des polymérases de Rhabdovirus.

Les différentes séquences alignées suivent le même ordre que l'alignement de la région conservée III.

Les acides aminés identiques et similaires sont représentés de la même façon.

Les neuf acides aminés les plus essentiels parmis les dix-sept acides aminés importants à la maturation de l'ARNm sont représentés avec des flèches et le numéro correspondant pour le virus de la stomatite vésiculaire à chaque acide aminé est indiqué en bas de la flèche (Li et al. 2008).

Deux régions conservées sont responsables à la formation de l'ARNm mature : les régions V et VI (Håkansson et al. 1997, Ogino et al. 2005, Li et al. 2005, Grdzelishvili et al. 2005, 2006, Li et al. 2006, 2008, Ogino & Banerjee 2007, Ogino et al. 2010) (figure 1.17 page 32).

Des alignements de séquences de la région conservée V des virus à ARN négatif non-ségmenté révèlent la présence de 17 acides aminés conservés pour l'activité d'addition de coiffe (Li et al. 2008, Ogino et al. 2010).

En 2007, le groupe de Whelan a montré par des tests de transcription et de formation de coiffe *in vitro* que la maturation de l'ARNm implique les acides aminés suivants : P-1104, Y-1152, G-1154, T-1157, H-1227, R-1228, H-1294, H-1296, C-1302 parmis les dix-sept précédemment identifiés (figure 1.20 page 36). La mutation d'un de ces résidus en alanine abolit complètement la maturation de l'ARNm (Li et al. 2008) et aboutit à une terminaison prématurée de la transcription (Li et al. 2008). (voir figure 1.21)



FIGURE 1.21 – La tétrade impliquée dans la formation de coiffe (Li et al. 2008).

A : Représente l'activité d'addition de coiffe des différents mutants de la polymérase du VSV. La réaction de transcription a été faite avec $[\alpha^{-32}P]$ GTP et la coiffe a été digérée avec terminator alcaline phosphatase (TAP). La migration est effectué en chromatographie à couche mince sur des plaques (TLC) avec un tampon de 1.2 M LiCl₂. Les mutants G-713A et D-714A du site catalytique « QGDNQ »sont utilisés comme contrôle négatif.

B : Représente une quantification avec un PhosphoImager de l'activité enzymatique présentée en "A".

Ces mutations ponctuelles et ces analyses d'activité transcriptionnelle révèlent un motif essentiel $G_{1154}xxT_{1157}[n]H_{1227}R_{1228}$ pour la synthèse d'ARNm. La mutation d'un de ces résidus provoque une disparition totale des ARNm (figure 1.21). La mutation des résidus $H_{1227}R_{1228}$ aboutit à une formation d'ARNm hétérogènes de petite taille (Li et al. 2008). L'H₁₂₂₇ est un des acides aminés conservés dans la région V qui est importante pour l'addition de coiffe. (figure 1.19 page 35, figure 1.20 page 36 et figure 1.21). Des mutations ponctuelles de cette histidine conduisent à la formation d'ARN de petites tailles (transcrits abortifs) sans coiffe.



FIGURE 1.22 – Alignement de séquence de la région conservée VI des polymérases de Rhabdovirus.

Les différentes séquences alignées suivent le même ordre que l'alignement de la région conservée III.

Les acides aminés identiques et similaires sont représentés de la même façon.

La région d'alignement est un peu élargie car les résidus comme D-1762, K-1795 et E-1833 se trouvant après la région conservée VI sont importants à l'activité de méthylation de la coiffe. (Li et al. 2008).

Le second domaine conservé VI se trouvant en C-terminal de la protéine est responsable de la méthylation du 7N-guanyl et 2-O' de l'adénosine, qui conduit à la formation de l'ARN mature (**7m**GpppA^m) (Rhodes et al. 1974, Abraham et al. 1975, Rose 1975, Rhodes & Banerjee 1975, Testa & Banerjee 1977, Li et al. 2005, Grdzelishvili et al. 2005, Li et al. 2006, Grdzelishvili et al. 2006, Galloway et al. 2008, Murphy & Grdzelishvili 2009).

Il existe plusieurs séquences conservées dans le domaine VI. L'une d'entre elles est impliquée dans

la fixation du donneur de méthyle (SAM) (Grdzelishvili et al. 2006). Ce motif (G₁₆₇₀ x G₁₆₇₂ x G₁₆₇₄ G₁₆₇₅) (figure 1.22) est conservé parmis toutes les polymérases de virus à ARN de polarité négative et positive qui fabriquent leur propre coiffe d'ARN messager mais également chez des eucaryotes comme A. thaliana, S. cerevisiae, H. sapiens et autres (Grdzelishvili et al. 2006). La méthylation des deux groupes 7N guanyl et 2O' ribose s'effectue après l'addition du groupe guanyl (Rhodes et al. 1974, Moyer et al. 1975, Rose 1975, Keene & Lazzarini 1976, Rhodes & Banerjee 1975, Moyer & Banerjee 1976, Testa & Banerjee 1977, Hercyk et al. 1988).

Dans les années 2000, la structure de la méthyltransférase (MTase) d'*E.coli* (RrmJ) a été résolue (Bügl et al. 2000). Les résidus responsables de l'activité catalytique appelés « tétrade » $K_{38}D_{124}K_{164}E_{199}$ sont primordiaux pour l'activité de transfert du groupe méthyle CH₃ à partir du donneur SAM (Schluckebier et al. 1995). Il est important de signaler que le repliement tridimensionnel de cette protéine est similaire à celui de la méthyltransférase VP39 du virus de la vaccine (Hodel et al. 1996).





A : Représente la structure cristallographique de la méthyl transférase de *E.coli* en complexe avec le donneur de méthyle (SAM) (code PDB : 1 EIZ). La structure est présentée en surface électrostatique, calculée avec Chimera colorée du bleu $(+5 \, \rm kT/e)$ au rouge $(-5 \, \rm kT/e)$, de manière à visualiser plus facilement la cavité où le SAM se positionne.

B : Représente le domaine de méthyltransférase du VSV construit par homologie d'après (Galloway et al. 2008). La figure en surface électrostatique permet de visualiser la cavité. Celle-ci ressemble à celle de la figure A et la tétrade catalytique KDKE tapisse la cavité.

C : Représente la méthyltransférase du VSV en structure secondaire. Les tubes violet représentent les hélices α et les brins β sont de couleur jaune. Les boucles sont grises. Les acides aminés conservés impliqués dans le site actif sont également montrés.

Depuis la résolution de la structure de la RrmJ, plusieurs structures cristallographiques de méthyltransférases isolées ou en complexe avec des petits ARN coiffés, avec des analogues de coiffe, ou des inhibiteurs de virus à ARN positif (Flavivirus, West Nile virus, le virus Murray Valley et autres) ont été résolues (Hodel et al. 1996, Egloff et al. 2002, Benarroch et al. 2004, Malet et al. 2007, Zhou et al. 2007, Assenberg et al. 2007, Dong et al. 2008, Geiss et al. 2009, Bollati et al. 2009, Milani et al. 2009, Dong et al. 2010, Yap et al. 2010). Cependant malgré la disponibilité de toutes ces structures de méthyltransférases de virus à ARN positif, aucune structure à haute résolution n'a pu être obtenue

pour un domaine méthylase d'une polymérase de virus à ARN négatif. Des analyses bioinformatiques d'alignement de séquences et de modélisation par homologie suggèrent que le domaine conservé VI possède une structure tridimensionnelle similaire à celle de RrmJ, évoquant l'hypothèse d'un ancêtre commun (Bujnicki & Rychlewski 2002, Ferron et al. 2002, Galloway et al. 2008) (figure 1.23 page 38).

Dans la voie classique de synthèse de la coiffe, les méthyltransférases de cellules eucaryotes (Muthukrishnan et al. 1976, Furuichi et al. 1977) effectuent une première méthylation du 7-N de la guanosine, suivi d'une méthylation en 2'-OH de l'adénosine. Les deux méthylations sont réalisées indépendamment par deux enzymes différentes (Furuichi & Shatkin 2000). Cette réaction de synthèse d'un ARNm mature est entièrement conservée parmis tous les eucaryotes (Shuman 2001), et l'activité enzymatique de chaque étape a été bien étudiée (Hâkansson & Wigley 1998, Lima et al. 1999, Fabrega et al. 2004).

Le virus de la stomatite vésiculaire possède son propre mécanisme de maturation de coiffe : une première méthylation s'effectue en position 2'OH du ribose suivi de 7-N du guanosine. (Abraham et al. 1975, Testa & Banerjee 1977, Li et al. 2006).

Dans les années 2005, par des mutations ponctuelles dans le domaine conservé VI, les groupes de Moyer et Whelan montrent que les résidus conservés (figure 1.22 page 37) K1651, D1671, G1674, D1762, K1795 et E1833 sont responsables de l'activité méthyltransférase (Grdzelishvili et al. 2005, Li et al. 2005, Grdzelishvili et al. 2006). Il est important de noter que la tétrade $K_{38}D_{124}K_{164}E_{199}$ de RrmJ, responsable de l'activité catalytique, est présente chez VSV et il est proposé que les résidus adoptent une conformation identique à celle de RrmJ dans le site actif (Bujnicki & Rychlewski 2002, Galloway et al. 2008) (figure 1.22 page 37 et figure 1.23). Toutes les analyses bioinformatiques et biochimiques confirment l'importance des acides aminés $K_{1651}D_{1762}xK_{1795}xE_{1833}$ dans la méthylation du 7-N et 2-O'. Il a été également démontré que des mutations dans le motif conservé $G_{1670} \times G_{1672} \times G_{1674} G_{1675}$ ont une influence sur l'activité de MTase. Quatre glycines sont conservées dans la mARN guanine-N7cap MTase des mammifères comme *H. sapiens, A. thaliana, S. cerevisiae*, mais également chez mARN nucléoside-2'cap MTase du virus de la vaccine (VP39), du virus West Nile, du virus de la fièvre jaune et du Réovirus $\lambda 2$ suggérant fortement que les deux activités (MTase en 7-N et 2'O du ribose) utilisent le même site de fixation du SAM (Grdzelishvili et al. 2005).

Le modèle structural par homologie a été obtenu *in silico* du domaine VI de la polymérase de VSV (figure 1.23 B,C page 38), responsable pour l'activité 2'-O-ribose méthyltransférase, qui a permis une analyse à *posteriori* de certaines données de mutagénèse (Galloway et al. 2008). Le modèle obtenu, homologue à la méthyltransférase RrmJ d'*E.coli* met en évidence les résidus importants dans la méthylation de la coiffe de l'ARNm. La mutation de ces résidus abolit la maturation de la coiffe (Li et al. 2008).

Deux publications récentes suggèrent que la polymérase est flexible. (Rahmeh et al. 2010, Tekes et al. 2011). Les classifications d'images de microscopie électronique présentées par Rahmeh et al. (2010) suggèrent des changements de conformations. Ces auteurs proposent un modèle selon lequel le domaine V serait responsable de l'addition de guanyl sur l'ARN (figure 1.17 page 32), puis dans un

deuxième temps un changement de conformation se produisait au niveau du C-terminal (domaine VI) qui permettrait de méthyler la coiffe.

• Le mécanisme de fabrication de la queue polyA.

Chez le virus de la stomatite vésiculaire, une séquence consensus conservée pour chaque gène est suffisante pour la terminaison de la transcription. A l'heure actuelle des protéines cellulaires responsables à la terminaison n'ont pas été identifiées. Une séquence poly U_7 est impliquée dans la formation de la queue polyA (Schubert et al. 1980). La polymérase produirait une queue polyA entre 100-300 polyA par un phénomène de dérapage au niveau de la séquence polyU. Des analyses de mutagénèse dirigée montrent qu'un allongement de la séquence polyU a un effet minime sur la terminaison de la transcription, par contre une réduction aboutit à un arrêt de la transcription (Barr et al. 1997*a*).

B. Régulation moléculaire de la transcription/réplication dans la cellule

La transcription des gènes viraux est gouvernée par des signaux internes propre au génome du virus. Durant la réplication, la polymérase se fixe à l'extrémité du génome viral et ignore tous les signaux de régulation de la transcription afin de produire un ARN de polarité positive qui sera encapsidé par la nucléoprotéine et servira par la suite pour la fabrication d'un nouveau génome qui sera également encapsidé (Banerjee 1987, Fields 1996, Knipe 2001).

Jusqu'en 2002 aucun complexe de transcriptase ou de réplicase, impliquant des protéines cellulaires, n'avait été mis en évidence dans les cellules infectées. Pourtant des mutations à des endroits spécifiques de la polymérase affectent soit le processus de transcription soit celui de réplication (Wertz 1978, Perlman & Huang 1973, Hunt et al. 1976). Chez VSV, il a été démontré que des phosphorylations de la phosphoprotéine seraient impliquées dans le processus de régulation (Pattnaik et al. 1997). Un changement de conformation du substrat N-ARN pourrait jouer aussi un rôle de régulation. Chuang et al. ont démontré que la mutation R146H au niveau de la nucléoprotéine perturbe le processus d'initiation en 3' vis-à-vis de l'initiation interne au début du gène de N (Chuang & Perrault 1997). Des analyses de cryo-électromicroscopie sur les nucléocapsides de Sendai montrent quatre formes différentes (Egelman et al. 1989). Un processus de régulation similaire que celui des virus de la famille des *Paramyxoviridae* pourrait être conservé dans la famille des *Rhabdoviridae*. Dans les années 80 une hypothèse a été émise que le complexe N°P jouerait un rôle dans la régulation de la transcription/réplication (Blumberg et al. 1983, Gupta & Banerjee 1997).

Récemment le groupe de Dr. Banerjee a isolé deux complexes polymérasiques différents. Un complexe formé uniquement de protéines virales qui possède une activité de réplication. L'autre impliquant également des protéines cellulaires a une activité de transcription (voir figure 1.24)



FIGURE 1.24 – Le complexe de transcriptase et réplicase du VSV (Qanungo et al. 2004).

Sur la figure est dessiné l'ARN de polarité négative du 3' vers le 5'. Le promoteur « leader » est indiqué en début de séquence et ensuite le gène codant pour la nucléoprotéine. De manière schématique sont représentés la polymérase (L) en jaune foncé, son cofacteur - la phosphoprotéine (P), le facteur d'élongation cellulaire EF1 α en rouge, une guanyltransférase cellulaire (GT) en violet et HSP60 en jaune.

Le complexe de réplication semblerait contenir uniquement la polymérase et la phosphoprotéine. L'ARN de polarité positive est dessiné avec un 5'ppp à son extrémité. Le complexe N°P est représenté avec deux boules une blanche pour N et une verte pour P.

Ces résultats suggèrent que la régulation de la transcription et de la réplication au cours du cycle viral passe par l'assemblage de différents complexes polymérasiques.

1.3 Objectifs du travail de thèse

Avant mon arrivée, le laboratoire avait débuté une caractérisation détaillée de la phosphoprotéine des *Rhabdoviridae*. La phosphoprotéine intervient dans plusieurs complexes protéiques avec la nucléoprotéine et avec la polymérase virale L et se trouve donc au coeur des processus de réplication et de transcription (Leyrat, Jensen, Ribeiro, Gérard, Ruigrok, Blackledge & Jamin 2011, Ivanov et al. 2011).

Ces travaux ont permis de déterminer que les phosphoprotéines des *Rhabdoviridae* existent sous la forme de dimères non-globulaires en solution (Gerard et al. 2007). Les structures de la région C-terminale de la phosphoprotéine du virus de la rage (Mavrakis et al. 2004) et du VSV (Ribeiro et al. 2008) ainsi que celle du domaine central de VSV avaient été déterminées. Plusieurs études avaient permis de comprendre l'organisation modulaire de la phosphoprotéine définissant ainsi les différentes régions structurales et fonctionnelles de la protéine. Plus récemment, des études par spectroscopie RMN et par diffusion des rayons X aux petits angles ontpermis la modélisation de la structure complète de la phosphoprotéine de VSV (Leyrat et al. 2010, Leyrat, Jensen, Ribeiro, Gérard, Ruigrok, Blackledge & Jamin 2011).

Une première partie de mon travail a consisté dans la détermination de la structure du domaine central de la phosphoprotéine du virus de la rage afin de compléter la caractérisation des domaines structurés des P du VSV et du RAV. Une seconde partie avait pour objet de caractériser la grande sous-unité du complexe polymérasique (L) du virus de la stomatite vésiculaire (souche Indiana). Ce travail consiste en plusieurs parties incluant une tentative d'expression de différentes constructions de domaines isolés en bactéries ainsi que de l'expression en cellule d'insectes de la polymérase seule ou en complexe avec son cofacteur, la phosphoprotéine. Au cours de la dernière année de ma thèse j'ai réussi à purifier la polymérase à partir de virus. Ceci nous a permis de réaliser une étude par microscopie électronique et des études fonctionnelles de la transcription à partir du complexe N-ARN viral. Chapitre 2

Caractérisation structurale et fonctionnelle de la phosphoprotéine du virus de la rage

Chapitre 2. Caractérisation structurale et fonctionnelle de la phosphoprotéine du virus 44 de la rage

Sommaire

.1 Introduction des phosphoprotéines de <i>Mononegavirales</i>	44
APER I : Modular Organization of the Rabies Virus Phosphoprotein	47
.2 Conclusion	66
.3 Approches envisagées pour la résolution de la structure du domaine central	
de la phosphoprotéine du virus de la rage	67
2.3.1 Détermination du groupe d'espace et collecte des jeux de données	69
.4 Stratégies utilisées pour obtention des phases	70
2.4.1 Construction du modèle	72
PAPER II : Structure of the Dimerization Domain of the Rabies Virus	
Phosphoprotein	76
.5 Conclusion	80

2.1 Introduction des phosphoprotéines de Mononegavirales

Des études antérieures de la P de VSV réalisées par des méthodes de co-immunoprécipitation et double-hybride combinées avec d'astucieuses approches pour la préparation de mutants de délétions suggéraient une organisation modulaire de cette protéine. Trois régions assurant de façon apparemment indépendante diverses fonctions ainsi qu'une région variable, dite « charnière »avaient été identifiées (Takacs et al. 1992). Donc, la phosphoprotéine avait été séparée en quatre régions :



FIGURE 2.1 – Organisation modulaire de la phosphoprotéine de *Rhabdoviridae* (Takacs et al. 1992).

La phosphoprotéine entière est présentée avec les trois domaines et la région charnière. Les points noirs dans la région I représente les sites de phosphorylations par la CKII. Les chiffres en bas de la figure montrent les bordures des régions. N signifie la partie N-terminale et C la partie C-terminale.

- La région I : du côté N-terminal de la protéine est riche en résidus acides et possède des sites de fixation pour la N et la L. Cette région renferme également des sites de phosphorylation essentiels pour la régulation de la transcription (figure 2.1 région I) (Barik & Banerjee 1992a,b)
- La région II : (211-240aa) fixe aussi la L et contient des sites de phosphorylation qui contrôlent la réplication (Hwang et al. 1999).
- La région III (241-265aa) : est riche en résidus basiques et est essentiel pour la fixation sur la matrice N-ARN (Das et al. 1997).

La région charnière (HINGE) est hypervariable et joue un rôle dans l'assemblage des particules virales (Das & Pattnaik 2005)

Plus récemment, l'organisation structurale des phosphoprotéines de RAV et de VSV ainsi que l'utilisation d'outils de prédictions ont permis de confirmer une organisation modulaire dans laquelle des domaines structurés sont assemblés avec des régions désordonnées (Gerard et al. 2009).

Des domaines structurées avaient été identifiées par protéolyses ménagées : le domaine central P_{CED} chez VSV (Ding et al. 2006), le domaine P_{CTD} de VSV et RAV (Mavrakis et al. 2004, Ribeiro et al. 2008) et les domaines P_{CED} et P_{CTD} du virus de Sendai (SV) (Tarbouriech, Curran, Ruigrok & Burmeister 2000, Blanchard et al. 2004).

Des études réalisées avec la phosphoprotéine de plusieurs Paramyxovirus indiquent que ces protéines contiennent des régions désordonnées et des prédictions de désordre suggéraient qu'il en était de même pour les P de Rhabdovirus (Karlin et al. 2003).

Depuis les années 1990, de nombreuses protéines contentant des régions intrinsèquement dépliées ont été découvertes. 30 % des protéines de cellules eucaryotes contiennent des régions désordonnées (Pérez et al. 2009).

Divers algorithmes de prédictions de désordre ont été développés, afin de localiser encore plus précisément les régions désordonnées et les régions structurées. Des programmes ont été développés pour prédire leur localisation à partir de la séquence en acide aminé. De tels serveurs ont été utilisés pour localiser les différentes régions de la phosphoprotéine du virus de la rage ainsi que celle du virus de la stomatite vésiculaire.

La publication intitulée « Modular Organization of Rabies Virus Phosphoprotein » décrit l'analyse informatique et les résultats de la dissection des P de RAV et VSV qui démontrent bien l'existence de domaines globulaires structurés et de régions désordonnées.

Ma contribution dans cette publication consistait à apporter la caractérisation biophysique par diffusion de lumière aux petits angles du domaine central de la phosphoprotéine du virus de la rage. Article I : Modular Organization of Rabies Virus Phosphoprotein

<u>Auteurs</u> : Francine C.A. Gerard, Euripedes de Almeida Ribeiro Jr, Cédric Leyrat, **Ivan Ivanov**, Danielle Blondel, Sonia Longhi, Rob W.H. Ruigrok and Marc Jamin

Publié dans : Journal of Molecular Biology (2009) 388, 978-996



Available online at www.sciencedirect.com





Modular Organization of Rabies Virus Phosphoprotein

Francine C. A. Gerard¹[†], Euripedes de Almeida Ribeiro Jr¹[†], Cédric Leyrat¹, Ivan Ivanov¹, Danielle Blondel², Sonia Longhi³, Rob W. H. Ruigrok¹ and Marc Jamin^{1*}

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

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

³Architecture 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

Received 12 November 2008; received in revised form 23 March 2009; accepted 25 March 2009 Available online 31 March 2009

A phosphoprotein (P) is found in all viruses of the Mononegavirales order. These proteins form homo-oligomers, fulfil similar roles in the replication cycles of the various viruses, but differ in their length and oligomerization state. Sequence alignments reveal no sequence similarity among proteins from viruses belonging to the same family. Sequence analysis and experimental data show that phosphoproteins from viruses of the Paramyxoviridae contain structured domains alternating with intrinsically disordered regions. Here, we used predictions of disorder of secondary structure, and an analysis of sequence conservation to predict the domain organization of the phosphoprotein from Sendai virus, vesicular stomatitis virus (VSV) and rabies virus (RV P). We devised a new procedure for combining the results from multiple prediction methods and locating the boundaries between disordered regions and structured domains. To validate the proposed modular organization predicted for RV P and to confirm that the putative structured domains correspond to autonomous folding units, we used two-hybrid and biochemical approaches to characterize the properties of several fragments of RV P. We found that both central and C-terminal domains can fold in isolation, that the central domain is the oligomerization domain, and that the C-terminal domain binds to nucleocapsids. Our results suggest a conserved organization of P proteins in the Rhabdoviridae family in concatenated functional domains resembling that of the P proteins in the Paramyxoviridae family.

© 2009 Elsevier Ltd. All rights reserved.

RNA viruses belonging to the Paramyxoviridae

(Sendai virus (SV), measles virus, respiratory syncytial virus, and Nipah virus), the Filoviridae (Ebola

virus and Marburg virus), and the Bornaviridae

(Borna disease virus), which are grouped together

with the Rhabdoviridae in the order of the Mono-

negavirales (MNV). In all MNV viruses, P is part of a

set of five common proteins encoded in the same order along the negative strand genome from the 3' extremity to the 5' extremity: the nucleoprotein (N),

the phosphoprotein (P), the matrix protein (M), the

glycoprotein (G) and the RNA-dependent RNA poly-

Edited by J. Karn

Keywords: phosphoprotein; rabies virus; Sendai virus; intrinsically disordered proteins; protein domains

Introduction

The phosphoprotein is an essential component of the replication machinery of viruses from the Rhabdoviridae family, which includes rabies virus (RV) and vesicular stomatitis virus (VSV). P has similar functions in other non-segmented negative strand

merase (L). The nucleoprotein binds to the viral RNA, forming long helical ribonucleoprotein complexes that serve as templates for RNA transcription and replication. Protein P has multiple roles at different stages of the viral cycle, forming a complex with the nascent nucleoprotein that prevents the

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

[†] F.C.A.G. and E.dA.R. contributed equally to this work. Abbreviations used: RV, rabies virus; VSV, vesicular stomatitis virus; SV, Sendai virus; MNV, Mononegavirales; N, nucleoprotein; P, phosphoprotein; M, matrix protein; G, glycoprotein; L, RNA-dependent RNA polymerase; aa, amino acid(s); CTD, C-terminal domain; SEC, size-exclusion chromatography; MALLS, multi-angle laser light-scattering; RI, refractometry.

association of the latter with host-cell RNA and forming a two-subunit viral RNA polymerase complex with protein L.

Recent studies have shown that P proteins from viruses of the Paramyxoviridae contain long disordered regions that alternate with structured domains,^{1–5} and various structural and functional data are available for either VSV P or RV P that suggest the existence of common functional regions. Also, a previous analysis of the amino acid sequence suggested that *Rhabdoviridae* P proteins contain dis-ordered or flexible sequences.¹ The P proteins from both viruses are highly sensitive to proteolysis; domains of RV and VSV P for which the structure were solved by X-ray crystallography were identified by limited proteolysis.^{6–8} The N-terminal fragments of both VSV P and RV P interact with N in the N^{0} -P complex, which is sufficient for maintaining N (referred to as N^0) in a soluble, RNA-free form.^{9,10} A structured domain in the central part of VSV \underline{P} is responsible for the dimerization of the protein,⁷ in agreement with our recent demonstration that recombinant, full-length P proteins from both VSV and RV form dimers in solution.¹¹ The oligomerization domain of RV P was shown to be located in the central region (amino acids (aa) 52–189),12 although its precise location and boundaries were not clearly determined.13 The C-terminal domains of RV and VSV P bind to N-RNA complexes, and their structures were solved by X-ray crystallography⁶ and NMR,¹⁴ respectively.

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

For many years after the first protein structures were determined,^{23–25} biological activity of proteins has been associated with folding of the polypeptide chain into a well-defined 3D structure. This idea was supported by the discovery that the biological activity disappears when proteins unfold.^{26,27} However, this structure/function paradigm was recently challenged with the discovery that many proteins are fully disordered or contain disordered regions under physiological conditions.^{28–31} Also, several proteins were found to exist in a disordered form in isolation

but to fold into a well-defined structure upon binding to a partner.^{32–36} More surprising was the realization that disorder was required for, or participated in, the biological function of some proteins.^{28,31,33,37–41} Several approaches were developed for sorting out disordered proteins from structured proteins on the basis of their amino acid sequence.^{29,38,42} Application of such algorithms to whole genomes suggested that as much as one-third of the proteins from eukaryotic organisms contain long disordered regions.^{38,42} This raised a new challenge in structural biology for locating disordered regions and folded domains within the protein. Various algorithms were developed for predicting the location of disordered regions from the amino acid sequence (for recent reviews, see Refs. 5 and 43). Some of these algorithms rely on the relationship between the amino acid sequence and the disordered state of the polypeptide chain (SEG, FoldIndex, Globplot, HCA plot, and NORSp).^{44,45} For example, it was recognized early that a protein region rich in charged residues and poor in hydrophobic residues is likely to be disordered.^{29,44} Other programs rely on trained neural networks and use sets of known disordered proteins (RONN, DisEMBL, and Disopred2), or combine both approaches (PONDR). However, different methods may yield different predictions and it is difficult to choose the method appropriate to a given protein.

Here, we analyzed the amino acid sequence of the phosphoproteins from SV, VSV, and RV using algorithms available through WEB servers for predicting the location of disordered regions and of secondary structure elements, and we analyzed the amino acid sequence conservation. To avoid having to choose among the prediction methods, we devised simple scoring procedures for integrating the predictions from the different algorithms, and for yielding consensus locations of boundaries between ordered and disordered regions. As test cases, we applied this approach to SV and VSV P for which the location of several structured domains and disordered regions are known from previous structural studies.^{2,3,7,14,46} Then, we obtained similar predictions from analysis of the amino acid sequence of RV P. The properties of the predicted structured domains have been tested in a cellular context using a yeast two-hybrid approach, and in vitro by producing and characterizing recombinant protein fragments. Circular dichroism was used to determine the secondary structure content and sizeexclusion chromatography (SEC) combined with detection by multi-angle laser light-scattering was used to determine the molecular size, molecular mass and the stoichiometry of the different protein fragments. Native gel mobility-shift experiments were used to test the interaction of the various domains with the nucleoprotein-RNA complex, which represents the physiological partner. Altogether, these results provided by the in vivo and in vitro characterizations of the various P fragments confirmed the *in silico* analysis and allowed us to propose a model for the modular organization of RV P.

Results

Predictions from the amino acid sequence of SV, VSV and RV P

Strategy for predicting the location of disordered and structured regions

We attempted to predict the boundaries between structured and unstructured regions in the SV, VSV and RV P proteins. The SV P was used as a test case, since knowledge about the positions of some of its structured and unstructured domains is available.^{2,3,7,8,14,47,48} The location of disordered regions in the sequence of the P proteins was predicted from the amino acid sequence by using 16 different algorithms available through WEB servers (Supplementary Data Figure Fig. S1). MeDor, a recently developed meta-server for disorder prediction, yielded similar pictures (Supplementary Data Fig. S2).⁴⁹ The predictions showed significant differences in the number, position and length of the predicted disordered regions. The discrepancies between the results from different algorithms argue for the existence of different types ("flavors") of disorder, 50,51 but raise the problem of choosing the best algorithm and/or strategy for locating the boundaries of disordered regions in a given protein.43,5

Recently, we devised a simple voting procedure that combines multiple predictions of disordered regions and predicts the boundaries between structured and unstructured regions. We used it to delimit the nucleoprotein-RNA (N-RNA) binding C-terminal domain of VSV P protein.¹⁴ Here, we improve this approach by including secondary structure predictions and amino acid sequence conservation, and use it to predict the modular organization of SV, VSV and RV P. Our method requires no advised input from the experimentalist. It yields a consensus prediction that integrates the results from multiple disorder predictions, from multiple secondary structure predictions, and from amino acid sequence conservation among homologous proteins. For disorder predictions and secondary structure predictions, we devised scoring methods based on simple binary voting procedures. For each disorder prediction, residues were simply ranked in a binary manner, ordered or disordered, using the default threshold set for each algorithm in the WEB servers. The statistical weights yielded by some of these algorithms were not accounted for, since they were not available for each predicting method and were difficult to compare. For each predicting algorithm, a score of 0 was attributed to each residue predicted to be disordered, and a score of 1 was attributed to each one predicted to be in a structured region. A D-score was calculated by adding the values for each residue and dividing by the number of used algorithms. We arbitrarily defined two threshold levels at 0.75 and 0.50; residues with a D-score <0.50 were assigned as

disordered, while residues with a D-score >0.75 were assigned as structured. The structural state of the intermediate zone, with a D-score between 0.50 and 0.75, was considered as ambiguous. Residues with intermediate D-scores were found to occur at the boundaries between ordered and disordered regions, but they could also indicate protein regions that are stabilized mainly by tertiary or intermolecular interactions.

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

Finally, as mutation rates tend to be higher in disordered regions than in structured domains,⁵ structured domains are expected to appear as conserved regions in multiple sequence alignments of homologous proteins. Exceptions could be found in disordered regions interacting with partners and hence presenting some sequence conservation. To score the conservation rate in multiple sequence alignments we used the score provided by the AL2CO software,⁵⁵ and we considered regions with a normalized score (C-score) >0.5 as conserved regions. Finally, we defined a global score (G-score), as a non-weighted, linear combination of the scores for disorder prediction (D-score), secondary structure prediction (SS-score) and sequence conservation (C-score). Again, we arbitrarily defined a threshold of 0.5 in order to discriminate between disordered (G-score < 0.5) and structured regions (G-score > 0.5).

Amino acid sequence analysis of SV P

To assess the reliability of the above-described scoring methods, we analyzed the amino acid sequence of the SV P protein. Figure 1a shows the location of functional regions and Fig. 1f shows the location of secondary structure elements found in high-resolution structures of these two domains.^{2,3,46} The structure of the central oligomerization domain (aa 320-429) was solved by X-ray crystallography. It is a tetramer in which each monomer is composed of a long C-terminal α -helix (aa 364–429) that forms a tetrameric coiled coil plus an N-terminal helical bundle of three short α -helices (Fig. 1f).³ The structure of the C-terminal N-RNA interacting domain (aa 516-568), solved by NMR spectroscopy, consists of a three-helix bundle.² NMR also showed that the structured C-terminal domain (aa 516–568) is linked to a flexible domain (aa 474-504) by an even more flexible linker (aa 505–515).^{2,46} There is no information on the structure of residues 1-319 or 430-473.



Although disorder predictions for SV P displayed significant discrepancies among the various algorithms (Supplementary Data Figs. S1A and S2A), the normalized D-score shown in Fig. 1b revealed the existence of three main disordered regions (score ≤ 0.5) ranging from residue 1 to residue 321 (with the exception of the 28–49 region), from residue 430 to residue 459, and from residue 494 to residue 520, which are thus considered as the consensus disordered regions (Fig. 1c). The D-score clearly

Fig. 1. Predicting boundaries between structured and disordered regions in Sendai virus (SV) P. (a) Location of regions of known function.^{47,74,75} The conventions used in this figure are shown in the inset below. (b) Score for disorder (D-score) as a function of residue number calculated as described in the text. (c) Location of consensus disordered regions with a D-score <0.50. Regions with a score in the range 0.50-0.75 are shown as striped boxes. Numbers above the boxes indicate the boundaries of the consensus disordered regions. (d) Score for secondary structure prediction calculated from 15 different predictions (see the text). (e) Consensus location of predicted secondary structure as obtained from the scoring procedure described in the text (SS-score > 0.5 or < -0.5, respectively). (f) Location of secondary structure elements in the known 3D structures of two domains,^{2,3} with the PDB filenames above the drawing. (g) Score for sequence conservation calculated with the AL2CO software from a multiple sequence alignment obtained with T-Coffee. Conserved residues are defined as regions with a C-score >0.5. (h) Consensus location of conserved regions as defined by the C-score. (i) Global score calculated as a linear combination of the D-score (b) the absolute value of the SS-score (d) and the C-score for sequence conservation (g). Structured regions have a score >0.5 and disordered regions have a score <0.5. (j) Consensus location of predicted disordered and structured domains as defined by the G-score.

highlighted the location of the two structured domains, namely a central region spanning residues 322–429, which closely corresponds to the tetramerization domain (aa 320–429),³ and a C-terminal region encompassing residues 521–568, which corresponds to the structured C-terminal domain (aa 516–568).²

Predictions of secondary structures (α -helices and β -strands) supported these conclusions. Figure 1d shows the normalized SS-score for α -helix/ β -strand



Fig. 2. Predicting boundaries between structured and disordered regions in vesicular stomatitis virus (VSV) P. (a) Location of regions of known function.^{7,8,10,58–66} The conventions used in this figure are shown in the inset. (b–j) see the legend to Fig. 1.

prediction and Fig. 1e shows the consensus locations of these secondary structure elements. Within the known structured domains (Fig. 1f),^{2,3} secondary structures were predicted correctly, with the exception of the first α -helix in the C-terminal domain that was predicted as a β -strand. This C-terminal domain is rather flexible in isolation and is stabilized upon binding to N–RNA complexes.⁵⁶ In particular, residues in the first helix undergo proton exchange,

and the helix is in conformational exchange with the unfolded form.⁴⁸ The long α -helix predicted in the central domain was also predicted to form a coiled coil (aa 375–420) by the analysis of the HCA plot (see Fig. 2 in Ref. 57), as well as by various coiled coil predictors (see Materials and Methods), although with a level of confidence generally lower than that set by default in the WEB servers (data not shown). The coiled coil of SV P exhibits a different packing

pattern and is more hydrophilic than typical coiled coils,³ which may explain the failure of these algorithms to clearly predict this structure from the sequence.

Å multiple sequence alignment was obtained for P proteins from four members of the *Respirovirus* genus in the *Paramyxoviridae* family. The normalized C-score shown in Fig. 1g indicated five short conserved regions in the N-terminal half of the protein (aa 27–42, 97–120, 162–174, 186–205, 211–230), and three long and more conserved regions in the C-terminal half (aa 326–381, 399–481, 500–564) (Fig. 1h). The three conserved regions in the C-terminal half of P cover most of the tetramerization domain and the entire C-terminal domain involved in binding to N–RNA complexes. However, part of the tetramerization domain is not conserved (aa 382–398) and two regions (aa 430–459 and 500–520) predicted to be disordered were found to be conserved.

Finally, the global score (G-score) shows that the C-terminal half of the protein is predicted to be more structured than the N-terminal half (Fig. 1i and j). The two regions with the highest G-score in the C-terminal half (aa 323-433, 520-566) match the domains of known 3D structure. In addition, our analyses revealed the presence of three regions in the N-terminal half (aa 26–36, 105–109 and 208–226) and one region in the C-terminal half of SV P (aa 458–488) with a G-score >0.5, suggesting that they are structured. Two of these regions (aa 26–36 and 458–488) with intermediate D-score are predicted to contain secondary structure elements (|SS-score| >0.5) and are conserved among homologous proteins (C-score >0.5). One (aa 26-36) corresponds to the binding site of N in the N⁰–P complex, while the other (aa 458-488) partially overlaps the flexible region (aa 474-504) identified in the NMR experiments.² The two other N-terminal regions (aa 105–109 and 206–226) that have D-scores <0.5 but that clearly stand out from the background (Fig. 1b), are predicted to form secondary structure elements (|SS-score| > 0.5) and are conserved among homologous proteins (C-score > 0.5). Until now, there is no experimental evidence for the presence of structured domains in these regions. The sequences are short and could correspond to flexible regions in isolation, which may fold upon interaction with viral or cellular partners.

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

Amino acid sequence analysis of VSV P

A similar analysis was carried out with the amino acid sequence of VSV P (Indiana serotype) because numerous experimental data about the structure and function of different regions of this protein are available. Figure 2a shows the location of functional regions along the sequence of P,^{7,8,10,58–66} and Fig. 2f shows the location of secondary structure ele-

ments in the high-resolution structures of the central dimeric domain,⁷ and in the C-terminal N-RNA binding domain¹⁴ determined by X-ray crystallography and NMR spectroscopy, respectively. The D-score (Fig. 2b) revealed the presence of three disordered regions (aa 40-90, 107-131 and 186-196) and four structured regions (aa 1-19, 99-104, 137-170 and 211–265 with the exception of 232–241). Figure 2c shows the location of consensus disordered regions. Figure 2d shows the SS-score for α helix/ β -strand prediction, and Fig. 2e shows the consensus location of these secondary structure elements. In the central domain, the α -helix and two β -strands found experimentally in the 3D structure (Fig. 2f)⁶ were predicted correctly, but two other β strands, one on each side of the α -helix were not predicted. In the C-terminal domain, all secondary structure elements were also predicted correctly apart from a short α -helix (aa 229–234), which was predicted as a β -strand. A coiled coil was predicted by four algorithms as well as by analysis of the HCA plot with a consensus location between residues 8 and 36, but these predictions had a level of confidence below the threshold value used in the various WEB servers and are, therefore, not indicated in Fig. 2. Secondary structures were predicted in two regions of the N-terminal part of VSV P. Firstly, α -helices (aa 4–34) are predicted in the region that binds to the N in the N⁰–P complex.¹⁰ Secondly, one short α -helix (aa 81–87) and one short β -strand (aa 101–104) are predicted in the region that binds to L.⁵

A multiple alignment of the amino acid sequences of P proteins from 12 members of the Vesiculovirus genus revealed an overall 5% identity and 20% similarity. The normalized C-score for sequence conservation calculated from the multiple sequence alignment is shown in Fig. 2g, and boundaries for consensus conserved regions are shown in Fig. 2h. Two clusters of conserved regions (C-score >0.5) are found in the sequence of VSV P. A first cluster (aa 8-34, 44-63, 68-70, 83-102 and 120-150) encompassing the N-terminal and central regions is separated from a second cluster corresponding to the C-terminal N-RNA binding domain (aa 216-235 and 239-265) by a long non-conserved region (aa 151-215). In the first cluster, one conserved region (aa 8-34) matched with the N-terminal N⁰ binding region (aa 11–30) and another one (aa 120– 150) with the main part of the central dimerization domain (aa 107-177). A third region (aa 44-63) contains the phosphorylation sites (Ser60, Thr62, and Ser64) that control the transcriptional activation,^{61,62} and a fourth region (aa 83–102) contains the α -helix and β -strand predicted in the L-binding region.58

Finally, the G-score calculated for VSV P suggests the presence of four main structured regions (aa 1–37, 81–105, 121–169 and 217–265) separated by three main disordered regions (aa 38–80, 106–120 and 170–216) (Fig. 2i and j). The predicted N-terminal domain (aa 1–37) encompasses the binding site of N in the N⁰–P complex (aa 11–30).¹⁰ The second predicted structured regions (aa 81–105)

contains one of the L-binding sites. The α -helix predicted in this region (Fig. 2d and e) is located in a conserved region (Fig. 2g and h) that is otherwise predicted to be disordered (Fig. 2b and c), suggesting that these secondary structure elements could fold only upon binding to the L polymerase. In the central dimerization domain (aa 121–169), the predictions indicate the presence of three structured regions (aa 121–126, 132–154, and 166–169) that corresponds closely to a large part of the second β -strand (aa 119–126), the middle kinked α -helix (aa

131–157) and the main part of fourth β -strand (aa 165–170) found in the crystal structure.⁷ The D-score predicts that the two N-terminal β -strands are disordered (aa 107–131) and our global prediction score fails to identify the first and third β -strands. The last predicted regions (aa 217–265) corresponds to most of the C-terminal N–RNA binding domain (aa 195–265), with the exception of the initial β -strand and the short aa 235–239 region that are predicted to be disordered. The aa 235–239 region corresponds precisely to the place where an addi-



Fig. 3. Predicting boundaries between structured and disordered regions in rabies virus (RV) P. (a) Location of regions of known function,^{6,9,12,16,18–22,68} The conventions used in this figure are shown in the inset. (b–j) see the legend to Fig. 1.

tional α -helix is inserted in the C-terminal N–RNA binding domain of RV P.¹⁴

A similar pattern of disordered and structured regions was found with the amino acid sequence of VSV New Jersey (NJ) serotype (data not shown). P protein from NJ serotypes are longer (274 aa) than those of Indiana (IND) serotypes (265 aa), containing two small gaps at positions 45 and 71–73, and a large insertion of 14 residues between residues 195 and 196 (numbering according to the IND serotype). This latter insertion is predicted to be disordered, resulting in a longer C-terminal disordered region than in the IND serotype.

Amino acid sequence analysis of RV P

Figure 3 shows the results of the same analyses carried out with the sequence of RV P. Figure 3a shows the location of functional regions along the sequence of RV P,^{12,13,21,22,67,68} and Fig. 3f shows the location of secondary structure elements in the high-resolution structure of the C-terminal N–RNA binding domain determined by X-ray crystallography.⁶ The consensus disordered regions (Fig. 3c) determined from the D-score (Fig. 3b) revealed the presence in RV P of two disordered regions (aa 53–86 and 133–194) and three structured regions (aa 1–29, 90–130 and 200–297 with the possible exception of 241–252). The consensus locations of several of the secondary structures (Fig. 3e) derived from the SS-score (Fig. 3d) indicate the presence of α -helices and β -strands in the three predicted structured

regions as well as of two α -helices in the second disordered regions. In the N-terminal region, which binds to N in the N⁰–P complex and to L,^{9,10,68} the analysis predicted the presence of one β -strand and three α -helices. Within this region, a coiled coil was also predicted by four algorithms and by analysis of the HCA plot with a consensus location between residues 9 and 42, but these predictions had a level of confidence below the threshold value used in the various WEB servers and are, therefore, not indicated in Fig. 3. In the Cterminal region, the analysis predicted most of the secondary structures experimentally found in the three-dimensional structure of the C-terminal domain of RV P (Fig. 3f),⁶ with the exception of one β -strand and three α -helices.

A multiple alignment of the amino acid sequences of P proteins from 13 members of the *Lyssavirus* genus revealed an overall 25% identity and 51% similarity. The consensus conserved regions (Fig. 3h) derived from the normalized C-score (Fig. 3g) revealed the presence in the sequence of RV P of four conserved regions (C-score >0.5: aa 1–50, 101–126, 194–244 and 249–292) that are located within or correspond to the otherwise predicted structured regions (D-score >0.5, Fig. 3a).

Finally, the G-score calculated for RV P points out the presence of three structured domains (aa 1–55, 92–131 and 195–293) separated by two disordered regions (aa 56–91 and 132–194) (Fig. 3i and j). The predicted N-terminal domain (aa 1–55) encompasses the binding site of N in the N⁰–P complex



Fig. 4. Binding of RV P to P or N as assessed by a yeast two-hybrid approach. (a and b) Deletion mutants. (c) Point mutations in the central domain. L40 yeast cells were co-transformed with the indicated combinations of plasmids. The interaction between mutants of P fused to the DNA-binding domain (BD) of Lex A (rows in the figure) and P or N genes fused to GAL4 AD (columns in the figure) was assessed by the appearance of blue colonies in the presence of X-Gal (a and c) and by assaying the β -galactosidase activity of yeast grown in liquid medium (b). Quantitative results were obtained from three independent yeast co-transformants assayed with ONPG as substrate. Error bars indicate the standard deviation.

(aa 4–40),⁹ the predicted central domain (aa 92–131) could correspond to the dimerization domain, since it is located within the region necessary for oligomerization that was previously identified by deletion (aa 52–189),¹² whereas the predicted C-terminal structured domain (aa 195–293) corresponds closely to the known N–RNA binding domain (aa 186–296). A small region (aa 242–252) in the middle of the C-terminal domain is predicted disordered by the different scores. It corresponds to the end of helix α 2, the entire helix α 3, and the loop connecting the two helices, a part of the C-terminal domain of RV P that is absent from the homologous C-terminal domain of VSV P.

Experimental analysis of domain structure in RV P

Yeast two-hybrid experiments

Yeast two-hybrid experiments were performed on agar plates and in liquid cultures for quantitative analysis of the interactions (Fig. 4). RV P protein associated with itself and with N (Fig. 4a and b). Deletion mutants were used to determine which region of RV P is responsible for the dimerization of P_{t}^{11} and for the interaction with N (Fig. 4a and b). Truncation of the first 172 N-terminal residues $(P_{\Delta N172})$ or of the central region ranging from residue 91 to residue 131 ($P_{\Delta 91-131}$) reduced the association of P with itself considerably, whereas truncation of the 52 N-terminal residues (aa 1-52, $P_{\Delta N52}$) slightly reduced β -galactosidase activity and that truncation of the 125 C-terminal residues (aa 172–297, $P_{\Delta C125}$) led to a twofold increase of β galactosidase activity (Fig. 4a). These results clearly show that the predicted central domain (aa 92-131 according to the G-score) is critical for the dimerization of RV P as in VSV P.7 To characterize the dimerization interface, the central region was further dissected. The deletion mutant $P_{\Delta 91-105}$ still associated with P, whereas the mutants $P_{\Delta 106-113}$, $P_{\Delta 114-121}$, and $P_{\Delta 122-131}$ showed a significant reduction of the association with P, indicating that



residues 106–131 are required for dimerization (data not shown). In addition, single amino acid mutations were introduced into the central region of P and were tested by using the two-hybrid approach. Five hydrophobic residues were mutated into alanine (F114, W118, I125, Y128, and V129). Although none of these mutations affected the binding to N, they reduced the association with P significantly (Fig. 4c).

Previous data showed that the C-terminal domain of RV P (aa 173–297) binds to N–RNA complexes,^{6,59} while the first 40 residues associate with N in the N⁰–P complex.⁹ In the yeast two-hybrid experiment, truncation of the 125 C-terminal residues of RV P reduced but did not abolish the interactions with N (Fig. 4a and b). The persisting interaction might reflect the interaction of the N-terminal domain of P with N in N⁰–P complexes. Truncation of the 172 N-terminal residues or of the central domain (aa 91–131) had little effect on the interactions with N,^{67,69} whereas truncation of the 52 N-terminal residues increased β-galactosidase activity slightly (Fig. 4b).

Characterization of recombinant proteins

The full-length P protein and different fragments (P_{NTD}, P_{CED}, P_{CTD}, P_{Δ C120}, P_{Δ N52}, and P_{Δ 91-131}) (Fig. 5) were prepared as recombinant proteins with a C-terminal His₆ tag and purified in a two-step procedure using Ni²⁺ affinity and SEC. P_{Δ N52} and P_{Δ C120} were poorly expressed and could be obtained in only small amounts.

The far-UV CD spectrum of full-length P, P_{CED} , P_{CTD} , and $P_{\Delta 91-131}$ exhibited two dichroic bands near 208 nm and 222 nm, indicating the presence of α -helical structure, while the spectrum obtained with P_{NTD} showed one minor dichroic band near 222 nm and one below 200 nm, suggesting a low α -helical content (Fig. 6). The proper folding of the central domain, P_{CED} , was confirmed by 1D NMR spectroscopy (data not shown). Because β -strand and random coil contribute little to the CD signal at

Fig. 5. A schematic representation of the different constructs used for producing recombinant proteins. The predicted disordered regions are labeled DR1 and DR2. NTD, amino-terminal domain; CED, central domain; CTD, carboxy-terminal domain.


Fig. 6. Circular dichroism spectroscopy of deletion mutants of RV P. CD spectra of different constructs of RV P. Protein spectra were recorded at 20 °C in 20 mM Tri-HCl pH 7.5 and 150 mM NaF.

222 nm, the average amount of α -helix was estimated grossly from the molar ellipticity value at 222 nm.⁷⁰ The values for P_{CED} (-21,100 deg cm² dmol⁻¹, 28 aa in α -helix), P_{CTD} (-14,000 deg cm² dmol⁻¹, 45 aa in α -helix), P_{Δ91-131} (-9,200 deg cm² dmol⁻¹, 67 aa in α -helix), and full-length P (-11,100 deg cm² dmol⁻¹, 94 aa in α -helix) (Fig. 6) corresponded closely to the fraction of α -helical structure



predicted from the amino acid sequence of these proteins (P_{CED} , 19 aa; P_{CTD} , 43 aa; $P_{\Delta 91-131}$, 73; and P_{2} 92 aa) (Fig. 3). The value for P_{NTD} (–3600 deg cm² dmol⁻¹, 7 aa in α -helix) indicated a lower α -helical content than that suggested by the secondary structure predictions $(P_{NTD}, 30 \text{ aa})$ (Fig. 3). The number of residues in α -helical conformation in P_{CED} and in $P_{\Delta 91-131}$ adds up to that calculated for full-length P (28 aa+67 aa=95 aa), supporting the hypothesis that P_{CED} and P_{CTD} are autonomous folding units. However, the number of α -helical residues in P_{CTD} calculated from the CD spectrum (45 aa) was significantly lower than that derived from the crystal structure (67 aa), and summing α helical residues in P_{NTD}, P_{CED}, and P_{CTD} yields a number of residues (7 aa+28 aa+45 aa=80 aa) significantly lower than that measured for the fulllength protein (94 aa). This discrepancy could be accounted for by the presence of additional structure in a region of P other than P_{NTD} , P_{CED} or P_{CTD} . It is possible also that the predicted α -helices in the Nterminal domain fold only in the context of a structure that is larger than the tested fragment of aa 1-68.

Analysis by SEC using on-line detection by multiangle laser light-scattering (MALLS) and refractometry (RI) indicated that full-length P, P_{CED} , P_{CTD} , and $P_{\Delta 91-131}$ were monodisperse. P_{NTD} , P_{CTD} , and

Fig. 7. Molecular mass and size of deletion mutants of RV P. (a) Molecular mass measured by SEC-MALLS-RI. The line shows the SEC elution profile as monitored by refractometry. The crosses show the molecular mass calculated from light-scattering and refractometry data. (b) Plot of Stokes' radius measured by SEC against the molecular mass measured by MALLS-RI. The filled and open circles show data taken from Ref. 71 for standard globular proteins in native or unfolded forms, respectively.

 $P_{\Delta 91-131}$ were monomeric with weight-averaged molecular mass of 7.6 ± 0.5 kDa, 15.6 ± 0.2 kDa, and 32.5±0.3 kDa (Fig. 7a), respectively, in agreement with the expected mass calculated from the amino acid sequence (P_{NTD} , 8696 Da; P_{CTD} , 13,989 Da; and $P_{\Delta 91-131}$, 29,306 Da). P_{CED} and P form dimers with an average molecular mass of 14.7 ± 0.2 kDa (calc. 6501 Da) and 69±2 kDa (calc. 33,616 Da), respectively.¹¹ The hydrodynamic radii (Stokes' radius, $R_{\rm S}$) of full-length P, P_{NTD}, P_{CED}, P_{CTD}, P_{Δ C120}, and $P_{\Delta 91-131}$ were determined by calibrating the SEC column with standard proteins.⁷¹ On a plot of the hydrodynamic radius as a function of molecular mass (Fig. 7b), P_{CED} and P_{CTD} appeared as native globular proteins, $P_{\Delta 91-131}$, $P_{\Delta C120}$, and full-length P were intermediate between native and unfolded proteins, indicating that these proteins are nonglobular, and P_{NTD} appeared as an unfolded protein. A previous analysis by small-angle neutron scattering revealed that full-length P forms elon-gated dimeric molecules.¹¹ These experiments confirmed that the predicted central domain of P constitutes the dimerization domain and, together



Fig. 8. Binding of deletion mutants of P to nucleoprotein–RNA complexes (N–RNA) using native gel electrophoresis. Recombinant N–RNA rings containing 10 N protomers purified from insect cells were incubated with the different constructs of RV P and loaded onto a native gel. Lane 1, N₁₀-RNA circular complex containing a small amount of contamination with N₉-RNA complex. Lanes 2–7, Interaction of N₁₀-RNA with different deletion mutants of RV P as indicated above the lanes. Proteins were visualized by staining with Coomassie brilliant blue. with the CD experiments, that P_{CED} and P_{CTD} fold as independent domains.

Binding to recombinant N-RNA rings

The binding of full-length P and its deletion mutants to a recombinant circular N-RNA complex containing 10 nucleoprotein protomers $(N_{10}-RNA)^{72}$ was tested by native gel electrophoresis (Fig. 8). Lane 1 shows the migration of N₁₀-RNA, with a contaminating band of N₉–RNA. In the presence of excess full-length P, a new band appeared corresponding to the complex, while the band corresponding to N₁₀-RNA alone disappeared, and the excess P appeared at the bottom of the gel. When N₁₀-RNA was incubated with $P_{\Delta 91-131}$ or with P_{CTD} , similar shifts were observed for the complex, indicating that, like full-length P, both protein fragments bound to N_{10} -RNA (Fig. 8 lanes 4 and 6). Conversely, when N_{10} -RNA was incubated with $P_{\Delta C120}$ or P_{CED} , the migration of N₁₀–RNA was not affected, indicating no association (Fig. 8 lanes 5 and 7). These results demonstrate that binding of P to N–RNA complexes requires only the presence of the C-terminal domain and is independent of the presence of P_{CED}, and thus, of the dimerization of P.

Discussion

Assessing the scoring procedure

Predicting the location of disordered regions in a protein from its amino acid sequence is an important challenge in structural biology. Many proteins contain disordered regions, which prevent structure determination by X-ray crystallography, although these regions are often found to participate in biological activity. Here, we have devised a simple method for predicting the location of boundaries between structured and disordered regions, which combines different disorder predictions, but which also takes into account secondary structure predictions and sequence conservation. Such a voting procedure integrates information from a large number of algorithms and avoids delicate choices between prediction methods. Clearly, this approach is simplistic and accounts neither for the specificity of the different algorithms nor for the confidence level associated with each amino acid by some algorithms. It may underscore some peculiar types of disorder predicted by some algorithms. Our selection of algorithms is not exhaustive, and the threshold level at 0.5 for defining disorder was chosen arbitrarily. Among the different algorithms used for predicting disordered regions, GlobPlot, PreLink, and the three DisEMBL predictors yielded the most divergent results (Supplementary Data Fig. S1). However, removal of these predictions from the score calculation had little effect on the overall location of disordered regions and modified the location of boundaries by only a few residues (data

not shown), demonstrating the robustness of the method. Therefore, despite possible objections, this procedure has predicted a cutting out of the phosphoproteins from SV, VSV and RV into disordered regions and structured domains that is in agreement with available experimental data, and it has proved successful in predicting the boundaries of folded domains in VSV P^{14} and in RV P (this study). Predictions for SV and RV P seemed more clear-cut than those for VSV P, emphasizing that such predictions might be complicated by the existence of regions with poor intrinsic stability that fold upon binding to partners or conserve an enhanced flexibility within a structured domain. Indeed, predictions in the central region of VSV P indicated a complex pattern of short disordered and structured regions, and parts of the domain solved by X-ray crystallography were predicted to be disordered (aa 106–120, 126–131, 155–165).⁷ However, a part of this region could fold only upon binding to the L polymerase, and the predicted disordered regions that are folded in the crystal structure exhibit high temperature factor values, which could reflect their dynamical properties.⁷ In addition, our analyses have predicted the existence of previously unsuspected structured regions in SV and VSV P, which can now be studied.

A common modular organization within Rhabdoviridae and Paramyxoviridae phosphoproteins

All *MNV* viruses encode a phosphoprotein that has similar roles in the replication cycle, but no similarity is found between P sequences from different *MNV* families or even between P sequences within members of the *Rhabdoviridae*. In the *Paramyxoviridae*, the 3D structure of the C-terminal domain of P is conserved between Sendai and measles viruses,^{2,73} and the sequence of the oligomerization domain is well-conserved, suggesting a common homotetrameric coiled coil.^{3,74} In the *Rhabdoviridae*, we showed recently that, despite a lack of sequence similarity, the structure of the C-terminal domain is conserved between VSV and RV P.^{6,14} However, this structure is different from that of the C-terminal domain of P from Sendai and measles viruses, and the structure of the central dimerization domain of VSV P, an α/β domain,⁷ is different from the tetrameric coiled-coil domain found in Sendai virus P.³

Despite this lack of similarity, the available structural and functional information combined with a computational analysis of the amino acid sequences of VSV and RV P suggest a modular organization similar to that proposed recently for P proteins from members of the *Paramyxoviridae*, with three structured domains alternating with two disordered regions (Fig. 9).¹

An N-terminal domain with a potential to form α -helices

The 40 to 50 N-terminal amino acids of VSV and RV P constitute an N-terminal hydrophobic region with α -helical potential, although the far-UV circular dichroism spectrum obtained with the P_{NTD} fragment of RV P (aa 1–68) and the hydrodynamic radius revealed a disordered domain with a low αhelical content in solution. This discrepancy between predictions and experiments might reflect a disordered state in isolation with folding induced upon binding to a physiological partner. In this regard, it is noteworthy that in VSV and RV P, the Nterminal domain contains a binding site for N⁰ and that in RV P it also contains a binding site for the L protein (aa 1–19).⁶⁸ The N-terminal region was previously proposed to constitute the oligomerization domain of RV P, but the first 52 residues were found to be dispensable for dimerization.¹³ Here, we confirm that deletion of the first 52 residues of RV P has no effect on the ability of P to self-associate, thus ruling out a role of this region in P dimerization. In P from SV, a small N-terminal domain (aa 1-60), is also predicted to adopt an $\alpha\text{-helical conformation, and}$ to be involved in binding to $N^{0,\,1.75}$ supporting a conserved role for this region of the protein in both MNV families.

An N-terminal disordered region

The aa 38–80 region in VSV P (Fig. 2j) and the 56–91 region in RV P (Fig. 3j) are predicted to be disordered. In RV P, a large part of this region (from aa 61) is



Fig. 9. Modular organization for *Rhabdoviridae* and *Paramyxoviridae* phosphoproteins. Structured regions are represented by wide boxes. The model for *Paramyxoviridae* P protein is reproduced from Ref. 1.

dispensable for transcription,¹² highlighting another similarity with the *Paramyxoviridae*, in which the corresponding disordered region (aa 42–319, Fig. 1a) is of variable length and dispensable for both transcription and replication.⁷⁶ In VSV P, this region has an unusually high content of acidic residues (net charge of 15 negative charges) and contains phosphorylation sites (Ser60, Thr62, and Ser64, Indiana serotype) for cellular kinases that control transcription during the viral replication cycle, providing a rational for the conservation of this disordered region among *Vesiculoviruses* (Fig. 2g and h). ^{61–63,66–78} In SV and RV P, phosphorylation sites for cellular kinases are also found in the corresponding region but, so far, no role in the regulation of viral replication has been found.^{13,79–82}

A central oligomerization domain

Recently, we have shown that RV P and VSV P form dimers in solution,¹¹ and here we demonstrate that the dimerization is mediated by the central domain in both proteins. The dimerization domain of VSV P (aa 107-170) was identified by limited proteolysis and its structure was solved by X-ray crystallography.⁷ It is formed by a kinked α -helix flanked on each side by two-stranded β -sheets and assembles into dimers through the packing of the α -helices, thereby forming two domain-swapped β -sheets.⁷ In RV P, the predicted dimerization domain appears shorter than that in VSV P. It is expressed in *Escherichia coli* as a soluble, structured protein domain, which is an autonomous folding unit. Secondary structure predictions suggest a different structural organization in RV P with respect to VSV P, with three short α -helices and one β -strand in each monomer (Fig. 2). Experiments with a minireplicon system revealed that the central region of RV P (aa 61–175), and thus the dimerization of P, are dispensable for transcription and replication,¹² precluding a mechanism where P dimers cartwheel along the nucleocapsid.⁷⁶ Our predictions reveal the existence of a second structured region (aa 81–87, 100–105) in the center of VSV P, which could be involved in binding to the L polymerase.⁵⁸ In *Paramyxoviridae*, the central folded domain forms tetramers through coiled coils,³ and the tetramerization domain is required for both transcription and replication. $^{76}\,$

A C-terminal disordered region

A second disordered region is predicted in the C-terminal part of both VSV and RV P. In RV P, this region contains a binding site for the cytoplasmic dynein light chain (LC8).^{21,22} This region is considered as a flexible linker in P proteins from the *Paramyxoviridae*.^{1,2}

A C-terminal N-RNA binding domain

This domain binds to N–RNA complexes,⁶ (and this study) and to proteins from the interferon-

induced cellular response, including the STAT1 factor,^{18,20} and the promyelocytic leukemia protein¹⁹. The structure of the C-terminal domain of RV P (aa 186–297), which was solved by X-ray crystallography,⁶ is conserved among *Rhabdoviridae*, as shown recently by comparison with the structure of the C-terminal domain of VSV P.¹⁴ A short structure of the C-terminal domain that bound to the N–RNA complex was also found in Sendai and measles viruses.^{56,73}

Disordered regions and functions of P proteins

Long disordered regions are found in many functional proteins,^{29,83} and have been associated with specific functions.^{31,33} In particular, disordered regions have major roles in exposing sites of posttranslational modifications, such as phosphorylation,^{31,40} and in promoting multi-molecular assemblies and interactions with different partners.^{31,33} The presence of long disordered regions in P proteins of members of the Rhabdoviridae family and of other MNV families may thus have functional implications. The transcriptional activity of VSV P is regulated by phosphorylation, and residues with a major role in this regulation are located in the long Nterminal disordered region.^{61,66,84–86} Phosphorylation sites are found also in the corresponding disordered region of SV and RV P, although no role could be assigned to these modifications in these two viruses.13,79-82 The phosphoprotein also has a key role in the assembly of the replication/transcription complex by binding to the L-polymerase,^{58,68} and to the nucleoprotein in both the N⁰–P complex and the N–RNA complex,^{9,67,87,88} as well as to cellular part-ners in the case of RV P.^{18–22} Intrinsically disordered proteins can have a role in the assembly of multimolecular complexes by simply bridging partners together, by regulating the spacing between binding sites, by providing large binding surface or multiple contact points for a large protein, by adopting different conformations in response to different stimuli, or by increasing the speed of interaction through a "fly-casting" mechanism.^{31,89} Binding sites for viral (L polymerase for VSV P) or cellular partners (dynein light chain LC8 for RV P) are also found in regions predicted to contain secondary structure elements but otherwise predicted to be disordered, suggesting that these regions could fold upon binding to their partners. Folding upon binding is accompanied by a large decrease in conformational entropy and can provide the high level of specificity associated with a moderate binding strength.³³ The importance of disordered regions in the assembly of the replication complexes of MNV viruses was nicely highlighted in the structure of the complex formed between the Cterminal domain of measles virus P protein and a peptide corresponding to a disordered region of N protein.5,56,90

In conclusion, in the modular organization proposed for the P protein from *Rhabdoviridae* and *Para-myxoviridae* (Fig. 9),¹ the structured domains share

similar functions, although the structures of some of the domains are different. The presence of disordered regions, however, seems well conserved and could provide the flexibility required for proper function of the proteins; for example, allowing the interaction between different partners within the replication complex. The presence of disordered regions and the function of the structured domains appear to be conserved during evolution.

Materials and Methods

Amino acid sequence analysis

The location of disordered regions within SV, VSV and RV P was predicted by submitting the amino acid sequences to 16 different algorithms accessible through WEB servers. Foldindex was run with a window size of 51 residues.⁴⁴ With the different PONDR predictors XL3, VL-XT, XL1-XT, and VSL1, residues with a score >0.5 were considered disordered.^{91,92} The NORSp server was used with a window size for disorder set to 25 residues, a structure content cut-off of 12% and with the number of consecutive exposed residues set to 10.93 The IUPred predictor for long disordered regions was used, and residues with a score >0.5 were considered disordered.⁹⁴ The DISOPRED server was run with the rate threshold for predictions of false positive set at 2%.42 The DisEMBL (Loops/Coils, Hot-loops and Remark465),95 the RONN,⁹⁶ and the PreLink⁹⁷ servers were run using the default parameters. Program SEG was used for analyzing sequence complexity.⁹⁸ HCA plots were drawn with DRAWHCA,⁹⁹ and disordered regions of >20 residues were identified by visual inspection. A simple scoring procedure was used to define consensus disordered regions. For each prediction, residues predicted to be part of a disordered region were assigned a score of 0 while other residues were assigned a score of 1. A D-score for each amino acid was calculated by adding the values from all predictions and normalizing the sum. Consensus disordered regions were defined as regions with a normalized score ≤ 0.50 , and structured domains were defined as regions with a normalized score ≥ 0.75 .

Similar scoring procedures were used for analyzing predictions of secondary structures. Secondary structure predictions were obtained from 15 different algorithms freely available through WEB servers: DPM, ¹⁰⁰ DSC, ¹⁰¹ GOR IV, ¹⁰² HHNC, ¹⁰³ PHD, ¹⁰⁴ PREDATOR, ¹⁰⁵ SIMPA96, ¹⁰⁶ SOPM, ¹⁰⁰ SOPMA, ¹⁰⁰ NNPREDICT, ¹⁰⁷ PORTER, ¹⁰⁸ JPRED, ¹⁰⁹ PSIPRED, ¹¹⁰ SSPRO, ¹¹¹ and PROF. ¹¹² All algorithms were run using the default parameters. For each prediction, residues predicted to be in α -helix conformation were assigned a score of 1 while residues predicted to be in extended conformation were assigned a score of –1. An SS-score was calculated for each amino acid by adding values from each prediction and normalizing the sum. α -Helices were defined as regions with an SS-score >0.5 and β -strands as regions with a SS-score <-0.5. Predicted locations of coiled coils were obtained from four different WEB servers: COILS, ¹¹³ Paircoil2, ¹¹⁴ Multicoil, ¹¹⁵ and Marcoil. ¹¹⁶

Multiple sequence alignments were obtained for P sequences from four different *Respiroviruses*: Sendai virus (BAC79134), human parainfluenza viruses 1 (NP_604435) and 3 (BAA00031), bovine parainfluenza virus 1

(NP_037642), from 13 different Lyssaviruses: rabies viruses strains CVS-11 (P22363) and Pasteur (P06747), Australian bat lyssavirus (Q9QSP3), Australian bat lyssavirus (Q8JTH2), European bat lyssaviruses 1 (A4UHP9) and 2 (A4UHQ4), Duvenhage virus (O56774), Lagos bat virus (O56773), Mokola virus (P0C569), Aravan virus (Q6X1D7), Irkut virus (Q5VKP5), Khujand virus (Q6X1D3), West Caucasian bat virus (Q5VKP1), or from 12 different Vesiculoviruses: VSV strains Indiana Orsay (ACK77581), Indiana San Juan (P03520), Indiana 94GUB Central America (Q8B0H3), Indiana Glasgow (P04879), Indiana Mudd-Summers (P04880), Indiana 98COE North America (Q8B0I3), Indiana 85CLB South America (Q8B0H8), New Jersey Missouri subtype Hazelhurst (P04878), New Jersey Ögden subtype Concan (P04877), Piry virus (Q01769), Chandipura virus (strain I653514) (P16380), and Isfahan virus (Q5K2K6). The multiple alignments were performed with T-Coffee using the default parameters of the server.¹¹⁷ The C-score characterizing the sequence conservation was obtained for each residue with the AL2CO software,55 using the Henikoff-Henikoff frequency estimation method, a sum of pairs conservation measure, an averaging window of 20 residues and a gap fraction of 0.5. The C-score corresponding to the standard deviation from the mean was normalized into values ranging from 0 to 1 according to:

for
$$Z - \text{score} \ge 0$$
, $1 - 0.5 \cdot \exp(-Z \text{ score})$ (1)

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

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

$$G - score = \frac{D - Score + |SS - Score| + C - score}{3}$$
(3)

Plasmid construction

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

Plasmids for expressing full-length RV P (CVS strain) and several fragments (P_{NTD} , aa 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 constructed by cloning the gene or the corresponding fragment into pET22b(+) using PCR. For each construct, a His₆ tag and a two amino acid linker (Glu-

Leu) were introduced at the C-terminal extremity. All plasmids were verified by standard dideoxy sequencing.

Two-hybrid system

Yeast L40 cells were co-transformed with the plasmids encoding wild-type P or N fused to the AD of GAL4 and the P mutants fused to the DB of Lex A. The interaction was tested by measuring the β -galactosidase activity of histidine-positive clones on plates and in liquid assays. On plates, an X-Gal mixture containing 0.5% (w/v) agar, 0.1% (w/v) SDS, 6% (v/v) dimethylformamide and 0.04% (w/v) X-Gal (5-bromo-4-chloro-3-indolyl- $\beta\Delta$ -galactosidase) was overlaid on freshly transformed cells grown on Trp⁻Leu⁻ dishes and blue clones were detected after 60 min–18 h at 30 °C. For the liquid assay, cultures were grown overnight and assayed for β -galactosidase activity with *o*-nitrophenylgalactoside (ONPG) as substrate.¹¹⁸ β -Galactosidase activity was calculated as:

Activity =
$$(1000 \ A_{420})/(A_{600}TV)$$
 (4)

where A_{420} is the absorbance of the reaction mixture at 420 nm, A_{600} is the cell density of the culture (measured as absorbance at 600 nm), *T* is the reaction time (in min) and *V* is the volume (in ml) used for the assay.¹¹⁸

Production and purification of RV P and its different fragments

The plasmids for expressing full-length P or its fragments were transformed into E. coli strain BL21 (DE3) or *E.coli* BL21(DE3)-RIL for expressing P_{CED}. Cells were grown in LB medium containing 100 µg mL⁻¹ ampicillin at 37 °C until A₆₀₀ reached 0.6–0.8 a.u. At this point, IPTG was added to a final concentration of 0.5 mM and the temperature was shifted to 24 °C. After an incubation for 5 h, cells were harvested by centrifugation and suspended in buffer A (20 mM Tris-HCl pH 7.5, 150 mM NaCl) supplemented with anti-proteases (Complete[™] Protease Inhibitor Cocktail Tablets, Roche), DNAse I (Sigma) and 20 mM MgSO₄, and cells were disrupted by sonication. The extract was centrifuged at 20,000 g for 1 h at 4 °C. The supernatant was filtered $(0.45 \ \mu m \text{ pore size})$ and loaded onto a Ni²⁺ resin column, pre-equilibrated with buffer A. The resin was washed with three bed volumes of buffer A, then with three bed volumes of 20 mM Tris-HCl, 1.5 M NaCl, 10 mM imidazole at pH 7.5, and the protein was eluted using buffer A supplemented with 400 mM imidazole. The recombinant protein was loaded onto an SEC column (HiLoad 16/60 Superdex 75 prep grade, GE Healthcare) equilibrated with buffer A. For PCED, the SEC was performed in 20 mM Tris-HCl, 500 mM NaCl at pH 8.5. Separations were performed at a flow rate of 1.0 mL.min⁻¹. The purified protein was concentrated (Centricon, Amicon, 10,000 MWCO) and store at 4 °C. The purity of the protein samples were assessed by SDS-PAGE. The identity and integrity of the proteins were confirmed by electrospray mass spectrometry. The experiments were performed on a Quattro II mass spectrometer (Micromass, Altricham, UK) by continuously injecting the sample with a Type 22 pump (Harvard Apparatus) at a flow rate of $5 \ \mu L.min^{-1}$. Protein concentrations were measured by spectrophotometry using extinction coefficients calculated from the amino acid sequence.

Circular dichroism spectroscopy

A JASCO J-810 CD spectropolarimeter equipped with a temperature-controller (Peltier system) was used to record the far-UV CD spectra at 20 °C, which were accumulated at least five times. Full-length RV P or fragments thereof were diluted to final concentrations ranging from 10–40 μ M in 20 mM Tris–HCl pH 7.5 containing 150 mM NaF using a cuvet with a path-length of 1 mm.

After subtracting the blank signal, the CD signal (in mdeg) was converted to mean molar residue ellipticity (in deg.cm² dmol⁻¹) using the following equation:

$$[\theta] = \frac{0.1 \left(CD \, signal\right)}{l \, C \, N_{aa}} \tag{5}$$

where *l* is the pathlength of the cuvette (in centimeters), *C* is the protein concentration (in molarity) and N_{aa} is the number of amino acids. The ellipticity at 222 nm was converted to fraction helix values using:

$$f_{\alpha} = \frac{[\theta]}{[\theta]_{helix}} \tag{6}$$

where $[\theta]_{helix} = -35,000 \text{ deg cm}^2 \text{ dmol}^{-1}$ represents the mean molar residue ellipticity for 100% helix calculated from the following expression derived for peptides:¹²⁰

$$\left[\theta\right]_{helix} = (-44000 + 250T) \left(1 - \frac{n}{N_r}\right) \tag{7}$$

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

SEC-MALLS-RI

SEC was performed with an S75 Superdex column (GE Healthcare) equilibrated with buffer A. Separations were performed at 20 °C with a flow rate of 0.5 mL min⁻¹. Typicallly, 50 μ L of a protein solution at a a concentration of 5-10 mg.mL⁻¹ was injected. On-line MALLS detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting 690 nm. Data were analyzed and weight-averaged molar masses (*Mw*) were calculated using ASTRA software (Wyatt Technology Corp., Santa Barbara, CA) as described.¹¹

The excluded (V_0) and total volumes (V_t) were measured with blue dextran and thymidine, respectively. The partition coefficients (K_{av}) were calculated as:

$$K_{\rm av} = \frac{V_{\rm e} - V_0}{V_{\rm t} - V_0} \tag{8}$$

where *V*_e is the elution volume of the protein. The column was calibrated using proteins of known Stokes' radius (*R*_S) and molecular mass (*M*_w):³⁵ bovine serum albumin, *R*_S 3.4 nm, *M*_w 67.0 kDa; RNAse A, *R*_S1.9 nm, *M*_w = 13.7 kDa; ovalbumin, *R*_S 3.0 nm, *M*_w 43.5 kDa; β-lactoglobulin, *R*_S 2,7 nm, *M*_w 6.8 kDa; and chymotrypsinogen, *R*_S 2.3 nm, *M*_w 25 kDa.³⁵

Nucleocapsid binding assay

The interaction between circular N–RNA complexes and the various P proteins was investigated by native gel electrophoresis. Circular N–RNA complexes containing 10 nucleoprotein protomers (N₁₀–RNA) were purified as described.⁷² The N₁₀–RNA complex was incubated in the presence of full-length RV P, P_{Δ91-131}, P_{ΔN52}, P_{ΔC120}, P_{CED} or P_{CTD} at 20 °C for 10 min. A native gel of 4% (w/v) polyacrylamide and with a 19:1 (w/w) acrylamide/bisacrylamide ratio was pre-run for 30 min at 200 V in the cold room. Typically, 10 µl of protein mixture at about 1 mg/ml (where P fragments were in molar excess with respect to N–RNA) was loaded, and the gel was run for 3 h at 200 V. Complexes were identified by staining with Coomassie brilliant blue.

Acknowledgements

This work was supported by the interdisciplinary program Maladies Infectieuses Emergentes from the CNRS, and by a grant from the French ANR (ANR-07-001-01 (ANRAGE)) and Lyonbiopôle. We thank Bertrand Lefèbvre, Denis Daveloos and Bernard Brasmes (CRSSA, Grenoble) for mass spectrometry measurements, and Baptiste Fouquet and Hélène Raux for the amino acid substitutions in P and the test in the two-hybrid system. We thank Noël Tordo for extensive discussions and critical reading of the manuscript. 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 fellowships 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.2009.03.061

References

- 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.
- 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.
- Tarbouriech, N., Curran, J., Ruigrok, R. W. & Burmeister, W. P. (2000). Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nature Struct. Biol.* 7, 777–781.
- Llorente, M. T., Garcia-Barreno, B., Calero, M., Camafeita, E., Lopez, J. A., Longhi, S. *et al.* (2006). Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an

alpha-helical domain involved in oligomerization. *J. Gen. Virol.* **87**, 159–169.

- Bourhis, J. M., Canard, B. & Longhi, S. (2006). Structural disorder within the replicative complex of measles virus: functional implications. *Virology*, 344, 94–110.
- Mavrakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. (2004). Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J. Mol. Biol. 343, 819–831.
- Ding, H., Green, T. J., Lu, S. & Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* 80, 2808–2814.
- Ding, H., Green, T. J. & Luo, M. (2004). Crystallization and preliminary X-ray analysis of a proteinase-Kresistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). *Acta Crystallogr. D*, 60, 2087–2090.
- Mavrakis, M., Mehouas, S., Real, E., Iseni, F., Blondel, D., Tordo, N. & Ruigrok, R. W. (2006). Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology*, 349, 422–429.
- Chen, M., Ogino, T. & Banerjee, A. K. (2007). Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. J. Virol. 81, 13478–13485.
- 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.
- Jacob, Y., Real, E. & Tordo, N. (2001). Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.* 75, 9613–9622.
- Gigant, B., Iseni, F., Gaudin, Y., Knossow, M. & 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.
- Ribeiro, E. A., Jr, Favier, A., Gerard, F. C., Leyrat, C., Brutscher, B., Blondel, D. *et al.* (2008). Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J. Mol. Biol.* 382, 525–538.
- Chenik, M., Chebli, K. & Blondel, D. (1995). Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism. *J. Virol.* 69, 707–712.
- Pasdeloup, D., Poisson, N., Raux, H., Gaudin, Y., Ruigrok, R. W. & Blondel, D. (2005). Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1dependent nuclear export signal. *Virology*, 334, 284–293.
- Moseley, G. W., Filmer, R. P., DeJesus, M. A. & Jans, D. A. (2007). Nucleocytoplasmic distribution of rabies virus P-protein is regulated by phosphorylation adjacent to C-terminal nuclear import and export signals. *Biochemistry*, 46, 12053–12061.
- Vidy, A., Bougrini, J., Cheldi-Alix, M. & Blondel, D. (2007). The nulceocytoplasmic rabies P protein counteracts interferons signaling by inhibiting both nulear accumulation and DNA binding of STAT1. *J. Virol.* 81, 4255–4263.

- Blondel, D., Regad, T., Poisson, N., Pavie, B., Harper, F., Pandolfi, P. P. *et al.* (2002). Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies. *Oncogene*, 21, 7957–7970.
- 20. Vidy, A., Chelbi-Alix, M. & Blondel, D. (2005). Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways. *J. Virol.* **79**, 14411–14420.
- 21. Raux, H., Flamand, A. & Blondel, D. (2000). Interaction of the rabies virus P protein with the LC8 dynein light chain. *J. Virol.* **74**, 10212–10216.
- Jacob, Y., Badrane, H., Ceccaldi, P. E. & Tordo, N. (2000). Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein. *J. Virol.* 74, 10217–10222.
- 23. Perutz, M. F. (1963). X-ray analysis of hemoglobin. *Science*, **140**, 863–869.
- Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C. & Sarma, V. R. (1965). Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature*, 206, 757–761.
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). A threedimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181, 662–666.
- Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, 181, 223–230.
- Tanford, C. (1968). Protein denaturation. *Adv. Protein Chem.* 23, 121–282.
- Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331.
- Uversky, V. N., Gillespie, J. R. & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Struct. Funct. Genet.* 41, 415–427.
- Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur. J. Biochem.* 269, 2–12.
- Tompa, P. (2002). Intrinsically unstructured proteins. Trends Biochem. Sci. 27, 527–533.
- Spolar, R. S. & Record, M. T., Jr (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science*, 263, 777–784.
- Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12, 54–60.
- Sugase, K., Dyson, H. J. & Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 447, 920–921.
- Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J. et al. (2002). Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*, 415, 549–553.
- Uversky, V. N., Segel, D. J., Doniach, S. & Fink, A. L. (1998). Association-induced folding of globular proteins. *Proc. Natl Acad. Sci. USA*, **95**, 5480–5483.
- Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N. & Dunker, A. K. (2006). Intrinsic disorder in transcription factors. *Biochemistry*, 45, 6873–6888.
- Bourhis, J. M., Johansson, K., Receveur-Brechot, V., Oldfield, C. J., Dunker, K. A., Canard, B. & Longhi, S. (2004). The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Res.* 99, 157–167.

- Romero, P., Obradovic, Z. & Dunker, A. K. (2004). Natively disordered proteins: functions and predictions. *Appl. Bioinformatics*, 3, 105–113.
- Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* 32, 1037–1049.
- 41. Tompa, P. & Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **18**, 1169–1175.
- 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.
- Bourhis, J. M., Canard, B. & Longhi, S. (2007). Predicting protein disorder and induced folding: from theoretical principles to practical applications. *Curr. Protein Pept. Sci.* 8, 135–149.
 Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T.,
- 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.
- 45. Promponas, V. J., Enright, A. J., Tsoka, S., Kreil, D. P., Leroy, C., Hamodrakas, S. *et al.* (2000). CAST: an iterative algorithm for the complexity analysis of sequence tracts. Complexity analysis of sequence tracts. *Bioinformatics*, **16**, 915–922.
- Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. & Blackledge, M. (2005). A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl Acad. Sci. USA*, **102**, 17002–17007.
- 47. Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R. W. & Burmeister, W. P. (2000). On the domain structure and the polymerization state of the sendai virus P protein. *Virology*, **266**, 99–109.
- Houben, K., Blanchard, L., Blackledge, M. & Marion, D. (2007). Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase co-factor P. *Biophys. J.* 93, 2930–2940.
- Lieutaud, P., Canard, B. & Longhi, S. (2008). MeDor: a metaserver for predicting protein disorder. BMC Genomics, 9, S25.
- Vucetic, S., Brown, C. J., Dunker, A. K. & Obradovic, Z. (2003). Flavors of protein disorder. *Proteins: Struct. Funct. Genet.* 52, 573–584.
- 51. Ferron, F., Longhi, S., Canard, B. & Karlin, D. (2006). A practical overview of protein disorder prediction methods. *Proteins: Struct. Funct. Genet.* **65**, 1–14.
- 52. King, R. D., Ouali, M., Strong, A. T., Aly, A., Elmaghraby, A., Kantardzic, M. & Page, D. (2000). Is it better to combine predictions? *Protein Eng.* **13**, 15–19.
- 53. Schulz, G. E., Barry, C. D., Friedman, J., Chou, P. Y., Fasman, G. D., Finkelstein, A. V. *et al.* (1974). Comparison of predicted and experimentally determined secondary structure of adenyl kinase. *Nature*, **250**, 140–142.
- Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J. *et al.* (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.* 55, 104–110.
- 55. Pei, J. & Grishin, N. V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, **17**, 700–712.
- 56. Kingston, R. L., Hamel, D. J., Gay, L. S., Dahlquist, F. W. & Matthews, B. W. (2004). Structural basis for

the attachment of a paramyxoviral polymerase to its template. *Proc. Natl Acad. Sci. USA*, **101**, 8301–8306.

- Ferron, F., Rancurel, C., Longhi, S., Cambillau, C., Henrissat, B. & Canard, B. (2005). VaZyMolO: a tool to define and classify modularity in viral proteins. J. Gen. Virol. 86, 743–749.
- 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.
- 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.
- 60. 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.
- Takacs, A. M., Barik, S., Das, T. & Banerjee, A. K. (1992). Phosphorylation of specific serine residues within the acidic domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro. J. Virol. 66, 5842–5848.
- Barik, S. & 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. USA*, 89, 6570–6574.
- 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.
- 64. Gao, Y., Greenfield, N. J., Cleverley, D. Z. & Lenard, J. (1996). The transcriptional form of the phosphoprotein of vesicular stomatitis virus is a trimer: structure and stability. *Biochemistry*, **35**, 14569–14573.
- 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.
- 66. Gao, Y. & Lenard, J. (1995). Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II. *EMBO J.* 14, 1240–1247.
- Chenik, M., Chebli, K., Gaudin, Y. & 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, 2889–2896.
- Chenik, M., Schnell, M., Conzelmann, K. K. & Blondel, D. (1998). Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J. Virol.* **72**, 1925–1930.
- Fu, Z. F., Zheng, Y., Wunner, W. W., Koprowski, H. & Dietzschold, B. (1994). Both the N- and C-terminal domains of the nominal phosphoprotein of rabies virus are involved in binding to the nucleoprotein. *Virology*, 200, 590–597.
- Scholtz, J. M., Qian, H., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991). Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers*, **31**, 1463–1470.
- 71. Uversky, V. N. (1993). Use of fast protein sizeexclusion liquid chromatography to study the

unfolding of proteins which denature through the molten globule. *Biochemistry*, **32**, 13288–13298.

- Albertini, A. A., Clapier, C. R., Wernimont, A. K., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. (2006). Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings. J. Struct. Biol. 158, 129–133.
- 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.
- Curran, J., Boeck, R., Lin-Marq, N., Lupas, A. & Kolakofsky, D. (1995). Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. *Virology*, 214, 139–149.
- Curran, J., Marq, J. B. & Kolakofsky, D. (1995). An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. J. Virol. 69, 849–855.
- Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. (2004). Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, 318, 463–473.
- Das, S. C. & Pattnaik, A. K. (2004). Phosphorylation of vesicular stomatitis virus phosphoprotein P is indispensable for virus growth. *J. Virol.* 78, 6420–6430.
- Pattnaik, A. K., Hwang, L., Li, T., Englund, N., Mathur, M., Das, T. & Banerjee, A. K. (1997). Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication. *J. Virol.* **71**, 8167–8175.
- Gupta, A. K., Blondel, D., Choudhary, S. & 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.
- Byrappa, S., Hendricks, D. D., Pan, Y. B., Seyer, J. M. & Gupta, K. C. (1995). Intracellular phosphorylation of the Sendai virus P protein. *Virology*, 208, 408–413.
- Byrappa, S., Pan, Y. B. & Gupta, K. C. (1996). Sendai virus P protein is constitutively phosphorylated at serine249: high phosphorylation potential of the P protein. *Virology*, **216**, 228–234.
- Huntley, C. C., De, B. P. & Banerjee, A. K. (1997). Phosphorylation of Sendai virus phosphoprotein by cellular protein kinase C zeta. *J. Biol. Chem.* 272, 16578–16584.
- Radivojac, P., Obradovic, Z., Smith, D. K., Zhu, G., Vucetic, S., Brown, C. J. *et al.* (2004). Protein flexibility and intrinsic disorder. *Protein Sci.* 13, 71–80.
- Lenard, J. (1999). Host cell protein kinases in nonsegmented negative-strand virus (mononegavirales) infection. *Pharmacol. Ther.* 83, 39–48.
- Chattopadhyay, D., Raha, T. & Chattopadhyay, D. (1997). Single serine phosphorylation within the acidic domain of Chandipura virus P protein regulates the transcription in vitro. *Virology*, 239, 11–19.
- Raha, T., Samal, E., Majumdar, A., Basak, S., Chattopadhyay, D. & Chattopadhyay, D. J. (2000). N-terminal region of P protein of Chandipura virus is responsible for phosphorylation-mediated homodimerization. *Protein Eng.* 13, 437–444.

- Masters, P. S. & Banerjee, A. K. (1988). Resolution of multiple complexes of phosphoprotein NS with nucleocapsid protein N of vesicular stomatitis virus. J. Virol. 62, 2651–2657.
- 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.
- 89. Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **579**, 3346–3354.
- 90. Bourhis, J. M., Receveur-Brechot, V., Oglesbee, M., Zhang, X., Buccellato, M., Darbon, H. *et al.* (2005). The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the Cterminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Sci.* 14, 1975–1992.
- 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.
- 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.
- 93. Liu, J., Tan, H. & Rost, B. (2002). Loopy proteins appear conserved in evolution. J. Mol. Biol. 322, 53–64.
- Dosztanyi, Z., Csizmok, 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.
- 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.
- 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.
- Coeytaux, K. & Poupon, A. (2005). Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics*, **21**, 1891–1900.
- Wooton, J. (1994). Non-globular doamins in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* 18, 269–285.
- Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., 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.
- Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem. Sci.* 25, 147–150.
- King, R. D., Saqi, M., Sayle, R. & Sternberg, M. J. (1997). DSC: public domain protein secondary structure predication. *Comput. Appl. Biosci.* 13, 473–474.
- 102. Garnier, J., Gibrat, J. F. & Robson, B. (1996). GOR method for predicting protein secondary structure

from amino acid sequence. *Methods Enzymol.* **266**, 540–553.

- 103. Guermeur, Y., Geourjon, C., Gallinari, P. & Deleage, G. (1999). Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics*, **15**, 413–421.
- Rost, B., Sander, C. & Schneider, R. (1994). PHD an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* 10, 53–60.
- 105. Frishman, D. & Argos, P. (1997). Seventy-five percent accuracy in protein secondary structure prediction. *Proteins: Struct. Funct. Genet.* **27**, 329–335.
- Levin, J. M. (1997). Exploring the limits of nearest neighbour secondary structure prediction. *Protein Eng.* 10, 771–776.
- Kneller, D. G., Cohen, F. E. & Langridge, R. (1990). Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.* 214, 171–182.
- Pollastri, G. & McLysaght, A. (2005). Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics*, 21, 1719–1720.
- 109. Rost, B. & Liu, J. (2003). The PredictProtein server. Nucleic Acids Res. 31, 3300–3304.
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202.
- 111. Pollastri, G., Przybylski, D., Rost, B. & Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins: Struct. Funct. Genet.* **47**, 228–235.
- 112. Ouali, M. & King, R. D. (2000). Cascaded multiple classifiers for secondary structure prediction. *Protein Sci.* **9**, 1162–1176.
- 113. Lupas, A., Van Dyke, M. & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
- Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M. & Kim, P. S. (1995). Predicting coiled coils by use of pairwise residue correlations. *Proc. Natl Acad. Sci.* USA, 92, 8259–8263.
- 115. Wolf, E., Kim, P. S. & Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.* 6, 1179–1189.
- Delorenzi, M. & Speed, T. (2002). An HMM model for coiled-coil domains and a comparison with PSSMbased predictions. *Bioinformatics*, 18, 617–625.
- Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217.
- Guarente, L. (1993). Strategies for the identification of interacting proteins. *Proc. Natl Acad. Sci. USA*, 90, 1639–1641.
- Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, 6, 1948–1954.
- 120. Luo, P. & Baldwin, R. (1997). Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry*, **36**, 8413–8421.

Chapitre 2. Caractérisation structurale et fonctionnelle de la phosphoprotéine du virus 66 de la rage

2.2 Conclusion

L'organisation modulaire des phosphoprotéines est conservée entre les genres Lyssavirus et Vesiculovirus, mais également entre les familles des *Rhabdoviridae* et des *Paramyxoviridae*.



FIGURE 2.2 – Alignements de séquences de phosphoprotéines de Lyssavirus et Vesiculovirus. Un alignement est présenté entre la souche Mudd Summers (VSV, Vesiculovirus) et souche CVS-11 (RAV, Lyssavirus). L'alignement a été réalisé avec ClustalW. Les résidus en couleur blanche dans les cases rouges sont identiques et les résidus encadrés dans des rectangles bleus sont similaires.

Malgré une conservation faible de séquence aux alentours de 15 %, (voir figure 2.2) les propriétés biochimiques des domaines de la phosphoprotéine au sein des différentes familles de l'ordre des *Mononegavirales* sont conservées.

Ces protéines possèdent une région N-terminale désordonnée fixant N° (Mavrakis et al. 2006, Chen et al. 2007), un domaine central d'oligomérisation - dimérique chez les *Rhabdoviridae* (Ding et al. 2006) et tétramérique chez les *Paramyxoviridae* (Tarbouriech, Curran, Ruigrok & Burmeister 2000, Llorente et al. 2006) et un domaine C-terminal responsable de l'interaction avec la matrice N-ARN viral (Jacob et al. 2001, Mavrakis et al. 2004, Ribeiro et al. 2008, Blanchard et al. 2004).

Concernant le N-terminal de P des Rhabdovirus, un travail détaillé montre que la partie N-terminale de la P contient un MoRE (Molecular Recognition Element) (Fuxreiter et al. 2004) désordonnée en solution avec une faible propension à former des hélices α (Leyrat et al. 2010), (Leyrat, Jensen, Ribeiro, Gérard, Ruigrok, Blackledge & Jamin 2011) mais qui se structure lors de sa fixation à son partenaire N° (Leyrat, Yabukarski, Tarbouriech, Ribeiro, Jensen, Blackledge, Ruigrok & Jamin 2011). Il en est de même pour la région N-terminale de la phosphoprotéine de Hendravirus (*Paramyxoviridae*) (Habchi & Longhi 2011), mais également pour les autres virus appartenant à la famille des *Paramyxoviridae*. En conclusion, l'organisation structurale de la protéine et la fonction des domaines de phosphoprotéine sont bien conservées au sein de différentes familles de virus.

Afin d'approfondir la comparaison entre la phosphoprotéine du VSV et celle de la rage, j'ai entrepris de déterminer la structure du domaine central de la phosphoprotéine du virus de la rage. Sur la base

2.3. Approches envisagées pour la résolution de la structure du domaine central de la phosphoprotéine du virus de la rage

de prédictions de désordre mentionnées dans l'introduction de ce chapitre 2.1 page 44, j'ai cloné et exprimé la partie prédite nécessaire à l'oligomérisation (95 à 131). La structure et la comparaison avec d'autres domaines centraux de phosphoprotéines de Mononegavirales sont présentées dans l'article intitulé « Structure of the dimerization domain of the rabies virus phosphoprotein ».

2.3Approches envisagées pour la résolution de la structure du domaine central de la phosphoprotéine du virus de la rage

L'étude bio-informatique précédente de la séquence en acides aminés de la P du RAV avait permis de localiser le domaine central (Gerard et al. 2009). Mon travail a consisté dans le clonage, la production et la purification de ce domaine central de la phosphoprotéine du virus de la rage.

Le domaine composé des acides aminés 95 à 131 se comporte comme une protéine globulaire en solution avec une masse expérimentale de 14 kDa indiquant la présence d'un dimère en solution. La purification du domaine est présentée sur le figure 2.3. Elle est effectuée en deux étapes : une chromatographie d'affinité (Ni-NTA, Qiagen), suivie d'une chromatographie d'exclusion de taille. Le rayon hydrodynamique est comparable à une protéine globulaire comparable à la taille de la RNAseA (Uversky 1993). Nous avons ensuite réussi à cristalliser la protéine et a déterminer la structure. La publication intitulée dans le Journal of Virology présente ces résultats et compare la structure du domaine de la rage au domaine central d'homo-oligomérisation de la P de VSV et du virus de Sendai.



FIGURE 2.3 - Caractérisation du domaine de dimérisation de la phosphoprotéine du virus de la rage.

A : Représente le domaine central de la phosphoprotéine du virus de la rage sur une colonne d'exclusion de taille S 75. Une fraction minoritaire d'agrégats de haut poids moléculaire sortent dans le volume mort de la colonne. Le pic majoritaire, à 80 ml d'élution est analysé en B. La colonne d'exclusion de taille S 75 a un volume total de 120 ml.

B : Représente l'analyse de la purification du domaine central de la P de la rage. Chaque puits correspond à une fraction du pic d'élution en A. La protéine est pure en solution avec une masse de \sim 7 kDa sur un gel 15 % SDS-PAGE.

C: Représente l'analyse en SEC-MALLS-RI de ce même domaine central. Sur le graphe est représenté uniquement le pic d'élution de la colonne S75. Sur l'axe des "y" est représenté l'excès d'indice de réfraction. Le barre rouge représente le calcul de la masse moléculaire de la protéine à chaque volume d'élution. La concentration protéique est calculée précisément avec un réfractomètre et le MALLS utilise la concentration pour calculer le poids moléculaire de la protéine en question.

Chapitre 2. Caractérisation structurale et fonctionnelle de la phosphoprotéine du virus 68 de la rage

Afin de résoudre la structure du domaine central de la phosphoprotéine du virus de la rage, j'ai envisagé simultanément deux approches.

- RMN : En raison de la petite taille du domaine central (14 kDa) il était raisonnable de résoudre la structure par Résonance Magnétique Nucléaire (RMN). A cette fin, j'ai préparé des échantillons doublement marqués ¹⁵N, ¹³C. Un premier spectre HSQC ¹⁵N-H a montré une dispersion des pics montrant que la protéine est repliée en solution.



FIGURE 2.4 – Spectre RMN HSQC ¹⁵N-H du domaine de dimérisation.

La fréquence de relaxation F1 allant de 108 à 130 ppm correspond à l'azote $^{15}\mathrm{N}$. La fréquence F2 allant de 6 à 10 ppm correspond à l'hydrogène. Chaque pic correspond à un acide aminé.

La qualité du spectre permettait d'espérer la détermination de la structure mais des résultats comme des mesures de Diffusion de Rayons X aux Petits Angles (SAXS) ont démontré la présence d'interactions intermoléculaires. Ceci a confirmé la présence d'espèces de taille supérieure à celle attendue pour un dimère de P_{CED} .

- Cristallisation : Une des difficultés de l'étude d'une macromolécule par cristallographie réside dans l'obtention de cristaux. La connaissance des propriétés phisico-chimiques de la macromolécule n'est pas suffisante pour établir une méthode rationnelle permettant d'obtenir des cristaux de protéine. La cristallogénèse reste pour une grande part très empirique. La platforme de cristallisation robotisée de l'EMBL-Grenoble a largement facilité cette étape, car elle a permis de cribler rapidement un grand nombre de conditions de cristallisation avec une faible quantité de protéine. Les cristaux obtenus avec le robot ont été reproduits manuellement dans des gouttes de plus grande taille ce qui a permis d'obtenir des cristaux utilisables pour des mesures de diffractions (voir figure 2.5 page 69).

Les cristaux ont été préparés en les trempant dans une solution de glycérol 20 % puis en les congelant dans l'azote liquide (100° K). Les premiers cristaux du robot ont été testés sur la ligne FIP de l'ESRF à Grenoble et ont donné des taches de diffractions à 2 Å ce qui a permis de montrer que le cristal contient une solution de protéine possédant un empilement régulier.

2.3. Approches envisagées pour la résolution de la structure du domaine central de la phosphoprotéine du virus de la rage 69



FIGURE 2.5 – Cristaux du domaine central de RAV P obtenus en nanogouttes et manuellement.

A 50 mM Na-Cacodylate pH 6.5, 2 M ammonium sulfate, 10 mM MgSO₄ $(0.1\mu l + 0.1\mu l)$ (Plate n° 3 Qiagen screen; well :C 09). B 50 mM Na-Cacodylate pH 6.5, 2 M ammonium sulfate, 10 mM MgSO₄ $(1\mu l + 1\mu l)$.

2.3.1

Les jeux de données utilisés pour la résolution de la structure ont été collectés sur les lignes ID14eh4 pour les cristaux natifs et ceux trempés avec les métaux lourds et la ligne BM14 pour les cristaux qui ont servi à la collecte des données anomales.

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

Le premier jeu de données (Natif1) à haute résolution collecté sur les cristaux natifs montrait des taches de diffraction à une résolution maximale de 1.5 Å. Après une collecte complète et l'analyse des statistiques, la résolution globale est restée proche de 1.5 Å (cf Tableau 2.1 page 70). Le groupe d'espace a été déterminé sur la ligne de lumière avec le programme MOSFLM (Leslie 2006) comme étant tétragonal. Après construction du modèle dans la carte de densité électronique et affinement du modèle avec ARP-wARP (Perrakis et al. 1999), il s'est avéré qu'un axe hélicoïdal est présent dans une direction. Une molécule de dimère est présente par unité asymétrique.

L'indexation et l'intégration des données ont été réalisées avec XDS (Kabsch 1988) et MOSFLM (Leslie 2006). Le programme XDS a été retenu pour sa rapidité d'indexation et d'intégration. L'indexation se fait image par image tout en affinant la distance du cristal au détecteur, la position du centre du détecteur, l'orientation du cristal et les paramètres de maille. Lors de l'intégration, l'intensité des taches de diffraction est mesurée en appliquant un « masque »autour de la tache. L'intérieur du masque délimite la surface à intégrer et l'extérieur permet d'estimer le bruit de fond à retrancher aux intensités intégrées.

Le programme Xscale (Kabsch 1988) a été utilisé pour la mise à l'échelle et la réduction des données. Une

Chapitre 2. Caractérisation structurale et fonctionnelle de la phosphoprotéine du virus 70 de la rage

mise à l'échelle des données est nécessaire car l'exposition aux rayons X et le pouvoir de diffraction du cristal sont différents sur chaque image. Lors de la réduction des données, le jeu de données est moyenné suivant les réflexions équivalentes collectées plusieurs fois ou reliées entre elles par des opérateurs de symétries. La qualité de jeu de données est estimée grâce au valeur de R_{sym} entre les réflexions équivalentes reliées par symétrie et au rapport $I/\sigma I$ (signal/bruit).

<u>Un deuxième jeu de données</u> a servi pour la collecte des données anomales à deux énergies (7 keV et 14 keV). C'est à partir de ces données de diffraction que j'ai réussi à résoudre les phases et à construire un premier modèle dans une carte de densité électronique à 2.5-3.0 Å de résolution. L'énergie à 7 keV correspond au pic d'absorption du soufre. En raison du faible signal anomal une multiplicité importante a été effectuée (voir Tableau 2.1). Afin d'encore améliorer la redondance, une collecte supplémentaire a été effectuée sur le même cristal à une énergie de 14 keV, loin du pic d'absorption du soufre.

Dans les collectes de données réalisées avec les cristaux trempés dans les métaux lourds l'intensité du signal anomal était insuffisante pour permettre de déterminer les premières phases. (voir Tableau 2.1 colonne Eu et Gd, ligne-SigAno). Les statistiques relatives à la collecte des jeux de données sont répertoriées dans le Tableau 2.1 :

General Data	Sulphur MAD-Anomalous data		Derivat	Native data	
	Set1	Set2	Eu	Gd	
X-Ray source	BM14	BM14	ID14-4	ID14-4	ID14-4
images	1010	960	90	90	100
Oscillation degree(°)	1	1	1	1	1
wavelength (Å)	1.771	0.885	0.976	0.976	0.976
energy (keV)	7.00	14.00	12.70	12.70	12.70
spacegroup	$I4_{1}22$	$I4_{1}22$	I4122	I4122	$I4_{1}22$
Cell dimensions					
a(Å)	43.15	43.23	43.35	43.07	43.29
b(Å)	43.15	43.23	43.35	43.07	43.29
c(Å)	190.90	191.44	192.13	188.97	192.02
$\alpha = \beta = \gamma$	90	90	90	90	90
Resolution range (Å) ^a	25-2.11(2.19-2.11)	25-2.00 (2.07-2.0)	15.0-2.04 (2.14-2.04)	15.0-1.96 (2.04-1.96)	25.0-1.5 (1.56-1.49)
completeness $(\%)^{a}$	97.6 (86.7)	97.9 (96.3)	80.0 (94.9)	94.1 (95.5)	93.6 (90.6)
R _{sym} (I)(%) ^{a,b}	5.7(31.7)	5.5(30.8)	6.4(35.1)	2.4(102.3)	5.1(42.5)
$I/\sigma I^a$	66.75 (11.53)	85.80 (21.81)	12.06 (2.53)	32.17 (1.06)	20.36 (4.75)
Total Reflections ^a	428,072 (24,834)	478,480 (45,631)	42,189 (5,441)	48,368 (4,818)	133,743 (13,727)
Unique reflections ^a	9,704 (904)	6,451 (595)	21,992 (2,814)	25,661 (2,668)	17,851 (1,755)
Multiplicity ^a	44,11 (27,47)	74,17 (76,69)	1,92(1,93)	1,88 (1,80)	7,49 (7,82)
SigAno ^{a,c}	4,8 (0,853)	1,006 (0,722)	1,078 (0,922)	$0,726\ (0,706)$	0,493 (0,715)

TABLE 2.1 – Statistiques relatives aux collectes de données.

^a Les valeurs entre parenthèses représentent les statistiques dans la dernière tranche de résolution. ^b $R_{sym}(I) = [\sum_{hkl} \sum i | \langle I_{hkl} \rangle - I_{hkl,i} |] / [\sum_{hkl} \sum i |I_{hkl}|]$, où i est le nombre de réflections hkl. ^c SigAno = F(+)-F(-)/ σ représente le signal anomal pour chaque tranche de résolution.

Une seconde étape cruciale est la détermination des phases. J'ai utilisé différentes stratégies de phasage pour résoudre la structure du domaine central de la phosphoprotéine du virus de la rage.

2.4 Stratégies utilisées pour obtention des phases

- A : Remplacement moléculaire

Dans un premier temps, disposant de la structure du domaine central de VSV (Ding et al. 2006), nous avons essayé sans succès de déterminer les phases des données de RAV par remplacement moléculaire. Les différents programmes de remplacement moléculaire utilisés, comme MOLREP (Rossmann 1990, Vagin & Teplyakov 2000), PHASER (McCoy et al. 2007), et MrBUMP (Keegan & Winn 2008), n'ont donné aucune solution de remplacement moléculaire. La raison probable de l'échec serait que le modèle initial et le modèle recherché ne sont pas identiques et possèdent de grosses différences de structure et d'identité de séquence.

Dans le laboratoire, le Dr. Cédric LEYRAT avait créé des modèles du domaine central de RAVP par modélisation moléculaire. La stratégie de modélisation a été conduite en deux étapes, en modélisant d'abord la structure du monomère et en reconstituant ensuite le dimère. La structure du monomère de RAV P_{CED} a été modélisée à partir de la séquence en acides aminés en utilisant le serveur LOMETS qui réalise une modélisation par homologie en utilisant une méta-analyse par enfilage (« threading ») (Wu & Zhang 2007). LOMETS réalise un consensus de 9 programmes différents d'enfilage. Un deuxième programme I-TASSER a été utilisé qui permet de combiner les premiers modèles obtenus par sélection d'enfilage avec une seconde étape de modélisation *ab initio*. Après une deuxième étape d'arrimage moléculaire (« docking ») et de raffinement les modèles de dimères ont été construits. Les meilleurs modèles parallèles et antiparallèles ont été raffinés en boucle par dynamique moléculaire en utilisant le programme GROMACS. En fin d'affinement, après une minimisation d'énergie, les 10 meilleurs modèles comprenant : 7 modèles avec une orientation parallèle et 3 avec une orientation anti-parallèle. Aucun de ces modèles n'a permis d'obtenir le phasage des données expérimentales.

– B : Trempage dans des sels de métaux lourds

Une seconde tentative a été envisagée par phasage avec des atomes lourds afin d'introduire le métal dans le cristal. nous avons fait plusieurs essais de trempage en utilisant le Gadolinium et l'Europium (Petsko 1985). Les cristaux natifs ont été incubés pendant 2 heures ou 5 heures à une concentration finale de 1 mM. Au bout de 5 heures de trempage le Gadolinium détruit les cristaux. Après le temps de trempage le cristal a été rincé dans une solution ne contenant pas de métal lourd de façon à enlever l'excès. Immédiatement après, le cristal est cryoprotégé dans une solution de 20 % glycérol et congelé dans l'azote liquide. Les seuls cristaux exploitables ont été obtenus avec un temps de trempage de 2 heures. Les statistiques de collecte des données après traitement par XDS, avec les deux métaux lourds, sont présentés dans le Tableau 2.1, page 70. Les résultats étant de qualité médiocre et sans signal anomal (SigAno) exploitable, cette méthode avait été abandonnée.

– C : Diffusion anomale du soufre - Sulfur-SAD/MAD

Finalement j'ai pu résoudre la structure avec la méthode Sulfur-MAD qui réside des propriétés intrinsèques d'absorption du soufre. La ligne de lumière BM14 (ESRF, Grenoble) est une ligne particulièrement utile pour un phasage expérimental avec des atomes intrinsèques de la protéine comme le soufre (S) des résidus méthionines. Ce type de phasage exige un faisceau de rayons

Chapitre 2. Caractérisation structurale et fonctionnelle de la phosphoprotéine du virus 72 de la rage

X peu fluctuant dans le temps. Cette technique est devenue populaire de par sa rapidité et sa simplicité. Une autre raison de sa popularité repose sur sa capacité à déterminer les phases des données cristallographiques sans marquage préalable de la protéine soit avec des atomes lourds soit avec des sélénométhionines. Un inconvénient majeur de cette technique est qu'elle exige une multiplicité importante à cause du faible nombre d'électron de soufre présent dans l'échantillon (voir tableau 2.1, page 70).

Pour la résolution des phases, l'expression d'une protéine sélénométhionylée en système d'expression bactérien n'a pas été envisagée à cause de la croissance lente des cristaux. Egalement des acides aminés sélénométhionines (Se-Met) sont généralement connus pour perturber l'empilement cristallin ce qui pourrait conduire à une mauvaise diffraction, ou bien même à empêcher la formation de cristaux. Pour éviter ces problèmes nous avons décidé d'utiliser la méthode de Sulfur-SAD/MAD. Dans le cas du soufre, l'énergie du faisceau de rayons X a été réglée au niveau du pic d'absorption du soufre (7 KeV). C'est une technique demandant des réglages minutieux au niveau de la ligne de lumière BM14. Afin d'avoir des bonnes statistiques et augmenter les chances d'enregistrer un signal anomal nécessaire au phasage nous avons collecté un jeu de données à une longueur d'onde éloignée du pic d'absorption du soufre (à 14 KeV). Les statistiques sont répertoriées dans le Tableau 2.1, page 70.

Après traitement des données avec XDS et Xscale, j'ai procédé au phasage. Plusieurs logiciels ont été testés car ils ont tous des algorithmes de calcul différents, mais uniquement le logiciel Auto-Rickshaw a donné des résultats exploitables (Panjikar et al. 2005). Les autres programmes comme SHARP, SHELXD n'ont pas permis de trouver les phases des données expérimentales.

Chaque monomère possède deux méthionines et il y a deux monomères, c'est-à-dire un dimère par unité asymétrique. SHELXD, dans la suite de Auto-Rickshaw, a trouvé 3 méthionines sur un total de quatre. Ceci pourrait s'expliquer par le fait qu'une méthionine se trouve dans une région flexible du cristal et en conséquence ne donne aucun signal anomal. Les meilleurs statistiques de SHELXD donnent les coefficients de corrélation suivant ($CC_{all}=33.91$ et $CC_{weak}=23.54$). Dans la littérature il est décrit qu'une solution est considérée correcte si la valeur du CC_{all} est supérieure à 30 et CC_{weak} est supérieure à 20. Suite au phasage, le programme density modification (DM) affine la carte de densité électronique en affinant la position des atomes lourds trouvés. Le programme DM calcule un paramètre appelé figure of merite (FOM) qui est une comparaison de la carte expérimentale et la carte calculée. Plus la valeur est grande (0 < FOM > 1), plus il y a des chances que la solution soit correcte. La valeur résultante étant de 0.670, la solution est considérée comme correcte.

2.4.1 Construction du modèle

La carte de densité électronique obtenue avec Auto-Rickshaw permettait de distinguer 2 hélices collées l'une à l'autre. On pouvait également observer des trous dans la densité entre les unités

asymétriques qui pouvaient correspondre à des molécules d'eau. Le programme ARP-wARP integré dans la suite de Auto-RickShaw a permis de construire 85 acides aminés en alanine sur un total de 108. La densité électronique avec le modèle construit est représentée sur la figure 2.6 A et 2.6 D page 74.

Le modèle, en alanine a ensuite été affiné par itération avec le programme ARP-wARP dans la suite CCP4. La nouvelle carte de densité électronique est présentée dans la figure 2.6 B et 2.6 E page 74. Suite à cet affinement certaines densités électroniques correspondant aux gros acides aminés tels que les phénylalanines, les tryptophanes et certaines chaînes latérales de lysines et d'isoleucines étaient clairement visibles. Néanmoins, certains acides aminés ne pouvaient pas être correctement positionnés dans la densité. Chaque boucle, reliant les deux hélices alpha des monomères correspondants, était manquante. Au fur et à mesure de la construction et de l'affinement du modèle avec ARP-wARP dans la suite CCP4, les paramètres d'affinement R_{work} et R_{free} ont nettement diminué. Une construction manuelle a ensuite été réalisée avec le programme COOT (Emsley & Cowtan 2004). Une difficulté majeure a été rencontrée lors de la construction de la chaîne principale et des chaînes latérales d'un des deux monomères (voir figure 2.6 C et 2.6 F page 74). Ce dernier se trouvant à proximité du solvant dans la maille cristalline est plus flexible que son voisin, ce qui conduit à une construction plus difficile de la chaîne. Afin de mieux visualiser les densités électroniques manquantes un modèle artificiel a été construit avec COOT en pivotant le monomère le mieux construit, de 180° et en le plaçant dans une densité électronique imaginaire peu visible. Ensuite un affinement avec ARP-wARP a fait apparaître la densité électronique de ce second monomère. Ceci m'a permis d'affiner au maximum le modèle cristallographique.

Le modèle final affiné par ARP-wARP représente une structure dimérique formée de deux hélices alpha pour chaque monomère (hélice n° 1 région 91-110, et hélice n° 2 région 113-133). Dans la structure cristallographique les monomères sont inclinés l'un par rapport à l'autre selon un angle de 46° proche de 50° généralement rencontrés lors de l'assemblage hélice-hélice dans des protéines globulaires mais très différent de l'angle de 20° généralement observés dans les faisceaux de 4 hélices (« 4 helix bundles ») (Tarbouriech, Curran, Ruigrok & Burmeister 2000). Cette orientation des hélices pourrait expliquer l'échec du remplacement moléculaire avec le modèle construit par modélisation moléculaire.



Chapitre 2. Caractérisation structurale et fonctionnelle de la phosphoprotéine du virus 74 de la rage

Article II : Strucure of the Dimerization Domain of the Rabies Virus Phosphoprotein

<u>Auteurs</u> : **Ivan Ivanov**, Thibaut Crépin, Marc Jamin and Rob W. H. Ruigrok

Publié dans : Journal of Virology (2010) 84, 3707-3710

Structure of the Dimerization Domain of the Rabies Virus Phosphoprotein[∇]

Ivan Ivanov,^{1,2} Thibaut Crépin,² Marc Jamin,² and Rob W. H. Ruigrok^{2*}

Institut Laue Langevin, 6 Rue Jules Horowitz, BP 156, 38042 Grenoble Cedex 9, France,¹ and UMI 3265 UJF-EMBL-CNRS, Unit for Virus Host Cell Interactions, 6 Rue Jules Horowitz, Grenoble, France²

Received 7 December 2009/Accepted 10 January 2010

The crystal structure of the dimerization domain of rabies virus phosphoprotein was determined. The monomer consists of two α -helices that make a helical hairpin held together mainly by hydrophobic interactions. The monomer has a hydrophilic and a hydrophobic face, and in the dimer two monomers pack together through their hydrophobic surfaces. This structure is very different from the dimerization domain of the vesicular stomatitis virus phosphoprotein and also from the tetramerization domain of the Sendai virus phosphoprotein, suggesting that oligomerization is conserved but not structure.

Rabies virus is a negative-strand RNA virus of the Rhabdovirus family. Its genomic RNA is encapsidated by numerous copies of the viral nucleoprotein (N) that binds to the sugar phosphate backbone of the RNA with a stoichiometry of nine nucleotides per N protomer (1). RNA replication and transcription take place on this N-RNA template (2) and are catalyzed by the viral RNA-dependent RNA polymerase (L). L binds to the N-RNA template with the help of the polymerase cofactor, the phosphoprotein (P) (7). The phosphoproteins of the Rhabdoviridae and of the Paramyxoviridae are oligomers (5, 8); P of the Rhabdoviridae (rabies virus and vesicular stomatitis virus [VSV]) form dimers (6, 9), whereas P of the Paramyxoviridae (Sendai virus) form tetramers (21). These phosphoproteins are modular proteins that have an N-terminal domain that keeps newly produced N (called N⁰) in a soluble, RNAfree form, a central oligomerization domain and a C-terminal domain that binds to N-RNA. The three domains are connected by two intrinsically disordered regions (10, 13). The atomic structures of the oligomerization domains of the phosphoproteins of VSV and Sendai virus have been determined, and it was found that these structures were quite different (6, 21). Because the secondary structure prediction of the oligomerization domain of rabies virus P was different again from that of Sendai virus and VSV (10), we decided to determine its structure.

The sequence analysis of rabies virus P suggests that the oligomerization domain stretches from amino acid residues 92 to 131 (10). The DNA corresponding to residues 91 to 133 was cloned into a pET22b vector and expressed in the *Escherichia coli* BL21 (DE3-RIL) strain as a His tag fusion protein. The protein was purified by nickel resin and size-exclusion chromatography (S75). In solution the protein behaves as a dimer with a molecular mass of 13.9 kDa (monomer, 6,501 Da) (10). Monomers or higher-order oligomers have never been observed.

The protein was crystallized at 20°C in 50 mM sodium caco-

dylate (pH 6.5)–10 mM MgSO₄ plus 2 M (NH₄)₂SO₄. Diffraction data on frozen crystals were obtained on beam lines ID14-4 and BM14 (ESRF, Grenoble, France). The crystals diffracted to a 1.5-Å resolution and belong to space group I4₁22 with one dimer in the asymmetric unit. The phases were solved by the Sulfur-MAD method (Table 1). Integration and scaling of the data were performed with XDS (12). Three methionine sites were found with the automated software Auto-RickShaw (17). An initial 2.5-Å resolution model was built, which was then used for molecular replacement with MOLREP (22) using the native data set. ARP-wARP (18) was

TABLE 1. Data collection and refinement statistics

Donomotor	Anomalo	Notivo orretol	
Parameter	Set 1	Set 2	Native crystal
General data			
X-ray source	BM14	BM14	ID14-4
Wavelength (Å)	1.771	0.885	0.976
Space group	I4122	I4122	I4122
Cell dimensions	-	-	-
a (Å)	43.15	43.23	43.29
b (Å)	43.15	43.23	43.29
c (Å)	190.90	191.44	192.02
$\alpha = \beta = \gamma$ (°)	90.00	90.00	90.00
Resolution range $(\text{\AA})^a$	25–2.11 (2.19–2.11)	25-2.0 (2.07-2.0)	25.0-1.5 (1.56-1.49)
Completeness (%) ^a	97.6 (86.7)	97.9 (96.3)	93.6 (90.6)
$R_{\rm sym}(I) (\%)^{a,b}$	5.7 (31.7)	5.5 (30.8)	5.1 (42.5)
$I/\sigma I^a$	66.75 (11.53)	85.80 (21.81)	20.36 (4.75)
Total reflections ^a	428,072 (24,834)	478,480 (45,631)	133,743 (13,727)
Unique reflections ^a	9,704 (904)	6,451 (595)	17,851 (1,755)
Multiplicity ^a	44.11 (27.47)	74.17 (76.69)	7.49 (7.82)
Refinement statistics			
R-factor (%)			19.9
$R_{\rm free}$ (%)			23.5
rmsd from ideal values			
Bond length (Å)			0.02
Bond angle (°)			1.46
Mean <i>B</i> -factor $(Å^2)$			22.4
Ramachandran plot Most favored regions (%)			100

^a Values in parentheses are for the highest-resolution shell. rmsd, root mean square deviations.

 $\int R_{\text{sym}}(I) = [\sum_{hkl} \Sigma_i | \langle I_{hkl} \rangle - I_{hkl,i} |] [\Sigma_{hkl} \Sigma_i | I_{hkl} |], \text{ where } i \text{ is the number of reflection } hkl.$

^{*} Corresponding author. Mailing address: UVHCI, UMI 3265 UJF-EMBL-CNRS, BP 181, 38042 Grenoble, Cedex 9, France. Phone: (11) 33-476-207273. Fax: (11) 33-476-297199. E-mail: ruigrok@embl.fr.

^v Published ahead of print on 20 January 2010.



FIG. 1. Structure of the monomer in the dimerization domain of the rabies virus phosphoprotein. (A) Front and back views of the helical hairpin monomer showing residues in the intramonomer interface. The amino acids forming the hydrophobic core holding the two helices together are indicated, as well as two hydrogen bonds (green stippled lines). (B) Surface potential (from -5 kT/e [red] to +5 kT/e [blue]) of the hydrophilic (left) and hydrophobic (right) surfaces of the monomer; the molecules have the same orientation as in panel A.

used to add water molecules. The model was improved by hand and finally refined using REFMAC5 (16), and the quality of the model was checked with PROCHECK (14) (Table 1).

Figure 1A shows two views of the monomer. Residues 90 to

109 form an N-terminal α -helix linked to a C-terminal helix (residues 114 to 133) by a loop (residues 110 to 113). The two helices are kept together by an extensive hydrophobic intramonomer interface, as indicated in Fig. 1A, completed by



FIG. 2. Structure of the dimerization domain of the rabies virus phosphoprotein. (A and B) Front and side views of the dimer. Residues in the interface are depicted in panel B. (C) Top-down view of the dimer showing that the N-terminal helix of one monomer interacts mainly with the C-terminal helix of the other. (D) Tryptophan 118 lodging in the hydrophobic cavity of the opposite monomer. Also indicated is the hydrogen bond between W118 and the main chain carbonyl oxygen of L97. (E) Fluorescence emission spectra of the intact phosphoprotein (P), the isolated C-terminal domain (P_{CTD}), and the dimerization domain (P_{CED}), showing the large blue shift of W118.

hydrogen bonds between Y96 and the main chain carbonyl of E124 and between Q103 and E124 (Fig. 1A). The two helices form a rather flat surface with a charged and a hydrophobic face (Fig. 1B). The dimer is formed by the interaction of the two hydrophobic faces, the two monomers crossing with an angle of about 46° (Fig. 2A and B), close to the value of 50° typically encountered in helix-to-helix packing in globular proteins, but different from the angle of 20° observed in four-helix bundles (4). The surface of the interface shows a conspicuous bulge next to a hydrophobic cavity (Fig. 1B). The bulge is formed by W118 and the cavity by L97, V100, G101, and T121 (Fig. 2D). When the two monomers lock together in the dimer, both tryptophans insert into the opposite cavities. The relative positions of W118 and the cavity contribute to or define the crossing angle. Residue W118 is deeply buried within the hydrophobic dimer interface, as shown by the large blue shift of the fluorescence emission spectrum of the dimerization domain ($\lambda_{max} = 318$ nm), compared to the emission spectrum of the C-terminal domain (containing W186 and W265; maximum at 344 nm) and to that of the full-length protein (maximum at 333 nm) (Fig. 2E). Other residues involved in the dimer interface are (chain A-chain B): M108-M108, D98-V122, I125/V122-L97, I104-F114, V129-F93/V129, and F93-V126, most of which are not involved in the intramonomer interface. The hydrogen bonds contributing to the dimer interface are (main chain interactions in italics): L97-W118, W118-L97, E123-Q94, and S119-Q94, the last two mediated by water molecules. In general, the N-terminal helix of one monomer interacts with the C-terminal helix of the other monomer (Fig. 2C).

Some of the amino acids in the hydrophobic interface were mutated to alanine, and the effects on dimerization were measured with the yeast two-hybrid system (10). Mutation to Ala of Y128 in the intramonomer interface or of different residues in the subunit interface (F114A, W118A, and I125A) led to loss of dimerization. An estimation of the contribution of each of these residues to the overall stability of the dimerization domain was obtained by using the FoldX web server (20). The predicted free-energy difference between the mutant and the wild-type proteins ($\Delta\Delta G$) is large for each of these mutants (Y128A, 4.9 kcal mol⁻¹; F114A, 10.0 kcal mol⁻¹; W118A, 9.6 kcal mol⁻¹; and I125A, 6.3 kcal mol⁻¹). The Y129A mutation also in the dimerization interface reduced but did not eliminate dimerization (calculated $\Delta\Delta G$ of 4.1 kcal mol⁻¹).

The structure of the oligomerization domain of the rabies virus phosphoprotein is very different from those of VSV and Sendai virus (Fig. 3). The VSV monomer consists of an α -helix preceded and followed by a two-stranded β -sheet. In the dimer (Fig. 3B), the two helices interact mainly through a hydrophobic cluster at their N-terminal ends. At both sides of the helices, the N-terminal sheet of one subunit forms a four-stranded sheet with the C-terminal sheet of the other subunit, stabilizing the dimer further by adding main chain hydrogen bonds. There are additional hydrophobic interactions between the two helices and the sheets (6). The crystal packing of the VSV dimerization domain suggests how higher-order oligomers might be formed, but this is not obvious from the packing of the rabies virus domain. The Sendai domain (Fig. 3C) forms a tetrameric coil that is stabilized by four times three N-terminal helices forming a small hydrophobic core (21). The largest functional

NOTES 3709



FIG. 3. Comparison of the oligomerization domains of the phosphoproteins of rabies virus, VSV, and Sendai virus. (A and B) Dimerization domains of the rabies virus and VSV phosphoproteins, respectively. (C) Tetramerization domain of Sendai virus P.

difference between the three oligomerization domains is that in the rabies virus P dimer the N-terminal end that is linked to the N^0 binding domain is positioned next to the C-terminal end that is linked to the N-RNA binding domain, whereas in both the VSV and the Sendai virus structures the N- and C-terminal ends are at opposite ends of the oligomerization domain. However, this difference may not have biological consequences because for all three phosphoproteins the N^0 and the N-RNA binding domains are connected to the oligomerization domain by extensive disordered regions (10).

There is no sequence similarity between the three oligomerization domains, but sometimes structural conservation can be recognized between proteins with different sequences, as was the case for the N-RNA binding domains of the rabies virus and VSV phosphoproteins (15, 19). In contrast, there is no structural conservation of the oligomerization domains. Similarly, the structures of the N-RNA binding domains of rabies virus and VSV P are very different from those of measles and Sendai virus (3, 11). For these phosphoproteins it seems that their modular organization consisting of three functional domains separated by two disordered regions is conserved rather than the structure of the functional domains.

The coordinates have been deposited in the RCSB PDB as PDB ID code 3L32.

This study was supported by a grant from the French ANR (ANR-07-001-01; ANRAGE).

We thank Giuseppe Zaccai (Institut Laue Langevin, Grenoble, France), Laurence Serre (PSB), and Nicolas Tarbouriech, Hassan Belrhali, and Nicolas Martinelli (UVHCI) for extensive discussions and

help; Francine Gérard for the fluorescence spectroscopy measurements; and the Partnership for Structural Biology in Grenoble for the excellent structural biology environment.

REFERENCES

- Albertini, A. A. V., A. K. Wernimont, T. Muziol, R. B. G. Ravelli, C. R. Clapier, G. Schoehn, W. Weissenhorn, and R. W. H. Ruigrok. 2006. Crystal structure of the rabies virus nucleoprotein-RNA complex. Science 313:357– 360.
- Arnheiter, H., N. L. Davis, G. Wertz, M. Schubert, and R. A. Lazzarini. 1985. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. Cell 41:259–267.
- Blanchard, L., N. Tarbouriech, M. Blackledge, P. Timmins, W. P. Burmeister, R. W. H. Ruigrok, and D. Marion. 2004. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. Virology 319:201–211.
- Chothia, C., M. Levitt, and D. Richardson. 1981. Helix-to-helix packing in proteins. J. Mol. Biol. 145:215–250.
- Curran, J., R. Boeck, N. Lin-Marq, A. Lupas, and D. Kolakofsky. 1995. Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. Virology 214:139–149.
- epitope dilution assay, via predicted coiled coils. Virology 214:139–149.
 Ding, H., T. J. Green, S. Lu, and M. Luo. 2006. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. J. Virol. 80:2808–2814.
- Emerson, S. U., and M. Schubert. 1987. Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. Proc. Natl. Acad. Sci. U. S. A. 84:5655–5659.
- Gao, Y., and J. Lenard. 1995. Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II. EMBO J. 14:1240–1247.
- 9. Gérard, F. C. A., E. de Almeido Ribeiro, Jr., A. A. V. Albertini, I. Gutsche, G. Zaccai, R. W. H. Ruigrok, and M. Jamin. 2007. Unphosphorylated *Rhabdoviridae* phosphoproteins form elongated dimers in solution. Biochemistry 46:10328–10338.
- Gérard, F. C. A., E. de Almeido Ribeiro, Jr., C. Leyrat, I. Ivanov, D. Blondel, S. Longhi, R. W. H. Ruigrok, and M. Jamin. 2009. Modular organization of rabies virus phosphoprotein. J. Mol. Biol. 388:978–996.

- Johansson, K., J. M. Bourhis, V. Campanacci, C. Cambillau, B. Canard, and S. Longhi. 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.
- Kabsch, W. 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. 26:795–800.
- Karlin, D., F. Ferron, B. Canard, and S. Longhi. 2003. Structural disorder and modular organization in *Paramyxovirinae* N and P. J. Gen. Virol. 84: 3239–3252.
- Laskowski, R. A., M. W. Mac Arthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structure. J. Appl. Crystallogr. 26:283–291.
- Mavrakis, M., A. A. McCarthy, S. Roche, D. Blondel, and R. W. H. Ruigrok. 2004. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J. Mol. Biol. 343:819–831.
- Murshudov, G. N. 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53:240– 255.
- Panjikar, S., V. Parthasarathy, V. S. Lamzin, M. S. Weiss, and P. A. Tucker. 2005. Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. Acta Crystallogr. D 61:449–457.
- Perrakis, A. R. Morris, and V. S. Lamzin. 1999. Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. 6:458–463.
- Ribeiro, E. A., Jr., A. Favier, F. C. A. Gérard, C. Leyrat, B. Brutscher, D. Blondel, R. W. H. Ruigrok, M. Blackledge, and M. Jamin. 2008. Solution structure of the C-terminal N-RNA binding domain of the vesicular stomatitis virus phosphoprotein. J. Mol. Biol. 382:525–538.
- Schymkowitz, J., J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano. 2005. The FoldX web server: an online force field. Nucleic Acids Res. 33:W382–W388.
- Tarbouriech, N., J. Curran, R. W. H. Ruigrok, and W. P. Burmeister. 2000. Tetrameric coiled coil domain of Sendai virus phosphoprotein. Nature Struct. Biol. 7:777–781.
- Vagin, A., and A. Teplyakov. 2000. An approach to multi-copy search in molecular replacement. Acta Crystallogr. D Biol. 229:105–124.

2.5 Conclusion

La comparaison des informations structurales collectées pour les phosphoprotéines de VSV et de RAV ainsi que pour les phosphoprotéines de Paramyxovirus (virus de la rougeole, des oreillons, de Sendai et du virus respiratoire syncytial) indique que les différentes parties de cette protéine présentes des taux différents de conservation. La structure du domaine C-terminal de VSV et RAV d'une part et des Paramyxovirus d'autre part semble être conservé entre les deux familles. Il est possible de le superposer partiellement indiquant une conservation du domaine entre ces familles de virus. Par opposition, le domaine central est différent. Celui des Paramyxovirus forme un tétramère alors que celui des Rhabdovirus forme un dimère. La structure est différente même au sein des Rhabdovirus, et la comparaison de structure ne permet pas de déterminer si ces domaines ont une origine commune. Les seules caractéristiques structurales de ce domaine qui semblent conservées sont la formation d'homooligomères et l'arrangement parallèle des monomères. Les autres régions sont désordonnées et de longueurs variables selon la protéine. La région N-terminale contient également un site de fixation de N° qui est localisé plus ou moins au même endroit dans la séquence en acides aminés et présente une propension à former une hélice α . L'organisation des domaines structurés et des régions désordonnées semblent également conservées dans la protéine (voir figure 2.7).



FIGURE 2.7 – Structure des phosphoprotéines des *Rhabdovirus*.

A : Représente une organisation modulaire de la phosphoprotéine du VSV. Les structures haute résolution sont représentées en cartoon du domaine central (VSVP_{CED}) et C-terminal domaine (VSVP_{CTD}) sont intégrées dans le modèle de la phosphoprotéine. Les régions désordonnées sont représentées en pointillées. Le cercle en pointillées représente la région MoRE responsable de la liaison à N°. Les codes pdb sont donnés sur la figure.

B : Représente une organisation modulaire de la phosphoprotéine de RAV. La région C-terminale intrinséquement dépliée (IDR_{CT}) en pointillées est plus longue de 60 acides aminés pour RAV que pour VSV. Les structures haute résolution et les codes pdb sont également présents sur la figure.

Cette organisation modulaire de la P commune aux Rhabdovirus et aux Paramyxovirus peut s'expliquer de différentes manières. Soit ces protéines ont un ancêtre commun mais les différentes régions de la protéine ont évoluées à des vitesses différentes parce que la pression de sélection n'était pas la même selon fonction. Si la fonction du domaine central est l'homo-oligomérisation, il est possible d'imaginer que cette structure ait évolué plus rapidement que le domaine C-terminal qui est impliqué dans la reconnaissance de la matrice et qui doit donc conserver une structure adéquate pour cette fonction. Une autre possibilité est que les protéines soient apparues de manière indépendante, mais il est alors difficile d'imaginer qu'une évolution convergente ait conduit à l'organisation des différents modules dans le même ordre dans toutes les protéines. La troisième possibilité est qu'il y ait des transferts de domaines à partir notamment de domaines de protéines cellulaires. Les différentes protéines seraient alors le résultat à la fois d'une évolution divergente et d'une évolution horizontal par échange de domaines.

A l'heure actuelle aucun résultat expérimental n'explique pourquoi la phosphoprotéine des Mononegavirales se présente sous forme oligomérique.

Au cours du processus de synthèse d'ARN, la P a pour rôles de maintenir la L associée à la matrice N-ARN et de permettre la progression de celle-ci le long de cette matrice. Pour les Paramyxovirus, une délétion du domaine de tétramérisation entraîne un défaut de transcription. D'autre part, l'attachement de la P sur la matrice N-ARN nécessite la fixation simultanée d'au moins deux domaines C-terminaux (Curran 1998). Sur la base de ces résultats, il a été proposé que la P progresse le long de la matrice par un mécanisme de rotation (cartwheeling) dont l'axe serait constitué par le domaine central (Kolakofsky et al. 2004).

Toutefois des résultats contradictoires ont été obtenus avec le RAV. La délétion du domaine central de la phosphoprotéine du RAV n'affecte pas la transcription virale (Jacob et al. 2001), alors que ce mutant est monomérique en solution et conserve sa capacité à se fixer sur le complexe N-ARN (Gerard et al. 2009). Ces résultats prouvent que la forme dimérique n'est pas nécessaire à la transcription virale. D'autre part, des expériences *in vitro* de résonance plasmonique de surface (SPR) ont permis d'étudier l'interaction entre des anneaux de RAV N₁₀ et le domaine P_{CTD} . Les cinétiques d'association et de dissociation sont lentes et ne semblent pas compatibles avec un mécanisme dans lequel la P devrait continuellement s'attacher et se dissocier de la matrice N-ARN. La fixation d'un P_{CTD} isolé est forte, avec une constante de dissociation de 140 nM (de Almeida Ribeiro et al. 2009). La constante d'affinité pour la P entière est dix fois plus grande, soit ~10 nM (Ribeiro et al. unpublished data). Toutes ces mesures d'affinité pour le C-terminal seul ou la protéine entière montrent qu'un mécanisme de rotation (cartwheeling) n'est pas possible.

Chapitre 3

Expression, purification et recherche de domaines solubles de la polymérase de VSV

Chapitre 3. Expression, purification et recherche de domaines solubles de la polymérase 84 de VSV

$\mathbf{\alpha}$			•
20	mr	\mathbf{na}	ire

3.1	Intro	oduction	84
3.2	Exp	ression de fragments en bactéries et cellules d'insectes	85
	3.2.1	Sélection des fragments	85
	3.2.2	Vecteurs d'expressions, milieux de culture et tests à petite échelle	86
	3.2.3	Expression et solubilité à petite échelle	87
3.3	Puri	fication du fragment N-terminal en fusion avec la MBP	88
	3.3.1	Purification sur chromatographie d'affinité (amylose)	88
	3.3.2	Purification par chromatographie d'exclusion de taille	90
	3.3.3	Purification par colonne Bleu sepharose	90
	3.3.4	Recherche de tampon de stabilisation du domaine N-terminal après clivage à la TEV	91
3.4	Exp	ression du complexe L-P en cellules d'insectes	93
	3.4.1	Introduction	93
	3.4.2	Clonage du complexe L-P et tests d'expression	93
	3.4.3	Expression du complexe L-P et purification sur colonne d'affinité Ni-NTA	94
	3.4.4	Purification du complexe L-P par gel filtration et analyse par Western Blot	96
	3.4.5	Caractérisation structurale et fonctionnelle du complexe	97
3.5	Con	clusions	98

3.1 Introduction

Que ce soit pour des tests de biochimie ou pour une étude structurale, l'obtention d'une protéine pure et homogène est indispensable. La protéine L des MNV est une protéine de grande taille de 2109 aa avec plusieurs domaines. A ce jour, 6 domaines ont été identifiés (Poch et al. 1989), mais aucune information structurale n'est connue. De plus, nous ne disposons pas d'information sur le repliement autonome de ces domaines et personne n'a jusqu'à ce jour pu les exprimer séparément. L'importance des domaines I, II et IV (figure 1.17 page 32) dans l'activité de la polymérase n'est également pas connue. Des hypothèses peuvent être émises sur le rôle de ces domaines. Ils pourraient notamment participer au recrutement de facteurs cellulaires essentiels ou participer directement à la régulation de l'activité de la polymérase.

Dans des cellules infectées comme dans le virus, la polymérase n'est jamais présente sous forme libre mais est toujours en complexe avec son cofacteur : la phosphoprotéine (P), formant ainsi un complexe L-P avec une stoechiométrie de deux polymérases pour un dimère de phosphoprotéine (Rahmeh et al. 2010). Le site d'interaction de la polymérase sur la phosphoprotéine a été cartographié dans la région N-terminale (Emerson & Schubert 1987, Castel et al. 2009). La phosphoprotéine joue le rôle important dans le positionnement de la polymérase sur la matrice N-ARN. Cependant de nombreuses questions restent encore en suspend : comment accède-t-elle au substrat ou encore comment s'effectue son déplacement le long de la nucléocapside. La fixation de P sur la matrice est effectuée par son domaine C-terminal. L'ensemble de ces informations en font un système de régulation complexe dont la biologie n'a pas encore été clairement établie.

L'objectif à long terme de cette étude est d'obtenir des informations structurales par cristallographie aux rayons X ou par microscopie électronique (cryo-microscopie) du complexe L-P mais aussi de comprendre le rôle de la phosphoprotéine, le mécanisme de progression de la polymérase sur l'ARN viral ainsi que l'origine moléculaire de la « processivité » de cette enzyme.

Une grande partie de ma recherche a été consacrée à l'identification et la détermination de domaines solubles de la polymérase, mais également à l'expression de la protéine L seule ou en complexe avec son cofacteur - P en cellules d'insectes (plateforme EEF, Dr. Imre Berger, EMBL, Grenoble). Le gène de la polymérase de VSV (souche Indiana) a été synthétisé par DNA2.0 et la séquence en acides nucléiques a été optimisée afin de substituer les codons rares par des codons plus abondants pour l'expression en bactéries et en cellules d'insectes.

3.2 Expression de fragments en bactéries et cellules d'insectes

3.2.1 Sélection des fragments

Les fragments clonés et exprimés, présentés dans la figure 3.1 page 85 et le tableau 3.1 page 87 ont été choisis sur les critères suivants :

- Sur la base des domaines conservés I-VI, j'ai cloné des fragments de la polymérase qui correspondaient aux domaines associés à une activité enzymatique.
- Prédictions de structures secondaires et prédiction de désordre (site web).
- Prédictions de désordre : une prédiction de désordre est calculée comme le consensus des résultats de seize différents programmes comme décrit dans (Gerard et al. 2009).

L'analyse des résultats sur la polymérase de VSV mais également sur la polymérase de RAV a donné des résultats similaires. La protéine entière est prédite ordonnée mais on peut observer des inflections dans le D-score qui pourraient correspondre aux limites des domaines.

Sur la base de ces analyses j'ai décidé de cloner les constructions présentées dans la figure (3.1 page 85).







La polymérase est présentée avec les six régions conservées comme indiquée dans Galloway et al. (2008). Les traits audessous représentent les constructions clonées et exprimées dans les systèmes d'expression eucaryote et bactérie. L'échelle sur la figure correspond à 50 aa.

3.2.2 Vecteurs d'expressions, milieux de culture et tests à petite échelle

Les fragments de la polymérase (figure 3.1) ont été clonés d'une part dans 2 vecteurs d'expression procaryote (pETM11 comportant une étiquette Histidine en N-terminal pour la purification sur une chromatographie d'affinité Histidine, et pETM40 portant une étiquette MBP - protéine de \sim 40kDa exprimée en fusion avec la protéine d'intérêt du côté N-terminal permettant d'augmenter la solubilisation des protéines recombinantes et donc des différents construits de la polymérase), et d'autre part dans un vecteur pFastBacHtb adapté pour les expressions en système eucaryote.

Le gène de la polymérase entière a été cloné par la compagnie « $DNA2.0 \gg$ dans le vecteur pFastbac1 de façon à pouvoir l'exprimer directement en cellules d'insectes. C'est ce vecteur d'origine que j'ai utilisé pour cloner manuellement les constructions présentées dans le tableau 3.1.

Une fois les séquences insérées dans les vecteurs d'expression procaryote, des tests d'expressions à petite échelle ont tout d'abord été réalisés dans des bactéries BL21(DE3) (vecteur pETM11, pETM40) en petit volume de 5 ml. L'induction de l'expression est réalisée avec 1 mM IPTG final et l'expression a été réalisée à 2 températures : 37° C et 16° C. Le temps d'expression a été de 5 heures pour l'expression à 37° C et O/N pour l'expression à 16° C. A la suite de cette étape de production, j'ai débuté les phases de purification en utilisant un tampon standard : 200 mM NaCl, 5% glycérol, 1 mM DTT ou bien 2 mM β ME et le pH du tampon a été choisi en fonction du pI de la protéine (20 mM Tris pH7-8 ou 20 mM MES pH 6.0). La lyse des bactéries est effectuée chimiquement en présence de BugBuster (Novagen) qui comporte un mélange de détergeant afin de lyser la parois bactérienne et d'obtenir l'extrait brut (cell extract-CE). Une centrifugation est réalisée à 11 200 g dans une centrifugeuse de paillasse afin de séparer la phase soluble (soluble fraction-SF) de la phase non-soluble. Un SDS-PAGE et un Western Blot avec un anticorps monoclonal α His sont réalisés afin de déterminer le degré de solubilité et de pureté de la protéine. La même méthodologie a été utilisée pour les tests d'expression des constructions avec une étiquette MBP.

La technique d'expression en système eucaryote nécessite des étapes supplémentaires. Un clonage doit être effectué dans un vecteur pFastBacHtb afin de faire la transposition dans un bacmide codant pour les protéines des cellules d'insectes. Une transformation dans des bactéries DH10Bac-YFP est effectuée et la sélection est effectuée sur une gélose comportant de l'IPTG, X-Gal et des antibiotiques. Les colonies positives ont une couleur blanche car les bactéries ne sont pas capables d'hydrolyser le X-Gal suite à l'insertion du gène provenant du pFastBac dans le génome du bacmide. Au contraire, les colonies qui ne possèdent pas le gène d'intérêt ont une couleur bleue car elles sont capables d'hydrolyser le X-Gal.

Les colonies blanches sont prélevées et des MiniPreps de bacmides sont réalisées avec un kit commercial (MiniPreps Qiagen). Le bacmide est dosé et transfecté dans des cellules d'insectes en plaque 6 puits en présence de lypofectamine (Invitrogen) qui permet le passage du bacmide à travers la membrane des cellules d'insectes par un processus de fusion et d'endocytose. Une incubation pendant 48 h est réalisée afin de produire 2 ml de virus (baculovirus) appelé V0 capable d'infecter d'autres cellules d'insectes qui vont pouvoir produire la protéine d'intérêt. Ces 2 ml de virus sont nécessaires à la réamplification de virus afin d'obtenir au final un stock viral de 50 ml appelé (V1). C'est ce stock viral qui nous servira par la suite pour exprimer les protéines à grande échelle. Les virus (V1) sont stockés à 4°C. Pour les expression à grande échelle, 400 ml de cellules d'insectes sont infectés avec le virus (V1) à 0.1 % (v/v). Les 400 ml de cellules sont centrifugés à basse vitesse (JA.10) et sont lavées deux fois avec un tampon isotonique PBS. La lyse est effectuée dans un tampon classique.

constructions hôte d'expression et étiquette	14-410	1595-2109	1050-1312	600-790	486-827	486-1847
bactéries (étiquette His)	-	-	Х	-	-	-
bactéries (étiquette MBP)	++	++	Х	++	++	+
cellules d'insectes (étiquette His)	Х	Х	Х	Х	Х	-
cellules d'insectes (étiquette MBP)	Х	Х	Х	Х	Х	+
constructions	855-1847	486-1059	1627-1847	1307-1847	1050-1847	1-2109
hôte d'expression et étiquette						
hôte d'expression et étiquette bactéries (étiquette His)	-	-	-		-	X
hôte d'expression et étiquette bactéries (étiquette His) bactéries (étiquette MBP)	- +	- +	-++	-++	- ++	X X
hôte d'expression et étiquette bactéries (étiquette His) bactéries (étiquette MBP) cellules d'insectes (étiquette His)	- + -	- + -	- ++ +	- ++ +	- ++ +	X X +

TABLE 3.1 – Tests de solubilisation des constructions polymérasiques en bactéries et cellules d'insectes.

Le "-" signifie que la protéine n'est pas du tout soluble et est exprimée en corps d'inclusion, le "+" représente une solubilité faible, "++" bonne solubilité et "X" la solubilité n'a pas été testée.

La plateforme d'Imre Berger (EMBL, Grenoble) pour l'expression de protéines ou de complexes protéiques permet de cultiver des cellules d'insectes en suspension, ce qui permet d'obtenir une plus grande masse de cellules pour un faible volume de culture. Un second avantage est le suivi de l'expression

Chapitre 3. Expression, purification et recherche de domaines solubles de la polymérase 88 de VSV

protéique grâce au gène YFP intégré dans le bacmide qui est sous le contrôle du même promoteur que la protéine d'intérêt. Les cellules d'insectes sont collectées dès que l'intensité de fluorescence de la YFP mesurée à l'aide d'un fluorimètre dans la fraction soluble atteint un plateau. Dès que les cellules arrêtent de se diviser, le signal de fluorescence de la YFP commence à augmenter signifiant un début d'expression protéique. En général le plateau de signal de YFP arrive après 72h de dpa (death post arrest).

3.2.3 Expression et solubilité à petite échelle

A l'analyse des résultats d'expression protéique, il est apparu que certaines constructions n'étaient pas du tout exprimées, d'autres avec une étiquette histidine étaient exprimées en corps d'inclusion et n'ont pas pu être solubilisées (voir tableau 3.1 page 87). Finalement, certaines constructions étaient très bien exprimées mais fixaient des chaperonnes.

Parmi ces constructions une protéine était bien exprimées avec un degré de solubilité satisfaisant. Par la suite, je vais donc vous présenter les résultats obtenus avec la construction MBP-F14-F410 en système d'expression bactérien.

3.3 Purification du fragment N-terminal en fusion avec la MBP

La construction N-terminale (F14-F410) a présenté pour moi un intérêt particulier, d'une part celle-ci n'est pas bien caractérisée dans la littérature, et d'autre part sa taille est acceptable pour une expression dans un système bactérien. Certaines publications suggèrent que ce domaine pourrait être impliqué dans l'interaction avec des protéines cellulaires. Notre objectif était donc de réaliser des expériences de « pull down » avec des extraits cellulaires afin d'identifier les partenaires cellulaires possibles.

3.3.1 Purification sur chromatographie d'affinité (amylose)

L'expression de la protéine a été effectuée comme expliqué plus haut à 16°C dans un système bactérien en fusion avec la MBP. Les cellules sont lysées dans le tampon classique (20 mM Tris pH8.0, 200 mM NaCl, 5% glycérol, 1 mM DTT) par sonication et le lysat cellulaire est centrifugé pour séparer les fractions soluble et non-soluble. La purification de la protéine est effectuée avec une chromatographie d'affinité amylose. Une étape de lavage est effectuée à haut sel afin d'éliminer les acides nucléiques fixés de manière non-spécifique et autres protéines. Une réequilibration à bas sel est effectuée avec le même tampon de lyse. L'élution est effectuée dans le même tampon additionné de 30 mM de maltose qui par compétition pour la résine amylose décroche la protéine de fusion MBP-F14-F410. Le gel de purification est présenté sur la figure 3.2.



FIGURE 3.2 – Analyse par gel SDS-PAGE de la purification de la construction MBP-F14-F410.

Les poids moléculaires (MW) sont présentés à gauche du gel SDS-PAGE. Les dépôts sont comme suit : extrait brut (CE), fraction soluble (SF), fraction non liée à la colonne d'affinité amylose (FT), lavage à haut sel (Wash). La protéine en sortie de colonne (Elution Fractions) est collectée par des fractions de 0.5 CV. La migration a été effectuée sur un gel 10% SDS-PAGE (Laemmli).

Le domaine N-terminal de la polymérase en fusion avec la MBP ($\sim 40 \text{ kDa}$) en gel SDS-PAGE migre avec un poids moléculaire apparent de 92 kDa. La fusion avec la MBP permet d'obtenir sous forme soluble le domaine recherché (puits CE et SF sur la figure 3.2). On peut bien noter que la première étape de purification permet d'avoir une protéine relativement pure avec peu de contaminants présents dans les fractions d'élution. Néanmoins des traces de chaperonne (GroEL) sont détectables à la limite de résolution du gel SDS-PAGE (voir figure 3.2).

3.3.2 Purification par chromatographie d'exclusion de taille

Une deuxième étape de purification a été envisagée à 4°C sur une chromatographie préparative d'exclusion de taille. Les fractions en sortie de la colonne d'affinité (amylose) (voir figure 3.2, elution fractions) ont été concentrées comme décrit auparavant par centrifugation avec un concentrateur Vivaspin (Sartorius) à une vitesse de 1560 g et la protéine concentrée a été injectée sur une colonne S200 (figure 3.3 A).



FIGURE 3.3 – Profil de gel filtration et analyse par gel SDS-PAGE du fragment MBP-F14-F410.

A : Représente le profil d'élution de la colonne d'exclusion de taille S200 (Amersham). Le trait rouge correspond à l'absorbance mesurée à 280 nm. La trait bleu correspond à la conductivité (mSv), constante lors d'une gel filtration, le tampon de gel filtration est 20 mM Tris pH 8.0, 150 mM NaCl, 10 mM β ME , 5% Glycérol et 1 mM DTT. B : Représente les fractions déposées sur gel 10% SDS-PAGE (Laemmli). La protéine fusion et la chaperonne sont montrées sur le gel.

La purification par chromatographie d'exclusion de taille n'a pas permis d'éliminer les chaperonnes qui restent accrochées à la protéine et sont éluées sur toute la largeur du pic d'élution, indiquant qu'elles sont liées très étroitement à la protéine d'intérêt. On observe également des produits de dégradation pouvant indiquer la présence de fragments non structurés ou dans un état non stable.

3.3.3 Purification par colonne Bleu sepharose

Une dernière tentative de purification de la protéine (MBP-F14-F410) a été réalisée avec une colonne Bleu (Amersham) qui est vendue comme une colonne multifonction (échangeuse d'anions, de cations, mais également elle peut être utilisée comme une colonne hydrophobe).



FIGURE 3.4 – Profil d'élution de la colonne Bleu et analyse par gel SDS-PAGE du fragment MBP-F14-F410.

A : Représente le profil d'élution par gradient allant de 100 à 1000 mM NaCl. L'absorbance à 280 nm (unité arbitraire) est présentée par la courbe rouge. La courbe noire en palier correspond au profil de gradient de sel crée par le système Biorad (Amercham). A chaque élution de protéine le gradient est arrêté de façon manuelle afin de séparer au maximum les différentes espèces. La courbe bleue représente la conductivité du tampon d'élution. Le pic I est élué à 217 mM NaCl, le pic II à 375 mM NaCl. Les pic I et II sont analysés en gel SDS-PAGE en B.

B : Les deux pics d'élution à 217 et 375 mM NaCl ainsi que le lavage en A sont analysés sur gel 10 % SDS-PAGE (Laemmli). Les dépôts sont indiqués en haut du gel : poids moléculaire-MW, échantillon chargé sur la colonne Bleu (load), lavage (W) et ensuite les deux pics d'élution, la coloration est effectuée en Blue de Coomassie.

La protéine purifiée sur la chromatographie d'exclusion de taille a été concentrée et purifiée par la colonne Bleu (Amersham). Le chargement de la protéine est effectué avec un tampon classique Tris-HCl pH 8.0, 100 mM NaCl, 5% glycérol, 1 mM DTT. Pour l'élution de la protéine un gradient de sel est effectué allant de 100 à 1000 mM NaCl.

La colonne Bleu nous a permis de nous débarrasser entièrement des produits de dégradation ainsi que la chaperonne qui restaient étroitement liée à la protéine après purification par chromatographie d'affinité, chromatographie échangeuse d'anions et gel filtration (figure 3.2 page 88, 3.3B page 90). Après élution de la protéine de la colonne, celle-ci a été dialysée dans le même tampon contenant 200 mM NaCl.

Après clivage de l'étiquette MBP avec la protéase TEV, le domaine (F14-F410) a précipité suggérant fortement que ce domaine n'est pas stable dans le tampon de purification. J'ai alors procédé à la recherche d'un tampon de stabilisation par une technique haut débit.

3.3.4 Recherche de tampon de stabilisation du domaine N-terminal après clivage à la TEV

Dans le laboratoire, une expérience haut débit a été mise en place permettant d'identifier des conditions pour la solubilisation d'une protéine. J'ai adapté cette technique que l'ai appliquée à la plupart des fragments de la polymérase de VSV peu solubles après clivage de la MBP.
Chapitre 3. Expression, purification et recherche de domaines solubles de la polymérase 92 de VSV

La protéine en sortie de colonne Bleu (Amersham) est concentrée par centrifugation avec un concentrateur Vivaspin (Sartorius). $20 \,\mu$ l de protéines à 5 mg/ml sont mélangés avec $180 \,\mu$ l de 96 tampons de stabilisation différents en présence de $10 \,\mu$ gr de TEV. Après clivage O/N, les résultats sont analysés par mesure de la turbidité des échantillons par une lecture à 600 nm et les résultats sont présentés dans le tableau 3.2. 96 tampons ont été testés utilisant une large gamme de pH allant de 5.0 à 8.5 avec des additifs tels que le Triton-X 100, le glycérol, l'EDTA, le DTT et 2 sels différents NaCl et Ammonium Sulfate.

	1	2	3	4	5	6
Tampons	NaCl 500 mM	NaCl 100 mM	Amm. Sulfate $500\mathrm{mM}$	Amm. Sulfate $100\mathrm{mM}$	NaCl 250 mM	
					5 % glycérol	10 % glycérol
Ac. citrique pH 5.0						
Ac. citrique pH 5.5						
MES pH 6.0						
MES pH 6.5						
Hepes pH 7.0						
Hepes pH 7.5						
Tris pH 8.0						
Tris pH 8.5						
	7	8	9	10	11	12
Tampons	7	8	9	10 NaCl 250 mM	11	12
Tampons	7 EDTA 10 mM	8 MgCl ₂ 10 mM	9 Triton-X 100 0.01 %	10 NaCl 250 mM DTT 10 mM	11 EDTA, DTT, glycérol	12 MgCl ₂ , DTT, glycérol
Tampons Ac. citrique pH 5.0	7 EDTA 10 mM	8 MgCl ₂ 10 mM	9 Triton-X 100 0.01 %	10 NaCl 250 mM DTT 10 mM	11 EDTA, DTT, glycérol	12 MgCl ₂ , DTT, glycérol
Tampons Ac. citrique pH 5.0 Ac. citrique pH 5.5	7 EDTA 10 mM	8 MgCl ₂ 10 mM	9 Triton-X 100 0.01 %	10 NaCl 250 mM DTT 10 mM	11 EDTA, DTT, glycérol	12 MgCl ₂ , DTT, glycérol
Tampons Ac. citrique pH 5.0 Ac. citrique pH 5.5 MES pH 6.0	7 EDTA 10 mM	8 MgCl ₂ 10 mM	9 Triton-X 100 0.01 %	10 NaCl 250 mM DTT 10 mM	11 EDTA, DTT, glycérol	12 MgCl ₂ , DTT, glycérol
Tampons Ac. citrique pH 5.0 Ac. citrique pH 5.5 MES pH 6.0 MES pH 6.5	7 EDTA 10 mM	8 MgCl ₂ 10 mM	9 Triton-X 100 0.01 %	10 NaCl 250 mM DTT 10 mM	11 EDTA, DTT, glycérol	12 MgCl ₂ , DTT, glycérol
Tampons Ac. citrique pH 5.0 Ac. citrique pH 5.5 MES pH 6.0 MES pH 6.5 Hepes pH 7.0	7 EDTA 10 mM	8 MgCl ₂ 10 mM	9 Triton-X 100 0.01 %	10 NaCl 250 mM DTT 10 mM	11 EDTA, DTT, glycérol	12 MgCl ₂ , DTT, glycérol
TamponsAc. citrique pH 5.0Ac. citrique pH 5.5MES pH 6.0MES pH 6.5Hepes pH 7.0Hepes pH 7.5	7 EDTA 10 mM	8 MgCl ₂ 10 mM	9 Triton-X 100 0.01 %	10 NaCl 250 mM DTT 10 mM	11 EDTA, DTT, glycérol	12 MgCl ₂ , DTT, glycérol
TamponsAc. citrique pH 5.0Ac. citrique pH 5.5MES pH 6.0MES pH 6.5Hepes pH 7.0Hepes pH 7.5Tris pH 8.0	7 EDTA 10 mM	8 MgCl ₂ 10 mM	9 Triton-X 100 0.01 %	10 NaCl 250 mM DTT 10 mM	11 EDTA, DTT, glycérol	12 MgCl ₂ , DTT, glycérol

TABLE 3.2 – Matrice de stabilisation de la construction F14-F410.

Présentation de la composition de la matrice de stabilisation de la protéine F14-F410 rassemblant 96 conditions testées qui sont constitués de 8 tampons, de 2 sels, de 1 agent réducteur, ainsi que 4 additifs de l'EDTA, du glycérol, du MgCl₂, ou encore du Triton-X 100. Les cases en couleur gris-foncé correspondent aux tampons qui ne donnent pas d'agrégats visibles après clivage de la MBP à la TEV. Les cases blanches, la solution protéique devient trouble et des précipités blanchâtres apparaissent.

Cette matrice de stabilisation permettant de tester 96 conditions différentes a permis d'identifier, quelques tampons prometteurs sur des tests en petit volume. Malheureusement les résultats n'ont pas pu être reproduits dans des tests à grande échelle. Après plusieurs tests de purification, je n'ai pas réussi à obtenir la construction N-terminale de la polymérase sous forme soluble et en grande quantité.

Malgré la précipitation de la construction F14-F410 après clivage de la MBP, une quantité faible mais suffisante est restée exploitable pour essayer de cristalliser le domaine. J'ai procédé à un criblage par le robot de cristallographie (EMBL, Grenoble) mais aucun cristal n'a été obtenu.

3.4 Expression du complexe L-P en cellules d'insectes

Puisque la polymérase est présente dans le virus toujours en complexe avec son cofacteur-P (Emerson & Schubert 1987), nous avons essayé d'exprimer le complexe L-P en cellule d'insectes (plateforme EEF, Dr. Imre Berger, EMBL, Grenoble). En parallèle des tentatives d'expression de la polymérase seule ont été réalisées mais celle-ci s'est avéré peu soluble.

3.4.1 Introduction

Récemment le groupe de Dr. Sean Whelan (Harvard Medical School, Boston) a rapporté la production de la polymérase de VSV fonctionnelle sous une forme soluble par expression en cellules d'insectes. Des changements de conformations y interviennent lorsque la phosphoprotéine (exprimer en système bactérien) s'accroche à la polymérase (Rahmeh et al. 2010)).



FIGURE 3.5 – Vecteur de clonage et suivi de l'expression du complexe L-P par mesure de YFP.

A : Représente les vecteurs utilisés pour le clonage des deux gènes codant pour L et P. Le gène de la polymérase est cloné dans un vecteur pFastBac1. La séquence de la protéine L est montrée allant de 1 à 2109 au avec les six régions conservées (rectangles jaunes) montrées de I à VI. Le vecteur pUCDM a été fabriqué par le groupe de Dr. Imre Berger (EMBL, Grenoble) avec un origine de réplication $R6K\gamma$ et une résistance au chloramphénicol (Cm) (Fitzgerald et al. 2007).

Le gène codant pour la phosphoprotéine entière est cloné dans la cassette d'expression (EC1) sous le promoteur PH. La région N-terminale et la région de liaison à L, le domaine de dimérisation et le domaine de liaison à N-ARN sont présentés de I à IV respectivement par des cases rouges. Le site LoxP pourrait être utilisé pour la fusion de deux vecteurs. Les deux gènes codant pour L et P possèdent une étiquette Histidine.

B : L'histogramme représente le suivi de fluorescence de YFP qui est directement corrélé avec l'expression de la protéine. Les cellules sont collectées lorsque la fluorescence atteint un plateau. L'histogramme représente le suivi d'expression du complexe L-P en fonction de death post arrest (dpa).

3.4.2 Clonage du complexe L-P et tests d'expression.

Après avoir testé l'expression de différents domaines de la polymérase en bactéries et cellules d'insectes, je me suis également focalisé sur l'expression du complexe L-P. La figure 3.5, page 93 représente la stratégie de clonage que j'ai entrepris et la suivie de l'expression protéique par la méthode de mesure de la YFP comme expliqué dans 3.4.3 (plateforme EEF, Dr. Imre Berger, EMBL, Grenoble).

Chapitre 3. Expression, purification et recherche de domaines solubles de la polymérase 94 de VSV

La phosphoprotéine entière de VSV (souche Indiana) a été clonée dans un vecteur pUCDM (plateforme, Dr. Imre Berger, EMBL, Grenoble), sous la dépendance du promoteur PH (polyhédrine). Ce vecteur est un vecteur navette (transitoire) qui sera fusionner par la suite avec le vecteur pFastBac1-L (gène codant pour la polymérase) afin d'obtenir un baculovirus possédant ainsi les deux gènes. Par la suite ceci nous permettra d'exprimer le complexe L-P en utilisant de ce fait qu'un seul virus. Après fusion des deux vecteurs grâce à un clonage classique, la technique de sélection, purification du bacmide, transfection des cellules d'insectes, obtention de V0 et V1 sont identiques comme expliqués dans les sous-chapitre 3.4.3.

Les tests d'expression du complexe L-P sont réalisés comme décrit dans le sous-chapitre 3.4.3. L'amplification du virus V0 est faite dans 50 ml de milieu de culture (SF 900-Invitrogen). Les cellules sont centrifugées, V1 est collectées et les cellules sont lysées afin de procéder aux tests d'expressions en petite échelle. L'analyse de l'expression sur 50 ml de cellules d'insectes a montré des résultats prometteurs et ceci nous a mené à procéder à des expressions à plus grande échelle.

3.4.3 Expression du complexe L-P et purification sur colonne d'affinité Ni-NTA.

Les expressions à plus grande échelle ont été réalisées dans deux flasques de 400 ml de milieu de culture (SF900-Invitrogen) chacune avec des cellules d'insectes SF 21. L'infection avec le virus V1 a été réalisé comme décrit à 0.1 % (v/v). Après l'infection, pendant deux générations de 24 h chacune, les cellules d'insectes ont continué à se diviser. Aucune infection virale n'a été observée sous le microscope (cellules de forme petite) et le signal de YFP était absent. Après 48 h de post-infection, les cellules ont commencé à grossir (observer par microscope photonique) et la division s'est arrêtée. C'est ce qu'on appelle le DPA 0. C'est à ce moment que l'expression protéique commence à apparaître.

Dans la figure 3.5 B, page 93 on peut observer la faible augmentation de YFP à DPA 0. Toutes les 12 h, la YFP est suivie sur la même quantité de cellule. Dès que la YFP atteint un plateau, les cellules sont collectées comme décrit dans .

Le culot de cellules est resuspendu dans un tampon de 20 mM Tris pH 8.0, 800 mM NaCl, 5% glycérol 1 mM DTT et un cocktail de protéase (Roche) et les cellules sont lysées avec un potter sur glace en effectuant 20 coups de piston, doucement, sans créer de bulles.

Le lysat est centrifugé à 13 000 g pendant 1h afin de séparer la phase soluble de la phase non-soluble. La colonne d'affinité est équilibrée dans le même tampon que celui pour la lyse. La fraction soluble (SF) est passée par la colonne afin de fixer les protéines ayant une étiquette Histidine (L et P) mais également d'autres protéines se fixent de manière non-spécifique. Un lavage à haut sel (2 M NaCl) avec 20 mM imidazole permet de décrocher des acides nucléiques et des protéines fixées de manière non-spécifiques sur la colonne. Une réequilibration à 800 mM NaCl est réalisée et l'élution du complexe L-P est réalisée à 400 mM imidazole. Les fractions d'élutions sont rassemblées et concentrées par centrifugation à 2560 g dans un concentrateur 50 000 MWCO (Vivaspin-Sartorius), afin d'être purifiées par une étape supplémentaire de gel filtration (Superose-12).



FIGURE 3.6 – Purification du complexe L-P par chromatographie d'affinité Ni-NTA.

Sur la figure ne sont présentées que les fractions d'élutions. La polymérase est indiquée avec la lettre L. La phosphoprotéine est également indiquée avec la lettre P. Les deux protéines L-P possèdent une étiquette Histidine. Les protéines additionnelles sont également indiquées sur le gel. Les poids moléculaires sont indiqués à gauche du gel. La migration est effectuée sur un gel 10% SDS-PAGE (Laemmli).

La figure 3.6 présente les fractions d'élution de la colonne d'affinité Ni-NTA. On peut bien observer que la polymérase (L) migre à une taille supérieure de 170 kDa indiquant qu'elle est exprimée dans sa forme entière. La phosphoprotéine (P) migre à sa taille attendue entre 43 et 55 kDa.

Du fait que les deux protéines contiennent une étiquette Histidine, il est difficile de conclure à leurs interactions. Les protéines additionnelles sont exprimées et semblent co-éluées de façon stoechiométrique de la colonne d'affinité ce qui indique une interaction stable.

3.4.3.1 Identification des protéines co-éluées par spéctrométrie de masse.

Les bandes du gel d'acrylamide (voir figure 3.6) contenant les protéines co-éluées avec le complexe L-P en sortie de purification Ni-NTA ont été découpées et les protéines ont été analysées par spectrométrie de masse (EMBL, Heidelberg). Les protéines majoritaires, coéluées en quantité stoechiométrique avec le complexe L-P correspondent à HSP 70, une protéine de liaison à un domaine de polyA et EF 2. Une co-expression de HSP 70 avec la protéine d'intérêt suggère un problème de repliement. Chapitre 3. Expression, purification et recherche de domaines solubles de la polymérase 96 de VSV

3.4.4 Purification du complexe L-P par gel filtration et analyse par Western Blot.

La colonne Superose-12 est équilibrée au préalable dans un tampon identique à celui du tampon de lyse. Les fractions sont collectées et analysées par SDS-PAGE.



FIGURE 3.7 – Chromatographie d'exclusion de taille et analyse du complexe L-P par Western Blot.

A : Représente le chromatogramme du complexe L-P purifié en colonne Superose-12. 250 μ l de protéine sont injectés. La colonne est préalablement équilibrée avec 20 mM Tris pH 8.0, 800 mM NaCl, 5% glycérol et 1 mM DTT et une tablette d'antiprotéase (Roche). Le débit de la colonne est à 0.25 ml/min. Le premier pic sort de la colonne à un volume de 8.75 ml et le second à 11.43 ml. Une équilibration de la colonne est effectuée avec des protéines globulaires : 7.9 ml-669 kDa, 9.6 ml-440 kDa et 11 ml-150 kDa. Le trait rouge représente l'absorbance à 280 nm, et le trait bleu représente la conductivité (mSv) du système. Les pics I et II sont analysés en gel SDS-PAGE dans la figure B.

B : Représente le gel 10 % SDS-PAGE de la purification. Les échantillons sur le gel sont déposés comme suit : standard de poids moléculaire (MW), échantillon purifié sur colonne de Ni-NTA, et les deux pics d'élution de la colonne Superose-12. Les protéines additionnelles préalablement identifiées par Mass Spectrométrie sont montrées sur le gel.

C : Représente le Western Blot avec un anticorps reconnaissant l'étiquette Histidine (Sigma-Aldrich). Les échantillons sont déposés sur un gel 10 % SDS-PAGE et le transfert est réalisé sur une membrane de nitrocellulose pendant une heure à ampérage constant de 20 mA. Les deux protéines ayant une étiquette Histidine sont reconnues par l'anticorps. La révélation est effectuée avec un kit commerçant et le Western Blot est révélé par exposition sur un film photo. Trois bandes non spécifiques sont présentes lors de la révélation.

Lors de la purification en gel filtration les protéines additionnelles sortent de la colonne dans les deux pics d'élution (voir figure 3.7 Å). La polymérase est éluée une taille supérieure à 170 kDa, complexée à EF2 et HSP70 (voir figure 3.7 B, pic I), avant la phosphoprotéine (voir figure 3.7 B pic II). Vraisemblablement les protéines HSP70 et EF2 sont donc attachées à la polymérase seule. Ce résultat, de co-élution d'EF2 et de HSP70 avec L, a été également observé lors de l'expression de la polymérase seule, suggérant donc que ces deux protéines, dans la purification du complexe L-P, sont attachées à la polymérase.

L'analyse des résultats de chromatographie de filtration sur gel par Western Blot avec un anticorps α Histidine (Sigma Aldrich) (voir figure 3.7 C), montrent bien que les deux protéines (L et P) sont bien exprimées. La fin du pic I (voir figure 3.7 C) on observe que le signal par WB de L et de P se chevauche, ce qui pourrait nous faire penser que les deux protéines interagissent ensemble. Dans ce cas de figure il serait difficile de conclure sur une interaction possible car la phosphoprotéine est désordonnée est en chromatographie d'exclusion de taille s'élue comme une protéine ayant une masse moléculaire plus grande, donc le pic d'élution de P se chevauchera avec la fin du pic d'élution de L. Afin de conclure

pour une interaction des tests de « pull down » doivent être effectués.

Des anticorps polyclonaux α L nous ont été gracieusement fournit par Dr. Banerjee. Donc, j'ai effectué une analyse de la chromatographie par WB avec ces anticorps. Ceci nous a permis de conclure que les anticorps α Histidine reconnaissent les protéines L et P et sont donc, bien exprimées.

Différentes publications provenant du groupe de Banerjee ont déjà rapportés la présence de protéines cellulaires associées à la polymérase sans que leur implication dans les activités enzymatiques n'ait pu être déterminée. La polymérase purifiée à partir de cellules d'insectes est active et donc la présence de ces protéines cellulaires n'est pas requise (Rahmeh et al. 2010).

Dans les années 2004, par des techniques de co-immunoprécipation et détection par Western Blot, le groupe de Banerjee a mis en évidence une différence entre les complexes de transcription et réplication lorsque la polymérase est purifiée en cellules de mammifères (BHK). Le complexe de transcription semble être formé par des protéines accessoires comme HSP60, EF-1 α et une guanyltransférase (GT). Au contraire le complexe de réplication n'est composé que du complexe L-P et le complexe N°P qui aurait pour rôle d'encapsider l'ARN nouvellement formé (Qanungo et al. 2004).

Dans mon cas, lorsque j'exprime la polymérase en complexe avec P en cellules d'insectes, elle s'associe a des protéines accessoires de la même famille (EF2, HSP70 et à une protéine de liaison à une séquence polyAdénine).

3.4.5 Caractérisation structurale et fonctionnelle du complexe.

Malgré la présence de protéines additionnelles on a voulu visualiser le complexe en microscopie électronique mais il semble qu'il est très hétérogène et ne peut pas être analysé pour une reconstruction 3D.

Nous avons alors utilisé un protocole de pontage suivant la publication de (Kastner et al. 2008). Il s'agit de rigidifier un complexe protéique constitué de plusieurs protéines. J'ai essayé d'adapter le protocole à mon problème en centrifugeant le complexe dans un gradient de glycérol en présence de glutaraldéhyde afin de ponter et de séparer en même temps les espèces de très haut poids moléculaire. Une analyse de chaque fraction du gradient a été effectuée par microscopie électronique, mais les résultats n'ont pas été prometteurs et j'ai abandonné cette technique.

Enfin j'ai voulu savoir si la polymérase exprimée en cellules d'insectes avec son cofacteur, même en présence de protéines additionnelles, est active. En collaboration avec Danielle BLONDEL (Gif sur Yvette, Paris) j'ai effectué des tests de transcription *in vitro*, comme décrit dans le papier intitulé « Structure of the Vesicular Stomatitis Virus L subunit of the RNA-dependent RNA polymerase by Electron Microscopy ». Après analyse des échantillons de transcription il s'est avéré que la polymérase ne possède aucune activité enzymatique. On peut donc conclure à un problème de repliement qui est cohérent avec l'expression de façon stoechiométrique de HSP70.

3.5 Conclusions

Dans ce chapitre j'ai présenté les difficultés que nous avons rencontrées lors des essais de purification de différentes constructions de la polymérase exprimées en bactéries et en cellules d'insectes avec des étiquettes différentes (histidine et MBP). Une des difficultés rencontrée réside dans le fait que la polymérase est associée à des protéines additionnelles que l'on peut difficilement éliminer (voir figure 3.6 page 95).

Pendant très longtemps j'ai réfléchi sur le problème rencontré et les échecs de purification des domaines seuls ou du complexe L-P. Récemment, la publication de Rahmeh et al. (2010) montre que la polymérase de VSV (souche Indiana) est exprimée en cellules d'insectes (SF21, monocouche) sans protéines additionnelles accrochées de manière stoechiométrique. Le gène utilisé dans notre travail pour l'expression de la polymérase en cellules d'insectes a été optimisé afin d'être mieux exprimé dans un système procaryote. La polymérase est une protéine complexe, avec plusieurs domaines, qui peut nécessiter lors de la traduction que le ribosome fasse des pauses afin de permettre le bon repliement des différents domaines. Le problème rencontré pourrait venir de l'optimisation des codons qui permettrait au ribosome de traduire plus rapidement la séquence d'ARNm sans faire de pauses ce qui conduirait à un mauvais repliement des domaines. Des protéines chaperonnes comme HSP70 en cellules d'insectes (figure 3.6 page 95) ou GroEL/GroES en bactéries (figure 3.3 page 90) peuvent alors se complexer à des parties hydrophobes afin d'empêcher une agrégation dans la cellule.

Chapitre 4

Structure of the vesicular stomatitis virus L subunit of the RNA-dependent RNA polymerase by electron microscopy

Article III : Structure of the vesicular stomatitis virus L subunit of the RNA-dependent RNA polymerase by electron microscopy

<u>Auteurs</u> : **Ivan Ivanov**^{1,2}, Danielle Blondel³, Julien Perard¹, Delphine Guilligay¹, Thibaut Crépin¹, Guy Schoehn¹, Rob W. H. Ruigrok¹ and Marc Jamin^{1*}

PAPER IN PREPARATION

Structure of the vesicular stomatitis virus L subunit of the RNA-dependent RNA polymerase by electron microscopy

Ivan Ivanov^{1,2}, Danielle Blondel³, Julien Pérard¹, Delphine Guilligay¹, Thibaut Crépin¹, Guy Schoehn¹, Rob W. H. Ruigrok¹ and Marc Jamin^{1*}

¹ UMI3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions

6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

² Institut Laue Langevin, 6 Rue Jules Horowitz, BP 156, 38042 Grenoble Cedex 9, France

³UPR 3296 CNRS, Virologie Moléculaire et Structurale, Gif sur Yvette, France.

Running title: Structure of VSV polymerase

Keywords: Vesicular stomatitis virus, Negative sense RNA virus, polymerase, activity, electron microscopy

* **Corresponding author:** Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France

E-mail : jamin@embl.fr, Phone: + 33 4 76 20 94 62, Fax : + 33 4 76 20 94 00

Introduction

Vesicular stomatitis virus is a prototypic member of the non-segmented negative sense RNA viruses. These viruses were grouped within the *Mononeqavirales* (MNV) because they share similar organizations of their genome and virion and similar mechanisms of replication. They include numerous human pathogens, such as rabies virus, measles virus, respiratory syncytial virus, Ebola virus and Nipah virus. Their genome is entirely coated by the viral nucleoprotein (N), forming a long helical N-RNA complex. This N-RNA complex, rather than naked RNA, is used serves as a template for both transcription and replication by a two-subunit viral RNA polymerase. The viral RNA-dependent RNA polymerase consists of a large subunit L (2109 aa in VSV) and a noncatalytic phosphoprotein (P). The P protein of VSV forms homodimers and is multi-functional protein made of concatenated disordered regions and structure domains. P has the ability to bind to both the N-RNA complex through its C-terminal domain and to the L subunit through its N-terminal region suggesting that the role of P in the polymerase complex is to position L onto its template. The L subunit of VSV catalyzes the template-based RNA synthesis, but also the capping of the mRNA, the methylation of the cap and the poly-adenylation of the mRNA. Amino acid sequence comparisons revealed the presence of six conserved regions shared by all MNV viruses (CR I-VI) (Figure 1). The CR III contains the active site for the RNA polymerase activity and is also required for the polyadenylation of the mRNA, which occurs by slippage of the polymerase. The CR V is responsible for the capping of the mRNA, whereas the CR VI catalyze the methylation of the cap. The mechanism of mRNA capping used by VSV polymerase is different than the mechanism found in eukaryotic cells.

1								2109	
	l 220-410	اا 491-594	III 595-816		IV 874-1045	V 1076-1307	16 <mark>37-17</mark> 12		
		RNA	polymeriza	tior	1	cap addition	cap methylation		

Illustration 1: **Conserved regions of VSV RNA dependent RNA polymerase.** The yellow boxes indicate the localization of the six conserved regions (CR I-VI) identified in VSV L by sequence alignment with the L proteins of other MNV viruses. The activities of RNA polymerization, cap addition and cap methylation were mapped to CR-III, CR-V and CR-VI, respectively.

Recently, the analysis of VSV L protein and some of its fragments by electron microscopy revealed the organization of the protein into a ring domain containing the RNA polymerase and three globular adjunct domains containing the capping and methylation activities. However, the structural variability precluded a three-dimensional reconstruction. In this paper, we have purified the L subunit directly from VSV and we report the three-dimensional reconstruction of the protein.

Results

Purification of L from virions

The L protein was extracted and purified from the virus. BSR cells were infected with the VSV Indiana serotype of the Mudd-Summers strain and viral progenies were harvested 24 h postinfection. The viral particles were disrupted and the ribonucleoprotein complex, containing the N-RNA complex and the L and the P proteins, was extracted by centrifugation through a glycerol step gradient (Figure 2A). The L/P complex was separated from the N-RNA template by centrifugation through a glycerol cushion in the presence of 250 mM 3-(1-Pyridino)-1-propane sulfonate (NDSB-201), and the L protein was purified by elution through a phosphocellulose C11 column.



Illustration 2: Purification of the viral L protein. Proteins were analyzed by 10% SDS/PAGE and stained with Coomassie blue. Lane 1: Molecular weight markers. Lane 2: Virus solubilized (SV) in 2% Triton X100. Lane 3: Ribonucleoprotein complex (RNP) collected after centrifugation through a glycerol step gradient. Lane 4: Molecular weight markers. Lane 5: L/P complex collected after centrifugation through a glycerol cushion in the presence of 250 mM 3-(1-Pyridino)-1-propane sulfonate. Lane 6: L protein after purification with a phosphocellulose column.

In vitro reconstitution assay showed that the purified L protein was catalytically functional in transcription. When combined with recombinant P protein and with N–RNA template purified from infected cells, the L protein catalyzed the synthesis of the five viral mRNAs (Figure 3A). These results demonstrate that the L protein is active in transcription and requires no cellular factors in stoichiometric amount. However, to for the requirement of cellular factors, we compared the transcriptional activity of the polymerase in the absence and presence of reticulocyte lysate (Figure 3B). The RNA synthesis activity is only slightly enhanced in the presence of cellular extract, confirming that the purified L, P and N-RNA components are sufficient for efficient RNA synthesis but suggesting that some cellular components may contribute to the activity by stabilizing the complex or by regulating its activity.



Illustration 3: **Functional assay of the purified L protein.** (A) Reconstitution of RNA transcription from purified L, P and N-RNA complex. Transcription reactions were performed in the presence of $[\alpha^{-32}P]$ UTP and the products were analyzed by electrophoresis on a 4% acrylamide gel containing 7 M urea. The identity of the different mRNAs is shown at the right. (B) Effect of rabbit reticulocyte lysate on transcriptional activity. Reactions were performed as in A in the presence of 25 µg.mL⁻¹ actinomycin D. RNA synthesis was measured with a scintillation counter.

Structure of the purified L protein

Figure shows a typical electron micrograph of the purified L subunit. The ring-like structure of the RNA polymerase domain is clearly visible.



Illustration 4: **Electron micrograph of purified VSV L protein.** Negative staining was performed with uranyl acetate. Protein concentration was 0.05 mg.ml⁻¹.

From the electron micrographs, images of 10,300 particles were collected and classified into 6*13 different class averages shown in figure 5A. The class averages indicated a compact shape and confirmed the overall organization of L, which comprises a large core ring-like domain associated with smaller globular domains.



Illustration 5: Three-dimensional reconstruction of VSV L protein. (A) Class averages of L. 10,300 images of individual particles were selected and classified into 6 different classes. The pixel dimension is 3Å. Each box has a size of 99*99 pixels. Each line represents members of one class. (B). **Image processing and three-dimensional reconstruction**. The first line of the figure shows 10 different single particles picked from the micrographs. The second line shows class averages after background subtraction. The third line shows the 2-D projection of each particle. A preliminary reconstruction of a single polymerase is shown in the black box. The best 5,000 particles were selected for the reconstruction. The resolution is estimated at 30 Å.

Analysis of the class averages of L (figure 5A) suggested that the polymerase from viral particles is a rigid body with additional domains attached to the core of the polymerase. The ring is approximately 6.5 nm in diameter. The distance between the center of the ring and the globular domains is around 7 nm and the diameter of a globular domain is around 3.7 nm. Most parts of the particles into the class averages were homogeneous and were used to build a three-dimensional reconstruction using another polymerase as an initial model. Figure 5B shows the reconstruction in different orientations.

Discussion

This structural characterization of VSV L protein confirms the presence of a ring-like domain typical of the other RNA-dependent RNA polymerases. In these polymerases, the central channel accommodates the RNA template. In accordance, for VSV, the RNA polymerase activity was previously mapped to this domain.

Three associated globular domains were shown to contain the CR V and VI, and therefore, the capping and methylase activities. In our reconstruction, these three domains are more closely packed onto the core of the polymerase than previously reported. We have no explanation for this difference, which may also explain the reduction in structural variability of the molecule, which allowed us to build a three-dimensional reconstruction. In this study, the L protein was folded and assembled in an infected cell and packaged into a virion. It is possible that the viral L protein has post-translational modification such as phosphorylations that allows it to adopt a more compact fold.

Materials and Methods

Purification of L protein from virions

BSR cells at 90% confluence were infected with VSV (Indiana serotype, Mudd Summers strain) at multiplicity of infection 0.3. The medium was collected 24h post-infection and filtered through a 0.3 µm filter (Millipore) to discard cell debris. The suspension was centrifuged for 150 min at 21,000 g to pellet the viral particles. The protocol for the purification of the L protein was adapted from that previously described (Das et al. 1996). The pellet was suspended in 4 mL buffer containing 20 mM Tris-HCl pH 8.0 and the suspension was centrifuged for 90 min at 274,000 g through a 30% glycerol cushion in 20 mM Tris-HCl buffer at pH 8.0 containing 150 mM NaCl and 1mM EDTA. The purified VSV was disrupted by incubation in 20 mM Tris-HCl buffer at pH 8.0 containing 400 mM NaCl, 1 mM DTT and 2% Triton X-100 at a final concentration near 1 mg.ml⁻¹ for 3hr at 25°C under continuous agitation on a mixing wheel. The solution was loaded onto a glycerol step gradient composed of 4.5 ml of 30% (v/v) glycerol 20 mM Tris-HCl buffer at pH 8.0 containing 400 mM NaCl and 1 mM DTT and 0.5 ml of 100% glycerol, and was centrifuged at 250,000 g for 1.5 hr at 4°C. The visible RNP band was collected in the smallest possible volume from the top of the 100% glycerol cushion. The collected fraction was diluted fivefold with 20 mM Tris-HCl buffer at pH 8.0 containing 1 mM EDTA, and then the solution was diluted twofold with 40 mM sodium phosphate buffer at pH 7.0 containing 2 M NaCl, 4 mM DTT and 500 mM mM 3-(1-Pyridino)-1-propane sulfonate (NDSB-201). The solution was incubated at 4°C for 4 hr under continuous agitation on a mixing wheel. Then, the solution was centrifuged through a 30% glycerol cushion for 5 hr at 274,000 g in 25 mM sodium phosphate buffer at pH 7.0 containing 300 mM NaCl, 1 mM DTT and 250 mM NDSB-201. The released LP complex was collected from the top of the 30% glycerol and was dialyzed for 12 hours against 1 L of 25 mM sodium phosphate buffer at pH 7.0 containing 200 mM NaCl, 1 mM DTT, 250 mM NDSB-201 and 5% glycerol with two changes. The dialyzed fraction was loaded onto a phosphocellulose (P11, Whatman) column (h = 1.1 cm, ϕ = 1 cm) equilibrated with the dialysis buffer. The column was then washed with five volumes of 25 mM sodium phosphate buffer at pH 7.0 containing 200 mM NaCl, 1 mM DTT, 250 mM NDSB-201 and 2% glycerol and the L protein was eluted in 25 mM sodium phosphate buffer at pH 7.0 containing 1 M NaCl, 1 mM DTT, 250 mM NDSB-201 and 2% glycerol. The protein was concentrated with centrifugal concentrators (Vivaspin 500 µL, cut-off 5 kDa) and dialyzed against 25 mM sodium phosphate buffer at pH 7.0 containing 500 mM NaCl, 1 mM DTT and 250 mM NDSB-201.

Viral N-RNA preparation

A lysate of VSV infected BSR cells was obtained by three cycles of freezing/thawing in liquid nitrogen in 20 mM Tris-HCl buffer at pH 7.5 containing 300 mM NaCl and 1 mM DTT. The cell lysate was centrifuged at 11,200g for 30 min. The supernatant was loaded onto a 20-40% CsCl gradient prepared in the same buffer and centrifuged at 274,000g for 16h at 12°C. The visible white band of N-RNA complex was collected from the tube in the smallest possible volume and was dialysed overnight against a 20 mM Tris-HCl buffer at pH 7.5 containing 150 mM NaCl, 1 mM DTT with several changes. The concentration was measured by a Bradford assay. The N-RNA template was then stored at -80°C without lost of activity.

RNA transcription activity assay

The transcription reaction was performed in a final volume of 40 µl in Tris-HCl pH 7.5, 150 mM NaCl, 1mM DTT, 1mM MgCl2, 0.01 % Triton-X100 and 1mM Spermidine, containing 1 [α^{32} P]UTP (250Ci/mmol) (Perkin-Elmer), 4 mM ATP, 4 mM GTP, 0.5 mM UTP, and 1 mM CTP, 1µg of viral N-RNA template, 1µg of purified L, 1 µg of P expressed in bacteria, 1 µL of RNAase inhibitors (40U/µl RNAzine). The transcription reaction was performed at 30°C for 3 hrs. After completion of the reaction an equal volume of RNA blue was added, and the sample was heated at 95°C during 1 min and loaded on a 4% acrylamid urea gel. The gel was run for 3hrs at constant power of 25 W. The gel was then exposed to an Amersham screen for 30 min. The picture of the photo was taken in a PhosphoImager machine (GE Healthcare, Typhoon).

For transcription tests with the present of 30% rabbit reticulocyte lisate, 25µg/ml Actinomycin D was added.

Electron Microscopy and 3D reconstruction

Four microliters of the protein sample at approximatively 0.05 mg.mL⁻¹ was adsorbed onto the clean face of a carbon film on a mica sheet (carbon/mica interface) and negatively stained with 2 % (w/v) uranyl acetate. Micrographs were taken under low-dose conditions with a FEI Polara electron microscope working at 300 kV and with a nominal magnification of 30,000X using an USC 4000 GATAN CCD camera (3.0A/pixel at the specimen level).

The image analysis procedure was started by selecting 10,300 particles in 100*100 pixel² from 26 micrographs by using boxer (Ludtke2010). The contrast transfer function was estimated using

CTFfind (Mindell2003) and the images corrected for the phases using bctf (Heymann2001). A starting model was created using one centered raw image which was back projected in a 100*100*100 box and the height was limited to 1.2 times the apparent diameter of the particle. The 3D starting model was then used to run a projection matching procedure as described in (Franzetti2002a). A further 100 cycles of refinement was performed using 317 reprojections (8 degrees step sampling). For the isosurface representation, we included the correct molecular weight of the complex using an average protein density of 0.84 Da/A³. To improve the quality of the 3D reconstruction we need to add more images (in progress).

REFERENCES

- Abraham, G., D. P. Rhodes, et al. (1975). "The 5' terminal structure of the methylated mRNA synthesized in vitro by vesicular stomatitis virus." <u>Cell</u> 5(1): 51-58.
- Arnheiter, H., N. L. Davis, et al. (1985). "Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis." <u>Cell</u> **41**(1): 259-267.
- Baltimore, D., A. S. Huang, et al. (1970). "Ribonucleic acid synthesis of vesicular stomatitis virus, II. An RNA polymerase in the virion." <u>Proc Natl Acad Sci U S A</u> **66**(2): 572-576.
- Barr, J. N. and G. W. Wertz (2001). "Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination." J Virol **75**(15): 6901-6913.
- Barr, J. N., S. P. Whelan, et al. (1997). "cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation." J Virol **71**(11): 8718-8725.
- Emerson, S. U. and M. Schubert (1987). "Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus." <u>Proc Natl Acad Sci U S A</u> **84**(16): 5655-5659.
- Emerson, S. U. and Y. Yu (1975). "Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus." <u>J Virol</u> **15**(6): 1348-1356.
- Ferrer-Orta, C., A. Arias, et al. (2006). "A comparison of viral RNA-dependent RNA polymerases." <u>Curr Opin Struct Biol</u> **16**(1): 27-34.
- Franzetti, B et al. (2002). "Tetrahedral aminopeptidase: a novel large protease complex from archaea." <u>EMBO J.</u> **21**: 2132-2138.
- Gérard, F. C. A., E. Ribeiro, et al. (2007). "Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution." <u>Biochemistry</u> **46**(36): 10328-10338.
- Gérard, F. C. A., E. A. Ribeiro , et al. (2009). "Modular organization of rabies virus phosphoprotein." J. Mol. Biol. **388**(5): 978-996.
- Grdzelishvili, V. Z., S. Smallwood, et al. (2005). "A single amino acid change in the L-polymerase protein of vesicular stomatitis virus completely abolishes viral mRNA cap methylation." J <u>Virol</u> **79**(12): 7327-7337.
- Grdzelishvili, V. Z., S. Smallwood, et al. (2006). "Identification of a new region in the vesicular stomatitis virus L polymerase protein which is essential for mRNA cap methylation." <u>Virology</u> **350**(2): 394-405.
- Hercyk, N., S. M. Horikami, et al. (1988). "The vesicular stomatitis virus L protein possesses the mRNA methyltransferase activities." <u>Virology</u> **163**(1): 222-225.
- Heymann, J.B. (2001). "Bsoft: image and molecular processing in electron microscopy." J. <u>Struct.Biol.</u> **133**: 156-169.
- Ivanov, I., F. Yabukarski, et al. (2011). "Structural insights into the rhabdovirus transcription/replication complex." <u>Virus Res.</u>
- Leyrat, C., E. A. Ribeiro , et al. (2011). "Structure, interactions with host-cell and and functions of rhabdovirus phosphoprotein." <u>Future Virol.</u> **6**(4): 465-481.
- Li, J., E. C. Fontaine-Rodriguez, et al. (2005). "Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity." J Virol **79**(21): 13373-13384.
- Li, J., A. Rahmeh, et al. (2008). "A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping." <u>J Virol</u> **82**(2): 775-784.
- Ludtke, S.J. (2010). "3D structures of macromolecules using single particle analysis in EMAN."

Methods Mol. Biol. 673: 157-173.

- Mellon, M. G. and S. U. Emerson (1978). "Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus." J Virol 27(3): 560-567.
- Moyer, S. A. and A. K. Banerjee (1975). "Messenger RNA species synthesized in vitro by the virion-associated RNA polymerase of vesicular stomatitis virus." <u>Cell</u> **4**(1): 37-43.
- Mindell, J.A. et al. (2003). "Accurate determination of local defocus and specimen tilt in electron microscopy." J. Struct. Biol. **142**: 334-347.
- Ogino, T. and A. K. Banerjee (2007). "Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus." <u>Mol Cell</u> **25**(1): 85-97.
- Poch, O., B. M. Blumberg, et al. (1990). "Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains." J Gen Virol **71** (**Pt 5**): 1153-1162.
- Pringle, C. R. (1997). "The order Mononegavirales-current status." <u>Arch Virol</u> **142**(11): 2321-2326.
- Rahmeh, A. A., J. Li, et al. (2009). "Ribose 2'-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein." J Virol **83**(21): 11043-11050.
- Rahmeh, A. A., A. D. Schenk, et al. (2010). "Molecular architecture of the vesicular stomatitis virus RNA polymerase." <u>Proc Natl Acad Sci U S A</u> **107**(46): 20075-20080.
- Sleat, D. E. and A. K. Banerjee (1993). "Transcriptional activity and mutational analysis of recombinant vesicular stomatitis virus RNA polymerase." J Virol **67**(3): 1334-1339.

Chapitre 5

CONCLUSIONS & PERSPECTIVES

Sommaire

5.1	Organisation modulaire de la phosphoprotéine				
5.2	La grande sous-unité (L) du complexe polymérasique (L-P)				
5.3	Fixation de la phosphoprotéine à la matrice N-ARN et le mécanisme de				
	transcription virale proposé	116			
5.4	Fixation du complexe N°P sur la matrice N-ARN et le mécanisme de				
	réplication proposé	119			

5.1 Organisation modulaire de la phosphoprotéine

Des prédictions bioinformatiques à partir de la séquence en acides aminés, combinées à la caractérisation biophysique et structurale des domaines structurés ont montré que la phosphoprotéine des *Rhabdoviridae* possède une organisation structurale conservée. Deux domaines capables de se replier de manière autonome (P_{CTD} et P_{CED}) sont connectés l'un à l'autre par une région désordonnée IDR_{CTD}.

Avant mon arrivé dans le laboratoire, le groupe disposait de la structure du domaine C-terminal de la phosphoprotéine (P) du virus de la stomatite vésiculaire obtenue par RMN (Ribeiro et al. 2008) et de celle du domaine C-terminal de P du virus de la rage (RAV), obtenue par cristallographie (Mavrakis et al. 2004). Malgré la faible identité de séquence ces deux structures sont similaires et toutes les deux portent la fonction de liaison à la matrice N-ARN.

Dans un autre laboratoire, le domaine central de la phosphoprotéine du virus de la stomatite vésuculaire (VSV) avait été résolu par cristallographie révélant que la phosphoprotéine est capable de former des dimères parallèles par son domaine central. Des mesures de diffusion de lumière couplée à un réfractomètre et une colonne d'exclusion de taille indiquaient que les deux phosphoprotéines (VSV et RAV) sont dimériques en solution.

Une partie de mon travail de thèse a consisté dans le clonage, l'expression et la purification du domaine central de la phosphoprotéine de RAV. Afin de déterminer la région à cloner nous avons utilisé les analyses de D-score présentées dans la publication de Gerard et al. (2009).

De par sa petite taille de 45 acides aminés, nous avons utilisé deux approches en parallèle : la Résonnance Magnétique Nucléaire (RMN), appliquée dans la plupart des cas pour des petites protéines et la cristallographie.

- Malheureusement, la RMN n'a pas donné de résultats satisfaisant car la protéine avait tendance à former des aggrégats solubles empêchant l'acquisition des données.
- La cristallographie présentait également une part de travail non négligeable car les cristaux poussaient très lentement (3-4 mois).

Après avoir obtenu les cristaux, nous avons utilisé plusieurs techniques pour la résolution de la structure. Comme expliquer dans le chapitre 2, la technique portant ses fruits s'est avérée être le phasage par la méthode de Sulfur-MAD.

La structure dimérique du domaine central de P de RAV est formée de deux hélices α par monomère interagissant avec l'autre monomère via des contacts hydrophobes (Ivanov et al. 2010). La structure révèle une organisation parallèle de la phosphoprotéine du virus de la rage, avec les domaines N-terminaux et C-terminaux très éloignés l'un de l'autre.

Cette structure semble être différente du domaine central de la P de VSV. Lors de l'évolution des virus, cette partie de la protéine a gardé sa fonction de dimérisation mais les deux structures ont divérgé.

Par des expériences d'analyses de désordre, cristallographie, RMN, diffusion de lumière statique, SAXS, SANS et dynamique moléculaire nous pouvons conclure que la phosphoprotéine possède trois parties fonctionnelles (P_{NTD} , P_{CED} , P_{CTD}) impliquées respectivement dans la fixation à N° (Leyrat et al. 2010) et à L (Emerson & Schubert 1987), à l'homodimérisation de la protéine (Ding et al. 2006, Ivanov et al. 2010), et à la fixation aux complexes N-ARN (Mavrakis et al. 2004, Ribeiro et al. 2008).

Au cours de ma thèse j'ai étudié également de la région 71-105 de la phosphoprotéine du virus de la stomatite vésiculaire, "candidat potentiel", responsable de la liaison à la polymérase (L). Le spectre ¹H ¹⁵N-HSQC par RMN révèle l'absence de structure secondaire.

Dans la publication de Canter & Perrault (1996), la région C-terminal de la polymérase (1595-2109) est décrite comme la région interagissant avec la phosphoprotéine.

Récemment nous avons obtenu des résultats préliminaires d'interaction (pull down) entre le domaine C-terminal de L (M1595-D2109) et le domaine N-terminal de la phosphoprotéine (1-105). Comme perspective, dans les mois à venir nous envisageons de déterminer plus exactement la région de liaison du C-terminal de L sur le N-terminal de P par RMN. Ensuite différents mutants seront utilisés afin de prouver le site d'interaction, et des mutants mimant des sites de phosphorylation seront utilisés pour investiguer les changements d'affinité entre les deux partenaires. Ceci dans le but de comprendre la régulation de la transcription et de la réplication par le phénomène de phosphorylation de la phosphoprotéine. Enfin, nous envisageons d'essayer de cristalliser le domaine C-terminal de L en complexe avec N-terminal de P.

5.2 La grande sous-unité (L) du complexe polymérasique (L-P)

Le deuxième but principal de ma thèse consistait à obtenir une information structurale sur l'ARN polymérase ARN dépendante. Cette protéine de 2109 acides aminés possède plusieurs activités enzymatiques décrites en détail dans le chapitre 1.

Dans le laboratoire aucun travail préliminaire n'existait sur la polymérase virale de VSV. J'ai suivi plusieurs stratégies en parallèle. L'expression de différentes constructions polymérasiques en bactéries et cellules d'insectes, l'expression de la polymérase entière mais également du complexe polymérasephosphoprotéine en cellules d'insectes. Deux constructions de la polymérase ont été identifiées qui s'expriment en bactéries et se purifient correctement.

Des expériences de cristallographie avec le robot cartésien à l'EMBL de Grenoble n'ont pas permis d'obtenir des cristaux de ces domaines. Comme perspectives, nous envisageons de cristalliser ce domaine en complexe avec la région de P se liant à L.

L'expression de la polymérase seule ou en complexe avec la phosphoprotéine n'a donné aucun résultat satisfaisant. Par analyses de FingerPrint (EMBL, Heidelberg), nous avons conclu que des protéines cellulaires comme HSP70, EF2 et une protéine de liaison à un domaine polyAdénine sont accrochées à la polymérase. Ces résultats en accord avec des données obtenues par un autre groupe (Qanungo et al. 2004), suggèrent l'implication de protéines cellulaires. Nous avons observé le complexe L-P-HSP70-EF2polyA binding protein en microscopie électronique mais il s'est avéré être très hétérogène et n'a pas pu être analysé par la suite.

Comme dernière tentative de purification de la polymérase entière (L) nous avons utilisé le virus entier (Dr. Danielle Blondel (Gif sur Yvettes)). J'ai procédé à l'optimisation du protocole de purification présenté dans la publication Das et al. (1996). J'ai pu obtenir une quantité infime de polymérase ($50\mu g$) suffisante pour procéder à l'analyse par microscopie électronique. Des clichés de coloration négative ont été enrégistré par le Dr. Guy Schoehn (IBS, Grenoble) comme expliqué dans le chapitre 4. La reconstruction préliminaire révèle la présence d'un domaine annulaire avec des domaines globulaires attachés à l'anneau. D'après le modèle obtenu, la polymérase semble être compacte et rigide.

En perspectives, afin d'améliorer le modèle préliminaire à 30 Å, nous envisageons d'utiliser la tomographie pour générer un modèle initial de meilleur qualité qui nous servira ensuite pour obtenir un modèle final à une meilleure résolution. Egalement des clichés supplémentaires de coloration négative ou de cryo-microscopie doivent être enrégistrés.

Afin de prouver si la polymérase extraite de virus entier est active j'ai réalisé des tests d'activité enzymatique par autoradiographie. Comme substrat de la polymérase nous avons utilisé le complexe N-ARN viral, purifié par un gradient de CsCl à partir de cellules de mammifères (BSR). Le cofacteur (P) a été produit en système d'expression bactérien (BL21). Des nucléotides de transcription sont utilisés, mélangés avec [³²P- α UTP] radioactif. Les transcrits d'ARN messager sont observés sur un gel dénaturant.

Par cette expérience *in vitro*, nous avons démontré qu'il est possible de reconstituer le système de transcription virale à partir de composants isolés et hautement purifiés.

5.3 Fixation de la phosphoprotéine à la matrice N-ARN et le mécanisme de transcription virale proposé

Les domaines C-terminaux du virus de la rage et du VSV responsables de la liaison à N-ARN sont homologues. Au cours de l'évolution des virus, ce domaine, essentiel à l'activité de transcription des virus, a gardé une structure tertiaire identique parmi les Rhabdovirus et les Paramyxovirus.

5.3. Fixation de la phosphoprotéine à la matrice N-ARN et le mécanisme de transcription virale proposé

Bien que le domaine C-terminal des Rhabdovirus et des Paramyxovirus, permette une fixation sur le complexe N-ARN, la constante de dissociation pour la matrice n'est pas la même dans les deux familles et pourrait expliquer la tétramérisation de la phosphoprotéine de Paramyxovirus. L'affinité plus faible nécessiterait la présence de quatre monomères au lieu de deux pour les Rhabdovirus.

Chez VSV et RAV, le domaine P_{CTD} interagit simultanément avec deux protomères de nucléoprotéine qui forment une série de liaison intermoléculaires avec les boucles C-terminales désordonnées de N. Des modèles moléculaires réalisés dans le laboratoire montrent que le reste de la phosphoprotéine est orienté en direction de la face interne de N-ARN, suggérant que la région P_{NTD} est localisée à proximité de la région centrale de fixation à l'ARN. Ainsi orienté la phosphoprotéine peut positionner la polymérase près de son substrat pour procéder aux processus de transcription/réplication.

Nous proposons un modèle de transcription (voir figure 5.1 à gauche page 118) qui se différencie du modèle proposé par Curran (1998) pour les *Paramyxoviridae*.

Dans ce modèle, la phosphoprotéine est constamment attachée sur la matrice N-ARN, avec une stoechiométrie de deux phosphoprotéines pour cinq nucléoprotéines. L'affinité pour la matrice est forte de l'ordre du nM (unpublished data, Dr. Euripedes Ribeiro). Des modélisations de dynamique moléculaire montrent que la phosphoprotéine ainsi attachée est mobile et occupe un espace conformationnel de 5nm de rayon. La polymérase virale pourrait ainsi effectuer des "sauts" d'une phosphoprotéine à une autre et procéder à la transcription virale.

Sur base de la structure du complexe N°P (Leyrat, Yabukarski, Tarbouriech, Ribeiro, Jensen, Blackledge, Ruigrok & Jamin 2011), nous avons proposé que le bras N-terminal de la phosphoprotéine dans le complexe L-P peut ouvrir la nucléocapside par compétition pour le site d'interaction avec l'ARN. La polymérase accrochée dans la région IDR_{NTD} de P sera à proximité de l'ARN viral et pourra commencer le processus de transcription.

Cette hypothèse de déplacement de l'ARN par le bras N-terminal de P nécessite une validation par des expériences suppélementaires.

Actuellement notre modèle de transcription n'explique pas comment la polymérase avance le long de la capside virale, mais nous avons émis l'hypothèse du mécanisme d'avancement par "sauts". Une structure cristallographique du complexe L-P pourra répondre à certaines questions afin de mieux comprendre le processus de transcription et réplication (figure 5.1 page 118).





Le substrat N-ARN est présenté en vert dans le sens 5'-3' (ARN de polarité négative). La phosphoprotéine par son domaine C-terminal (rouge) est accrochée à la nucléocapside. Une phosphoprotéine est représentée sur l'extrémité 3' qui fait compétition pour le site d'interaction entre la N et l'ARN via son bras N-terminal et libère l'ARN viral. La polymérase (jaune) en complexe avec le cofacteur P commence à transcrire l'ARN à partir du premier nucléotide.

La figure de réplication de droite est présentée avec les mêmes couleurs. Un echange entre le complexe N°P et la phosphoprotéine est effectué d'encapsider l'ARN de polarité positive nouvellement formé. La phosphoprotéine déplacée par N°P est indiquée sur la figure.

Fixation du complexe N°P sur la matrice N-ARN 5.4et le mécanisme de réplication proposé

Les mécanismes de la transcription et de la réplication du virus de la stomatite vésiculaire demeurent à l'heure actuelle mal compris. Les avancées effectuées dans les deux dernières années nous ont permis de comprendre comment la phosphoprotéine protège la nucléoprotéine et la garde sous forme soluble et monomérique. Ceci étant indispensable pour l'encapsidation de nouveaux génomes viraux.

Les connaissances acquises nous poussent à penser que le basculement entre les processus de transcription et réplication doit se faire en partie par une augmentation de la quantité du complexe N°P dans la cellule infectée. Cette augmentation de la concentration de N°P, couplée à un processus de régulation par phosphorylation de la phosphoprotéine pourront déclencher le processus de réplication virale. Il est évident qu'il doit y avoir suffisamment de molécules de nucléoprotéines sous forme monomérique afin de pouvoir encapsider tout le génome viral. Donc, après infection, le virus doit avoir suffisamment de polymérases, et phosphoprotéines afin de pouvoir transcrire le génome viral pour fabriquer le complexe N°P qui servira par la suite pour l'encapsidation du génome de polarité positive et négative.

Sur la figure 5.1 à droite page 118 est présenté le modèle du mécanisme de réplication viral. Le complexe N°P est produit à un taux suffisant dans la cellule, il déplace la phosphoprotéine attachée sur N-ARN. Ainsi la polymérase pourra utiliser la nucléoprotéine nouvellement formée comme substrat pour l'encapsidation de l'ARN de polarité positive et négative. Ce modèle ainsi que le modèle de transcription n'expliquent pas comment la polymérase se déplace le long de la nucléocapside. Dans le processus de réplication le mécanisme de « sauts » pourrait faire intervenir un autre complexe, pas mentionné jusqu'à présent, le complexe N°P-L (site de fixation à L disponible dans le complexe N°P).

Bibliographie

- Abraham, G. & Banerjee, A. K. (1976), 'Sequential transcription of the genes of vesicular stomatitis virus.', Proc Natl Acad Sci U S A 73(5), 1504–1508.
- Abraham, G., Rhodes, D. P. & Banerjee, A. K. (1975), 'The 5' terminal structure of the methylated mrna synthesized in vitro by vesicular stomatitis virus.', Cell 5(1), 51–58.
- Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K. & Miyano, M. (1999), 'Crystal structure of the rna-dependent rna polymerase of hepatitis c virus.', *Structure* 7(11), 1417–1426.
- Albertini, A. A. V., Clapier, C. R., Wernimont, A. K., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. H. (2007), 'Isolation and crystallization of a unique size category of recombinant rabies virus nucleoprotein-rna rings.', J Struct Biol 158(1), 129–133. URL: http://dx.doi.org/10.1016/j.jsb.2006.10.011
- Albertini, A. A. V., Schoehn, G. & Ruigrok, R. W. H. (2005), 'Structures impliquées dans la réplication et le transcription des virus à arn non semgmenté de sens négatif', *Virologie* **9**(2), 83–92.
- Albertini, A. A. V., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. H. (2008), 'Structural aspects of rabies virus replication.', Cell Mol Life Sci 65(2), 282–294. URL: http://dx.doi.org/10.1007/s00018-007-7298-1
- Albertini, A. A. V., Wernimont, A. K., Muziol, T., Ravelli, R. B. G., Clapier, C. R., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. H. (2006), 'Crystal structure of the rabies virus nucleoprotein-rna complex.', *Science* **313**(5785), 360–363.
 URL: http://dx.doi.org/10.1126/science.1125280
- Ammar, E.-D. & Hogenhout, S. A. (2008), 'A neurotropic route for maize mosaic virus (rhabdoviridae) in its planthopper vector peregrinus maidis.', Virus Res 131(1), 77–85. URL: http://dx.doi.org/10.1016/j.virusres.2007.08.010
- Ammar, E.-D., Tsai, C.-W., Whitfield, A. E., Redinbaugh, M. G. & Hogenhout, S. A. (2009), 'Cellular and molecular aspects of rhabdovirus interactions with insect and plant hosts.', Annu Rev Entomol 54, 447–468. URL: http://dx.doi.org/10.1146/annurev.ento.54.110807.090454
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. & Yeh, L.-S. L. (2004), 'Uniprot : the universal protein knowledgebase.', Nucleic Acids Res 32(Database issue), D115–D119. URL: http://dx.doi.org/10.1093/nar/gkh131
- 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(1), 259–267.
- Assenberg, R., Ren, J., Verma, A., Walter, T. S., Alderton, D., Hurrelbrink, R. J., Fuller, S. D., Bressanelli, S., Owens, R. J., Stuart, D. I. & Grimes, J. M. (2007), 'Crystal structure of the murray valley encephalitis virus ns5 methyltransferase domain in complex with cap analogues.', J Gen Virol 88(Pt 8), 2228–2236. URL: http://dx.doi.org/10.1099/vir.0.82757-0
- Astic, L., Saucier, D., Coulon, P., Lafay, F. & Flamand, A. (1993), 'The cvs strain of rabies virus as transneuronal tracer in the olfactory system of mice.', *Brain Res* **619**(1-2), 146–156.
- Badrane, H. & Tordo, N. (2001), 'Host switching in lyssavirus history from the chiroptera to the carnivora orders.', J Virol **75**(17), 8096–8104.
- Ball, L. A., Pringle, C. R., Flanagan, B., Perepelitsa, V. P. & Wertz, G. W. (1999), 'Phenotypic consequences of rearranging the p, m, and g genes of vesicular stomatitis virus.', J Virol 73(6), 4705–4712.
- Ball, L. A. & White, C. N. (1976), 'Order of transcription of genes of vesicular stomatitis virus.', Proc Natl Acad Sci U S A 73(2), 442–446.
- Baltimore, D., Huang, A. S. & 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), 572–576.

- Bandyopadhyay, A., Kopperud, K., Anderson, G., Martin, K. & Goodin, M. (2010), 'An integrated protein localization and interaction map for potato yellow dwarf virus, type species of the genus nucleorhabdovirus.', Virology 402(1), 61–71. URL: http://dx.doi.org/10.1016/j.virol.2010.03.013
- Banerjee, A. D., Abraham, G. & Colonno, R. J. (1977), 'Vesicular stomatitis virus : mode of transcription.', J Gen Virol 34(1), 1-8.
- Banerjee, A. K. (1987), 'Transcription and replication of rhabdoviruses.', Microbiol Rev 51(1), 66-87.
- Banerjee, A. K., Moyer, S. A. & Rhodes, D. P. (1974), 'Studies on the in vitro adenylation of rna by vesicular stomatitis virus.', Virology 61(2), 547–558.
- Barbosa, E. & Moss, B. (1978), 'mrna nucleoside 2 methyltransferase from vaccinia virus. caracteristics and substrate specificity', J Biol Chem 253(21), 7698–7702. URL: http://dx.doi.org/10.1006/viro.1996.0272
- Barik, S. & Banerjee, A. K. (1992a), 'Phosphorylation by cellular casein kinase ii is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein p.', Proc Natl Acad Sci U S A 89(14), 6570–6574.
- Barik, S. & Banerjee, A. K. (1992b), 'Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation.', J Virol 66(2), 1109–1118.
- Barik, S., Rud, E. W., Luk, D., Banerjee, A. K. & Kang, C. Y. (1990), 'Nucleotide sequence analysis of the l gene of vesicular stomatitis virus (new jersey serotype) : identification of conserved domains in l proteins of nonsegmented negative-strand rna viruses.', Virology 175(1), 332–337.
- Barr, J. N., Tang, X., Hinzman, E., Shen, R. & Wertz, G. W. (2008), 'The vsv polymerase can initiate at mrna start sites located either up or downstream of a transcription termination signal but size of the intervening intergenic region affects efficiency of initiation.', Virology 374(2), 361–370. URL: http://dx.doi.org/10.1016/j.virol.2007.12.023
- Barr, J. N., Whelan, S. P. & Wertz, G. W. (1997*a*), 'cis-acting signals involved in termination of vesicular stomatitis virus mrna synthesis include the conserved auac and the u7 signal for polyadenylation.', *J Virol* **71**(11), 8718–8725.
- Barr, J. N., Whelan, S. P. & Wertz, G. W. (1997b), 'Role of the intergenic dinucleotide in vesicular stomatitis virus rna transcription.', J Virol 71(3), 1794–1801.
- Baudin, F., Bach, C., Cusack, S. & 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), 3158–3165.
- Benarroch, D., Egloff, M.-P., Mulard, L., Guerreiro, C., Romette, J.-L. & Canard, B. (2004), 'A structural basis for the inhibition of the ns5 dengue virus mrna 2'-o-methyltransferase domain by ribavirin 5'-triphosphate.', J Biol Chem 279(34), 35638–35643. URL: http://dx.doi.org/10.1074/jbc.M400460200
- Bergmann JE, Tokuyasu KT, S. S. (1981), 'Passage of an integral membrane protein, the vesicular stomatitis virus glycoprotein, through the golgi apparatus en route to the plasma membrane.', *Proc Natl Acad Sci U S A* **78**(3), 1746–1750.
- Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007), 'Structural characterization of flexible proteins using small-angle x-ray scattering.', J Am Chem Soc 129(17), 5656–5664. URL: http://dx.doi.org/10.1021/ja069124n
- Bhella, D., Ralph, A., Murphy, L. B. & Yeo, R. P. (2002), 'Significant differences in nucleocapsid morphology within the paramyxoviridae.', *J Gen Virol* 83(Pt 8), 1831–1839.
- Bhella, D., Ralph, A. & Yeo, R. P. (2004), 'Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction.', J Mol Biol 340(2), 319–331. URL: http://dx.doi.org/10.1016/j.jmb.2004.05.015
- Bicout, D. J. & Zaccai, G. (2001), 'Protein flexibility from the dynamical transition : a force constant analysis.', Biophys J 80(3), 1115– 1123. URL: http://dx.doi.org/10.1016/S0006-3495(01)76089-4
- Bishop, D. H. & Roy, P. (1972), 'Dissociation of vesicular stomatitis virus and relation of the virion proteins to the viral transcriptase.', J Virol 10(2), 234–243.
- Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W. H. & Marion, D. (2004), 'Structure and dynamics of the nucleocapsid-binding domain of the sendai virus phosphoprotein in solution.', Virology **319**(2), 201–211. URL: http://dx.doi.org/10.1016/j.virol.2003.10.029

- Blondel, D., Regad, T., Poisson, N., Pavie, B., Harper, F., Pandolfi, P. P., Thé, H. D. & Chelbi-Alix, M. K. (2002), 'Rabies virus p and small p products interact directly with pml and reorganize pml nuclear bodies.', Oncogene 21(52), 7957–7970. URL: http://dx.doi.org/10.1038/sj.onc.1205931
- Blumberg, B. M., Crowley, J. C., Silverman, J. I., Menonna, J., Cook, S. D. & Dowling, P. C. (1988), 'Measles virus l protein evidences elements of ancestral rna polymerase.', *Virology* **164**(2), 487–497.
- Blumberg, B. M., Giorgi, C. & Kolakofsky, D. (1983), 'N protein of vesicular stomatitis virus selectively encapsidates leader rna in vitro.', Cell 32(2), 559–567.
- Blumberg, B. M., Giorgi, C., Rose, K. & Kolakofsky, D. (1984), 'Preparation and analysis of the nucleocapsid proteins of vesicular stomatitis virus and sendai virus, and analysis of the sendai virus leader-np gene region.', J Gen Virol 65 (Pt 4), 769–779.
- Blumberg, B. M., Leppert, M. & Kolakofsky, D. (1981), 'Interaction of vsv leader rna and nucleocapsid protein may control vsv genome replication.', Cell 23(3), 837–845.
- Bollati, M., Milani, M., Mastrangelo, E., de Lamballerie, X., Canard, B. & Bolognesi, M. (2009), 'Crystal structure of a methyltransferase from a no-known-vector flavivirus.', *Biochem Biophys Res Commun* 382(1), 200–204. URL: http://dx.doi.org/10.1016/j.bbrc.2009.03.008
- Bouloy, M., Plotch, S. J. & Krug, R. M. (1978), 'Globin mrnas are primers for the transcription of influenza viral rna in vitro.', Proc Natl Acad Sci U S A 75(10), 4886–4890.
- Bourhy, H., Kissi, B. & Tordo, N. (1993), 'Molecular diversity of the lyssavirus genus.', Virology 194(1), 70–81. URL: http://dx.doi.org/10.1006/viro.1993.1236
- Bourhy, H., Tordo, N., Lafon, M. & Sureau, P. (1989), 'Complete cloning and molecular organization of a rabies-related virus, mokola virus.', J Gen Virol 70 (Pt 8), 2063–2074.
- Brun, G., Bao, X. & Prevec, L. (1995), 'The relationship of piry virus to other vesiculoviruses : a re-evaluation based on the glycoprotein gene sequence.', *Intervirology* **38**(5), 274–282.
- Brzozka, K., Finke, S. & 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), 7673–7681. URL: http://dx.doi.org/10.1128/JVI.79.12.7673-7681.2005
- Brzozka, K., Finke, S. & Conzelmann, K.-K. (2006), 'Inhibition of interferon signaling by rabies virus phosphoprotein p : activationdependent binding of stat1 and stat2.', J Virol 80(6), 2675–2683. URL: http://dx.doi.org/10.1128/JVI.80.6.2675-2683.2006
- Buchen-Osmond, C. (2003), 'The universal virus database ictvdb', Computing in Science and Engineering 5(3), 16-25.
- Bujnicki, J. M. & Rychlewski, L. (2002), 'In silico identification, structure prediction and phylogenetic analysis of the 2'-o-ribose (cap 1) methyltransferase domain in the large structural protein of ssrna negative-strand viruses.', *Protein Eng* **15**(2), 101–108.
- Bügl, H., Fauman, E. B., Staker, B. L., Zheng, F., Kushner, S. R., Saper, M. A., Bardwell, J. C. & Jakob, U. (2000), 'Rna methylation under heat shock control.', Mol Cell 6(2), 349–360.
- Calain, P. & Roux, L. (1993), 'The rule of six, a basic feature for efficient replication of sendai virus defective interfering rna.', J Virol 67(8), 4822–4830.
- Canter, D. M. & Perrault, J. (1996), 'Stabilization of vesicular stomatitis virus l polymerase protein by p protein binding : a small deletion in the c-terminal domain of l abrogates binding.', Virology 219(2), 376–386. URL: http://dx.doi.org/10.1006/viro.1996.0263
- Cantlon, J. & Gordy, PW. Bowen, R. (2000), 'Immune responses in mice, cattle and horses to a dna vaccine for vesicular stomatitis', Vaccine 18(22), 2368–2374.
- Carneiro, F. A., Bianconi, M. L., Weissmüller, G., Stauffer, F. & Poian, A. T. D. (2002), 'Membrane recognition by vesicular stomatitis virus involves enthalpy-driven protein-lipid interactions.', J Virol **76**(8), 3756–3764.
- Carneiro, F. A., Ferradosa, A. S. & Poian, A. T. D. (2001), 'Low ph-induced conformational changes in vesicular stomatitis virus glycoprotein involve dramatic structure reorganization.', J Biol Chem 276(1), 62–67. URL: http://dx.doi.org/10.1074/jbc.M008753200

- Carneiro, F. A., Lapido-Loureiro, P. A., Cordo, S. M., Stauffer, F., Weissmüller, G., Bianconi, M. L., Juliano, M. A., Juliano, L., Bisch, P. M., Poian, A. T. D. & Poian, A. T. D. (2006), 'Probing the interaction between vesicular stomatitis virus and phosphatidylserine.', Eur Biophys J 35(2), 145–154. URL: http://dx.doi.org/10.1007/s00249-005-0012-z
- Castel, G., Chtéoui, M., Caignard, G., Préhaud, C., Méhouas, S., Réal, E., Jallet, C., Jacob, Y., Ruigrok, R. W. H. & Tordo, N. (2009), 'Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity.', *J Virol* 83(20), 10808–10820.

URL: http://dx.doi.org/10.1128/JVI.00977-09

- Chamberlin, M. & Ring, J. (1973), 'Characterization of t7-specific ribonucleic acid polymerase. 1. general properties of the enzymatic reaction and the template specificity of the enzyme.', J Biol Chem 248(6), 2235–2244.
- Chattopadhyay, D. & Banerjee, A. K. (1987), 'Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro.', *Cell* **49**(3), 407–414.
- Chelbi-Alix, M. K., Vidy, A., Bougrini, J. E. & Blondel, D. (2006), 'Rabies viral mechanisms to escape the ifn system : the viral protein p interferes with irf-3, stat1, and pml nuclear bodies.', J Interferon Cytokine Res 26(5), 271–280. URL: http://dx.doi.org/10.1089/jir.2006.26.271
- Chen, J. L., Das, T. & Banerjee, A. K. (1997), 'Phosphorylated states of vesicular stomatitis virus p protein in vitro and in vivo.', Virology 228(2), 200–212. URL: http://dx.doi.org/10.1006/viro.1996.8401
- Chen, M., Ogino, T. & Banerjee, A. K. (2007), 'Interaction of vesicular stomatitis virus p and n proteins : identification of two overlapping domains at the n terminus of p that are involved in n0-p complex formation and encapsidation of viral genome rna.', J Virol 81(24), 13478–13485. URL: http://dx.doi.org/10.1128/JVI.01244-07
- Chenik, M., Schnell, M., Conzelmann, K. K. & Blondel, D. (1998), 'Mapping the interacting domains between the rabies virus polymerase and phosphoprotein.', J Virol 72(3), 1925–1930.
- Choi, T. J., Kuwata, S., Koonin, E. V., Heaton, L. A. & Jackson, A. O. (1992), 'Structure of the l (polymerase) protein gene of sonchus yellow net virus.', *Virology* 189(1), 31–39.
- Choudhary, S. K., Malur, A. G., Huo, Y., De, B. P. & Banerjee, A. K. (2002), 'Characterization of the oligomerization domain of the phosphoprotein of human parainfluenza virus type 3.', *Virology* **302**(2), 373–382.
- Chow, T.L. Chow, F. R. H. (1954), 'Morphology of vesicular stomatitis virus', J Bacteriol 68(6), 724–726.
- 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. & Mahy, B. W. (2000), 'Nipah virus : a recently emergent deadly paramyxovirus.', *Science* 288(5470), 1432–1435.
- Chuang, J. L. & Perrault, J. (1997), 'Initiation of vesicular stomatitis virus mutant polr1 transcription internally at the n gene in vitro.', J Virol 71(2), 1466–1475.
- Colonno, R. J., Abraham, G. & Banerjee, A. K. (1976), 'Blocked and unblocked 5'termini in vesicular stomatitis virus product rna in vitro : their possible role in mrna biosynthesis.', *Prog Nucleic Acid Res Mol Biol* **19**, 83–87.
- Colonno, R. J. & Banerjee, A. K. (1977), 'Mapping and initiation studies on the leader rna of vesicular stomatitis virus.', *Virology* **77**(1), 260–268.
- Colonno, R. J. & Banerjee, A. K. (1978), 'Complete nucleotide sequence of the leader rna synthesized in vitro by vesicular stomatitis virus.', Cell 15(1), 93–101.
- Comer, J. A., Tesh, R. B., Modi, G. B., Corn, J. L. & Nettles, V. F. (1990), 'Vesicular stomatitis virus, new jersey serotype : replication in and transmission by lutzomyia shannoni (diptera : Psychodidae).', Am J Trop Med Hyg 42(5), 483–490.
- Compans, R. W. & Choppin, P. W. (1967), 'Isolation and properties of the helical nucleocapsid of the parainfluenza virus sv5.', Proc Natl Acad Sci U S A 57(4), 949–956.
- Condra, J. H. & Lazzarini, R. A. (1980), 'Replicative rna synthesis and nucleocapsid assembly in vesicular stomatitis virus-infected permeable cells.', J Virol **36**(3), 796–804.

Cotton, W. (1927), 'Vesicular stomatitis', Vet. Med 22(6), 169-175.

- Cureton, D. K., Massol, R. H., Saffarian, S., Kirchhausen, T. L. & Whelan, S. P. J. (2009), 'Vesicular stomatitis virus enters cells through vesicles incompletely coated with clathrin that depend upon actin for internalization.', *PLoS Pathog* 5(4), e1000394. URL: http://dx.doi.org/10.1371/journal.ppat.1000394
- Curran, J. (1998), 'A role for the sendai virus p protein trimer in rna synthesis.', J Virol 72(5), 4274–4280.
- Curran, J., Marq, J. B. & 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), 849–855.
- Curran, J., Pelet, T. & Kolakofsky, D. (1994), 'An acidic activation-like domain of the sendai virus p protein is required for rna synthesis and encapsidation.', Virology 202(2), 875–884. URL: http://dx.doi.org/10.1006/viro.1994.1409
- Das, S. C., Panda, D., Nayak, D. & Pattnaik, A. K. (2009), 'Biarsenical labeling of vesicular stomatitis virus encoding tetracysteinetagged m protein allows dynamic imaging of m protein and virus uncoating in infected cells.', J Virol 83(6), 2611–2622. URL: http://dx.doi.org/10.1128/JVI.01668-08
- Das, S. C. & Pattnaik, A. K. (2004), 'Phosphorylation of vesicular stomatitis virus phosphoprotein p is indispensable for virus growth.', J Virol 78(12), 6420–6430. URL: http://dx.doi.org/10.1128/JVI.78.12.6420-6430.2004
- 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(13), 8101–8112. URL: http://dx.doi.org/10.1128/JVI.79.13.8101-8112.2005
- Das, S., Nayak, D., Zhou, Y. & Pattnaik, A. (2006), 'Visualization of intracellular transport of vesicular stomatitis virus nucleocapsids in living cells.', J Virol 80(13), 103–114.
- Das, T., De, B. P. & Banerjee, A. K. (1996), 'Expression, purification, and characterization of rhabdovirus polymerase.', Methods Enzymol 275, 99–122.
- Das, T., Gupta, A. K., Sims, P. W., Gelfand, C. A., Jentoft, J. E. & 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), 24100–24107.
- Das, T., Pattnaik, A. K., Takacs, A. M., Li, T., Hwang, L. N. & Banerjee, A. K. (1997), 'Basic amino acid residues at the carboxyterminal eleven amino acid region of the phosphoprotein (p) are required for transcription but not for replication of vesicular stomatitis virus genome rna.', Virology 238(1), 103–114. URL: http://www.ncbi.nlm.nih.gov/pubmed?term=Basicid0ofbut0RNA
- David, A. (1973), 'Assembly of the vesicular stomatitis virus envelope : incorporation of viral polypeptides into the host plasma membrane', J.Mol.Biol 76(1), 135–148.
- de Almeida Ribeiro, E., Leyrat, C., Gérard, F. C. A., Albertini, A. A. V., Falk, C., Ruigrok, R. W. H. & Jamin, M. (2009), 'Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-rna complexes.', J Mol Biol 394(3), 558–575. URL: http://dx.doi.org/10.1016/j.jmb.2009.09.042
- De, B. & Banerjee, A. K. (1984), 'Specific interactions of vesicular stomatitis virus l and ns proteins with heterologous genome ribonucleoprotein template lead to mrna synthesis in vitro.', J Virol 51(3), 628–634.
- De, B. & Banerjee, A. K. (1985), 'Requirements and functions of vesicular stomatitis virus l and ns proteins in the transcription process in vitro.', *Biochem Biophys Res Commun* **126**(1), 40–49.
- Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. (1990), 'An attempt to unify the structure of polymerases.', *Protein Eng* **3**(6), 461–467.
- Delmas, O., Holmes, E. C., Talbi, C., Larrous, F., Dacheux, L., Bouchier, C. & Bourhy, H. (2008), 'Genomic diversity and evolution of the lyssaviruses.', *PLoS One* 3(4), e2057. URL: http://dx.doi.org/10.1371/journal.pone.0002057
- Deng, M., Bragg, J. N., Ruzin, S., Schichnes, D., King, D., Goodin, M. M. & Jackson, A. O. (2007), 'Role of the sonchus yellow net virus n protein in formation of nuclear viroplasms.', J Virol 81(10), 5362–5374. URL: http://dx.doi.org/10.1128/JVI.02349-06
- Dias, A., Bouvier, D., Crépin, T., McCarthy, A. A., Hart, D. J., Baudin, F., Cusack, S. & Ruigrok, R. W. H. (2009), 'The capsnatching endonuclease of influenza virus polymerase resides in the pa subunit.', *Nature* 458(7240), 914–918. URL: http://dx.doi.org/10.1038/nature07745

- 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(6), 2808–2814. URL: http://dx.doi.org/10.1128/JVI.80.6.2808-2814.2006
- Domingo, E. & Holland, J. J. (1997), 'Rna virus mutations and fitness for survival.', Annu Rev Microbiol 51, 151–178. URL: http://dx.doi.org/10.1146/annurev.micro.51.1.151
- Doms, R. W., Keller, D. S., Helenius, A. & 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), 1957–1969.
- Dong, H., Liu, L., Zou, G., Zhao, Y., Li, Z., Lim, S. P., Shi, P.-Y. & Li, H. (2010), 'Structural and functional analyses of a conserved hydrophobic pocket of flavivirus methyltransferase.', J Biol Chem 285(42), 32586–32595. URL: http://dx.doi.org/10.1074/jbc.M110.129197
- Dong, H., Zhang, B. & Shi, P.-Y. (2008), 'Flavivirus methyltransferase : a novel antiviral target.', Antiviral Res 80(1), 1–10. URL: http://dx.doi.org/10.1016/j.antiviral.2008.05.003
- Doster, W., Cusack, S. & Petry, W. (1989), 'Dynamical transition of myoglobin revealed by inelastic neutron scattering.', Nature 337(6209), 754–756. URL: http://dx.doi.org/10.1038/337754a0
- Ebert, O., Shinozaki, K., Huang, T.-G., Savontaus, M. J., García-Sastre, A. & Woo, S. L. C. (2003), 'Oncolytic vesicular stomatitis virus for treatment of orthotopic hepatocellular carcinoma in immune-competent rats.', *Cancer Res* **63**(13), 3605–3611.
- Egelman, E. H., Wu, S. S., Amrein, M., Portner, A. & Murti, G. (1989), 'The sendai virus nucleocapsid exists in at least four different helical states.', J Virol 63(5), 2233–2243.
- Egloff, M.-P., Benarroch, D., Selisko, B., Romette, J.-L. & Canard, B. (2002), 'An rna cap (nucleoside-2'-o-)-methyltransferase in the flavivirus rna polymerase ns5 : crystal structure and functional characterization.', *EMBO J* 21(11), 2757–2768. URL: http://dx.doi.org/10.1093/emboj/21.11.2757
- Egloff, M.-P., Decroly, E., Malet, H., Selisko, B., Benarroch, D., Ferron, F. & Canard, B. (2007), 'Structural and functional analysis of methylation and 5'-rna sequence requirements of short capped rnas by the methyltransferase domain of dengue virus ns5.', J Mol Biol 372(3), 723–736.
 URL: http://dx.doi.org/10.1016/j.jmb.2007.07.005
- Emerson, S. U. (1982), 'Reconstitution studies detect a single polymerase entry site on the vesicular stomatitis virus genome.', Cell **31**(3 Pt 2), 635–642.
- Emerson, S. U. & Schubert, M. (1987), 'Location of the binding domains for the rna polymerase l and the ribonucleocapsid template within different halves of the ns phosphoprotein of vesicular stomatitis virus.', Proc Natl Acad Sci U S A 84(16), 5655–5659.
- Emerson, S. U. & 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. & Yu, Y. (1975), 'Both ns and l proteins are required for in vitro rna synthesis by vesicular stomatitis virus.', J Virol **15**(6), 1348–1356.
- Emsley, P. & Cowtan, K. (2004), 'Coot : model-building tools for molecular graphics.', Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1), 2126–2132. URL: http://dx.doi.org/10.1107/S0907444904019158
- Fabrega, C., Hausmann, S., Shen, V., Shuman, S. & Lima, C. D. (2004), 'Structure and mechanism of mrna cap (guanine-n7) methyltransferase.', Mol Cell 13(1), 77–89.
- Feldhaus, A. L. & Lesnaw, J. A. (1988), 'Nucleotide sequence of the l gene of vesicular stomatitis virus (new jersey) : identification of conserved domains in the new jersey and indiana l proteins.', *Virology* **163**(2), 359–368.
- Ferron, F., Longhi, S., Henrissat, B. & Canard, B. (2002), 'Viral rna-polymerases a predicted 2'-o-ribose methyltransferase domain shared by all mononegavirales.', *Trends Biochem Sci* 27(5), 222–224.
- Fields (1996), 'Les rhabdovirus', Book Fields Chapitre 37-38.
- Finke, S., Cox, J. H. & Conzelmann, K. K. (2000), 'Differential transcription attenuation of rabies virus genes by intergenic regions : generation of recombinant viruses overexpressing the polymerase gene.', J Virol 74(16), 7261–7269.

- Fitzgerald, D. J., Schaffitzel, C., Berger, P., Wellinger, R., Bieniossek, C., Richmond, T. J. & Berger, I. (2007), 'Multiprotein expression strategy for structural biology of eukaryotic complexes.', *Structure* 15(3), 275–279. URL: http://dx.doi.org/10.1016/j.str.2007.01.016
- Flanagan, E., Zamparo, JM Ball, L., Rodriguez, L. & Wertz, G. (2001), 'Rearrangement of the genes of vesicular stomatitis virus eliminates clinical disease in the natural host : new strategy for vaccine development', J Virol. **75**(13).
- Fujinami, R. S. & Oldstone, M. B. (1983), 'Antigenic modulation : a mechanism of viral persistence.', Prog Brain Res 59, 105–111. URL: http://dx.doi.org/10.1016/S0079-6123(08)63856-X
- Furuichi, Y., LaFiandra, A. & Shatkin, A. J. (1977), '5'-terminal structure and mrna stability.', Nature 266(5599), 235-239.
- Furuichi, Y. & Shatkin, A. J. (2000), 'Viral and cellular mrna capping : past and prospects.', Adv Virus Res 55, 135-184.
- Fuxreiter, M., Simon, I., Friedrich, P. & Tompa, P. (2004), 'Preformed structural elements feature in partner recognition by intrinsically unstructured proteins.', J Mol Biol 338(5), 1015–1026. URL: http://dx.doi.org/10.1016/j.jmb.2004.03.017
- Galinski, M. S., Mink, M. A. & Pons, M. W. (1988), 'Molecular cloning and sequence analysis of the human parainfluenza 3 virus gene encoding the l protein.', *Virology* **165**(2), 499–510.
- Galloway, S. E., Richardson, P. E. & Wertz, G. W. (2008), 'Analysis of a structural homology model of the 2'-o-ribose methyltransferase domain within the vesicular stomatitis virus l protein.', Virology 382(1), 69–82. URL: http://dx.doi.org/10.1016/j.virol.2008.08.041
- Gaudin, Y. (2000), 'Rabies virus-induced membrane fusion pathway.', J Cell Biol 150(3), 601-612.
- Gaudin, Y., Ruigrok, R. W., Tuffereau, C., Knossow, M. & Flamand, A. (1992), 'Rabies virus glycoprotein is a trimer.', Virology 187(2), 627–632.
- Ge, P., Tsao, J., Schein, S., Green, T. J., Luo, M. & Zhou, Z. H. (2010), 'Cryo-em model of the bullet-shaped vesicular stomatitis virus.', Science 327(5966), 689–693. URL: http://dx.doi.org/10.1126/science.1181766
- Geiss, B. J., Thompson, A. A., Andrews, A. J., Sons, R. L., Gari, H. H., Keenan, S. M. & Peersen, O. B. (2009), 'Analysis of flavivirus ns5 methyltransferase cap binding.', J Mol Biol 385(5), 1643–1654. URL: http://dx.doi.org/10.1016/j.jmb.2008.11.058
- Gerard, F. C. A., de Almeida Ribeiro, E., Albertini, A. A. V., Gutsche, I., Zaccai, G., Ruigrok, R. W. H. & Jamin, M. (2007), 'Unphosphorylated rhabdoviridae phosphoproteins form elongated dimers in solution.', *Biochemistry* 46(36), 10328–10338. URL: http://dx.doi.org/10.1021/bi7007799
- Gerard, F. C. A., de Almeida Ribeiro, E., Leyrat, C., Ivanov, I., Blondel, D., Longhi, S., Ruigrok, R. W. H. & Jamin, M. (2009), 'Modular organization of rabies virus phosphoprotein.', J Mol Biol 388(5), 978–996. URL: http://dx.doi.org/10.1016/j.jmb.2009.03.061
- 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 U S A* 83(23), 8873–8877.
- Gode, G. R. & Bhide, N. K. (1988), 'Two rabies deaths after corneal grafts from one donor.', Lancet 2(8614), 791.
- Goodin, M. M., Austin, J., Tobias, R., Fujita, M., Morales, C. & Jackson, A. O. (2001), 'Interactions and nuclear import of the n and p proteins of sonchus yellow net virus, a plant nucleorhabdovirus.', J Virol 75(19), 9393–9406. URL: http://dx.doi.org/10.1128/JVI.75.19.9393-9406.2001
- Goodin, M., Yelton, S., Ghosh, D., Mathews, S. & Lesnaw, J. (2005), 'Live-cell imaging of rhabdovirus-induced morphological changes in plant nuclear membranes.', Mol Plant Microbe Interact 18(7), 703–709. URL: http://dx.doi.org/10.1094/MPMI-18-0703
- Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999), 'Espript : analysis of multiple sequence alignments in postscript.', Bioinformatics 15(4), 305–308.
- Grdzelishvili, V. Z., Smallwood, S., Tower, D., Hall, R. L., Hunt, D. M. & Moyer, S. A. (2005), 'A single amino acid change in the l-polymerase protein of vesicular stomatitis virus completely abolishes viral mrna cap methylation.', J Virol 79(12), 7327–7337. URL: http://dx.doi.org/10.1128/JVI.79.12.7327-7337.2005
- Grdzelishvili, V. Z., Smallwood, S., Tower, D., Hall, R. L., Hunt, D. M. & Moyer, S. A. (2006), 'Identification of a new region in the vesicular stomatitis virus l polymerase protein which is essential for mrna cap methylation.', Virology 350(2), 394–405. URL: http://dx.doi.org/10.1016/j.virol.2006.02.021
- Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. (2006), 'Structure of the vesicular stomatitis virus nucleoprotein-rna complex.', Science 313(5785), 357–360. URL: http://dx.doi.org/10.1126/science.1126953
- Gupta, A. K. & Banerjee, A. K. (1997), 'Expression and purification of vesicular stomatitis virus n-p complex from escherichia coli : role in genome rna transcription and replication in vitro.', J Virol 71(6), 4264–4271.
- Gupta, A. K., Blondel, D., Choudhary, S. & 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), 91–98.
- Gupta, A. K., Mathur, M. & Banerjee, A. K. (2002), 'Unique capping activity of the recombinant rna polymerase (l) of vesicular stomatitis virus : association of cellular capping enzyme with the l protein.', *Biochem Biophys Res Commun* 293(1), 264–268. URL: http://dx.doi.org/10.1016/S0006-291X(02)00217-6
- Habchi, J. & Longhi, S. (2011), 'Structural disorder within paramyxovirus nucleoproteins and phosphoproteins.', Mol Biosyst . URL: http://dx.doi.org/10.1039/c1mb05204g
- Haglund, K., Leiner, I., Kerksiek, K., Buonocore, L., Pamer, E. & Rose, J. K. (2002), 'High-level primary cd8(+) t-cell response to human immunodeficiency virus type 1 gag and env generated by vaccination with recombinant vesicular stomatitis viruses.', J Virol 76(6), 2730–2738.
- Halpin, K., Bankamp, B., Harcourt, B. H., Bellini, W. J. & Rota, P. A. (2004), 'Nipah virus conforms to the rule of six in a minigenome replication assay.', J Gen Virol 85(Pt 3), 701–707.
- Hanson, R. (1952), 'The natural history of vesicular stomatitis', Bacteriol Rev. 16(3), 179-204.
- Harty, R. N., Paragas, J., Sudol, M. & Palese, P. (1999), 'A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with ww domains of cellular proteins : implications for viral budding.', J Virol **73**(4), 2921–2929.
- Haupt, W. (1999), 'Rabies-risk of exposure and current trends in prevention of human cases', Vaccine 17(13-14), 1742–1749.
- Heggeness, M. H., Scheid, A. & Choppin, P. W. (1980), 'Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus : reversible coiling and uncoiling induced by changes in salt concentration.', Proc Natl Acad Sci U S A 77(5), 2631– 2635.
- Hemachudha, T., Laothamatas, J. & Rupprecht, C. E. (2002), 'Human rabies : a disease of complex neuropathogenetic mechanisms and diagnostic challenges.', *Lancet Neurol* 1(2), 101–109.
- Hercyk, N., Horikami, S. M. & Moyer, S. A. (1988), 'The vesicular stomatitis virus l protein possesses the mrna methyltransferase activities.', Virology 163(1), 222–225.
- Hodel, A. E., Gershon, P. D., Shi, X. & Quiocho, F. A. (1996), 'The 1.85 a structure of vaccinia protein vp39 : a bifunctional enzyme that participates in the modification of both mrna ends.', *Cell* 85(2), 247–256.
- Hogenhout, S. A., Redinbaugh, M. G. & Ammar, E.-D. (2003), 'Plant and animal rhabdovirus host range : a bug's view.', Trends Microbiol 11(6), 264–271.
- Holmes, D. E. & 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), 3078–3083.
- Horikami, S. M. & Moyer, S. A. (1982), 'Host range mutants of vesicular stomatitis virus defective in in vitro rna methylation.', Proc Natl Acad Sci U S A 79(24), 7694–7698.
- Howard, M. & Wertz, G. (1989), 'Vesicular stomatitis virus rna replication : a role for the ns protein.', J Gen Virol 70 (Pt 10), 2683–2694.
- Howatson, A. F. & Whitmore, G. F. (1962), 'The development and structure of vesicular stomatitis virus.', Virology 16, 466-478.
- Huang AS, Baltimore D, S. M. (1970), 'Ribonucleic acid synthesis of vesicular stomatitis virus. 3. multiple complementary messenger rna molecules.', Virology 42(4), 946–957.
- Hunt, D. M., Emerson, S. U. & Wagner, R. R. (1976), 'Rna- temperature-sensitive mutants of vesicular stomatitis virus : L-protein thermosensitivity accounts for transcriptase restriction of group i mutants.', J Virol 18(2), 596–603.

- Hwang, L. N., Englund, N., Das, T., Banerjee, A. K. & 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(7), 5613–5620.
- Hâkansson, K. & Wigley, D. B. (1998), 'Structure of a complex between a cap analogue and mrna guanylyl transferase demonstrates the structural chemistry of rna capping.', *Proc Natl Acad Sci U S A* **95**(4), 1505–1510.
- Håkansson, K., Doherty, A. J., Shuman, S. & Wigley, D. B. (1997), 'X-ray crystallography reveals a large conformational change during guanyl transfer by mrna capping enzymes.', Cell 89(4), 545–553.
- Irie, T., Licata, J. M., Jayakar, H. R., Whitt, M. A., Bell, P. & 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), 7823–7827. URL: http://dx.doi.org/10.1128/JVI.78.14.7823-7827.2004
- Iseni, F., Barge, A., Baudin, F., Blondel, D. & Ruigrok, R. W. (1998), 'Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures.', J Gen Virol 79 (Pt 12), 2909–2919.
- Iseni, F., Baudin, F., Blondel, D. & Ruigrok, R. W. (2000), 'Structure of the rna inside the vesicular stomatitis virus nucleocapsid.', RNA 6(2), 270–281.
- Iseni, F., Baudin, F., Garcin, D., Marq, J.-B., Ruigrok, R. W. H. & Kolakofsky, D. (2002), 'Chemical modification of nucleotide bases and mrna editing depend on hexamer or nucleoprotein phase in sendai virus nucleocapsids.', RNA 8(8), 1056–1067.
- Ito, N., Moseley, G. W., Blondel, D., Shimizu, K., Rowe, C. L., Ito, Y., Masatani, T., Nakagawa, K., Jans, D. A. & Sugiyama, M. (2010), 'Role of interferon antagonist activity of rabies virus phosphoprotein in viral pathogenicity.', J Virol 84(13), 6699–6710. URL: http://dx.doi.org/10.1128/JVI.00011-10
- Ivanov, I., Crépin, T., Jamin, M. & Ruigrok, R. W. H. (2010), 'Structure of the dimerization domain of the rabies virus phosphoprotein.', J Virol 84(7), 3707–3710. URL: http://dx.doi.org/10.1128/JVI.02557-09
- Ivanov, I., Yabukarski, F., Ruigrok, R. W. H. & Jamin, M. (2011), 'Structural insights into the rhabdovirus transcription/replication complex.', Virus Res. URL: http://dx.doi.org/10.1016/j.virusres.2011.09.025
- Iverson, L. E. & Rose, J. K. (1981), 'Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription.', Cell 23(2), 477–484.
- Jacob, Y., Badrane, H., Ceccaldi, P. E. & Tordo, N. (2000), 'Cytoplasmic dynein lc8 interacts with lyssavirus phosphoprotein.', J Virol 74(21), 10217–10222.
- Jacob, Y., Real, E. & Tordo, N. (2001), 'Functional interaction map of lyssavirus phosphoprotein : identification of the minimal transcription domains.', J Virol 75(20), 9613–9622. URL: http://dx.doi.org/10.1128/JVI.75.20.9613-9622.2001
- Johannsdottir, H. K., Mancini, R., Kartenbeck, J., Amato, L. & Helenius, A. (2009), 'Host cell factors and functions involved in vesicular stomatitis virus entry.', J Virol 83(1), 440–453. URL: http://dx.doi.org/10.1128/JVI.01864-08
- Johnson, K. M., Tesh, R. B. & 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), 2133–2140.
- Johnson, K. M., Vogel, J. E. & Peralta, P. H. (1966), 'Clinical and serological response to laboratory-acquired human infection by indiana type vesicular stomatitis virus (vsv).', Am J Trop Med Hyg 15(2), 244–246.
- Jones, O. W. & Berg, P. (1966), 'Studies on the binding of rna polymerase to polynucleotides.', J Mol Biol 22(2), 199-209.
- Kabsch, W. (1988), 'Automatic indexing of rotation diffraction patterns', Journal of Applied Crystallography 21(1), 67–72. URL: http://dx.doi.org/10.1107/S0021889887009737
- Kalmakoff, J., Lewandowski, L. J. & Black, D. R. (1969), 'Comparison of the ribonucleic acid subunits of reovirus, cytoplasmic polyhedrosis virus, and wound tumor virus.', J Virol 4(6), 851–856.
- Kamer, G. & Argos, P. (1984), 'Primary structural comparison of rna-dependent polymerases from plant, animal and bacterial viruses.', *Nucleic Acids Res* **12**(18), 7269–7282.

- Karlin, D., Ferron, F., Canard, B. & Longhi, S. (2003), 'Structural disorder and modular organization in paramyxovirinae n and p.', J Gen Virol 84(Pt 12), 3239–3252.
- Kastner, B., Fischer, N., Golas, M. M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J. M., Poerschke, D., Lührmann, R. & Stark, H. (2008), 'Grafix : sample preparation for single-particle electron cryomicroscopy.', Nat Methods 5(1), 53–55. URL: http://dx.doi.org/10.1038/nmeth1139
- Kawai, A., Toriumi, H., Tochikura, T. S., Takahashi, T., Honda, Y. & Morimoto, K. (1999), 'Nucleocapsid formation and/or subsequent conformational change of rabies virus nucleoprotein (n) is a prerequisite step for acquiring the phosphatase-sensitive epitope of monoclonal antibody 5-2-26.', Virology 263(2), 395–407. URL: http://dx.doi.org/10.1006/viro.1999.9962
- Kawano, M., Okamoto, K., Bando, H., Kondo, K., Tsurudome, M., Komada, H., Nishio, M. & Ito, Y. (1991), 'Characterizations of the human parainfluenza type 2 virus gene encoding the l protein and the intergenic sequences.', Nucleic Acids Res 19(10), 2739–2746.
- Keegan, R. M. & Winn, M. D. (2008), 'Mrbump : an automated pipeline for molecular replacement.', Acta Crystallogr D Biol Crystallogr 64(Pt 1), 119–124. URL: http://dx.doi.org/10.1107/S0907444907037195
- Keene, J. D. & Lazzarini, R. A. (1976), 'A comparison of the extents of methylation of vesicular stomatitis virus messenger rna.', Virology 69(1), 364–367.
- Keene, J. D., Thornton, B. J. & Emerson, S. U. (1981), 'Sequence-specific contacts between the rna polymerase of vesicular stomatitis virus and the leader rna gene.', Proc Natl Acad Sci U S A 78(10), 6191–6195.
- Kelly, R. M. & Strick, P. L. (2000), 'Rabies as a transneuronal tracer of circuits in the central nervous system.', J Neurosci Methods 103(1), 63–71.
- Kiley, M. P. & Wagner, R. R. (1972), 'Ribonucleic acid species of intracellular nucleocapsids and released virions of vesicular stomatitis virus.', J Virol 10(2), 244–255.
- Knipe, D. . H. (2001), 'Fields virology', Book Fields N/A(4th Ed.), 1221-1244.
- Knipe, D. . H. (2006), 'Fields virology (downloaded from internet)', Book Fields N/A(5th Ed.).
- Kobayashi, Y., Sato, G., Mochizuki, N., Hirano, S., Itou, T., Carvalho, A. A. B., Albas, A., Santos, H. P., Ito, F. H. & Sakai, T. (2008), 'Molecular and geographic analyses of vampire bat-transmitted cattle rabies in central brazil.', BMC Vet Res 4, 44. URL: http://dx.doi.org/10.1186/1746-6148-4-44
- Kolakofsky, D., Mercier, P. L., Iseni, F. & Garcin, D. (2004), 'Viral dna polymerase scanning and the gymnastics of sendai virus rna synthesis.', Virology 318(2), 463–473.
- Kolakofsky, D., Pelet, T., Garcin, D., Hausmann, S., Curran, J. & Roux, L. (1998), 'Paramyxovirus rna synthesis and the requirement for hexamer genome length : the rule of six revisited.', J Virol 72(2), 891–899.
- Kolakofsky, D., Roux, L., Garcin, D. & Ruigrok, R. W. H. (2005), 'Paramyxovirus mrna editing, the "rule of six" and error catastrophe : a hypothesis.', J Gen Virol 86(Pt 7), 1869–1877. URL: http://dx.doi.org/10.1099/vir.0.80986-0
- Kouznetzoff, A., Buckle, M. & Tordo, N. (1998), 'Identification of a region of the rabies virus n protein involved in direct binding to the viral rna.', J Gen Virol 79 (Pt 5), 1005–1013.
- Kurath, G Leong, J. (1985), 'Characterization of infectious hematopoietic necrosis virus mrna species reveals a nonvirion rhabdovirus protein.', J Virol. 53(2), 462–468.
- Kurilla, M. G., Piwnica-Worms, H. & 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), 5240–5244.
- Kuzmin, I. V., Franka, R. & Rupprecht, C. E. (2008), 'Experimental infection of big brown bats (eptesicus fuscus) with west caucasian bat virus (webv).', Dev Biol (Basel) 131, 327–337.
- Köhler, G. & Milstein, C. (1975), 'Continuous cultures of fused cells secreting antibody of predefined specificity.', *Nature* **256**(5517), 495–497.
- Lesburg, C. A., Cable, M. B., Ferrari, E., Hong, Z., Mannarino, A. F. & 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), 937–943. URL: http://dx.doi.org/10.1038/13305

- Leslie, A. G. W. (2006), 'The integration of macromolecular diffraction data.', Acta Crystallogr D Biol Crystallogr 62(Pt 1), 48–57. URL: http://dx.doi.org/10.1107/S0907444905039107
- Lewandowski, L. J., Kalmakoff, J. & 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.
- Leyrat, C., Gérard, F. C. A., de Almeida Ribeiro, E., Ivanov, I., Ruigrok, R. W. H. & Jamin, M. (2010), 'Structural disorder in proteins of the rhabdoviridae replication complex.', Protein Pept Lett 17(8), 979–987.

Leyrat, C., Jensen, M. R., Ribeiro, E. A., Gérard, F. C. A., Ruigrok, R. W. H., Blackledge, M. & Jamin, M. (2011), 'The n(0)-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient alpha helices.', Protein Sci 20(3), 542–556. URL: http://dx.doi.org/10.1002/pro.587

- Leyrat, C., Yabukarski, F., Tarbouriech, N., Ribeiro, E. A., Jensen, M. R., Blackledge, M., Ruigrok, R. W. H. & Jamin, M. (2011), 'Structure of the vesicular stomatitis virus n-p complex.', *PLoS Pathog* 7(9), e1002248. URL: http://dx.doi.org/10.1371/journal.ppat.1002248
- Li, J., Chorba, J. S. & Whelan, S. P. J. (2007), 'Vesicular stomatitis viruses resistant to the methylase inhibitor sinefungin upregulate rna synthesis and reveal mutations that affect mrna cap methylation.', J Virol 81(8), 4104–4115. URL: http://dx.doi.org/10.1128/JVI.02681-06
- Li, J., Fontaine-Rodriguez, E. C. & Whelan, S. P. J. (2005), 'Amino acid residues within conserved domain vi of the vesicular stomatitis virus large polymerase protein essential for mrna cap methyltransferase activity.', J Virol 79(21), 13373–13384. URL: http://dx.doi.org/10.1128/JVI.79.21.13373-13384.2005
- Li, J., Rahmeh, A., Brusic, V. & Whelan, S. P. J. (2009), 'Opposing effects of inhibiting cap addition and cap methylation on polyadenylation during vesicular stomatitis virus mrna synthesis.', J Virol 83(4), 1930–1940. URL: http://dx.doi.org/10.1128/JVI.02162-08
- Li, J., Rahmeh, A., Morelli, M. & Whelan, S. P. J. (2008), 'A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense rna viruses that is essential for mrna capping.', J Virol 82(2), 775–784. URL: http://dx.doi.org/10.1128/JVI.02107-07
- Li, J., Wang, J. T. & Whelan, S. P. J. (2006), 'A unique strategy for mrna cap methylation used by vesicular stomatitis virus.', Proc Natl Acad Sci U S A 103(22), 8493–8498. URL: http://dx.doi.org/10.1073/pnas.0509821103
- Li, M. L., Rao, P. & Krug, R. M. (2001), 'The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits.', EMBO J 20(8), 2078–2086. URL: http://dx.doi.org/10.1093/emboj/20.8.2078
- Li, T. & Pattnaik, A. K. (1999), 'Overlapping signals for transcription and replication at the 3' terminus of the vesicular stomatitis virus genome.', J Virol 73(1), 444–452.
- Libersou S, Albertini AA Ouldali, M., Maury, V., Maheu, C., Raux, H., de Haas, F., Roche, S., Gaudin, Y. & Lepault, J. (2010), 'Distinct structural rearrangements of the vsv glycoprotein drive membrane fusion.', *J Cell Biochem Biophys Res Commun* **191**(1), 1930–1940.
- Lichty, B. D., Power, A. T., Stojdl, D. F. & Bell, J. C. (2004), 'Vesicular stomatitis virus : re-inventing the bullet.', Trends Mol Med 10(5), 210–216. URL: http://dx.doi.org/10.1016/j.molmed.2004.03.003
- Lima, C. D., Wang, L. K. & Shuman, S. (1999), 'Structure and mechanism of yeast rna triphosphatase : an essential component of the mrna capping apparatus.', Cell 99(5), 533–543.
- Llorente, M. T., García-Barreno, B., Calero, M., Camafeita, E., López, J. A., Longhi, S., Ferrón, F., Varela, P. F. & Melero, J. A. (2006), 'Structural analysis of the human respiratory syncytial virus phosphoprotein : characterization of an alpha-helical domain involved in oligomerization.', J Gen Virol 87 (Pt 1), 159–169. URL: http://dx.doi.org/10.1099/vir.0.81430-0
- Lotti, L., Torrisi, M., Pascale, M. & Bonatti, S. (1992), 'Immunocytochemical analysis of the transfer of vesicular stomatitis virus g glycoprotein from the intermediate compartment to the golgi complex', *J Cell Biol* **11**(1), 43 -50. URL:http ://dx.doi.org/10.1128/JVI.02034-09

- Luo, M., Green, T. J., Zhang, X., Tsao, J. & Qiu, S. (2007), 'Conserved characteristics of the rhabdovirus nucleoprotein.', Virus Res 129(2), 246–251. URL: http://dx.doi.org/10.1016/j.virusres.2007.07.011
- LWOFF, A., HORNE, R. & TOURNIER, P. (1962a), 'A system of viruses.', Cold Spring Harb Symp Quant Biol 27, 51-55.
- Lwoff, A., Horne, R. & Tournier, P. (1962b), 'A virus system', C R Hebd Seances Acad Sci 254, 4225-4227.
- Malet, H., Egloff, M.-P., Selisko, B., Butcher, R. E., Wright, P. J., Roberts, M., Gruez, A., Sulzenbacher, G., Vonrhein, C., Bricogne, G., Mackenzie, J. M., Khromykh, A. A., Davidson, A. D. & Canard, B. (2007), 'Crystal structure of the rna polymerase domain of the west nile virus non-structural protein 5.', J Biol Chem 282(14), 10678–10689. URL: http://dx.doi.org/10.1074/jbc.M607273200
- Martins, C. R., Johnson, J. A., Lawrence, D. M., Choi, T. J., Pisi, A. M., Tobin, S. L., Lapidus, D., Wagner, J. D., Ruzin, S., McDonald, K. & Jackson, A. O. (1998), 'Sonchus yellow net rhabdovirus nuclear viroplasms contain polymerase-associated proteins.', J Virol 72(7), 5669–5679.
- Masters, P. S. & Banerjee, A. K. (1988), 'Complex formation with vesicular stomatitis virus phosphoprotein ns prevents binding of nucleocapsid protein n to nonspecific rna.', J Virol 62(8), 2658–2664.
- Mathur, M., Das, T., Chen, J. L., Chattopadhyay, D. & Banerjee, A. K. (1997), 'Display of disparate transcription phenotype by the phosphorylation negative p protein mutants of vesicular stomatitis virus, indiana serotype, expressed in e. coli and eucaryotic cells.', Gene Expr 6(5), 275–286.
- Matlin, K. S., Reggio, H., Helenius, A. & Simons, K. (1982), 'Pathway of vesicular stomatitis virus entry leading to infection.', J Mol Biol 156(3), 609–631.
- Mavrakis, M., Iseni, F., Mazza, C., Schoehn, G., Ebel, C., Gentzel, M., Franz, T. & Ruigrok, R. W. H. (2003), 'Isolation and characterisation of the rabies virus n degrees-p complex produced in insect cells.', *Virology* **305**(2), 406–414.
- Mavrakis, M., Kolesnikova, L., Schoehn, G., Becker, S. & Ruigrok, R. W. H. (2002), 'Morphology of marburg virus np-rna.', Virology 296(2), 300–307.
 URL: http://dx.doi.org/10.1006/viro.2002.1433
- Mavrakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. H. (2004), 'Structure and function of the c-terminal domain of the polymerase cofactor of rabies virus.', J Mol Biol 343(4), 819–831. URL: http://dx.doi.org/10.1016/j.jmb.2004.08.071
- Mavrakis, M., Méhouas, S., Réal, E., Iseni, F., Blondel, D., Tordo, N. & Ruigrok, R. W. H. (2006), 'Rabies virus chaperone : identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific rna.', *Virology* **349**(2), 422–429.

URL: http://dx.doi.org/10.1016/j.virol.2006.01.030

- McColl, K. A., Tordo, N. & Setién, A. A. A. (2000), 'Bat lyssavirus infections.', Rev Sci Tech 19(1), 177-196.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007), 'Phaser crystallographic software.', J Appl Crystallogr 40(Pt 4), 658–674. URL: http://dx.doi.org/10.1107/S0021889807021206
- Mebatsion, T., Weiland, F. & Conzelmann, K. K. (1999), 'Matrix protein of rabies virus is responsible for the assembly and budding of bullet-shaped particles and interacts with the transmembrane spike glycoprotein g.', J Virol 73(1), 242–250.
- Mellon, M. G. & Emerson, S. U. (1978), 'Rebinding of transcriptase components (l and ns proteins) to the nucleocapsid template of vesicular stomatitis virus.', J Virol 27(3), 560–567.
- Milani, M., Mastrangelo, E., Bollati, M., Selisko, B., Decroly, E., Bouvet, M., Canard, B. & Bolognesi, M. (2009), 'Flaviviral methyltransferase/rna interaction : structural basis for enzyme inhibition.', Antiviral Res 83(1), 28–34. URL: http://dx.doi.org/10.1016/j.antiviral.2009.03.001
- Mire, C. E., Dube, D., Delos, S. E., White, J. M. & Whitt, M. A. (2009), 'Glycoprotein-dependent acidification of vesicular stomatitis virus enhances release of matrix protein.', J Virol 83(23), 12139–12150. URL: http://dx.doi.org/10.1128/JVI.00955-09
- Mire, C. E., White, J. M. & Whitt, M. A. (2010), 'A spatio-temporal analysis of matrix protein and nucleocapsid trafficking during vesicular stomatitis virus uncoating.', *PLoS Pathog* 6(7), e1000994. URL: http://dx.doi.org/10.1371/journal.ppat.1000994

Miyamoto, K Matsumoto, S. (1965), 'The nature of the negri body', J. Cell. Biol. 27(3), 677-782.

Moseley, G. W., Lahaye, X., Roth, D. M., Oksayan, S., Filmer, R. P., Rowe, C. L., Blondel, D. & Jans, D. A. (2009), 'Dual modes of rabies p-protein association with microtubules : a novel strategy to suppress the antiviral response.', J Cell Sci 122(Pt 20), 3652– 3662.

URL: http://dx.doi.org/10.1242/jcs.045542

- Moyer, S. A., Abraham, G., Adler, R. & Banerjee, A. K. (1975), 'Methylated and blocked 5' termini in vesicular stomatitis virus in vivo mrnas.', Cell 5(1), 59–67.
- Moyer, S. A. & Banerjee, A. K. (1976), 'In vivo methylation of vesicular stomatitis virus and its host-cell messenger rna species.', Virology 70(2), 339–351.
- Moyer, S., Smallwood-Kentro, S., Haddad, A. & Prevec, L. (1991), 'Assembly and transcription of synthetic vesicular stomatitis virus nucleocapsids.', J Virol. 65(5), 2170–2178.
- Munyon, W., Paoletti, E. & Grace, J. T. (1967), 'Rna polymerase activity in purified infectious vaccinia virus.', Proc Natl Acad Sci U S A 58(6), 2280–2287.
- Murphy, A. M. & Grdzelishvili, V. Z. (2009), 'Identification of sendai virus l protein amino acid residues affecting viral mrna cap methylation.', J Virol 83(4), 1669–1681. URL: http://dx.doi.org/10.1128/JVI.01438-08
- Muthukrishnan, S., Morgan, M., Banerjee, A. K. & Shatkin, A. J. (1976), 'Influence of 5'-terminal m7g and 2'-o-methylated residues on messenger ribonucleic acid binding to ribosomes.', *Biochemistry* 15(26), 5761–5768.
- Navaza, J., Lepault, J., Rey, F. A., Alvarez-Rúa, C. & Borge, J. (2002), 'On the fitting of model electron densities into em reconstructions : a reciprocal-space formulation.', Acta Crystallogr D Biol Crystallogr 58(Pt 10 Pt 2), 1820–1825.
- Newcomb, W. W. & Brown, J. C. (1981), 'Role of the vesicular stomatitis virus matrix protein in maintaining the viral nucleocapsid in the condensed form found in native virions.', J Virol 39(1), 295–299.
- Newcomb, W. W., Tobin, G. J., McGowan, J. J. & Brown, J. C. (1982), 'In vitro reassembly of vesicular stomatitis virus skeletons.', J Virol 41(3), 1055–1062.
- Ogino, T. & 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. URL: http://dx.doi.org/10.1016/j.molcel.2006.11.013
- Ogino, T. & Banerjee, A. K. (2011), 'An unconventional pathway of mrna cap formation by vesiculoviruses.', Virus Res. URL: http://dx.doi.org/10.1016/j.virusres.2011.09.012
- Ogino, T., Kobayashi, M., Iwama, M. & Mizumoto, K. (2005), 'Sendai virus rna-dependent rna polymerase l protein catalyzes cap methylation of virus-specific mrna.', J Biol Chem 280(6), 4429–4435. URL: http://dx.doi.org/10.1074/jbc.M411167200
- Ogino, T., Yadav, S. P. & Banerjee, A. K. (2010), 'Histidine-mediated rna transfer to gdp for unique mrna capping by vesicular stomatitis virus rna polymerase.', Proc Natl Acad Sci U S A 107(8), 3463–3468. URL: http://dx.doi.org/10.1073/pnas.0913083107
- Palin, A., Chattopadhyay, A., Park, S., Delmas, G., Suresh, R., Senina, S., Perlin, D. S. & Rose, J. K. (2007), 'An optimized vaccine vector based on recombinant vesicular stomatitis virus gives high-level, long-term protection against yersinia pestis challenge.', *Vaccine* 25(4), 741–750. URL: http://dx.doi.org/10.1016/j.vaccine.2006.08.010
- Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005), 'Auto-rickshaw : an automated crystal structure determination platform as an efficient tool for the validation of an x-ray diffraction experiment.', Acta Crystallogr D Biol Crystallogr 61(Pt 4), 449–457.

URL: http://dx.doi.org/10.1107/S0907444905001307

- Parks, G. D., Ward, C. D. & Lamb, R. A. (1992), 'Molecular cloning of the np and l genes of simian virus 5 : identification of highly conserved domains in paramyxovirus np and l proteins.', Virus Res 22(3), 259–279.
- Pasdeloup, D., Poisson, N., Raux, H., Gaudin, Y., Ruigrok, R. W. H. & Blondel, D. (2005), 'Nucleocytoplasmic shuttling of the rabies virus p protein requires a nuclear localization signal and a crm1-dependent nuclear export signal.', Virology 334(2), 284–293. URL: http://dx.doi.org/10.1016/j.virol.2005.02.005

- Pattnaik, A. K., Hwang, L., Li, T., Englund, N., Mathur, M., Das, T. & Banerjee, A. K. (1997), 'Phosphorylation within the aminoterminal acidic domain i of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication.', J Virol 71(11), 8167–8175.
- Pattnaik, A. K. & 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), 2948–2957.
- Patton JT, Davis NL, W. G. (1984), 'N protein alone satisfies the requirement for protein synthesis during rna replication of vesicular stomatitis virus.', J Virol 49(2), 303–309.
- Paul, P. R., Chattopadhyay, D. & Banerjee, A. K. (1988), 'The functional domains of the phosphoprotein (ns) of vesicular stomatitis virus (indiana serotype).', Virology 166(2), 350–357.
- Peeters, B. P., Gruijthuijsen, Y. K., de Leeuw, O. S. & Gielkens, A. L. (2000), 'Genome replication of newcastle disease virus : involvement of the rule-of-six.', Arch Virol 145(9), 1829–1845.
- Pelet, T., Delenda, C., Gubbay, O., Garcin, D. & Kolakofsky, D. (1996), 'Partial characterization of a sendai virus replication promoter and the rule of six.', Virology 224(2), 405–414. URL: http://dx.doi.org/10.1006/viro.1996.0547
- Perlman, S. M. & Huang, A. S. (1973), 'Rna synthesis of vesicular stomatitis virus. v. interactions between transcription and replication.', J Virol 12(6), 1395–1400.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999), 'Automated protein model building combined with iterative structure refinement.', Nat Struct Biol 6(5), 458–463. URL: http://dx.doi.org/10.1038/8263
- Perrault, J., Clinton, G. M. & McClure, M. A. (1983), 'Rnp template of vesicular stomatitis virus regulates transcription and replication functions.', Cell 35(1), 175–185.
- Petsko, G. A. (1985), 'Preparation of isomorphous heavy-atom derivatives.', Methods Enzymol 114, 147-156.
- Phillips, R. J., Samson, A. C. & Emmerson, P. T. (1998), 'Nucleotide sequence of the 5'-terminus of newcastle disease virus and assembly of the complete genomic sequence : agreement with the "rule of six".', Arch Virol 143(10), 1993–2002.
- Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. (1981), 'A unique cap(m7gpppxm)-dependent influenza virion endonuclease cleaves capped rnas to generate the primers that initiate viral rna transcription.', Cell 23(3), 847–858.
- Poch, O., Blumberg, B. M., Bougueleret, L. & Tordo, N. (1990), 'Sequence comparison of five polymerases (l proteins) of unsegmented negative-strand rna viruses : theoretical assignment of functional domains.', J Gen Virol 71 (Pt 5), 1153–1162.
- Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989), 'Identification of four conserved motifs among the rna-dependent polymerase encoding elements.', EMBO J 8(12), 3867–3874.
- Poisson, N., Real, E., Gaudin, Y., Vaney, M. C., King, S., Jacob, Y., Tordo, N. & 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), 2691–2696.
- Pringle, C. R. (1997), 'The order mononegavirales-current status.', Arch Virol 142(11), 2321–2326.
- Pérez, Y., Gairí, M., Pons, M. & Bernadó, P. (2009), 'Structural characterization of the natively unfolded n-terminal domain of human c-src kinase : insights into the role of phosphorylation of the unique domain.', J Mol Biol 391(1), 136–148. URL: http://dx.doi.org/10.1016/j.jmb.2009.06.018
- Qanungo, K., Shaji, D., Mathur, M. & Banerjee, A. (2004), 'Two rna polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome rna.', Proc Natl Acad Sci U S A 101(16), 5952–5957.
- Qi, X., Lan, S., Wang, W., Schelde, L. M., Dong, H., Wallat, G. D., Ly, H., Liang, Y. & Dong, C. (2010), 'Cap binding and immune evasion revealed by lassa nucleoprotein structure.', *Nature* 468(7325), 779–783. URL: http://dx.doi.org/10.1038/nature09605
- Quiroz, E., Moreno, N., Peralta, P. H. & Tesh, R. B. (1988), 'A human case of encephalitis associated with vesicular stomatitis virus (indiana serotype) infection.', Am J Trop Med Hyg 39(3), 312–314.
- Rahmeh, A. A., Li, J., Kranzusch, P. J. & Whelan, S. P. J. (2009), 'Ribose 2'-o methylation of the vesicular stomatitis virus mrna cap precedes and facilitates subsequent guanine-n-7 methylation by the large polymerase protein.', J Virol 83(21), 11043–11050. URL: http://dx.doi.org/10.1128/JVI.01426-09

- Rahmeh, A. A., Schenk, A. D., Danek, E. I., Kranzusch, P. J., Liang, B., Walz, T. & Whelan, S. P. J. (2010), 'Molecular architecture of the vesicular stomatitis virus rna polymerase.', Proc Natl Acad Sci U S A 107(46), 20075–20080. URL: http://dx.doi.org/10.1073/pnas.1013559107
- Ramsburg, E., Rose, N. F., Marx, P. A., Mefford, M., Nixon, D. F., Moretto, W. J., Montefiori, D., Earl, P., Moss, B. & Rose, J. K. (2004), 'Highly effective control of an aids virus challenge in macaques by using vesicular stomatitis virus and modified vaccinia virus ankara vaccine vectors in a single-boost protocol.', J Virol 78(8), 3930–3940.
- Rao, B. L., Basu, A., Wairagkar, N. S., Gore, M. M., Arankalle, V. A., Thakare, J. P., Jadi, R. S., Rao, K. A. & Mishra, A. C. (2004), 'A large outbreak of acute encephalitis with high fatality rate in children in andhra pradesh, india, in 2003, associated with chandipura virus.', *Lancet* 364(9437), 869–874. URL: http://dx.doi.org/10.1016/S0140-6736(04)16982-1
- Raux, H., Flamand, A. & Blondel, D. (2000), 'Interaction of the rabies virus p protein with the lc8 dynein light chain.', J Virol 74(21), 10212–10216.
- Raymond, D. D., Piper, M. E., Gerrard, S. R. & Smith, J. L. (2010), 'Structure of the rift valley fever virus nucleocapsid protein reveals another architecture for rna encapsidation.', Proc Natl Acad Sci U S A 107(26), 11769–11774. URL: http://dx.doi.org/10.1073/pnas.1001760107
- Rhodes, D. P. & Banerjee, A. K. (1975), '5'-terminal sequence of vesicular stomatitis virus mrna's synthesized in vitro.', J Virol 17(1), 33–42.
- Rhodes, D. P., Moyer, S. A. & Banerjee, A. K. (1974), 'In vitro synthesis of methylated messenger rna by the virion-associated rna polymerase of vesicular stomatitis virus.', Cell 3(4), 327–333.
- Ribeiro, E. A., Favier, A., Gerard, F. C. A., Leyrat, C., Brutscher, B., Blondel, D., Ruigrok, R. W. H., Blackledge, M. & Jamin, M. (2008), 'Solution structure of the c-terminal nucleoprotein-rna binding domain of the vesicular stomatitis virus phosphoprotein.', J Mol Biol 382(2), 525–538.
 URL: http://dx.doi.org/10.1016/j.jmb.2008.07.028
- Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A. & Honig, B. (2002), 'Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies : applications to the molecular systems and geometric objects.', J Comput Chem 23(1), 128–137. URL: http://dx.doi.org/10.1002/jcc.1161
- Roche, S. & Gaudin, Y. (2004), 'Evidence that rabies virus forms different kinds of fusion machines with different ph thresholds for fusion.', J Virol 78(16), 8746–8752. URL: http://dx.doi.org/10.1128/JVI.78.16.8746-8752.2004
- Rose, J. K. (1975), 'Heterogneeous 5'-terminal structures occur on vesicular stomatitis virus mrnas.', J Biol Chem 250(20), 8098–8104.
- Rose, J. K. (1978), 'Complete sequences of the ribosome recognition sites in vesicular stomatitis virus mrnas : recognition by the 40s and 80s complexes.', Cell 14(2), 345–353.
- Rose, J. K. & Iverson, L. (1979), 'Nucleotide sequences from the 3'-ends of vesicular stomatitis virus mrna's as determined from cloned dna.', J Virol **32**(2), 404–411.
- Rossmann, M. G. (1990), 'The molecular replacement method', Acta Crystallographica Section A 46(2), 73–82. URL: http://dx.doi.org/10.1107/S0108767389009815
- Rudolph, M. G., Kraus, I., Dickmanns, A., Eickmann, M., Garten, W. & Ficner, R. (2003), 'Crystal structure of the borna disease virus nucleoprotein.', *Structure* 11(10), 1219–1226.
- Ruigrok, R. W., Crépin, T. & Kolakofsky, D. (2011), 'Nucleoproteins and nucleocapsids of negative-strand rna viruses.', Curr Opin Microbiol 14(4), 504–510. URL: http://dx.doi.org/10.1016/j.mib.2011.07.011
- Ruigrok, T. J. H., Pijpers, A., Goedknegt-Sabel, E. & Coulon, P. (2008), 'Multiple cerebellar zones are involved in the control of individual muscles : a retrograde transneuronal tracing study with rabies virus in the rat.', Eur J Neurosci 28(1), 181–200. URL: http://dx.doi.org/10.1111/j.1460-9568.2008.06294.x
- Rupprecht, CE Smith, J., Fekadu, M. & Childs, J. (1995), 'The ascension of wildliffe rabies : a cause for public health concern or intervention?', Emerg Infect Dis 1(4), 107–114. URL: http://www.ncbi.nlm.nih.gov/pubmed/15486256

- Sabeta, C. T., Markotter, W., Mohale, D. K., Shumba, W., Wandeler, A. I. & Nel, L. H. (2007), 'Mokola virus in domestic mammals, south africa.', Emerg Infect Dis 13(9), 1371–1373.
- Samal, S. K. & Collins, P. L. (1996), 'Rna replication by a respiratory syncytial virus rna analog does not obey the rule of six and retains a nonviral trinucleotide extension at the leader end.', J Virol 70(8), 5075–5082.
- Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. (1995), 'Universal catalytic domain structure of adomet-dependent methyltransferases.', J Mol Biol 247(1), 16–20. URL: http://dx.doi.org/10.1006/jmbi.1994.0117
- Schnell, M., Buonocore, L., Boritz, E., Ghosh, H., Chernish, R. & Rose, J. (1998), 'Requirement for a non-specific glycoprotein cytoplasmic domain sequence to drive efficient budding of vesicular stomatitis virus.', EMBO J. 17(5), 1289–1296.
- Schnell, M. J., Buonocore, L., Whitt, M. A. & Rose, J. K. (1996), 'The minimal conserved transcription stop-start signal promotes stable expression of a foreign gene in vesicular stomatitis virus.', J Virol 70(4), 2318–2323.
- Schnell, M. J., McGettigan, J. P., Wirblich, C. & Papaneri, A. (2010), 'The cell biology of rabies virus : using stealth to reach the brain.', Nat Rev Microbiol 8(1), 51–61. URL: http://dx.doi.org/10.1038/nrmicro2260
- Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D. & Ruigrok, R. W. (2001), 'Structure of recombinant rabies virus nucleoprotein-rna complex and identification of the phosphoprotein binding site.', J Virol 75(1), 490–498. URL: http://dx.doi.org/10.1128/JVI.75.1.490-498.2001
- Schoehn, G., Mavrakis, M., Albertini, A., Wade, R., Hoenger, A. & Ruigrok, R. W. H. (2004), 'The 12 a structure of trypsin-treated measles virus n-rna.', J Mol Biol 339(2), 301–312. URL: http://dx.doi.org/10.1016/j.jmb.2004.03.073
- Schubert, M., Harmison, G. G. & Meier, E. (1984), 'Primary structure of the vesicular stomatitis virus polymerase (l) gene : evidence for a high frequency of mutations.', J Virol 51(2), 505–514.
- Schubert, M., Keene, J. D., Herman, R. C. & Lazzarini, R. A. (1980), 'Site on the vesicular stomatitis virus genome specifying polyadenylation and the end of the l gene mrna.', J Virol **34**(2), 550–559.
- Sellers, R. & Maarouf, A. (1990), 'Trajectory analysis of winds and vesicular stomatitis in north america, 1982-5.', *Epidemiol Infect.* **104**(2), 313–328.
- Shatkin, A. J. & Sipe, J. D. (1968), 'Rna polymerase activity in purified reoviruses.', Proc Natl Acad Sci U S A 61(4), 1462–1469.
- Shioda, T., Iwasaki, K. & Shibuta, H. (1986), 'Determination of the complete nucleotide sequence of the sendai virus genome rna and the predicted amino acid sequences of the f, hn and l proteins.', *Nucleic Acids Res* **14**(4), 1545–1563.
- Shuman, S. (2001), 'Structure, mechanism, and evolution of the mrna capping apparatus.', Prog Nucleic Acid Res Mol Biol 66, 1-40.
- Simonsen, C. C., Batt-Humphries, S. & Summers, D. F. (1979), 'Rna synthesis of vesicular stomatitis virus-infected cells : in vivo regulation of replication.', J Virol **31**(1), 124–132.
- Simpson, R. W. & Hauser, R. E. (1966), 'Structural components of vesicular stomatitis virus.', Virology 29(4), 654-667.
- Sleat, D. E. & Banerjee, A. K. (1993), 'Transcriptional activity and mutational analysis of recombinant vesicular stomatitis virus rna polymerase.', J Virol 67(3), 1334–1339.
- Sokol, F., Schlumberger, H. D., Wiktor, T. J. & Koprowski, H. (1969), 'Biochemical and biophysical studies on the nucleocapsid and on the rna of rabies virus.', Virology 38(4), 651–665.
- Soria, M., Little, S. P. & Huang, A. S. (1974), 'Characterization of vesicular stomatitis virus nucleocapsids. i. complementary 40 s rna molecules in nucleocapsids.', *Virology* **61**(1), 270–280.
- Srinivasan, A., Burton, E. C., Kuehnert, M. J., Rupprecht, C., Sutker, W. L., Ksiazek, T. G., Paddock, C. D., Guarner, J., Shieh, W.-J., Goldsmith, C., Hanlon, C. A., Zoretic, J., Fischbach, B., Niezgoda, M., El-Feky, W. H., Orciari, L., Sanchez, E. Q., Likos, A., Klintmalm, G. B., Cardo, D., LeDuc, J., Chamberland, M. E., Jernigan, D. B., Zaki, S. R. & in Transplant Recipients Investigation Team, R. (2005), 'Transmission of rabies virus from an organ donor to four transplant recipients.', N Engl J Med 352(11), 1103–1111.
- Steinhauer, D. A. & Holland, J. J. (1986), 'Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral rna.', J Virol 57(1), 219–228.

- Stillman, E. A. & Whitt, M. A. (1999), 'Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription.', J Virol 73(9), 7199–7209.
- Strous, G., Willemsen, R., van Kerkhof, P., Slot, J., Geuze, H. & Lodish, H. (1983), 'Vesicular stomatitis virus glycoprotein, albumin, and transferrin are transported to the cell surface via the same golgi vesicles.', J Cell Biol. 97(6), 1815–1822.
- Stähli, C., Staehelin, T., Miggiano, V., Schmidt, J. & Häring, P. (1980), 'High frequencies of antigen-specific hybridomas : dependence on immunization parameters and prediction by spleen cell analysis.', J Immunol Methods 32(3), 297–304.
- Summers, W. C. & Siegel, R. B. (1970), 'Transcription of late phage rna by t7 rna polymerase.', Nature 228(5277), 1160-1162.
- Sun, X., Yau, V. K., Briggs, B. J. & Whittaker, G. R. (2005), 'Role of clathrin-mediated endocytosis during vesicular stomatitis virus entry into host cells.', Virology 338(1), 53–60. URL: http://dx.doi.org/10.1016/j.virol.2005.05.006
- Sutton, G., Grimes, J. M., Stuart, D. I. & Roy, P. (2007), 'Bluetongue virus vp4 is an rna-capping assembly line.', Nat Struct Mol Biol 14(5), 449–451. URL: http://dx.doi.org/10.1038/nsmb1225
- Swinteck, B. D. & Lyles, D. S. (2008), 'Plasma membrane microdomains containing vesicular stomatitis virus m protein are separate from microdomains containing g protein and nucleocapsids.', J Virol 82(11), 5536–5547. URL: http://dx.doi.org/10.1128/JVI.02407-07
- 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(2), 821–826. URL: http://dx.doi.org/10.1006/viro.1995.1219
- Takacs, A. M., Barik, S., Das, T. & Banerjee, A. K. (1992), 'Phosphorylation of specific serine residues within the acidic domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro.', J Virol 66(10), 5842–5848.
- Takacs, A. M., Das, T. & Banerjee, A. K. (1993), 'Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system.', *Proc Natl Acad Sci U S A* **90**(21), 10375–10379.
- Takacs, A. M., Perrine, K. G., Barik, S. & Banerjee, A. K. (1991), 'Alteration of specific amino acid residues in the acidic domain i of vsv phosphoprotein (p) converts a gal4-p(i) hybrid into a transcriptional activator.', New Biol 3(6), 581–591.
- Tang, Y., Rampin, O., Giuliano, F. & Ugolini, G. (1999), 'Spinal and brain circuits to motoneurons of the bulbospongiosus muscle : retrograde transneuronal tracing with rabies virus.', *J Comp Neurol* **414**(2), 167–192.
- Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R. W. & Burmeister, W. P. (2000), 'On the domain structure and the polymerization state of the sendai virus p protein.', Virology 266(1), 99–109. URL: http://dx.doi.org/10.1006/viro.1999.0066
- Tarbouriech, N., Curran, J., Ruigrok, R. W. & Burmeister, W. P. (2000), 'Tetrameric coiled coil domain of sendai virus phosphoprotein.', Nat Struct Biol 7(9), 777–781. URL: http://dx.doi.org/10.1038/79013
- Tawar, R. G., Duquerroy, S., Vonrhein, C., Varela, P. F., Damier-Piolle, L., Castagné, N., MacLellan, K., Bedouelle, H., Bricogne, G., Bhella, D., Eléouët, J.-F. & Rey, F. A. (2009), 'Crystal structure of a nucleocapsid-like nucleoprotein-rna complex of respiratory syncytial virus.', Science 326(5957), 1279–1283. URL: http://dx.doi.org/10.1126/science.1177634
- Tekes, G., Rahmeh, A. A. & Whelan, S. P. J. (2011), 'A freeze frame view of vesicular stomatitis virus transcription defines a minimal length of rna for 5' processing.', *PLoS Pathog* 7(6), e1002073. URL: http://dx.doi.org/10.1371/journal.ppat.1002073
- Tesh, R. B., Boshell, J., Modi, G. B., Morales, A., Young, D. G., Corredor, A., de Carrasquilla, C. F., de Rodriguez, C., Walters, L. L. & Gaitan, M. O. (1987), 'Natural infection of humans, animals, and phlebotomine sand flies with the alagoas serotype of vesicular stomatitis virus in colombia.', Am J Trop Med Hyg 36(3), 653–661.
- Tesh, R. B., Chaniotis, B. N. & Johnson, K. M. (1972), 'Vesicular stomatitis virus (indiana serotype) : transovarial transmission by phlebotomine sandlies.', *Science* **175**(29), 1477–1479.
- Testa, D. & Banerjee, A. K. (1977), 'Two methyltransferase activities in the purified virions of vesicular stomatitis virus.', J Virol 24(3), 786–793.

- Testa, D., Chanda, P. K. & Banerjee, A. K. (1980), 'Unique mode of transcription in vitro by vesicular stomatitis virus.', *Cell* **21**(1), 267–275.
- Thomas, D., Newcomb, W. W., Brown, J. C., Wall, J. S., Hainfeld, J. F., Trus, B. L. & Steven, A. C. (1985), 'Mass and molecular composition of vesicular stomatitis virus : a scanning transmission electron microscopy analysis.', J Virol 54(2), 598–607.
- Thoulouze, M., Bouguyon, E., Carpentier, C. & Brémont, M. (2004), 'Essential role of the nv protein of novirhabdovirus for pathogenicity in rainbow trout.', J Virol **78**(8), 4098–4107.
- Tirawatnpong, S., Hemachudha, T., Manutsathit, S., Shuangshoti, S., Phanthumchinda, K. & Phanuphak, P. (1989), 'Regional distribution of rabies viral antigen in central nervous system of human encephalitic and paralytic rabies.', *J Neurol Sci* **92**(1), 91–99.
- Tordo, N. & Kouknetzoff, A. (1993), 'The rabies virus genome : an overview.', Onderstepoort J Vet Res 60(4), 263–269.
- Tordo, N., Poch, O., Ermine, A., Keith, G. & 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), 3914–3918.
- Tordo, N., Poch, O., Ermine, A., Keith, G. & 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), 565–576.
- Toriumi, H., Honda, Y., Morimoto, K., Tochikura, T. S. & Kawai, A. (2002), 'Structural relationship between nucleocapsid-binding activity of the rabies virus phosphoprotein (p) and exposure of epitope 402-13 located at the c terminus.', *J Gen Virol* 83(Pt 12), 3035–3043.
- Toriumi, H. & Kawai, A. (2004), 'Association of rabies virus nominal phosphoprotein (p) with viral nucleocapsid (nc) is enhanced by phosphorylation of the viral nucleoprotein (n).', *Microbiol Immunol* **48**(5), 399–409.
- Ugolini, G. (1995), 'Specificity of rabies virus as a transneuronal tracer of motor networks : transfer from hypoglossal motoneurons to connected second-order and higher order central nervous system cell groups.', J Comp Neurol 356(3), 457–480. URL: http://dx.doi.org/10.1002/cne.903560312
- Uversky, V. N. (1993), 'Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule.', *Biochemistry* **32**(48), 13288–13298.
- Vagin, A. & Teplyakov, A. (2000), 'An approach to multi-copy search in molecular replacement', Acta Crystallographica Section D 56(12), 1622–1624. URL: http://dx.doi.org/10.1107/S0907444900013780
- Vidy, A., Bougrini, J. E., Chelbi-Alix, M. K. & Blondel, D. (2007), 'The nucleocytoplasmic rabies virus p protein counteracts interferon signaling by inhibiting both nuclear accumulation and dna binding of stat1.', J Virol 81(8), 4255–4263. URL: http://dx.doi.org/10.1128/JVI.01930-06
- Vidy, A., Chelbi-Alix, M. & Blondel, D. (2005), 'Rabies virus p protein interacts with stat1 and inhibits interferon signal transduction pathways.', J Virol 79(22), 14411–14420. URL: http://dx.doi.org/10.1128/JVI.79.22.14411-14420.2005
- Villarreal, L. P., Breindl, M. & Holland, J. J. (1976), 'Determination of molar ratios of vesicular stomatitis virus induced rna species in bhk21 cells.', *Biochemistry* 15(8), 1663–1667.
- Wagner, R. R., Schnaitman, T. C., Snyder, R. M. & Schnaitman, C. A. (1969), 'Protein composition of the structural components of vesicular stomatitis virus.', J Virol 3(6), 611–618.
- Wang, J. T., McElvain, L. E. & Whelan, S. P. J. (2007), 'Vesicular stomatitis virus mrna capping machinery requires specific cisacting signals in the rna.', J Virol 81(20), 11499–11506. URL: http://dx.doi.org/10.1128/JVI.01057-07
- Wertz, G. W. (1978), 'Isolation of possible replicative intermediate structures from vesicular stomatitis virus-infected cells.', Virology **85**(1), 271–285.
- Wertz, G. W., Whelan, S., LeGrone, A. & Ball, L. A. (1994), 'Extent of terminal complementarity modulates the balance between transcription and replication of vesicular stomatitis virus rna.', *Proc Natl Acad Sci U S A* **91**(18), 8587–8591.
- Whelan, S. P. J. & 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), 9178–9183. URL: http://dx.doi.org/10.1073/pnas.152155599

- Whelan, S. P. & Wertz, G. W. (1999a), 'The 5' terminal trailer region of vesicular stomatitis virus contains a position-dependent cis-acting signal for assembly of rna into infectious particles.', J Virol 73(1), 307–315.
- Whelan, S. P. & Wertz, G. W. (1999b), '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.
- White, J., Matlin, K. & Helenius, A. (1981), 'Cell fusion by semliki forest, influenza, and vesicular stomatitis viruses.', J Cell Biol 89(3), 674–679.
- Whitt MA, Chong L, R. J. (1989), 'Glycoprotein cytoplasmic domain sequences required for rescue of a vesicular stomatitis virus glycoprotein mutant.', J Virol 63(9), 3569–3578.
- Wu, S. & Zhang, Y. (2007), 'Lomets : a local meta-threading-server for protein structure prediction.', Nucleic Acids Res **35**(10), 3375–3382.

URL: http://dx.doi.org/10.1093/nar/gkm251

- Wu, X., Lei, X. & Fu, Z. F. (2003), 'Rabies virus nucleoprotein is phosphorylated by cellular casein kinase ii.', *Biochem Biophys Res* Commun **304**(2), 333–338.
- Yang, J., Hooper, D. C., Wunner, W. H., Koprowski, H., Dietzschold, B. & Fu, Z. F. (1998), 'The specificity of rabies virus rna encapsidation by nucleoprotein.', Virology 242(1), 107–117. URL: http://dx.doi.org/10.1006/viro.1997.9022
- Yang, J., Koprowski, H., Dietzschold, B. & Fu, Z. F. (1999), 'Phosphorylation of rabies virus nucleoprotein regulates viral rna transcription and replication by modulating leader rna encapsidation.', J Virol 73(2), 1661–1664.
- Yap, L. J., Luo, D., Chung, K. Y., Lim, S. P., Bodenreider, C., Noble, C., Shi, P.-Y. & Lescar, J. (2010), 'Crystal structure of the dengue virus methyltransferase bound to a 5'-capped octameric rna.', *PLoS One* 5(9). URL: http://dx.doi.org/10.1371/journal.pone.0012836
- Yap, T. L., Xu, T., Chen, Y.-L., Malet, H., Egloff, M.-P., Canard, B., Vasudevan, S. G. & Lescar, J. (2007), 'Crystal structure of the dengue virus rna-dependent rna polymerase catalytic domain at 1.85-angstrom resolution.', J Virol 81(9), 4753–4765. URL: http://dx.doi.org/10.1128/JVI.02283-06
- Ye, Q., Krug, R. M. & Tao, Y. J. (2006), 'The mechanism by which influenza a virus nucleoprotein forms oligomers and binds rna.', Nature 444(7122), 1078–1082. URL: http://dx.doi.org/10.1038/nature05379
- Yilma, T Breeze, R., Ristow, S., Gorham, J. & Leib, S. (1985), 'Immune responses of cattle and mice to the g glycoprotein of vesicular stomatitis virus.', Vet Clin North Am Food Anim Pract. 185, 101–115.
- Yusoff, K., Millar, N. S., Chambers, P. & Emmerson, P. T. (1987), 'Nucleotide sequence analysis of the l gene of newcastle disease virus : homologies with sendai and vesicular stomatitis viruses.', Nucleic Acids Res 15(10), 3961–3976.
- Zafrullah, M., Ozdener, M. H., Panda, S. K. & Jameel, S. (1997), 'The orf3 protein of hepatitis e virus is a phosphoprotein that associates with the cytoskeleton.', J Virol 71(12), 9045–9053.
- Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., Guo, Y., Bernard, K. A., Shi, P.-Y. & Li, H. (2007), 'Structure and function of flavivirus ns5 methyltransferase.', J Virol 81(8), 3891–3903. URL: http://dx.doi.org/10.1128/JVI.02704-06

Annexe A ANNEXE

La diffusion de neutrons pour l'étude des macromolécules

A.1 Neutrons en Biologie

Les mouvements atomiques dans une protéine sont essentiels à la fonction biologique. Certaines caractéristiques des neutrons, font de ces particules une sonde d'intérêt majeur pour étudier ces mouvements. Compte tenu de son énergie et de son absence de charge, son interaction avec l'échantillon est non destructive.

Durant mes trois ans de thèse j'ai utilisé les neutrons comme un outil pour l'étude des protéines virales. Tout d'abord je vais commencer par la diffusion de neutrons aux petits angles - cette technique permet d'obtenir des informations structurales à basse résolution. Elle est idéale pour l'étude de grosses molécules flexibles qui ne peuvent pas être caractérisées par RMN ou cristallographie. Egalement la technique de variation de contraste est idéale pour l'étude de complexes protéiques et des changements de conformations qui y parviennent au cours des processus cellulaires.

A.1.1 La diffusion des neutrons

Un échantillon biologique est composé d'une multitude de noyaux différents qui sont hétérogènes par la présence d'isotopes différents. Les neutrons ont la particularité d'interagir non pas avec les électrons du nuage électronique mais avec les noyaux des atomes ce qui les rend utiles pour l'étude des macromolécules.

La diffusion des neutrons peut être divisée en deux parties : la diffusion *cohérente* et la diffusion *incohérente*.

- La diffusion *cohérente* nous renseigne sur le comportement collectif des atomes. Les ondes sphériques diffusées par les différents noyaux interfèrent entre elles. C'est cette diffusion qui est associée à l'obtention d'un modèle à basse résolution par diffusion de Rayons X ou neutrons.
- La diffusion *incohérente*, les neutrons interagissent indépendamment avec chaque noyau, les ondes diffusées n'interfèrent pas. La diffusion *incohérente* peut nous renseigner sur les mouvements individuels des atomes i.e. la dynamique de la protéine d'intérêt.

La diffusion *incohérente* peut être élastique, i.e. sans perte d'énergie. Elle permet, dans ce cas, d'extraire la géométrie du mouvement atomique. Les neutrons peuvent aussi être diffusés de façon quasiélastique ou inélastique. Il y a alors transfert d'énergie entre le neutron et le noyau. L'information obtenue renseigne sur l'amplitude et la fréquence des mouvements.



FIGURE A.1 – **Diffusion** cohérente de certains atomes.

A. Représente la diffusion cohérente des atomes constituant les protéines obtenus par Rayons X
B. Représente la diffusion cohérente des même atomes par diffusion de neutrons.

La figure A.1 indique la différence de la diffusion *cohérente* du proton (H) en Rayons X et en neutrons. Le neutron interagissant avec le noyau des atomes et non pas avec les électrons donne un signal plus important que le proton. Egalement le Deutérium (D) ayant un neutron en plus que H dans le noyau aurait comme propriété de diffuser plus intensement les neutrons.

A chaque élément constituant les protéines est associée une longeur de diffusion (b) qui traduit l'intensité de l'interaction neutron-atome. Cette longueur de diffusion est reliée à la section efficace (σ) par la relation $\sigma = 4\pi b^2$. A chaque élement sont associées une section efficace *cohérente* et une section efficace *incohérente*.

Le tableau suivant nous donne les valeurs des longueurs de diffusion des éléments les plus abondants constituant les protéines.

Element	$b_{coh} \ (10^{-15} \mathrm{m})$	$b_{inc} \ (10^{-15} \mathrm{m})$	σ_{coh} (barns)	σ_{inc} (barns)	σ_{tot} (barns)
Н	-3.74	25.21	1.76	79.91	81.67
D	6.67	4.03	5.60	2.04	7.64
С	6.65	0	5.55	0	5.55
Ν	9.37	1.98	11.01	00.49	11.52
0	5.81	0	4.24	0	4.24

TABLE A.1 – Sections efficaces et longueurs de diffusion des principaux éléments composant les protéines.

On peut noter deux choses :

- La section efficace *incohérente* de l'hydrogène est presque deux ordres de grandeur plus élevée que celle des autres éléments. Les spectres de diffusion *incohérente* seront donc dominés par cette contribution.
- La longueur de diffusion *cohérente* de l'hydrogène est négative alors que celle du deutérium est positive. Cette différence est à la base de la technique de variation de contraste en diffusion de neutrons aux petits angles.

A.1.1.1 Diffusion aux petits angles

La diffusion aux petits angles est une technique exploitant la diffusion *cohérente* d'un échantillon, elle nous donne donc une information structurale. Elle est basée sur des phénomènes d'interférence.

L'intensité de neutrons diffusée (I) en fonction de l'angle (Q) nous permet de connaitre certains paramètres de la macromolécule en solution comme son rayon de giration (Rg) et la masse moléculaire (MW).

A.1.1.2 La variation de contraste

Lorsqu'on a une solution de N particules identiques n'intéragissant pas entre elles, l'intensité mesurée est due au contraste des particules par rapport au solvant. Le solvant peut être considéré comme un milieu ayant une amplitude de diffusion constante.

La technique de variation de contraste s'avère donc très intéressante lorsque l'on utilise un rayonnement neutronique : en effet, l'amplitude de diffusion des isotopes de l'hydrogène sont très différentes (voir tableau A.1). On peut donc, dans certaines proportions de H_2O/D_2O de rendre invisible une protéine donnée. Cette technique est très intéressante lorsque l'on a des complexes protéines-protéines, car on peut produire une version deutérée d'une protéine et former un complexe avec une protéine non deutérée. A certains rapports H_2O/D_2O , on va donc faire "disparaître" la protéine-D ou la protéine-H (figure A.2).



FIGURE A.2 – Longueurs de diffusion de différents composants en fonction du pourcentage de D_2O .

La ligne bleue représente la densité de diffusion d'une protéine deutérée (protéine-D) allant de 0 à 100 % de D₂O dans le solvent. La ligne rouge est la diffusion de l'eau et la ligne verte la diffusion d'une protéine hydrogénée (protéine-H) dans un solvent d'eau allant de 0 à 100 % D₂O.

Expérimentalement, on enregistre des courbes d'un complexe protéine-D/protéine-H à différentes proportions H_2O/D_2O .

Dans le laboratoire, nous nous sommes intéressés à étudier le complexe N°P par variation de contraste. En collaboration étroite avec le laboratoire DLAB de l'Institut Laue Langevin nous avons produit en système d'expression bactérien la phosphoprotéine sous forme deutérée (protéine-D) (Dr. Martin Moulin, DLAB, ILL, Grenoble).

Les échantillons du complexe N°P (nucléoprotéine est hydrogénée et la phosphoprotéine est deutérée) ont été produits par Filip Yabukarski. Toutes les protéines sont exprimées dans des cellules *E. coli* BL21. Le domaine N-terminal (P-NTD)de la phosphoprotéine a été purifié sur colonne Ni-NTA puis par chromatographie d'exclusion de taille. Un mutant de la nucléoprotéine (N Δ 21) (Leyrat, Yabukarski, Tarbouriech, Ribeiro, Jensen, Blackledge, Ruigrok & Jamin 2011) fusioné avec un tag MBP clivable à la TEV a été purifié sur colonne d'amylose. Ces deux échantillons sont mélangés, puis purifiés sur colonne Ni-NTA et colonne d'exclusion de taille. La phosphoprotéine entière, produite en milieu deutéré (au D-lab de l'ILL) est aussi purifiée sur colonne Ni-NTA suivie d'une chromatographie d'exclusion de taille. Le complexe N°P_{FL} est obtenu en mélangeant le complexe N°P_{NTD} avec la phosphoprotéine entière dans un ratio 1 pour 5. Ce complexe est ensuite passé sur une colonne d'amylose et incubé pendant 40h avec la protéase TEV et purifié sur colonne Ni-NTA, suivi d'une chromatographie d'exclusion de taille. Le complexe N°P_{FL} a ensuite été dialysé dans des concentrations de 20, 40 et 100% de D₂O.

La collecte des données a été réalisée sur l'instrument D22 de l'ILL à deux distances échantillondétecteur différentes. L'intégration radiale et la soustraction du tampon et du bruit électronique du détecteur ont été faites avec le programme SPOLLY.

La modélisation et l'analyse des ensembles ont été générés par le Dr. Cédric Leyrat à partir des structures haute résolution du domaine central de P (P-CED) (Ding et al. 2006), du domaine C-terminal de P (P-CTD) (Ribeiro et al. 2008) du complexe N-ARN/P-CTD (Green et al. 2006), et du complexe N⁰-P. Les linkers flexibles ont été générés avec le programme Gajoe (Bernadó et al. 2007), et les différentes parties ont été reliées entre elles manuellement.

Ensuite, des courbes de diffusion théoriques ont été calculées pour chaque modèle avec le programme Cryson dans les 4 conditions de contraste (0, 20 40 et 100% de D_2O). Le programme Gajoe (Bernadó et al. 2007) a été utilisé pour faire une sélection et trouver un ensemble permettant de reproduire les données expérimentales.

La contribution de mon travail dans l'étude du complexe N^0 -P par variation de contraste était l'acquisition des données sur l'instrument D22. Ensuite Dr. Cédric Leyrat et Nicolas Martinez ont traité les données comme expliquer plus haut. Donc, je ne présenterai que les courbes de diffusion de neutrons ainsi que les modèles préliminaires qui ont été générés par Dr. Cédric Leyrat.

Les courbes de variation de contraste du complexe N⁰-P sont présentées dans la figure A.3 :



FIGURE A.3 – Données expérimentales de diffusion de neutrons aux petits angles du complexe N^0 -P.

Comme indiqué sur la figure la courbe noire représente la diffusion à 0% D₂O, la coube rouge - 20% D₂O, la courbe verte - 40% D₂O et la courbe bleue - 100% D₂O.

Les expériences de SEC-MALLS-RI sur le complexe indiquent que la stoechiométrie du complexe dépend de la manière dont l'échantillon est préparé. Selon le rapport N :P utilisé dans la procédure de reconstitution du complexe, l'expérience de SEC-MALLS-RI indique que l'échantillon contient soit un mélange de 2N par dimère de P, soit 1 N par dimère de P.

Nous avons donc fait deux types de modèles : des modèles avec une stoechiométrie 1N :2P (figure A.4a), et des modèles avec une stoechiométrie 2N :2P (figure A.4b).



(a) Modèle avec 1 seule N par dimère de P
 (b) Modèle avec 2 N par dimère de P
 FIGURE A.4 – Différents types de modèles utilisés lors de la modélisation du complexe N⁰-P.
 Les domaines centraux de P, PCTD, P-NTD et N sont représentés en orange, jaune, bleu et rouge respectivement.

Les ensembles sélectionnées pour chaque condition de contraste (figure A.3) contiennent une forte proportion de modèles avec une stoechiométrie de 1 molécule de N par dimère de P correspondant approximativement à 70 % des modèles.

Si l'on se repère par rapport au domaine central, on observe que la zone explorée par la nucléoprotéine est très large, tout comme la partie N-terminale de la P libre. Aussi, on remarque que N est plutôt éloignée de la partie N-terminale du dimère libre (figure A.4a). Le domaine C-terminal

de P a tendance à se retrouver proche du domaine central, mais plûtot dans une location opposée à la partie N-terminale, avec une distance moyenne de 9 nm entre leurs centres de masse.

Pour les modèles ayant deux N par dimère de P, qui représentent donc environ 30% de l'ensemble, on obtient des modèles ou les 2 N sont plus proches (figure A.4b), la distance moyenne entre les centres de masses des parties N-terminales de P est de 11 nm, contre 15 nm lorsque l'on a une seule N. La partie C-terminale de P est là aussi à 9nm en moyenne de la partie N-terminale.

Ces résulats nous montrent que N explore un espace conformationnel très large. Ce volume très large exploré par N et la partie C-terminale de N pourrait permettre au complexe, si celui-ci est fixé à une nucléocapside de faciliter l'encapsidation d'un ARN néo-synthétisé, mais aussi permettre à la partie C-terminale de P de chercher un monomère de N distant de la nucléocapside et de libérer l'ARN pour qu'il soit pris en charge par la polymérase.

A.1.1.3 Etude de la dynamique de la phosphoprotéine et sa partie globulaire C-terminale

La deuxième partie de ce chapitre est consacrée à l'étude de la dynamique de la phosphoprotéine du virus de la rage par diffusion élastique *incohérente*.

Lorsqu'un échantillon de protéine se trouve placé dans un faisceau de neutrons, l'intensité diffusée permet d'analyser ses mouvements sur des temps moléculaires. La dépendance en température des mouvements atomiques caractérise le comportement dynamique de la protéine, et permet de corréler celui-ci à la fonction et à l'activité biologique de la macromolécule étudiée.

La dynamique est quantifiée par deux paramètres : le déplacement atomique carré moyen (u^2) , qui rend compte de la flexibilité de la structure, et la constante de force (k'), appelée « résilience », qui informe sur la rigidité du système. L'intensité *incohérente* diffusée élastiquement a une forme gaussienne pour des mouvements localisés dans la fenêtre de l'instrument. L'intensité gaussienne dépend du déplacement atomique carré moyen. Ce premier paramètre peut être directement extrait des mesures expérimentales. Une analyse en terme de constante de force effective (k') a été proposée par (**Bicout** & Zaccai 2001). Ceci est la force résultante agissant sur un atome lorsqu'il s'éloigne de sa position d'équilibre peut s'exprimer sous la forme d'une force de rappel due à son environnement. La constante de force (k') est inversement proportionnelle à la pente de la dépendance de température du déplacement atomique carré moyen.

Les protéines sont très sensibles à la température. Si elles sont rigides à très basses températures, elles deviennent flexibles à température physiologique (Doster et al. 1989). En dessous de 200 K environ, elles se comportent comme des solides harmoniques. Les atomes sont fermement ancrés dans la structure, bloqués dans un état énergétique donné. Ils vibrent autour de leur position d'équilibre. Au-delà de 200 K, ils acquièrent une énergie d'activation suffisante et sont susceptibles de sauter d'un puits énergétique dans un autre. Ainsi, la protéine peut changer de sous-état conformationnel. La température permet cette transition dynamique.

Le neutron est sensible à l'atome d'hydrogène, il le « voit » 40 fois plus que celui de tout autre atome (figure A.5).



A l'Institut Laue Langevin (ILL), nous disposons du réacteur à haut flux neutronique. Plusieurs instruments sont dédiés à l'étude de la dynamique des protéines couvrant des temps de mouvements allant de la picoseconde à 100 ns.

Le choix du spectromètre détermine la gamme de temps et d'amplitude des mouvements étudiés.

La fenêtre en temps est déterminée par la résolution instrumentale en énergie et l'amplitude par les angles de diffusion accessibles aux détecteurs.

Pour l'étude de la dynamique de la phosphoprotéine (P) et son domaine globulaire (C-terminal), nous avons utilisé le spectromètre à rétrodiffusion IN13 (figure A.6).



FIGURE A.6 – Principe du spectromètre à rétrodiffusion IN13 de l'ILL.

Dans le cas d'une expérience de diffusion élastique, le faisceau de neutron est envoyé sur un monochromateur qui sélectionne la longueur d'onde incidente. Un déflecteur en graphite focalise ensuite le faisceau monochromateur sur l'échantillon. Les neutrons diffusés par l'échantillon sont analysés par des cristaux analyseurs. Ceux-ci sélectionnent ceux qui ont conservé leur énergie et les rétrodiffusent avec des angles de diffraction de 90°. Les neutrons sont finalement détectés par un jeu de 32 détecteurs à ³He. IN13 a une résolution en énergie de $8\mu eV$, qui permet d'avoir accès à des temps caractéristiques d'environ 100ps. Il permet de couvrir une gamme large d'amplitude de mouvement.

L'instrument IN13 permet de détecter des mouvements atomiques de quelques Angströms sur une centaine de picoseconde. Sur ces gammes de temps et d'amplitude, les atomes d'hydrogène ancrés dans la structure reflètent les mouvements des chaines latérales et principales auxquels ils sont liés, donc ils reflètent la dynamique globale de la protéine. Par ailleurs, les atomes d'hydrogène du solvent, qui pourraient fortement dominer le signal, diffusent hors de la fenêtre et ne contribuent pas à l'intensité mesurée.

Comme décrit dans le chapitre 2, la phosphoprotéine de *Rhabdoviridae* est une protéine intrinséquement dépliée avec une région N-terminale désordonnée, un domaine central et C-terminal structurés qui sont interconnectés par des régions désordonnées (figure A.7).



FIGURE A.7 – Représentation schématique de la phosphoprotéine et son domaine Cterminal.

A. Représente la protéine entière avec ses régions désordonnées en trait marron, en jaune, le domaine central responsable à la dimérisation et en rouge, le domaine Cterminal.

B. Représente la structure cristallographique du domaine Cterminal globulaire, responsable à la liaison à la matrice N-ARN. Les lettres N et C signifient N-terminal et C-terminal respectivement. Les hélices α sont représentées en rouge et les brins β en vert.

On a donc deux échantillons avec des propriétés différentes :

- le domaine C-terminal est globulaire constitué exclusivement de résidues ordonnés. Il forme une unité de repliement autonome de 9 kDa dont la structure est connue.
- La protéine entière possédant de nombreux sites de fixations différents : à N-ARN par sa partie C-terminal et à N° par sa partie N-terminal. Elle forme des dimères de 69 kDa. Sa structure désordonnée intercoupée par des domaine structurés (figure A.7A) suggèrent des phénomènes d'interaction avec des différents partenaires cellulaires ou viraux.

Par mesure de diffusion élastique *incohérente* de neutrons, nous avons essayé de comparer la dynamique des deux protéines. Notre hypothèse est que la protéine entière doit être plus mobile qu'une protéine globulaire afin d'interagir avec ses parteniares.

Les deux protéines ont été produits en système d'expression bactérien (*E.coli*). Les cellules ont été ensuite lysées par sonication, et la fraction soluble passée sur une colonne Ni-NTA. Les fractions contenant la protéine ont été vérifiées par SDS-PAGE, rassemblées, concentrées et passées sur gel filtration. Ensuite, les fractions de gel filtration contenant la protéine ont été concentrées puis dialysées dans un tampon. La protéine est liophylisée et séchée dans un dessicateur avec du P_2O_5 pendant 48 heures. Ensuite environ 200 mg de protéine par échantillon sont hydratés avec D_2O à 0.4g/g de manière à obtenir une couche d'hydratation autour de la protéine. Après hydratation la cellule d'alluminium est sélée. La mesure sur l'instrument IN13 est effectuée de 20K à 300K par des palliers de 10K.

L'analyse des résultats a été effectuée par Dr. Andreas Stadler.



FIGURE A.8 – Courbes de diffusion élastique *incohérente* sur IN13 du domaine C-terminal et la phosphoprotéine entière.

(a) Représente l'approximation Gaussienne $(\ln I=f(Q^2))$, expliquée au début de A.1.1.3) à différentes température du domaine C-terminal de la phosphoprotéine.

(b) Représente l'approximation Gaussienne pour la phosphoprotéine entière.

Pour plus de clarté, seulement les températures 30K, 80K, 130K, 180K, 230K et 280K sont présentées sur le graphique.

Les déplacements carrés moyens ainsi analysés pour les deux protéines semblent identiquer que la phosphoprotéine entière n'est pas plus flexible que le domaine globulaire.

Suite à cette analyse, plusieurs hypothèses peuvent être émises :

- La phosphoprotéine étant figée dans une poudre hydratée à $0.4g/g D_2O$ ne peut pas explorer des mouvements importants ce qui la rend aussi rigide qu'une protéine globulaire et structurée.
- Une autre hypothèse pourrait suggérer que la phosphoprotéine même ayant des parties désordonnées se comporte comme une protéine rigide.

Nous avons calculé la résilience (k') pour les deux protéines (figure A.9).



FIGURE A.9 – Evolution des déplacements carrés moyens en fonction de la température.

En rouge est présenté la courbe de dynamique du domaine globulaire (C-terminal) et en noir pour la phosphoprotéine entière. Les barres d'erreurs sont représentés pour chaque température sur le graphique.

La résilience exprimée en N/m est calculée par l'inverse da la pente des déplacements carrés moyens.

On peut remarquer que la résilence des deux protéines de 20K à 220K est identique. Dans cette gamme de température non physiologique, la protéine est figée est présente une dynamique rigide. Les faibles mouvements observés sont dûs aux mouvements des atomes autour de leurs centres d'équilibres.

Après 220K on observe un changement de dynamique marqué par une cassure de la courbe. La résilience pour le domaine C-terminal est de 0.112N/m et pour la protéine entière de 0.144N/m indique que les deux protéines se comportent de la même manière en poudre hydratée. Aucune dynamique de la protéine entière n'est observée.

Ces résultats vont à l'encontre de ce qu'on avait émis comme hypothèse que la phosphoprotéine entière doit être plus flexible qu'une protéine globulaire. Comme discuter plus haut, ils pourraient être dûs à l'environnement de poudre hydratée.

Afin de répondre avec plus de certitude, nous devons réaliser des expériences en solution hautement concentré (200-300mg/ml) afin de mimer un environnement cellulaire.

Annexe B ANNEXE

Production d'anticorps monoclonaux de $$\mathrm{VSV}$$

Avant mon arrivé dans le laboratoire aucun anticorps monoclonal pour l'ARN polymérase ARN dépendante (L) n'était disponible. De tels anticorps couplés à des billes d'or pourraient être utilisés pour l'étude structurale des nucléocapsides virales (N, P et L) par microscopie électronique et en particulier la localisation de la polymérase dans ce complexe. Nous avons donc entrepris de générer des anticorps contre les cinq protéines virales. Pour toutes les immunisations de souris nous avons utilisé le virus de la stomatite vésiculaire provenant du laboratoire de Dr. Danielle Blondel (Gif sur Yvettes, Paris).

L'obtention d'anticorps monoclonaux est réalisée par la technique d'hybridome décrite par (Köhler & Milstein 1975). C'est une technique dont le principe est basé sur la fusion d'un lymphocyte B et d'une cellule myélomateuse qui confère l'immortalité à l'hybride. Les anticorps obtenus sont tous identiques puisqu'ils dérivent de la même cellule.

Immunisation des souris

Des souris sont immunisées par voie intra-péritoniale avec 0,50 ml d'une solution contenant $25 \,\mu \text{g}$ d'antigène. La première immunisation a été réalisée avec le virus entier à une concentration de $50 \,\mu \text{g/ml}$. Les deux immunisations suivantes sont réalisées avec les nucléocapsides virales, composées de la nucléoprotéine, la phosphoprotéine et l'ARN polymérase ARN dépendante (L), à une concentration de $10 \,\mu \text{g/ml}$. Afin d'augmenter les chances d'obtention d'hybridomes exprimant des Anticorps monoclonaux de la polymérase, la dernière immunisation est réalisée qu'avec la polymérase virale (L). L'immunisation des souris est réalisée selon la méthode de (Stähli et al. 1980) : l'échantillon d'immunisation est dilué par un volume égal d'adjuvent complet de Freund pour une première injection ou d'ajuvent incomplet pour les injections postérieures. Les différentes immunisations ont été faites à J0, J15, J30 et J90. Chaque souris est saignée par ponction dans la veine caudale, avant immunisation puis à J14, J29 et J89 et les différents sérums sont conservés à -20°C. Les souris répondant positivement aux contrôles sérologiques reçoivent une injection de rappel de 100 μ l de polymérase à une concentration de 10 μ g/ml sans adjuvent, trois jours avant la fusion cellulaire.

Prélèvement des rates

Les souris immunisées sont sacrifiées et leur rate est prélevée stérilement. Après lavage en milieu sans sérum celle-ci est parée à l'aide d'un scalpel puis lavée trois fois de suite avec du milieu complet sans sérum. La suspension cellulaire est homogénéisée en écrasant la rate sous une grille à l'aide du piston d'une seringue. Après, un dernier lavage dans un milieu complet sans sérum, le culot cellulaire est remis en suspension dans un volume de ce milieu et conservé à l'étuve. Après vérification de la viabilité cellulaire par Bleu Trypan et leur dénombrement au microscope optique, les cellules SP2 et les lymphocytes B sont mélangées selon un rapport SP2/B égale 1/5. Le mélange cellulaire est centrifugé comme précédemment et au culot cellulaire 0,5 ml de PEG 4000 sont ajoutés goutte à goutte, sous

agitation douce pendant 60 sec. Après 90 sec d'incubation, la suspension est diluée goutte à goutte dans un milieu sans sérum et centrifugée à 200 g pendant 10 min. Le surnageant est éliminé et 10 ml de milieu complet sont ajoutés au culot cellulaire. Le mélange est homogénéisé délicatement et réparti dans des boites de culture à raison de 10⁷ cellules par boite. 9-100 ml de mileu complet sont ajoutés dans chaque boite et ces dernières sont déposées à l'étuve à 37°C sous 5% de CO_2 .

Répartition dans des plaques de 24 cupules

Après 24 h, le contenu des boites est dilué au demi avec du milieu hypoxanthine-azaserine (HA) et $1ml (10^5 \text{ cellules})$ de cette suspension est déposé dans chaque puits. Les plaque sont portées à l'étuve et l'apparition des hybridomes est observée après un délai de 7 à 21 jours.

Criblage des surnageants

Le criblage des surnageants s'effectue dès l'apparition des hybridomes : le surnageant des puits contenant des colonies sont analysés par Western Blot. La membrane de nitrocellulose (Whatmann) contenant les protéines virales (N, P, G, M et L) est saturée avec une solution PBS-lait 5%. Ensuite 3 ml de chaque surnageant sont incubés pendant 1 h avec une bandelette de membrane nitrocellulose. La membrane est ensuite lavée avec une solution PBS-0.1% Tween20 et réincubée avec un anticorps secondaire anti-Ig-Gs (H+L) (Sigma-Aldrich) totale conjugué à l'HRP. Ensuite la membrane est révélée avec un substrat luminescent (SuperSignal West Pico Chemiluminescent Substrate, Thermo Scientific).

Clonage des hybridomes

Le clonage des hybridomes est effectué par la méthode de dilutions limites. On effectue des dilutions des hybridomes jusqu'à obtenir théoriquement 1 cellule par puits dans des plaques de 96 puits. On laisse les cultures se développer à l'étuve et on surveille l'apparition des colonies. Ces colonies sont remises en culture après avoir vérifié l'activité secrétante des anticorps spécifiques. Les hybridomes positifs sont soit congelés rapidement dans l'azote liquide, soit amplifiés en culture ou produit en ascite chez la souris.

Conclusions

Nous avons obtenus de nombreux hybridomes exprimant des anticorps dirigés contre les cinq protéines du virus. Il faut maintenant les caractériser en détails et mettre au point le marquage avec les billes d'or.

Annexe C ANNEXE

Virus Research 162 (2011) 126-137

Contents lists available at SciVerse ScienceDirect



Virus Research



journal homepage: www.elsevier.com/locate/virusres

Review

Structural insights into the rhabdovirus transcription/replication complex

Ivan Ivanov^{a,b}, Filip Yabukarski^a, Rob W.H. Ruigrok^a, Marc Jamin^{a,*}

^a UMI 3265 UJF-EMBL-CNRS Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France ^b Institut Laue-Langevin, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

ARTICLE INFO

ABSTRACT

Article history: Available online 22 September 2011

Keywords: Mononegavirales Rhabdoviridae Rabies virus Vesicular stomatitis virus Transcription Replication The rhabdoviruses have a non-segmented single stranded negative-sense RNA genome. Their multiplication in a host cell requires three viral proteins in addition to the viral RNA genome. The nucleoprotein (N) tightly encapsidates the viral RNA, and the N–RNA complex serves as the template for both transcription and replication. The viral RNA-dependent RNA polymerase is a two subunit complex that consists of a large subunit, L, and a non-catalytic cofactor, the phosphoprotein, P. P also acts as a chaperone of nascent RNA-free N by forming a N⁰–P complex that prevents N from binding to cellular RNAs and from polymerizing in the absence of RNA. Here, we discuss the recent molecular and structural studies of individual components and multi-molecular complexes that are involved in the transcription/replication complex of these viruses with regard to their implication in viral transcription and replication.

© 2011 Elsevier B.V. All rights reserved.

Contents

1.	The rhabdoviruses	126
2.	The viral transcription/replication machinery	127
3.	The nucleoprotein and the N–RNA complex	129
4.	The phosphoprotein	130
5.	Complexes between the nucleoprotein and the phosphoprotein	130
	5.1. The N–RNA–P complex	130
	5.2. The N ⁰ – P complex	131
6.	The L subunit of the viral polymerase	134
7.	The viral particle	134
8.	Future work	135
	Acknowledgments	135
	References	135

1. The rhabdoviruses

The *Rhabdoviridae* is a family of enveloped viruses that have a non-segmented genome ((–)RNA) made of a single stranded negative-sense RNA molecule. The *Rhabdoviridae* family is grouped in the order *Mononegavirales* (*MNV*) (the name is a composition of three elements: *Mono* – single; *nega* – negative sense; *virales* – virus) with the *Filoviridae* (Ebola and Marburg viruses), the *Paramyxoviridae* (measles, mumps, respiratory syncytial viruses) and the *Bornaviridae* (Borna disease virus). All these viruses, share

E-mail address: jamin@embl.fr (M. Jamin).

a similar organization of their genome and virion as well as similar strategies of RNA replication and transcription (Pringle, 1997). Viruses classified as *Rhabdoviridae* are currently grouped into six genera on the basis of structural properties, antigenicity and phylogenetic analyses (Delmas et al., 2010; Tordo et al., 2004). The *Lyssaviruses, Vesiculoviruses* and *Ephemeroviruses*, infect a variety of animal hosts, including mammals, fishes and arthropods, the *Novirhabdoviruses* infect numerous fishes and the *Cytorhabdoviruses* and *Nucleorhabdoviruses* are arthropod-borne and infect plants (Jackson et al., 2005; Tordo et al., 2004). Numerous other rhabdoviruses remain unclassified. The rabies virus (RAV) and the vesicular stomatitis virus (VSV), respectively, a *Lyssavirus* and a *Vesiculovirus*, are prototypes of this large family of viruses.

RAV is the main causative agent of rabies. It is a neurotropic virus that causes an incurable human brain disease with a complex pathogenesis (Dietzschold et al., 2008; Hemachudha et al., 2002). It

^{*} Corresponding author at: UMI 3265 UJF-EMBL-CNRS Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France. Tel.: +33 4 76 20 94 62; fax: +33 4 76 20 94 00.

^{0168-1702/\$ –} see front matter 0 2011 Elsevier B.V. All rights reserved. doi:10.1016/j.virusres.2011.09.025

leads to death if it remains untreated before the disease is declared, but the infection can be prevented. Since its first successful use by Louis Pasteur in 1885, an anti-rabies vaccine is available as well as a post-infection treatment (Rupprecht et al., 2006). Currently, rabies is under control in Europe and North America because of efficient vaccination campaigns of dogs and wild animals (Warrell and Warrell, 2004), but, it still kills more than 50,000 people every year mainly in Africa and Asia (see WHO Fact Sheet N° 99, 2009) and causes an important economic burden in affected populations (Knobel et al., 2005). Recent reports suggest that the number of cases is even substantially underestimated in poor rural areas (Cleaveland et al., 2002). The worldwide persistence of the disease is mainly due to the cost of efficient vaccination campaigns and of post-exposure treatments. Of the 10 million people receiving post-exposure treatments each year after being exposed to rabiessuspect animals (WHO estimation in 2004), only a few benefit from local instillation of anti-rabies immunoglobulins, which has been demonstrated to be an indispensable complement to vaccination. The main natural reservoirs of rabies virus are bats or carnivores such as dogs, foxes, raccoons and skunks, but the mechanism of cross-species transmission remains poorly understood (Streicker et al., 2010). New strains, potentially unsusceptible to vaccination against classical rabies could emerge, cross the species barrier and spread in the human population. Indeed, several isolates of bat Lyssavirus are already less sensitive or insensitive to the current vaccine and post-infection treatments against rabies (Hanlon et al., 2005; Weyer et al., 2008). The vesicular stomatitis virus (VSV) is an arthropod-transmitted virus that causes a mild disease in horses, cattle and pigs and rarely transmits to human. It has economical impacts in the farm industry. For many years, VSV has been used as a model system for studying the replication cycle of the Rhadoviridae and, more generally, the interactions between viruses and their host cells. More recently, VSV has been used as a vaccine vector and as an oncolytic virus (Lichty et al., 2004; Rose et al., 2008). Other Vesiculoviruses such as Chandipura virus or Piry virus also infect humans. In recent outbreaks in India, Chandipura virus produced acute encephalitis with high fatality ratio (Basak et al., 2007; Rao et al., 2004). Finally, Vesiculoviruses infecting fishes or plants are of major economic importance to fish farming and agricultural industries (Jackson et al., 2005).

VSV and RAV use different strategies for regulating their replication in their host cell that lead to the different pathologies caused by these viruses. VSV replicates rapidly to high levels of progenies in a minimum of time and strongly interferes with the host's cell metabolism. The virus strongly inhibits host gene expression and translation (Ahmed et al., 2003; Connor and Lyles, 2002, 2005), interferes with the host cell innate immune response (Durbin et al., 1996) and induces apoptosis of infected cells (Koyama, 1995). In contrast, RAV replicates more slowly, interferes much less with the host cell metabolism and is less cytopathic than VSV (Schnell et al., 2010). Recently, recombinant RAV and VSV have been recovered from cDNA (Lawson et al., 1995; Schnell et al., 1994; Whelan et al., 1995), but most studies analyzing the processes of transcription and replication and their regulations have been carried out for VSV, because functional transcription/replication complexes of VSV can be reconstituted from isolated components (Emerson and Yu, 1975) but not of RAV. However, the structural organization of the transcription/replication complex is similar, and comparative studies provide useful information about the mechanisms of action of this molecular machinery.

2. The viral transcription/replication machinery

The genome of the *Rhabdoviridae* is of 9–18 kb in length and comprises up to ten genes flanked by untranslated leader (le) and

trailer (tr) RNA regions. RAV and VSV genomes contain only five genes that are common to all members of the MNV and encode successively from the 3' terminus, the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the large subunit of the RNA-dependent RNA polymerase (L) (Fig. 1A).

The replication cycle of both RAV and VSV occurs entirely in the cytoplasm of their host cells (Follett et al., 1974) and leads to the formation of intracytoplasmic inclusions (Heinrich et al., 2010; Lahaye et al., 2009). The G and M proteins are involved in the entry of the virus in its host cell and in the assembly and budding of new viral particles. The viruses attach to their host cells through the G protein and are internalized by endocytosis. The ribonucleoprotein complex containing the genomic RNA together with N, P and L proteins, named hereafter the nucleocapsid (NC), is then delivered into the cytoplasm by fusion of the viral envelope with the lipid bilayer of the endocytic vesicle. The fusion is induced by a pHdependent conformational change of the G protein (Gaudin et al., 1999; Libersou et al., 2010). The M protein maintains the structural integrity of the virus particle by connecting the lipid envelope to the NC, and in infected cells, the M protein spontaneously associates with the cytoplasmic membrane, where it recruits newly synthesized nucleocapsids (Solon et al., 2005). The sequence of the M protein contains a motif of interaction with the host cell machinery involved in multi-vesicular body (MVB) formation and it mediates the release of viral particles by redirecting this machinery to the cytoplasmic membrane (Harty et al., 1999; Irie et al., 2004).

The NC constitutes the active machinery that carries out both the transcription and replication of the genome and the actual infectious part of the virus (Szilagyi and Uryvayev, 1973). For VSV, transcription could be reconstituted in vitro using a mixture of L, P and N-RNA complex isolated from the virus (Emerson and Yu, 1975), and both transcription and replication could be reconstituted in vivo using a three-plasmid (L, P and N) inverse genetic system (Pattnaik et al., 1997). For VSV, no host cell protein is necessary in stoichiometric amount relative to the viral proteins, but the involvement of low amounts of cellular proteins cannot be ruled out. Numerous results in the literature support a role of the phosphorylation of P or N in the activation and regulation of the viral transcription/replication machinery (Pattnaik et al., 1997; Toriumi and Kawai, 2004), and it is likely that cellular kinases associated with the NC are required for the proper action of this machinery (Lenard, 1999).

In the NC (Fig. 1B), the N protein entirely coats the RNA genome, and it is the N–RNA complex, rather than the naked RNA, that serves as a template for the two-subunit RNA-dependent RNA polymerase comprising the P and L proteins (Moyer et al., 1991). However, the structural organization of the polymerase complex on the N–RNA complex and the role of N and its interactions with the polymerase complex remain unknown. In addition, the encapsidation by N also protects the genome from cellular nucleases and hides it from the intracellular receptors of the innate immune response. The 5′ end of the viral RNA bears a triphosphate that is recognized by RIG-I (Hornung et al., 2006) and would certainly trigger an immune response if it were freely accessible.

At first, when it is delivered into a host cell, the NC carries out the primary transcription of the genome, generating a short uncapped and untranslated leader RNA (leRNA) and five capped and polyadenylated messenger RNAs (mRNA) that encode the viral proteins. The amounts of leRNA and of the five mRNAs produced in this process decrease in the order of the genes on the genome, suggesting that the polymerase initiates at the 3' end of the genome (Abraham and Banerjee, 1976; Ball and White, 1976; Emerson, 1982) and transcribes the genes sequentially from the 3' end to the 5' end, although it can also initiate at the N gene start sequence (Chuang and Perrault, 1997; Whelan and Wertz, 2002). It is assumed that the polymerase complex proceeds by a


Fig. 1. The rhabdovirus transcription/replication machinery. (A) Schematic organization of the genome. The gene products are denoted as follows: N, nucleoprotein; P, phosphoprotein; M, matrix protein; G, surface glycoprotein and L, large subunit of the RNA-dependent RNA polymerase. Le and tr indicates the leader and trailer RNA sequences. (B) Schematic representations of the transcription/replication machinery. The nucleoprotein (in green) forms with the RNA genome (blue line) the active template for the RNA-dependent RNA polymerase (RdRP) complex comprising the L (in yellow) and P (in orange and red) proteins. P forms dimers and each monomer of P contains a central dimerization domain (in orange), a C-terminal N-RNA binding domain (red bean-like shape) and two intrinsically disordered regions (red lines). During the replication of the genome, the newly synthesized antigenomic or genomic RNA copies are encapsidated by nascent N molecules that are transferred from the soluble N⁰–P complex. In the N⁰–P complex, the N-terminal MoRE of P (shown as a red cylinder in the N⁰–P complex) prevents host-cell RNA binding by obstructing the RNA binding groove and the self-assembly of N by interfering with the docking of the N_{NT}-arm of another N. Upon transfer of N to the growing NC P is released, the binding groove for the N_{NT}-arm is freed in the RNA-bound form and can accept the next incoming N molecule.

stop-and-start mechanism remaining attached to the template after pausing and re-initiating at the next gene start (Iverson and Rose, 1981). In this process, the polymerase terminates the transcription at each intergenic region by iterative slippage on a poly-U tract that generates the poly-A tail (Barr and Wertz, 2001; Barr et al., 1997; Hwang et al., 1998; Stillman and Whitt, 1999). The mRNAs are capped and methylated at their 5' end and poly-adenylated at their 3' end by the L protein. The mRNAs are then used by the cellular ribosomes for synthesizing viral proteins. In a second stage, the viral polymerase complex switches from transcription to replication. During the replication process the viral polymerase initiates at the 3' end of the genome, ignores all regulatory signals at intergenic regions and produces full-length complementary (+)RNA intermediates that then serves as a template for generating new copies of the (-)RNA genome. These antigenomes and genomes, but not mRNAs, are encapsidated by N, and therefore the replication process requires a continuous supply of nascent RNA-free N molecules (Howard and Wertz, 1989; Patton et al., 1984; Peluso and Moyer, 1988). Finally, at a later stage of the viral cycle, a secondary phase of transcription occurs, thus increasing the amounts of viral proteins present in the cell, possibly to permit the assembly of viral progenies (Simonsen et al., 1979).

The phosphoprotein (P) is an essential non-catalytic cofactor of the polymerase (Emerson and Yu, 1975), but the L subunit carries out the multiple enzymatic activities that are required for the transcription and the replication of the genome (Abraham et al., 1975; Barr and Wertz, 2001; Barr et al., 1997; Grdzelishvili et al., 2005; Hercyk et al., 1988; Li et al., 2006; Moyer and Banerjee, 1975). L catalyzes the template-based RNA synthesis, the capping and methylation of the 5' extremity of the mRNAs and the polyadenylation of their 3' extremity. The role of P as a cofactor of the polymerase is unclear. P may stabilize L (Canter and Perrault, 1996), but its main known role is in the attachment of L to the N–RNA template. By itself, the L subunit has a poor affinity for the N–RNA template (Mellon and Emerson, 1978), whereas P binds to the N–RNA template and to the polymerase, suggesting that P allows the correct positioning of the polymerase onto its template (Fig. 1B). It is possible that P acts as a processivity factor that keeps L attached to its template during the process at intergenic regions and thereby facilitates the re-initiation of RNA synthesis at the next start signal.

The second major role of P is to chaperone N and allows the assembly of new NCs from nascent N molecules. N has a strong affinity for RNA and must be prevented from encapsidating cellular RNAs. P binds to nascent N molecules, forming a soluble N/P complex, named the N⁰–P complex, with the superscript 0 indicating the absence of RNA (Fig. 1B). The continuous production of the N⁰–P complex is required for the replication of the viral genome and it has been hypothesized that the accumulation of this complex triggers the switch between transcription and replication (Howard and Wertz, 1989; Masters and Banerjee, 1988; Peluso and Moyer, 1988).

Here, we review recent structural studies of the components of the transcription/replication machinery of either VSV or RAV and the new hypotheses that these results raise about the structural organization of the NC and its dynamic remodeling during the functioning of the viral polymerase. Low- and high-resolution structures of N–RNA complexes revealed how N molecules assemble into NCs and how they interact with and protect the RNA molecule. Various biophysical and structural studies demonstrated that the phosphoproteins are highly flexible, made of concatenated disordered regions and structured domains. The structures of

A VSV B 50 nm С RAV D E NCTD N_{cT}loop NCTD N_{NT} arm NNTD

Fig. 2. Structure of the nucleoproteins and N–RNA complexes. (A) Typical electron micrograph of negatively stained VSV nucleocapsids. The viral nucleocapsids were isolated with centrifugation in CsCl gradient and negative staining was done with sodium silicotungstate. (B) Structure of the circular decameric N–RNA complex of VSV (PDB ID: 2gic). One nucleoprotein is shown in green and the RNA is shown in blue. (C) Electron micrographs of negatively stained RAV nucleocapsid. Negative staining was done with sodium silicotungstate. (D) Structure of the circular 11-mer N–RNA complex of RAV (PDB ID: 2gtt). One nucleoprotein is shown in green and the RNA is shown in blue. (E) Ribbon representation of one protomer of VSV N bound to 9 nucleotides. The N protein, shown in green, has an N-terminal domain (N_{NTD}), a C-terminal domain (N_{CTD}) and two subdomains, the N-terminal arm (N_{NT}–arm, aa 1–21) and the C-terminal loop (N_{CT}–loop, aa 340–375). The RNA, shown in blue, is completely enwrapped by N_{NTD} and N_{CTD} and is not accessible for the polymerase.

two types of complexes between N and P, revealed how different regions of P specifically binds to different regions of N and carry out different functions in the transcription/replication complex. Finally, electron micrographs of purified VSV L protein provided a first glimpse at the shape of this large protein.

3. The nucleoprotein and the N-RNA complex

In rhabdovirus NCs, each N molecule binds nine nucleotides (Iseni et al., 1998; Thomas et al., 1985). The number of N molecules present in a virion of VSV (1250 N) (Ge et al., 2010; Thomas et al., 1985) corresponds closely to the theoretical number required for

covering the entire genome (VSV: 11161 nt \rightarrow 1240 N), although the length of the genomic RNA molecule is not an entire multiple of nine nucleotides. By contrast, in *Paramyxoviridae* each N molecule binds six nucleotides and the length of the genome corresponds to a multiple of six nucleotides (Kolakofsky et al., 2005).

The N–RNA complex extracted from infected cells or from virus particles forms a long and irregular helix (Fig. 2A and C). The electron micrographs reveal a striated structure corresponding to the side by side arrangement of the nucleoproteins along the RNA molecule. The dimensions of the nucleoprotein in the NCs from VSV and RAV are similar as well as the diameter of individual coils (Iseni et al., 1998), but NCs of RAV appear more tightly coiled (Fig. 2A and C). In the NC, the RNA is tightly bound to N, resisting centrifugation

in CsCl density gradient (Lynch and Kolakofsky, 1978). It is highly protected against digestion by nucleases independently of the salt concentration, although the bases are accessible to small molecules (Iseni et al., 2000). The RNA can however be removed by treatment with very high concentrations of RNAse A at 60 °C (Green et al., 2010).

The expression of the N as a recombinant protein in bacteria or in insect cells leads to the formation of long N–RNA complexes that resemble the natural NCs and to circular N–RNA complexes that contain short unidentified RNAs (Green et al., 2000; Iseni et al., 1998; Schoehn et al., 2001). The recombinant circular RAV N–RNA complexes contain from 9 to 13 N-protomers with a larger proportion of rings with 10 protomers (Albertini et al., 2006a; Iseni et al., 1998), while recombinant VSV N–RNA complexes contain predominantly 10 protomers (Blumberg et al., 1983; Ge et al., 2010; Green et al., 2000). The N–RNA rings of RAV and VSV are less flexible than the long NCs and their structure was reconstructed from electron micrographs (Chen et al., 2004; Iseni et al., 1998; Schoehn et al., 2001) and, more recently, solved at high-resolution by X-ray crystallography (Albertini et al., 2006b; Green et al., 2006) (Fig. 2B and D).

The N protein has two domains, an N-terminal domain (N_{NTD}) and a C-terminal domain (N_{CTD}), that are connected by a hinge region (Fig. 2E). The RNA molecule is bound in a groove of the protein located at the interface between N_{NTD} and N_{CTD} and is completely sequestered from the solvent. In addition N has two subdomains, an N-terminal arm (N_{NT} -arm) and a C-terminal flexible loop (N_{CT} -loop), which exchange in opposite direction between adjacent subunits of the N–RNA complex. The polymer of N in complex with RNA is stabilized by multiple salt bridges between the sugar-phosphate backbone of the RNA and basic residues of N, by contacts between neighboring N molecules involving hydrophobic side-to-side interactions, mainly between adjacent N_{CTD} , and by the exchange of N- and C-terminal sub-domains between adjacent N protomers (Albertini et al., 2006b; Green et al., 2006).

4. The phosphoprotein

Early studies showed that P contains independent functional regions (Das et al., 1997; Takacs et al., 1993) and its hydrodynamic properties suggested that it is a non-globular molecule (Gérard et al., 2007). These characteristic features are explained by the recent findings that P is made of concatenated disordered regions and structured domains. The structure and functions of rhabdovirus phosphoproteins have been reviewed recently (Leyrat et al., 2010, 2011b), and therefore, their structural organization is only briefly discussed here (Fig. 3).

The structured domains have been localized in the amino acid sequence by limited proteolysis (Ding et al., 2004; Mavrakis et al., 2004) or by a meta-prediction of disorder (Gérard et al., 2009). The protein is dimeric (Gérard et al., 2007) and a central autonomous folding domain (P_{CED}) is responsible for the dimerization of the protein (Ding et al., 2006; Ivanov et al., 2010). In RAV and VSV P dimers, the monomers are oriented in a parallel orientation, but the structure of the dimerization domain is different, suggesting that oligomerization is conserved but not the structure (Ivanov et al., 2010). The structure of the C-terminal folded domain (P_{CTD}) of VSV and RAV P clearly adopt a similar fold clearly showing that they are homologous (Mavrakis et al., 2004; Ribeiro et al., 2008). VSV P_{CTD} lacks two α -helices of RAV P_{CTD} , but the common parts can be superimposed. In VSV, the junction between P_{CED} and P_{CTD} (Ribeiro et al., 2008) and the N-terminal region of both P proteins were shown to be globally disordered (Gérard et al., 2009; Leyrat et al., 2011a). The N-terminal intrinsically disordered region (IDR_{NT}) contains the binding sites for the RNA-free N⁰ (Chen et al.,



Fig. 3. Structure of the phosphoproteins. (A) Schematic organization of the P proteins. P proteins of rhabdoviruses consist of a long N-terminal intrinsically disordered region (IDR_{NT}), a central oligomerization domain (P_{CED}), a C-terminal IDR (IDR_{CT}) and a C-terminal N–RNA binding domain (P_{CTD}). In VSV, IDR_{CT} is about 16 aa long, whereas in RAV, it is about 60 aa long. The N-terminal region contains the binding site for RNA-free N. In VSV P, this region consists of a MoRE (molecular recognition element) that forms transient α -helices in the isolated protein and adopts a rigid fold in the RNA-free N⁰–P complex. (B) Structured domains of VSV and RAV P. The high resolution structures of two domains corresponding closely to the predicted central domain of P (P_{CED}) and the C-terminal domain (P_{CTD}) have been determined at high resolution. The known three-dimensional structure of P_{CED} and P_{CTD} are shown as cartoon models and the disordered regions are shown as dotted lines. The dotted circles represent the localization of N⁰ binding MoRE. The PDB codes are given in figure.

2007; Mavrakis et al., 2006). A detailed NMR characterization of the N-terminal region of VSV revealed the transient presence of two α -helices in the N⁰-binding region, suggesting that this part of P is a molecular recognition element (MoRE) (Leyrat et al., 2011a). A MoRE is a small segment of a protein that is disordered in its isolated form but folds upon binding to a molecular partner (Fuxreiter et al., 2004). The binding site for L has also been localized in the N-terminal part of P but has not yet been precisely mapped (Castel et al., 2009; Emerson and Schubert, 1987).

5. Complexes between the nucleoprotein and the phosphoprotein

During the replication cycle, P forms two different complexes with N, the N^0 -P complex and the N-RNA-P complex. These interactions are independent of each other, and the recent structural characterization of these two complexes reveals the molecular origin of the dual behavior of P.

5.1. The N–RNA–P complex

The attachment of the L protein onto the N–RNA template requires the presence of P (Mellon and Emerson, 1978), suggesting that one function of P in the polymerase complex is to attach the L protein to its N–RNA template. For both VSV and RAV, the interaction between P and L was mapped to the N-terminal region of P (Castel et al., 2009; Emerson and Schubert, 1987) and the C-terminal region of L (Castel et al., 2009; Chenik et al., 1998), and

the interaction between P and the N–RNA complexes was mapped to the C-terminal domain of P (Emerson and Schubert, 1987; Jacob et al., 2001; Takacs et al., 1993) and the C-terminal domain of N (Kouznetzoff et al., 1998; Schoehn et al., 2001). For both viruses, only the C-terminal domain of P is required for the attachment onto the N–RNA template (Green et al., 2000; Mavrakis et al., 2004), although the dimerization of P could modulate the affinity of these two proteins.

For VSV, a complex between P_{CTD} and an N-RNA rings containing 10 protomers of N was produced by co-expressing both proteins, and the structure was solved by X-ray crystallography showing one P_{CTD} bound to each N protomer (Green and Luo, 2009) (Fig. 4A). For RAV, titration experiments showed that only two P_{CTD} could bind simultaneously to an N-RNA rings containing 10 or 11 protomers of N and that the binding of these P_{CTD} is non-cooperative with a dissociation constant of 160 ± 20 nM (Ribeiro et al., 2009). A model of the complex between an N-RNA ring and P_{CTD} was obtained by a flexible cross-docking approach that used biochemical and small-angle X-ray scattering (SAXS) data as constraints (Fig. 4B). A previous analysis of mutations in the P_{CTD} of Mokola virus revealed the existence of two important patches of amino acid residues on opposite faces of the P_{CTD} that are critical for the interaction with N, a hydrophobic pocket with an exposed tryptophan residue (Whole) and a patch of positive residues (Jacob et al., 2001; Mavrakis et al., 2004). In accordance, the modeling procedure included the selection of models that maximize the number of contacts between these residues of P_{CTD} and N (Ribeiro et al., 2009).

In the best model of the RAV complex, like in the VSV structure, each P_{CTD} is bound at the C-terminal top of one N protomer (N_i) (Fig. 4A and B). The C-terminal helix of P_{CTD} lies along helix α_{14} of N_{CTD} but, because P_{CTD} of VSV is shorter than that of RAV, the helix is oriented in the opposite direction (in VSV, the C-terminal side of the α -helix points toward the outside of the ring, and in RAV, it points toward the centre of the ring), while in both complexes, the N-terminal extremity of P_{CTD} points toward the centre of the ring (Fig. 4C and D). Also in both complexes, the N_{CT}-loops of protomer N_i and N_{i+1} act like a pair of pincers that grasp the P_{CTD} domain (Fig. 4C and D). In the VSV complex, the N_{CT}-loop is slightly shifted upon binding to P_{CTD} and several side-chains are reoriented, while in the RAV model, a large part of the flexible N_{CT}-loop that was not visible in the N–RNA complex alone interacted with P_{CTD} .

The binding site of P_{CTD} is distributed over two adjacent N protomers and thus requires the polymerization of N, nicely explaining why P_{CTD} can bind only to polymerized N-RNA complexes but not to the isolated N⁰ (Mavrakis et al., 2006). A main difference is that in VSV N-RNA complex, every N is capable of simultaneously accommodating a P_{CTD} whereas in the RAV complex, the N_{CT} loop on each side of P_{CTD} adopted different conformations and the N-terminal extremity of P_{CTD} protruded in the centre of the ring, thus preventing the binding of PCTD on adjacent N molecules. Also, the NCT-loop of RAV is longer than that of VSV, and the two flanking loops bury a larger part of P_{CTD} surface area in the RAV complex than in the VSV complex (950 Å² in the VSV complex and 1700 Å² in the RAV complex). In the RAV model, hydrophobic residues of the N_{CT}-loop of the N_i protomer inserts in the W-hole, while acidic residues of the N_{CT}-loop of the N_{i+1} protomer form salt-bridges with the basic patches on the opposite face of P_{CTD}. In VSV, the N_{CT} loops of protomers N_i and N_{i+1} interact with P_{CTD} mainly by forming side-chain H-bonds, and a hydrophobic cavity at the surface of P_{CTD} that was postulated to be a potential binding site is capped by the N_{CT}-loop of protomer N_{i+1} but no side-chain of N penetrates within this cavity. In the VSV complex, however, an additional stabilization of the complex may result from the interactions between P_{CTD}s bound to the adjacent protomers (Fig. 4C).

In the N–RNA complexes of RAV and VSV, the RNA bases are completely enwrapped by $N_{\rm NTD}$ and $N_{\rm CTD}$ (Albertini et al., 2006b;

Green et al., 2006). The binding of P_{CTD} to the N–RNA ring revealed no significant conformational change in N that would suggest that binding of P triggers the opening of the N jaws in order to allow the RNA polymerase access to the RNA (Green and Luo, 2009; Ribeiro et al., 2009). There is, however, no evidence that the circular N–RNA complexes are active templates for RNA synthesis, and it is possible that ring closure blocks a potential conformational change. Alternatively, the conformational change required for giving access for the polymerase to the RNA could be induced by the polymerase itself.

Two different models have been proposed to explain how P keeps L attached to the N-RNA complex and how the polymerase complex moves along its template. A first model posits that P progresses along the template by a cartwheeling mechanism in which each C-terminal domain alternatively associates with and dissociates from N (Curran, 1998; Kolakofsky et al., 2004). This model was proposed on the basis of results obtained for Sendai virus (SEV), a prototype member of the Paramyxoviridae (Curran, 1998). The P protein of paramyxoviruses shares a similar modular organization with that of the rhabdoviruses (Gérard et al., 2009; Karlin et al., 2003), except that SEV P forms a tetramer rather than a dimer (Tarbouriech et al., 2000). For SEV, the affinity of a single P_{CTD} for the N–RNA template is weak, in the micromolar range $(60 \pm 20 \,\mu\text{M})$ (Houben et al., 2007), and both the central tetramerization domain and C-terminal domain of P are required for binding to the N-RNA template in pull-down experiments performed at sub-micromolar concentrations of P and for efficient RNA synthesis (Curran, 1998). For rhabdoviruses, however, several results are in apparent contradiction with the cartwheeling model. Firstly, a mutant of RAV P deleted from its central dimerization domain and that remained functional in transcription (Jacob et al., 2001) was found to behave as a monomer in solution (Gérard et al., 2009). The oligomerization of P is not required for proper motion of the polymerase, but a monomer cannot cartwheel along the template. Secondly, a single RAV P_{CTD} is able to bind to the N–RNA rings with a good affinity $(160 \pm 20 \text{ nM})$ and the kinetics of association and dissociation of this complex on the second time range are incompatible with a mechanism in which PCTDs continuously associate and dissociate to progress along the N-RNA template at a rate compatible with the rate of RNA transcription and replication (Ribeiro et al., 2009). A second model proposes that P molecules are bound at regular intervals along the N-RNA complex and that L jumps between bound Ps, helped in this process by the high flexibility of P (Ribeiro et al., 2009). According to this model, P molecules remain continuously attached to the N-RNA complex in the virion and in the cytoplasm of infected cells, and the oligomerization state of P is selected to obtain the adequate affinity for its partner. The number of P molecules bound to the N-RNA template in an active transcription/replication complex is unknown, but because of the dimensions and flexibility of the P protein, the binding of one P dimer every five N protomers is sufficient to allow one P dimer to catch the L protein from the previous P dimer and to transfer it to the next one. This model needs now to be tested.

5.2. The N^0 –P complex

The encapsidation of newly synthesized antigenomes and genomes is an essential step in the replication process that requires the continuous production of nascent N molecules in a soluble, RNA-free form (Masters and Banerjee, 1988; Peluso and Moyer, 1988). It has been established that N first forms a soluble complex with P and is then transferred from P to the newly made RNA molecules and that the growing NC assembles unidirectionally from the 5' extremity to the 3' extremity of the RNA (Blumberg and Kolakofsky, 1981; Dillon and Gupta, 1988). How the transfer occurs and is regulated are questions that need to be addressed. During



Fig. 4. Structure of the N–RNA–P_{CTD} complexes. (A) Space-filling model of VSV decameric N–RNA–P_{CTD} complex (PDB ID: 3hhz). N protomers are shown in green. The P_{CTD} are shown alternatively in red and pink. (B) Space-filling model of RAV N–RNA–P_{CTD} complex. The model, generated by molecular modeling, contains two P_{CTD} (in red) bound to a 11-mer N–RNA ring (in green). (C) Side-view of the N–RNA–P_{CTD} complex of VSV. The complex is seen from outside the ring. P_{CTD} is shown in red. The N protomer on which P_{CTD} lies is shown in green and is labeled N_i. The adjacent protomer on the 3' side of the N_i protomer is shown in grave (N_{i-1}) and that on the 5' side is shown in cyan (N_{i+1}). The adjacent P_{CTD} are shown in pink. (D) Side-view of the N–RNA–P_{CTD} complex of RAV. The N–RNA–P_{CTD} the same orientation, at the same scale and with the same color scheme than that of VSV. In RAV complex, only two P_{CTD} binds to a N-RNA ring and they are not interacting with each other.

the replication process, RNA synthesis is initiated at the 3' end of the template and the commitment to synthesize full-length RNA antigenomes and genomes rather than mRNAs must be taken by the polymerase complex before or at the moment where it reaches the end of the leader or trailer sequences, respectively. The concentration of N⁰–P plays a role in switching from transcription to replication (Arnheiter et al., 1985; Gupta and Banerjee, 1997) and the transfer of N could be controlled by its relative affinity for P and for encapsidation signals present in the terminal untranslated regions (Blumberg et al., 1981). Conversely, it could depend on the modification of the polymerase complex (Qanungo et al., 2004).

Because of the difficulty to prepare the N⁰–P complex (Mavrakis et al., 2003), this complex escaped structural characterization for a long time. The region of P involved in the formation of the N⁰–P complex was localized to a short segment of its N-terminal region (Chen et al., 2007; Mavrakis et al., 2006). For RAV, a fragment containing residues 4–40 of P was identified by limited proteolysis (Mavrakis et al., 2006), while for VSV, residues 11–30 were found to be essential for keeping N soluble (Chen et al., 2007). For VSV, this N⁰-binding region encompasses the transient α -helices that were discovered by NMR spectroscopy in the N-terminal region of P (Leyrat et al., 2011a).

Recently, a N⁰–P complex was reconstituted between an armless N, lacking the 21 N-terminal residues, $N_{\Delta 21}$, and either a peptide corresponding to the 60 first residues of P (P₆₀) or a dimer of full length P (Leyrat et al., 2011c). These purified complexes are soluble, lack RNA and correspond to a heterodimer in the case of P₆₀ or to a mixture of hetero-trimers (1N:2P) and hetero-tetramers (2N:2P) in the case of full length P (Leyrat et al., 2011c). The structural characterization of the $N_{\Delta 21}^0$ –P₆₀ complex in solution by SAXS revealed a particle of similar size and shape as those of an N monomer extracted from the N–RNA crystal structure, while NMR spectroscopy showed that only residues 6–40 of P are directly

involved in the complex with N⁰, whereas the flanking regions (aa 1–5 and 41–60) remain flexible in the complex (Leyrat et al., 2011c). The $N^0_{\Delta 21}$ – P_{60} complex was crystallized in the form of a multimeric circular complex, in which the N molecules are arranged as in the N-RNA complex (Leyrat et al., 2011c). In each protomer, residues 6-35 of P were clearly visible in the crystal structure. Residues 17–30 form an amphipathic α -helix that is docked on one side of N in a groove at the interface between N_{NTD} and N_{CTD}. The peptide is oriented in a roughly perpendicular direction relative to the RNA binding groove (Fig. 5A and B), and none of the basic residues of VSV N directly contacting the RNA backbone in the N-RNA complex are involved in the interaction with the MoRE of P. The second transient α -helix (aa 25–31) present in the isolated N⁰-binding region is stabilized upon binding to its N⁰ partner and extends from residue 17 to residue 31. In contrast with RNA, P binds mainly through intermolecular hydrophobic interactions although the bipolar distribution of charges on the surface of P, with a positive pole at its N-terminus that docks on the negatively charged backside of N_{CTD} and a negative pole at its C-terminus that localizes in the highly positive RNA binding groove, could play a role in orientating P before binding. The hinge region of N (aa 200-300) is more conserved among VSV isolates as well as between the more evolutionary distant VSV and RAV (Luo et al., 2007) than the rest of the protein (Fig. 5C) and participates in the regulation of the transcription and replication (Nayak et al., 2009) in accordance with its central roles in binding both the RNA and MoRE of P. The conservation of several hydrophobic residues lining the binding groove of P, as well as of Arg³¹², which forms a salt bridge with an acidic residue of P, suggests that a similar N⁰–P complex forms in RAV.

The N⁰-binding region of P must prevent two processes, the polymerization of N and the non-specific encapsidation of host cell RNAs (Chen et al., 2007; Howard and Wertz, 1989; Masters and Banerjee, 1988; Mavrakis et al., 2006; Peluso and Moyer, 1988),



Fig. 5. Structure of the N⁰–P complex. (A) Representations of one protomer from VSV N–RNA complex (PDB ID: 2gtt). The N protomer is shown as a space filling model in green and the RNA is shown in blue. The 3' terminal nucleotide of the RNA molecule is shown in magenta. (B) Representations of one protomer from the N⁰_{Δ21}–P₆₀ complex (PDB ID: 3pmk). The N protomer is shown as a space filling model in green and P₆₀ is shown as a cartoon representation in red. The N– and C-terminal extremities of P₆₀ that are visible in the crystal are labeled. (C) Sequence conservation between VSV and RAV N mapped on a protomer of VSV N. Conserved residues are shown in red, similar residues are shown in orange and yellow, according to clustalW2 similarity scheme (Larkin et al., 2007). (D, E) Close-up of the interactions between exchangeable sub-domains in the circular N–RNA and N⁰_{Δ21}–P₆₀ complexs of VSV. In the VSV N–RNA complex (D), the N_{NT}-arm of protomer N_{i-1} contacts the N_{CT}-loop of protomer N_i (in green). In the N⁰_{Δ21}–P₆₀ complex (E), the N-terminal extremity of P₆₀ docks on protomer N_i (in green) at the position of the N_{NT}-arm of protomer N_{i-1} and contacts the N_{CT}-loop of protomer N_{i+1}. These representations suggest that P₆₀ interferes with the assembly of N in the absence of RNA.

which are independent of each other (Zhang et al., 2008). The structure of the $N^0_{\Lambda 21}$ -P₆₀ complex unveils how P fulfills its two chaperone functions. The C-terminal extremity of the MoRE overlaps with the RNA binding groove and directly competes with RNA binding (Fig. 5A and B), whereas its N-terminal part interferes with the binding of the N_{NT} arm of the neighboring N_{i+1} molecule and thereby prevents the polymerization of N (Fig. 5C and D). Also, the binding of the MoRE of P notably reduces the positive surface potential on one side of the N molecule, and may thereby affect the side-by-side interaction with another N molecule. In addition, the flanking parts of the MoRE, by fluctuating between multiple conformations, may also contribute to the chaperone activities of P by acting as entropic bristles that repel incoming RNA or N molecules or by masking their binding interfaces. In the complex with the full-length P dimer, these effects may be exacerbated by the high flexibility and the bulkiness of the remainder of the protein.

These results suggested a hypothesis to explain how P controls the encapsidation of a newly synthesized RNA molecule during the replication of the viral genome (Fig. 1B). Little is known about the mechanism by which N is transferred from N^0 –P to a RNA molecule or about the role played by the polymerase complex in this process. The N^0 –P complex conserves its ability to bind to the N–RNA complex (Gupta and Banerjee, 1997) more likely through the binding of P_{CTD} and it is possible that the P is transferred to the growing NC from a bound N⁰–P complex. However, the recent experiments showed that the N_{NT}-arm stabilizes the multimeric assembly of N molecules since its removal allowed reversing the assembly process (Leyrat et al., 2011c) and suggested that the multimeric N–RNA complex is more stable than the monomeric N⁰–P complex. The assembly of the growing NC could thus simply be driven by the higher stability of the multimeric N complex. In the soluble N⁰–P complex, P blocks the backside groove of N in which the N_{NT} arm of a neighboring N molecule binds and thereby prevents the assembly of soluble N molecules in the absence of RNA. Upon transfer of N onto a RNA molecule, P is released and the backside groove of the last added N molecule is liberated and becomes available for accepting the N_{NT}-arm of the next incoming N molecule (Fig. 1B).

In addition to its role in RNA encapsidation, the binding of the N-terminal region of P to N may also provide a mechanism for the initiation of (+)RNA synthesis at the genome 3' end and of (-)RNA synthesis at the antigenome 3' end (Leyrat et al., 2011c). Encapsidated RNA genome and antigenome are completely covered by the nucleoprotein and are not accessible to the RNA polymerase. However, the first N molecule at the 3' extremity of nucleocapsids



Fig. 6. Structure of the virion. (A) Typical negative staining electron microscopy micrographs of VSV virions. Negative staining was done with sodium silicotungstate. (B) Typical negative staining electron microscopy micrograph of RAV virions. (C) Transverse section in the cryo-electron microscopy density map of a reconstructed trunk of a VSV particle. The electron density map corresponding to 10 turns of 37.5 N molecules was reconstructed from the electron density map deposited in the Electron Microscopy Databank (EMD ID: EMD_1663). It is displayed at a threshold of 2.9 above the mean. The trunk is seen from its basal face. A decamer of N adapted from the high resolution structure of the N–RNA complex (PDB ID: 2wyy) was fitted into the electron density map with the program VEDA (graphical interface of UROX, Navaza et al., 2002) and is shown as a cartoon representation. N protomers are shown in green and the RNA molecule in blue. (D) Close up of the transverse section. A slice 5 nm thick of the reconstructed electron density map shows the organization of the N molecules in the trunk of the virion. The volumes of electron density on the outside of the virion correspond to the M protein (intermediate layer) and to the cytoplasmic tail of the G protein (outer layer).

exposes its binding site for the N-terminal MoRE of P. By binding to this surface, P may destabilize the N–RNA complex sufficiently to displace several nucleotides from the first N protomer and allow the polymerase access to the 3' extremity of the viral RNA.

6. The L subunit of the viral polymerase

The L protein is a 250 kDa multi-enzymatic protein, which catalyzes RNA synthesis as well as mRNA capping, methylation and polyadenylation. Its sequence is well conserved among all nonsegmented (-)RNA viruses, and sequence alignments revealed the existence of six conserved regions, numbered I-VI (Poch et al., 1990). The conserved region III contains different motives involved in the RNA-dependent RNA polymerase activity, while the conserved regions V and VI carry out mRNA capping. The capping mechanism of the non-segmented (-)RNA viruses is different from that of their eukaryotic hosts. GDP, rather than GMP, is transferred to a 5' monophosphate RNA by a polyribonucleotidyltransferase (PRNTase) (Ogino and Banerjee, 2007, 2008) that was mapped to region V (Li et al., 2008). This reaction proceeds through the formation of a covalent intermediate between the 5' monophosphate mRNA and a histidine residue of L (Ogino et al., 2010). After its formation, the cap is methylated successively at ribose 2'-O and guanine-N-7 positions by region VI (Grdzelishvili et al., 2005; Li et al., 2005, 2006; Rahmeh et al., 2009; Zhang et al., 2010). The polyadenylation also depends on region III and is performed by slippage of the L protein on a U-rich region present at the 5' end of each gene (Barr et al., 1997).

The first glimpse at the molecular structure of this large protein has been obtained recently by negative-staining electron microscopy (Rahmeh et al., 2010). Different regions of the protein were localized in the structure by deletion studies. Conserved domains I-IV encompassing the site of RNA synthesis were localized in a ring-like domain, similar to that found in other RNA polymerases (Ferrer-Orta et al., 2006). Conserved regions V and VI involved in cap formation and methylation were mapped to three globular domains linked by flexible linkers to the ring-like core (Rahmeh et al., 2010). The binding of P was also shown to induce a conformational change in the L protein that results in the rearrangement of the globular domains at its C-terminus and to promote the formation of dimers (Rahmeh et al., 2010). The activities of L are controlled by P (Das et al., 1997; Hwang et al., 1998; Pattnaik et al., 1997) and this rearrangement of L may represent a mechanism of this regulation. In addition, the various activities catalyzed by the L protein are coordinated and regulated (Li et al., 2009; Stillman and Whitt, 1999) and it has been shown that a minimal length of the transcript is required to allow its capping, suggesting that the spatial organization of the different domains of L could control its activities (Tekes et al., 2011).

7. The viral particle

Animal rhabdoviruses have an overall bullet shape with one conical end and one flat end, while plant rhabdoviruses have a bacillus shape with two conical ends. Fig. 6A and B shows negativestaining electron micrographs of RAV and VSV particles. Recently, the cryo-electron microscopy (EM) reconstruction of the virion of VSV, revealed the molecular organization of the viral proteins in three concentric layers (Ge et al., 2010). The outer shell is a host cell-derived lipid membrane decorated with trimeric G spikes. The middle layer is formed by M proteins which are organized in a regular network onto which the trimeric G spikes attach. The NC constitutes the inner shell and has itself a bullet shape particle (Ge et al., 2010). Its striated appearance comes from the regular arrangement of multimeric N (Schoehn et al., 2001). In the trunk of the bullet, N packs the genomic RNA into a regular helical N-RNA complex that comprises 37.5 subunits of N per turn (Fig. 6C). The N molecules are tilted upward by 27° from the horizontal plane. The tip of the bullet is formed of seven turns containing varying numbers of subunits, in which the tilt from the horizontal plane draws closer to that in the recombinant N-RNA rings. The upper turn which encapsidates the 5' extremity of the RNA molecule and may represent the nucleation centre from which the particle assembles, resembles a decameric ring, suggesting that a ten-membered ring or spiral corresponds to an optimal side-by-side orientation between adjacent N subunits. The N subunits are oriented with their N-terminal domain pointing toward the exterior of the virion and their C-terminal domain pointing toward the centre of the particle. The N_{NTD} connects the NC to the M network providing rigidity to the assembly, whereas on the inside of the NC the N_{CTD} is available for binding P molecules. A similar organization has been found for the RAV particle, although the RAV nucleocapsid forms helical turn of about 53 protomers (Albertini et al., 2008; Iseni et al., 1998; Schoehn et al., 2001).

The bullet-like particle contains a cavity of 31 nm in diameter and about 150 nm in length that was previously described as a "cigar" (Barge et al., 1993). This central cavity is large enough to accommodate the 450 molecules of P, 55 molecules of L protein and about 500 M molecules that are involved in the middle shell (Thomas et al., 1985), but its content could not be reconstructed from the cryo-electron micrographies because its components are not organized in a highly symmetrical fashion (Ge et al., 2010). The exact position of the L/P complexes in the viral particle remains, however, unknown. They may remain attached to the NC at randomly distributed localizations as they were captured by condensation of the NC during budding and as they are usually represented in drawings of rhabdovirus particles (Jackson et al., 2005; Schnell et al., 2010; Wertz, 1978). Alternatively, the M and perhaps P proteins may form a scaffold onto which the polymerase molecules await the unpacking of the NC in the host-cell cytoplasm to successively initiate RNA synthesis at the 3' end of the genome. The M protein has the ability to self-assemble reversibly into filaments that share some structural features with the observed "cigar" (Gaudin et al., 1995), and the P protein was shown to form filamentous aggregates upon incubation at 37 °C (Gérard et al., 2007).

8. Future work

The recent progress made in the preparation of highly purified components of the rhabdovirus transcription/replication complex and in the characterization of their stoichiometry, size, shape and structure opens new avenues for deciphering the molecular mechanisms of this machinery and the subtleties of their regulations. In the near future, it should be possible to reconstitute a functional system from purified components and to test hypotheses concerning the motion of the polymerase on its template, the exact role of the dimerization of P or the effect of concentration of N⁰-P and of the phosphorylation of P on the switch between the transcription and replication of the genome. The structural characterization of the L protein and of its complex with P is a long-sought grail that may find its epilogue in the near future. One can also hope that these new structural data will lead to the development of new antiviral therapies. Finally, the recent progresses made in in vivo imaging should also allow us to investigate the organization of the rhabdovirus transcription/replication machinery in its cellular context and to a better understanding of the connections with host cell components.

Acknowledgments

Research activities in the authors' laboratory were supported by grants from the French ANR (ANR-07–001-01 (ANRAGE)), the FINOVI foundation and Lyonbiopôle. Ivan Ivanov was supported by a PhD fellowship from the ILL and Filip Yabukarski was supported by a MENRT fellowship from the French government. We thank Guy Schoehn and Leandro Estrozi for their help in preparing Figs. 2 and 6, our colleagues for extensive discussions and the Partnership for Structural Biology for the excellent structural biology environment.

References

- Abraham, G., Banerjee, A.K., 1976. Sequential transcription of the genes of vesicular stomatitis virus. Proc. Natl. Acad. Sci. U.S.A. 73, 1504–1508.
- Abraham, G., Rhodes, D.P., Banerjee, A.K., 1975. The 5' terminal structure of the methylated mRNA synthesized *in vitro* by vesicular stomatitis virus. Cell 5, 51–58.
- Ahmed, M., McKenzie, M.O., Puckett, S., Hojnacki, M., Poliquin, L., Lyles, D.S., 2003. Ability of the matrix protein of vesicular stomatitis virus to suppress beta interferon gene expression is genetically correlated with the inhibition of host RNA and protein synthesis. J. Virol. 77, 4646–4657.
- Albertini, A.A., Clapier, C.R., Wernimont, A.K., Schoehn, G., Weissenhorn, W., Ruigrok, R.W., 2006a. Isolation and crystallization of a unique size category of recombinant rabies virus nucleoprotein–RNA rings. J. Struct. Biol. 158, 129–133.
- Albertini, A.A., Schoehn, G., Weissenhorn, W., Ruigrok, R.W., 2008. Structural aspects of rabies virus replication. Cell Mol. Life Sci. 65, 282–294.
- Albertini, A.A., Wernimont, A.K., Muziol, T., Ravelli, R.B., Clapier, C.R., Schoehn, G., Weissenhorn, W., Ruigrok, R.W., 2006b. Crystal structure of the rabies virus nucleoprotein–RNA complex. Science 313, 360–363.
- Arnheiter, H., Davis, N.L., Wertz, G., Schubert, M., Lazzarini, R.A., 1985. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. Cell 41, 259–267.
- Ball, L.A., White, C.N., 1976. Order of transcription of genes of vesicular stomatitis virus. Proc. Natl. Acad. Sci. U.S.A. 73, 442–446.
- Barge, A., Gaudin, Y., Coulon, P., Ruigrok, R.W., 1993. Vesicular stomatitis virus M protein may be inside the ribonucleocapsid coil. J. Virol. 67, 7246–7253.
- Barr, J.N., Wertz, G.W., 2001. Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination. J. Virol. 75, 6901–6913.
- Barr, J.N., Whelan, S.P., Wertz, G.W., 1997. cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. J. Virol. 71, 8718–8725.
- Basak, S., Mondal, A., Polley, S., Mukhopadhyay, S., Chattopadhyay, D., 2007. Reviewing Chandipura: a vesiculovirus in human epidemics. Biosci. Rep. 27, 275–298.
- Blumberg, B.M., Giorgi, C., Kolakofsky, D., 1983. N protein of vesicular stomatitis virus selectively encapsidates leader RNA in vitro. Cell 32, 559–567.
- Blumberg, B.M., Kolakofsky, D., 1981. Intracellular vesicular stomatitis virus leader RNAs are found in nucleocapsid structures. J. Virol. 40, 568–576.
- Blumberg, B.M., Leppert, M., Kolakofsky, D., 1981. Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication. Cell 23, 837–845.
- Canter, D.M., Perrault, J., 1996. Stabilization of vesicular stomatitis virus L polymerase protein by P protein binding: a small deletion in the C-terminal domain of L abrogates binding. Virology 219, 376–386.
- Castel, G., Chteoui, M., Caignard, G., Prehaud, C., Mehouas, S., Real, E., Jallet, C., Jacob, Y., Ruigrok, R.W., Tordo, N., 2009. Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. J. Virol. 83, 10808–10820.
- Chen, M., Ogino, T., Banerjee, A.K., 2007. Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N-terminus of P that are involved in N⁰–P complex formation and encapsidation of viral genome RNA. J. Virol. 81, 13478–13485.
- Chen, Z., Green, T.J., Luo, M., Li, H., 2004. Visualizing the RNA molecule in the bacterially expressed vesicular stomatitis virus nucleoprotein–RNA complex. Structure 12, 227–235.
- Chenik, M., Schnell, M., Conzelmann, K.K., Blondel, D., 1998. Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. J. Virol. 72, 1925–1930.
- Chuang, J.L., Perrault, J., 1997. Initiation of vesicular stomatitis virus mutant polR1 transcription internally at the N gene *in vitro*. J. Virol. 71, 1466–1475.
- Cleaveland, S., Fevre, E.M., Kaare, M., Coleman, P.G., 2002. Estimating human rabies mortality in the United Republic of Tanzania from dog bite injuries. Bull. World Health Organ. 80, 304–310.
- Connor, J.H., Lyles, D.S., 2002. Vesicular stomatitis virus infection alters the eIF4F translation initiation complex and causes dephosphorylation of the eIF4E binding protein 4E-BP1. J. Virol. 76, 10177–10187.
- Connor, J.H., Lyles, D.S., 2005. Inhibition of host and viral translation during vesicular stomatitis virus infection. eIF2 is responsible for the inhibition of viral but not host translation. J. Biol. Chem. 280, 13512–13519.
- Curran, J., 1998. A role for the Sendai virus P protein trimer in RNA synthesis. J. Virol. 72, 4274–4280.

- Das, T., Pattnaik, A.K., Takacs, A.M., Li, T., Hwang, L.N., Banerjee, A.K., 1997. Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. Virology 238, 103-114.
- Delmas, O., Assenberg, R., Grimes, J.M., Bourhy, H., 2010. The structure of the nucleoprotein binding domain of Lyssavirus phosphoprotein reveals a structural relationship between the N-RNA binding domains of Rhabdoviridae and Paramyxoviridae. RNA Biol. 7, 322-327.
- Dietzschold, B., Li, J., Faber, M., Schnell, M., 2008. Concepts in the pathogenesis of rabies. Future Virol. 3, 481-490.
- Dillon, P.J., Gupta, K.C., 1988. Early steps in the assembly of vesicular stomatitis virus nucleocapsids in infected cells. J. Virol. 62, 1582–1589.
- Ding, H., Green, T.J., Lu, S., Luo, M., 2006. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. J. Virol. 80, 2808-2814
- Ding, H., Green, T.J., Luo, M., 2004. Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). Acta Crystallogr. D: Biol. Crystallogr. 60, 2087-2090.
- Durbin, J.E., Hackenmiller, R., Simon, M.C., Levy, D.E., 1996. Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease. Cell 84, 443-450.
- Emerson, S.U., 1982. Reconstitution studies detect a single polymerase entry site on the vesicular stomatitis virus genome. Cell 31, 635–642. Emerson, S.U., Schubert, M., 1987. Location of the binding domains for the RNA
- polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. Proc. Natl. Acad. Sci. U.S.A. 84, 5655-5659.
- Emerson, S.U., Yu, Y., 1975. Both NS and L proteins are required for in vitro RNA
- Synthesis by vesicular stomatitis virus. J. Virol. 15, 1348–1356.
 Ferrer-Orta, C., Arias, A., Escarmis, C., Verdaguer, N., 2006. A comparison of viral RNA-dependent RNA polymerases. Curr. Opin. Struct. Biol. 16, 27–34.
 Follett, E.A., Pringle, C.R., Wunner, W.H., Skehel, J.J., 1974. Virus replication in
- enucleate cells: vesicular stomatitis virus and influenza virus. J. Virol. 13, 394-399.
- Fuxreiter, M., Simon, I., Friedrich, P., Tompa, P., 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J. Mol. Biol. 338.1015-1026.
- Gaudin, Y., Barge, A., Ebel, C., Ruigrok, R.W., 1995. Aggregation of VSV M protein is reversible and mediated by nucleation sites: implications for viral assembly. Virology 206, 28-37.
- Gaudin, Y., Tuffereau, C., Durrer, P., Brunner, J., Flamand, A., Ruigrok, R., 1999. Rabies virus-induced membrane fusion. Mol. Membr. Biol. 16, 21-31.
- Ge, P., Tsao, J., Schein, S., Green, T.J., Luo, M., Zhou, Z.H., 2010. Cryo-EM model of the bullet-shaped vesicular stomatitis virus. Science 327, 689-693.
- Gérard, F.C.A., Ribeiro, E., Albertini, A., Zaccai, G., Ebel, C., Ruigrok, R., Jamin, M., 2007. Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. Biochemistry 46, 10328-10338.
- Gérard, F.C.A., Ribeiro, E.A., Leyrat, C., Ivanov, I., Blondel, D., Longhi, S., Ruigrok, R.W.H., Jamin, M., 2009. Modular organization of rabies virus phosphoprotein. J. Mol. Biol. 388, 978-996.
- Grdzelishvili, V.Z., Smallwood, S., Tower, D., Hall, R.L., Hunt, D.M., Moyer, S.A., 2005. A single amino acid change in the L-polymerase protein of vesicular stomatitis virus completely abolishes viral mRNA cap methylation. J. Virol. 79, 7327-7337.
- Green, T.L. Luo, M., 2009. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. Proc. Natl. Acad. Sci. U.S.A. 106, 11721-11726.
- Green, T.J., Macpherson, S., Qiu, S., Lebowitz, J., Wertz, G.W., Luo, M., 2000. Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. J. Virol. 74, 9515-9524.
- Green, T.J., Rowse, M., Tsao, J., Kang, J., Ge, P., Zhou, Z.H., Luo, M., 2010. Access of RNA encapsidated in the nucleocapsid of vesicular stomatitis virus, J. Virol, 85, 2714-2722
- Green, T.J., Zhang, X., Wertz, G.W., Luo, M., 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science 313, 357-360.
- Gupta, A.K., Banerjee, A.K., 1997. Expression and purification of vesicular stomatitis virus N-P complex from Escherichia coli: role in genome RNA transcription and replication in vitro. J. Virol. 71, 4264-4271.
- Hanlon, C.A., Kuzmin, I.V., Blanton, J.D., Weldon, W.C., Manangan, J.S., Rupprecht, C.E., 2005. Efficacy of rabies biologics against new Lyssaviruses from Eurasia. Virus Res. 111, 44-54.
- Harty, R.N., Paragas, J., Sudol, M., Palese, P., 1999. A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding, J. Virol. 73, 2921-2929.
- Heinrich, B.S., Cureton, D.K., Rahmeh, A.A., Whelan, S.P., 2010. Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. PLoS Pathog. 6, e1000958.
- Hemachudha, T., Laothamatas, J., Rupprecht, C.E., 2002. Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges. Lancet Neurol. 1. 101-109.
- Hercyk, N., Horikami, S.M., Moyer, S.A., 1988. The vesicular stomatitis virus L protein possesses the mRNA methyltransferase activities. Virology 163, 222-225.
- Hornung, V., Ellegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K.K., Schlee, M., Endres, S., Hartmann, G., 2006. 5'-Triphosphate RNA is the ligand for RIG-I. Science 314, 994–997.

- Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R.W., Blanchard, L., 2007. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. J. Virol 81 6807-6816
- Howard, M., Wertz, G., 1989. Vesicular stomatitis virus RNA replication: a role for the NS protein. J. Gen. Virol. 70, 2683-2694.
- Hwang, L.N., Englund, N., Pattnaik, A.K., 1998. Polyadenylation of vesicular stomatitis virus mRNA dictates efficient transcription termination at the intercistronic gene junctions. J. Virol. 72, 1805-1813.
- Irie, T., Licata, J.M., McGettigan, J.P., Schnell, M.J., Harty, R.N., 2004. Budding of PPxYcontaining rhabdoviruses is not dependent on host proteins TGS101 and VPS4A. J. Virol. 78, 2657-2665.
- Iseni, F., Barge, A., Baudin, F., Blondel, D., Ruigrok, R.W., 1998. Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. J. Gen. Virol. 79. 2909-2919.
- Iseni, F., Baudin, F., Blondel, D., Ruigrok, R.W., 2000. Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. RNA 6, 270–281. Ivanov, I., Crepin, T., Jamin, M., Ruigrok, R., 2010. Structure of the dimerisation
- domain of the rabies virus phosphoprotein. J. Virol. 84, 3707-3710.
- Iverson, L.E., Rose, J.K., 1981. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. Cell 23, 477-484
- Jackson, A.O., Dietzgen, R.G., Goodin, M.M., Bragg, J.N., Deng, M., 2005. Biology of plant rhabdoviruses. Annu. Rev. Phytopathol. 43, 623–660.
- Jacob, Y., Real, E., Tordo, N., 2001. Functional interaction map of Lyssavirus phosphoprotein: identification of the minimal transcription domains. J. Virol. 75, 9613-9622.
- Karlin, D., Ferron, F., Canard, B., Longhi, S., 2003. Structural disorder and modular organization in Paramysoviridae N and P. J. Gen. Virol. 84, 3239–3252. Knobel, D.L., Cleaveland, S., Coleman, P.G., Fevre, E.M., Meltzer, M.I., Miranda, M.E.,
- Shaw, A., Zinsstag, J., Meslin, F.X., 2005. Re-evaluating the burden of rabies in Africa and Asia. Bull. World Health Organ. 83, 360–368.
- Kolakofsky, D., Le Mercier, P., Iseni, F., Garcin, D., 2004. Viral RNA polymerase scanning and the gymnastics of sendai virus RNA synthesis. Virology 318, 463-473.
- Kolakofsky, D., Roux, L., Garcin, D., Ruigrok, R.W., 2005. Paramyxovirus mRNA editing, the "rule of six" and error catastrophe: a hypothesis. J. Gen. Virol. 86, 1869-1877.
- Kouznetzoff, A., Buckle, M., Tordo, N., 1998. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. J. Gen. Virol. 79, 1005-1013
- Koyama, A.H., 1995. Induction of apoptotic DNA fragmentation by the infection of vesicular stomatitis virus. Virus Res. 37, 285–290. Lahaye, X., Vidy, A., Pomier, C., Obiang, L., Harper, F., Gaudin, Y., Blondel, D., 2009.
- Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: evidence that NBs are sites of viral transcription and replication. J. Virol. 83, 7948-7958.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.
- Lawson, N.D., Stillman, E.A., Whitt, M.A., Rose, J.K., 1995. Recombinant vesicular stomatitis viruses from DNA. Proc. Natl. Acad. Sci. U.S.A. 92, 4477-4481.
- Lenard, J., 1999. Host cell protein kinases in nonsegmented negative-strand virus (mononegavirales) infection. Pharmacol. Ther. 83, 39-48.
- Leyrat, C., Gerard, F.C., de Almeida Ribeiro Jr., E., Ivanov, I., Ruigrok, R.W., Jamin, M., 2010. Structural disorder in proteins of the Rhabdoviridae replication complex. Protein Pept. Lett. 17, 979–987.
- Leyrat, C., Jensen, M.R., Ribeiro, E.A., Gérard, F., Ruigrok, R., Blackledge, M., Jamin, M., 2011a. The N⁰-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α -helices. Protein Sci. 20, 542–556.
- Leyrat, C., Ribeiro, E.A., Gérard, F.C., Ivanov, I., Ruigrok, R.W.H., Jamin, M., 2011b. Structure, interactions with host-cell and functions of rhabdovirus phosphoprotein. Future Virol. 6, 465-481.
- Leyrat, C., Yabukarski, F., Tarbouriech, N., Ribeiro, E.A., Jensen, M.R., Blackledge, M., Ruigrok, R.W., Jamin, M., 2011c. Structure of the vesicular stomatitis virus N⁰-P Complex. PLoS Pathog. 7, e1002248.
- Li, J., Fontaine-Rodriguez, E.C., Whelan, S.P., 2005. Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity. J. Virol. 79, 13373–13384.
- Li, J., Rahmeh, A., Brusic, V., Whelan, S.P., 2009. Opposing effects of inhibiting cap addition and cap methylation on polyadenylation during vesicular stomatitis virus mRNA synthesis. J. Virol. 83, 1930-1940.
- Li, J., Rahmeh, A., Morelli, M., Whelan, S.P., 2008. A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. J. Virol. 82, 775-784.
- Li, J., Wang, J.T., Whelan, S.P., 2006. A unique strategy for mRNA cap methylation used by vesicular stomatitis virus. Proc. Natl. Acad. Sci. U.S.A. 103, 8493-8498.
- Libersou, S., Albertini, A.A., Ouldali, M., Maury, V., Maheu, C., Raux, H., de Haas, F., Roche, S., Gaudin, Y., Lepault, J., 2010. Distinct structural rearrangements of the VSV glycoprotein drive membrane fusion. J. Cell Biol. 191, 199–210. Lichty, B.D., Power, A.T., Stojdl, D.F., Bell, J.C., 2004. Vesicular stomatitis virus: re-
- inventing the bullet. Trends Mol. Med. 10, 210-216.
- Luo, M., Green, T.J., Zhang, X., Tsao, J., Qiu, S., 2007. Conserved characteristics of the rhabdovirus nucleoprotein. Virus Res. 129, 246-251.
- Lynch, S., Kolakofsky, D., 1978. Ends of the RNA within sendai virus defective interfering nucleocapsids are not free. J. Virol. 28, 584-589.

- Masters, P.S., Banerjee, A.K., 1988. Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. J. Virol. 62, 2658–2664.
- Mavrakis, M., Iseni, F., Mazza, C., Schoehn, G., Ebel, C., Gentzel, M., Franz, T., Ruigrok, R.W., 2003. Isolation and characterisation of the rabies virus N degrees–P complex produced in insect cells. Virology 305, 406–414.
- Mavrakis, M., McCarthy, A.A., Roche, S., Blondel, D., Ruigrok, R.W., 2004. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J. Mol. Biol. 343, 819–831.
- Mavrakis, M., Mehouas, S., Real, E., Iseni, F., Blondel, D., Tordo, N., Ruigrok, R.W., 2006. Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. Virology 349, 422–429.
- Mellon, M.G., Emerson, S.U., 1978. Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus. J. Virol. 27, 560–567.
- Moyer, S.A., Banerjee, A.K., 1975. Messenger RNA species synthesized *in vitro* by the virion-associated RNA polymerase of vesicular stomatitis virus. Cell 4, 37–43. Moyer, S.A., Smallwood-Kentro, S., Haddad, A., Prevec, L., 1991. Assembly and
- Moyer, S.A., Smallwood-Kentro, S., Haddad, A., Prevec, L., 1991. Assembly and transcription of synthetic vesicular stomatitis virus nucleocapsids. J. Virol. 65, 2170–2178.
- Navaza, J., Lepault, J., Rey, F.A., Alvarez-Rua, C., Borge, J., 2002. On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. Acta Crystallogr. D: Biol. Crystallogr. 58, 1820–1825.
- Nayak, D., Panda, D., Das, S.C., Luo, M., Pattnaik, A.K., 2009. Single-amino-acid alterations in a highly conserved central region of vesicular stomatitis virus N protein differentially affect the viral nucleocapsid template functions. J. Virol. 83, 5525–5534.
- Ogino, T., Banerjee, A.K., 2007. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. Mol. Cell 25, 85–97.
- Ogino, T., Banerjee, A.K., 2008. Formation of guanosine(5')tetraphospho(5') adenosine cap structure by an unconventional mRNA capping enzyme of vesicular stomatitis virus. J. Virol. 82, 7729–7734.
- Ogino, T., Yadav, S.P., Banerjee, A.K., 2010. Histidine-mediated RNA transfer to GDP for unique mRNA capping by vesicular stomatitis virus RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 107, 3463–3468.
- Pattnaik, A.K., Hwang, L., Li, T., Englund, N., Mathur, M., Das, T., Banerjee, A.K., 1997. Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication. J. Virol. 71, 8167–8175.
- Patton, J.T., Davis, N.L., Wertz, G.W., 1984. N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus. J. Virol. 49, 303–309.
- Peluso, R.W., Moyer, S.A., 1988. Viral proteins required for the *in vitro* replication of vesicular stomatitis virus defective interfering particle genome RNA. Virology 162, 369–376.
- Poch, O., Blumberg, B.M., Bougueleret, L., Tordo, N., 1990. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. J. Gen. Virol. 71, 1153–1162.
- Pringle, C.R., 1997. The order *Mononegavirales* current status. Arch. Virol. 142, 2321–2326.
- Qanungo, K.R., Shaji, D., Mathur, M., Banerjee, A.K., 2004. Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. Proc. Natl. Acad. Sci. U.S.A. 101, 5952–5957.
- Rahmeh, A.A., Li, J., Kranzusch, P.J., Whelan, S.P., 2009. Ribose 2'-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein. J. Virol. 83, 11043–11050.
- Rahmeh, A.A., Schenk, A.D., Danek, E.I., Kranzusch, P.J., Liang, B., Walz, T., Whelan, S.P., 2010. Molecular architecture of the vesicular stomatitis virus RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 107, 20075–20080.
- Rao, B.L., Basu, A., Wairagkar, N.S., Gore, M.M., Arankalle, V.A., Thakare, J.P., Jadi, R.S., Rao, K.A., Mishra, A.C., 2004. A large outbreak of acute encephalitis with high fatality rate in children in Andhra Pradesh, India, in 2003, associated with Chandipura virus. Lancet 364, 869–874.
- Ribeiro Jr., E.A., Favier, A., Gerard, F.C., Leyrat, C., Brutscher, B., Blondel, D., Ruigrok, R.W., Blackledge, M., Jamin, M., 2008. Solution structure of the C-terminal nucleoprotein–RNA binding domains of the vesicular stomatitis virus phosphoprotein. J. Mol. Biol. 382, 525–538.

- Ribeiro, E.A., Leyrat, C., Gérard, F.C., Albertini, A.A., Falk, C., Ruigrok, R.W., Jamin, M., 2009. Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein–RNA complexes. J. Mol. Biol. 394, 558–575.
- Rose, N.F., Publicover, J., Chattopadhyay, A., Rose, J.K., 2008. Hybrid alphavirus-rhabdovirus propagating replicon particles are versatile and potent vaccine vectors. Proc. Natl. Acad. Sci. U.S.A. 105, 5839–5843.
- Rupprecht, C.E., Willoughby, R., Slate, D., 2006. Current and future trends in the prevention, treatment and control of rabies. Expert Rev. Anti Infect. Ther. 4, 1021–1038.
- Schnell, M.J., McGettigan, J.P., Wirblich, C., Papaneri, A., 2010. The cell biology of rabies virus: using stealth to reach the brain. Nat. Rev. Microbiol. 8, 51–61.
- Schnell, M.J., Mebatsion, T., Conzelmann, K.K., 1994. Infectious rabies viruses from cloned cDNA. EMBO J. 13, 4195–4203.
- Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D., Ruigrok, R.W., 2001. Structure of recombinant rabies virus nucleoprotein–RNA complex and identification of the phosphoprotein binding site. J. Virol. 75, 490–498.
- Simonsen, C.C., Batthumphries, S., Summers, D.F., 1979. RNA-synthesis of vesicular stomatitis virus-infected cells – *in vivo* regulation of replication. J. Virol. 31, 124–132.
- Solon, J., Gareil, O., Bassereau, P., Gaudin, Y., 2005. Membrane deformations induced by the matrix protein of vesicular stomatitis virus in a minimal system. J. Gen. Virol. 86, 3357–3363.
- Stillman, E.A., Whitt, M., 1999. Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription. J. Virol. 73, 7199–7209.
- Streicker, D.G., Turmelle, A.S., Vonhof, M.J., Kuzmin, I.V., McCracken, G.F., Rupprecht, C.E., 2010. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science 329, 676–679.
- Szilagyi, J.F., Uryvayev, L., 1973. Isolation of an infectious ribonucleoprotein from vesicular stomatitis virus containing an active RNA transcriptase. J. Virol. 11, 279–286.
- Takacs, A.M., Das, T., Banerjee, A.K., 1993. Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. Proc. Natl. Acad. Sci. U.S.A. 90, 10375–10379.
- Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R.W., Burmeister, W.P., 2000. On the domain structure and the polymerization state of the sendai virus P protein. Virology 266, 99–109.
- Tekes, G., Rahmeh, A.A., Whelan, S.P., 2011. A freeze frame view of vesicular stomatitis virus transcription defines a minimal length of RNA for 5' processing. PLoS Pathog. 7, e1002073.
- Thomas, D., Newcomb, W.W., Brown, J.C., Wall, J.S., Hainfeld, J.F., Trus, B.L., Steven, A.C., 1985. Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. J. Virol. 54, 598–607.
- Tordo, N., Benmansour, A., Calisher, C., Dietzgen, R.G., Fang, R.-X., Jackson, A.O., Kurath, G., Nadin-Davis, S., Tesh, R.B., Walker, P.J., 2004. Rhabdoviridae. In: Fauquet, M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), Virus Taxonomy, VIIIth Report of the ICTV. Elsevier/Academic Press, London, pp. 623–644.
- Toriumi, H., Kawai, A., 2004. Association of rabies virus nominal phosphoprotein (P) with viral nucleocapsid (NC) is enhanced by phosphorylation of the viral nucleoprotein (N). Microbiol. Immunol. 48, 399–409.
- Warrell, M.J., Warrell, D.A., 2004. Rabies and other Lyssavirus diseases. Lancet 363, 959–969.
- Wertz, G.W., 1978. Isolation of possible replicative intermediate structures from vesicular stomatitis virus-infected cells. Virology 85, 271–285.
- Weyer, J., Kuzmin, I.V., Rupprecht, C.E., Nel, L.H., 2008. Cross-protective and cross-reactive immune responses to recombinant vaccinia viruses expressing full-length *Lyssavirus* glycoprotein genes. Epidemiol. Infect. 136, 670–678.
- Whelan, S.P., Ball, L.A., Barr, J.N., Wertz, G.T., 1995. Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. Proc. Natl. Acad. Sci. U.S.A. 92, 8388–8392.
- Whelan, S.P., Wertz, G.W., 2002. Transcription and replication initiate at separate sites on the vesicular stomatitis virus genome. Proc. Natl. Acad. Sci. U.S.A. 99, 9178–9183.
- Zhang, X., Green, T.J., Tsao, J., Qiu, S., Luo, M., 2008. Role of intermolecular interactions of vesicular stomatitis virus nucleoprotein in RNA encapsidation. J. Virol. 82, 674–682.
- Zhang, X., Wei, Y., Ma, Y., Hu, S., Li, J., 2010. Identification of aromatic amino acid residues in conserved region VI of the large polymerase of vesicular stomatitis virus is essential for both guanine-N-7 and ribose 2'-O methyltransferases. Virology 408, 241–252.

Structural Disorder in Proteins of the Rhabdoviridae Replication Complex

Cédric Leyrat, Francine C.A. Gérard, Euripedes de Almeida Ribeiro Jr., Ivan Ivanov, Rob W.H. Ruigrok and Marc Jamin*

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

Abstract: *Rhabdoviridae* are single stranded negative sense RNA viruses. The viral RNA condensed by the nucleoprotein (N), the phosphoprotein (P) and the large subunit (L) of the RNA-dependent RNA polymerase are the viral components of the transcription/replication machineries. Both P and N contain intrinsically disordered regions (IDRs) that play different roles in the virus life cycle. Here, we describe the modular organization of P based on recent structural, biophysical and bioinformatics data. We show how flexible loops in N participate in the attachment of P to the N-RNA template by an induced-fit mechanism. Finally, we discuss the roles of IDRs in the mechanism of replication/transcription, and propose a new model for the interaction of the L subunit with its N-RNA template.

Keywords: Rhabdoviridae, replication complex, phosphoprotein, nucleoprotein, intrinsically disordered regions.

THE RHABDOVIRIDAE

Rabies virus (RV) and vesicular stomatitis virus (VSV) are prototypic members of the large family Rhabdoviridae that includes numerous pathogens of plants and animals, including man, and are grouped in the Mononegavirales (MNV) order with the Filoviridae (Ebola and Marburg viruses), the Paramyxoviridae (measles, mumps, respiratory syncytial viruses) and the Bornaviridae (Borna disease virus). Rhabdoviridae are enveloped viruses with an overall bullet or bacillus shape, whose genome is made of a linear single-stranded molecule of negative sense RNA. They vary widely in terms of cellular interactions and replicate either in the cytoplasm or in the nucleus, but they share similar genome and structural organizations as well as similar modes of RNA replication and transcription [1, 2]. The viral genome (9 to 18 kb) of the Rhabdoviridae comprises up to 10 genes, but both RV and VSV genomes contain only the five genes that are common to all members of the MNV (Fig. 1a). They encode successively from the 3' terminus, the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the large subunit of the RNA polymerase (L). Only three of these viral proteins, N, P and L, are required to synthesize viral RNA in an efficient and regulated manner [3, 4].

The viral genomic RNA of *Rhabdoviridae* is always encapsidated by N (Fig. **1b**) forming a helical ribonucleocapsid (NC) in which each N protomer binds to nine nucleotides. This N-RNA complex serves as the active template for both transcription and replication by the two-subunit P-L RNA-dependent RNA polymerase [5]. When expressed in bacteria or insect cells, recombinant N protein binds to cellular RNAs and forms circular N_m -RNA complexes in addition to long

NCs [6-8] (the subscript m denotes the number of N protomers *per* ring). Circular N_m-RNA complexes containing different numbers of N protomers can be separated by preparative gel electrophoresis [9]. The crystal structure of circular N_m-RNA complexes from RV (m = 11) and VSV (m = 10) revealed that N is a two-domain protein that enwraps completely the viral RNA molecule and hides it from the polymerase and from the innate immune system [10, 11].

P is essential for viral replication and plays multiple roles at different stages of the viral cycle [12]. Firstly, P acts as a chaperone of N, by forming N⁰-P complexes (where the superscript⁰ denotes the absence of RNA) thus preserving N in a RNA-free form until N is transferred, by an unknown mechanism, from P to the nascent viral RNA in order to form new NCs [13-15]. Secondly, P forms with the L protein a two-subunit RNA-dependent RNA polymerase complex [12, 16]. L carries out all enzymatic activities of transcription and replication, namely RNA synthesis, mRNA cap synthesis and mRNA polyadenylation, while P acts as an essential cofactor although its precise role remains largely unknown. The only identified role of P in this complex is to mediate the attachment of the L subunit to the N-RNA template. L makes no direct interaction with N, but P binds to the N-RNA template through its C-terminal domain and to the L subunit through its N-terminal part [17], and thus provides a physical link between the polymerase and its N-RNA template. In order to synthesize both messenger and genomic RNA, the L polymerase must move along its N-RNA template. L is a processive enzyme [18] and P must maintain the attachment of L to its template during its displacements.

Recent studies have shown that P and N proteins from viruses of the *Paramyxoviridae* contain long Intrinsically Disordered Regions (IDRs) that alternate with structured domains [19-24]. Both amino acid sequence analysis and experimental data also suggested the presence of IDRs in P and N proteins from RV and VSV [19, 25-27]. The P protein from both viruses is highly sensitive to proteolysis and struc-

0929-8665/10 \$55.00+.00

© 2010 Bentham Science Publishers Ltd.

^{*}Address correspondence to this author at the Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France; Tel: + 33 4 76 20 94 62; Fax: + 33 4 76 20 94 00; E-mail: jamin@embl.fr



Figure 1. Genome organization and morphology of the *Rhabdoviridae*. (a) Genome organization. The gene products are denoted as follows: N, nucleoprotein, P, phosphoprotein, M, matrix protein, G, surface glycoprotein and L, large subunit of the RNA-dependent RNA polymerase complex. Le and Tr indicate the leader and trailer RNA sequences. (b) Schematic representation of a *Rhabdoviridae* particle. G is a glycoprotein embedded in the lipidic membrane, involved both in virus entry into host cells and in virus budding. M is the matrix protein, involved in assembly of viral particles and in virus budding. N, P and L plus the genomic RNA constitute the replication unit.

tured domains in RV and VSV P were identified by limited proteolysis [25-27]. In RV N, limited proteolysis eliminates the C-terminal part of the protein and abrogates its interaction with P, although it has no effect on the N-N interactions, also suggesting the presence of a disordered region in N [6, 28].

DISORDER PREDICTIONS USING THE D-SCORE

Recently, the structure/function paradigm was challenged with the discovery that many proteins are intrinsically disordered (IDPs) or contain intrinsically disordered regions (IDRs) under physiological conditions [29-32], and that disorder is required for, or participates in the biological function of proteins [29, 32-39]. Some proteins are disordered in isolation but fold into a well-defined structure upon binding to a partner [38, 40-43]. The recent appraisal of the importance of IDPs and IDRs in the protein world raised new issues such as understanding the role and influence of disorder in the stability, dynamics and interaction with molecular partners, and predicting the boundaries between disordered and structured regions from the amino acid sequence. To solve this latter problem, various algorithms were developed that rely on different approaches (for recent reviews, see [44, 45]). However, these different predictors yield different results, with large discrepancies in the number, position and length of the predicted disordered regions [46, 47]. These discrepancies argue for the existence of different types ("flavors") of disorder [45, 48], but raise the problem of choosing the best algorithm and/or strategy for locating the boundaries of disordered regions in a given protein [44, 45]. To avoid this choice, we recently devised a simple procedure that integrates predictions from multiple algorithms and relies on a simple binary voting procedure [17, 49]. The amino acid sequence is submitted to different algorithms available through WEB servers. For each prediction, each residue is ranked in a binary manner as ordered or disordered using the default thresholds set in the WEB servers and without accounting for the statistical weights provided by some servers [17]. 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. For each residue, the values obtained from the different predictors are added and divided by the number of used algorithms in order to normalize the score. The calculated D-score (Figs. 2a, 3a and 4c) varies from 0 to 1, and consensus disordered regions are arbitrarily defined as regions with a normalized D-score \leq 0.50. This approach was successful for locating known structured domains within the phosphoproteins of Sendai virus [21, 22, 25, 49, 50] and rabies virus [26] and for identifying unknown structured domains in VSV [49] and RV [51]. This analysis, backed by biochemical and biophysical data led to a model of modular organization of Rhabdoviridae P proteins [17, 52].

MODULAR ORGANIZATION OF *RHABDOVIRIDAE* P

On the basis of disorder predictions and of the structural characterization of different domains, a structural map is emerging. The D-score (Figs. **2a** and **3a**) calculated from the amino acid sequence suggests the presence of four structured regions separated by three IDRs in VSV P (Figs. **2b** and **2c**) and of three structured regions separated by two IDRs in RV

Structural Disorder in Proteins of the Rhabdoviridae

P (Figs. **3b** and **3c**), respectively. A comparison between VSV and RV P revealed the existence of common functional and structural regions, and suggested a modular organization for *Rhabdoviridae* P [17] similar to that previously proposed for *Paramyxoviridae* P [19].



Figure 2. Modular organization of VSV P. (a) Score for disorder (D-score) as a function of residue number. Consensus disordered regions are defined as regions with a D-score < 0.50. (b) Location of consensus disordered regions with a D-score < 0.50. Numbers below the boxes indicate the boundaries of the consensus disordered regions. (c) Structural model showing the known three-dimensional structures as ribbon diagrams within space-filled model, the IDR as dotted lines and the regions that putatively fold upon binding to partners as dotted ovals. The small circles show the position of phosphorylation sites involved in transcriptional regulation.

In both RV and VSV P, the D-score predicts a structured N-terminal region (P_{NTD}) encompassing the first 40 to 60 residues. This region contains the binding site for N⁰ as protein fragments corresponding to P_{NTD} of both VSV P and RV P are sufficient for maintaining N in a soluble, RNA-free form [14, 15]. However, in isolation these fragments are disordered [17]. The Stokes's radius of P_{NTD} measured by size exclusion chromatography is that of a fully unfolded protein of this size, and circular dichroism spectroscopy reveals low amounts of secondary structure [17]. This part of the protein is highly acidic and could mimic the RNA in its electrostatic interactions with N⁰ and could fold only upon binding to N⁰. The RNA binding groove in N is narrow (Fig. **4a**) and inserting P_{NTD} into this cavity might require it to be unfolded in the initial stages of the binding process.

The D-score predicts two structured domains in the central part of VSV P (P_{CED1} : amino acids 91-106 and P_{CED2} :



Figure 3. Modular organization of RV P. (a) Score for disorder (D-score) as a function of residue number. Consensus disordered regions are defined as regions with a D-score < 0.50. (b) Location of consensus disordered regions with a D-score < 0.50. Numbers below the boxes indicate the boundaries of the consensus disordered regions. (c) Structural model showing the known three-dimensional structures as ribbon diagrams within space-filled model, the IDR as dotted lines and the regions that putatively fold upon binding to partners as dotted ovals.

amino acids 132-185), and one structured region in that of RV P (P_{CED}: amino acids 87-132) (Fig. 3a and 3b). VSV P_{CED1} is conserved among VSV strains [17] and is part of one of the binding sites proposed for L [53]. There is no experimental evidence that P_{CED1} is folded in free P [27], and it may fold only upon binding to its partner. No equivalent structured or conserved region is found in RV P [17], although a major binding site for L has been similarly located in the 40-100 region by two-hybrid analysis [54]. In addition, peptides corresponding to the first 57 or 60 residues of RV P, and thus overlapping the N⁰ and L binding sites, inhibit the transcription/replication of a rabies virus minigenome as well as rabies infection in kidney cells and in two neuronal cell lines [54]. This study however did not clearly demonstrate whether the inhibition is caused by inhibiting the formation of N⁰-P or by preventing the binding to L. Both P proteins form oligomers [52]. For VSV, it has been proposed that phosphorylation of P controls transcription by modifying the oligomeric state of P [55, 56]. Size exclusion chromatography combined with detection by multi-angle laser light scattering (SEC-MALLS) and small-angle neutron scattering (SANS) revealed that recombinant P proteins form dimers in solution [52]. Previous studies showed that the region responsible for the oligomerization is localized in the central part of P (VSV P, amino acid 107-177; RV P, amino acid 52



Figure 4. The Flexible N_{CTD} loop participates in binding P_{CTD} . (a) Ribbon diagram within a space-filled model of the N protomer in a lateral view showing that the RNA binding groove is closed. The N_{NTD} and N_{CTD} loops are shown as dotted lines. (b) Structure of the complex formed between RV P_{CTD} and the N_{11} -RNA complex obtained by molecular docking and modeling. The ribbon representation shows the location of P_{CTD} at the surface of the circular N_{11} -RNA ring. The meshed structure shows the circular N_{11} -RNA complex. The light gray ribbon representations show the three N_{CTD} domains used in the molecular modeling and the dark gray ribbon representation shows the P_{CTD} . The N_{CTD} loop of protomers N_i and N_{i-1} are shown, as well as helix α_6 of P_{CTD} and helix α_{14} of the N_i protomer. (c) D-score calculated for RV N.

to 189) [3, 57, 58]. The central region of VSV P corresponding to residues 107 to 177 was identified as a folded domain by limited proteolysis [27] and the crystal structure revealed that it forms dimers [25]. The monomer is made of a central α -helix preceded and followed by a two-stranded β -sheet. In the dimer, the helices from each monomer pack against each other in a parallel orientation, and at both sides of the helices, the N-terminal B-sheet of one monomer forms a four stranded β -sheet with the C-terminal β -sheet of the other monomer [52] (Fig. 2c). The first part of the dimerization domain, forming the N-terminal β -sheet, is predicted to be disordered by the D-score (amino acids 107-131). However, in the crystal structure, it appears to be folded, stabilized by intermolecular interactions, suggesting that it could fold only upon forming the dimer. The first β -strand is rich in charged residues and poor in hydrophobic residues, sequence features that are typically associated with disorder, and which could explain why most disorder predictors, and consequently the D-score, stated this part as disordered. In RV P, the D-score locates more precisely the central domain within residues 87 to 132. Two-hybrid experiments confirmed that this region contains the dimerization domain [17]. The isolated central region of RV P forms an autonomous folding unit and SEC-MALLS experiments confirmed that it forms dimers in solution, whereas the protein deleted from this central region $(P_{\Delta 91-131})$ forms monomers [17]. Recently, the structure of the dimerization domain (amino acids 91-133) was solved by Xray crystallography [51]. The monomer is made of two α helices forming a hairpin, and the dimer is assembled by packing two monomers in a parallel orientation with an angle of about 50°. This structure is so different from that of VSV P_{CED2} that it is impossible to determine whether these domains have evolved from a common ancestor [51]. The largest functional difference between the two dimerization domains is that in the RV P dimer the N-terminal end is positioned next to the C-terminal end whereas in the VSV P dimer the N- and C-terminal ends are on opposite sides of the dimerization domain. However, this difference may not have biological consequences because the N- and C-terminal domains are connected to the dimerization domain by disordered regions [17], allowing a similar topological arrangement of the domains (Figs. **2c** and **3c**).

The last predicted structured region in RV P (amino acid 195-297) corresponds closely to the C-terminal N-RNA binding domain (P_{CTD}), whose structure was solved by X-ray crystallography (amino acids 186-296) [26] (Fig. **3c**). The equivalent domain was correctly localized in VSV P from the D-score analysis (amino acids 197-265), the isolated domain (amino acids 195-265) behaved as an autonomous folding unit and its structure was solved by NMR [49]. A longer form of this domain (amino acids 185-265) showed that the preceding ten residues (amino acids 185-194) are disordered, confirming the existence of the IDR_{CTD} [49]. Although multiple sequence alignments indicate almost no sequence similarity between P proteins from VSV and RV, the structure of VSV P_{CTD} is homologous to that of RV P. The topology of the backbone is conserved, although VSV P_{CTD} lacks helices

Leyrat et al.

Structural Disorder in Proteins of the Rhabdoviridae

FLEXIBLE LOOPS IN N PARTICIPATE IN BINDING OF P

In the crystal structure of circular N-RNA complexes from RV and VSV, the N protomer consists of two main domains, which completely enwrap the single-stranded RNA [10, 11] (Fig. 4a). The N-terminal core domain (NTD) folds into a helical arrangement composed of helices connected by large loops. The C-terminal core domain (CTD) is composed of helices joined by tighter loops. Both the NTD and CTD act as jaws that clamp down onto the RNA strand and enclose it completely, an observation which is consistent with the fact that the RNA remains bound to N in CsCl gradients [60]. In addition to the core domains, two smaller subdomains participate in domain exchange between protomers and stabilize polymeric forms [10, 11].

In the RV structure, two regions, the N_{NTD} loop (amino acids 105 to 118) and the N_{CTD} loop (amino acids: 376 to 397), are flexible and absent in the crystal structure. Limited trypsin digestion cleaves the protein at Lys376 [28]. Trypsinated N-RNA complexes lose their ability to bind P_{CTD} although they retain their overall structure [6, 28]. The calculated D-score predicts a disordered region between residue 371 and residue 399 (Fig. **4c**), in good agreement with the location of the flexible N_{CTD} loop.

Recently, the structures of the complexes between the N-RNA template and the C-terminal domain of P were obtained for both RV and VSV. A multiple stage flexible docking procedure was used to locate the binding site for the Cterminal domain of P on the N-RNA complex of RV and to predict the structure of the complex [61] in agreement with biological data [57]. The structure of the equivalent N-RNA-PCTD complex of VSV was determined by X-crystallography [62]. In both VSV and RV complexes, the P_{CTD} lies on the top of a N protomer (Ni) with the C-terminal helix of PCTD (α_6 in RV, α_4 in VSV) lying on the antepenultimate helix (α 14 in RV, α_{13} in VSV) of Ni [10, 11, 63]. By an induced-fit mechanism, the NCTD loops of the same protomer (Ni) and of the adjacent one (Ni-1) mold around the PCTD, making extensive protein-protein contacts that could explain the submicromolar affinity of P for its template [61]. In both VSV and RV, the P_{CTD} exhibits the same orientation on the top side of the N-RNA ring, its N-terminal extremity pointing toward the centre of the ring. For the RV complex, titration experiments revealed that binding sites for isolated P_{CTD} on the circular N_m-RNA complexes are independent.

The main differences between the RV and VSV complexes reside in the numbers of residues involved in the interaction and in the surface area buried in the complex. In the RV complex, 34 residues of P_{CTD} and 39 residues of N (28 in N_i and 11 in N_{i+1}) participate in the binding against 18 and 17 residues in the VSV P_{CTD} and N, respectively. The N_{CTD} loop is longer in RV than in VSV, and the loops from two adjacent N protomers form a larger clamp in RV that hides a larger part of P_{CTD} than in VSV. In the RV complex, the total surface area of P_{CTD} (total surface area = 6,465 Å²) buried upon interaction with two adjacent N molecules is 1,739 Å² (~27% of the total surface area), significantly larger than in the VSV complex for which only 956 Å² are buried (~19% of the total surface area). The other important difference is in the occupancy of the potential $P_{\rm CTD}$ binding sites on the N-RNA ring. In the VSV crystal structure, one $P_{\rm CTD}$ is bound to each N protomer, although reduced occupancy was observed in some positions [62]. In the RV complex, the involvement of N_{CTD} loops from adjacent protomers in binding one $P_{\rm CTD}$ makes it unlikely to reach a stoichiometry larger than one $P_{\rm CTD}$ for two N protomers.

Folding of an IDR upon binding to its partner is also involved in the attachment of P to the N-RNA template in *Paramyxoviridae* [64-66]. In this other *MNV* family, the C-terminal part of N, named N_{TAIL}, is an IDR containing a short region that transiently populates α -helical conformers [65] and that folds upon binding to the C-terminal P_X domain [21] (also, see reviews by Longhi and Oglesbee and by Jensen *et al.* in this issue).

ROLE OF STRUCTURAL DISORDER IN PROTEINS OF THE *RHABDOVIRIDAE* REPLICATION COM-PLEX

Disordered regions have been found in many functional proteins [30] and have been associated with specific functions [32, 38]. Some of these roles can be assigned to disordered regions found in P and N proteins of both RV and VSV. Firstly, disordered regions have major roles in exposing sites of posttranslational modifications, such as phosphorylation [32, 37]. In VSV, phosphorylation of residues located in the IDR_{NTD} regulates transcriptional activation [55, 56]. Other phosphorylation sites located in the structured P_{CTD} have no known functional roles [67] (Fig. 2c). Phosphorylation sites are also found in the IDR_{NTD} of RV P, although no role could be assigned to these modifications [58, 68]. Secondly, IDRs and IDPs have been shown to participate in multi-molecular assemblies by simply bridging together different partners, by regulating the spacing between binding sites, by providing a large binding surface or multiple contact points for a large protein or by adopting different conformations in response to different stimuli [32, 38]. Often IDRs fold upon binding to their partners, a process accompanied by a large decrease in conformational entropy that can provide a high specificity associated with a moderate binding strength [38, 69]. RV and VSV P play key roles in the assembly of the replication/transcription complex by binding to the L-polymerase, [53, 54, 70], to N in the N^0 -P complex and to N in the N-RNA template [14, 15, 71-73]. The attachment of P to the N-RNA template involves the closure around P_{CTD} of the N_{CTD} loops from two adjacent N protomers [61, 62]. This peculiar binding site provides an explanation to why binding of P to N through P_{CTD} can only occur with the N-RNA complex. The IDR_{NTD} of RV and VSV P contains two binding sites for viral partners, one for N₀ and the other for the L polymerase. The binding site for N_0 (P_{NTD}) in RV and VSV and the binding site for L in VSV are predicted to be structured but appear as disordered in the absence of their partners, while the binding site for L in RV is located in the IDR_{NTD} and was thus not predicted as a structured region [17]. The involvement of IDRs in the interactions with these viral partners suggests that these regions

984 Protein & Peptide Letters, 2010, Vol. 17, No. 8

could fold upon binding. Because these interactions are transient during the replication process, folding upon binding could provide the high specificity required for the correct assembly of the viral machineries and a modest or moderate affinity allowing the rearrangement of these multi-molecular complexes [38, 69]. In the case of RV P, the IDR_{CTD} also contains a binding site for the cellular dynein light chain LC8 [74, 75]. Finally, IDPs and IDRs can provide a mean for establishing long-range interactions between partner proteins or for increasing the speed of interaction through the socalled "fly-casting" mechanism [41]. The location of binding sites for N⁰ and L in the long disordered N-terminal region (P_{NTD} + IDR_{NTD}) could serve for catching these viral components and for bringing them to the site of RNA synthesis on the N-RNA template.

Recent data led us to propose a new model for the Rhabdoviridae replication complex [17, 49, 52, 61]. The affinity of P_{CTD} for the N_m-RNA template is high, with dissociation constants in the submicromolar range $(160 \pm 20 \text{ nM})$ [61]. Moreover, the association and dissociation kinetics measured by surface plasmon resonance (SPR) are slow (tens of seconds time range) as compared to the time scale of viral replication [61]. Both observations are difficult to conciliate with a mechanism involving continuous association and dissociation of P_{CTD} from the N_m-RNA template. In addition, in Rhabdoviridae the central dimerization seems to be dispensable for RNA synthesis [57, 76]. Recently, we showed that RV P deleted from this central domain, $P_{\Delta 91-131}$, is monomeric [17] and this mutant is functional in a minigenome transcription assay (Tordo N., personal communication). These results suggest that P dimers are continuously attached to the N-RNA template in the virion and in the cytoplasm of infected cells (Fig. 5). The model obtained for the RV N-RNA-P_{CTD} complex indicates that two protomers of N are required for binding one P_{CTD}, suggesting that one P dimer could attach every four N protomers, whereas the estimated number of P dimers in the VSV virion [77] would result in one P dimer attached every five N protomers. When P is attached to the N-RNA template through its C-terminal domains, the binding sites for N⁰ and L, located in the Nterminal part of P could thus extend far from the N-RNA template to catch their partners (Fig. 5). This could allow the polymerase to move along the N-RNA template by jumping between adjacent P dimers and also to bring N⁰ to the site of assembly of the new NC. This mechanism contrast with the cartwheeling mechanism proposed for the Paramyxoviridae [1, 78]. In this model, P, which forms tetramers through its central coiled-coil domain, would rotate around its central coiled-coil axle with the four C-terminal PX domains continuously associating and dissociating from the N-RNA template and thus carrying along their L cargo. There are two main differences in this part of the replication machinery between Rhabdoviridae and Paramyxoviridae. Firstly, the oligomerization domain of P seems to be dispensable for the synthesis of RNA in Rhabdoviridae, whereas a mutant lacking the central tetramerization domain in Sendai virus is unable to synthesize RNA [1, 78]. Secondly, affinities of PX for N_{TAIL} measured by NMR titration or isothermal titration calorimetry yielded dissociation constants in the range of 10 to 50 µM for both measles virus and Sendai virus proteins [21, 64, 79-82] and are thus weaker that that found for RV.

However, for measles virus values near 100 nM were also obtained by SPR [83] suggesting a more complex picture (see the review by Longhi and Oglesbee in this issue).



Figure 5. Schematic model for the attachment of P, L and N⁰ to the N-RNA template. The modular P protein is attached at regular intervals along the N-RNA template through its C-terminal N-RNA binding domain. The N-terminal IDRs that contain the binding sites for both N⁰ and L, are thus stretching away from the template, "flycasting" their partner proteins and bringing them to the template. Binding to their partners could be coupled to folding of the interacting regions (light gray circles). The bars represent the sub-domains of N interacting with the adjacent protomer on each side. The circled P indicate the putative phosphorylations of Ser389.

ACKNOWLEDGMENTS

This work was supported by the interdisciplinary program "Maladies Infectieuses Emergentes" from the CNRS and by a grant from the French ANR (ANR-07-001-01 (ANRAGE)) and Lyonbiopôle. The authors thank Danielle Blondel, Noël Tordo, Martin Blackledge, Malene Jensen, Thibaut Crépin, Giuseppe Zaccai and Sonia Longhi for fruitful discussions and collaborations, and the Partnership for Structural Biology for the excellent structural biology environment.

ABBREVIATIONS

MNV	=	Mononegavirales order
VSV	=	Vesicular stomatitis virus
RV	=	Rabies virus
RNA	=	Ribonucleic acid
Ν	=	Nucleoprotein
N-RNA	=	Nucleoprotein-ribonucleic acid complex
Р	=	Phosphoprotein
L	=	Large subunit of the RNA-dependent RNA polymerase
NC	=	Nucleocapsid
N ⁰ -P	=	Complex between the RNA-free nucleoprotein and the phosphoprotein
P _{CTD}	=	C-terminal domain of P
P _{CED}	=	Central domain of P

Structural Disorder in Proteins of the Rhabdoviridae

P _{NTD}	=	N-terminal domain of P
N _{NTD}	=	N-terminal domain of N
N _{CTD}	=	C-terminal domain of N
IDR	=	Intrinsically disordered region
IDP	=	Intrinsically disordered protein
SEC	=	Size exclusion chromatography
MALLS	=	Multi-angle laser light scattering
SANS	=	Small-angle neutron scattering
SAXS	=	Small-angle X-ray scattering

- NMR = Nuclear magnetic resonance
- SPR = Surface plasmon resonance

REFERENCES

- Kolakofsky, D.; Le Mercier, P.; Iseni, F.; Garcin, D. Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, 2004, 318, 463-473.
- [2] Whelan, S. P.; Barr, J. N.; Wertz, G. W. Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr. Top. Microbiol. Immunol.*, 2004, 283, 61-119.
- [3] Pattnaik, A. K.; Hwang, L.; Li, T.; Englund, N.; Mathur, M.; Das, T.; Banerjee, A. K. 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.*, **1997**, 71, 8167-8175.
- [4] Gupta, A. K.; Shaji, D.; Banerjee, A. K. Identification of a novel tripartite complex involved in replication of vesicular stomatitis virus genome RNA. *J. Virol.*, 2003, 77, 732-738.
- [5] Arnheiter, H.; Davis, N. L.; Wertz, G.; Schubert, M.; Lazzarini, R. A. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell*, **1985**, *41*, 259-267.
- [6] Schoehn, G.; Iseni, F.; Mavrakis, M.; Blondel, D.; Ruigrok, R. W. Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J. Virol.*, 2001, 75, 490-498.
- [7] Green, T. J.; Macpherson, S.; Qiu, S.; Lebowitz, J.; Wertz, G. W.; Luo, M. Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. J. Virol., 2000, 74, 9515-9524.
- [8] Iseni, F.; Barge, A.; Baudin, F.; Blondel, D.; Ruigrok, R. W. Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. J. Gen. Virol., 1998, 79 (Pt 12), 2909-2919.
- [9] Albertini, A. A.; Clapier, C. R.; Wernimont, A. K.; Schoehn, G.; Weissenhorn, W.; Ruigrok, R. W. Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings. J. Struct. Biol., 2006, 158, 129-133.
- [10] Albertini, A. A.; Wernimont, A. K.; Muziol, T.; Ravelli, R. B.; Clapier, C. R.; Schoehn, G.; Weissenhorn, W.; Ruigrok, R. W. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science*, 2006, 313, 360-363.
- [11] Green, T. J.; Zhang, X.; Wertz, G. W.; Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science*, 2006, 313, 357-360.
- [12] Emerson, S. U.; Yu, Y. Both NS and L proteins are required for *in vitro* RNA synthesis by vesicular stomatitis virus. *J Virol*, 1975, 15, 1348-1356.
- [13] Masters, P. S.; Banerjee, A. K. Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. J. Virol., 1988, 62, 2658-2664.
- [14] Mavrakis, M.; Mehouas, S.; Real, E.; Iseni, F.; Blondel, D.; Tordo, N.; Ruigrok, R. W. Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology*, **2006**, *349*, 422-429.
- [15] Chen, M.; Ogino, T.; Banerjee, A. K. Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. J. Virol., 2007, 81, 13478-13485.

Protein & Peptide Letters, 2010, Vol. 17, No. 8 985

- [16] De, B. P.; Banerjee, A. K. Specific interactions of vesicular stomatitis virus L and NS proteins with heterologous genome ribonucleoprotein template lead to mRNA synthesis *in vitro*. J. Virol., 1984, 51, 628-634.
- [17] Gérard, F. C. A.; Ribeiro, E. A.; Leyrat, C.; Ivanov, I.; Blondel, D.; Longhi, S.; Ruigrok, W. H.; Jamin, M. Modular organization of rabies virus phosphoprotein. J. Mol. Biol., 2009, 388, 978-996.
- [18] Stillman, E. A., and Whitt, M. Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription. J. Virol., 1999, 73, 7199-7209.
- [19] Karlin, D.; Ferron, F.; Canard, B.; Longhi, S. Structural disorder and modular organization in Paramyxovirinae N and P. J. Gen. Virol., 2003, 84, 3239-3252.
- [20] Longhi, S. Nucleocapsid structure and function. Curr. Top. Microbiol. Immunol., 2009, 329, 103-128.
- [21] Blanchard, L.; Tarbouriech, N.; Blackledge, M.; Timmins, P.; Burmeister, W. P.; Ruigrok, R. W.; Marion, D. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*, 2004, 319, 201-211.
- [22] Tarbouriech, N.; Curran, J.; Ruigrok, R. W.; Burmeister, W. P. Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat. Struct. Biol.*, 2000, 7, 777-781.
- [23] Llorente, M. T.; Garcia-Barreno, B.; Calero, M.; Camafeita, E.; Lopez, J. A.; Longhi, S.; Ferron, F.; Varela, P. F.; Melero, J. A. Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. J. Gen. Virol., 2006, 87, 159-169.
- [24] Bourhis, J. M.; Canard, B.; Longhi, S. Structural disorder within the replicative complex of measles virus: functional implications. *Virology*, 2006, 344, 94-110.
- [25] Ding, H.; Green, T. J.; Lu, S.; Luo, M. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. J. Virol., 2006, 80, 2808-2814.
- [26] Mavrakis, M.; McCarthy, A. A.; Roche, S.; Blondel, D.; Ruigrok, R. W. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J. Mol. Biol., 2004, 343, 819-831.
- [27] Ding, H.; Green, T. J.; Luo, M. Crystallization and preliminary Xray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). Acta Crystallogr. D Biol. Crystallogr., 2004, 60, 2087-2090.
- [28] Kouznetzoff, A.; Buckle, M.; Tordo, N. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. J. Gen. Virol., 1998, 79(Pt 5), 1005-1013.
- [29] Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. J. Mol. Biol., 1999, 293, 321-331.
- [30] Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, 2000, 41, 415-427.
- [31] Uversky, V. N. What does it mean to be natively unfolded? *Eur. J. Biochem.*, **2002**, *269*, 2-12.
- [32] Tompa, P. Intrinsically unstructured proteins. *Trends Biochem Sci.*, 2002, 27, 527-533.
- [33] Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. Intrinsic disorder in transcription factors. *Biochemistry*, 2006, 45, 6873-6888.
- [34] Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry*, 2002, 41, 6573-6582.
- [35] Dunker, A. K.; Brown, C. J.; Obradovic, Z. Identification and functions of usefully disordered proteins. *Adv. Protein Chem.*, 2002, 62, 25-49.
- [36] Romero, P.; Obradovic, Z.; Dunker, A. K. Natively disordered proteins: functions and predictions. *Appl. Bioinformatics*, 2004, 3, 105-113.
- [37] Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.*, 2004, 32, 1037-1049.
- [38] Dyson, H. J.; Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.*, 2002, 12, 54-60.
- [39] Tompa, P., and Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.*, 2004, 18, 1169-1175.

986 Protein & Peptide Letters, 2010, Vol. 17, No. 8

- [40] Spolar, R. S., and Record, M. T., Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **1994**, 263, 777-784.
- [41] Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 2007, 447, 920-921.
- [42] Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M., and Wright, P. E. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*, **2002**, *415*, 549-553.
- [43] Uversky, V. N.; Segel, D. J.; Doniach, S.; Fink, A. L. Associationinduced folding of globular proteins. *Proc. Natl. Acad. Sci. USA*, 1998, 95, 5480-5483.
- [44] Bourhis, J. M.; Canard, B.; Longhi, S. Predicting protein disorder and induced folding: from theoretical principles to practical applications. *Curr. Protein Pept. Sci.*, 2007, 8, 135-149.
- [45] Ferron, F., Longhi, S., Canard, B., and Karlin, D. A practical overview of protein disorder prediction methods. *Proteins*, 2006, 65, 1-14.
- [46] Lieutaud, P.; Canard, B.; Longhi, S. MeDor: a metaserver for predicting protein disorder. *BMC Genomics*, 2008, 9(Suppl. 2), S25.
- [47] Ishida, T.; Kinoshita, K. Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics*, 2008, 24, 1344-1348.
- [48] Vucetic, S.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. Flavors of protein disorder. *Proteins*, 2003, 52, 573-584.
- [49] Ribeiro, E. A.; Jr., Favier, A.; Gérard, F. C.; Leyrat, C.; Brutscher, B.; Blondel, D.; Ruigrok, R. W.; Blackledge, M.; Jamin, M. Solution Structure of the C-Terminal Nucleoprotein-RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein. J. Mol. Biol., 2008, 382, 525-538.
- [50] Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W.; Blackledge, M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 17002-17007.
- [51] Ivanov, I.; Crepin, T.; Jamin, M.; Ruigrok, R. Structure of the dimerisation domain of the rabies virus phosphoprotein. J. Virol., 2010, 84, 3707-3710.
- [52] Gérard, F.; Ribeiro, E.; Albertini, A.; Zaccai, G.; Ebel, C.; Ruigrok, R.; Jamin, M. Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry*, 2007, 46, 10328-10338.
- [53] Emerson, S. U.; Schubert, M. Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA*, **1987**, *84*, 5655-5659.
 [54] Castel, G.; Chteoui, M.; Caignard, G.; Prehaud, C.; Mehouas, S.;
- [54] Castel, G.; Chteoui, M.; Caignard, G.; Prehaud, C.; Mehouas, S.; Real, E.; Jallet, C.; Jacob, Y.; Ruigrok, R. W.; Tordo, N. Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. J. Virol., 2009, 83, 10808-10820.
- [55] Gao, Y.; Lenard, J. Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II. *EMBO J.*, **1995**, *14*, 1240-1247.
- [56] Takacs, A. M.; Barik, S.; Das, T.; Banerjee, A. K. Phosphorylation of specific serine residues within the acidic domain of the phosphoprotein of vesicular stomatitis virus regulates transcription *in vitro. J. Virol.*, **1992**, *66*, 5842-5848.
- [57] Jacob, Y.; Real, E.; Tordo, N. Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. J. Virol., 2001, 75, 9613-9622.
- [58] Gigant, B.; Iseni, F.; Gaudin, Y.; Knossow, M.; Blondel, D. Neither phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is required for its oligomerization. J. Gen. Virol., 2000, 81, 1757-1761.
- [59] Assenberg, R.; Delmas, O.; Ren, J.; Vidalain, P. O.; Verma, A.; Larrous, F.; Graham, S. C.; Tangy, F.; Grimes, J. M.; Bourhy, H. The Structure of the N-RNA Binding Domain of the Mokola virus Phosphoprotein. J. Virol., 2009, 84, 1089-1096.
- [60] Lynch, S.; Kolakofsky, D. Ends of the RNA within Sendai virus defective interfering nucleocapsids are not free. J. Virol., 1978, 28, 584-589.
- [61] Ribeiro, E. D.; Jr., Leyrat, C.; Gerard, F. C.; Albertini, A. A.; Falk, C.; Ruigrok, R. W.; Jamin, M. Binding of Rabies Virus Polymerase Cofactor to Recombinant Circular Nucleoprotein-RNA Complexes. J. Mol. Biol., 2009, 394, 558-575.

- [62] Green, T. J.; Luo, M. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc. Natl. Acad. Sci. U S A*, 2009, 106, 11721-11726.
- [63] Albertini, A. A.; Schoehn, G.; Weissenhorn, W.; Ruigrok, R. W. Structural aspects of rabies virus replication. *Cell Mol Life Sci*, 2008, 65, 282-294.
- [64] Kingston, R. L.; Hamel, D. J.; Gay, L. S.; Dahlquist, F. W.; Matthews, B. W. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc. Natl. Acad. Sci. U S A*, 2004, 101, 8301-8306.
- [65] Jensen, M. R.; Houben, K.; Lescop, E.; Blanchard, L.; Ruigrok, R. W.; Blackledge, M. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. J. Am. Chem. Soc., 2008, 130, 8055-8061.
- [66] Longhi, S.; Receveur-Brechot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S.; Canard, B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J. Biol. Chem., 2003, 278, 18638-18648.
- [67] Gao, Y.; Lenard, J. Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly. J. Virol., 1995, 69, 7718-7723.
- [68] Gupta, A. K.; Blondel, D.; Choudhary, S.; Banerjee, A. K. The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C. J. Virol., 2000, 74, 91-98.
- [69] Wright, P. E.; Dyson, H. J. Linking folding and binding. Curr. Opin. Struct. Biol., 2009, 19, 31-38.
- [70] Chenik, M.; Schnell, M.; Conzelmann, K. K.; Blondel, D. Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. J. Virol., 1998, 72, 1925-1930.
- [71] Chenik, M.; Chebli, K.; Gaudin, Y.; Blondel, D. *In vivo* interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein. *J. Gen. Virol.*, **1994**, 75 2889-2896.
- [72] Masters, P. S.; Banerjee, A. K. Resolution of multiple complexes of phosphoprotein NS with nucleocapsid protein N of vesicular stomatitis virus. J. Virol., 1988, 62, 2651-2657.
- [73] Takacs, A. M.; Das, T.; Banerjee, A. K. 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*, **1993**, *90*, 10375-10379.
- [74] Raux, H.; Flamand, A.; Blondel, D. Interaction of the rabies virus P protein with the LC8 dynein light chain. J. Virol., 2000, 74, 10212-10216.
- [75] Jacob, Y.; Badrane, H.; Ceccaldi, P. E.; Tordo, N. Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein. J. Virol., 2000, 74, 10217-10222.
- [76] Das, S. C.; Pattnaik, A. K. 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.*, 2005, 79, 8101-8112.
- [77] Thomas, D.; Newcomb, W. W.; Brown, J. C.; Wall, J. S.; Hainfeld, J. F.; Trus, B. L.; Steven, A. C. Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. J. Virol., **1985**, 54, 598-607.
- [78] Curran, J.; Kolakofsky, D. Replication of paramyxoviruses. Adv. Virus Res., 1999, 54, 403-422.
- [79] Kingston, R. L.; Baase, W. A.; Gay, L. S. Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins. J. Virol., 2004, 78, 8630-8640.
- [80] Bernard, C.; Gely, S.; Bourhis, J. M.; Morelli, X.; Longhi, S.; Darbon, H. Interaction between the C-terminal domains of N and P proteins of measles virus investigated by NMR. *FEBS Lett.*, 2009, 583, 1084-1089.
- [81] Houben, K.; Blanchard, L.; Blackledge, M.; and Marion, D. Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase co-factor P. *Biophys. J.*, 2007, 93, 2930-2940.
- [82] Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R. W.; Blanchard, L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. J. Virol., 2007, 81, 6807-6816.

Structural Disorder in Proteins of the Rhabdoviridae

[83] Bourhis, J. M.; Receveur-Brechot, V.; Oglesbee, M.; Zhang, X.; Buccellato, M.; Darbon, H.; Canard, B.; Finet, S.; Longhi, S. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phos-

 Received: ??????????
 Revised: ??????????
 Accepted: ???????????

Protein & Peptide Letters, 2010, Vol. 17, No. 8 987

phoprotein via two distinct sites and remains predominantly unfolded. *Protein Sci.*, 2005, 14, 1975-1992.

Structure, interactions with host cell and functions of rhabdovirus phosphoprotein

Cédric Leyrat, Euripedes A Ribeiro Jr, Francine CA Gérard, Ivan Ivanov, Rob WH Ruigrok & Marc Jamin[†]

*Author for correspondence: UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France = jamin@embl.fr

Rabies is an incurable albeit preventable disease that remains an important human health issue, with the number of deaths exceeding 50,000 people each year. Its causative agent, the rabies virus, is a negative-sense RNA virus, the genome of which encodes five proteins. Three of these proteins, the nucleoprotein, the phosphoprotein (P) and the large protein, are required to synthesize viral RNA in an efficient and regulated manner. P plays multiple roles during the transcription and replication of the RNA genome. It acts as a noncatalytic cofactor of the large protein polymerase and it chaperones nucleoprotein. Recent structural characterizations of rabies virus P revealed that P forms elongated and flexible dimers and uncovered the structural basis of its modular organization, revealing the existence of two independent structured domains and two long intrinsically disordered regions. In addition, recent studies also revealed that P interacts with nucleocytoplasmic trafficking carriers and with the host cell cytoskeleton, probably allowing viral components to be transported within the host cell and blocking the innate immune response by inhibiting different steps of the interferon pathway. With multiple binding sites for different viral and cellular partners located in either its structured or disordered regions, P appears to be a flexible 'hub' protein that connects viral or cellular proteins and allows their assembly into multimolecular complexes. These new findings shed light on the mechanism of replication of the virus and on the intimate interactions between the virus and its host cell, and will also help to identify new targets for the development of antiviral treatments.

Rabies & the family *Rhabdoviridae* Rabies

Rabies is a multihost zoonotic disease. Its transmission by dog bite and the fatality of the disease have been known since the 23rd century BC [1]. The main causative agent of rabies is the rabies virus (RV; genotype 1), the prototypic member of the genus Lyssavirus, although members of several other genotypes can also elicit the disease. RV is a neurotropic virus that causes an encephalitis with a complex pathogenesis leading to different forms of the disease ('furious' or encephalitic, 'dumb' or paralytic, nonclassic rabies) [2,3]. The encephalitis is incurable and fatal in 100% of the cases but the infection can be prevented. Since the first successful use of rabies vaccine by Louis Pasteur in 1885, preexposure and post-exposure treatments have been available [4] and the transmission of the disease to humans is under control in Western Europe and North America, mainly because of vaccination campaigns targeting dogs and wild animals [5]. However, rabies remains an important, albeit neglected, human health problem, killing more than 50,000 people each year worldwide (one person every 10 min on average), mainly in Asia and Africa [201] and imposing an important economic burden on affected populations [6]. Moreover, recent reports suggest that the number of cases is substantially underestimated in poor rural area [7]. The main natural reservoirs of RV are in the orders *Carnivora* (dog, fox, raccoon and skunk) and *Chiroptera* (bats). The mechanism of cross-species transmission remains poorly understood [8], but new strains, potentially not susceptible to vaccination against classical rabies [9,10] could emerge and spread again in human populations [11,12].

Rhabdoviridae

Rabies virus is a negative RNA virus that belongs to the family *Rhabdoviridae*. This large family of RNA viruses includes numerous pathogens of plants and animals, which vary widely in terms of cellular interactions and replicate either in the cytoplasm or in the nucleus of the host cell. The vesicular stomatitis virus (VSV) in the genus *Vesiculovirus* is the other prominent member of the *Rhabdoviridae* family. It is an arthropodborne virus that causes a mild disease in horses,

Keywords

intrinsically disordered
 region = phosphoprotein
 rabies = replication = virus



cattle and pigs and rarely transmits to humans. However, it has an important economical impact in the farm industry in the Americas where the virus is endemic. For many years, VSV served as a model system for studying the replication cycle of the rhabdoviruses and, more generally, the interactions between viruses and their host cells. More recently, VSV has been used as a vaccine vector and as an oncolytic virus [13,14]. Chandipura virus is another vesiculovirus, which is widespread in India, infects humans [15] and caused recent outbreaks of acute encephalitis with a high fatality ratio [16]. Finally, numerous rhabdoviruses infect fish or plants and cause economic damage in industrial fisheries and agricultural farms [17,18]. The family Rhabdoviridae is grouped in the order Mononegavirales (MNV) with the Filoviridae (Ebola and Marburg viruses), the Paramyxoviridae (measles, mumps and respiratory syncytial viruses) and the Bornaviridae (Borna disease virus) because all these viruses share similar genome and structural organizations as well as similar strategies of RNA replication and transcription [19].

The rabies virus & its life cycle RNA genome

The viral genome of RV is made of a single molecule of negative-sense RNA of approximately 12 kb, flanked on both sides by untranslated



Figure 1. The rabies virus. (A) Rabies virus genome. The genome encodes five proteins and contains untranslated 3' le and 5' tr sequences. **(B)** Structure of the virion. Bullet-shaped rabies virus showing the dimensions of the particle and the relative organization of the G, M, nucleoprotein–RNA complex, P and L. The L and P proteins are localized inside the N–RNA complex but are shown here outside for presentation purposes.

G: Glycoprotein; L: Large subunit of the RNA-dependent RNA polymerase; le: Leader; M: Matrix protein; P: Phosphoprotein; tr: Trailer. leader and trailer RNA regions (FIGURE 1A). It is comprised of only five genes that are common to all members of the MNV and that encode successively from the 3' terminus: the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the large subunit of the RNA-dependent RNA polymerase (L) (FIGURE 1A).

Structure of the viral particle

Rabies virus is a bullet-shaped particle 180 nm in length and 90 nm in diameter (FIGURE 1B) [20]. Cryoelectron microscopy (EM) reconstruction of the VSV particle revealed the molecular organization of the viral proteins in three concentric layers [21]. The outer shell is a host cell-derived lipid membrane decorated with trimeric G spikes. The middle layer is formed by M proteins that are organized in a regular helix, onto which the trimeric G spikes attach. The inner layer is the nucleocapsid (NC), the actual infectious part of the virus, composed of the genomic RNA molecule associated with the N, P and L proteins. N has two domains forming a groove into which the RNA molecule binds through interactions with the sugar-phosphate backbone [20,22,23]. Every N protein binds to nine nucleotides and completely enwraps the genomic RNA, forming a long N-RNA complex organized in a regular helix [20]. P and L proteins attach to this N-RNA complex but cannot be visualized in the cryo-EM reconstruction because they are not organized in a highly symmetrical fashion.

Viral life cycle

The intricacies of the rabies life cycle have not yet been fully unravelled and recent findings have been reviewed [24]. FIGURE 2 summarizes the main events of the viral life cycle, which is divided into three main phases. Numerous features of the cellular replication process are common to RV and VSV, although RV replicates more slowly and is less cytopathic than VSV, indicating different regulatory mechanisms and different modes of interference with the host cell. The first phase includes the attachment of the viral particle onto the surface of the host cell, followed by its internalization into an endosome and its intra-cellular axonal transport towards the cell body. These steps depend on the G protein. Once the viral particle has reached the neuronal cell body, the NC is released into the cytoplasm through fusion of the viral membrane with the endosomal membrane. The fusion is induced by a pH-dependent conformational change of the G protein [25]. The recent



Figure 2. Rabies virus replication cycle. The replication of rabies virus occurs in three main phases. **(A)** The attachment onto the host cell, the internalization by endocytosis and the release of the infectious part of the virus in the cytoplasm. **(B)** The replication of the genome and the production of viral proteins. **(C)** The transport and assembly of new viral particles and their budding through the cell membrane.

ER: Endoplasmic reticulum; G: Glycoprotein; L: Large subunit of the RNA-dependent RNA polymerase; M: Matrix protein; N⁰: Nucleoprotein (the subscript zero indicates the absence of RNA); P: Phosphoprotein.

crystal structures of the G protein of VSV in preand post-fusion conformations have provided a glimpse of the conformational changes of G that underlie the fusion process [26,27]. In the second phase, the viral genome is transcribed and replicated. Both processes occur entirely in the cytoplasm, in inclusion bodies named Negri bodies that are typical of RV-infected neurons [28], but were also found recently in VSV-infected cells [29]. Only three of the viral proteins, N, P and L, are required to synthesize viral RNA in an efficient and regulated manner [30]. In the early stages of the replication cycle, transcription of the incoming viral genome takes place in a sequential process, producing a positive-stranded leader RNA and five capped and poly-adenylated mRNAs. The abundance of the mRNAs decreases with distance from the 3' end of the genome because the polymerase initiates transcription at a single entry site and transcribes the genes in order from the 3' to the 5' end. The viral proteins are then produced by the host cell translation machinery. In contrast with VSV infection, in which the expression of host cell genes is completely shut off, allowing the rapid and efficient production of viral proteins, RV infection maintains the expression of host cell genes and preserves the cellular functions required for the correct spread of the virus. In RV infection, the M protein binds to eIF3h and downregulates the canonical cap-dependent translation initiation process [31]. Since the cap structure of rhabdoviruses is different from the cellular cap structure [32], it is probable that the expression of RV proteins is less affected. Later, the viral polymerase switches from transcription to replication of the RNA genome and starts producing a full-length positive-sense antigenome that subsequently serves as a template for the synthesis of new viral RNA genomes. In VSV, two multimolecular complexes that contain the viral L, P and N proteins, but differ in their content in host cell proteins, have been postulated to separately catalyze transcription and replication [33]. The mRNAs are not encapsidated by N [34], but both antigenomes and newly synthesized genomes are encapsidated by N and therefore replication requires the continuous production of soluble N. The switch from transcription to replication could be regulated by the accumulation of soluble N proteins, but it also depends on the M protein [35] and the precise mechanism of this regulation remains unknown. In the third phase, the viral components migrate to the host cell plasma membrane where they assemble. New viral particles bud out, taking away a piece of the host cell membrane containing the G protein. The M protein interacts with both the N protein and host membranes [36] and can therefore serve as a scaffold for assembling the viral components at the inner face of the cellular membrane. The membrane must then deform and pinch off in a membrane fission step. The RV M protein contains late domains involved in recruiting cellular machineries involved in the formation of vesicles that bud away from the cytoplasm, such as the proteins involved in the vacuolar protein-sorting process. In particular, binding of M to Nedd4, an E3 ubiquitin ligase that recruits proteins in the vacuolar protein sorting pathway, is required for efficient budding of RV [37,38].

Structure of the P & its post-translational modifications

Phosphoprotein is a multifunctional protein that plays multiple roles in viral transcription and replication processes. First, P plays an essential role in the encapsidation of both genomic and antigenomic RNA molecules. It chaperones nascent N molecules by forming a complex termed N⁰–P (the superscript zero indicates the absence of RNA). Secondly, P associates with L to form an active RNA-dependent RNA polymerase (L-P complex), which uses the N-RNA complex rather than the naked RNA as a functional template. Host cell cofactors are also found in association with the VSV polymerase complex, but their roles and mechanisms of action have not vet been determined [33]. In addition, P interferes with the host cell intracellular trafficking systems and RV P acts at several stages of the interferon pathway, inhibiting the cellular innate immune response. Recent studies of both RV and VSV P have shed new light on the structural organization of these proteins and on the way they achieve their multiple functions. In this article, data on VSV P are included since this protein shares functional and structural features with RV P and might provide insights into RV P.

Truncated forms of P

In infected cells, P is heterogeneous and present in several truncated forms and in different phosphorylation states (Figure 3). In cells infected with RV of the CVS strain as well as in the virion, four N-terminally truncated forms of RV P (P₂: aa 20–297, P₃: aa 53–297, P₄: aa 69–297 and P₅: aa 83-297) are produced in addition to full-length P, from internal in-frame AUG initiation codons by a ribosomal leaky scanning mechanism (FIGURE 3) [39]. P_2 and P_3 are present in infected cells and in the virion in slightly lower amounts than full-length P, whereas P_4 and P_5 are significantly less populated [39]. These forms of P have different subcellular localizations resulting from the interplay between nuclear import and export processes mediated by a C-terminal nuclear localization signal (NLS) and nuclear export signals (NESs) in the N- and C-terminal parts of P [40-42]. Full-length P and P, are found in the cytoplasm, whereas P_3 , P_4 and P_5 are found in the nucleus. The exact biological role of the different forms of RV P has yet to be determined, and the Met residues exhibit a limited conservation among lyssaviruses [43]. In the cytoplasm of cells infected with the CVS strain of RV, fulllength P participates in the transcription and replication processes, whereas in the nucleus, P_{3} interferes with the interferon response pathway [41,44,45], but the functions of P_2 , P_4 and P_5 remain to be characterized. In VSV, as in other members of the MNV, truncated forms of P are also produced from the same mRNA by internal initiation of translation [46,47]. In the paramyxoviruses, a truncated form of P blocks the innate immune response [48,49].

Phosphorylation of P

Rabies virus P and VSV P are phosphorylated in the virion and in infected cells, and numerous studies, in particular for VSV, have been devoted to deciphering the role of these phosphorylations. In both RV and VSV, phosphorylation occurs in two regions of P, in the N-terminal acidic region and in the C-terminal domain (FIGURE 3). In both viruses, phosphorylation is mediated by two different kinases and occurs sequentially, resulting in the existence of two main forms of the protein [50,51]. Phosphorylation of VSV P regulates transcription and replication, although the precise mechanism of these regulations is unknown. Phosphorylation in the N-terminal acidic region by casein kinase II is required for transcription [52]. The subsequent phosphorylations in the C-terminal domain by a kinase associated with the L protein regulate the binding of the L-P complex to its N-RNA template and are thus required for RNA synthesis [53]. It has been proposed that phosphorylation in the



Figure 3. Functional and structural domains of rabies virus protein. The dark orange and light orange boxes represent the localization along the amino acid sequence of the known 3D structures and of the different functional regions. Numbers indicate the defined boundaries. The three blue areas correspond to the localization of structured domains predicted from the amino acid sequence by a recent meta-prediction [70]. NESs are depicted by the light orange boxes and NLSs by the dark orange boxes. Pi sites indicate the localization of phosphorylation sites.

L: Large protein; LC8: Light chain 8 of the dynein complex; N: Nucleoprotein; NES: Nuclear export signal; NLS: Nuclear localization signal; P: Phosphoprotein; PML: Pro-myelocytic leukemia protein; STAT: Signal transducer and activator of transcription 1.

N-terminal region induces the self-assembly of P and thereby regulates the association with L and the activation of transcription [54-57]. However, a double mutation (S60D/T62D) that mimicks phosphorylation and renders P constitutively active in transcription [57] has no effect on the oligomerization of P [58]. In RV, protein kinase C phosphorylates residues in the C-terminal part of P (S162, S210 and S271), whereas a RV-specific kinase phosphorylates residues in the N-terminal part (S63 and S64 of CVS) (FIGURE 3) [50]. The phosphorylation sites in the N-terminal region of RV P are not conserved among lyssaviruses and phosphorylation of these residues has no effect on the oligomerization of the protein and no known function [59], whereas phosphorylation in the P C-terminal domain (P_{CTD}), in particular of S210, appears to modulate the nucleocytoplasmic distribution of P and plays a role in blocking the innate immune response (see later) [42].

Functional modules

The existence of independent functional modules in P was first revealed by deletions and mutations in VSV P. Early studies identified three distinct functional regions directly involved in transcription and showing a high conservation rate among VSV strains [60-62]. The first N-terminal region (aa 1–137) is highly acidic and contains binding sites for L [62] and N⁰ [63]. It also contains three phosphorylation sites involved in transcriptional regulation [51,64]. The second and third functional regions form the C-terminal part of P (aa 211-240 and aa 245-265). The second region binds to L and contains phosphorylation sites that control replication [53], while the third region (aa 245–265) is basic and contains the binding site for the N-RNA template [65,66]. The remaining region (aa 138-210), named the 'hinge' region, is hypervariable and its distal part (aa 201-220) plays a role in the assembly of infectious particles [61]. A central region of VSV P, at the junction of the acidic N-terminal region and of the 'hinge' region, contains a self-association domain [67]. In RV P, similar functions were mapped to corresponding N-terminal and C-terminal regions [20,59,68–71] and a self-association domain was mapped in the central part of the molecule (FIGURE 3) [69,72,73], suggesting a common modular organization. Comparison with the P of Sendai virus and measles virus of the *Paramyxoviridae* and amino acid sequence analysis suggested that RV P contains structured domains separated by intrinsically disordered regions (IDRs) [69,74].

Intrinsically disordered regions

Since the early days of protein-folding studies, biological activities have been associated with the adoption by the polypeptide chain of a welldefined 3D structure. However, this structurefunction paradigm was recently challenged with the discovery that many proteins are intrinsically disordered or contain IDRs under physiological conditions [75]. Various algorithms were developed for identifying and locating IDRs in proteins [76] and complete-genome surveys revealed that as many as a third of the proteins from eukaryotes contain IDRs [77]. Even more surprisingly, disorder appears to be required for the biological function of these proteins [78]. IDRs are found in numerous proteins involved in molecular recognition, cell signaling and regulation processes [79] and many of them serve as a flexible scaffold for the assembly of multimolecular complexes [80,81]. IDRs exhibit different modes of binding to their partners; some fold upon binding, whereas others remain partially disordered and thereby can have a wide range of effects on the energy of the interactions.

Prediction of disordered regions

A recent meta-prediction of IDRs in RV P hinted at a modular structural organization of RV P with three folded domains, an N-terminal domain (P_{NTD}) , a central domain (P_{CED}) and a $\mathrm{P}_{\mathrm{CTD}}$, separated by two IDRs (IDR_{\mathrm{NTD}} and IDR_{CTD}) (FIGURE 4A & C) [69]. A multiple alignment of the amino acid sequences of Ps from 13 members of the genus Lyssavirus revealed that the three predicted structured domains were more conserved than the predicted IDRs (FIGURE 4B) [69], in accordance with a previous phylogenetic analysis [43]. The precise localization of the boundaries between disordered regions and folded domains was confirmed by dissecting the protein and characterizing its fragments [69]. The predicted P_{CTD} had been previously identified as an autonomous folding domain and its structure solved by x-ray crystallography (FIGURE 4D) [68]. The predicted P_{CED} also appeared as an autonomous folding domain and its structure was recently solved (FIGURE 4D) [73]. The N-terminal part of RV P is acidic, as in the VSV P, and the predicted $P_{_{\rm NTD}}$ appeared globally disordered [69]. A similar analysis of VSV P also predicted the presence of folded domains in the N- and C-terminal parts that closely correspond to some previously identified functional regions. The predicted P_{NTD} corresponds to the N⁰-binding site and, in isolation, it is globally disordered but contains two transiently populated α -helical elements [82]. The identified C-terminal domain (aa 194–265), whose structure could be solved by NMR spectroscopy [70], encompasses the second and third functional regions as well as the distal part of the hinge region required for encapsidation of the RNA. Predictions in the central part of VSV P are not as clear and are discussed later.

Self-assembly of P

Rabies virus P self-assembles both in vivo and in vitro, and the oligomerization domain perfectly matches with the predicted P_{CED} (FIGURES 3 & 4) [59,69,72]. Nonphosphorylated RV P as well as isolated $\mathbf{P}_{_{\rm CED}}$ produced in bacteria form dimers [58], whereas the protein deleted from the central region (P_{91-131}) forms monomers [69]. In the crystal structure, P_{CED} also forms dimers (FIGURE 4D). Similarly, unphosphorylated VSV P forms dimers as well [58], and an autonomous folding domain isolated by limited proteolysis $(VSV P_{CED})$, aa 107–170) is dimeric in the crystal structure [67,83]. However, the meta-prediction of IDRs in VSV P failed to identify the crystallized domain and instead predicted a bipartite folded domain in the center of P (aa 98-105 and aa 136–186) [69].

A comparison of the dimeric structures of RV and VSV P_{CED} showed that they are so different from each other that it is impossible to determine whether these domains have evolved from a common ancestor [73]. The monomer of VSV P_{CED} is made of a central α -helix flanked by two two-stranded β -sheets. In the dimer, the helices from the two monomers pack against each other in a parallel orientation, and on each side of the helical bundle, the N-terminal β -sheet of one monomer forms a four-stranded β -sheet with the C-terminal β -sheet of the other monomer [67]. The monomer of RV P_{CED} is made of two α -helices forming a hairpin and the dimer is assembled by packing two monomers in a parallel orientation at an angle of approximately 50° [73].

By comparison with the mechanism of transcription/replication of Sendai virus (*Paramyxoviridae*), it was speculated that oligomerization of P allows the protein to move along the N–RNA template by a cartwheeling mechanism [84]. In Sendai virus, P forms tetramers through a central coiled-coil domain that is required for transcription. In RV, the self-assembly domain of P is dispensable for transcription [72], but a recent study revealed that dimerization is necessary for the association of P₃ with the microtubules and its nuclear import [85,86].

Overall structure of RV P dimers

The P dimer is elongated, with overall dimensions much larger than those expected for a globular protein of the same molecular mass [58]. P has a dimeric central core from which four flexible 'arms' emerge (FIGURE 5). So far, no interaction involving the central domain has been described, but both the N- and C-terminal flexible arms contain multiple binding sites. The N-terminal arm $(P_{_{\rm NTD}}$ + $IDR_{_{\rm NTD}})$ contains no stable structured regions and behaves as a random coil chain. The C-terminal arm contains a terminal folded domain $(\mathrm{P}_{\rm CTD})$ tethered to the P_{CED} by a long flexible linker (IDR_{CTD}). As in the N-terminal arm, the C-terminal flexible linker (IDR_{CTD}) contains a binding site for a cellular partner (see later). Conversely, the C-terminal domain is a well-structured domain, involved in the interaction with the N-RNA template, but it also contains binding sites for various cellular proteins (importin, exportin, kinase, STATs, promyelocytic leukemia [PML] and microtubules). The overlap of binding sites in both N- and C-terminal arms certainly modulates the capacity of RV to replicate and might be part of the strategy developed by the virus for limiting the impact of its replication on the cellular metabolism.

Interactions of P with its host cell Nucleo-cytoplasmic trafficking

During the life cycle of viruses, viral components need to be transported in the cytoplasm or in and out of the nucleus. Numerous viruses hijack the cellular motors and the cytoskeleton tracks for intracellular transport as well as the cellular nucleo-cytoplasmic shuttles for proceeding through the nuclear pore complex [87]. Although RV replicates in the cytoplasm, P shuttles between the cytoplasm and the



Figure 4. Modular organization of rabies virus phosphoproteins.

(A) Meta-prediction of intrinsically disordered regions (IDRs). A consensus prediction for the localization of IDRs was obtained by combining results from 16 independent predictors available through web servers [70]. For each disorder prediction, residues were simply ranked in a binary manner as ordered or disordered using the default threshold set for each algorithm. For each predicting algorithm, a score of 0 was attributed to a residue when it was predicted to be disordered, whereas a score of 1 was attributed when it was predicted to be in a structured region. The D-score was calculated by adding the values for each residue and dividing by the number of algorithms used. We arbitrarily defined a threshold at 0.50, above which residues were assigned as structured and below which residues were assigned as disordered. (B) Sequence conservation. The amino acid sequences of the P protein from 13 lyssavirus variants were aligned and the conservation rate for each amino acid was scored by using the AL2CO software [137]. A C-score above 0.5 indicates a conserved position. This analysis revealed the presence in the sequence of three conserved regions (aa 1-50, 101-126 and 194-292) that are located within or correspond to the otherwise predicted structured regions. (C) Consensus localization of structured and disordered regions. The consensus disordered regions determined from the D-score suggest the presence of two disordered regions (aa 53-89 and 131-194) and three structured domains (aa: 1-52, 90-130 and 195-297). (D) Structured domains: the high-resolution structures of two domains corresponding closely to the predicted P_{CED} and P_{CTD} have been determined by x-ray crystallography [69,74]. CED: Central domain; CTD: C-terminal domain; IDR: Intrinsically disordered region; NTD: N-terminal domain; P: Phosphoprotein.

Review Leyrat, Ribeiro, Gérard, Ivanov, Ruigrok & Jamin





L: Large subunit of the RNA-dependent RNA polymerase; N^o: Nucleoprotein (the subscript zero indicates the absence of RNA).

nucleus through interactions with importins and exportins. The protein contains a bipartite NLS [88], consisting of a conserved basic motif (K₂₁₁KYK₂₁₄), and Arg₂₆₀, which form a patch on the surface of the C-terminal domain (FIGURE 3) [40]. The protein also contains two conserved chromosome region maintenance protein 1 (CRM1)-dependent NESs [89], one in the N-terminal region (aa 49-58) and the other in P_{CTD} (aa 224–234) (FIGURE 3) [40,42]. The N-terminal NES is localized in the disordered N-terminal part of P, whereas the C-terminal NES is close to the import sequence and to a phosphorylation site (S210). The residues forming the second NES are buried inside the P_{CTD} structure, and the activation of this NES is controlled by the phosphorylation of the adjacent residue [42].

Interactions with the cytoskeleton

Various studies have revealed that RV and VSV Ps interact with the microtubules, but the precise role of these interactions remains poorly understood. RV P binds to the light chain 8 (LC8) of the host cell dynein complex [90,91]. Dynein is a multimolecular complex that transports various cellular cargos towards the minus-end of cytoskeletal microtubules, usually oriented towards the cell center. This interaction thus suggests that RV uses the host cell machinery for intracellular transport [90,91] and could possibly explain the retrograde axonal transport of RV. However, a recent study showed that the whole virus, rather than the NC, is transported to the cell body, a process that cannot therefore depend on P, which is located inside the viral particle [92]. Moreover, deletion of the LC8-binding region (aa 138–149) of P did not attenuate the spread of the virus from a peripheral entry site to the CNS [93,94]. Dynein could also help the polymerase to move along the N-RNA template since deletion of the LC8-binding region (aa 138-149) reduces the production of viral RNA in the CNS [93,94]. However, mutation of two residues in the LC8binding site P abolishes the interaction with LC8, but has no effect on transcription [95]. More recently, it was shown that the association of P₃ with microtubules participates in its NLSdependent nuclear import [42]. In VSV, no binding site for LC8 has been found, but the P also interacts with microtubules and participates in the transport of NCs towards the cell periphery via a microtubule-mediated process [96].

P is a multifunctional antagonist of the host cell interferon response

The innate immune response represents the first line of defense of an organism against viral infection. The recognition of viral components by Toll-like receptors or retinoic acid-inducible-like receptors triggers the production of interferons. Interferons activate signaling cascades and induce the expression of various interferonstimulated genes with antiviral properties. Many viruses, including RV, have developed strategies for evading these cellular antiviral responses, in particular, the production of type I interferon [97–99]. RV induces an innate immune response [100,101] and is sensitive to interferon treatment [102,103]. P plays a major role in modulating this

response [104,105]. An engineered RV strain producing insufficient amounts of P to inhibit the host cell response could replicate in interferondeficient cells but not in interferon-competent cells [102,106].

Phosphoprotein acts at different stages of the interferon pathway. First, P interferes with the production of interferon by acting on the IRF-3 [102] or on signal transducer and activator of transcription 1 and 2 (STAT1 and STAT2) [46,106]. The interaction with STAT involves the C-terminal domain of P (FIGURE 3), occurs both in the cytoplasm and in the nucleus [42,86] and reduces the pathogenicity of RV, probably by allowing a more efficient interferon response [45]. Second, P interacts through its C-terminal domain with an interferon-stimulated genes product, the PML protein (also known as the tripartite motif protein 19 [TRIM19]) (FIGURE 3). The PML protein is part of the nuclear multiprotein complexes, named nuclear bodies, that are involved in host cell defense mechanisms [107], and although the mechanism of antiviral action of these cellular components remains unknown, nuclear bodies are involved in the interferon-stimulated host cell response against RV infection [108,109].

Roles of P in the transcription/ replication complexes P is a chaperone of nascent nucleoprotein The N⁰-P complex

The RNA genome of rhabdoviruses is protected by the N and, thus, during viral replication, the newly synthesized RNA genomes and antigenomes must be encapsidated by N. N has a strong affinity for RNA and binds nonspecifically to cellular RNA. The first role of P is to act as a chaperone of nascent N molecules. P forms with the N, resulting in a complex named the N⁰-P complex [72,110], with a stoichiometry of one N plus two Ps [111], that maintains nascent N molecules in a soluble and RNA-free form and that is competent for encapsidation of the newly synthesized RNA genomes [112,113].

Binding site for N^o

Limited proteolysis showed that the N-terminal 40 residues of RV P are necessary and sufficient to maintain the integrity of the soluble N^0 –P complex [72]. Similarly, in VSV, the binding site for N^0 is localized in the N-terminal region (aa 11–30) [110]. The binding site for N^0 is thus localized within the predicted N-terminal domain, named P_{NTD} . The RNA-binding cavity

of N has a highly positive electrostatic surface potential [23,114], whereas $P_{\rm NTD}$ is rich in negatively charged residues, suggesting that $P_{\rm NTD}$ binds in the same cavity of N as RNA [72,115].

Although our meta-prediction of IDRs suggested that the first 52 amino acids of RV P form a structured domain (FIGURE 4A) and that they are highly conserved among the lyssaviruses, a peptide corresponding to the first 60 residues is globally disordered [70]. Its Stokes's radius measured by size exclusion chromatography is that of a fully unfolded protein of the same molecular mass, and the narrow width of resonances in the NMR spectrum and the poor chemical shift dispersion imply a globally disordered molecule. Similar results were obtained with VSV P_{NTD}, but circular dichroism spectroscopy indicates the presence of low amounts of secondary structure and NMR relaxation measurements reveal the existence of two transient α -helical elements (aa 2-8 and aa 25-38 [83]; i.e., α -helices that are not stable enough to be present all the time in the protein, but that fold and unfold continuously). This local order in the N-terminal part of P could indicate that P_{NTD} contains a molecular recognition element, for example, a short motif within a disordered protein that promotes association with a partner by folding upon binding [116,117].

Importance of disorder for the activity of P

The N-terminal part of RV P (aa 1–92) not only contains the binding site for N⁰, but also contains binding sites for L, host cell kinases and importins. IDRs have the ability to interact with multiple structurally diverse partners and their binding energetics favor exchange between different partners. During the viral replication cycle, the Nº-P complex is not an end product. It forms only transiently, as an intermediate during the synthesis of new NCs. The Nº-bound P must be outcompeted by the newly synthesized genomic RNA and therefore the binding affinity of P for Nº cannot be too high. The mechanism of folding upon binding provides a means of specific recognition without the corollary of high affinity [82,118]. Indeed, the folding or the adoption of a rigid structure by the ligand when it binds to its receptor will generally lead to the formation of multiple specific intermolecular interactions that confer a great specificity and consequently a high affinity. If the ligand is disordered in its unbound form, this strong binding energy is opposed by the high entropy of the disordered protein. We might thus speculate that by being disordered in its free form, P_{NTD} recognizes N⁰ with high

specificity but moderate affinity, thus allowing its displacement, not only by the newly synthesized viral RNA, but also by host cell proteins.

P is a cofactor of the L polymerase P links L to the N-RNA template

Phosphoprotein associates with the L protein to form an active RNA-dependent RNA polymerase complex [119]. In this complex, L carries out all enzymatic activities involved in the transcription and replication of the viral genome, including RNA synthesis, mRNA cap synthesis and methylation, and mRNA polyadenylation, while P is an essential noncatalytic cofactor [31]. In the absence of P, L initiates RNA synthesis and produces short oligonucleotides of two to four ribonucleotides, but P is required for RNA elongation [120]. L is a processive enzyme, which must remain attached as it moves along its template [121], but L alone has a weak binding affinity for the N-RNA complex and is unable to re-associate with its template [122]. P has binding sites for both L and the N-RNA complex and is presumed to correctly position L onto its N-RNA template, maintaining the attachment of L during its displacements along the template.

Binding site for L

The binding site for the L protein is located in the N-terminal part of P, although its precise localization is unclear [73,123,124]. The binding site was initially localized in the 60 first amino acids of P, which are required for transcription, whereas residues 61-173 were thought to be dispensable [73]. Later, the binding site was suggested to be contained in the 19 N-terminal amino acids [124], but a recent study showed that the 100 N-terminal amino acids are required for a strong interaction with L and that the region 40-70 plays a major role [123]. In VSV, a binding site for L was also localized in the distal part of the N-terminal IDR (aa 79–123) [63]. The requirement of a globally disordered acidic region for the correct assembly of the transcription/replication complex is reminiscent of eukaryotic transcription-regulator proteins, such as GAL4 and GCN4. These proteins contain disordered acidic regions that activate transcription by recruiting transcription factors to the promoter and helping in the assembly of the transcription complex [125]. The acidic character of these regions, and not their specific amino acid sequence, is the major determinant of their activity, and transcriptional activation can be achieved with random polypeptides of identical

composition [126]. In VSV, for transcription, the N-terminal part of P can be replaced by β -tubulin, a protein with an acidic C-terminal region [127]. It is also known that β -tubulin binds to the L [128]. The N-terminal disordered region of P (P_{NTD} + IDR_{NTD}) likely recruits the L to the site of RNA synthesis and allows the assembly of the multimolecular transcription and replication complexes.

Binding site for the N-RNA template

The binding site for the N-RNA template was mapped to the C-terminal region of P [73,129,130]. This region is critical for RNA synthesis and corresponds closely to the last predicted structured region in RV P (FIGURE 4D). The crystal structures of P_{CTD} from RV and the related Mokola virus (Lyssavirus, genotype 3) revealed that this domain (aa 186-297 in RV) has the shape of a half pear, the pear being cut lengthwise [69,131]. Several residues involved in binding the N-RNA template, previously identified by a mutational analysis in Mokola virus, form patches on the opposite faces of the half pear [73]. Conserved hydrophobic residues line a hydrophobic cavity on the rounded face, named the W-hole, and four basic residues form a patch on the flat face. The corresponding P_{CTD} was correctly identified in VSV P from a meta-prediction analysis (aa 195-267) and its structure was solved by NMR spectroscopy [71]. A longer form of this domain (aa 185-267) confirmed that $\boldsymbol{P}_{\text{CTD}}$ is preceded by a flexible linker. Although multiple sequence alignments indicate almost no sequence similarity between Ps from VSV and RV, the topology of the peptide chains is conserved and the secondary structure elements could be superimposed [71]. A comparison of these structures with those of Mokola virus [131] and of paramyxoviruses suggests that the structure of the C-terminal N-RNA-binding domains is conserved among these virus families [132].

Structure of the complex between P & the N-RNA template

To understand how P attaches to the N–RNA template, we built a model of the complex formed by a recombinant circular N–RNA complex and $P_{\rm CTD}$ by computational modeling using SAXS and biochemical data as constraints [133]. In this model, $P_{\rm CTD}$ lies on the top of the C-terminal domain of one N protomer, while by an induced-fit mechanism, otherwise flexible loops from two adjacent N protomers [23] mold around $P_{\rm CTD}$ and form extensive intermolecular

contacts with the W-hole and the conserved basic patch, respectively. The involvement of two adjacent protomers indicates that the complete binding site for $P_{\rm CTD}$ only exists in multimeric N–RNA complexes. The binding of $P_{\rm CTD}$ to the N–RNA complex is noncooperative, with a dissociation constant near 100 nM [133]. The homologous complex of VSV solved by x-ray crystallography showed a similar binding mode for $P_{\rm CTD}$ [134]. Again, despite the absence of sequence similarity, a hydrophobic cavity on one side of $P_{\rm CTD}$ accommodates a hydrophobic side chain of the flexible loop of N, while basic residues on the opposite side of $P_{\rm CTD}$ interact with acidic residues of the corresponding flexible

loop of the adjacent N subunit, suggesting a conservation of the binding modes. The structures of both complexes revealed no conformational change in N that would significantly modify the accessibility to the RNA.

Role of P in RNA synthesis

In RV, phosphorylation at S389 of N enhances the binding affinity for P_{CTD} [135]. Our model suggests that phosphorylation of this serine residue creates an additional network of stabilizing electrostatic interactions [133] and recent data show that phosphorylation of N achieves a 1000-fold increase in binding affinity [RIBEIRO E & JAMIN M, UNPUBLISHED DATA].



Figure 6. Mechanisms of transcription and replication of rabies virus. During RNA synthesis, the polymerase must move along its RNA template and must displace the RNA from the nucleoproteins. P provides the physical link by which the polymerase attaches to its template. In rabies virus (RV), the dimerization domain of P is dispensable for transcription and cannot therefore move by a cartwheeling mechanism. Owing to the strong binding affinity of the C-terminal domain of P for the N–RNA template, we suppose that P dimers are bound at regular intervals along the N–RNA complex. During transcription, the polymerase attaches to the N-terminal region of P and transcribes a stretch of RNA. Since P is elongated and flexible, the N-terminal arm of the adjacent P molecule can extend and catch the polymerase, which is then transferred downstream. The polymerase can thus move along the N–RNA complex by jumping between P molecules. During replication, the polymerase can move by a similar mechanism, but in addition, the newly synthesized genomic RNA is encapsidated by nascent N molecules. Nascent N molecules form soluble N⁰–P complexes, which can also attach to the N–RNA complex through their P_{CTD}. N⁰ can then be delivered to the site of RNA synthesis, yielding new N–RNA complexes.

N: Nucleoprotein; NC: Nucleocapsid; P: Phosphoprotein.

In the transcription/replication complexes, the high flexibility of P and the high binding affinity of P_{CTD} for the N–RNA template suggests an alternative to the cartwheeling mechanism for the movement of the polymerase along its N-RNA template, in which dimerization of P is dispensable (FIGURE 6). Highly flexible P molecules tethered to the N-RNA template through their P_{CTD} will sample large conformational spaces, acting like brushes that repel each other. The estimated number of P dimers in the VSV virion [136], would result in one dimer attached to every five N protomers. The small size of the structured domains of P and the high flexibility of the molecule precludes the localization of P dimers along the N-RNA template by EM. However, if P molecules are distributed at regular intervals along the template, displaying their Nº- and L-binding sites, the polymerase (~50 molecules packed in each VSV virion [136]) could move along the template by jumping from one P to the next. The long N-terminal arms of P could capture the polymerase upstream, keeping it attached to the template while it replicates a length of RNA, and then delivering it downstream to the next P molecule. The switch from transcription to replication of the viral genome could be triggered by the progressive accumulation of the Nº-P complexes, which could, above a critical concentration, displace the P dimers attached to N-RNA and thereby deliver the nascent Nº molecules to the site of RNA synthesis. The availability of Nº at the site of RNA synthesis would result in the encapsidation of the 5' extremity of the newly synthesized RNA molecule and thereafter allow the polymerase to read through the intergenic regions without stopping. This would thus lead to the synthesis of full-length antigenomes and genomes. The binding sites for Nº and for L on P are near each other or overlap and, currently, it is not known whether a single molecule of P can bind both partners concomitantly or whether separate molecules of P are required to attach both partners on the template.

Future perspective

Recent studies of RV P revealed the peculiar structural properties of this protein. The existence of two folded domains and two long IDRs confers the possibility of continuous extension and contraction of the polypeptide chain. The presence of binding sites for multiple viral and cellular partners, including in the IDRs, backs the role of P as an adapter in the formation of several multimolecular complexes. This new picture provides a first glimpse at the way P plays its role in the viral transcription/replication machineries. However, numerous questions regarding the molecular mechanisms of P remain unanswered and it is now necessary to investigate more deeply into the mechanisms of action of P. This will require combining more structural and biophysical studies of isolated proteins and protein complexes with *in vitro* reconstitutions of functional machineries and *in vivo* observation of these machineries in their physiological environment.

The structure of the complex between P_{CTD} and a circular N-RNA complex shows how P attaches to the transcription/replication template. We must now determine if P moves along the N-RNA template, serving as a carrier of L or if it remains static and forms a chain where L, is transmitted from one to the other. In both cases, the L-P polymerase moves along the template and we must determine what drives P or L. Resolving the structure of the N-RNA-P_{CTD} complex showed that binding of $P_{_{\rm CTD}}$ to the N-RNA template does not modify the accessibility of the RNA. We now need to understand how the polymerase obtains access to the RNA. We found that the N-terminal N⁰-binding region of P is globally disordered but possesses some residual structure that could fold upon binding to nascent N molecules. We need to find out how P binding to Nº prevents the binding of cellular RNAs and allows the specific encapsidation of viral RNAs. We must also understand how N is transferred from P to the viral RNA and how the new NCs assemble. In the host cell, transcription and replication of the viral genome occur in Negri bodies. We need to characterize the organization of the transcription/replication machineries within these cellular compartments and their interactions with cellular components, and understand how viral transcription and replication occur in the host cell. Recent studies disclosed important interactions between RV P and host cell components. We need to delve more deeply into the molecular mechanisms of

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

these interactions to understand how RV precisely diverts these cellular machineries. Finally, deciphering the molecular mechanisms by which P interacts with viral and cellular partners could also uncover new targets for the development of an antiviral treatment.

Executive summary

Rabies & the rabies virus

- Rabies has been feared by mankind for more than 20 centuries and it remains an important but neglected infectious disease. The resulting encephalitis remains incurable but is preventable by pre- and post-infection treatments.
- The main causative agent is the rabies virus (RV), a nonsegmented negative-sense RNA virus.

The rabies virus replication complex

- Transcription and replication of RV occurs in the cytoplasm.
- The RNA genome and three viral proteins, nucleoprotein (N), phosphoprotein (P) and large protein (L), constitute the infectious part of the virus and are required for efficient and regulated RNA synthesis.
- The N–RNA complex serves as a template for both processes.
- L carries out all enzymatic activities, including RNA synthesis, mRNA capping, methylation and poly-adenylation.
- P is an essential noncatalytic cofactor for both transcription and replication.

Phosphoprotein functions

- P acts as a chaperone of nascent N molecules by forming a N⁰-P complex that prevents N from binding nonspecifically to cellular RNA and maintains it in a soluble form until it is used to encapsidate newly synthesized RNA genomes or antigenomes.
- P attaches the polymerase complexes to the N–RNA template and is required for the elongation of RNA molecules.
- P shuttles between the cytoplasm and the nucleus, hijacking cellular importins and exportins.

7.

- P binds to the molecular motor dynein through direct interaction with the light chain 8.
- P blocks the host cell innate immune response by interacting at different stages of the interferon pathway.

Phosphoprotein structure

- P forms elongated dimers.
- P contains two folded domains, a central dimerization domain and a C-terminal N–RNA-binding domain and long intrinsically disordered regions.
- Binding sites for several viral and cellular partners are located both in the disordered and folded regions.
- Folding upon binding to a partner could confer a high specificity with a moderate binding affinity, which might be important for the dynamic functioning of the transcription/replication complexes.

Bibliography

- Papers of special note have been highlighted as: • of interest
- of considerable interest
- 1. Théodoridès J: *Histoire de la Rage*. Masson, Paris, France (1986).
- Dietzschold B, Li J, Faber M, Schnell M: Concepts in the pathogenesis of rabies. *Future Virol.* 3, 481–490 (2008).
- Hemachudha T, Laothamatas J, Rupprecht CE: Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges. *Lancet Neurol.* 1, 101–109 (2002).
- Rupprecht CE, Willoughby R, Slate D: Current and future trends in the prevention, treatment and control of rabies. *Expert Rev. Anti Infect. Ther.* 4, 1021–1038 (2006).
- Warrell MJ, Warrell DA: Rabies and other lyssavirus diseases. *Lancet* 363, 959–969 (2004).
- Knobel DL, Cleaveland S, Coleman PG et al.: Re-evaluating the burden of rabies in Africa and Asia. Bull. World Health Organ. 83, 360–368 (2005).

- Cleaveland S, Fevre EM, Kaare M, Coleman PG: Estimating human rabies mortality in the United Republic of Tanzania from dog bite injuries. *Bull. World Health Organ.* 80, 304–310 (2002).
- Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE: Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science* 329, 676–679 (2010).
- Hanlon CA, Kuzmin IV, Blanton JD, Weldon WC, Manangan JS, Rupprecht CE: Efficacy of rabies biologics against new lyssaviruses from Eurasia. *Virus Res.* 111, 44–54 (2005).
- Weyer J, Kuzmin IV, Rupprecht CE, Nel LH: Cross-protective and cross-reactive immune responses to recombinant vaccinia viruses expressing full-length lyssavirus glycoprotein genes. *Epidemiol. Infect.* 136, 670–678 (2008).
- Fooks AR: Rabies remains a 'neglected disease'. *Euro Surveill.* 10, 211–212 (2005).
- Leslie MJ, Messenger S, Rohde RE et al.: Bat-associated rabies virus in skunks. Emerg. Infect. Dis. 12, 1274–1277 (2006).

- Lichty BD, Power AT, Stojdl DF, Bell JC: Vesicular stomatitis virus: re-inventing the bullet. *Trends Mol. Med.* 10, 210–216 (2004).
- Rose NF, Publicover J, Chattopadhyay A, Rose JK: Hybrid alphavirus–rhabdovirus propagating replicon particles are versatile and potent vaccine vectors. *Proc. Natl Acad. Sci.* USA 105, 5839–5843 (2008).
- Basak S, Mondal A, Polley S, Mukhopadhyay S, Chattopadhyay D: Reviewing Chandipura: a vesiculovirus in human epidemics. *Biosci. Rep.* 27, 275–298 (2007).
- Rao BL, Basu A, Wairagkar NS *et al.*: A large outbreak of acute encephalitis with high fatality rate in children in Andhra Pradesh, India, in 2003, associated with Chandipura virus. *Lancet* 364, 869–874 (2004).
- Jackson AO, Dietzgen RG, Goodin MM, Bragg JN, Deng M: Biology of plant rhabdoviruses. *Annu. Rev. Phytopathol.* 43, 623–660 (2005).
- Hoffmann B, Beer M, Schutze H, Mettenleiter TC: Fish rhabdoviruses: molecular epidemiology and evolution. *Curr. Top. Microbiol. Immunol.* 292, 81–117 (2005).

Review Leyrat, Ribeiro, Gérard, Ivanov, Ruigrok & Jamin

- Pringle CR: The order Mononegavirales current status. Arch. Virol. 142, 2321–2326 (1997).
- Schoehn G, Iseni F, Mavrakis M, Blondel D, Ruigrok RW: Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J. Virol.* 75, 490–498 (2001).
- Ge P, Tsao J, Schein S, Green TJ, Luo M, Zhou ZH: Cryo-EM model of the bulletshaped vesicular stomatitis virus. *Science* 327, 689–693 (2010).
- Molecular organization of the helical part of the vesicular stomatitis virus particle was revealed by 3D reconstruction at 11 Å resolution by cryoelectron microscopy.
- 22. Albertini AA, Wernimont AK, Muziol T *et al.*: Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 313, 360–363 (2006).
- Structure of a recombinant circular nucleoprotein (N)–RNA complex of rabies virus (RV) revealed how N encapsidates the RNA genome.
- 23. Iseni F, Baudin F, Blondel D, Ruigrok RW: Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *RNA* 6, 270–281 (2000).
- Schnell MJ, McGettigan JP, Wirblich C, Papaneri A: The cell biology of rabies virus: using stealth to reach the brain. *Nat. Rev. Microbiol.* 8, 51–61 (2010).
- Comprehensive review of the replication cycle of rabies virus and of the interactions of the virus with its host cells.
- Gaudin Y, Tuffereau C, Durrer P, Brunner J, Flamand A, Ruigrok R: Rabies virus-induced membrane fusion. *Mol. Membr. Biol.* 16, 21–31 (1999).
- Roche S, Bressanelli S, Rey FA, Gaudin Y: Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science* 313, 187–191 (2006).
- Roche S, Rey FA, Gaudin Y, Bressanelli S: Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. *Science* 315, 843–848 (2007).
- Lahaye X, Vidy A, Pomier C et al.: Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: evidence that NBs are sites of viral transcription and replication. J. Virol. 83, 7948–7958 (2009).
- Provides clear evidences that, in RVinfected cells, RNA synthesis occurs in well-defined cytoplasmic inclusion bodies, suggesting the compartmentalization of the viral replication machineries.

- Heinrich BS, Cureton DK, Rahmeh AA, Whelan SP: Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. *PLoS Pathog.* 6, E1000958 (2010).
- Barr JN, Whelan SP, Wertz GW: Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Biochim. Biophys. Acta* 1577, 337–353 (2002).
- Komarova AV, Real E, Borman AM *et al.*: Rabies virus matrix protein interplay with eIF3, new insights into rabies virus pathogenesis. *Nucleic Acids Res.* 35, 1522–1532 (2007).
- Ogino T, Banerjee AK: Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol. Cell* 25, 85–97 (2007).
- Qanungo KR, Shaji D, Mathur M, Banerjee AK: Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. *Proc. Natl Acad. Sci. USA* 101, 5952–5957 (2004).
- Patton JT, Davis NL, Wertz GW: N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus. *J. Virol.* 49, 303–309 (1984).
- Finke S, Conzelmann KK: Dissociation of rabies virus matrix protein functions in regulation of viral RNA synthesis and virus assembly. J. Virol. 77, 12074–12082 (2003).
- Dancho B, McKenzie MO, Connor JH, Lyles DS: Vesicular stomatitis virus matrix protein mutations that affect association with host membranes and viral nucleocapsids. *J. Biol. Chem.* 284, 4500–4509 (2009).
- Harty RN, Paragas J, Sudol M, Palese P: A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding. J. Virol. 73, 2921–2929 (1999).
- Wirblich C, Tan GS, Papaneri A *et al.*: PPEY motif within the rabies virus (RV) matrix protein is essential for efficient virion release and RV pathogenicity. *J. Virol.* 82, 9730–9738 (2008).
- Chenik M, Chebli K, Blondel D: Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism. *J. Virol.* 69, 707–712 (1995).
- Pasdeloup D, Poisson N, Raux H, Gaudin Y, Ruigrok RW, Blondel D: Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1dependent nuclear export signal. *Virology* 334, 284–293 (2005).

- Vidy A, El Bougrini J, Chelbi-Alix MK, Blondel D: The nucleocytoplasmic rabies virus P protein counteracts interferon signaling by inhibiting both nuclear accumulation and DNA binding of STAT1. *J. Virol.* 81, 4255–4263 (2007).
- Moseley GW, Filmer RP, DeJesus MA, Jans DA: Nucleocytoplasmic distribution of rabies virus P-protein is regulated by phosphorylation adjacent to C-terminal nuclear import and export signals. *Biochemistry* 46, 12053–12061 (2007).
- Nadin-Davis SA, Abdel-Malik M, Armstrong J, Wandeler AI: Lyssavirus P gene characterisation provides insights into the phylogeny of the genus and identifies structural similarities and diversity within the encoded phosphoprotein. *Virology* 298, 286–305 (2002).
- Ito N, Moseley GW, Blondel D *et al.*: Role of interferon antagonist activity of rabies virus phosphoprotein in viral pathogenicity. *J. Virol.* 84, 6699–6710 (2010).
- 45. Vidy A, Chelbi-Alix M, Blondel D: Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways. J. Virol. 79, 14411–14420 (2005).
- Kretzschmar E, Peluso R, Schnell MJ, Whitt MA, Rose JK: Normal replication of vesicular stomatitis virus without C proteins. *Virology* 216, 309–316 (1996).
- Spiropoulou CF, Nichol ST: A small highly basic protein is encoded in overlapping frame within the *P* gene of vesicular stomatitis virus. *J. Virol.* 67, 3103–3110 (1993).
- Childs KS, Rejeva J, Randall RE, Goodbourn S: Mechanism of MDA-5 inhibition by paramyxovirus V proteins. *J. Virol.* 83, 1465–1473 (2009).
- Poole E, He B, Lamb RA, Randall RE, Goodbourn S: The V proteins of simian virus 5 and other paramyxoviruses inhibit induction of interferon-β. *Virology* 303, 33–46 (2002).
- Gupta AK, Blondel D, Choudhary S, Banerjee AK: 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 (2000).
- Barik S, Banerjee AK: 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 (1992).
- Chattopadhyay D, Banerjee AK: Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription *in vitro*. *Cell* 49, 407–414 (1987).

- 53. Hwang LN, Englund N, Das T, Banerjee AK, Pattnaik AK: 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 (1999).
- Gao Y, Lenard J: Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly. *J. Virol.* 69, 7718–7723 (1995).
- Pattnaik AK, Hwang L, Li T *et al.*: 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 (1997).
- Das T, Gupta AK, Sims PW, Gelfand CA, Jentoft JE, Banerjee AK: 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 (1995).
- Gao Y, Lenard J: Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II. *EMBO J.* 14, 1240–1247 (1995).
- Gérard FC, Ribeiro E, Albertini A *et al.*: Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry* 46, 10328–10338 (2007).
- Gigant B, Iseni F, Gaudin Y, Knossow M, Blondel D: Neither phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is required for its oligomerization. *J. Gen. Virol.* 81, 1757–1761 (2000).
- Gill DS, Chattopadhyay D, Banerjee AK: 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 (1986).
- Das SC, Pattnaik AK: 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 (2005).
- 62. Emerson SU, Schubert M: Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl Acad. Sci.* USA 84, 5655–5659 (1987).
- 63. Takacs AM, Das T, Banerjee AK: 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 (1993).

- Barik S, Banerjee AK: Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P. *Proc. Natl Acad. Sci.* USA 89, 6570–6574 (1992).
- Paul PR, Chattopadhyay D, Banerjee AK: The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype). *Virology* 166, 350–357 (1988).
- 66. Das T, Pattnaik AK, Takacs AM, Li T, Hwang LN, Banerjee AK: Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. Virology 238, 103–114 (1997).
- Ding H, Green TJ, Lu S, Luo M: Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* 80, 2808–2814 (2006).
- Mavrakis M, McCarthy AA, Roche S, Blondel D, Ruigrok RW: Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J. Mol. Biol.* 343, 819–831 (2004).
- Gérard FC, Ribeiro EA, Ede A Jr *et al.*: Modular organization of rabies virus phosphoprotein. *J. Mol. Biol.* 388, 978–996 (2009).
- Provides evidences for the modular organization of RV phosphoprotein at the structural level and clearly defines the boundaries between intrinsically disordered regions and structured domains.
- Ribeiro EA Jr, Favier A, Gerard FC *et al.*: Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J. Mol. Biol.* 382, 525–538 (2008).
- Mavrakis M, Mehouas S, Real E *et al.*: Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology* 349, 422–429 (2006).
- Jacob Y, Real E, Tordo N: Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.* 75, 9613–9622 (2001).
- Ivanov I, Crepin T, Jamin M, Ruigrok R: Structure of the dimerisation domain of the rabies virus phosphoprotein. *J. Virol.* 84, 3707–3710 (2010).
- Karlin D, Ferron F, Canard B, Longhi S: Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* 84, 3239–3252 (2003).

- Wright PE, Dyson HJ: Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331 (1999).
- He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK: Predicting intrinsic disorder in proteins: an overview. *Cell Res.* 19, 929–949 (2009).
- Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS: Do viral proteins possess unique biophysical features? *Trends Biochem. Sci.* 34, 53–59 (2009).
- 78. Tompa P: Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27, 527–533 (2002).
- Uversky VN, Oldfield CJ, Dunker AK: Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* 18, 343–384 (2005).
- Spolar RS, Record MT Jr: Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263, 777–784 (1994).
- Wright PE, Dyson HJ: Linking folding and binding. *Curr. Opin. Struct. Biol.* 19, 31–38 (2009).
- Leyrat C, Jensen MR, Ribeiro EA *et al.*: The N⁰-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α-helices. *Protein Sci.* 20(3), 542–556 (2011).
- Ding H, Green TJ, Luo M: Crystallization and preliminary X-ray analysis of a proteinase-Kresistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). *Acta Crystallogr. D Biol. Crystallogr.* 60, 2087–2090 (2004).
- Kolakofsky D, Le Mercier P, Iseni F, Garcin D: Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology* 318, 463–473 (2004).
- Moseley GW, Lahaye X, Roth DM *et al.*: Dual modes of rabies P-protein association with microtubules: a novel strategy to suppress the antiviral response. *J. Cell Sci.* 122, 3652–3662 (2009).
- Moseley GW, Roth DM, DeJesus MA *et al.*: Dynein light chain association sequences can facilitate nuclear protein import. *Mol. Biol. Cell* 18, 3204–3213 (2007).
- Leopold PL, Pfister KK: Viral strategies for intracellular trafficking: motors and microtubules. *Traffic* 7, 516–523 (2006).
- Dingwall C, Laskey RA: Nuclear targeting sequences – a consensus? *Trends Biochem. Sci.* 16, 478–481 (1991).
- Bogerd HP, Fridell RA, Benson RE, Hua J, Cullen BR: Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel *in vivo* randomization-selection assay. *Mol. Cell. Biol* 16, 4207–4214 (1996).

Review Leyrat, Ribeiro, Gérard, Ivanov, Ruigrok & Jamin

- 90. Raux H, Flamand A, Blondel D: Interaction of the rabies virus P protein with the LC8 dynein light chain. *J. Virol.* 74, 10212–10216 (2000).
- Jacob Y, Badrane H, Ceccaldi PE, Tordo N: Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein. *J. Virol.* 74, 10217–10222 (2000).
- Klingen Y, Conzelmann KK, Finke S: Double-labeled rabies virus: live tracking of enveloped virus transport. *J. Virol.* 82, 237–245 (2008).
- Tan GS, Preuss MA, Williams JC, Schnell MJ: The dynein light chain 8 binding motif of rabies virus phosphoprotein promotes efficient viral transcription. *Proc. Natl Acad. Sci. USA* 104, 7229–7234 (2007).
- 94. Mebatsion T: 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, 11496–11502 (2001).
- Das SC, Nayak D, Zhou Y, Pattnaik AK: Visualization of intracellular transport of vesicular stomatitis virus nucleocapsids in living cells. J. Virol. 80, 6368–6377 (2006).
- Poisson N, Real E, Gaudin Y *et al.*: Molecular basis for the interaction between rabies virus phosphoprotein P and the dynein light chain LC8: dissociation of dynein-binding properties and transcriptional functionality of P. *J. Gen. Virol.* 82, 2691–2696 (2001).
- Katze MG, Fornek JL, Palermo RE, Walters KA, Korth MJ: Innate immune modulation by RNA viruses: emerging insights from functional genomics. *Nat. Rev. Immunol.* 8, 644–654 (2008).
- Chelbi-Alix MK, Vidy A, El Bougrini J, Blondel D: Rabies viral mechanisms to escape the IFN system: the viral protein P interferes with IRF-3, Stat1, PML nuclear bodies. J. Interferon Cytokine Res. 26, 271–280 (2006).
- Rieder M, Conzelmann KK: Rhabdovirus evasion of the interferon system. J. Interferon Cytokine Res. 29, 499–509 (2009).
- 100. Wang ZW, Sarmento L, Wang Y et al.: Attenuated rabies virus activates, while pathogenic rabies virus evades, the host innate immune responses in the central nervous system. J. Virol. 79, 12554–12565 (2005).
- Hornung V, Ellegast J, Kim S et al.:
 5'-triphosphate RNA is the ligand for RIG-I. Science 314, 994–997 (2006).
- 102. Brzozka K, Finke S, Conzelmann KK: Identification of the rabies virus α/β interferon antagonist: phosphoprotein P interferes with phosphorylation of interferon regulatory factor 3. J. Virol. 79, 7673–7681 (2005).

- Delhaye S, Paul S, Blakqori G *et al.*: Neurons produce type I interferon during viral encephalitis. *Proc. Natl Acad. Sci. USA* 103, 7835–7840 (2006).
- 104. Shimizu K, Ito N, Sugiyama M, Minamoto N: Sensitivity of rabies virus to type I interferon is determined by the phosphoprotein gene. *Microbiol. Immunol.* 50, 975–978 (2006).
- 105. Finke S, Brzozka K, Conzelmann KK: Tracking fluorescence-labeled rabies virus: enhanced green fluorescent protein-tagged phosphoprotein P supports virus gene expression and formation of infectious particles. J. Virol. 78, 12333–12343 (2004).
- 106. Brzozka K, Finke S, Conzelmann KK: Inhibition of interferon signaling by rabies virus phosphoprotein P: activation-dependent binding of STAT1 and STAT2. *J. Virol.* 80, 2675–2683 (2006).
- 107. Bernardi R, Pandolfi PP: Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat. Rev. Mol. Cell Biol.* 8, 1006–1016 (2007).
- 108. Blondel D, Regad T, Poisson N *et al.*: Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies. *Oncogene* 21, 7957–7970 (2002).
- 109. Blondel D, Kheddache S, Lahaye X, Dianoux L, Chelbi-Alix MK: Resistance to rabies viral infection conferred by PMLIV isoform. J. Virol. 84(20), 10719–10726 (2010).
- 110. Chen M, Ogino T, Banerjee AK: Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N⁰–P complex formation and encapsidation of viral genome RNA. *J. Virol.* 81, 13478–13485 (2007).
- 111. Mavrakis M, Iseni F, Mazza C *et al.*: Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells. *Virology* 305, 406–414 (2003).
- 112. Howard M, Wertz G: Vesicular stomatitis virus RNA replication: a role for the NS protein. *J. Gen. Virol.* 70(Pt 10), 2683–2694 (1989).
- 113. Peluso RW, Moyer SA: Viral proteins required for the *in vitro* replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology* 162, 369–376 (1988).
- Green TJ, Zhang X, Wertz GW, Luo M: Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 313, 357–360 (2006).
- 115. Hudson LD, Condra C, Lazzarini RA: Cloning and expression of a viral phosphoprotein: structure suggests vesicular stomatitis virus NS may function by mimicking an RNA template. *J. Gen. Virol.* 67(Pt 8), 1571–1579 (1986).

- Mohan A, Oldfield CJ, Radivojac P *et al.*: Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* 362, 1043–1059 (2006).
- 117. Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RW, Blackledge M: Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. J. Am. Chem. Soc. 130, 8055–8061 (2008).
- Dyson HJ, Wright PE: Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12, 54–60 (2002).
- Emerson SU, Yu Y: Both NS and L proteins are required for *in vitro* RNA synthesis by vesicular stomatitis virus. *J. Virol.* 15, 1348–1356 (1975).
- 120. De BP, Banerjee AK: Requirements and functions of vesicular stomatitis virus L and NS proteins in the transcription process *in vitro. Biochem. Biophys. Res. Commun.* 126, 40–49 (1985).
- 121. Stillman EA, Whitt M: Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription. J. Virol. 73, 7199–7209 (1999).
- 122. Mellon MG, Emerson SU: Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus. J. Virol. 27, 560–567 (1978).
- 123. Castel G, Chteoui M, Caignard G et al.: Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. J. Virol. 83, 10808–10820 (2009).
 - Reveals that a peptide corresponding to the 60 first residues of RV phosphoprotein and encompassing the №-binding region prevents the replication of the virus in neuronal cells.
- 124. Chenik M, Schnell M, Conzelmann KK, Blondel D: Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J. Virol.* 72, 1925–1930 (1998).
- 125. Ptashne M, Gann A: Transcriptional activation by recruitment. *Nature* 386, 569–577 (1997).
- 126. Ma J, Ptashne M: Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* 48, 847–853 (1987).
- 127. Chattopadhyay D, Banerjee AK: NH2terminal acidic region of the phosphoprotein of vesicular stomatitis virus can be functionally replaced by tubulin. *Proc. Natl Acad. Sci. USA* 85, 7977–7981 (1988).
- 128. Moyer SA, Baker SC, Lessard JL: Tubulin: a factor necessary for the synthesis of both Sendai virus and vesicular stomatitis virus RNAs. *Proc. Natl Acad. Sci. USA* 83, 5405–5409 (1986).

Structure, interactions with host-cell & functions of rhabdovirus phosphoprotein Review

- 129. Chenik M, Chebli K, Gaudin Y, Blondel D: *In vivo* interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein. *J. Gen. Virol.* 75 2889–2896 (1994).
- 130. Fu ZF, Zheng Y, Wunner WH, Koprowski H, Dietzschold B: Both the N- and the C-terminal domains of the nominal phosphoprotein of rabies virus are involved in binding to the nucleoprotein. *Virology* 200, 590–597 (1994).
- Assenberg R, Delmas O, Ren J *et al.*: The structure of the N–RNA binding domain of the Mokola virus phosphoprotein. *J. Virol.* 84, 1089–1096 (2009).
- 132. Delmas O, Assenberg R, Grimes JM, Bourhy H: The structure of the nucleoprotein binding domain of lyssavirus phosphoprotein reveals a structural relationship between the N–RNA binding domains of Rhabdoviridae and Paramyxoviridae. *RNA Biol.* 7, 322–327 (2010).

- Ribeiro EA, Leyrat C, Gérard FC *et al.*: Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes. *J. Mol. Biol.* 394, 558–575 (2009).
- This model of the complex between the rabies virus N–RNA complex and the C-terminal domain of the phosphoprotein reveals a binding site on the template that requires two adjacent Ns and the folding of their flexible C-terminal loops.
- 134. Green TJ, Luo M: Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc. Natl Acad. Sci.* USA 106, 11721–11726 (2009).
- Crystal structure of the vesicular stomatitis virus complex reveals a very similar structural organization than in the modeled RV complex.

- 135. Toriumi H, Kawai A: Association of rabies virus nominal phosphoprotein (P) with viral nucleocapsid (NC) is enhanced by phosphorylation of the viral nucleoprotein (N). *Microbiol. Immunol.* 48, 399–409 (2004).
- 136. Thomas D, Newcomb WW, Brown JC et al.: Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. J. Virol. 54, 598–607 (1985).
- Pei J, Grishin NV: AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* 17, 700–712 (2001).

Website

201. WHO: rabies

www.who.int/mediacentre/factsheets/fs099/ en/
Résumé :

Les rhabdovirus, dont les virus de la stomatite vésiculaire (VSV) et de la rage (RAV) constituent des prototypes, sont des virus enveloppés dont le génome est constitué d'une seule molécule d'ARN simple brin de polarité négative qui font partie de l'ordre des *Mononegavirales* (MNV). La machinerie de transcription/réplication de ces virus est constituée de l'ARN génomique et de trois protéines qui sont communes à tous les virus de l'ordre des MNV, la (N) qui encapside le génome viral, la grande sous-unité de l'ARN polymérase ARN dépendante (L) et la phosphoprotéine (P) qui est un cofacteur non-catalytique de la L et sert de chaperonne à la N.

Le premier objectif de mon travail de thèse consistait à déterminer la structure cristallographique du domaine de dimérisation de la phosphoprotéine du virus de la rage. La P des rhabdovirus est une protéine modulaire qui contient deux régions intrinsèquement désordonnée, un domaine central responsable de la dimérisation et un domaine C-terminal responsable de la fixation sur la matrice N-ARN. Le modèle atomique obtenu à une résolution de 1.5 Å montre que la structure est très différente de celle du domaine correspondant chez VSV.

Le second objectif de mon travail était la caractérisation structurale de la grande sous-unité L de la polymérase du virus de la stomatite vésiculaire. Cette enzyme de 2109 aa, possède six régions conservées. Le domaine conservé III comprend les régions impliquées dans l'activité de polymérisation et les domaines V et VI sont responsables de la formation de la coiffe des ARNm. Plusieurs stratégies ont été envisagées successivement. (1) Sur la base de prédictions de structures secondaires et de prédictions de désordre, nous avons essayé d'exprimer différents fragments en système d'expression bactérien. Les constructions testées se sont avérées insolubles et certaines d'entre elles fixaient GroEL, indiquant un problème de repliement. (2) Nous avons alors essayé d'exprimer la L seule ou en complexe avec la P en système d'expression eucaryote. La purification s'est avérée impossible, la protéine L restant toujours associée à des protéines cellulaires visibles par coloration au bleu de Coomassie. (3) Finalement nous avons réussi à purifier la polymérase à partir de virus entier. La préparation de la polymérase était très homogène et a permis d'entreprendre une caractérisation par microscopie électronique. Une classification d'images a permis de construire un premier modèle à basse résolution. Le modèle révèle la présence d'un domaine annulaire avec plusieurs domaines structurés attachés au coeur de la polymérase. La cryo-microscopie électronique et la tomographie permettront d'obtenir plus de détails sur cette protéine.

Abstract :

Rhabdoviruses, including vesicular stomatitis virus (VSV) and rabies virus (RAV), are enveloped viruses which genome is made of a single molecule of negative-sense RNA and are classified in the order *Mononegavirales* (MNV). The transcription/replication machinery of these viruses consists of the genomic RNA and of three proteins, which are common to all other viruses of the order MNV, a nucleoprotein (N) that encapsidates the viral genome, a large subunit of the RNA-dependent RNA polymerase (L) and a phosphoprotein (P) that acts as a non-catalytic cofactor of L and a chaperone of N.

The first goal of my research project was to determine the crystallographic structure of the dimerization domain of the rabies virus phosphoprotein. The P protein of the rhabdoviruses is a modular protein, which contains two intrinsically disordered regions, a central dimerization domain and a C-terminal domain involved in binding to the N-RNA template. The atomic model obtained at a resolution of 1.5 Å showed that the structure is different from that of the corresponding domain of VSV.

The second goal was the structural characterization of the large subunit L of VSV polymerase. The enzyme of 2109 aa has six conserved regions. Conserved region III includes the residues involved in the RNA synthesis activity, whereas domains V and VI are involved in mRNA capping formation. Three strategies were successively developed : (1) On the basis of secondary structure and disorder predictions, we tried to express different fragments in bacterial expression systems. These constructions appeared to be insoluble and some of them bound GroEL suggesting a folding problem; (2) We tried to express L alone or co-express it with P in eukaryotic expression system. The purification appeared to be impossible, the L protein always remaining associated with host-cell proteins in amounts detectable by Coomassie staining; (3) We succeeded in purifying the L protein from the virus. The L samples were homogeneous and allowed a characterization by electron microscopy. Image classes allowed the reconstruction of a first low-resolution model. This model revealed the presence of a large ring-like domain and several globular domains. Cryo-electron microscopy and tomography should lead to a more detailed description of this protein.