



**HAL**  
open science

# Caractérisation structurale des protéines P et C encodées par des séquences chevauchantes chez les Orthoparamyxovirinae

Ada Roy

► **To cite this version:**

Ada Roy. Caractérisation structurale des protéines P et C encodées par des séquences chevauchantes chez les Orthoparamyxovirinae. Microbiologie et Parasitologie. Université Grenoble Alpes [2020-..], 2023. Français. NNT : 2023GRALV014 . tel-04146755

**HAL Id: tel-04146755**

**<https://theses.hal.science/tel-04146755>**

Submitted on 30 Jun 2023

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

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

THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

École doctorale : CSV- Chimie et Sciences du Vivant

Spécialité : Virologie - Microbiologie - Immunologie

Unité de recherche : Institut de Biologie Structurale

**Caractérisation structurale des protéines P et C encodées par des séquences chevauchantes chez les Orthoparamyxovirinae**

**Structural characterization of Orthoparamyxovirinae P and C proteins encoded by overlapping sequences**

Présentée par :

**Ada ROY**

Direction de thèse :

**Marc JAMIN**  
Université Grenoble Alpes

Directeur de thèse

**Jean Marie BOURHIS**  
Université Grenoble Alpes

Co-encadrant de thèse

Rapporteurs :

**Branka HORVAT**  
DIRECTRICE DE RECHERCHE, INSERM délégation Auvergne-Rhône-Alpes

**Bernado PAU**  
DIRECTEUR DE RECHERCHE, INSERM délégation Occitanie Méditerranée

Thèse soutenue publiquement le **9 mars 2023**, devant le jury composé de :

**Marc JAMIN**  
PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Directeur de thèse

**Branka HORVAT**  
DIRECTRICE DE RECHERCHE, INSERM délégation Auvergne-Rhône-Alpes

Rapporteuse

**Bernado PAU**  
DIRECTEUR DE RECHERCHE, INSERM délégation Occitanie Méditerranée

Rapporteur

**Denis GERLIER**  
DIRECTEUR DE RECHERCHE, CNRS délégation Rhône Alpes Auvergne

Examineur

**Emmanuel DROUET**  
PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Président

**Jean-Marie BOURHIS**  
MAITRE DE CONFERENCES HDR, Université Grenoble Alpes

Co-directeur de thèse





# Remerciements

L'écriture des remerciements représente le point final de ma rédaction et, à fortiori, d'une aventure humaine et scientifique exaltante au cours de laquelle j'ai eu la chance de recevoir l'aide bienveillante de nombreuses personnes. Ces pages vous sont adressées et j'espère réussir à y exprimer toute la gratitude que je ressens de vous avoir croisées sur mon chemin et d'avoir partagé avec vous cette expérience.

Je voudrais tout d'abord adresser mes remerciements à mon directeur de thèse, le professeur, et récemment élu directeur de l'École doctorale Chimie et Science du Vivant, **Marc Jamin**, pour avoir fait le pari de donner sa chance à une ingénieure agronome en reconversion en thèse au sein de son groupe « Viral Replication Machine » (VRM) à l'Institut de Biologie Structurale (IBS). Nous n'aurions pas pu prévoir que cette expérience serait perturbée par une pandémie mondiale mais, malgré tout, tu as su faire preuve de patience, de disponibilité et de rigueur tout au long de ces travaux, merci !

Je tiens très sincèrement à exprimer toute ma reconnaissance à mon co-directeur de thèse, le professeur **Jean-Marie Bourhis**, présent tout au long de ce projet pour me former au laboratoire et pour me pousser à aller « au bout des manip ». Il m'aura guidé à travers les méandres du métier de la recherche avec ses joies et ses déceptions. Il a été à l'initiative de nombreuses collaborations me poussant sans cesse à tester diverses approches. C'est ton énergie et ta bonne humeur qui m'auront permise, à de nombreuses reprises, à ne rien abandonner et à toujours voir l'envers positif des difficultés traversées. Bien que la mémoire de mon ordinateur risque de ne jamais te pardonner, je te remercie de m'avoir guidé dans l'apprentissage des (trop) nombreux logiciels de biologie structurale.

Je tiens à exprimer toute ma gratitude aux membres du jury qui me font l'honneur d'avoir accepté d'évaluer mon travail de thèse. Je remercie particulièrement le **Dr. Branka Horvat** et **Dr. Pau Bernadó** qui ont accepté de relire mon manuscrit et dont les travaux m'ont régulièrement accompagnée tout au long de ce projet. Mes remerciements s'adressent également au **Dr. Emmanuel Drouet**, professeur à l'Université Grenoble-Alpes, et le **Dr. Denis Gerlier** qui ont accepté de faire partie de ce jury en tant qu'examineurs.

Mes remerciements chaleureux sont également adressés aux membres de mon comité de suivi de thèse, le **Dr. Carlo Petosa**, **Dr. Pascal Fender** et **Dr. Luis Martinez-Gil**, qui ont suivi mon travail tout au long de ces trois années et qui ont prêté une oreille attentive et un regard neuf sur mes résultats et, notamment, sur les difficultés traversées. Ils m'ont offert des paroles rassurantes qui soulagent quand le poids de l'exercice devient un peu trop lourd à porter.

Je remercie le professeur **Winfried Weissenhorn**, directeur de l'IBS, de m'avoir accueillie dans les locaux du CIBB pour la réalisation de ce projet. J'adresse aussi tous mes remerciements à mon École doctorale, Chimie et Sciences du Vivant, et à son directeur au cours de ce projet de thèse, le professeur **Guy Royal**, pour m'avoir acceptée dans les rangs des doctorants de l'université Grenoble Alpes.

Je souhaite également remercier les nombreux collaborateurs de ce projet, qu'ils soient proches géographiquement ou virtuellement. Merci aux membres de l'équipe « Base moléculaire de la pathogénicité virale » du professeur **Viktor Volchkov** pour leur rôle dans le

cadre de ce projet « Nipah-C » co-financé par l'Agence Nationale de la Recherche. Un immense merci au **Dr. David Karlin** pour son aide précieuse dans l'interprétation et la relecture, ses conseils avisés pour l'écriture de notre publication sur la protéine C du virus de la toupaye. Merci au **Dr. Luis Martinez-Gil** pour ta disponibilité, notamment durant la « période COVID », pour discuter de mon projet et pour ton aide à la réalisation de ce travail. Merci à l'équipe « Fonctionnalité de l'Écosystème Intestinal » du **Dr. Nicolas Lapaque** d'avoir accepté de m'aider pour la réalisation de l'interactome de la protéine C du virus Nipah. Merci également aux « local contact » des synchrotrons SOLEIL de Paris-Saclay, les **Dr. Aurélien Thureau** et **Dr. Pierre Legrand**, qui m'ont accueillie à plusieurs reprises et qui m'ont accordé du temps pour l'analyse de mes échantillons et le traitement des données. Merci également aux membres du synchrotron de l'ESRF pour m'avoir permise de travailler sur les beamline BM29 et ID30A-1. Enfin, je souhaitais également remercier les membres de la plateforme HTX qui m'ont conseillé à de nombreuses reprises lors de mes nombreux essais de cristallographie.

Bien évidemment, ce travail n'aurait pas abouti sans l'accueil généreux des membres du CIBB à qui j'adresse mes chaleureux remerciements. Vous m'avez permise de me sentir la bienvenue dans une ville inconnue et un environnement nouveau (vivre à côté des montagnes, quelle idée merveilleuse) et pour cela je vous remercie. Je souhaite tout particulièrement remercier le professeur **Nicolas Tarbouriech** qui, non content de réparer, avec toute la patience et la gentillesse qui le caractérisent, toutes les machines de notre laboratoire usées par la présence de nombreux doctorants maladroits, a également répondu à mes nombreuses questions qui a accepté de corriger en partie mon manuscrit. Je souhaite également remercier le **Dr. Caroline Mas**, ingénieure en charge de la plateforme de biophysique (ISBG), sans qui nos travaux ne seraient tout simplement pas réalisables. Merci également à **Philippe Mas**, ingénieur en charge de la plateforme ESPRIT, pour le temps investi sur ce projet et pour les nombreux conseils et astuces prodigués en biologie moléculaire. Merci à tous les deux pour votre bonne humeur et pour toujours accorder du temps aux doctorants dans les moments difficiles et moins difficiles. Merci au **Dr. Darren Hart** d'avoir accepté d'allouer la plateforme ESPRIT à mon projet et d'avoir relu et suivi ma candidature à la mise en place d'une collaboration avec le consortium INSTRUCT-ERIC. Je souhaite également remercier le **Dr. Pascal Fender** pour son aide précieuse lors des expérimentations cellulaires au L2. Merci également au **Dr. Thibault Crépin** pour le partage de son matériel de cristallographie. Enfin, je souhaite remercier les équipes « de l'ombre », ceux sans qui nos laboratoires tomberaient bien vite dans le chaos, **Ebru Oner**, **Pierre Métais**, **José Barranco**. Ebru, j'en profite pour t'adresser mes excuses pour les bacs de vaisselle bien trop nombreux et bien trop remplis en provenance du labo 008. On ne dira à personne que tu t'ennuies depuis qu'on a commencé la rédaction...

S'agissant des doctorants du CIBB, mes remerciements s'adressent tout particulièrement aux co-thésards du CIBB et notamment, à vous deux, **Solène Besson** et **Fella Bouchama**. Vous avez été présentes tout au long de ces dernières années pour partager cette montagne russe émotionnelle que représente l'exercice de la thèse. Fella, merci pour toute ces heures partagées entre le laboratoire 008 et la chambre froide à discuter, à débriefer, à rigoler et à réaliser nos fameux karaokés qui ont permis au temps de passer plus rapidement malgré les nocturnes. Solène, mille mercis pour avoir été présente à mes côtés avec autant d'attention pour tous les nombreux rebondissements de ma petite vie (je crois qu'on en est à l'épisode 5243, saison 235, si je ne me trompe pas ?). Je souhaitais également remercier du fond du cœur les stagiaires qui sont venus affronter ce projet (parfois plusieurs fois, c'est à n'y rien

comprendre...), **Élodie**, **Paco** et **Ardita**. Vous encadrer a été un réel plaisir et j'espère que vous en gardez un bon souvenir.

Ces travaux de thèse représentent l'aboutissement d'un long parcours rempli de rencontres et il m'est bien évidemment impossible de tous vous citer. Cependant, je voudrais tout de même adresser une mention spéciale aux Supagroïens (les SLB qui se reconnaîtront, les anciens de toujours et H4iens que j'ai retrouvés près des palmiers, **Célia** et **André**, et mes fidèles de la pause-café, **Eva**, **Émeline** et **Déborah**) qui, malgré mes absences des dernières années, sont restés présents. Sachez que je souhaite de tout mon cœur être plus disponible pour vous aujourd'hui. Une pensée spéciale est adressé à **Sélim**, sans qui je ne serais probablement pas là aujourd'hui (littéralement et métaphoriquement), merci à toi. Merci à vous deux, **Loïc** et **Maya**, mes colocataires lyonnais, de m'avoir accueillie à vos côtés. Merci également aux retrouvailles de ces dernières années mais dont le soutien a toujours été indéfectible et que j'espère recroiser prochainement sur ma route, **Zina Lanseur**. Merci à toi **Adèle**, ma déménageuse de l'extrême, ma belle rencontre grenobloise, de m'avoir permise de me sentir chez moi dans ce nouvel endroit.

Bien évidemment, je ne serais pas arrivée jusqu'ici sans le soutien inconditionnel de ma famille, et tout particulièrement de mes parents, **Agnès** et **Pierre**, qui m'ont toujours encouragée à poursuivre mes rêves et soutenue tout au long de mes études sans douter. Je ne suis pas vétérinaire aujourd'hui mais je suis fière d'être arrivée jusqu'ici, et c'est bien grâce à vous. Vous m'avez donné le goût du challenge, et vous voir investis dans vos métiers a construit la personne que je suis aujourd'hui. Merci tout particulièrement à toi maman pour tes encouragements sur cette dernière ligne droite.

Merci à mes frères, **Quentin**, **Axel** et **Willem**, et à ma sœur, **Lola**. Grandir dans une famille nombreuse comme la nôtre peut présenter ses difficultés mais c'est avant tout une vraie chance de se construire entourée de personnes passionnantes. Lola, tu as été très présente ces dernières années pour moi et je t'en remercie du fond du cœur, ces années nous auront rapprochées et j'espère que les prochaines ne feront que renforcer cela.

Pour ceux qui me connaissent, vous savez à quel point j'attache de l'importance à nos animaux qui sont pour moi des membres à part entière de ma famille. C'est pourquoi je ne peux m'empêcher d'adresser une pensée à un absent qui nous aura quittés un peu trop tôt, **Juju**. Merci à toi d'avoir été là tout au long de mes études et de mes nuits blanches de révision.

Je terminerai en remerciant une personne dont la présence à mes côtés ces deux dernières années a rendu le quotidien plus doux. **Thibault**, ton soutien a été sans faille et ta présence un véritable pilier sur cette dernière ligne droite. J'ai vraiment commencé à apprécier la vie à Grenoble à tes côtés et j'ai hâte de découvrir ce que le futur nous réserve.



# Résumé

La famille des *Paramyxoviridae* est une grande famille de virus à ARN négatif non-segmenté dont certains sont pathogènes pour l'homme ou les animaux domestiques. Leur génome de taille réduite comprend six unités de transcription, dont celle codant pour la phosphoprotéine P qui est polycistronique et permet la production, d'une part, de la phosphoprotéine, un composant essentiel pour le fonctionnement de la machine de synthèse d'AR et, d'autre part, d'une ou de plusieurs protéines accessoires, notamment impliquées dans l'échappement au système immunitaire. En particulier, le génome de la plupart des membres de la sous-famille des *Orthoparamyxovirinae* contient un gène surimprimé dans la séquence du gène codant pour la phosphoprotéine mais dans un cadre de lecture différent. Les cellules infectées par ces virus expriment ainsi une protéine accessoire C par un mécanisme d'initiation alternative de la traduction. Cette protéine est multifonctionnelle, elle est impliquée dans la régulation de la synthèse d'ARN viral, dans l'antagonisme de la réponse immunitaire innée et dans le bourgeonnement de nouvelles particules virales. Pour certains virus, cette protéine est un facteur de virulence. Mais, nous manquons d'informations sur la structure et les mécanismes d'action de la protéine C.

Mon travail a eu pour objectif de caractériser la structure de la protéine C de deux paramyxovirus, seule ou en complexe avec des partenaires cellulaires, afin de mieux comprendre comment cette protéine peut agir dans les cellules infectées.

Des études bio-informatiques ont confirmé la conservation d'une architecture commune à l'ensemble des protéines C des *Orthoparamyxovirinae* et l'existence d'une région N-terminale désordonnée et d'un domaine structuré en hélice- $\alpha$  dans la partie C-terminale de la protéine. Nous avons exprimé et purifié la protéine C du paramyxovirus de la toupe (TPMV) ainsi qu'un fragment C-terminal et nous avons déterminé la structure cristallographique de ce dernier. En dépit d'une faible similarité de séquence, la structure tridimensionnelle indique un repliement similaire à celui du domaine C-terminal du virus Sendai. À partir de données de diffusion de rayons X aux petits angles, nous avons pu établir un modèle d'ensemble de la protéine entière. En collaboration, nous avons montré que la protéine C du TPMV inhibe la machinerie de synthèse d'ARN du virus Nipah. Il a notamment été démontré dans le cas du virus de la rougeole que la protéine C inhibe la synthèse d'ARN par une interaction avec la phosphoprotéine. Dans cette optique, nous avons entrepris la caractérisation structurale de la phosphoprotéine du TPMV et nous avons cherché à déterminer si elle interagit avec la protéine C.

Nous avons utilisé différentes stratégies pour exprimer et purifier la protéine C du virus Nipah, seule ou en complexe avec des partenaires cellulaires. Par une méthode de criblage à haut débit, nous avons défini plusieurs constructions solubles et nous avons utilisé une de ces protéines pour générer des anticorps à domaine unique de lama qui seront utilisés comme chaperons de cristallisation. En collaboration, nous avons montré par les méthodes de double



hybride en levure et de TAP-TAG MS que la protéine C du NiV se lie à la protéine VPS37B (Vacuolar Protein Sorting 37 Homolog B) du sous-complexe cellulaire ESCRT-I de la machinerie ESCRT (Endosomal Sorting Complexes Required for Transport I). Nous avons coexprimé la protéine C avec le sous-complexe ESCRT-I et montré que la protéine C s'associe à ces protéines.

En conclusion, mes travaux démontrent que les protéines C des *Orthoparamyxovirinae* partagent une structure commune et peuvent agir de manière croisée entre espèces appartenant à des genres différents suggérant qu'elles ont une origine commune.

# Abstract

The *Paramyxoviridae* family is a large family of non-segmented negative RNA viruses, some of which are pathogenic to humans or domestic animals. Their small genome comprises six transcription units, including the one coding for the phosphoprotein, which is polycistronic and allows the production of both the phosphoprotein, an essential component for the functioning of the RNA synthesis machinery, and one or more accessory proteins, notably involved in the escape from the immune system. In particular, the genome of most members of the *Orthoparamyxovirinae* subfamily contains a superimposed gene in the sequence of the gene coding for the phosphoprotein but in a different reading frame. Cells infected with these viruses thus express an accessory protein C by a mechanism of alternative translation initiation. This protein is multi-functional and is involved in regulating viral RNA synthesis, antagonizing the innate immune response and budding new viral particles. For some viruses, this protein is a virulence factor. However, we lack information on the structure and mechanisms of action of protein C.

The aim of my work was to characterize the structure of the C protein of two paramyxoviruses, alone or in complex with cellular partners, in order to better understand how this protein can act in infected cells.

Bioinformatics studies confirmed the conservation of a common architecture to all *Orthoparamyxovirinae* C proteins and the existence of a disordered N-terminal region and a helix-structured domain in the C-terminal part of the protein. We expressed and purified the C protein of Tupaia paramyxovirus (TPMV) as well as a C-terminal fragment and determined the crystallographic structure of the latter. Despite low sequence similarity, the three-dimensional structure indicates a folding similar to that of the Sendai virus C-terminal domain. Using small-angle X-ray scattering data, we were able to establish an overall model of the entire protein. In collaboration, we have shown that the TPMV C protein inhibits the RNA synthesis machinery of Nipah virus. In particular, it has been shown in the case of measles virus that protein C inhibits RNA synthesis through an interaction with the phosphoprotein. With this in mind, we undertook the structural characterization of the TPMV phosphoprotein and investigated whether it interacts with the protein C.

We used different strategies to express and purify the Nipah virus C protein, alone or in complex with cellular partners. Using a high-throughput screening method, we defined several soluble constructs and used one of these proteins to generate llama nanobodies to use as crystallization chaperones. In collaboration, we have shown by yeast two-hybrid and TAP-TAG MS methods that NiV C protein binds to the VPS37B (Vacuolar Protein Sorting 37 Homolog B) protein of the ESCRT-I cellular subcomplex of the ESCRT (Endosomal Sorting Complexes Required for Transport) machinery. We co-expressed the protein C with the ESCRT-I subcomplex and showed that the protein C associates with these proteins.

In conclusion, my work demonstrates that the C proteins of the *Orthoparamyxovirinae* subfamily share a common structure and can cross-act with species belonging to different genera suggesting that they have a common origin.

Remerciements .....	3
Résumé .....	7
Abstract .....	9
Liste des abréviations.....	15
Liste des figures .....	17
<b>CHAPITRE 1 : INTRODUCTION .....</b>	<b>21</b>
I. Généralités.....	21
(A) Les virus émergents.....	21
(B) Les Orthoparamyxovirus : un sujet d'inquiétude ?.....	21
II. Découvertes du virus Nipah et du virus de la toupaye .....	22
(A) Le virus Nipah.....	23
(1) Émergence du virus Nipah .....	23
(2) Les espèces réservoirs.....	24
(3) Les hôtes intermédiaires.....	26
(4) Les différentes souches.....	27
(5) Tableau clinique.....	27
(6) Traitements.....	29
(B) Le virus de la toupaye .....	31
(1) Émergence.....	31
(2) Tropisme.....	32
(3) Utilisation médicale.....	32
III. Aperçu des spécificités du génome, des protéines et de la machinerie de réplication/transcription des <i>Orthoparamyxovirinae</i> , le cas du virus Nipah .....	33
(A) Classification.....	33
(B) Le génome viral .....	34
(C) Le Cycle viral.....	35
(1) Dissémination et transmission : structure du virion.....	36
(2) Entrée et tropisme.....	37
(3) La multiplication du virus : le complexe de réplication/transcription.....	40
(4) Assemblage et bourgeonnement .....	46
(F) Le contrôle de la réponse immunitaire innée antivirale .....	48
(1) Détection des pathogènes viraux et mise en place de l'immunité innée .....	48
(2) Reconnaissance des virus à ARN .....	51
(3) Réponse à l'IFN de type I.....	51
<b>CHAPITRE 2 : L'IMPORTANCE DES PROTEINES NON STRUCTURALES DES <i>ORTHOPARAMYXOVIRINAE</i> .....</b>	<b>55</b>
I. Les protéines non structurales des <i>Orthoparamyxovirinae</i> .....	55
(A) Les gènes chevauchants chez les virus .....	55
(B) Mécanismes d'expression des protéines accessoires.....	55
(1) Édition co-transcriptionnelle du gène P.....	57
(2) La réplication du génome et le changement de mode d'édition du gène P .....	57
(3) Surimpression translationnelle des ARNm codant pour P/V/W.....	58
(C) Structures et fonctions des protéines V, W et C.....	58
(1) La protéine V .....	58
(2) Protéine W.....	59
(3) Protéine C.....	60
II. Les fonctions des protéines C des <i>Orthoparamyxovirinae</i> .....	61
(A) Localisation cellulaire : navette entre le noyau et le cytoplasme ? .....	62
(B) Rôle dans la pathogénicité virale - immuno-évasion virale.....	63
(C) Rôle dans la réplication virale et transcription virale .....	64
(1) Modulation de l'activité de l'ARN polymérase virale .....	65
(2) Altération du gradient de transcription .....	65
(3) Modulation du rapport génome/antigénome.....	66
(4) Interaction avec le complexe polymérase .....	66
(5) Rôle dans l'inhibition de la production de DI RNA immunostimulateurs.....	67
(D) Interaction directe avec la réponse immunitaire innée .....	67
(1) Induction des IFN.....	67

(2) Interaction avec la cascade de signalisation de la réponse IFN.....	68
(3) Inflammation.....	68
(E) Rôle dans l'assemblage des particules virales et le bourgeonnement.....	69
III. Étude bio-informatique des protéines C des <i>Orthoparamyxovirinae</i> .....	70
(A) Quelle histoire évolutive les protéines C partagent-elles ?.....	70
(B) Une évidente absence de conservation de séquence au sein des protéines C.....	71
(C) Une organisation structurale commune à l'ensemble des protéines C.....	74

### CHAPITRE 3 : METHODES DE CARACTERISATION STRUCTURALE DES PROTEINES APPLIQUEES A CES PROJETS.....79

I. La chromatographie d'exclusion de taille couplée à la diffusion de la lumière à plusieurs angles et à la réfractométrie (SEC-MALLS-RI).....	79
II. Évaluation de la structure tertiaire d'un échantillon par dichroïsme circulaire (CD). ....	80
III. La Cristallographie aux rayons X.....	82
(A) Cristallogenèse.....	82
(B) La diffraction et la densité électronique.....	83
(C) Collecte et traitement des données.....	85
(D) La dérivation des protéines.....	87
(E) La diffusion anormale.....	87
(F) Le remplacement moléculaire.....	89
(G) Construction et affinement des modèles.....	90
IV. Diffusion aux rayons X à petits angles (SAXS).....	92
(A) Approximation de Guinier et rayon de giration.....	93
(B) Fonction de distribution des paires de distances, P(r).....	95
(C) Le graphique sans dimension de Kratky.....	96
(D) La méthode d'Optimisation d'Ensembles (EOM).....	97
V. La génération à haut débit de clones de protéines par la méthode ESPRIT.....	98
VI. La production d'anticorps à domaine unique spécifiques de la protéine C $\Delta$ 28 du virus Nipah.....	100
VII. Identification de partenaires cellulaires par la méthode du double-hybride en levure.....	101

### CHAPITRE 4 : ÉTUDE BIOPHYSIQUE ET STRUCTURALE DE LA PROTEINE C DU VIRUS DE LA TOUPAYE..... 105

I. Introduction.....	105
II. Article : Les protéines C des <i>Orthoparamyxovirinae</i> partagent une origine et une organisation structurale communes.....	105
III. Conclusions.....	132

### CHAPITRE 5 : ÉTUDE STRUCTURALE DE LA PHOSPHOPROTEINE DU VIRUS DE LA TOUPAYE..... 135

I. Introduction.....	135
II. Article en cours de préparation : Caractérisation structurale de domaines conservés de la phosphoprotéine du virus de la toupaye.....	135
III. Conclusions.....	141

### CHAPITRE 6 : ÉTUDE BIOPHYSIQUE ET STRUCTURALE DE LA PROTEINE C DU VIRUS NIPAH ..... 167

I. Stratégies élaborées pour étudier cette protéine difficile.....	167
(A) Approche préliminaire et analyse bio-informatique.....	167
(B) Approche de mutagenèse haut-débit.....	167
(C) Reconstitution de complexes pour stabiliser la protéine C du NiV.....	168
(D) Co-expression en complexes pour solubiliser la protéine C du NiV.....	169
II. Résultats.....	169
(A) Étude bio-informatique de la protéine C du virus Nipah.....	169
(B) Approche de mutagenèse haut-débit pour déterminer des constructions solubles de la protéine C du virus Nipah : le clone C $\Delta$ 28.....	171
(C) Identification de partenaires des protéines C <sub>FL</sub> et C $\Delta$ 28 par double-hybride.....	172
(D) Expression et purification des protéines C, P et SHCBP1.....	175
(1) Purification de la protéine pleine longueur C <sub>FL</sub> .....	175
(2) Purification de la protéine pleine longueur C $\Delta$ 45.....	175
(3) Purification de la protéine pleine longueur C $\Delta$ 91.....	176
(4) Purification des clones sélectionnés par la méthode ESPRIT.....	176

(5) Purification des protéines P du NiV, SHCBP1 .....	179
(6) Purification de la protéine C <sub>Δ28</sub> coexprimées avec VPS37B ou le complexe ESCRT-I .....	180
<b>(E) Caractérisation biophysique de la protéine C<sub>Δ28</sub> du virus Nipah .....</b>	<b>181</b>
(1) Évaluation du repliement tridimensionnel et de l'oligomérisation de C <sub>Δ28</sub> en solution en présence de faible quantité de détergent.....	181
(2) Essais de cristallographie de C <sub>Δ28</sub> en présence de détergent .....	184
<b>(F) Etude de l'interaction de C<sub>Δ28</sub> avec les protéines P du NiV et SHCBP1 .....</b>	<b>185</b>
(1) Étude du statut d'oligomérisation de la protéine P du NiV par SEC-MALLS-RI .....	185
(2) Étude du statut d'oligomérisation de la protéine SHCBP1 par SEC-MALLS-RI.....	186
(3) Caractérisation structurale de la protéine SHCBP1 par SEC-SAXS .....	187
(4) interactions entre les protéines C et P du virus Nipah.....	189
<b>(G) Production d'anticorps à domaine unique (SdAb) anti-C<sub>Δ28</sub> : vers de futurs cristaux pour la protéine C du virus Nipah ? .....</b>	<b>192</b>
<b>(H) Modèle AlphaFold2 de la protéine C du virus Nipah.....</b>	<b>193</b>
<b>III. Conclusions et perspectives.....</b>	<b>194</b>
<b>IV. Matériels appliqués à cette partie .....</b>	<b>195</b>
<b>(A) Prédiction de désordre.....</b>	<b>195</b>
<b>(B) Clonage.....</b>	<b>196</b>
(1) Clones d'expression de la protéine C du virus Nipah.....	196
(2) Clones d'expression des protéines P du virus Nipah et du facteur cellulaire humain SHCBP1.....	196
(3) Clones de co-expression de C <sub>Δ28</sub> avec les protéines du complexe ESCRT-I.....	196
<b>(C) Expression et purification .....</b>	<b>197</b>
(1) Expression et purification des protéines C de NiV en système bactérien .....	197
(2) Expression et purification de la protéine C <sub>Δ28</sub> de NiV en cellules d'insecte.....	198
(3) Expression et purification de la protéine P de NiV et SHCBP1.....	198
(4) Co-expression et purification de la protéine C <sub>Δ28</sub> avec les protéines du complexe ESCRT-I.....	198
 <b>CHAPITRE 7 : CONCLUSIONS ET PERSPECTIVES .....</b>	 <b>201</b>
<b>I. Comment expliquer l'insolubilité de la protéine C du virus Nipah ? .....</b>	<b>201</b>
<b>II. Caractérisation structurale des protéines P et C du virus de la toupeye .....</b>	<b>202</b>
<b>III. Quelles sont les liens de phylogénie partagés par les protéines C des <i>Orthoparamyxovirinae</i> ? .....</b>	<b>203</b>
<b>(A) La difficulté de retracer la phylogénie récente des <i>Orthoparamyxovirinae</i> .....</b>	<b>203</b>
<b>(B) Les scénarios évolutifs de l'émergence de la région codante de la protéine C chez les <i>Orthoparamyxovirinae</i> .....</b>	<b>204</b>
<b>(C) Les scénarios évolutifs a et b confirmés par nos résultats.....</b>	<b>206</b>
<b>ANNEXE 1.....</b>	<b>211</b>
<b>ANNEXE 2.....</b>	<b>212</b>
<b>ANNEXE 3.....</b>	<b>213</b>
<b>ANNEXE 4.....</b>	<b>214</b>
<b>Références bibliographiques .....</b>	<b>215</b>



# Liste des abréviations

Å : Angstrom  
aa : Acide aminé  
ADN : Acide DésoxyriboNucléique  
ARN : Acide RiboNucléique  
ARNm : ARN messenger  
C : protéine C  
CD : Dichroïsme Circulaire  
CLR : C-Type Lectin Receptor  
CTD : Domaine C-terminal  
Da/kDa : Dalton/kilodalton  
EDTA : Acide éthylènediaminetétraacétique  
*E.coli* : *Escherichia coli*  
ESRF : European Synchrotron Radiation Facility  
F : protéine de fusion  
G : glycoprotéine  
GE: gene end  
GS : gene start  
HTX : High-Throughput Crystallography / Cristallographie haut débit  
IBS : Institut de Biologie Structurale  
ICTV : International Committee on Taxonomy of Viruses  
IDR : Intrinsically Disordered Region / Région intrinsèquement désordonnée  
IFN : InterFeroN  
ISG : Interferon-Stimulated Gene  
kb: kilobase  
L: grande sous unité de l'ARN polymérase ARN dépendante virale  
M/mM : Molaire/miliMolaire  
M : protéine matrice  
MALLS : Multi Angle Laser Light Scattering / diffusion de lumière laser à angle multiple  
MM : Masse Moléculaire  
MoRE : Molecular Recognition Element / élément de reconnaissance moléculaire  
nm : nanomètre  
nt: nucléotide  
N<sub>CORE</sub> : domaine globulaire N-terminal de la nucléoprotéine chez les *Paramyxoviridae*  
N<sub>CTA</sub> : bras C-terminal de la nucléoprotéine (sous domaine C-terminal)  
NC : nucléocapside  
NCR : non coding regions / régions non codantes  
N<sub>NTA</sub> : bras N-terminal de la nucléoprotéine (sous domaine N-terminal)  
N<sub>TAIL</sub>: domaine désordonné C-terminal de la Nucléoprotéine  
NTD : domaine N-terminal  
N : nucléoprotéine  
OMS : Organisation Mondiale de la Santé  
ORF : Open Reading Frame / Cadre de lecture ouvert  
P : phosphoprotéine  
P<sub>MD</sub> : domaine d'oligomérisation de la phosphoprotéine  
P<sub>XD</sub> : domaine C-terminal de la phosphoprotéine, appelé domaine X  
PAGE : Poly-Acrylamide GEl  
PDB : Protein Data Bank  
PEG : polyethylene glycol  
pH: potentiel Hydrogène  
PSB : Platform for Structural Biology / Plateforme de Biologie Structurale  
RE : glutamate 50 mM / arginine 50mM



RI : Index de réfraction  
Rg : Rayon de giration  
Rh : Rayon hydrodynamique  
RLR : RIG-I Like Receptor  
RMN : Résonance Magnétique Nucléaire  
RNP : RiboNucléoProtéine  
RMSD : Root Mean Square Deviation  
SAD : single wavelength anomalous diffraction / diffraction anormale à une longueur d'onde  
SAXS : Small Angle X-ray Scattering / diffusion des rayons X aux petits angles  
SdAb : Single domain Antibody / Anticorps à domaine unique  
SEC: Size Exclusion Chromatography / chromatographie d'exclusion de taille  
T : température  
TLR : Toll-Like Receptor  
UV : ultraviolet  
vRNAP : viral RNA polymerase / ARN polymérase virale  
WB : Western Blot

# Liste des figures

- Figure 1 : Récurrence épidémiologie du virus Nipah
- Figure 2 : Localisation géographique des épidémies causées par les Henipavirus et aires de distribution de leurs hôtes réservoirs, les chauves-souris du genre *Pteropus*
- Figure 3 : Transmission du virus Nipah
- Figure 4 : Tableau Clinique d'une infection par le virus
- Figure 5 : La pathogénèse du virus Nipah chez l'humain
- Figure 6 : Découverte du virus de la toupaye et de son hôte *Tupaia belangeri*
- Figure 7 : Classification des *Orthoparamyxovirinae*
- Figure 8 : Organisation des génomes des *Orthoparamyxovirinae*
- Figure 9 : Représentation schématique d'une particule de paramyxovirus
- Figure 10 : Représentation schématique du cycle viral des *Orthoparamyxovirinae* sur le modèle de NiV
- Figure 10 : Représentation schématique d'une particule de paramyxovirus (sur le modèle du NiV).
- Figure 11 : Structure tridimensionnelle de la glycoprotéine G du virus Nipah seule et liée à son récepteur Éphrine B3 (PDB 3D12).
- Figure 12 : La protéine de fusion F du virus Nipah (PDB 5EVM).
- Figure 13 : La protéine de fusion F du virus Nipah (PDB 5EVM).
- Figure 14 : Représentation schématique du mécanisme de fusion dans le cas du virus Nipah.
- Figure 15 : La nucléocapside du virus Nipah.
- Figure 16 : La phosphoprotéine (P) du virus Nipah.
- Figure 17 : Représentation schématique de l'encapsidation de l'ARN viral par la nucléocapside et le complexe RdRp.
- Figure 18 : La protéine de matrice (M) de virus Nipah (PDB 7SKT et 7SKU).
- Figure 19 : Représentation schématique de la modulation de la réponse immunitaire IFN de type I suite à l'infection par un paramyxovirus.
- Figure 20 : Mécanismes d'expression des protéines non structurales V, W par édition co-transcriptionnelle depuis le gène P et C et par initiation de la traduction par balayage depuis les ARNm codant pour P/V et W pour le virus Nipah.
- Figure 21 : Structure de la protéine C de Sendai en complexe.
- Figure 22 : Fonctions des protéines C des *Orthoparamyxoviridae*.
- Figure 23 : Alignement des protéines C du groupe I.
- Figure 24 : Alignement des protéines C du groupe II.
- Figure 25 : Alignement des protéines C du groupe III.
- Figure 26 : Analyse des structures secondaires des protéines C des *Orthoparamyxovirinae*.
- Figure 27 : Dscore calculé à partir de la séquence en acides aminés des protéines C des *Orthoparamyxovirinae* répartis par genre.
- Figure 28 : Le dichroïsme circulaire.
- Figure 29 : Diagramme de phase bidimensionnel de la cristallisation d'une macromolécule en fonction de sa concentration et de la concentration de la solution en agent cristallisant.
- Figure 30 : La méthode de cristallisation par la diffusion de vapeur pour le dispositif de la goutte suspendue.
- Figure 31 : Paramètres de la maille cristalline.
- Figure 32 : Illustration de la condition de Bragg : réflexion sélective sur une famille de plan (hkl).
- Figure 33 : Le phasage expérimental par remplacement isomorphe et diffusion anormale.
- Figure 34 : Projection de la fonction de Patterson dérivée d'un cristal à trois atomes.
- Figure 35 : Représentation schématique d'une expérience de SAXS.
- Figure 36 : Courbe caractéristique des intensités de diffusion en fonction de la forme et de l'organisation des corps géométriques.
- Figure 37 : Graphiques de Guinier caractéristiques.
- Figure 38 : Courbes caractéristiques des fonctions de distribution des paires de distances.
- Figure 39 : Représentation schématique des courbes du graphique de Kratky.

Figure 40 : Représentation schématique de la stratégie EOM pour l'analyse de données SAXS dans le cas du Rg.

Figure 41 : ADN plasmidique utilisé par la méthode ESPRIT.

Figure 42 : Étapes de la méthode ESPRIT.

Figure 43 :

Figure 44 : Représentation schématique des constructions  $C_{FL}$ ,  $C_{H132}$  et  $C_{M150}$ .

Figure 45 : Analyse bio-informatique des structures secondaires et de désordre de la protéine C de NiV.

Figure 46 : Alignement multiple des séquences des protéines C du genre *Henipavirus*.

Figure 47 : Clones de la protéine C du NiV sélectionnés par la méthode ESPRIT.

Figure 48 : Résultats du double-hybride en levure pour  $C_{FL}$  et  $C_{\Delta 28}$ .

Figure 49 : Structure cristalline de la tête globulaire du complexe ESCRT-I.

Figure 50 : Construction et expression du complexe  $C_{\Delta 28}$ -VPS37B.

Figure 51 : Construction et expression du complexe TSG101-VPS37B-VPS28-MVB12- $C_{\Delta 28}$ .

Figure 52 : Résultats des purifications de la protéine  $C_{FL}$  de NiV en présence de différents détergents.

Figure 53 : Résultats des purifications de la protéine  $C_{\Delta 45}$  de NiV en Triton X-100 0.1%.

Figure 54 : Test d'expression des mutants  $C_{\Delta 28}$ ,  $C_{\Delta 18}$ ,  $C_{\Delta 12}$ ,  $C_{\Delta 35}$  et  $C_{\Delta 23}$ .

Figure 55 : Résultats des purifications des protéines  $C_{\Delta 28}$ ,  $C_{\Delta 23}$ ,  $C_{\Delta 35}$  et  $C_{\Delta 12}$  de NiV.

Figure 56 : Solubilisation de la protéine  $C_{\Delta 28}$  de NiV par addition de détergent.

Figure 57 : Présence de la protéine constitutive BirA lors de la solubilisation et purification des protéines  $C_{\Delta 28}$  et  $C_{\Delta 12}$  de NiV.

Figure 58 : Purification de la protéine  $C_{\Delta 28}$  exprimée en cellule d'insecte Hi5 en présence de 0.5% de Brij-58.

Figure 59 : Purification des protéines P et SHCBP1.

Figure 60 : Construction et expression du complexe  $C_{\Delta 28}$ -VPS37B.

Figure 61 : Construction et expression du complexe TSG101-VPS37B-VPS28-MVB12- $C_{\Delta 28}$ .

Figure 62 : Analyse SEC-MALLS de la protéine  $C_{\Delta 28}$ .

Figure 63 : Deuxième analyse de  $C_{\Delta 28}$  par SEC-MALLS.

Figure 64 : Spectre de dichroïsme circulaire de la protéine  $C_{\Delta 28}$ .

Figure 65 : Spectre 1D standard par RMN pour la protéine  $C_{\Delta 28}$ .

Figure 66 : Analyse SAXS de  $C_{\Delta 28}$ .

Figure 67 : Analyse SEC-MALLS-RI de la protéine P du virus Nipah.

Figure 68 : Étude de l'oligomérisation de la protéine SHCBP1 par SEC-MALLS.

Figure 69 : Étude de la protéine SHCBP1 par SEC-SAXS.

Figure 70 : Image de microscopie électronique à transmission de la protéine SHCBP1.

Figure 71 : Comparaison des profils d'élution des protéines SHCBP1 et P de NiV.

Figure 72 : Comparaison des profils d'élution des protéines SHCBP1 et  $C_{\Delta 28}$  de NiV.

Figure 73 : Comparaison des profils d'élution des protéines P et  $C_{\Delta 28}$  de NiV.

Figure 74 : Comparaison des profils d'élution des protéines P,  $C_{\Delta 28}$  et SHCBP1.

Figure 75 : Purification des SdAb CA18773 et CA18774.

Figure 76 : Modèle AlphaFold2 de la structure tridimensionnelle pour la protéine C du NiV.

Figure 77 : Superposition des domaines C-terminaux de la protéine C du virus Sendai et des modèles AlphaFold2 des protéines C du genre *Henipavirus*.

Figure 78 : Hypothèses évolutives de l'acquisition du cadre de lecture codant pour la protéine C chez les *Paramyxoviridae*.

Figure 79 : Les protéines C des *Orthoparamyxovirinae* sont homologues sur base de la modélisation.

Tableau 1 : MM moyenne des tronctions de la protéine C du NiV par la méthode ESPRIT.

Tableau 2 : Résultats des premiers essais de cristallisation de la protéine  $C_{\Delta 28}$

Tableau 3 : Présentation des constructions de C utilisées dans l'étude.

Tableau 4 : Abréviations des virus cités dans l'étude et numéros d'accessions des protéines virales citées.

Tableau 5 : Résumé des conditions d'expression et de purification testées pour les constructions de C de NiV de l'étude.





# Chapitre 1 : Introduction

## I. Généralités

### *(A) Les virus émergents*

Les virus sont des parasites intracellulaires obligatoires qui établissent des relations étroites avec les organismes qu'ils infectent. N'ayant pas la capacité de se reproduire eux-mêmes, ils sont capables de détourner les machineries cellulaires de leur cellule hôte à leur avantage. Le premier virus humain, l'agent de la fièvre jaune, fut découvert en 1901 par Walter Reed (Knipe and Howley, 2013). Les émergences virales ont par la suite profondément marqué l'histoire humaine et sont toujours d'actualité, comme l'a montré la pandémie mondiale de COVID-19 liée à l'émergence d'un coronavirus dont les cas continuent à apparaître à l'heure actuelle et qui a paralysé la planète pendant de nombreux mois. Plus de 1400 espèces de virus, parasites et microbes ont été énumérées comme pathogènes chez l'humain (Cleaveland, Laurenson and Taylor, 2001) et, parmi elles, plus de 60% sont d'origine zoonotique, c'est-à-dire provenant des animaux. Les pathogènes responsables de ces épidémies sont majoritairement des virus et des bactéries. Aujourd'hui, il est estimé que plus de 320 000 virus infectant des mammifères attendent encore d'être découverts (Anthony *et al.*, 2013).

Les maladies infectieuses émergentes sont des maladies « nouvelles » à plusieurs titres. Les mécanismes biologiques dont elles relèvent peuvent être nouveaux, comme leur rythme de propagation et leur extension géographique (exacerbés du fait de la mondialisation et de la rapidité des échanges internationaux). Les processus en jeu, à la fois biologiques, sociologiques, politiques et économiques, sont de plus en plus complexes. L'historien William McNeil dans son ouvrage *Plagues and People* (1976) a été le premier à poser l'hypothèse que les maladies infectieuses sont des agents majeurs, même si contingents, de l'histoire humaine et qu'il existe une relation positive entre le temps de domestication animale et le nombre de maladies que les humains partagent avec chaque espèce animale domestique.

Le virus Nipah (NiV), qui a été à l'origine de mes travaux de thèse, est l'un des premiers virus pour lesquels les liens entre écologie, économie et santé ont été théorisés. Il fait partie des nombreux virus dont l'émergence a permis de développer le concept *OneHealth*. Ce terme englobe une série d'initiatives regroupant de nombreuses collaborations entre divers secteurs et disciplines qui reconnaissent l'interconnexion entre la santé des humains, des animaux et de l'environnement. Il est reconnu, aujourd'hui, qu'il est primordial d'accroître les investissements internationaux de surveillance, de recherches et de développement dans le domaine de la santé afin, notamment, de mieux comprendre les interactions complexes qui existent entre les virus et leurs réservoirs naturels pour faciliter le développement de nouvelles thérapies virales et vaccins et l'anticipation des futures pandémies.

### *(B) Les Orthoparamyxovirus : un sujet d'inquiétude ?*

Les Orthoparamyxovirus sont étroitement reliés au concept *OneHealth*. Cette sous-famille de virus appartient à la plus large famille des Paramyxoviridae et l'ordre des Mononegavirales qui regroupe les virus hautement mutagènes à ARN monocaténaire de polarité négative et à génome non-segmenté. Nombre d'entre eux, comme le NiV ou le virus de la rougeole (MeV), sont des virus zoonotiques et ont, à plusieurs occasions infectés et causé de nombreux décès chez l'Homme. Les Orthoparamyxovirus infectent une grande variété d'animaux sauvages (chauves-souris, rongeurs, poissons, serpents, oiseaux,...) et domestiques (chèvres, chats, chiens, cochons, chevaux,...). Dans ce contexte, la surveillance accrue et systématique des virus hébergés par la faune sauvage, et plus particulièrement par les chauves-souris et les rongeurs, a été mise en place depuis deux décennies et est essentielle pour anticiper les futures émergences de ces virus zoonotiques notamment à l'échelle politique (Cunningham, Daszak and Wood, 2017; Wells *et al.*, 2022; Su *et al.*, 2023).

Certains de ces virus ont été capables de traverser la barrière de l'espèce très récemment. C'est le cas du virus Langya (LayV), étroitement lié au virus Nipah, identifié pour la première fois en 2018 dans les provinces de Shandong et d'Henan en Chine. Ce virus ne sera caractérisé formellement que le 4 août 2022 (Choudhary *et al.*, 2022). Depuis son émergence, il a infecté 35 personnes, majoritairement des fermiers, sans aucun décès. L'ARN viral de LayV a été détecté dans plus d'un quart de 262 toupayes, un petit mammifère arboricole (voir partie II.B), testées, 2% des chèvres domestiques et 5% des chiens, suggérant que la toupaye soit le réservoir naturel de ce virus. Bien que le risque de transmission interhumaine soit très faible, en raison de sa proximité phylogénétique avec le mortel virus Nipah, le virus Langya peut être aujourd'hui considéré comme un virus à potentiel pandémique et doit être fortement surveillé. Cet exemple met en valeurs l'importance de la compréhension des mécanismes à l'origine de la pathogénicité des virus émergents pour le développement d'outils efficaces de lutte contre les maladies infectieuses.

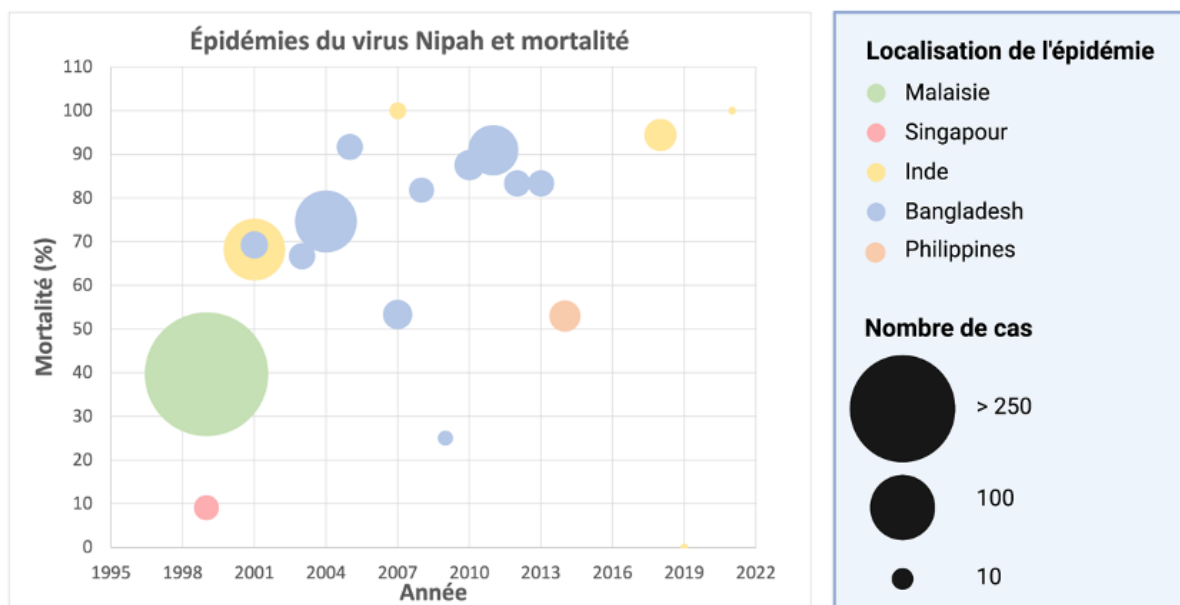
Au cours de ce travail, nous avons cherché à résoudre la structure et à identifier les fonctions de la protéine C, une protéine virale propre à la sous-famille des *Orthoparamyxovirinae*, dont le rôle clé dans la pathogénicité de ces virus en fait une cible prometteuse pour le développement de futures thérapies antivirales. La résolution de la structure de telle protéines permet d'ouvrir la voie au développement de futurs antiviraux et composés inhibiteurs, notamment grâce à l'évolution des outils bio-informatiques qui permettent le criblage de nombreux composés chimiques sur la base des structures. Cette approche a été récemment mise en application pour l'identification de nouveaux composés thérapeutiques ciblant la glycoprotéine G de surface du virus Hendra (Ahmad *et al.*, 2022). Nous avons choisi, pour cette étude, de nous concentrer d'une part, sur le virus Nipah, qui constitue une menace grandissante pour la santé publique, et d'autre part, sur le virus de la toupaye. Ce dernier n'a pas encore infecté l'Homme à l'heure actuelle mais partage son réservoir naturel avec le virus LayV et constitue donc un virus d'intérêt dans le cadre de l'approche *OneHealth*.

## II. Découvertes du virus Nipah et du virus de la toupaye

## (A) Le virus Nipah

### (1) Émergence du virus Nipah

Le virus Nipah appartient à l'ordre des *Mononegavirales*, la famille des *Paramyxoviridae* et la sous-famille des *Orthoparamyxovirinae*. Cette sous-famille comprend huit genres dont les *Henipavirus* qui incluent le virus Nipah (Kuhn *et al.*, 2019).



**Figure 1 : Réurrence épidémiologie du virus Nipah.** Graphique à bulles illustrant les taux de mortalité des multiples épidémies du virus Nipah de 1999 à 2021. Chaque couleur de cercle correspond à un pays où s'est produit des épidémies. La circonférence des cercles est proportionnelle aux taux de mortalité des épidémies (Figure adaptée de (Alam, 2022)).

La première flambée épidémique de cas d'encéphalites attribuée au virus Nipah se produit en septembre 1998 près de la ville d'Ipoh en Malaisie (**Figure 1**). Une étude épidémiologique menée par le CDC (Centers for Disease Control and Prevention, USA) observe que la majorité des personnes atteintes par cette maladie a été en contact direct avec des élevages porcins et présente des antigènes similaires à ceux de personnes ayant été contaminées par le virus Hendra (HeV), un paramyxovirus apparu chez les chevaux en 1994 en Australie (Lam and Chua, 2002; Ksiazek, Rota and Rollin, 2011; Alam, 2022). Le virus Nipah est finalement isolé en mars 1999 et nommé selon le village éponyme de Kampung Sungai Nipah d'où était originaire le patient chez lequel l'échantillon ayant servi à la caractérisation avait été prélevé (Lam and Chua, 2002). On recense plus de 265 cas humains dont 105 décès. La propagation de ce virus ne sera contrôlée que par l'abattage systématique des animaux dans les élevages porcins contaminés (plus d'1.1 millions de porcs seront abattus) (Lam and Chua, 2002), par la formation des éleveurs concernés à différentes pratiques d'hygiène et par la généralisation du port d'équipements individuels de protection. Le porc aurait agi comme un hôte intermédiaire d'amplification virale (Nor, Gan and Ong, 2000; Lam and Chua, 2002). Le développement de l'élevage intensif sans contrôle des pratiques aurait facilité l'émergence du virus. Le cas du virus Nipah est remarquable et souligne les liens étroits entre les virus zoonotiques et l'émergence (Kessler *et al.*, 2018; Eby



*et al.*, 2022). On estime les pertes économiques liées à cette première épidémie à près de 100 millions de dollars US. La mise en place de mesures de prévention sera concomitante au développement d'un second foyer d'infection à Singapour causé par l'importation de porcs vivants contaminés. Cette seconde épidémie causera 11 contaminations humaines et 1 mort mais ne touchera que des personnes travaillant dans des abattoirs (Skowron *et al.*, 2022).

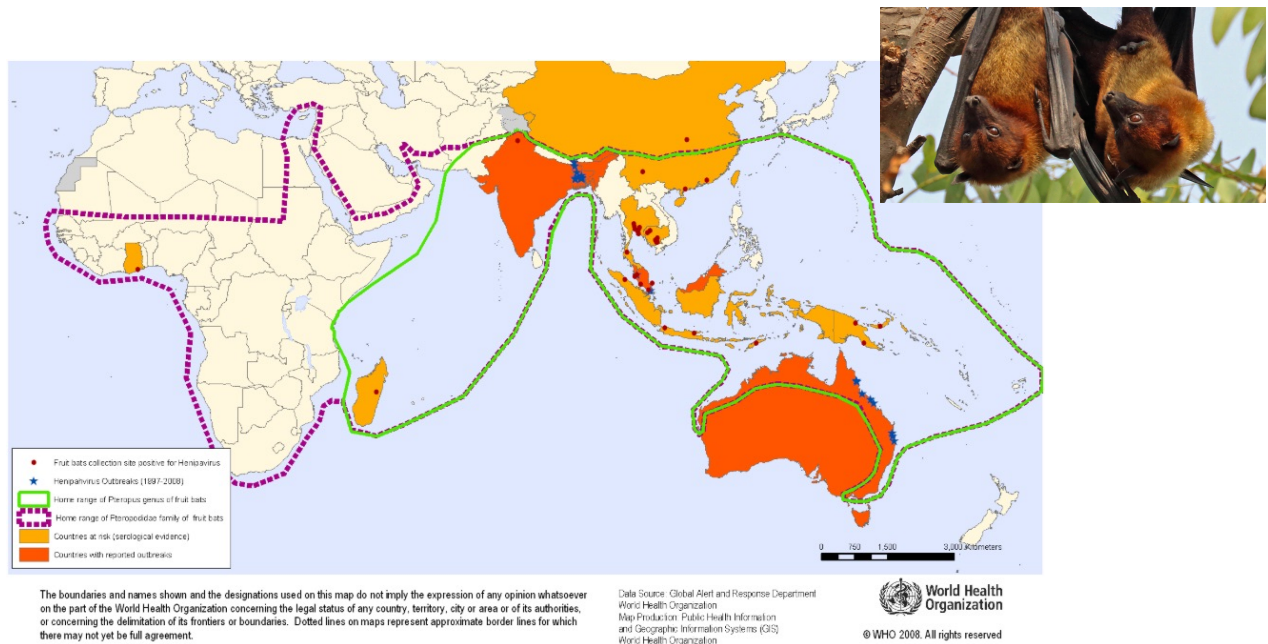
Le virus Nipah réapparaît ensuite en 2001 en Inde dans la province de Siliguri sous la forme d'une souche plus virulente qui sera par la suite appelée NiV-BD et s'avère identique à la souche détectée au Bangladesh (Chadha *et al.*, 2006; Singh *et al.*, 2019). Cette épidémie est remarquable car elle présente les premiers cas de transmission interhumaine du virus. Par la suite, 4 épidémies se sont succédées de 2001 à 2004 au Bangladesh pour lesquelles le taux de mortalité du NiV atteindra près de 80% sans identification d'hôtes intermédiaires (Nikolay *et al.*, 2019). Cependant, plusieurs études relèvent la mention de la présence récurrente à proximité des foyers d'infection virale de chauves-souris frugivores du genre *Pteropus* toutes communément appelées roussettes ou renards volants. Ces études suggèrent l'existence d'une transmission virale par contact direct ou indirect avec les agriculteurs (Luby *et al.*, 2006; Daszak *et al.*, 2011). Il est notamment observé que ces chauves-souris viennent s'abreuver et se nourrir au niveau des cultures de palmiers, contaminant ainsi directement les fruits ou la sève en cours de collecte par leurs fèces, urine et salive et, en conséquence, les agriculteurs consommant des dattes ou du lait de palme sous la forme de jus frais (Luby *et al.*, 2009). Le virus sera isolé par la suite chez ces chauves-souris en Malaisie, au Cambodge et en Thaïlande. L'hypothèse de la contamination des fruits et de la sève sera finalement confirmée lors de l'épidémie de Tangail de décembre 2004 à janvier 2005 pour laquelle le seul point commun entre les malades était la présence de chauves-souris *Pteropus giganteus* à proximité des lieux de collecte.

Depuis 2001, plusieurs épidémies sporadiques ont été recensées au Bangladesh, en Inde et aux Philippines. La plus récente a eu lieu en Inde en 2021 et a provoqué le décès d'un enfant dans une zone rurale de l'état du Kerala où des épidémies avaient déjà eu lieu en 2018 et 2019 (Thakur, Thakur and Ratho, 2022) (**Figure 1**). L'épidémie de 2018 est classée en troisième position en terme d'intensité pour le virus Nipah avec un taux de mortalité de 91% (Arunkumar *et al.*, 2019). 23 cas de personnes contaminées par le NiV ont été recensées, dont 21 morts. Les foyers malaysiens représentent 43% des cas de contaminations au NiV et le Bangladesh et l'Inde représentent respectivement 42% et 15% du total de cas reportés dans le monde (Soman Pillai, Krishna and Valiya Veetil, 2020).

## (2) Les espèces réservoirs

Les chauves-souris de la famille des *Pteropodidae* appartenant au genre *Pteropus*, sont les réservoirs naturels principaux du virus Nipah (Hayman *et al.*, 2008; Daszak *et al.*, 2011). Ces chauves-souris sont répandues dans l'hémisphère sud et leur séropositivité au NiV a été reportée en Asie du Sud-Est, notamment au Cambodge (Reynes *et al.*, 2005), en Thaïlande (Wacharapluesadee *et al.*, 2005) et à Madagascar (Iehlé *et al.*, 2007). Des chauves-souris n'appartenant pas à la famille des *Pteropodidae* séropositives aux Henipavirus ont également

été recensées en Afrique de l'Ouest (Hayman *et al.*, 2008) (**Figure 2**). Leur large répartition géographique ainsi que le développement du commerce et du tourisme international vers ces pays à risques représentent une menace potentielle d'émergence du virus Nipah à l'échelle mondiale (Luby, 2013).

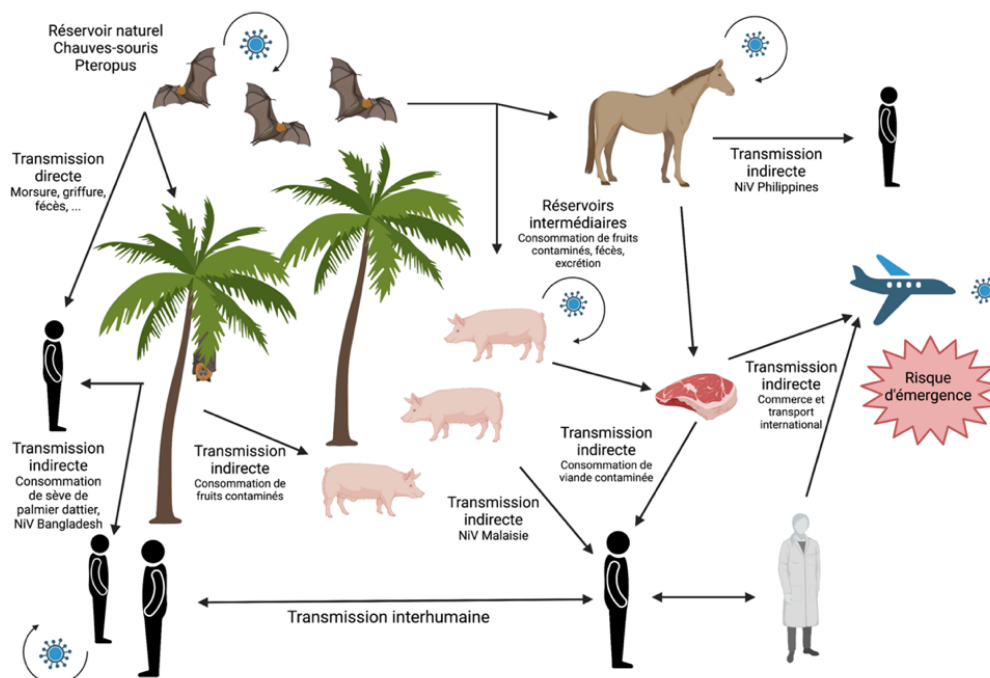


**Figure 2 : Localisation géographique des épidémies causées par les *Henipavirus* et aires de distribution de leurs hôtes réservoirs, les chauves-souris du genre *Pteropus*.** En vert est détournée l'aire de répartition des chauves-souris du genre *Pteropus* (photo en haut à droite, J. Sharp Photography). En violet est détournée l'aire de répartition des chauves-souris de la famille *Pteropodidae*. En orange figurent les pays à risque (évidences sérologiques) et en rouge les pays où des épidémies ont déjà eu lieu. La localisation précise de ces épidémies est indiquée par des étoiles bleues. Les points rouges représentent les zones où ont été collectées des chauves-souris porteuses d'*Henipavirus* (WHO, 2008).

L'infection par le virus Nipah ne provoque que peu ou pas de symptômes chez ces chauves-souris en raison probablement d'une longue coexistence et coévolution des henipavirus avec leurs réservoirs naturels (Daszak *et al.*, 2011). Le système immunitaire des chauves-souris *Pteropodidae* est remarquable par sa capacité à restreindre la réplication virale précoce par une réponse interféron (IFN) et par une forte expression de cytokines pro-inflammatoires unique à la famille (Baker, Schountz and Wang, 2013). La particularité de ce mode d'action chez les chauves-souris est qu'il est associé avec plusieurs gènes de réponse aux interférons (ISG) impliqués notamment dans des voies non-inflammatoires. L'élévation de la réponse interféron n'est donc pas associée à une inflammation chronique comme c'est le cas chez les humains (Daszak *et al.*, 2011; Lawrence and Escudero-Pérez, 2022). Les chauves-souris sont donc capables de réagir très rapidement à une infection en limitant mais sans éliminer les pathogènes viraux. Cela entraîne la persistance des virus chez ces espèces pour de longues périodes (Wellehan Jr. *et al.*, 2009; Baker, Schountz and Wang, 2013; Schountz *et al.*, 2017; Singh *et al.*, 2019).

### (3) Les hôtes intermédiaires

Le NiV est transmis par les chauves-souris à de nombreuses espèces de mammifères tels que les porcs, les chiens ou les chevaux, qui ont été les hôtes intermédiaires « amplificateurs » dans plusieurs épidémies. Expérimentalement, il a été montré que le NiV pouvait également infecter les chats, les cochons d'Inde, les hamsters, les furets et plusieurs espèces de singes ainsi que des bovins et des chèvres (Vigant and Lee, 2011; Lawrence and Escudero-Pérez, 2022). Cette capacité à infecter de nombreuses espèces de mammifères différentes repose notamment sur l'utilisation des récepteurs cellulaires Éphrine B2 et B3 (Eph B2/B3) pour l'entrée du virus dans les cellules. Ces récepteurs sont conservés chez de nombreuses espèces de mammifères (Xu *et al.*, 2008; Xu, Broder and Nikolov, 2012). À la différence des chauves-souris qui semblent insensibles aux infections par ce virus, ces hôtes intermédiaires peuvent présenter des symptômes caractéristiques d'infections respiratoires et de maladies neurologiques (WHO, OIE). L'uniformisation de la diversité génétique des porcs dans l'élevage intensif a permis en Malaisie une augmentation de la probabilité d'émergence du virus vers l'homme en favorisant l'amplification virale du NiV (Soman Pillai, Krishna and Valiya Veetil, 2020) (**Figure 3**).



**Figure 3 : Transmission du virus Nipah.** Les chauves-souris frugivores du genre *Pteropus* constituent un réservoir naturel du virus Nipah. Les chauves-souris frugivores porteuses du NiV se nourrissent des fruits du palmier-dattier et contaminent le sol, les fruits et la sève en cours de collecte. Le virus peut survivre dans des solutions riches en sucre, c'est-à-dire dans la pulpe des fruits ou le lait de palme fabriqué à partir de la sève. Le virus est transmis à l'homme par la consommation de ces produits. Les fruits contaminés sont également consommés par les porcs et d'autres animaux qui ont un rôle d'hôte intermédiaire et amplificateur. La viande de porc infectée par le NiV est exportée vers d'autres régions. La consommation de viande de porc infectée peut être une autre source d'infection pour l'homme. Un contact étroit avec un humain affecté par le NiV peut entraîner la propagation du NiV à d'autres personnes (schéma inspiré de Singh *et al.*, 2019).

#### (4) Les différentes souches

Les analyses phylogénétiques ont confirmé la circulation de deux clades majeurs de Nipah : NiV-BD et NiV-MY. Ces analyses sont basées sur l'étude des séquences complètes des gènes N et G. Les isolats de Nipah de Malaisie et du Cambodge forment le clade NiV-MY tandis que les isolats du Bangladesh et d'Inde font partie du clade NiV-BD (Chan *et al.*, 2001; Harcourt *et al.*, 2005; Singh *et al.*, 2019).

La variabilité génétique entre les différentes souches de NiV est liée au mode de transmission du virus. En effet, durant l'épidémie de NiV-MY de 1998-1999 en Malaisie, les analyses phylogénétiques ont montré que les séquences des isolats viraux de porcs étaient presque identiques à ceux des patients, suggérant que le virus avait été amplifié par son hôte intermédiaire avant d'être transmis à l'homme. En revanche, l'émergence de NiV-BD au Bangladesh a montré, en comparaison, une relative hétérogénéité des isolats viraux suggérant de multiples transmissions du virus à l'homme par la chauve-souris par le biais de la consommation de sève de palmier-dattier contaminée (Chan *et al.*, 2001; Luby *et al.*, 2009; Rahman and Chakraborty, 2012). NiV-BD diffère de NiV-MY pour 8.2% de sa séquence (91,8% d'identité de séquence) et possède 6 nucléotides de plus en 5'-UTR du gène codant pour la protéine F. La souche NiV-BD est la plus mortelle pour l'homme (90% de mortalité contre 40% pour la malaisienne) (Harcourt *et al.*, 2005). La transmission interhumaine n'a été démontrée que dans le cas de la souche NiV-BD.

#### (5) Tableau clinique

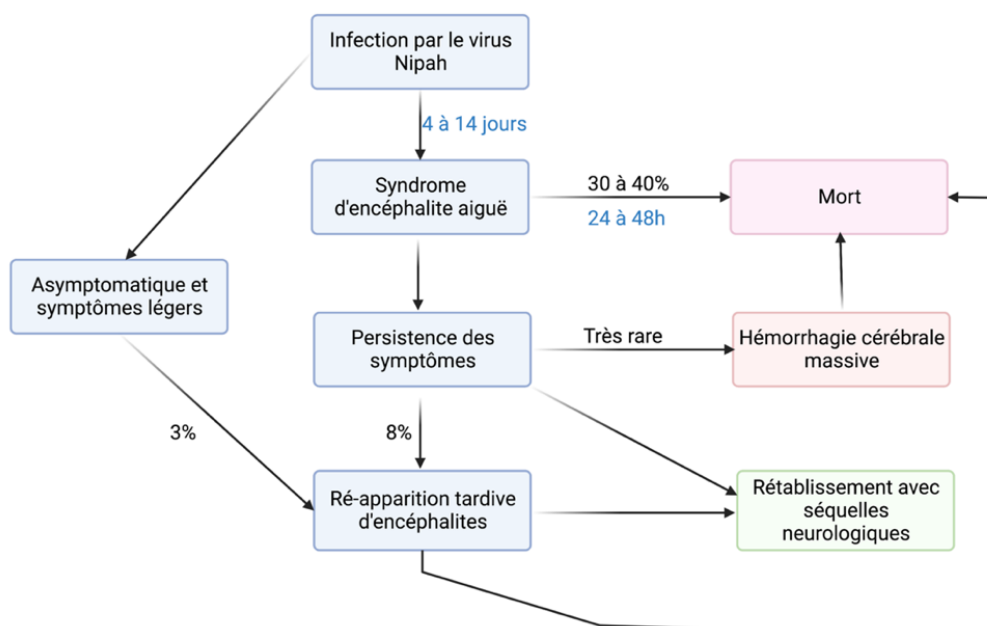
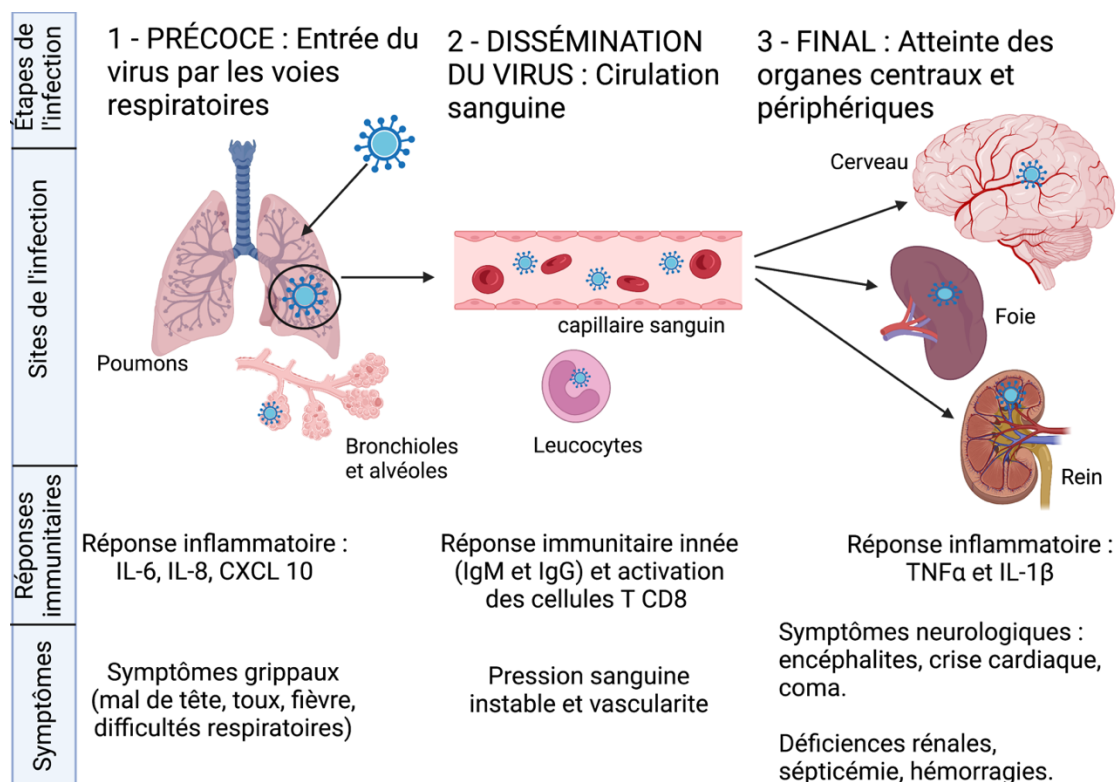


Figure 4 : Tableau clinique d'une infection par le virus Nipah (reproduit de (Wong *et al.*, 2002))

L'entrée naturelle du virus dans l'organisme se fait majoritairement par voie aérienne. La durée d'incubation des patients infectés par le NiV s'étale sur une période de 4 jours à 2 mois, avec un décès qui survient en moyenne 9,5 jours après l'infection (Luby *et al.*, 2006,

2009; Vigant and Lee, 2011; Singh *et al.*, 2019). Les patients présentent des symptômes grippaux avec de la fièvre, des céphalées intenses, un étourdissement, des vomissements et douleurs musculaires. De nombreux patients souffrent de perte de conscience, des somnolences et présentent des signes de dysfonctionnements cérébraux diverses. Près de 40% des patients développent une atteinte respiratoire (Nikolay *et al.*, 2019; Alam, 2022). L'une des caractéristiques du virus Nipah est le développement d'une encéphalite tardive, parfois plusieurs mois ou années après l'encéphalite initiale. Le délai de réactivation le plus long connu est de 11 ans (Ang, Lim and Wang, 2018). Ces rechutes sont associées à des phénomènes de paralysie des muscles oculomoteurs, un nystagmus<sup>1</sup> ou un affaissement du visage. De nombreux patients sont également touchés par des troubles psychiatriques incluant des dépressions, des changements de personnalité et des déficits de l'attention ou de la mémoire visuelle et verbale caractéristiques de l'affect neurologique causé suite à l'infection par le NiV. Ces patients présentent dans la majorité des cas un état de fatigue sévère persistant pouvant parfois s'accompagner de troubles moteurs. Ces symptômes seraient des conséquences directes des encéphalites aiguës (**Figure 4**).



**Figure 5 : La pathogénèse du virus Nipah chez l'humain.** La pathogénèse du virus Nipah chez l'humain peut être divisée en 3 grandes étapes : (1) l'étape précoce correspondant à l'entrée du virus par les voies respiratoires, (2) l'étape de dissémination du virus par la circulation sanguine et (3) l'étape finale correspondant à l'atteinte des organes centraux et périphériques. Chacune de ces étapes est caractérisée par un site d'infection ou de dissémination du virus (décrit sous chaque étape), une réponse immunitaire spécifique (détaillée à chaque étape) et l'apparition de symptômes divers (détaillés à chaque étape) (Figure adaptée de (Lawrence and Escudero-Pérez, 2022). Figure créée avec Biorender.

<sup>1</sup> Le nystagmus est un **mouvement rythmique des yeux, qui est involontaire**. Il peut être congénital ou acquis.

Les informations collectées suite aux différentes épidémies permettent d'établir la chronologie suivante (**Figure 5**). L'infection démarrerait par une réplication primaire du virus dans les cellules dendritiques de l'épithélium respiratoire ou digestif qui, une fois infectées, migreraient alors vers les organes lymphoïdes *via* les tissus lymphatiques. On peut alors observer une nécrose de ces tissus et l'apparition d'une charge virale dans le sang du patient. La propagation du virus se poursuit ensuite dans les cellules endothéliales, deuxième site de réplication virale. On observe alors la formation de *syncytia* de cellules endothéliales, signe clinique caractéristique d'une infection par le NiV (Wong *et al.*, 2002), menant au développement d'une vascularite, de thromboses, de nécroses et de l'infection de cellules parenchymateuses de divers tissus. De nombreux organes présentent des dommages liés à l'infection par le NiV, cependant les vaisseaux sanguins du système nerveux central sont les plus affectés (Wong *et al.*, 2002). Les lésions occasionnées dans ce tissu vasculaire permettraient au virus d'atteindre les neurones où l'on trouve la plus forte charge d'antigènes (Mathieu *et al.*, 2011).

## (6) Traitements

Il n'existe à l'heure actuelle aucun traitement ni vaccin contre le virus Nipah. Le virus Nipah est aujourd'hui classé dans la catégorie d'agent pathogène de niveau 4 sur l'échelle de dangerosité microbiologique de l'OMS, suite au constat de l'existence d'une transmission interhumaine et d'une forte mortalité causée par ce virus. Il doit être manipulé dans des laboratoires de haute sécurité. L'essentiel des options disponibles et efficaces pour lutter contre la propagation du NiV aujourd'hui consiste en des mesures préventives et de traitement des symptômes.

- **Mesures préventives**

Les épidémies qui ont eu lieu en Malaisie auront permis la mise en place de deux mesures préventives principales : (i) la formation des fermiers et des populations concernées aux règles de bonne pratique d'hygiène, (ii) la limitation de la surpopulation du bétail dans les zones fréquentées par les chauves-souris et de la superposition des vergers avec les zones d'élevages afin de prévenir de futurs contacts du bétail avec les chauves-souris contaminées (Chua, 2003). En plus de ces mesures, l'élevage de porc a été interdit dans sept provinces malaisiennes sur treize.

Dans d'autres pays, l'utilisation de systèmes de protection (nattes de bambous), empêchant les chauves-souris de venir se nourrir sur les sites de saignée des palmiers, a été réinstaurée (Nahar *et al.*, 2010; Salah Uddin Khan *et al.*, 2010).

Aujourd'hui, plusieurs études révèlent les relations fortes entre la nette diminution dans l'abondance et la disponibilité des ressources alimentaires des chauves-souris à la suite des modifications des usages du territoire et les changements climatiques sur les événements de transmission des henipavirus (Kessler *et al.*, 2018; Eby *et al.*, 2022). Les chauves-souris auraient ainsi envahi les vergers qui accueillait des élevages porcins, conduisant à la première épidémie en Malaisie. Ces études proposent des modèles prédictifs qui pourraient permettre d'anticiper ces événements en adaptant notre agriculture.

- **Vaccins**

Bien qu'il n'existe aucun vaccin autorisé contre le NiV, plusieurs études suggèrent que le développement d'un vaccin efficace est possible. Plusieurs plateformes vaccinales ont été développées basées soit, sur l'utilisation de la glycoprotéine G du virus Hendra (Bossart *et al.*, 2012), soit sur des vecteurs viraux recombinants du virus de la stomatite vésiculeuse (Mire *et al.*, 2013; Prescott *et al.*, 2015), du virus Measles (Yoneda *et al.*, 2013) et du virus de la rage exprimant la protéine G du virus Nipah. Ces plateformes ont montré qu'il était possible de cibler l'une des protéines de la membrane extérieure de NiV telle que la glycoprotéine G ou la protéine de fusion F pour stimuler une réponse immunitaire protectrice chez différents modèles animaux incluant les hamsters, les chats, les furets, les cochons et les singes verts Africains (Satterfield, Geisbert and Mire, 2016).

Plusieurs vaccins candidats se sont avérés capables de fournir une protection complète contre le NiV pour des modèles animaux (Broder, Weir and Reid, 2016; Satterfield, Dawes and Milligan, 2016). Ces plateformes vaccinales seraient capables de déclencher une réponse anticorps neutralisante après une injection à la fois contre les souches NiV-BD et NiV-MY du virus. De nouvelles approches ont été envisagées récemment et consistent à développer des vaccins exprimant de multiples sous-unités du virus Nipah afin notamment d'anticiper les futures mutations, très élevées pour les virus à ARN (Ojha *et al.*, 2019).

D'autres plateformes utilisent des particules pseudo-virales du NiV non infectieuses, sans génome viral, obtenues par l'assemblage spontané des protéines de la capsid virale du NiV (Kalodimou *et al.*, 2019). Ces particules permettent le déclenchement d'une réaction immunitaire très forte. Elles ont permis l'immunisation d'hamsters Syriens qui ont développé de forts titres d'anticorps neutralisants (Walpita *et al.*, 2017).

Enfin, des études développent également des vaccins dérivés d'antigènes peptidiques conformationnels synthétiques permettant de stimuler après administration une réponse immunitaire aux virus Nipah et Hendra (Saha *et al.*, 2017).

Cependant, les épidémies de NiV sont à l'heure actuelle relativement bien contrôlées par les mesures de prévention mises en place, ce qui constitue un frein au financement d'un vaccin efficace par les compagnies pharmaceutiques. C'est pourquoi, en 2017, un effort collaboratif des gouvernements et des compagnies pharmaceutiques, appelé la coalition pour les innovations en matière de préparation aux épidémies (CEPI), a vu le jour (Brende *et al.*, 2017). Elle a pour objectif de financer le développement de vaccins abordables, efficaces et sûrs contre diverses maladies à potentiel pandémique telles que celles causées par le virus Nipah. Grâce aux avancées dans le domaine des vaccins à ARN messager (ARNm) depuis la pandémie de COVID-19, un essai clinique de phase 1 chez l'homme d'un vaccin ARN messager (mRNA-1215) contre le virus Nipah a été lancé en juillet 2022 par le centre de recherche vaccinal du NIAID et la compagnie Moderna. Cet essai permettra d'évaluer la sûreté, la tolérance et la capacité de ce vaccin à générer une réponse immunitaire contre le

virus Nipah chez 40 adultes humains de 18 à 60 ans par l'administration de deux doses de vaccins.

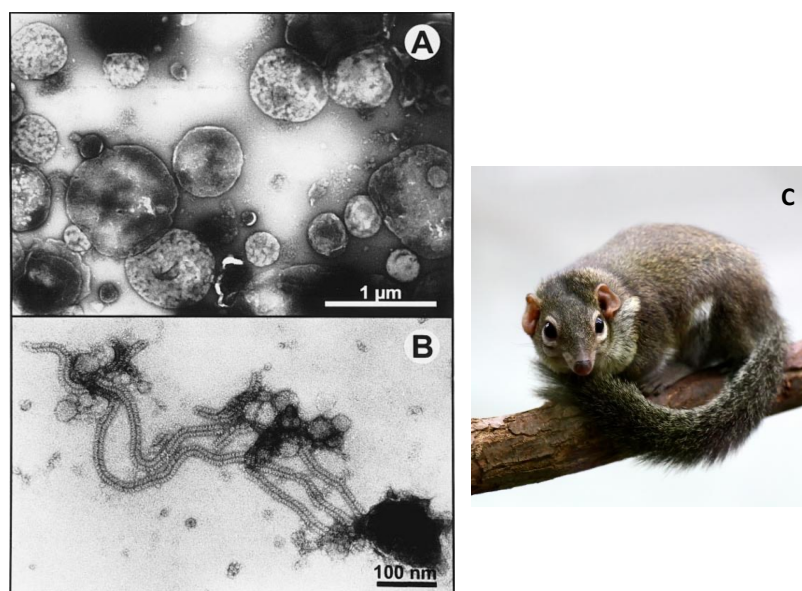
- **Antiviraux**

Aucun traitement spécifique au virus Nipah n'a été autorisé à l'heure actuelle. Malgré des essais de traitements basés sur le développement d'anticorps monoclonaux ciblant spécifiquement la glycoprotéine G, lesquels se sont révélés prometteurs *in vivo* (Bossart *et al.*, 2009), le prix de tels traitements ne justifie encore une fois pas leur développement au regard du faible nombre de cas annuels enregistrés. Il serait plus souhaitable de développer des antiviraux à larges spectres ou des traitements basés sur des petits ARN interférents. Dans ce contexte, certains composés ont montré leur efficacité dans des modèles animaux tels que le favipiravir (T-705) (Dawes *et al.*, 2018). Les peptides thérapeutiques comme antiviraux capables soit d'inhiber la fusion de l'enveloppe du virus avec celle de la cellule hôte (Porotto *et al.*, 2010; Mathieu *et al.*, 2015), soit de prévenir la formation du complexe entre la nucléoprotéine non-assemblée et la phosphoprotéine (Yabukarski *et al.*, 2014), sont toujours à l'étude et pourraient représenter une innovation intéressante pour le traitement de virus émergents de cette famille.

## **(B) Le virus de la toupaye**

### **(1) Émergence**

Le virus de la toupaye (TupV) a été isolé et caractérisé en 1999 en Asie du Sud-Est à partir de cellules rénales dégénérantes d'un petit mammifère arboricole (*Tupaia belangeri*), apparemment sain, capturé près de Bangkok (Tidona *et al.*, 1999) (**Figure 6**). L'habitat naturel de la toupaye est la forêt tropicale d'Asie du Sud-Est où elle se nourrit de fruits, d'insectes et de petits vertébrés.



**Figure 6 : Découverte du virus de la toupaye et de son hôte *Tupaia belangeri*. (A) Micrographies électroniques à transmission de particules du virus de la toupaye colorées négativement et (B) de**



nucléocapsides libres purifiées à partir de surnageants de cultures cellulaires infectées par centrifugation en gradient de densité de saccharose (Tidona et al 1999). (C) La toupaye (*Tupaia Belangerii*), espèce réservoir du TupV.

Depuis plusieurs années, la toupaye attire une attention croissante en tant qu'animal modèle pour diverses maladies humaines et pour le développement de traitements thérapeutiques. En effet, les toupayes sont proches des primates dont elles ont divergé il y a environ 85 millions d'années (Kumar and Hedges, 1998) et sont aujourd'hui classifiées dans la famille des *Tupaiaidae* dans l'ordre des *Scandentia*. Elles représentent, de ce fait, un modèle prometteur pour la recherche biomédicale étant bien plus proches des humains que n'importe quel autre modèle animal non-primate étudié en laboratoire, mais qui présente les avantages d'être un petit animal et de se reproduire rapidement. Elles sont notamment sensibles aux virus de l'hépatite B et C comme le chimpanzé (Tsukiyama-Kohara and Kohara, 2014).

Le TupV appartient à la sous-famille des *Orthoparamyxovirinae* et au genre *Narmovirus*. L'analyse de la séquence nucléotidique des gènes codant pour les protéines P et N du TupV a révélé que le virus est plus étroitement lié aux genres *Morbillivirus* et *Henipavirus* (Springfeld et al., 2005). Il est intéressant de noter que cette taxonomie est aujourd'hui remise en question, notamment par la découverte d'un nombre certain de nouveaux virus dans la sous-famille des *Orthoparamyxovirinae*. La proposition, par certains auteurs, de la création d'un nouveau genre pour le virus de la toupaye est aujourd'hui mise en avant grâce au développement de méthodes modernes d'étude de la phylogénie des virus (Wells et al., 2022).

## (2) Tropisme

Le tropisme de ce virus est restreint aux cellules de la toupaye et aucune réactivité croisée avec d'autre virus de la famille n'a été détectée (Tidona et al., 1999). Il a été suspecté que ce tropisme résultait de la présence d'un site de clivage au sein de la glycoprotéine de fusion F de la surface du virus qui serait spécifique d'une protéase présente uniquement chez la toupaye et dont l'action permettrait l'activation de la protéine de fusion et l'entrée du virus dans les cellules de la toupaye. Cependant, il a été montré par la suite que cette protéase était ubiquiste et que le facteur limitant du tropisme du TupV serait lié à l'absence de récepteur spécifique permettant l'entrée du virus dans les cellules d'autres mammifères (Springfeld et al., 2005). Ce récepteur n'est pas connu à l'heure actuelle.

## (3) Utilisation médicale

L'absence d'anticorps reconnaissant le TupV chez les humains, combinée à l'incapacité de réplication efficace de ce virus dans des cellules humaines, a encouragé le développement de vecteurs lentiviraux pseudotypés pour la thérapie génique (Springfeld et al., 2005). Plusieurs études ont cherché à démontrer l'intérêt des glycoprotéines de surface du TupV dans le développement de vecteurs oncolytiques efficaces dans l'échappement à la réponse des anticorps neutralisants. Notamment, des virus chimères portant des glycoprotéines du virus de la rougeole (MeV) et du TupV ont été développés. Bien que la

fusion soit conservée et que des particules virales recombinantes puissent être secourues, la réplication de ces virus chimériques a été fortement compromise (Hudacek, Navaratnarajah and Cattaneo, 2013). En réponse, des études ont cherché à développer des vecteurs TupV recombinants capables de cibler des récepteurs de surface spécifiques en incorporant, au sein du brin d'ARN antigénomique codant pour le génome du TupV, une unité de transcription supplémentaire codant pour des anticorps scFv (Single Chain Variable Fragment) de spécificité choisie. Cette méthode permet, à condition qu'un scFV approprié soit disponible, de générer des particules virales du TupV de tropisme spécifique (Engeland *et al.*, 2017).

### III. Aperçu des spécificités du génome, des protéines et de la machinerie de réplication/transcription des *Orthoparamyxovirinae*, le cas du virus Nipah

#### (A) Classification

La sous-famille des *Orthoparamyxovirinae* appartient à la famille des *Paramyxoviridae* et à l'ordre des *Mononegavirales*, un ordre de virus regroupant les virus à ARN monocaténaire de polarité négative (groupe V de la classification Baltimore) et à génome non segmenté (Figure 7). La sous-famille est composée de huit genres : les *Aquaparamyxovirus*, les *Ferlavirus*, les *Henipavirus*, les *Jeilongvirus*, les *Morbillivirus*, les *Narmovirus*, les *Respirovirus* et les *Salemvirus* (Kuhn *et al.*, 2019).

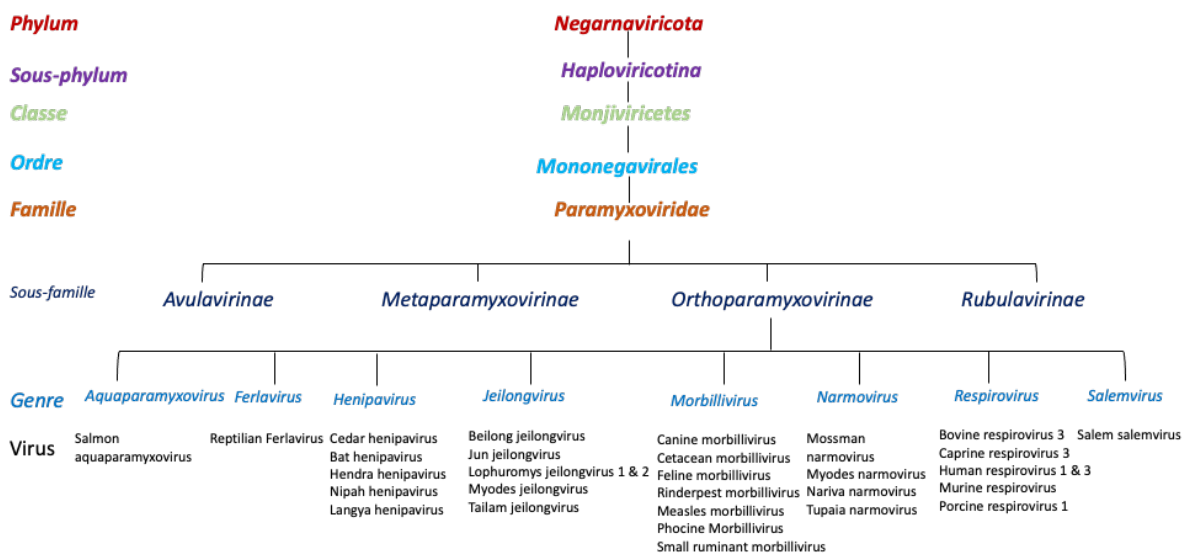
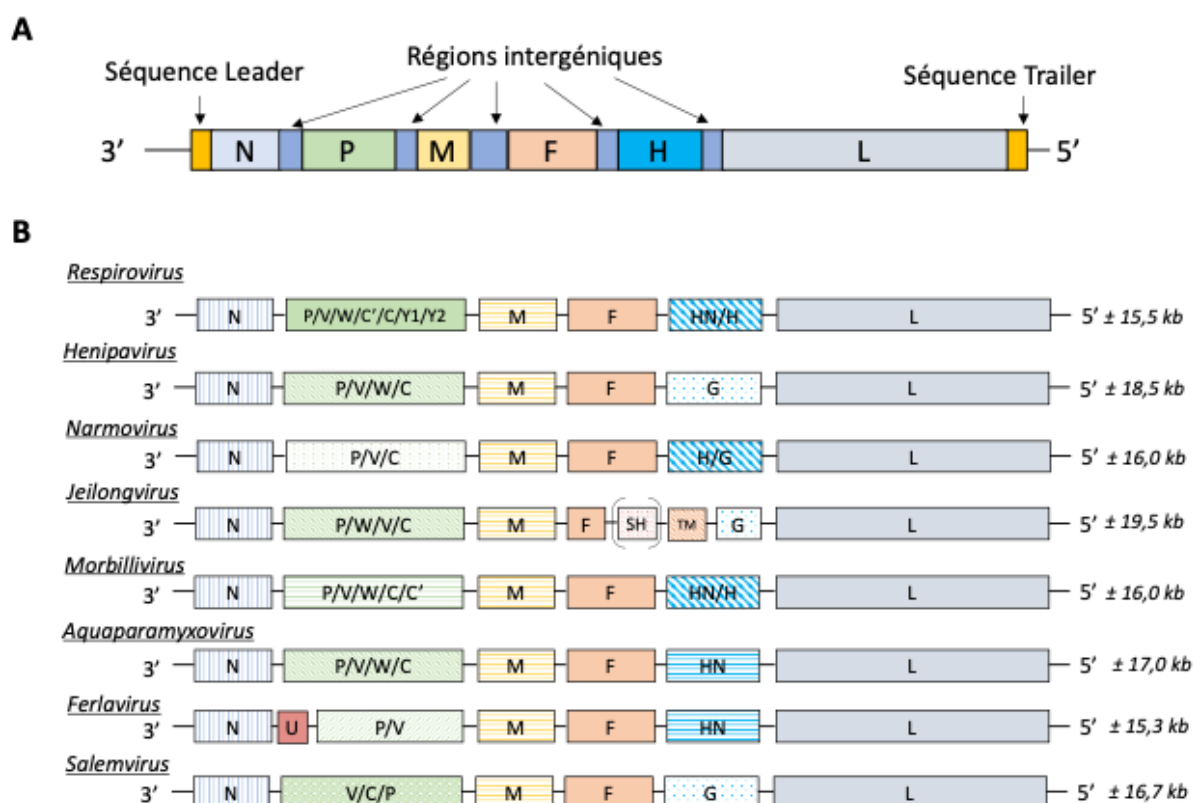


Figure 7 : Classification des *Orthoparamyxovirinae*. (d'après (Kuhn *et al.*, 2019)).

Très variables dans leur spécificité d'hôtes (mammifères, poissons, reptiles, oiseaux), leurs modes de transmission et les maladies dont ils sont responsables, les *Orthoparamyxovirinae* partagent avec les autres *Paramyxoviridae* de nombreuses similitudes. Parmi elles, on peut noter une organisation similaire de leur génome et de leur virion, une stratégie similaire de réplication de l'ARN viral ainsi que des mécanismes d'échappement à l'immunité innée similaires déclenchés suite à l'infection virale (Knipe and Howley, 2013).

## (B) Le génome viral

La taille des génomes des virus des *Orthoparamyxovirinae* est comprise entre 15 kb et 20 kb. Ils comprennent entre 6 et 8 gènes et sont codants sur plus de 90% de leur longueur.



**Figure 8 : Organisation des génomes des *Orthoparamyxovirinae*.** (A) Exemple du génome viral du virus Nipah où chaque rectangle correspond à un gène codant pour une protéine virale (N-P-M-F-H-L). Les rectangles jaunes situés aux extrémités 3' et 5' du génome correspondent, aux séquences 3' leader et 5' trailer. Les rectangles bleus correspondent aux régions intergéniques non traduites contenant les séquences Gene End (GE) et Gene Start (GS) à l'origine de l'interruption de la transcription par la polymérase virale entre chaque gène. (B) Description des génomes viraux et de leur spécificité au sein des *Orthoparamyxovirinae*. Chaque rectangle correspond à un gène codant pour une protéine virale : (N) Nucléoprotéine, (P) Phosphoprotéine, (V/W/C'/Y1/Y2) Protéines non structurales encodées au sein de la séquence du gène P, (M) Matrice, (F) Fusion, (SH) Petite protéine hydrophobe (Small Hydrophobic), (TM) Protéine TransMembranaire, (HN/H/G) Glycoprotéine, (L) grande sous-unité de la polymérase ARN-ARN dépendante. Les différents motifs permettent d'identifier les similarités entre les genres pour une même cassette de transcription. La longueur moyenne des génomes est indiquée à droite.

Le génome code, d'une part, pour des protéines structurales<sup>1</sup> (N-P-M-F-G/H/HN-L) requises pour former la particule virale et le complexe de réplication, mais également pour

<sup>1</sup> Les protéines structurales ont été initialement définies comme les protéines présentes dans les virions. Des études récentes ont montré que diverses protéines, y compris des protéines virales non-structurales, se retrouvent en faible quantité dans les virions de certains virus (Vera-Velasco *et al.*, 2018), remettant en question la dénomination de protéine structurale. Dans le manuscrit, j'utilise le terme « protéine structurale » pour dénommer les protéines présentes en grand nombre dans les particules virales et le terme « protéine non-structurale » pour celles qui sont produites en grand nombre dans les cellules infectées et peuvent se retrouver parfois dans les particules.

des protéines non-structurales (V-W-C<sup>1</sup>) dont les fonctions, toujours à l'étude chez de nombreux virus, semblent majoritairement tournées vers l'échappement viral. L'agencement de six des gènes codant pour les protéines structurales le long du génome est commun aux *Mononegavirales* et comprend dans le sens 3' vers 5' : la nucléoprotéine (N), la phosphoprotéine (P), la protéine de matrice (M), la glycoprotéine de fusion (F), une glycoprotéine d'attachement (G/H ou HN) et l'ARN polymérase ARN-dépendante (RdRp) aussi appelée protéine L pour « large » car codant pour la grande sous-unité de la polymérase fonctionnelle (**Figure 8**). Chaque région codante est encadrée par des régions intergéniques et chaque gène est encadré par des séquences START et STOP permettant la transcription d'ARN messagers individuels codant pour chaque protéine à partir d'un promoteur unique situé à l'extrémité 3' du génome.

L'ARN génomique présente également une complémentarité entre ses extrémités 3' et 5'. L'organisation du génome est linéaire de sens 3' vers 5' avec deux régions non-traduites en 3' (3'-UTR) et en 5' (5'-UTR) : une région en 3' d'approximativement 50 nucléotides dénommée « leader » et une région en 5' de 50 à 161 nucléotides dénommée « trailer ».

### ***(C) Le Cycle viral***

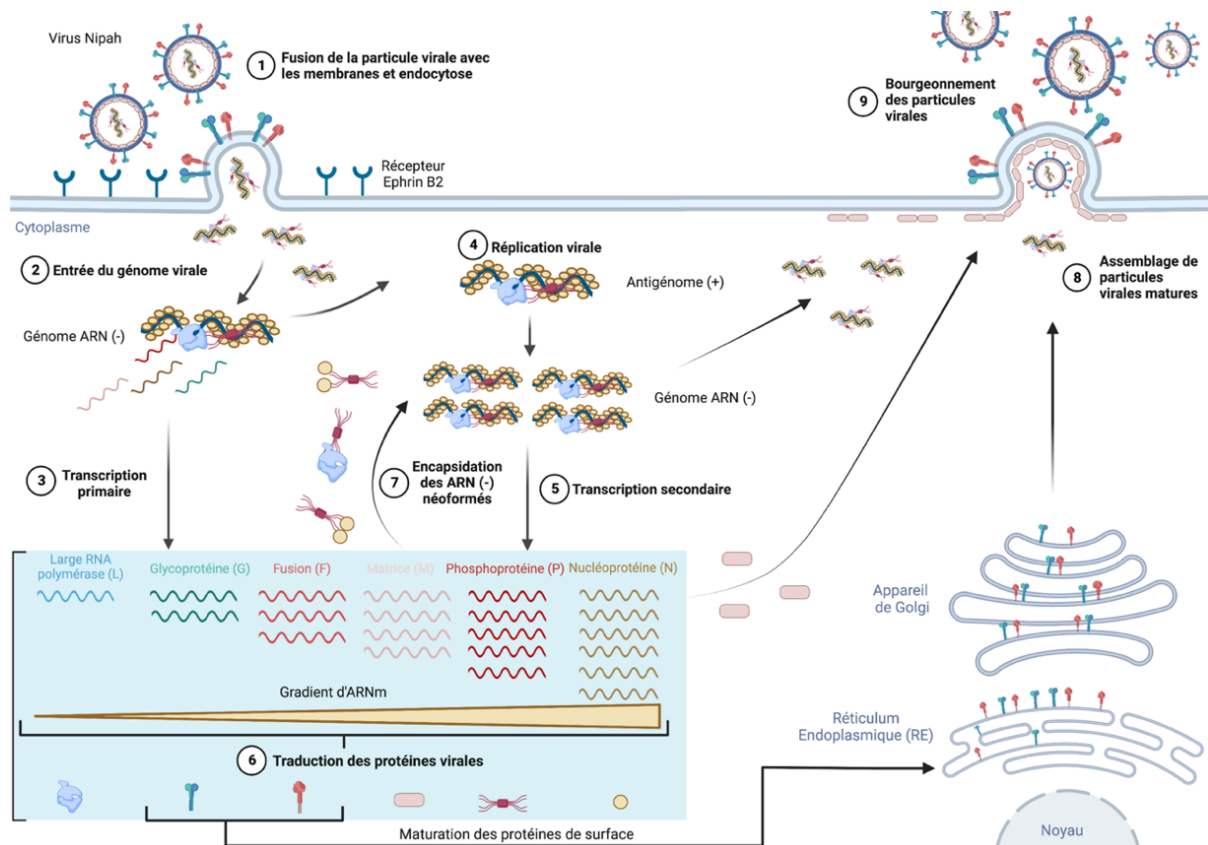
Le cycle de réplication viral peut être divisé en quatre parties présentées en **Figure 9** :

- (1) La phase extracellulaire de dissémination des particules virales et leur transmission à de nouveaux individus ou cellules non-infectées.
- (2) L'attachement du virus à son récepteur, la pénétration dans la cellule hôte et le transport jusqu'au site de réplication.
- (3) La multiplication du virus (production d'ARN messagers et de protéines virales et la réplication du génome).
- (4) L'assemblage de nouvelles particules virales et leur bourgeonnement à partir de la cellule hôte.

Pour les *Orthoparamyxovirinae*, la réplication s'effectue dans le cytoplasme de la cellule hôte. Tout au long du cycle, les virus interagissent avec des machineries et des protéines cellulaires pour réguler le cycle viral et déjouer le système immunitaire.

---

<sup>1</sup>Certains *Orthoparamyxovirinae* expriment des ensembles de protéines C en utilisant plusieurs codons initiateurs alternatifs en amont ou en aval pour l'initiation de la traduction. Ceci a entraîné de multiples dénominations pour ces protéines : C, C', Y1 ou Y2. Nous reviendrons sur ces notions dans le chapitre 2.



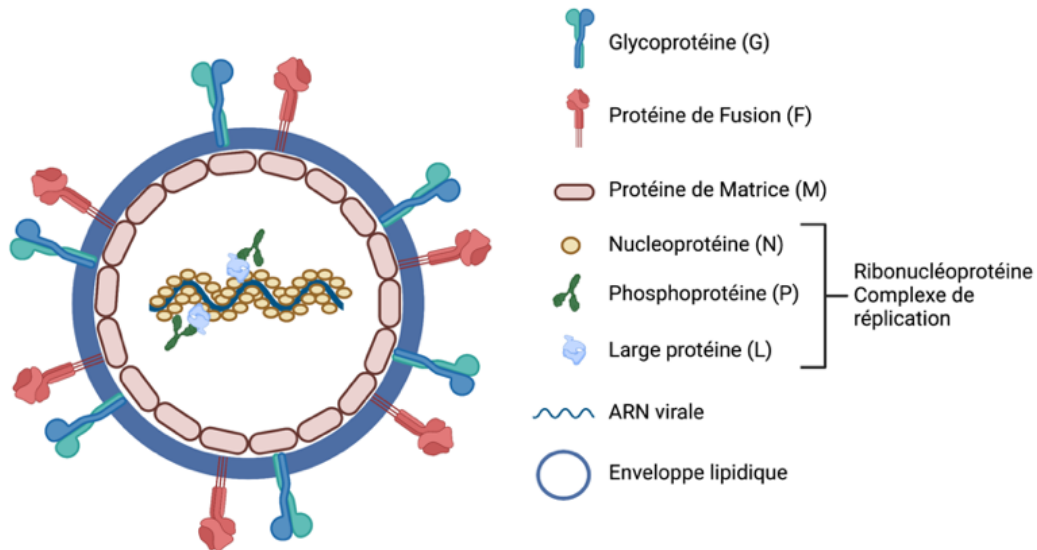
**Figure 9 : Représentation schématique du cycle viral des *Orthoparamyxovirinae* sur le modèle de NiV.** (1) Après attachement aux récepteurs cellulaires Éphrine B2/B3, la membrane de la particule virale fusionne avec la membrane cellulaire, (2) libérant le complexe de transcription dans le cytoplasme. (3) Les ARNm viraux sont produits par la polymérase virale directement à partir du génome et ces ARNm viraux sont traduits par les ribosomes de la cellule (transcription primaire). (4) La polymérase initie la production d'antigénomes à partir du génome, puis de nouvelles copies du génome à partir d'anti-génomes. (5) Les protéines néosynthétisées contribuent à la production de nouveaux ARNm à partir de l'anti-génome synthétisé (transcription secondaire) et (6) de nouvelles protéines virales sont traduites depuis ces ARNm. (7) Les copies d'ARN génomique sont encapsidées par les protéines N synthétisées. (8) La protéine de matrice M permet l'assemblage à la membrane des différents composants de la particule virale avant (9) le bourgeonnement et la libération des nouveaux virions.

### (1) Dissémination et transmission : structure du virion

Les *Orthoparamyxovirinae* sont majoritairement des virus respiratoires qui se transmettent soit par les contacts directs entre patients infectés (féces, sang, urine,...) mais également par excrétion de gouttelettes liquides contenant des particules virales.

Les virions des *Orthoparamyxovirinae* sont pléiomorphes, grossièrement sphériques (150-350nm) (**Figure 10**). Ils sont organisés en trois couches concentriques. La couche externe est une bicouche lipidique dérivée de la membrane plasmique de la cellule hôte présentant à sa surface une rangée de glycoprotéines formant des spicules visibles dans les images de microscopie électronique. La couche interne est formée par les nucléocapsides composées du génome d'ARN encapsidé par les nucléoprotéines N polymérisées en association avec la phosphoprotéine P et la protéine L qui constituent le complexe de réplication ou ribonucléoprotéine. Entre l'enveloppe et le cœur de la particule se trouve une couche

intermédiaire constituée d'un réseau tridimensionnel formés par les protéines de matrice M assurant l'architecture du virion et la cohésion entre les couches internes et externes.



**Figure 10 : Représentation schématique d'une particule de paramyxovirus (sur le modèle du NiV).** Ce virus enveloppé est entouré de glycoprotéines virales (F et G) à la surface de la particule. La protéine de matrice (M) recouvre le feuillet interne de la bicouche lipidique et forme des interactions avec les queues cytoplasmiques des glycoprotéines. Le complexe ribonucléoprotéique (RNP) est constitué du génome viral encapsidé par la nucléoprotéine (N) associée à la polymérase virale ARN-ARN dépendante (RdRP) constituée de la polymérase virale (L) et de la phosphoprotéine (P).

## (2) Entrée et tropisme

Pour débiter le cycle viral, le virus s'introduit dans la cellule-hôte et libère le génome viral et la machinerie de réplication de l'ARN viral. La fixation du virus à la membrane cellulaire est assurée par la fixation via les spicules des glycoprotéines tétramériques HN, H ou G à un récepteur cellulaire. La nature du récepteur cellulaire du virus détermine le tropisme cellulaire. Cette interaction active alors la fusion de l'enveloppe virale avec la membrane plasmique de la cellule hôte par l'intermédiaire de la protéine F (Chang and Dutch, 2012).

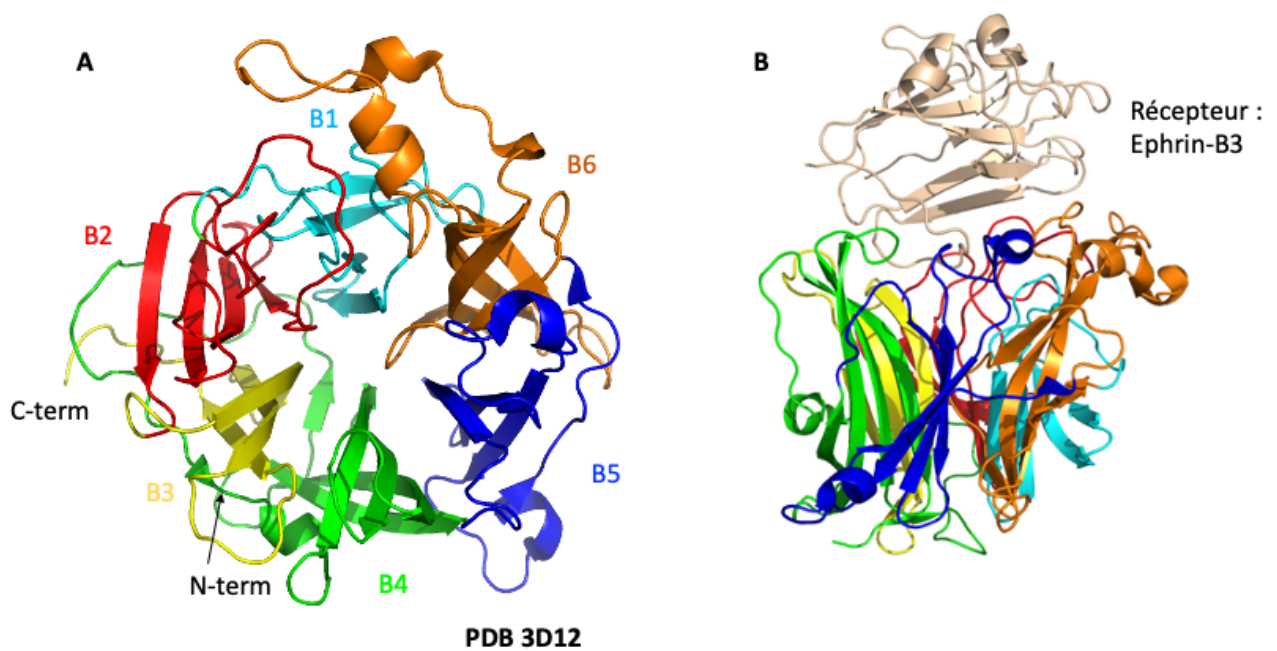
### (2.a) Les glycoprotéines G, H ou HN : liaison au récepteur cellulaire

Tous les *Orthoparamyxovirinae* possèdent une glycoprotéine responsable de la reconnaissance et de l'attachement du virus à la cellule hôte. Toutes ces protéines sont des glycoprotéines membranaires de type II. Elles présentent soit une activité hémagglutinine et neuraminidase (HN), soit seulement une activité d'hémagglutination (H) ou aucune activité (G) enzymatique.

Dans le cas des henipavirus NiV et HeV (**Figure 9**), les récepteurs cellulaires de la protéine G sont les protéines membranaires Éphrine B2/B3 (Negrete *et al.*, 2005; Xu, Broder and Nikolov, 2012), omniprésents chez les vertébrés et exprimés à la surface d'une variété de

cellules (neurones, cellules endothéliales et épithéliales, fibres de muscle lisse de la *tunica media*) présentes dans des organes divers (poumon, rate, rein, cerveau, placenta) expliquant le large spectre d'hôtes sensibles à ces virus et la dissémination systémique du virus chez les personnes infectées (Mathieu *et al.*, 2011).

Les protéines G forment des tétramères (Bowden *et al.*, 2008; Bowden, Jones and Stuart, 2011) dont chaque monomère est composé d'une queue cytoplasmique N-terminale, d'un domaine transmembranaire, d'une tige et d'une tête globulaire extracellulaire C-terminale (**Figure 11A**). La tête globulaire de la protéine G est responsable de la liaison au récepteur (**Figure 11B**). Elle présente une structure en  $\beta$ -propeller formée par l'arrangement de 6 feuillets  $\beta$  en forme de lame autour d'un axe central selon une configuration toroïdale. Chaque feuillet est composé de quatre brins  $\beta$  antiparallèles. Cette structure est stabilisée par des ponts disulfures connectant les extrémités N- et C-terminales du  $\beta$ -propeller ainsi qu'au sein de chacun des 6 feuillets.



**Figure 11 : Structure tridimensionnelle de la glycoprotéine G du virus Nipah seule et liée à son récepteur Éphrine B3 (PDB 3D12).** (A) Représentation de la structure tridimensionnelle, vue de face, du monomère de glycoprotéine G de NiV. Chaque  $\beta$ -propeller est représenté par une couleur et numéroté de B1 à B6. (B) Représentation de la structure tridimensionnelle, vue de côté, du complexe formé par la glycoprotéine G de NiV et son récepteur cellulaire (Xu *et al.*, 2008).

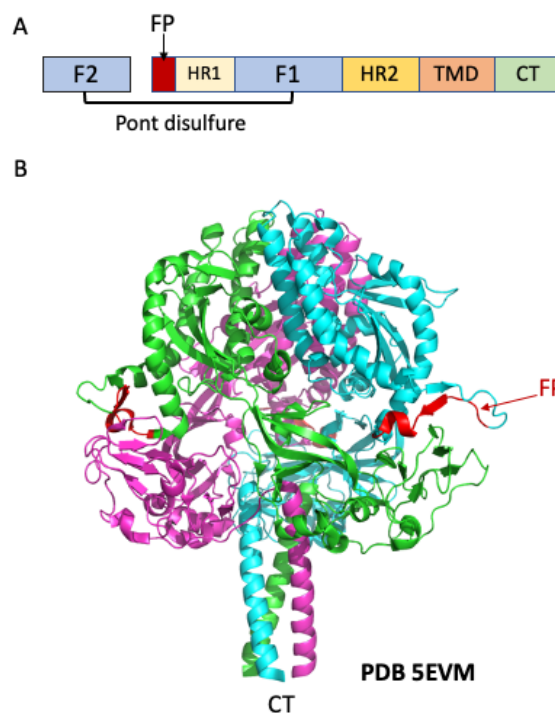
## (2.b) La protéine F

La protéine de fusion F est, pour tous les paramyxovirus, une glycoprotéine transmembranaire de type I responsable de la fusion entre la membrane virale et la membrane cellulaire.

Ces protéines sont produites sous la forme d'homo-trimères de précurseurs F0 inactifs métastables. Cependant, pour la plupart des paramyxovirus, la protéine F est présente à la

surface cellulaire sous une forme biologiquement active suite au clivage par une protéase générant une structure stable présentant deux sous-unités F1 et F2 liées par un pont disulfure.

Dans le cas des *Henipavirus*, la protéine F est présente à la surface du virion sous une forme F0 et une endocytose ultérieure est nécessaire, après interaction de la protéine G avec son récepteur cellulaire, pour l'activation par clivage par l'endoprotéase cathepsine L dans des endosomes acides avant un recyclage vers la surface cellulaire (Diederich, Thiel and Maisner, 2008 ; Pager *et al.*, 2006). Chaque hétérodimère F1+F2 est composé d'une courte queue cytoplasmique C-terminale impliquée dans l'interaction avec les protéines de matrice M lors du bourgeonnement, d'un domaine transmembranaire inséré dans la membrane virale, de deux régions de répétition en heptade (HR1 et HR2) et d'un peptide de fusion hydrophobe replié sur la structure (**Figure 13A&B**).

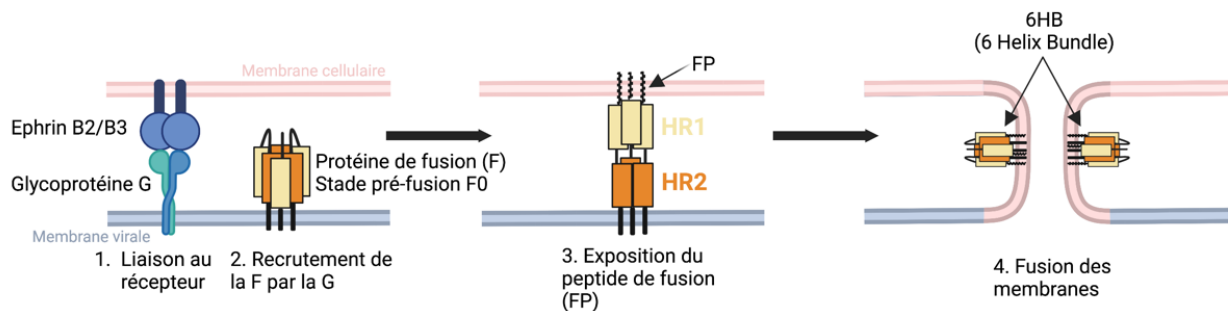


**Figure 13 : La protéine de fusion F du virus Nipah (PDB 5EVM).** (A) Représentation schématique des domaines formant la protéine F du NiV après clivage par la cathepsine L des deux domaines F1 et F2 (rectangles bleus) reliés par un pont disulfure. Les autres rectangles représentent, du N-terminal au C-terminal, le peptide de fusion (FP), les domaines de répétitions en heptades HR1 et HR2, le domaine transmembranaire (TMD), la queue cytoplasmique (CT). (B) Structure tridimensionnelle du trimère formé par la protéine F du NiV dans sa conformation de pré-fusion après clivage par la cathepsine. Les monomères sont colorés en cyan, magenta et vert. Les peptides de fusion (FP) des trois monomères sont colorés en rouge (Figure adaptée de El Najjar, Schmitt and Dutch, 2014).

Bien que différents modèles soient proposés en fonction des glycoprotéines d'attachement exprimé par les virus, l'entrée dans la cellule, pour tous les paramyxovirus, repose sur un mécanisme central commun (Lee and Ataman, 2011; Navaratnarajah *et al.*, 2020) : la fixation des glycoprotéines sur leurs récepteurs cellulaires respectifs déclenche une cascade de changements conformationnels dans la protéine F entraînant la fusion des membranes (**résumé dans le cas du NiV sur la Figure 14**) (Aguilar *et al.*, 2010 ; Navaratnarajah *et al.*, 2020). Sans stimulus, la protéine F se trouve sous une forme dite de



pré-fusion constituée d'un assemblage de trimères d'hétérodimères métastables matures F1 + F2 où les deux domaines de répétitions en heptade HR1 et HR2 sont associés (**Figure 13B**). Après activation, les domaines HR1 et HR2 peuvent se déployer pour former une structure de « pré-épingle à cheveux » permettant l'exposition du peptide de fusion hydrophobe FP qui vient s'enchâsser dans la membrane de la cellule hôte. Par la suite, les domaines HR1 et HR2 s'associent en un ensemble stable de 6 hélices (Helix Bundle 6HB) permettant le rapprochement et la fusion des membranes puis l'entrée de la nucléocapside virale dans la cellule.



**Figure 14 : Représentation schématique du mécanisme de fusion dans le cas du virus Nipah.** L'interaction de la protéine G avec les récepteurs cellulaires Éphrine B2/B3 (étape 1) entraîne l'activation de la protéine F au stade de pré-fusion (étape 2) par changement de conformation permettant aux 2 domaines HR1 et HR2 de se déplier exposant le peptide de fusion FP (étape 3). Le peptide s'ancre alors dans la membrane cellulaire. Par souci de simplicité, une seule protéine F est représentée sur les deux premiers encarts (gauche – étape 1 à 3) de ce schéma mais de nombreuses protéines sont concentrées au niveau de la zone de fusion. Les protéines F subissent alors un nouveau changement de conformation vers une forme stable de 6 hélices appelée « Helix Bundle » qui permet la fusion des membranes (4) (adaptée de Aguilar et al., 2010).

Un second mécanisme est observé pour la propagation des paramyxovirus et consiste en la fusion entre des cellules exprimant à leur surface les glycoprotéines d'attachement et des cellules exprimant le récepteur spécifique à ces glycoprotéines. Ce processus entraîne la formation de cellules géantes multinucléées appelées syncytia.

### (3) La multiplication du virus : le complexe de réplication/transcription

#### (3.a) La protéine L

La protéine L constitue la grande sous-unité de la polymérase virale (L pour Large). C'est un composant essentiel de la machinerie de réplication virale. Elle est systématiquement encodée par le gène le plus distant du promoteur et est donc invariablement synthétisée en plus faible proportion. Elle est majoritairement colocalisée avec la protéine P sur les nucléocapsides (Knipe and Howley, 2013). Les protéines L sont généralement composées de plus de 2 000 aa (250kDa).

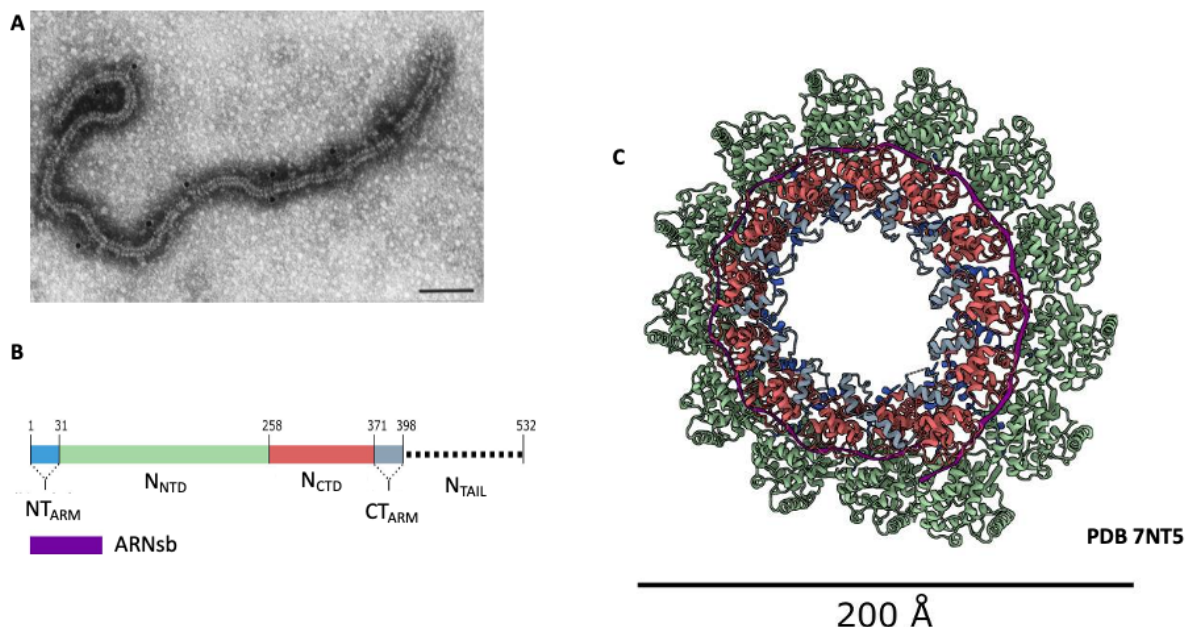
La protéine L est une protéine multienzymatique qui possède les différentes activités catalytiques nécessaires à la synthèse d'ARNm viraux (polymérisation des nucléotides,

synthèse de la coiffe 5', méthylation et polyadénylation 3' des ARNm). La protéine L est aussi responsable de la réplication virale de l'ARN génomique et antigénomique.

### (3.b) La nucléoprotéine ou protéine N

La protéine N est la première protéine transcrite pour presque tous les paramyxovirus. Les protéines N sont capables d'encapsider des molécules d'ARN génomique et antigénomique en s'associant en homopolymères pour former un complexe ribonucléoprotéique (RNP) N-ARN hélicoïdal appelé nucléocapside (NC) (**Figure 15A**). Les NCs servent de matrice à la polymérase virale pour la transcription et la réplication. La structure à haute résolution de la nucléocapside du virus Nipah et les mécanismes de son assemblage ont été résolus récemment par cryo-microscopie électronique (CryoEM) à une résolution de 3,5 Å (**Figure 15C&D**) (Ker *et al.*, 2021).

Il a été montré *in vitro*, dans le cas du virus Sendai (SeV), que cet assemblage en NC permet la protection contre l'action des ribonucléases cellulaires (Lynch and Kolakofsky, 1978; Iseni *et al.*, 2002). Il prévient également la formation de structures secondaires par l'ARN viral ainsi que l'hybridation des brins complémentaires 3' et 5' d'ARN viraux ou de génome/antigénome au cours de la réplication (Nilsson-Payant *et al.*, 2021). Il a été montré dans le cas du virus de la rougeole (MeV) que cet assemblage permettait également de limiter la détection des ARN viraux par les récepteurs intracellulaires de l'immunité innée (Mura *et al.*, 2017).



**Figure 15 : La nucléocapside du virus Nipah.** (A) Préparation en coloration négative et marquage à l'or des nucléocapsides de NiV. Les barres noires correspondent à 100nm (Figure de Goldsmith *et al.*, 2003). (B&C) Ces figures sont issues de (Ker *et al.*, 2021). (B) Représentation schématique des régions de la nucléoprotéine de NiV, les rectangles correspondent aux domaines structurés et la ligne en pointillés correspond à une région désordonnée. (C) Résolution par cryoEM de la structure de la nucléocapside du NiV lié à l'ARN viral. Les conventions utilisées correspondent à (B).

Chez les *Paramyxoviridae*, chaque protéine N recouvre exactement 6 nucléotides d'ARN et le génome des paramyxovirus est obligatoirement un multiple de 6 (« règle des 6 »). Il est supposé que cela résulte de la nécessité de positionner les séquences promotrices nécessaires à l'initiation de la synthèse de l'ARN dans la phase correcte par rapport à la nucléoprotéine formant la NC (Kolakofsky *et al.*, 2005; Ruigrok, Crépin and Kolakofsky, 2011; Longhi *et al.*, 2017; le Mercier and Kolakofsky, 2019). Une déviation à cette règle des six entraîne une réduction considérable de la réplication.

De plus, il a été montré que les virus dont le génome respecte cette « règle des 6 » produisent des ARN messagers issus du gène P dit « édités » par l'insertion de 1 ou 2 résidus guanine décalant le cadre de lecture (voir Chapitre 2 – Partie I.B. pour les détails du mécanisme d'édition). L'application de la « règle des 6 » permet d'éviter la propagation de génomes ou d'antigénomes qui auraient été édités (Iseni *et al.*, 2002).

Les protéines N des *Paramyxoviridae* sont composées de deux régions principales (Yabukarski *et al.*, 2014; Bloyet, 2021; Ker *et al.*, 2021) (**Figure15B&C**) :

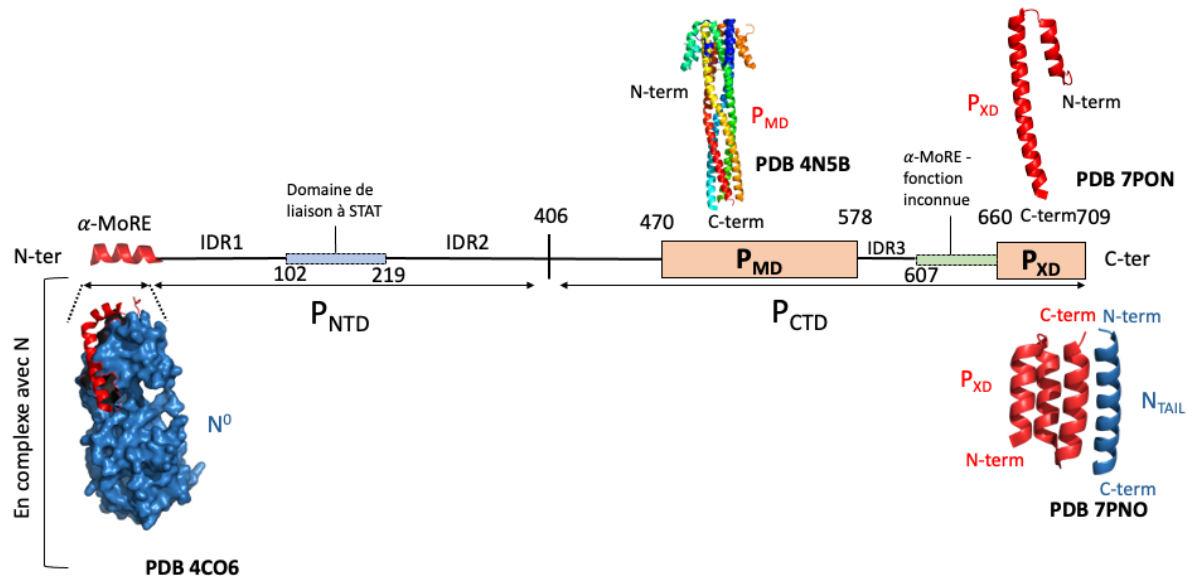
- Une région centrale N<sub>CORE</sub> formée de deux domaines globulaires N-terminal (N<sub>NTD</sub>) et C-terminal (N<sub>CTD</sub>) reliés par une charnière créant un sillon de liaison à l'ARN (Yabukarski *et al.*, 2014). Deux sous-domaines permettent à chaque N de s'arrimer aux sous-unités voisines, une région N-terminale (N<sub>TARM</sub>) et une C-terminale (C<sub>TARM</sub>) (Gutsche *et al.*, 2015). Ils enferment la molécule d'ARN au niveau d'une région riche en résidus basiques.
- Une longue région intrinsèquement désordonnée C-terminale, appelée N<sub>TAIL</sub>, capable d'interagir avec le domaine C-terminal de la phosphoprotéine P (P<sub>XD</sub>) (Bourhis *et al.*, 2022) et la polymérase L (Longhi *et al.*, 2017). Ces interactions jouent des rôles clés dans les mécanismes de transcription, de réplication et d'encapsidation en connectant P et L à la nucléocapside. Cette interaction permettrait le maintien de la polymérase à la matrice nucléocapside durant la transcription mais également lors du passage des régions intergéniques durant la réplication (Bloyet *et al.*, 2016).

### (3.c) La protéine P

Le gène codant pour la phosphoprotéine P est un exemple remarquable de la façon dont les virus exploitent la capacité de codage d'un gène viral. En effet, il comprend de multiples informations en codant, en plus de la protéine P, plusieurs produits différents, appelées protéines accessoires ou non-structurales, par deux mécanismes (Lo *et al.*, 2009; Knipe and Howley, 2013) : l'édition de l'ARNm (protéines P, V et W) ou le chevauchement du cadre de lecture (protéines C).

La protéine P est la seule protéine codée par le gène P qui est essentielle dans la synthèse de l'ARN viral. Chez les *Orthoparamyxovirinae*, la protéine P des henipavirus possède plus de 100 aa supplémentaires aux autres protéines P (Jensen *et al.*, 2020). Les protéines P jouent notamment un rôle de cofacteur indispensable de la polymérase virale L

pour la transcription/réplication et stabilisent les nucléoprotéines libres  $N^0$ . Les protéines P interagissent également avec le domaine N-terminal de la protéine L pour permettre son ancrage à la nucléocapside.



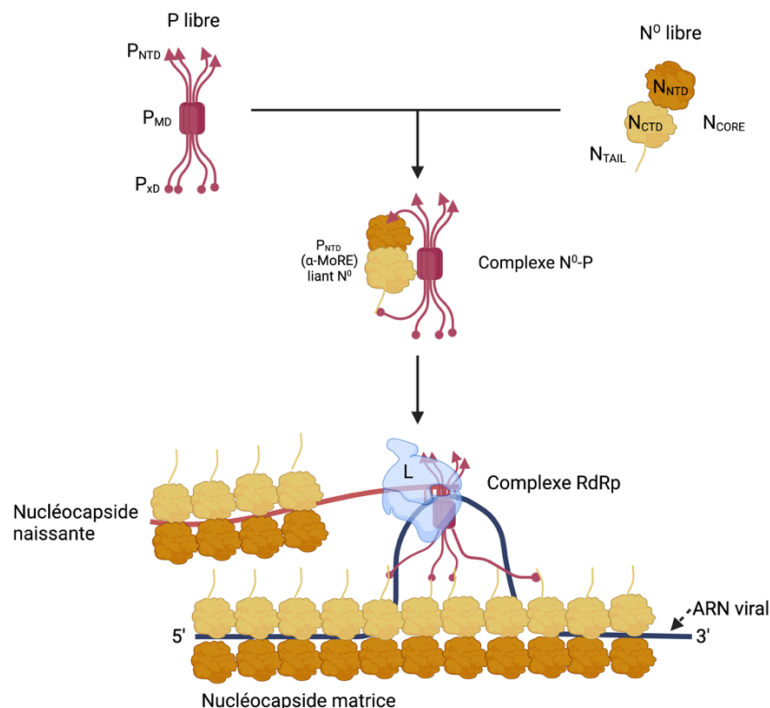
**Figure 16 : La phosphoprotéine (P) du virus Nipah.** Les rectangles orange entourés par des lignes continues représentent les domaines structurés de la protéine. Les lignes noires représentent les régions désordonnées. Le rectangle bleu représente le domaine de liaison à STAT1. Le rectangle vert représente une région de fonction inconnue prédite ordonnée, probablement capable de se replier en  $\alpha$ -MoRE. Les structures du domaine d'oligomérisation  $P_{MD}$  (PDB 4N5B) (Bruhn *et al.*, 2014; Jensen *et al.*, 2020) et du domaine X  $P_{XD}$  (PDB 7PON) (Bourhis *et al.*, 2022) résolues par cristallographie sont représentées ci-dessus. Le domaine d'oligomérisation  $P_{MD}$  forme un tétramère d'hélices parallèles super enroulées (Coiled-coil). La structure du domaine  $P_{XD}$  (en rouge), représentée ci-dessous, a été résolue seule et en complexe avec la queue  $N_{TAIL}$  de la nucléoprotéine (en bleue) (PDB 7PNO) (Bourhis *et al.*, 2022). En complexe, le domaine  $P_{XD}$  se replie en 3 hélices  $\alpha$  en conformation « up-down-up » à la différence des 2 hélices anti-parallèles de la structure seule. En N-terminal se trouve le domaine  $\alpha$ -MoRE dont la structure (en rouge) en complexe avec la nucléoprotéine (en bleue) est représentée ci-dessous (PDB 4C06) (Yabukarski *et al.*, 2014) (Figure inspirée de Longhi *et al.*, 2017; Jensen *et al.*, 2020).

Les protéines P présentent des régions de désordre intrinsèque alternant avec des domaines structurés lui permettant d'assurer différentes fonctions à différents moments du cycle viral. Ces domaines ont été caractérisés, dans le cas du virus Nipah, à la fois par cristallographie, par résonance magnétique nucléaire et par modélisation et sont représentés sur la **figure 16** (Bruhn *et al.*, 2014; Jensen *et al.*, 2020; Bourhis *et al.*, 2022). Elles sont composées de 4 régions :

- Une région N-terminale désordonnée contenant un site  $\alpha$ -MoRE en N-terminale (Molecular Recognition Pattern). Les motifs MoRE sont des régions intrinsèquement désordonnées ayant tendance à se replier lors de la liaison avec un partenaire. Dans ce cas précis, cette région N-terminale est capable de se replier en hélice- $\alpha$  suite à l'interaction avec les nucléoprotéines libre  $N^0$ . Cette région désordonnée, commune aux protéines P, V et W, contient, pour plusieurs paramyxovirus dont le NiV (Jensen *et al.*, 2020 ; Ciancanelli *et al.*, 2009) et le HeV (Rodriguez, Wang and Horvath, 2003), une séquence

d'interaction avec STAT1 et STAT2, facteurs cellulaires induits par la cellule en réponse à l'infection virale.

- Une région centrale ordonnée hélicoïdale, appelée  $P_{MD}$  (Multimerization domain), responsable de l'oligomérisation de la phosphoprotéine. Les domaines  $P_{MD}$  sont constitués de longues hélices- $\alpha$  s'assemblant parallèlement en tétramères pour former un faisceau d'hélice  $\alpha$  (coiled coil) (Bruhn *et al.*, 2014; Jensen *et al.*, 2020).
- Une région connectrice désordonnée et flexible qui semble également contenir des motifs  $\alpha$ -MoRE.
- Un domaine C-terminal appelé  $P_{XD}$  qui forme un complexe avec la région  $\alpha$ -MoRE C-terminale désordonnée de N ( $N_{TAIL}$ ). L'interaction forme un complexe de quatre hélices  $\alpha$  au sein duquel le  $\alpha$ -MoRE de  $N_{TAIL}$ , replié en hélice suite à l'interaction avec  $P_{XD}$ , s'encastre à l'intérieur d'une fente hydrophobe créée par les hélices  $\alpha_2$  et  $\alpha_3$  de  $P_{XD}$ . Les structures des domaines  $P_{XD}$  ont été étudiées pour le virus Hendra et le virus Nipah (Communie *et al.*, 2013; Bourhis *et al.*, 2022). L'interaction de  $P_{XD}$  avec  $N_{TAIL}$  est faible, avec une constante de dissociation ( $K_D$ ) de l'ordre du  $\mu M$  (Bloyet *et al.*, 2016; Longhi *et al.*, 2017; Bourhis *et al.*, 2022) et est supposée réguler la processivité du complexe polymérase sur la nucléocapside. Le modèle actuel propose que cette interaction maintient la polymérase, par le biais de la P, ancrée au nucléocapside durant la lecture de l'ARN viral et notamment lors du passage des régions IGs ou lors de la ré-initiation de la transcription (Bloyet *et al.*, 2016; Longhi *et al.*, 2017; Bloyet, 2021).



**Figure 17 : Représentation schématique de l'encapsidation de l'ARN viral par la nucléocapside et le complexe RdRp.** Les domaines C-terminal  $P_{XD}$  et N-terminal  $P_{NTD}$  d'un tétramère de phosphoprotéine interagissent, respectivement, avec la queue  $N_{TAIL}$  désordonnée d'une nucléoprotéine

*N<sup>o</sup> libre et le domaine N<sub>NTD</sub> du core de la nucléoprotéine (illustration centrale). Le tétramère de phosphoprotéine est capable d'interagir avec la polymérase virale ARN-ARN dépendante L formant le complexe RdRP (illustration du bas). Ce complexe est ensuite capable de répliquer et/ou transcrire l'ARN viral selon un mécanisme qui n'est pas encore complètement élucidé à l'heure actuelle (Figure adaptée de Yabukarski et al., 2014).*

Une étude biochimique récente a montré que seule une copie du domaine P<sub>XD</sub> est nécessaire et limitante pour les interactions N-P et L-P tandis que le motif N<sub>TAIL</sub> n'est pas essentielle pour la liaison préliminaire de la polymérase à la matrice d'ARN (Du Pont, Jiang and Plemper, 2019). Cependant, le mécanisme par lequel la polymérase progresse le long de l'ARN encapsidé est encore à l'étude et soulève de nombreuses questions. Il est supposé que la la phosphoprotéine sous la forme d'un tétramère joue le rôle d'un médiateur entre la protéine L et la nucléocapside (**Figure 17**). Il s'agit du modèle dit de "cartwheeling" dans lequel le complexe polymérase progresse sur sa matrice grâce à des cycles d'association/dissociation entre le domaine P<sub>XD</sub> et la nucléocapside (Curran, 1998; Kolakofsky et al., 2004; Longhi et al., 2017; Bloyet, 2021).

### (3.d) Transcription primaire et secondaire, gradient des ARNm viraux, traduction et répllication du génome

Les virus à ARN négatif non segmenté sont connus pour former des corps d'inclusion sans membranes comprenant des protéines impliquées dans la machinerie de répllication. Il a été récemment observé, par exemple, que les protéines N et P du virus de la rage (Nikolic et al., 2017) et du MeV (Guseva et al., 2020) co-localisent dans ces corps d'inclusion, également appelés usines virales, après une infection. Il est aujourd'hui supposé que ce mécanisme est partagé à travers l'ordre des *Mononegavirales*. La transcription et la répllication virale se produisent dans ces structures. Les génomes viraux, ainsi que les protéines N, P et L, sont concentrés dans ces inclusions.

La transcription intracellulaire est initiée par les protéines contenues dans la particule virale (Noton and Fearn, 2015). Elle commence avec la transcription primaire par la polymérase virale ARN dépendante (vRNAP) du génome encapsidé par les protéines N en ARN messenger coiffé en 5' et polyadénylé en 3'. La vRNAP est formée au minimum d'un homo-tétramère de protéine P et d'une protéine L. La vRNAP commence la synthèse des ARNm par l'extrémité 3' de l'ARN génomique en reconnaissant la séquence promotrice leader et transcrit les gènes de manière séquentielle et polaire (3' vers 5') en terminant et réinitiant la transcription à la fin de chaque gène viral au niveau de séquences intergéniques (séquences non codantes, NCR). Ces séquences intergéniques sont composées :

- D'une séquence d'initiation (gene-start, GS) permettant l'initiation de la transcription et la synthèse d'une coiffe ARN.
- D'une courte séquence intergénique non transcrite (IG)

- D'une séquence de terminaison (gene-end, GE) de la transcription qui contient une série de 4 à 7 résidus uridine monophosphates à l'origine de la polyadénylation de l'ARN messenger.

Au moment du passage sur chacune de ces NCR, la polymérase peut, avec une fréquence variable, redémarrer la transcription au début du gène suivant (séquence GS), se désengager de l'ARN matrice et réinitier une nouvelle fois la transcription du génome au niveau de la séquence leader. De cette façon-là se crée un gradient de concentrations d'ARNm qui est inversement proportionnel à la distance entre le gène et l'extrémité 3' du génome.

La traduction des ARNm produits est ensuite assurée par les ribosomes cellulaires. On observe un gradient de protéines virales qui découle directement du gradient de transcription de l'ARN génomique. La taille du génome des *henipavirus* est 10% plus grande que les autres genres, notamment par la présence de séquences GS plus grandes qui semblent jouer un rôle essentiel dans la mise en place du gradient protéique (Sugai *et al.*, 2017).

La production de nouvelles protéines virales, notamment de protéines L, entraîne la formation de nouveaux complexes de transcription capables d'initier une transcription secondaire permettant une augmentation de la production d'ARNm et de protéines virales.

Dès lors que suffisamment de protéine N est produite, la polymérase virale est capable de passer d'un mode « transcriptase » à un mode « réplique » (Kolakofsky *et al.*, 2005). Elle génère alors des copies entières de l'ARN viral sans reconnaître les signaux intergéniques et donc sans s'arrêter aux extrémités des gènes. Ces copies de polarité positive appelées anti-génomes seront par la suite utilisées comme matrices pour la synthèse de génomes complémentaires de polarité négative. Les ARN génomiques et antigénomiques sont encapsidés par des nucléoprotéines, formant des nucléocapsides compétentes pour la polymérase.

Nous verrons par la suite que de nombreux mécanismes, reposant sur les protéines C, régulent différents points clés de la multiplication virale ayant en conséquence un rôle essentiel dans la pathogénicité de ces virus.

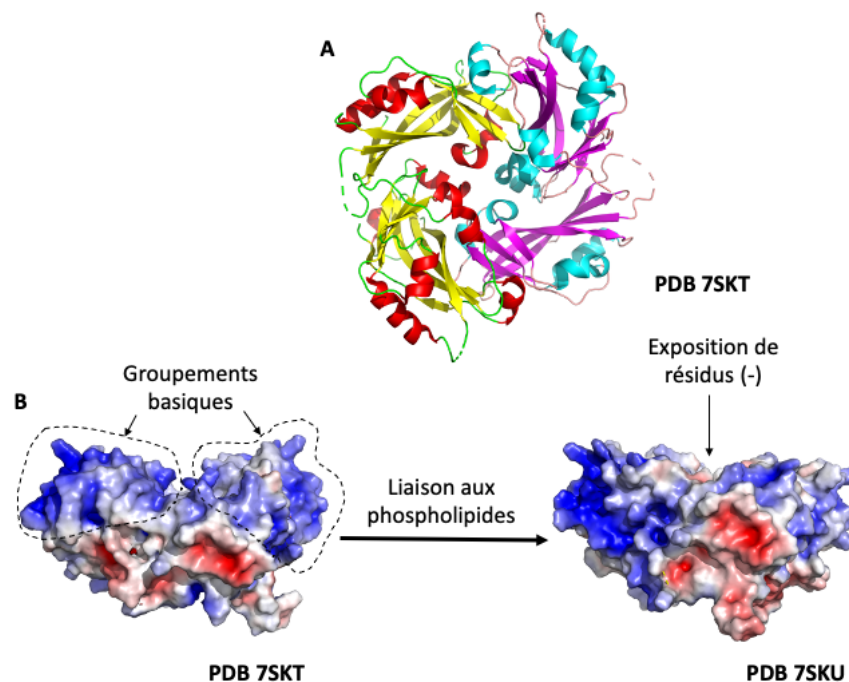
#### (4) Assemblage et bourgeonnement

La dernière étape du cycle viral est constituée par l'assemblage et le bourgeonnement des particules virales (Takimoto and Portner, 2004). Afin de produire des particules virales infectieuses, les paramyxovirus ont besoin de trois composants clés qui sont concentrés à la membrane plasmique avant le bourgeonnement : les glycoprotéines d'enveloppe, les protéines de matrice M et les complexes ribonucléoprotéiques (RNPs). Le RNP est constitué du génome viral encapsidé par la nucléoprotéine (N) associée à la polymérase virale ARN-ARN dépendante (RdRP) constituée de la polymérase virale (L) et de la phosphoprotéine (P). Des interactions entre les trois composants M, RNPs et glycoprotéines d'enveloppe sont essentielles à la formation de particules virales et de nombreuses études, revues dans (El

Najjar, Schmitt and Dutch, 2014), ont montré le rôle déterminant joué par les protéines M dans ce processus.

#### (4.a) La protéine M

La protéine de matrice (M) est la plus abondante dans le virion. Ces protéines sont majoritairement basiques, hydrophobes, sans domaine d'accrochage à la membrane lipidique et visibles en microscopie électronique par la présence d'une couche dense d'électrons sous la membrane lipidique. Elles interagissent avec la queue cytoplasmique des protéines virales membranaires G et F ainsi qu'avec les nucléocapsides et la membrane lipidique. Elles sont garantes de la morphogénèse virale par leur autoassemblage en un réseau tridimensionnel et par des interactions avec les nucléocapsides qui guident la formation des particules virales au moment du bourgeonnement et assurent la cohésion entre les couches externes et internes de la particule virale (Harrison, Sakaguchi and Schmitt, 2010).



**Figure 18 : La protéine de matrice (M) de virus Nipah (PDB 7SKT et 7SKU).** (A) Structure tridimensionnelle du dimère stable formé par la protéine de matrice M. Les feuillets  $\beta$  et hélices  $\alpha$  des deux monomères sont représentés, respectivement, en jaune ou magenta et en rouge ou cyan. Les boucles désordonnées sont colorées en vert et rose. (B) Surface de potentiel électrostatique du bleu (+5KbT) au rouge (-5KbT) du dimère formé par la protéine M du NiV avant liaison aux phospholipides membranaires (gauche) et après liaison (droite). On observe une ouverture de l'angle formé par deux groupements basiques (entourés sur la structure 7SKT) permettant l'exposition de résidus chargés négativement au centre (flèche de la structure 7SKU). Ce changement de conformation est à l'origine de la courbure de la membrane plasmique par la protéine M lors du bourgeonnement viral (Figure adapté de Norris et al., 2022).

Malgré une relative faible similarité de séquence (environ 20%), les structures tridimensionnelles des protéines M connues des paramyxovirus partagent une architecture commune (Liu, Grusovin and Adams, 2018; Donnelly et al., 2021) (**Figure 18A**) : elles comprennent deux domaines en feuillets  $\beta$  repliés décorés par quelques hélices  $\alpha$  exposées



reliées par une boucle désordonnée. Les protéines M présentent de larges surfaces de résidus chargés positivement et plusieurs résidus hydrophobes exposés qui sont responsables de l'association des membranes pendant le bourgeonnement. Elles s'autoassemblent en dimères stables qui eux-mêmes forment des pseudo-tétramères sous la membrane plasmique.

#### (4.b) Interactions entre trois composants clés lors de l'assemblage et bourgeonnement de nouveaux virions

Le rôle de la protéine M dans le bourgeonnement a été étudié pour de nombreux paramyxovirus dont, entre autres, SeV (Takimoto *et al.*, 2001), NiV (Ciancanelli and Basler, 2006; Patch *et al.*, 2007; Norris *et al.*, 2022) et MeV (Runkler *et al.*, 2007; Norris *et al.*, 2022) pour lesquels son expression transitoire suffit à promouvoir le bourgeonnement de particules virales.

Il a été récemment montré que les protéines M du NiV et de MeV à elles seules sont capables de s'associer aux membranes, d'induire la courbure de la membrane plasmique et de provoquer sa scission par leur dimérisation qui permet d'exposer une surface concave chargée positivement capable d'interagir avec les lipides chargés négativement des membranes (**Figure 18B**) (Norris *et al.*, 2022). Ceci entraîne alors un changement de conformation du dimère provoquant la courbure de la membrane. Les dimères de protéines M s'autoassemblent ensuite, entraînant la formation du virion. Les protéines M sont également capables d'interagir avec la queue cytoplasmique des glycoprotéines d'enveloppe et les RNPs permettant leur concentration au site de bourgeonnement.

Chez certains *Orthoparamyxovirinae*, les protéines non-structurales encodées par le gène P jouent un rôle régulateur dans l'assemblage et le bourgeonnement des particules virales. Ce sujet sera abordé ultérieurement.

### ***(F) Le contrôle de la réponse immunitaire innée antivirale***

#### **(1) Détection des pathogènes viraux et mise en place de l'immunité innée**

L'entrée d'un pathogène dans une cellule hôte est détectée par de nombreux senseurs appelés PRRs (Pattern Recognition Receptors - PRRs) reconnaissant des motifs conservés parmi les pathogènes, appelés PAMPs (Pathogen-Associated Molecular PatternS) (Carty, Guy and Bowie, 2021). Nous nous concentrerons dans cette partie sur les voies activées par la détection d'un pathogène viral et, notamment, les voies régulées par la suite par divers mécanismes (détaillés dans la suite du texte) par les *Orthoparamyxovirinae* (**Figure 19**).

Les PAMPs viraux sont de natures variées : ARNs viraux triphosphates sans coiffe, ARNs viraux double et simple brin issus de la réplication virale présent dans le cytoplasme<sup>1</sup> et

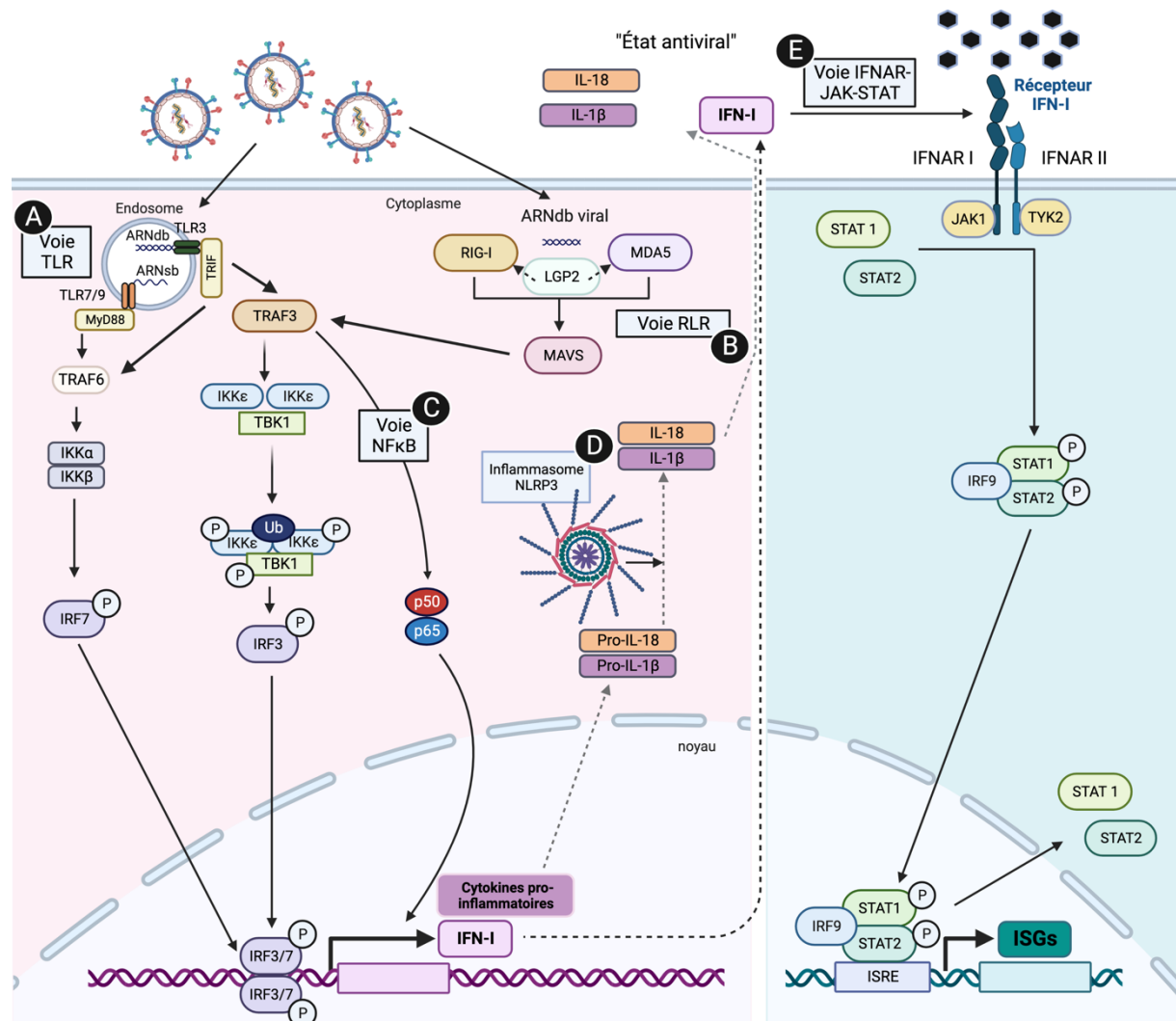
---

<sup>1</sup> La transcriptase et la réplicase des *Orthoparamyxovirinae* peut faire des erreurs conduisant à la production d'ARNm coiffés ne possédant pas de queue polyA pouvant s'associer à l'ARN trailer pour former une séquence double-brin. De même, certains ARN simple brin peuvent adopter une structure secondaire conduisant à des configurations partiellement double brin.

protéines virales de surface (Gerlier and Lyles, 2011). Ils sont reconnus par différents récepteurs classés dans plusieurs catégories : les TLRs (Toll-Like Receptors), les RLRs (RIG-I-Like Receptors), les CLR (C type Lectin Receptors) et les inflammasomes. La détection d'un virus par les PRRs entraîne le déclenchement de cascades de signalisation aboutissant à l'expression d'interféron de type I (IFN-I) entraînant la transcription de gènes stimulés par l'IFN-I (ISG) impliqués dans la réponse immunitaire innée.

Les mécanismes de l'immunité innée sont non-spécifiques et ne produisent pas d'effet mémoire vis-à-vis de l'agent pathogène.

La deuxième ligne de défense mise en place à la suite d'une infection par un agent pathogène est l'immunité acquise basée sur les lymphocytes B et T. Les lymphocytes B étant responsables de la production d'anticorps et de la réponse immunitaire à médiation cellulaire. L'immunité acquise est hautement spécifique et requiert un temps de mise en place plus long mais avec un effet mémoire.



**Figure 19 : Représentation schématique de la modulation de la réponse immunitaire IFN de type I suite à l'infection par un paramyxovirus.** L'infection par un paramyxovirus est suivie par la production d'ARN viraux qui peuvent être reconnus par divers récepteurs : (A) Les récepteurs TLRs (TLR3 et TLR7/9) qui déclenchent une cascade d'activation entraînant la phosphorylation et l'entrée dans le noyau des facteurs de transcription IRF3 et IRF7. Ces facteurs de transcription favorisent la production d'IFN-I et de cytokines pro-inflammatoires ; (B) Les récepteurs RLRs (LGP2, RIG-I et

*MDA5) qui peuvent également activer la cascade de réactions de IRF3 mais aussi (C) la voie NFκB permettant la translocation dans le noyau des facteurs de transcription p50 et p65. Ces voies permettent la production d'IFN-I et de cytokines pro-inflammatoires. (D) Les cytokines pro-inflammatoires sont envoyées vers l'inflammasome NLRP3 qui permet leur maturation. Ces cytokines, ainsi que les IFN-I sont sécrétées afin de stimuler le recrutement des cellules immunitaires et de mettre en place un état « anti-viral » des cellules. L'inflammasome, une fois activé suite à la détection d'un pathogène, peut également déclencher des programmes de mort cellulaire ; (E) Les IFN-I sécrétés activent la réponse IFN-I avec notamment la voie IFNAR-JAK-STAT. Cette voie entraîne la phosphorylation et translocation du complexe de facteur de transcription STAT1-STAT2 qui stimule la transcription des gènes ISGs. Les produits de ces gènes sont divers et permettent la généralisation de l'état « anti-viral » des cellules (Figure inspirée de Pelissier, Iampietro and Horvat, 2019; Wang et al., 2022 et réalisée dans BioRender)*

Les TLRs sont des récepteurs membranaires de type I répartis en deux catégories : les TLR endosomaux (TLR3, TLR7, TLR8 et TLR9), capables de détecter la présence d'acides nucléiques viraux, et les TLRs de surface (TLR2 et TLR4), capables de détecter les protéines virales de surface (Fitzgerald and Kagan, 2020). Selon plusieurs études et revues, les paramyxovirus peuvent être détectés par les récepteurs TLR3, TLR7, TLR8 (Kitagawa *et al.*, 2011; Caignard, Vidalain and Tangy, 2012; Wang *et al.*, 2022). La détection des paramyxovirus par ces TLRs entraîne le recrutement des récepteurs Toll/interleukin-1 (TIR) qui incluent le facteur 88 de différenciation myéloïde (MyD88) et l'adaptateur contenant un domaine TIR induisant les IFN-β (TRIF) (**Figure 19A**). Chacun de ces récepteurs est responsable de l'activation de cascades de signalisation conduisant à la production d'IFN-I, de cytokines pro-inflammatoires et de chemokines médiées par les facteurs de régulation TRAF3 et TRAF6 (TNF receptor-associated factor 3) de la voie de signalisation IRF3/7 et à l'activation des ISG par la voie NFκB (**Figure 19B**).

Les RLRs sont une famille de récepteurs cytoplasmiques incluant les récepteurs RIG-I, (Retinoic acid-inducible gene I), MDA5 (Melanoma Differentiation-Associated protein 5) et LGP2 (Laboratory of Genetics and Physiology 2) (Rehwinkel and Gack, 2020). RIG-I et MDA5, dont les fonctions sont modulées par LGP2, sont capables de se lier à la protéine MAVS (Mitochondrial AntiViral Signaling) induisant alors la production d'IFN-I par la voie de signalisation IRF3 et la production de cytokines pro-inflammatoires par la voie NFκB (**Figure 19C**).

Les inflammasomes sont des complexes cytosoliques de multiples protéines assemblées suite à un stress cellulaire ou à une infection par un récepteur de la famille NLRP (Nucleotide-binding oligomerization domain, Leucine rich Repeat and Pyrin domain containing). Ils recrutent l'activation de caspases inflammatoires induisant la production d'interleukines IL-1β et IL-18 qui seront sécrétées et participent à la mise en place de la réponse immunitaire par l'activation de certaines cellules immunitaires (Broz and Dixit, 2016). L'inflammasome est responsable de l'activation des processus inflammatoires, et peut induire un phénomène de pyroptose, programme de mort cellulaire différent de l'apoptose (**Figure 19D**).

## (2) Reconnaissance des virus à ARN

Dans le cas des virus à ARN, un certain nombre de structures virales spécifiques sont à l'origine de la détection du pathogène par les cellules infectées (Carty, Guy and Bowie, 2021):

- Les ARN doubles (ARNdb) et simples brins (ARNsb) viraux sont reconnus par les TLR3 et TLR7/8.
- MDA5 reconnaît l'accumulation des ARNdb viraux dans les cellules infectées.
- RIG-I reconnaît les ARNdb viraux 5'-triphosphates non-coiffés, les ARNdb viraux 5'-diphosphates non-coiffés et les ARN présentant un nucléotide 5'-terminal non méthylé en position 2'-O.

Dans le cas des paramyxovirus, les ARNm viraux sont coiffés et polyadénylés et les ARN génomiques et antigénomiques viraux sont encapsidés, ce qui prévient normalement leur détection par des PRRs. Cependant, durant la transcription et la réplication des virus à ARN, la transcriptase et la réplicase virale peuvent faire des erreurs conduisant à la synthèse d'ARN défectueux qui joue un grand rôle dans la détection par la cellule infectée. En effet, certains ARNm pourront, suite à une erreur, être coiffés mais non polyadénylés ou formés des séquences double brins reconnus par la cellule. Certains ARN simples brins peuvent aussi adopter des structures secondaires partiellement double brin également reconnues par la machinerie de détection cellulaire.

Les particules virales déficientes interférentes (DI) du génome viral sont des copies incomplètes du génome viral pouvant également déclencher une détection par la cellule hôte (Pfaller *et al.*, 2014). De plus, des ARN courts peuvent s'associer à ces DI et former des ARN double-brins reconnus alors par la cellule.

## (3) Réponse à l'IFN de type I

La réponse interféron de type I constitue la première ligne de défense de l'immunité innée suite à l'infection par un virus. MyD88, TRIF et MAVS induisent l'expression de cascades de signalisation permettant l'activation de facteurs de transcription IRF3, IRF7 (**Figure 20A&B**) et la voie NFκB avec les facteurs p50 et p65 (**Figure 20C**). Ces facteurs sont transloqués dans le noyau où ils vont induire la transcription d'IFN-I et de cytokines pro-inflammatoires.

Les cytokines et IFN-I sont secrétées et pourront activer les cellules voisines par la voie de signalisation IFNAR-JAK-STAT en se liant au récepteur membranaire IFNAR (**Figure 20E**). La liaison entraîne alors la phosphorylation des protéines tyrosines kinases Janus (TYK1 et JAK1) liées à la partie cytosolique du récepteur. Ces protéines activent à leur tour les facteurs de transcription STAT1 et STAT2 (Signal Transducers and Activators of Transcription 1 and 2). Ces facteurs peuvent se complexer sous la forme d'hétérodimères avec le facteur de régulation 9 (IRF9), se déplacer dans le noyau et se lier à une séquence d'activation génique des gènes de réponse à l'interféron (ISRE) afin de stimuler des centaines

d'ISGs (IFN-Stimulated Genes) responsables de la généralisation de l'état antiviral des cellules. Ces gènes jouent un rôle important dans le contrôle de l'infection virale.

L'activation du système immunitaire suite à l'infection par un virus peut également entraîner la sous- ou sur-expression de cytokines impliquées dans la signalisation autocrine et paracrine des chimiokines pro-inflammatoires (CXCL 1, 2, 3, 8, 10 et CCL5). Ces molécules ont la capacité de guider les cellules effectrices du système immunitaire vers le site d'infection provoquant notamment une inflammation pathologique et destructrice très représentative de la pathogénèse de certains paramyxovirus.

**En réponse aux mécanismes de défenses cellulaires, les paramyxovirus ont développé de nombreuses stratégies d'échappement à l'immunité innée et aux défenses antivirales. Ces stratégies passent notamment par la synthèse de protéines accessoires encodées sur le gène P. Nous accorderons une attention particulière à la protéine C exprimée par presque tous les *Orthoparamyxovirinae*.**





# Chapitre 2 : L'importance des protéines non structurales des *Orthoparamyxovirinae*

## I. Les protéines non structurales des *Orthoparamyxovirinae*

### (A) Les gènes chevauchants chez les virus

La plupart des virus possède des gènes codant pour plus d'une protéine. Lorsque ces protéines émergent de la traduction d'une séquence commune de nucléotides depuis différents cadres de lecture, ce phénomène est nommé chevauchement de gène (« gene overlap ») ou surimpression (« overprinting »). Chez les virus, ce phénomène a souvent été relié aux nombreuses contraintes exercées sur le génome viral et est supposé donner un avantage évolutif (Sabath, Wagner and Karlin, 2012) car ils permettraient de maximiser la quantité d'informations contenues sur un génome de petite taille. Ils représentent un mécanisme extraordinaire pour l'évolution de nouvelles protéines et fonctions. Pour les virus à ARN qui mutent rapidement, les gènes chevauchants représentent un moyen de fournir un avantage en terme de fitness en imposant des contraintes évolutives sur la séquence et en limitant les mutations nuisibles. Nombre de ces gènes n'ont pas de fonctions connues, plusieurs ont été associés à des fonctions de transmission du virus et de sévérité de la maladie et à des fonctions de régulation de l'expression des gènes.

### (B) Mécanismes d'expression des protéines accessoires

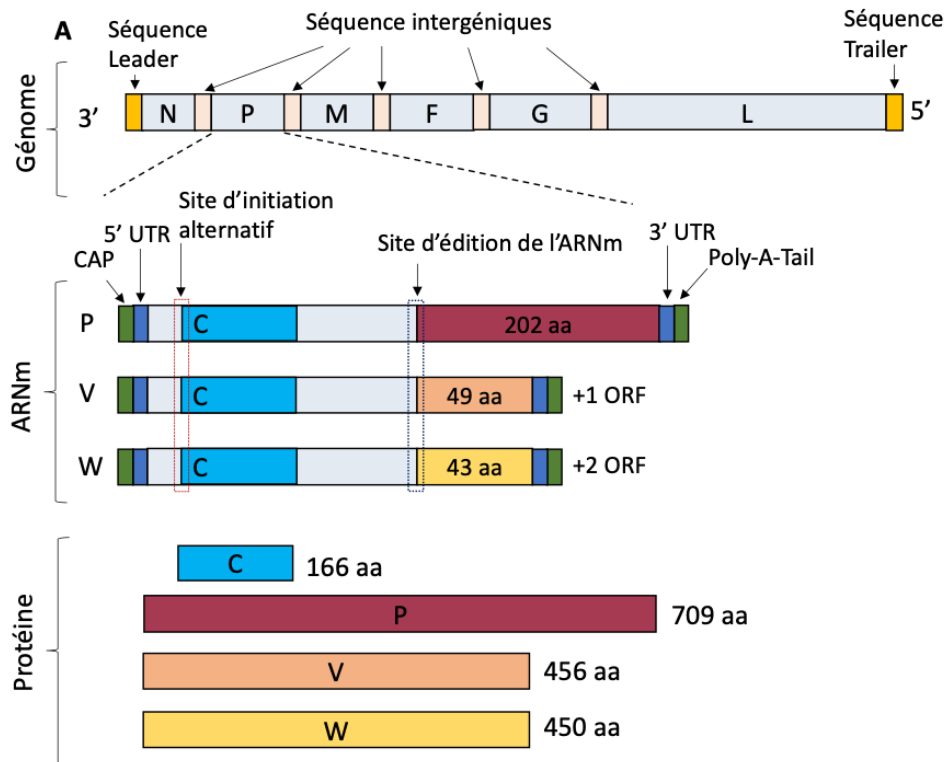
Pour exprimer de nouvelles protéines, les *Paramyxoviridae* utilisent deux stratégies principales décrites dans la **figure 20** :

- (1) l'insertion de nouveaux gènes dans le génome à condition de respecter la règle des 6.
- (2) l'expression de plus d'une protéine à partir de la même séquence d'acide nucléique.

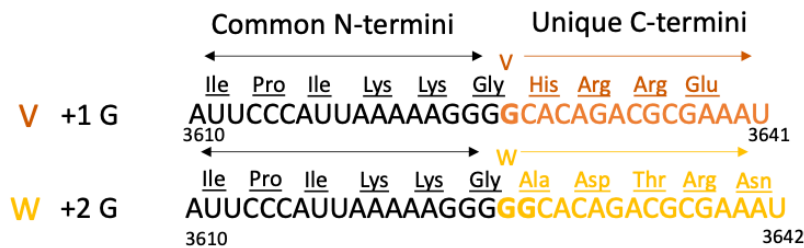
Le gène P des *Paramyxoviridae* est un exemple remarquable de la façon dont les virus exploitent la capacité de codage d'un gène viral. En utilisant la deuxième stratégie mentionnée, le gène P exprime de multiples informations et code plusieurs produits différents via deux mécanismes (Lo *et al.*, 2009; Schlub and Holmes, 2020) (**Figure 20A**) :

- (a) édition co-transcriptionnelle au niveau d'une séquence spécifique où la polymérase bégaie (codage des protéines non structurales V, W) (**Figure 20B**).
- (b) initiation alternative de la traduction dans des cadres de lecture différents (protéines non structurales C, C', Y1, Y2) (**Figure 20C**).

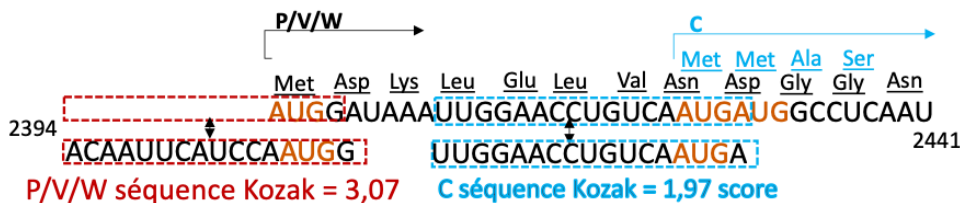




**B Mécanisme d'édition co-transcriptionnelle de l'ARNm : protéines V/W**



**C Mécanisme d'initiation de la traduction par balayage :**



**Figure 20 : Mécanismes d'expression des protéines non structurales V, W par édition co-transcriptionnelle depuis le gène P et C et par initiation de la traduction par balayage depuis les ARNm codant pour P/V et W pour le virus Nipah. (A) Le panneau supérieur montre une représentation schématique du génome du NiV, contenant de l'extrémité 3' à l'extrémité 5', les gènes codant pour les protéines N, P, M, F, G et L. Le panneau central montre l'architecture des ARNm générés à partir du gène P par édition co-transcriptionnelle. L'ARNm codant pour la protéine P n'est pas édité, alors que celui codant pour les protéines V et W a, respectivement, 1G ou 2G insérés au niveau du site d'édition. Les deux ARNm contiennent l'UTR 3' et 5' et sont traités aux deux extrémités avec une structure CAP à l'extrémité 3' et une queue polyA à l'extrémité 5'. Le panneau inférieur montre les protéines exprimées à partir de ces ARNm. Les protéines P, V et W sont exprimées par**

initiation de la traduction au premier codon AUG de leur ARNm respectif, tandis que la protéine C est exprimée par initiation alternative de la traduction au codon AUG suivant en aval des 3 ARNm. **(B)** Le mécanisme d'édition co-transcriptionnelle de l'ARN viral du gène P amenant à la synthèse des ARNm des protéines V et W est détaillé dans ce panneau. Les couleurs sont les mêmes que pour la figure A. **(C)** Le mécanisme d'initiation de la traduction par balayage des ARNm issus de la transcription du gène P amenant à la traduction de la protéine C est détaillé dans ce panneau. Les couleurs sont les mêmes que pour la figure A. Les séquences Kozak calculées selon la méthode de (Iida and Masuda, 1996) pour les protéines P/V et W sont détaillées dans l'encadré rouge et pour la protéine C dans l'encadré bleu (Figure inspirée de Satterfield, Geisbert and Mire, 2016).

## (1) Édition co-transcriptionnelle du gène P

Le premier mécanisme d'expression produit des ARNm dont les ORFs sont modifiés par l'insertion d'un nombre  $n$  de résidus guanosines G par la polymérase virale dans une région spécifique riche en uracile U du gène codant pour la P, ce qui entraîne le décalage du cadre de lecture traductionnel (**Figure 19B**). En conséquence, les protéines P, V et W partagent une région N-terminale identique (407 acides aminés (aa) identiques dans le cas des *Henipavirus*) mais diffèrent dans leurs régions C-terminales suivant le site d'insertion des résidus G. La spécificité de la séquence du site d'édition détermine la fréquence et le nombre d'insertions G conduisant à la production de protéines V ou W.

Dans le cas des *Orthoparamyxovirinae*, ce mécanisme opère via un mode appelé le mode P où P est encodé par le gène non édité et où l'insertion de  $3k+1$  G permet la synthèse de l'ARNm de V et de  $3k+2$  G celle de W (Douglas, Drummond and Kingston, 2021) ( $k$  étant un nombre entier). Le site d'insertion des G correspond à une région intrinsèquement désordonnée des protéines P, V et W (Longhi *et al.*, 2017), c'est pourquoi il est supposé que la taille de la séquence homo-polymérique de G insérée, traduite en une séquence poly-glycine désordonnée, n'a pas de conséquences sur les fonctions des protéines. L'analyse des produits de la transcription des ARNm du gène P par amplification RT-PCR (Lo *et al.*, 2009) à partir de lysats cellulaires infectées à 24h pour plusieurs paramyxovirus a montré que les *Henipavirus* produisent des transcriptions éditées à une fréquence deux fois plus élevée que celle des autres virus (Hausmann *et al.*, 1999; Mebatsion *et al.*, 2001; Bankamp *et al.*, 2008; Lo *et al.*, 2009). Les transcrits du gène P sont éditées selon une fréquence variable dépendant du virus (82% pour le NiV, 42% pour le MeV et 31% dans le cas du SeV) (Kulkarni *et al.*, 2009). Dans le cas du NiV, 33% des transcrits généreront effectivement une protéine P et les transcrits restants généreront des ARNm modifiés au niveau d'un site d'édition donnant lieu à la synthèse des protéines dites accessoires ou non-structurales V (36% des transcrits) et/ou W (31% des transcrits).

## (2) La répllication du génome et le changement de mode d'édition du gène P

Le mécanisme d'édition du gène P n'est présent que dans la famille des *Paramyxoviridae*. Il a été observé que le système d'édition co-transcriptionnelle fonctionne avec deux modes (Douglas, Drummond and Kingston, 2021) :

(a) le mode P où la protéine P est codée par le gène non-édité comme décrit dans la partie 1.

(b) Le mode V, observé pour le genre *Ferlavirus*, où la protéine non éditée correspond à la protéine V. La protéine W dérive alors de l'insertion d'une seule guanosine tandis que la protéine P dérive de l'insertion de 2 guanosines au niveau du site d'édition.

### (3) Surimpression translationnelle des ARNm codant pour P/V/W

Le deuxième mécanisme code pour les protéines C par l'intermédiaire de cadres de lecture ouverts alternatifs. Ces protéines sont exprimées à partir des ARNm P, V ou W par un mécanisme d'initiation de la traduction par balayage (leaky scanning).

Dans l'initiation par balayage, les sous-unités ribosomiques se lient à proximité de la coiffe 5' des ARNm viraux et balayent linéairement jusqu'au premier codon AUG. L'efficacité d'initiation de la traduction par le ribosome est modulée par l'identité des nucléotides précédant et suivant le codon initiateur, la séquence Kozak (Kozak, 1984; Iida and Masuda, 1996) (**Figure 19C** pour le cas du NiV). Pour la protéine C, le site primaire d'initiation de la traduction de la protéine P se trouve parfois dans une séquence consensus de Kozak de score faible (3.07 pour la séquence codante de la P de NiV en comparaison aux autres gènes pour lesquelles le score de la séquence Kozak est compris entre 4.1 et 5.2 à l'exception de la séquence codante pour la protéine C), ce qui entraîne parfois l'initiation de la traduction à des sites alternatifs en aval et dans un cadre de lecture différent de la protéine P.

Pour certains orthoparamyxovirus, notamment le virus Sendai, plusieurs codons initiateurs peuvent être décomptés le long de la séquence codant pour le gène P amenant à la synthèse d'une multiplicité de protéines de type C : les protéines C, C', Y1 et Y2. Les ORFs initiateurs de ces protéines sont situés majoritairement en aval de celui de la protéine P, à l'exception de la protéine C' qui est légèrement en amont de l'ORF de P. Il a été montré notamment, dans le cas du virus Sendai, que la mutation de la séquence d'initiation de la protéine P entraîne des modifications des niveaux d'expression des protéines C suggérant une interdépendance dans l'expression des protéines P et C et, donc, soulevant l'hypothèse d'une interdépendance de leur évolution (Curran, Richardson and Kolakofsky, 1986).

## ***(C) Structures et fonctions des protéines V, W et C***

### (1) La protéine V

Dans l'ensemble, la protéine V est la deuxième plus grande des protéines produites depuis le gène P. C'est une protéine multifonctionnelle impliquée dans l'échappement à la réponse immunitaire innée et l'une des actrices déterminantes de la pathogénicité virale.

Les fonctions des protéines V sont multiples et varient en fonction des genres et des espèces. Les protéines V sont capables d'interagir avec une multitude de facteurs cellulaires pour interférer dans la réponse IFN. L'ensemble de ces fonctions a été largement décrite dans la revue de (Wang *et al.*, 2022), et nous nous concentrerons sur les fonctions concernant le virus Nipah.

La grande majorité des protéines V des *Orthoparamyxovirinae* sont capables d'induire la dégradation par le protéasome des facteurs de transcription STAT1 et STAT2. Dans le cas du NiV, la protéine V induit la formation dans le cytoplasme de complexes de hauts poids moléculaires constitués de STAT1 et STAT2 non phosphorylés (Rodriguez, Parisien and Horvath, 2002) via un signal d'exportation nucléaire localisé dans la partie N-terminale de la protéine V (Rodriguez, Cruz and Horvath, 2004). Cette inhibition contrecarre la signalisation de la réponse IFN-I.

Dans le cas du NiV, les protéines V interagissent par le biais d'un motif riche en tryptophane avec le facteur cellulaire IRF7 pour contrer la réponse par la voie TLR7/9 (Kitagawa *et al.*, 2011).

En plus de partager une fonction d'antagoniste de la voie de régulation du RLR MDA5 avec d'autre *Orthoparamyxovirinae* (Andrejeva *et al.*, 2004; Uchida *et al.*, 2018), la protéine V du NiV interagit également, par le biais de son domaine C-terminal, avec le récepteur RIG-I et avec le facteur cellulaire TRIM25 perturbant la voie de signalisation des MAVS de la réponse IFN-I (Sánchez-Aparicio *et al.*, 2018). Les protéines V du virus Nipah jouent également un rôle dans l'inhibition de la réplication du minigénome de Nipah et d'autre *Orthoparamyxovirinae* comme HPIV3 et MeV (Sleeman *et al.*, 2008).

L'étude de l'histoire évolutive de la protéine V des paramyxovirus (Douglas, Drummond and Kingston, 2021) a montré que la protéine V a été perdue à deux occasions. Chez ces virus, il est supposé que la protéine C serait capable de compenser la perte de la protéine V dans l'échappement à la réponse immunitaire innée.

## (2) Protéine W

À la différence des protéines P et V, la séquence C-terminale de W n'est pas conservée à travers la famille des paramyxovirus et c'est pourquoi une variété de noms a été attribuée historiquement à cette protéine : W, D ou I.

La protéine W possède un signal de localisation nucléaire (NLS) dans sa région C-terminale et se concentre en conséquence dans le noyau des cellules infectées (Shaw *et al.*, 2005; Audsley, Jans and Moseley, 2016a; Karsunke *et al.*, 2019). Les structures, en complexe avec des isoformes de l'importine  $\alpha$ , des régions NLS des protéines W du NiV et de HeV ont été résolues (Smith *et al.*, 2018).

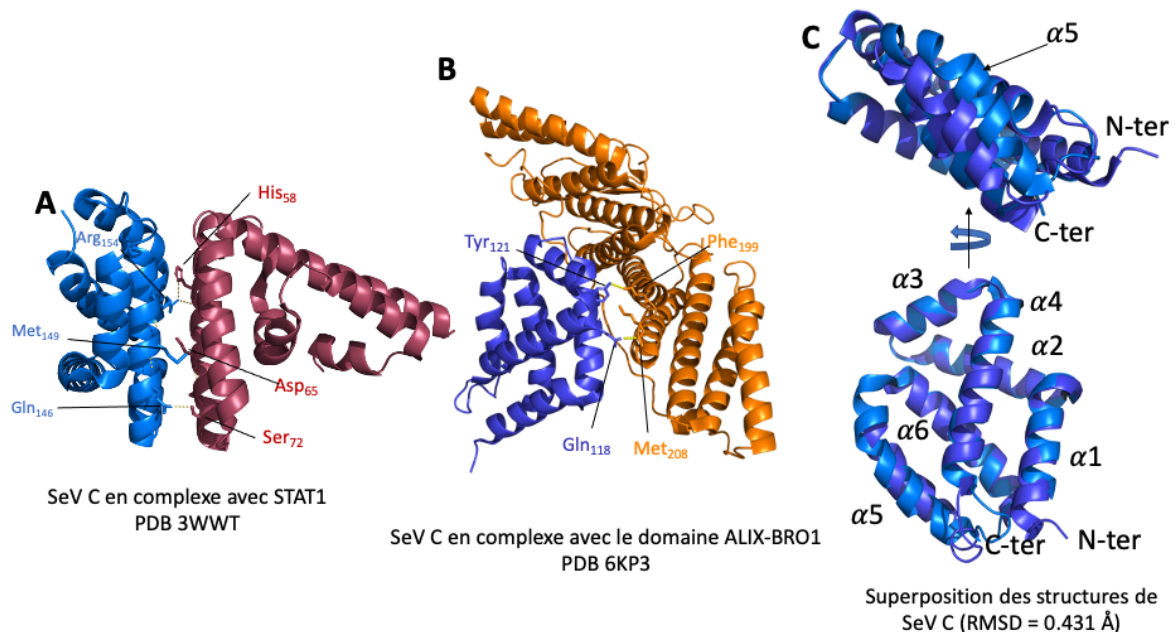
Les fonctions des protéines W sont variées et, pour de nombreux virus, sont encore peu comprises. Dans le cas des *Henipavirus*, il a été démontré pour des modèles animaux que la protéine W joue un rôle dans la pathogénicité du virus notamment via la subversion de la réponse IFN, soit par la séquestration des protéines STAT dans le noyau, soit par une interaction avec la famille des protéines régulatrices 14-3-3 (Ciancanelli *et al.*, 2009; Enchéry *et al.*, 2021). D'autres études ont également montré que, pour le virus parainfluenza humain 3

(HPIV-3), les protéines W sont capables d'impacter la transcription et réplication virale (Durbin *et al.*, 1999; Roth *et al.*, 2013).

### (3) Protéine C

Les protéines C sont présentes dans 7 des 8 genres composant la sous-famille des *Orthoparamyxovirinae*.

Peu d'informations sont aujourd'hui connues sur la structure des protéines C. Cependant, les prédictions de structures secondaires basées sur leur séquence en aa semblent indiquer que tous les domaines C-terminaux des protéines C sont riches en hélices- $\alpha$  tandis que les domaines N-terminaux sont majoritairement désordonnés à l'exception de potentielles régions  $\alpha$ -MoRE (Lo, Sogaard and Karlin, 2014). Les poids moléculaires des protéines C en prédiction iraient de 17 à 27 kDa (Siering, Cattaneo and Pfaller, 2022).



**Figure 21 : Structure de la protéine C de Sendai en complexe. (A) avec STAT1 (PDB 3WWT) et (B) avec le domaine Bro1 de la protéine ALIX (PDB 6KP3)** (Oda *et al.*, 2015, 2021). (C) Les deux structures précédentes de la protéine C du SeV ont été superposées par PYMOL et le RMSD obtenu est de 0,431 Å.

À l'heure actuelle, seule la structure du domaine C-terminal ordonné (appelée Y3 comprenant les résidus 98 à 204) de la protéine C du SeV a été résolue par cristallographie à deux occasions en complexe avec deux partenaires cellulaires différents : avec la protéine STAT1, un facteur de transcription impliqué dans la réponse IFN (PDB 3WWT) (Oda *et al.*, 2015), et avec le domaine Bro1 de la protéine ALIX (ALG2-interacting protein X), un composant du complexe de tri endosomal requis pour le transport (ESCRT) (PDB 6KP3) (Oda *et al.*, 2021).

Dans le premier cas, il a été montré que deux molécules Y3 se lient au dimère formé par le domaine N-terminal de STAT1 (STAT1ND) (**Figure 20A**). La structure cristallographique du complexe a été obtenue à une résolution de 2.0 Å. La protéine Y3 est

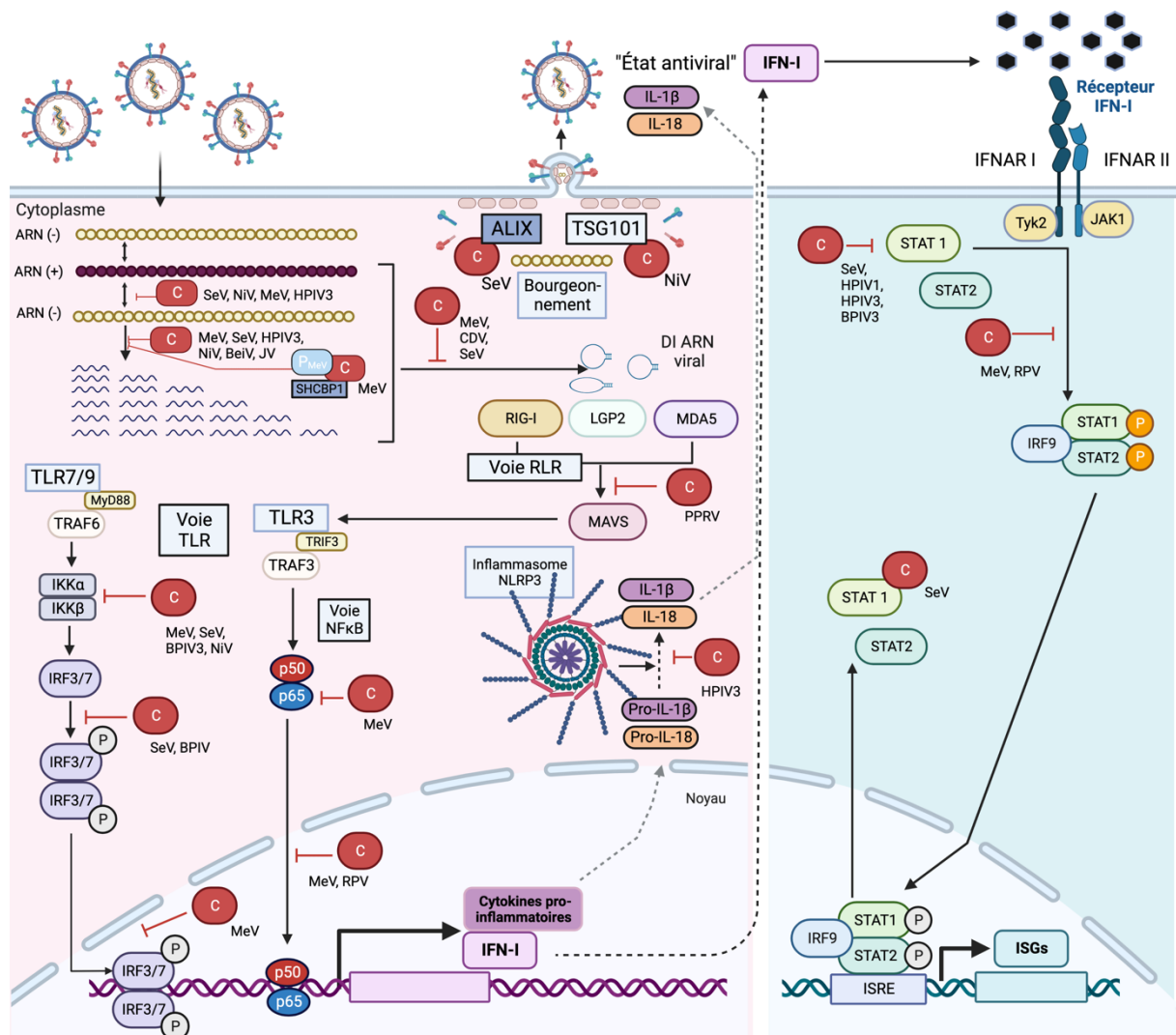
composée de 6 hélices  $\alpha$  ( $\alpha 1$  à  $\alpha 6$ ) formant une bicouche structurale maintenue majoritairement par des interactions hydrophobes. Le premier niveau regroupe les hélices  $\alpha 1$ ,  $\alpha 3$  et  $\alpha 4$  et le second les hélices  $\alpha 2$ ,  $\alpha 5$  et  $\alpha 6$ . Cette étude suggère un rôle clé de la protéine Y3 sur l'inhibition de l'activité de STAT1 durant l'infection par SeV.

La deuxième structure montre que la protéine C du virus Sendai est capable d'interagir avec le domaine Bro1 de la protéine Alix (**Figure 21B**). La structure cristallographique du complexe a été obtenue à une résolution de 2.2 Å. La structure de Y3 associée au domaine Bro1 peut être super-imposée à celle du complexe avec STAT1ND, à l'exception des hélices  $\alpha 4$  et  $\alpha 5$  dont la position varie légèrement. La superposition génère une valeur de RMSD de 0.431 Å suggérant que la protéine ne change pas de conformation après liaison à ALIX ou à Bro1 (**Figure 21C**). L'analyse par mutagenèse de l'effet de l'interaction de SeV C avec ALIX a montré que la protéine C est capable de favoriser le bourgeonnement des particules virales par cette interaction.

Les protéines C des *Orthoparamyxovirinae* partagent de nombreuses fonctions communes malgré une indéniable faible similarité de séquences (< 20%).

## II. Les fonctions des protéines C des *Orthoparamyxovirinae*

Les protéines C sont des facteurs clés de la pathogénicité des *Orthoparamyxovirinae*. Leurs fonctions s'expriment à divers moments du cycle viral incluant la régulation de la transcription et de la réplication virale, la régulation de la processivité de la polymérase virale, le contrôle de la réponse immunitaire innée et l'assemblage et le bourgeonnement des particules virales. À travers cette revue bibliographique, nous avons cherché d'une part à organiser et résumer la multitude de fonctions jouées par les protéines C répertoriées mais également à insister sur les similarités de ces fonctions et des voies cellulaires où le rôle des protéines C est essentiel pour le cycle viral. Nous avons, pour cela, construit un schéma résumé vulgarisant les rôles des protéines C dans la cellule au cours d'une infection par un paramyxovirus (**Figure 22 et annexe 2**). Ce schéma est disponible en annexe 2 et peut être découpé pour servir de support pour la lecture de ce chapitre.



**Figure 22 : Fonctions des protéines C des *Orthoparamyxoviridae* : dans la régulation et l'inhibition de la réplication et transcription virale, dans le contrôle du ratio génome/antigénome, dans l'antagonisme du système immunitaire et dans le bourgeonnement viral (figure adaptée de (Siering, Cattaneo and Pfaller, 2022) et (Pelissier, Iampietro and Horvat, 2019)). Les protéines C sont identifiées par les cercles rouges et le virus où la fonction décrite a été identifiée est précisé à droite sous le cercle. Figure réalisée dans BioRender.**

### **(A) Localisation cellulaire : navette entre le noyau et le cytoplasme ?**

Le cycle de réplication des *Orthoparamyxovirinae* se produit entièrement dans le cytoplasme. Cependant, plusieurs études ont souligné le rôle de plus en plus important du trafic nucléo-cytoplasmique et/ou de la localisation nucléaire de certaines protéines virales durant la réplication (Audsley, Jans and Moseley, 2016b). La navette des protéines accessoires entre le noyau et le cytoplasme s'est révélée jouer un rôle majeur permettant aux protéines accessoires d'interagir dans le noyau des cellules avec les facteurs de transcription impliqués dans la signalisation antivirale. Néanmoins, la localisation cellulaire des protéines C soulève encore de nombreuses questions.

Il a notamment été montré que la protéine C du MeV présente une localisation dynamique. Elle est retrouvée dans le noyau au début de l'infection et transite par la suite entre le noyau et le cytoplasme (Nishie, Nagata and Takeuchi, 2007). Des signaux de localisation nucléaire (NLS) (<sup>41</sup>PPARKRRQ<sup>48</sup>) et d'exportation nucléaire (NES) (<sup>75</sup>DLEKAMTTLKLWE<sup>87</sup>) ont été cartographiés dans la séquence. Des mutations de la séquence NLS peuvent altérer la capacité de la protéine C du MeV à antagoniser l'induction des IFN pendant l'infection (Shaw *et al.*, 2005; Sparrer, Pfaller and Conzelmann, 2012; Irie, Yoshida and Sakaguchi, 2013). Ces mutations ont été identifiées dans une souche vaccinale atténuée (Schwarz MeV) pour laquelle on observe une réduction de la localisation nucléaire de MeV C (Nakatsu *et al.*, 2009).

Cette observation a été retrouvée avec la protéine C du NiV qui est d'abord détectée dans le cytoplasme lors d'une infection (Lo *et al.*, 2009). Des études récentes ont ensuite observé la capacité de cette protéine à circuler entre le noyau et le cytoplasme (Audsley, Jans and Moseley, 2016b; Horie *et al.*, 2016). Les régions C et N-terminales de la protéine C du NiV ainsi que la conformation des résidus 1 à 24 et 110 à 139 sont nécessaires et suffisantes pour maintenir la protéine C dans le cytoplasme (Horie *et al.*, 2016). Cependant, il a également été montré que les résidus 60 à 67 et 72 à 75 de cette protéine peuvent induire son transfert vers le noyau. Ces motifs NES et NLS ont été également retrouvés dans la séquence de la protéine C du SeV. Cependant ils ne semblent pas conservés à travers les différents genres et pourraient suggérer une évolution indépendante de la localisation de ces protéines.

La protéine C du virus Rinderpest (RPV) semble, quant à elle, exclusivement localisée dans le noyau (Boxer, Nanda and Baron, 2009) de cellules transfectées. Cette localisation a été mise en parallèle avec sa capacité à inhiber l'activation des promoteurs IFN-I.

### ***(B) Rôle dans la pathogénicité virale - immuno-évasion virale***

Le développement des technologies de génétique inverse a permis d'identifier plusieurs fonctions des protéines C dans l'immuno-évasion virale. Leur importance dans la pathogénicité virale a également été mise en évidence pour divers modèles animaux.

Il a notamment été observé qu'un virus NiV recombinant portant une mutation empêchant l'expression de la protéine C (NiVΔC) pouvait se répliquer en cellule, suggérant que C pourrait être une protéine non essentielle à la réplication virale (Mathieu *et al.*, 2012). Cette observation est retrouvée pour d'autres virus recombinants de la famille tels que les *Respirovirus* et plusieurs *Morbillivirus* (Garcin, Itoh and Kolakofsky, 1997; Kurotani *et al.*, 1998; Baron and Barrett, 2000; Bartlett *et al.*, 2010) pour certaines lignées cellulaires.

Ce virus NiVΔC est sévèrement atténué *in vivo*. Une première étude a observé 100% de survie chez des hamsters infectés (Yoneda *et al.*, 2010), résultat modéré par une seconde étude qui n'observe qu'une atténuation partielle du virus avec 16.7% à 66.7% de survie en fonction de la dose de virus utilisée pour l'infection (Mathieu *et al.*, 2012). Cette étude



démontre que NiV $\Delta$ C induit la génération d'une réaction pro-inflammatoire et d'une réponse immunitaire adaptative adéquate à la suite de l'infection. Ce résultat suggère que la protéine C du NiV joue un rôle essentiel dans la pathogénicité virale par la régulation de l'expression des cytokines pro-inflammatoires. Cette conclusion fait écho à de nombreux autres résultats montrant que les virus déficients pour la protéine C montrent généralement une forte atténuation *in vitro* et *in vivo* (Kurotani *et al.*, 1998; Gotoh *et al.*, 1999; Takeuchi *et al.*, 2005; Komatsu; Siering, Sawatsky and Pfaller, 2021). De manière intéressante, nous avons observé que les protéines C des souches BD et MY de NiV responsables des épidémies, respectivement, au Bangladesh et en Malaisie sont identiques à près de 95%. Des différences sont observés dans la région N-terminale et pourraient être à l'origine de la variation des modes de transmission, de la durée d'incubation et de la pathogénicité des souches virales (Mire *et al.*, 2016). En effet, sous des conditions expérimentales identiques, après exposition de singes verts africains à l'une des deux souches virales, la souche NiV-BD s'est avéré plus pathogénique causant 100% de mortalité en modèle animale pour 50% pour la souche NiV-MY.

À l'inverse de ces résultats, une étude de la contribution de la protéine C du NiV à la pathogénèse chez les furets a produit des résultats contradictoires aux deux études précédentes. Il a été montré que des furets infectés par un virus NiV $\Delta$ C ne développaient pas de symptômes respiratoires mais uniquement de forts symptômes neurologiques avec 100% de mortalité à long terme (Satterfield *et al.*, 2016). Cette étude conclut que la protéine C du NiV ne joue pas un rôle majeur dans la composante respiratoire de la maladie dans le cas du modèle animal furet.

Le rôle essentiel de la protéine C du virus de la maladie de Carré, le canine distemper virus (CDV), a été étudié *in vitro* et *in vivo* dans le modèle furet (Siering, Sawatsky and Pfaller, 2021). Cette étude a démontré qu'il existe, de la même façon que pour le virus Sendai, de multiples codons initiateurs le long des ARNm codant pour les protéines P permettant l'expression d'une diversité de protéines C de tailles différentes. Cette observation pourrait expliquer le manque d'homogénéité de certaines études pour lesquelles l'entièreté de l'expression des fragments de protéines C n'aurait pas été correctement éteinte. L'infection de furets par un CDV recombinant C KO pour tous les fragments de protéines C est suivie du rétablissement des animaux trois semaines après infection. Ces résultats sont similaires à ceux observés pour un virus recombinants MeV déficients pour la protéine C suggérant l'existence de fonctions communes dans l'inhibition de la réponse immunitaire et la réplication de l'ARN viral (Pfaller *et al.*, 2019). Cette étude a également démontré que les 43 résidus N-terminaux de la protéine C de MeV ne sont pas essentiels à l'expression de ces fonctions et il est suggéré que le même phénomène existe pour la protéine C du CDV, soulevant l'existence d'un rôle essentiel du domaine C-terminal des protéines C des morbillivirus pour la pathogénicité (Pfaller *et al.*, 2019; Siering, Sawatsky and Pfaller, 2021; Siering, Cattaneo and Pfaller, 2022).

### ***(C) Rôle dans la réplication virale et transcription virale***

Les protéines C sont impliquées dans plusieurs voies de régulation de la réplication et de la transcription virale. Les mécanismes, toujours à l'étude, restent encore incompris.

### (1) Modulation de l'activité de l'ARN polymérase virale

La protéine C du NiV est capable de moduler l'activité de la polymérase virale ARN-ARN dépendante. L'analyse de l'évolution de l'expression du gène rapporteur du minigénome de NiV révèle que les protéines C, V et W du NiV inhibent son expression de manière dose-dépendante. La quantité totale d'ARN (+) poly(A) et d'ARN résistants aux nucléases micrococcales diminuent en présence de la protéine C du NiV (Sleeman *et al.*, 2008). De plus, cette fonction est partagée par certains paramyxovirus puisque la protéine C du NiV est également capable d'inhiber la réplication du minigénome de MeV mais pas celle de HPIV3. La protéine C du MeV inhibe également l'expression du gène rapporteur du minigénome de MeV (Reutter *et al.*, 2001).

Dans le cas du SeV, l'existence de multiples protéines C complexifie l'interprétation de leur rôle dans la régulation de la synthèse de l'ARN viral. En effet, l'étude de la transcription dans le cas du SeV suggère que les protéines C, bien que jouant un rôle dans la régulation et la réplication, ont des rôles qui ne sont pas fonctionnellement équivalents (Latorre *et al.*, 1998). C' et C semblent capables de réguler négativement la synthèse de l'ARN viral, fonction qui ne semble pas partagée pour les protéines Y1 et Y2. L'infection de cellules par des virus recombinants n'exprimant pas C' (SeVC'KO) ou C (SeVCKO) entraîne un retard dans l'accumulation d'ARN viraux dans les premières phases de l'infection. Seuls les virus exprimant C' conservent une haute virulence en souris. De manière intéressante, les codons initiateurs des protéines C et C' sont situés en amont des protéines Y1 et Y2. C' et C possèdent donc une région N-terminale plus longue, probablement impliquée dans la régulation de la synthèse de l'ARN viral.

### (2) Altération du gradient de transcription

Il a été observé que les résidus conservés en position 147 et 166 de la protéine C du MeV sont impliqués dans la modulation de l'activité de la polymérase virale ARN-ARN dépendante. Cette modulation altère le gradient de transcription des différents ARNm viraux (Reutter *et al.*, 2001; Bankamp *et al.*, 2005).

Les protéines C du SeV (Curran, Richardson and Kolakofsky, 1986; Tapparel *et al.*, 1997; Garcin, Curran and Kolakofsky, 2000; Kato *et al.*, 2004) et du HPIV3 (Smallwood and Moyer, 2004) sont également capables d'inhiber spécifiquement la transcription virale. Sur le même sujet, lors d'infections par des SeV recombinants déficients ou partiellement déficients pour la protéine C, il est constaté que la production d'ARNm augmente par rapport au virus sauvage.

Des MeV recombinants déficients pour la protéine C présentent un gradient de transcription plus abrupt que le virus parental (Pfaller *et al.*, 2014). Des altérations des

niveaux de transcriptions ont été observées, de la même façon, dans le cas du NiV, du HPIV3, du virus Jeilong (JV) et du virus Beilong (BeiV) (Malur, Hoffman and Banerjee, 2004; Magoffin, Mackenzie and Wang, 2007; Sleeman *et al.*, 2008; Malur *et al.*, 2009; Wells, Addington-Hall and Malur, 2012).

### (3) Modulation du rapport génome/antigénome

Il a été observé, dans le cas du SeV, que pour des virus recombinants n'exprimant pas soit la protéine C', soit la protéine C ou aucune des 4 protéines C, l'expression de virions non-infectieux contenant des ARN antigénomiques était prédominant en comparaison aux virions infectieux contenant de l'ARN génomique. Ce résultat indique que les protéines C sont également capables de réguler l'équilibre génome-antigénome viral au cours de la réplication (Irie *et al.*, 2008a).

Cette fonction peut être mise en parallèle avec des résultats précédents qui suggéraient une action de la protéine C sur le contrôle de la sélectivité du complexe de réplication pour les éléments régulateurs en cis du promoteur (séquence *leader* et *trailer*) (Tapparel *et al.*, 1997). Plusieurs mécanismes, décrits dans la partie suivante, sont proposés à l'heure actuelle.

### (4) Interaction avec le complexe polymérase

Un premier modèle suggérant une interaction directe entre les protéines C et L de SeV a été proposé (Grogan and Moyer, 2001). En effet, la protéine C de SeV colocalise avec les ribonucléocapsides et cette association, démontrée par essais pull-down (Horikami *et al.*, 1997), est nécessaire à l'inhibition de la synthèse de l'ARN viral. L'interaction a été cartographiée à plusieurs aa C-terminaux chargés indépendants du site d'interaction de la P et de la L (Grogan and Moyer, 2001). L'hypothèse retenue aujourd'hui est que l'interaction de la protéine C pourrait stabiliser ou réarranger le complexe L-P affectant ainsi les fonctions de réplication de l'ARN viral et de synthèse du génome-antigénome.

Un autre modèle a été proposé dans le cas de la protéine C du MeV dont la régulation de la synthèse de l'ARN viral semble reposer sur une interaction indirecte avec le facteur de croissance de la cellule hôte SHCBP1 (SHC SH2-binding protein 1). Ce facteur peut interagir avec la protéine P du MeV (Ito *et al.*, 2013). Il est donc supposé que l'interaction SHCBP1-C peut empêcher l'interaction de P avec SHCBP1 causant l'inhibition de la synthèse de l'ARN viral. Les mécanismes par lesquels l'interaction P-SHCBP1 favorise la synthèse de l'ARN viral sont inconnus. De manière étonnante, l'étude récente de l'interactome de la protéine C du MeV n'a pas retrouvé SHCBP1 dans les interactants cellulaires potentiels de cette protéine (Meignié *et al.*, 2021).

Par ailleurs, il a été montré que la protéine C du MeV colocalise avec les LLPS formées par les protéines N, P et L notamment au travers d'une interaction physique directe avec la phosphoprotéine et donc avec le complexe de réplication de l'ARN viral (Pfaller *et al.*, 2019). Le site d'interaction avec la P de MeV a notamment été cartographié aux hélices C-terminales de la protéine C (Pfaller *et al.*, 2019).

Cette interaction régule probablement le processus d'encapsidation de l'ARN néoformé (Nishie and Nagata, 2015; Pfaller *et al.*, 2019). En déstabilisant l'encapsidation, la protéine C pourrait provoquer la terminaison abrupte de la synthèse de l'ARN viral.

## (5) Rôle dans l'inhibition de la production de DI RNA immunostimulateurs

Il a été observé que l'infection par un MeV déficient pour la protéine C entraîne une augmentation de l'expression d'ARN défectifs interférents. Ces ARN forment des ARN double-brins *in vitro*. Ces structures sont reconnues par la réponse immunitaire innée de l'hôte, ce qui entraîne alors l'activation de mécanismes de défense comme la réponse au stress médiée par la protéine kinase R (PKR) et la production d'interféron de type I médiée par le gène MDA-5 (Sparrer, Pfaller and Conzelmann, 2012; Pfaller *et al.*, 2014, 2015).

Cette observation a été faite pour d'autres *Morbillivirus* comme le CDV (Siering, Sawatsky and Pfaller, 2021). Le résultat est également observé *in vitro* et *in vivo* (modèle souris) pour un virus SeV recombinant dont l'expression des 4 formes de la protéine C est éteinte (Kato *et al.*, 2007).

### ***(D) Interaction directe avec la réponse immunitaire innée***

Les protéines C interagissent à de nombreux niveaux avec le système immunitaire. Elles interfèrent notamment directement avec les protéines responsables de l'induction de la réponse IFN de type I prévenant de cette façon l'installation d'une réponse inflammatoire.

## (1) Induction des IFN

La participation de la protéine C à l'inhibition de la production d'IFN a été observée chez les *Henipavirus*, les *Respirovirus* et les *Morbillivirus*.

Les protéines C des *Orthoparamyxovirinae* sont capables de contrecarrer la production d'IFN en interagissant avec IKK $\alpha$ , ce qui entraîne le blocage de l'induction des IFN dépendant des récepteurs TLR7/9 (Yamaguchi *et al.*, 2014).

La première activité antagoniste des IFN des protéines C de paramyxovirus a été découverte lors d'une étude sur la protéine C du SeV. Un SeV déficient pour les 4 protéines C induit une forte réponse IFN liée à la surproduction d'ARNdb par ce virus (Koyama *et al.*, 2003). L'expression ectopique de la protéine C du SeV inhibe la phosphorylation et la dimérisation du facteur de régulation des IFN (IRF3) sur des cellules soumises à des stimuli étrangers activateurs de la réponse IFN (traitement au poly(I:C) (synthetic polyinosinic-polycytidylic acid double-stranded RNA)) (Komatsu *et al.*, 2004; Komatsu, Takeuchi and Gotoh, 2007; Strahle *et al.*, 2007). Le pré-traitement des cellules à la protéine C favorise la croissance de virus sensibles aux interférons comme le NDV ou le virus de la stomatite vésiculaire (VSV).

Dans le cas des *Morbillivirus*, les protéines C sont également capables d'interférer avec la production d'IFN par l'interaction avec les gènes de transcription des IFN-I dans le

noyau des cellules infectées (Boxer, Nanda and Baron, 2009; Sanz Bernardo, Goodbourn and Baron, 2017). Il a été montré, par exemple, que des souches virales de MeV incapables d'entrer dans le noyau sont également incapables d'inhiber la transcription IRF3-dépendante des IFN-I (Sparrer, Pfaller and Conzelmann, 2012). De plus, la protéine C du MeV serait capable d'interagir avec la sous-unité p65 du facteur de transcription NFκB (Meignié *et al.*, 2021).

## (2) Interaction avec la cascade de signalisation de la réponse IFN

De nombreux paramyxovirus interagissent avec la cascade de signalisation des IFN prévenant l'expression des ISGs dans les cellules infectées.

Pour le SeV, il a été montré que la protéine C interférait avec la cascade de signalisation de l'IFN en interagissant directement avec STAT1 mais aussi de manière indépendante à STAT1, possiblement en interagissant avec le récepteur IFNAR2. Cette interaction empêcherait l'activation des voies de signalisation IFNAR-JAK-STAT et TYK2 (Gotoh *et al.*, 1999; Takeuchi *et al.*, 2001).

La résolution de la structure de SeV C en interaction avec STAT1 (PDB 3WWT) a permis d'étudier les mécanismes de l'inhibition de la réponse IFN (Oda *et al.*, 2017).

La protéine C se lie à STAT1, empêchant la phosphorylation et la déphosphorylation de son résidu tyrosine 701, étape critique pour l'activation de STAT1 (Takeuchi *et al.*, 2001; Garcin *et al.*, 2002; Komatsu *et al.*, 2004; Oda *et al.*, 2017). Le dimère formé par STAT1 présente deux arrangements différents : une forme parallèle et une forme antiparallèle. Le dimère STAT1 sous forme antiparallèle peut facilement être déphosphorylé par la phosphatase cellulaire. Sous la forme monomérique, le résidu Tyr 701 est phosphorylé de façon stable. Il a été montré que le dimère STAT1 interagissant avec une seule molécule Y3 (résidus 98 à 204 de la protéine C du SeV) serait majoritairement sous la forme antiparallèle et celui avec deux molécules Y3 serait majoritairement sous la forme parallèle. La concentration de protéine C produite durant une infection serait donc un facteur clé dans le contrôle de la réponse immunitaire par la voie IFNAR-JAK-STAT. Les protéines C' et C du SeV sont capables d'induire la dégradation de STAT1, tandis que Y1 et Y2 empêchent la phosphorylation de STAT1 sans activer sa dégradation (Garcin *et al.*, 2001, 2004; Gotoh *et al.*, 2003).

Pour le NiV, ces fonctions sont régulées par les protéines V et W du NiV (Malur *et al.*, 2005; Kitagawa *et al.*, 2011). Cette fonction ferait intervenir le domaine N-terminal des protéines C tandis que la liaison avec STAT1 se fait par le domaine C-terminal (Mao, Chattopadhyay and Banerjee, 2009).

## (3) Inflammation

La réponse inflammatoire est essentielle pour l'élimination des virus. Lors d'infection de cellules endothéliales par un virus recombinant NiVΔC, une augmentation de l'expression de chimiokines et de cytokines pro-inflammatoires est observée (Mathieu *et al.*, 2012)

entraînant une pathogénicité amoindrie de ce virus en comparaison au virus sauvage chez le hamster.

La capacité de la protéine C de HPIV3 à se lier à l'inflammasome NLRP3 provoquant sa dégradation par le protéasome limite son activation et donc la réponse inflammatoire menant à la mise en place d'un état anti-viral des cellules (Shil *et al.*, 2018).

### ***(E) Rôle dans l'assemblage des particules virales et le bourgeonnement***

Les protéines C des genres *Respirovirus* et *Henipavirus* jouent un rôle dans le bourgeonnement viral (Sugahara *et al.*, 2004; Park *et al.*, 2016).

Il a été suggéré que la protéine C du NiV favorise un assemblage efficace des particules virales et leur bourgeonnement par le biais d'interactions avec les complexes de tri endosomal requis pour le transport (ESCRT) et la protéine de matrice (M) (Park *et al.*, 2016). La recherche de modèles d'homologie structurale sur la protéine C du NiV a révélé qu'un segment important de la protéine C, appelé domaine MiD (résidus 47 à 127), s'aligne avec le domaine C-terminal (CTD) du facteur cellulaire VPS28, un adaptateur majeur des sous-unités des sous-complexes ESCRT-I et ESCRT-II du système ESCRT. Ces complexes facilitent le transport de protéines via des corps multivésiculaires et le tri endosomal. Il a également été découvert que le domaine CTD de la protéine C pouvait interagir avec le domaine CTD de TSG101, un composant du sous-complexe ESCRT-I dont le rôle critique dans le bourgeonnement d'autres virus a déjà été démontré (Park *et al.*, 2016). Cette interaction repose sur deux résidus (W103 et L104) conservés chez les *Henipavirus* et également certains *Respirovirus*.

La protéine C du SeV interagit avec le facteur cellulaire ALIX, une protéine adaptatrice connectant le complexe ESCRT-I avec l'ESCRT-III (Sakaguchi *et al.*, 2005, Irie *et al.*, 2008b), via un motif LXXW conservé au sein des *Respirovirus*. Cette interaction a été révélée par la résolution de la structure cristallographique du domaine CTD de la protéine Y3 de SeV en complexe avec le domaine BRO-1 de ALIX (PDB 6KP3) (Irie *et al.*, 2008b). Par ailleurs, il a été observé que les 23 résidus N-terminaux des protéines C et C' du SeV sont nécessaires pour leur adressage et ancrage à la membrane plasmique. Ces résidus pourraient donc jouer un rôle dans l'induction du bourgeonnement des particules virales par ces protéines (Sakaguchi *et al.*, 2005; Marq *et al.*, 2007). L'interaction avec la protéine ALIX a été également retrouvée pour le virus HPIV1 (Boonyaratanakornkit *et al.*, 2013).

Pour de nombreux paramyxovirus, il a été observé que la communication directe d'une cellule infectée à une autre repose également sur la formation de syncytia par fusion entre cellules infectées exprimant les glycoprotéines de surface et les cellules naïves exprimant les récepteurs viraux (Aguilar *et al.*, 2016).

**Le constat de la revue bibliographique précédente souligne, que malgré une diversité de séquence apparente, les protéines C sont impliquées dans près de 12 fonctions pour seulement 4 genres viraux. Toutes ces fonctions semblent, par différents**

**mécanismes, jouer un rôle essentiel dans la pathogénicité par l'inhibition de la réponse immunitaire et/ou par la modulation des étapes du cycle viral (transcription, réplication ou bourgeonnement viral).**

### **III. Étude bio-informatique des protéines C des *Orthoparamyxovirinae***

Que l'on considère la séquence nucléotidique du gène de la P dont dérive la protéine C ou les séquences en aa des protéines P et C, les similarités de séquences de ces protéines sont très faibles. Les résultats précédents, qui démontrent que les protéines C semblent toutes impliquées dans des voies d'antagonisme de la réponse immunitaire commune sont, de ce fait, intrigants. Nous nous sommes demandés si ces similarités pouvaient être expliquées par le partage de motifs structuraux communs à ces protéines.

De plus, nombre des nouveaux paramyxovirus découverts récemment se sont révélés difficiles à classer au sein de la phylogénie actuelle de la sous-famille. Les phylogénies sont traditionnellement établies sur des critères historiques comme la séquence en nucléotide de gènes conservés comme le gène L, la nature des glycoprotéines de surface ou le spectre d'hôte. Ces critères de classification ont tous été remis en cause récemment, car leur utilisation ne permet plus de classer avec certitude ces nouveaux virus (Rima *et al.*, 2018; Wells *et al.*, 2022). Or, le nombre potentiel de paramyxovirus à reclasser pourrait bien être très élevé dans les années à venir, car il existe une grande diversité de séquences dans les réservoirs de la faune sauvage. Améliorer la classification des familles de virus comme les paramyxovirus s'avère de plus en plus pressant car cela nous aidera à déterminer les facteurs qui peuvent contribuer à l'émergence de nouveaux paramyxovirus par la caractérisation de leurs mécanismes de transmission inter-espèces. Ces connaissances sont essentielles pour limiter les impacts futurs de ces agents pathogènes.

La résolution de la structure de protéines telles que les protéines C, protéines qui a priori possèderaient une histoire évolutive plus récente que le gène P dont elles sont issues, pourrait donc permettre d'établir un nouveau critère de classification pour comprendre l'histoire évolutive récente des *Orthoparamyxovirinae*.

#### ***(A) Quelle histoire évolutive les protéines C partagent-elles ?***

À l'heure actuelle, seule une structure de protéine C en complexe a été résolue. Les protéines C possédant, d'une part, des séquences distantes les unes des autres et, d'autre part, étant exprimées par chevauchement depuis le gène P, il a été initialement supposé qu'elles auraient pu émerger à plusieurs reprises dans la sous-famille. L'absence de protéines C dans le genre *Ferlavirus* des *Orthoparamyxovirinae* renforce de plus cette hypothèse.

Cependant, une autre hypothèse a déjà été évoquée dans plusieurs travaux précédents : celle de l'existence d'un ancêtre commun aux protéines C qui, soit serait apparu à l'embranchement des *Orthoparamyxovirinae*, soit qui serait antérieur et commun aux

*Paramyxoviridae* et aurait été perdu chez les lignées des avulavirus et pneumovirus (Jordan, Sutter IV and McClure, 2000; Sabath, Wagner and Karlin, 2012; Lo, Søggaard and Karlin, 2014). Une étude réalisée par Lo *et al* (Lo, Søggaard and Karlin, 2014) dévoilait pour la première fois l'existence de fortes similarités de structure partagées par toutes les protéines C des *Orthoparamyxovirinae*. Cette analyse démontre l'existence de liens d'homologie entre les protéines C de plusieurs genres renforçant l'hypothèse d'un ancêtre commun. Pour autant, Lo et al souligne que, bien qu'il ne soit pas possible de distinguer avec une complète certitude qui de P ou de C est ancestrale, il est plus probable que la phosphoprotéine (P) ait une origine antérieure à la protéine C au vu de son rôle indispensable à la réplication des paramyxovirus.

Il a été constaté que la séquence protéique des protéines C de la famille est plus conservée que celle des protéines P (Jordan, Sutter IV and McClure, 2000). L'analyse du ratio dN/dS (proportion de mutation non synonyme/proportion de substitution synonymes) qui permet d'évaluer la sélection naturelle qui s'applique à une séquence codante suggère une sélection négative s'opérant sur les codons du cadre de lecture de la protéine C par rapport à ceux du cadre de lecture de P. L'évolution de la protéine C est donc à priori plus contrainte que l'évolution de la région correspondante de la protéine P. Cette observation est contre-intuitive et va à l'encontre des résultats de l'étude de Sabath et al (Sabath, Wagner and Karlin, 2012) qui démontre qu'en général, les gènes ancestraux sont plus contraints sur leurs séquences que les gènes *de novo* qui les chevauchent.

Nous avons souhaité, sur la base de la méthodologie mise en œuvre dans les études de phylogénie précédentes, mettre à jour les connaissances en terme de similarités de séquence et de structure des protéines C, en nous basant sur la taxonomie de l'ICTV (International Committee on Taxonomy of Viruses) de la sous-famille datant de 2019 incluant des paramyxovirus découverts dans la dernière décennie. Ces analyses ont également été complétées par des paramyxovirus découvert récemment (Wells *et al.*, 2022). Ces analyses nous permettront, d'une part, de vérifier la conservation des conclusions phylogénétiques des études précédentes et, d'autre part, d'établir une méthodologie pour l'étude structurale des protéines C prévue dans le cadre de notre projet.

### ***(B) Une évidente absence de conservation de séquence au sein des protéines C***

Nous avons cherché en première approche à vérifier si les homologies détectées par Lo *et al* (Lo, Søggaard and Karlin, 2014) entre certains genres étaient retrouvées pour les paramyxovirus récents. Les outils classiques d'alignements de séquence ne sont pas basés sur la recherche d'homologies et présentent de fortes limitations lorsque les séquences comparées sont trop éloignées les unes des autres, comme c'est le cas pour les protéines C. Cependant, les séquences de protéines homologues ne partagent pas toujours des similarités de séquences significatives et il est possible, soit de trouver des séquences dans des bases de données avec lesquelles les protéines en question partagent de fortes similarités, soit de retrouver des homologies, par la construction d'alignements de profils par la stratégie d'extension d'homologie.



Cette stratégie permet d'identifier des homologies entre séquences éloignées. Elle repose sur la génération d'une librairie d'alignements multiples de séquences par la méthode BLAST suite à la recherche préalable, pour chacune des séquences de l'alignement, de séquences homologues dans les bases de données de séquences disponibles. Les alignements BLAST qui en résultent sont ensuite transformés en profils. Par rapport aux séquences individuelles, la génération de profils permet d'accumuler des informations sur quel acide aminé est toléré à chaque position et avec quelle probabilité (Lo, Søgaaard and Karlin, 2014). Ces profils sont alors comparés deux à deux et l'alignement le plus cohérent est ajouté à la librairie. L'alignement de profils permet de mieux refléter les contraintes structurales et fonctionnelles d'une famille de protéines lors de la construction de l'alignement multiple de séquence. Cette librairie est ensuite transformée en un alignement multiple de séquences auquel est attribuée une valeur d'index de fiabilité appelée CORE (Consistency of Overall Residue Evaluation). L'alignement multiple présentant le score de cohérence le plus élevé est ensuite conservé.

Nous avons utilisé la méthode de calcul PSI-coffee (Position Specific Iterative) permettant de générer l'alignement progressif le plus cohérent pour des séquences de protéines éloignées les unes des autres en combinant la méthode d'extension d'homologie de BLAST et l'alignement progressif basé sur la cohérence (Poirot, O'Toole and Notredame, 2003; Floden *et al.*, 2016). Les abréviations utilisées pour les noms des virus dans les figures suivantes et le reste de ce manuscrit ainsi que les identifiants des séquences utilisées sont détaillées en annexe 1.

Nous avons généré des alignements de séquences intra-genre et intergenre pour les protéines C des *Orthoparamyxovirinae*. La comparaison des valeurs de CORE des alignements générés nous a permis de regrouper les séquences des protéines C en trois groupes, concordant avec ceux de l'étude évolutive réalisé par Lo *et al* (Lo, Søgaaard and Karlin, 2014) :

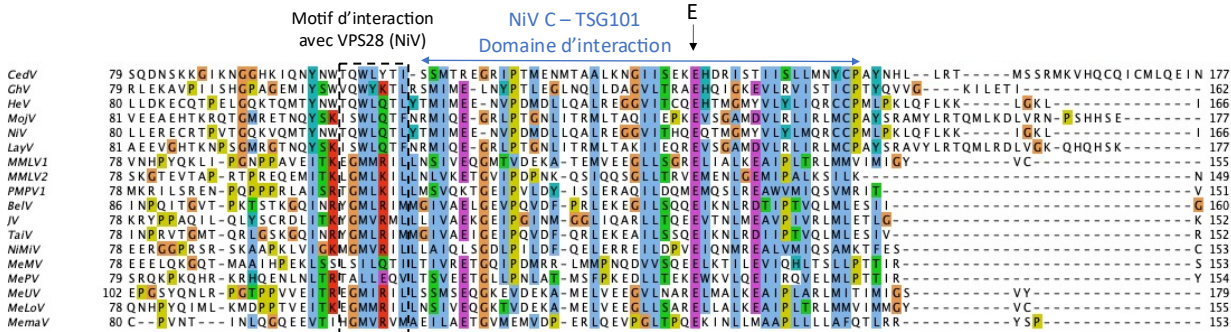
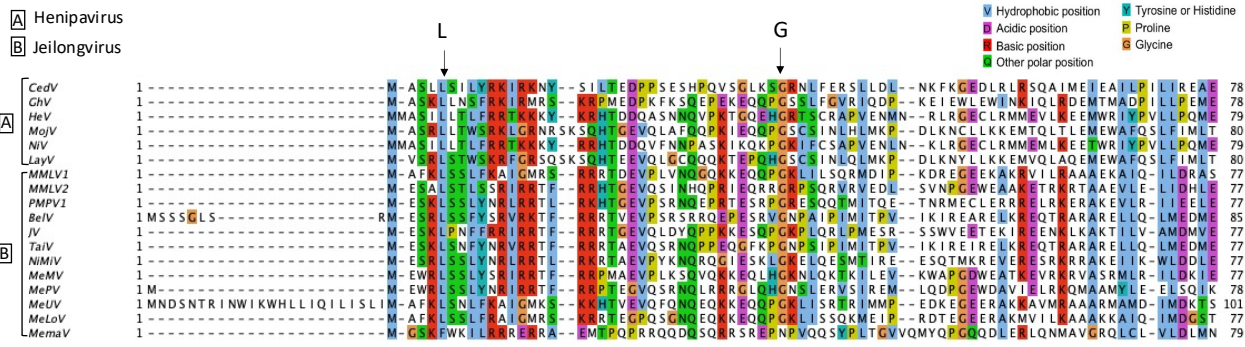
- le groupe I : *Henipavirus* et *Jeilongvirus*
- le groupe II : *Narmovirus*, *Salemvirus* et *Morbillivirus*
- le groupe III : *Respirovirus* et *Aquaparamyxovirus*.

Les alignements des protéines C de chacun des groupes ont ensuite été comparés deux à deux par la méthode d'alignement des profils HHpred de la suite HHalign (Soding, 2005). HHpred suggère que les protéines C du premier et du deuxième groupe sont significativement similaires ( $E_{\text{value}} = 3,5E10^{-8}$ ) sur une région allant de la position 24 à 148. La méthode ne détecte aucune similarité de ces deux groupes avec le troisième groupe.

Les alignements des protéines au sein des groupes I et II sont présentés dans les **Figure 23 et 24**.

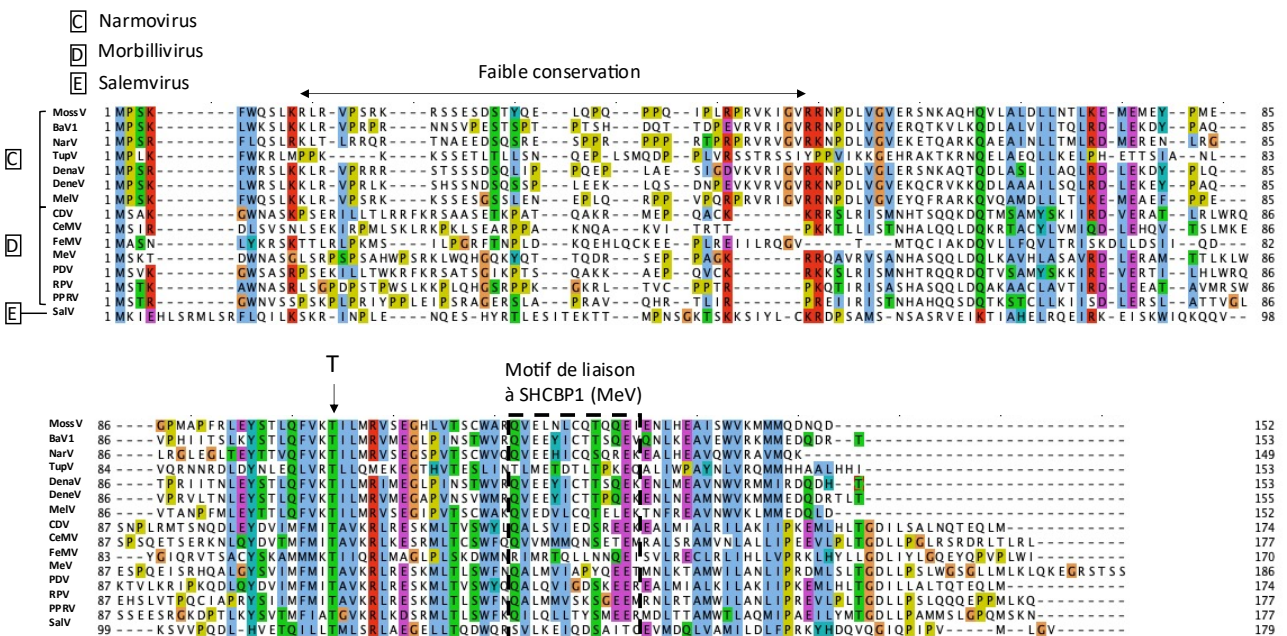
Dans le cas du groupe I (**Figure 23**), trois résidus sont strictement conservés, une leucine (L) N-terminale, une glycine (G) (à l'exception du virus Memana) au niveau de la région connectrice et un acide glutamique (E) C-terminal. Le groupe I présente une région N-

terminale supplémentaire, par rapport au groupe II, bien conservée, comprenant un ensemble de résidus hydrophobes suivis d'une région basique. Cette région est retrouvée dans l'alignement du groupe III (Figure 25).



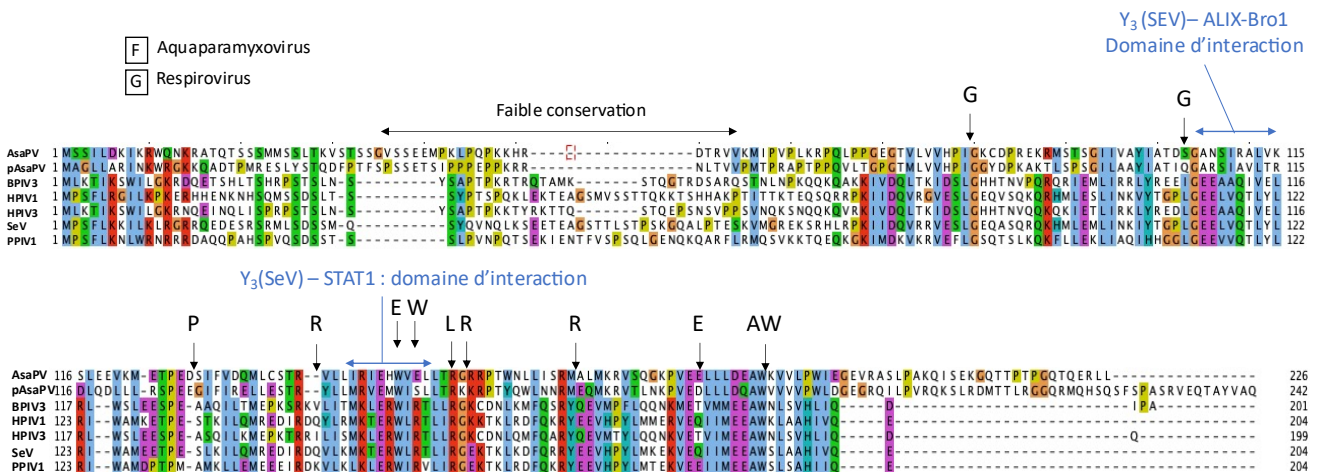
**Figure 23 : Alignement des protéines C du groupe I.** Les conventions utilisées pour les virus de l'alignement sont présentées en annexe. Le surlignage en bleu représente des résidus hydrophobes, en violet des résidus acides, en rouge des résidus basiques et en vert des résidus polaires. Les résidus strictement conservés sont indiqués par des flèches.

Dans le cas du groupe II (Figure 24), l'alignement ne semble fiable que pour la région CTD comme le souligne l'absence d'insertions dans l'alignement. Une région supplémentaire par rapport au groupe I, peu conservée (nombreuse insertions), reliant les domaines conservés NTD et CTD est présente. Seule une thréonine (T) est strictement conservée en position 134.



**Figure 24 : Alignement des protéines C du groupe II.** Mêmes conventions que pour la figure 25.

L'alignement réalisé pour les protéines C du groupe III (Figure 25) présente également deux régions conservées : une courte région de 15 résidus N-terminale incluant une région basique puis un domaine ordonné C-terminal incluant 10 résidus strictement conservés. Ces protéines C sont globalement plus longues que pour les groupes I et II notamment par la présence d'une longue boucle désordonnée d'environ 50 résidus entre le CTD et le NTD. Une deuxième région basique, riche en résidus lysine (K) et arginine (R), est présente entre les résidus 165 et 179.



**Figure 25 : Alignement des protéines C du groupe III.** Mêmes conventions que pour la Figure 25.

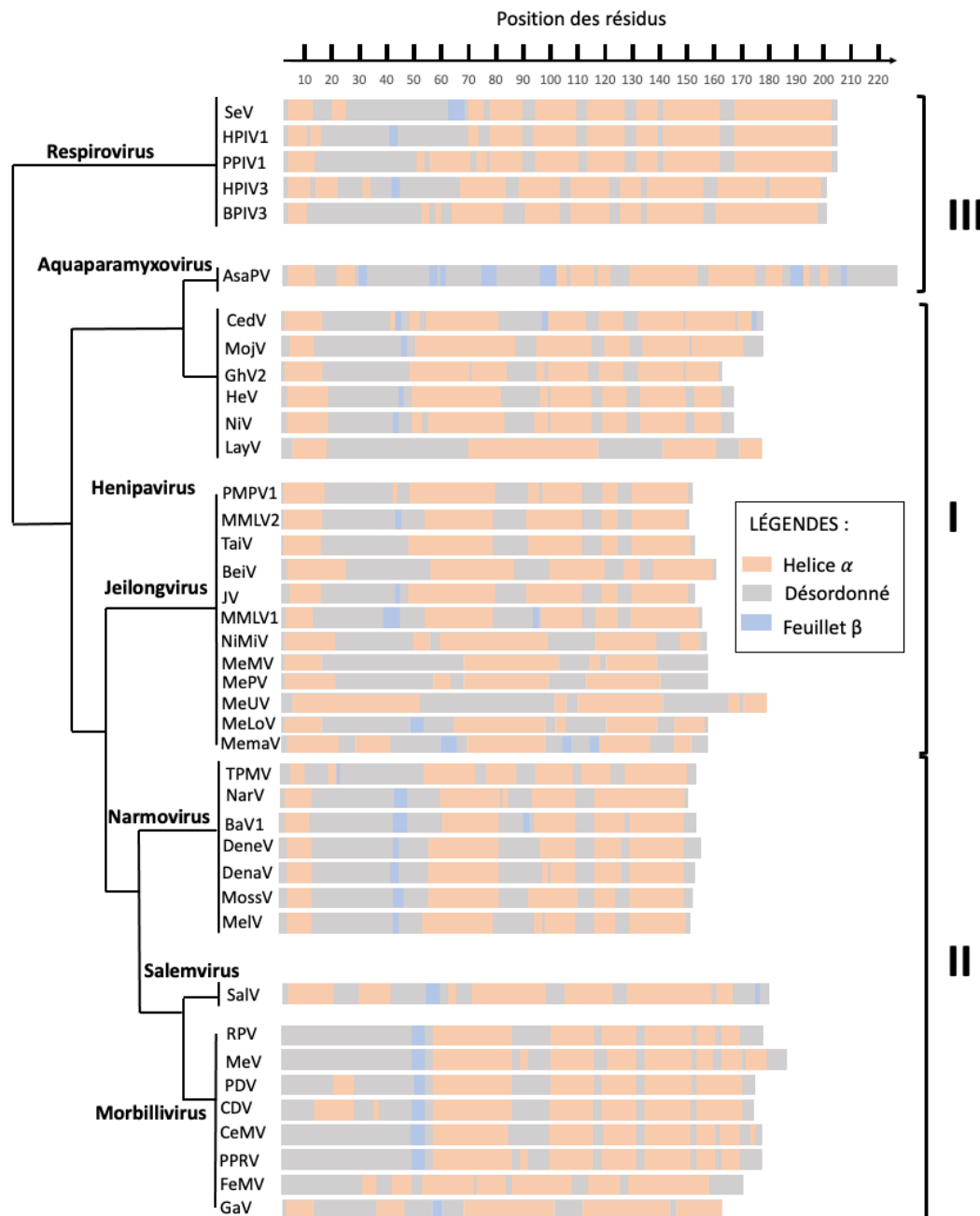
Sur l'alignement des protéines C du groupe II (Figure 24), nous avons indiqué dans un encadré noir, le motif de 12 acides aminés interagissant, dans le cas du MeV, avec la protéine SHCBP1 (Ito *et al.*, 2013). Deux positions du motif sont conservées au sein du groupe, un résidu hydrophobe et un acide glutamique strictement conservé soulignant une possible conservation de la fonction pour les autres virus du groupe.

De la même façon, nous avons indiqué sur l'alignement des protéines C du groupe I (Figure 23) le domaine d'interaction de la protéine C du NiV avec le CTD de la protéine TSG101 du sous-complexe ESCRT-I (Park *et al.*, 2016). Il a été montré que la déplétion du facteur TSG101 inhibe l'augmentation par la protéine C du bourgeonnement et empêche la libération de particules virales. On peut remarquer que la région délimitée de l'interaction avec TSG101 est très conservée avec les autres protéines C du groupe et suggère également le partage d'une fonction commune.

**(C) Une organisation structurale commune à l'ensemble des protéines C**

Des outils de prédictions de structures secondaires sur la base de la séquence en acides aminés ont été combinés à l'échelle de la sous-famille et les résultats sont présentés en Figure 26. Les prédictions de structures secondaires des protéines C ont été réalisées via l'interface web RPBS (Ressource Parisienne en Bioinformatique Structurale) qui met à disposition l'analyse PSIPRED 4.0.1 (Protein Secondary Structure Prediction) (Jones, 1999).

Cette méthode permet, à partir de la séquence en aa, de positionner les limites des structures secondaires des protéines. Ces résultats sont présentés dans la figure suivante.

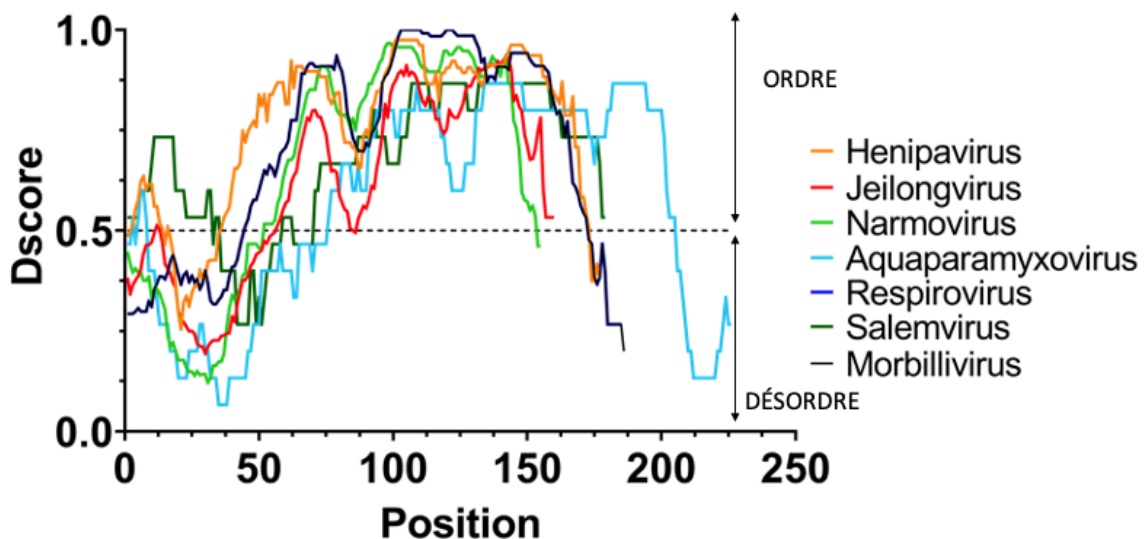


**Figure 26 : Analyse des structures secondaires des protéines C des *Orthoparamyxovirinae*.** Représentation graphique des prédictions de structures secondaires des protéines C des *Orthoparamyxovirinae*. Les rectangles gris représentent les régions prédites non structurées, en bleu les régions repliées en feuillet  $\beta$  et en orange en hélices  $\alpha$ . Les positions des virus ont été choisies selon l'arbre phylogénétique construit sur la séquence nucléotidique des protéines L. Les groupes I, II et III représentés à droite correspondent aux résultats des alignements présentés précédemment. Le groupe I comprend les Jeilongvirus et Henipavirus, le groupe II comprend les Narmovirus, Salemvirus et Morbillivirus et le groupe III comprend les Respirovirus et Aquaparamyxovirus (Figure inspirée de Siering, Cattaneo and Pfaller, 2022).

Ces prédictions ont été complétées par la recherche des emplacements des régions intrinsèquement désordonnées (IDR) des protéines C des *Orthoparamyxovirinae* (Figure 27)

en soumettant les séquences d'acides aminés à 16 algorithmes différents accessibles via des serveurs web et en calculant un score de désordre consensuel (Dscore) comme décrit précédemment (Gerard *et al.*, 2009; Jensen *et al.*, 2020; Roy *et al.*, 2023). Pour chaque serveur, les résidus prédits comme faisant partie d'une région désordonnée ont reçu un score de 0, tandis que les autres résidus ont reçu un score de 1. Pour chaque acide aminé, une valeur de Dscore a été calculée en ajoutant les valeurs de toutes les prédictions et en normalisant la somme obtenue. Les régions dont le Dscore normalisé est inférieur à 0,5 sont considérées comme intrinsèquement désordonnées, tandis que les régions dont le score est supérieur à 0,5 sont considérées comme structurées.

Pour plus de clarté, nous présentons uniquement la moyenne des analyses de Dscore réalisées pour chacun des virus au sein d'un même genre. En effet, nous avons constaté que les Dscore des protéines C des virus d'un même genre étaient globalement identiques. Seuls les *Henipavirus* présentaient des variations dans leur région N-terminale où le Dscore était supérieur à 0.5 pour CeV, NiV et HeV et inférieur à 0.5 pour MojV et GhV2. Cette inconsistance au sein du genre pourrait être mise en parallèle avec l'étude récemment publiée de Wells et al (Wells *et al.*, 2022) qui propose la création d'un nouveau genre pour le MojV qui, en fonction de la rigueur des critères de sélection choisis, présente des dissimilitudes avec les autres *Henipavirus*.



**Figure 27 :** Dscore calculé à partir de la séquence en acides aminés des protéines C des *Orthoparamyxovirinae* répartis par genre : *Henipavirus*, *Narmovirus*, *Jeilongvirus*, *Morbillivirus*, *Respirovirus*, *Aquaparamyxovirus*, *Salemvirus*.

Ces deux analyses confirment l'organisation modulaire évoquée précédemment que partagent les protéines C des *Orthoparamyxovirinae* (Lo, Søgaard and Karlin, 2014; Siering, Cattaneo and Pfaller, 2022) :

(i) toutes les régions CTD des protéines C de la sous-famille ont un Dscore supérieur à 0.5 et sont prédites structurées par une série d'hélices- $\alpha$  et par la présence occasionnelle d'un potentiel feuillet  $\beta$  (**Figure 26&27**) ;

(ii) les régions NTD sont globalement intrinsèquement désordonnées (Dscore < 0.5) à l'exception de la présence de petits domaines NTD structurés en hélices- $\alpha$  que l'on retrouve pour tous les genres sauf les *Morbillivirus* (**Figure 26&27**). Ces hélices N-terminales présentent un Dscore globalement faiblement supérieur à 0.5 (sauf dans le cas des *Salemvirus*) et nous réitérons l'hypothèse évoquée précédemment qu'elles pourraient correspondre à une région  $\alpha$ -MoRE.

En conclusion, la séquence en aa des protéines C des *Orthoparamyxovirinae* est faiblement conservée et l'homologie n'est généralement pas identifiable à partir de la séquence au-delà du genre. Curieusement, malgré une apparente dissemblance, les groupes I et II formés par les protéines C des *Orthoparamyxovirinae* possède une homologie. De plus, l'ensemble des protéines C possède une organisation structurale commune. À l'heure actuelle, seule la structure de la protéine C du virus Sendai a été résolue. Dans l'état des connaissances, il est donc impossible de déterminer si ces homologies proviennent d'une évolution convergente de la structure ou si ces protéines ont divergé en séquences depuis un ancêtre commun.

Dans cette optique, nous avons souhaité illustrer cette hypothèse évolutive par la caractérisation structurale de deux protéines C supplémentaires, seules ou en complexes, appartenant au premier et deuxième groupe, la protéine C du virus Nipah et la protéine C du virus de la toupe. Nous avons également souhaité identifier des interactions avec des partenaires cellulaires ou viraux afin de mieux comprendre comment cette protéine peut agir, d'un point de vue structural, dans les cellules infectées.

De plus, un certain nombre de thérapies antivirales ont été développées sur la base d'informations structurales et sur la caractérisation des interfaces d'interaction entre les protéines virales et les protéines cellulaires. C'est pourquoi il apparaît intéressant, au vu du rôle clé de la protéine C dans l'échappement au système immunitaire et à la pathogénicité, de confirmer leur apparente identité de structures. Ces protéines pourraient représenter des cibles de choix pour le développement de thérapies antivirales globales contre les paramyxovirus car les possibilités de mutations permettant au virus de développer des résistances sont réduites sur les gènes chevauchants en raison des contraintes évolutives liées aux deux protéines. **Une approche structurale permettant de caractériser les protéines C des *Orthoparamyxovirinae* pourrait donc permettre à terme de cibler thérapeutiquement plusieurs virus de la même famille.**



# Chapitre 3 : Méthodes de caractérisation structurale des protéines appliquées à ces projets

## I. La chromatographie d'exclusion de taille couplée à la diffusion de la lumière à plusieurs angles et à la réfractométrie (SEC-MALLS-RI)

Une fois les échantillons de protéines recombinantes purifiés, nous procédons à l'évaluation de leur polydispersité, de leur qualité et nous mesurons leur masse moléculaire en solution grâce à la chromatographie d'exclusion de taille (SEC) couplée à la détection par diffusion de la lumière à plusieurs angles (Multi-angle Laser Light Scattering - MALLS) et à la réfractométrie différentielle (Differential Refractometry Index - dRI).

Par cette méthode, il est possible de déterminer une mesure absolue de la masse moléculaire des particules. L'utilisation de la colonne de chromatographie d'exclusion de taille permet de séparer, si besoin, les molécules présentes en solution afin de les analyser au niveau du MALLS en fonction de leur volume d'élution.

Le détecteur dRI permet de mesurer la concentration de l'échantillon à partir de l'index de réfraction de la solution dans le cas des protéines solubles. Cette mesure permet de s'affranchir de la composition en aa de la protéine étudiée car l'index de réfraction varie très peu en fonction des aa à la différence du coefficient d'extinction molaire. Dans le cas des protéines membranaires en détergent, les UV<sub>280</sub> et le dRI sont utilisés de manière concomitante pour déterminer la concentration et la masse moléculaire de la protéine. Une valeur générique d'index de réfraction peut être utilisée pour l'ensemble des protéines et des acides nucléiques (0.185 mL.g<sup>-1</sup>). L'excès d'indice de réfraction est mesuré par un réfractomètre différentiel RI2000 (Schambeck).

Enfin, le MALLS mesure l'intensité de la lumière diffusée par la substance à analyser à de multiples angles par rapport au faisceau lumineux. Le signal de diffusion de lumière est mesuré à 18 angles à l'aide d'un détecteur DAWN-EOS II (Wyat Technology Corp., Santa Barbara, CA) en utilisant un laser de longueur d'onde de 690nm. Les données sont analysées avec le logiciel ASTRA (Wyatt).

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

$$MM = \frac{R_{\theta}}{KC \left(\frac{dn}{dc}\right)^2}$$

où MM est la masse moléculaire de l'échantillon (g.mol<sup>-1</sup>), R<sub>θ</sub> le rapport de Rayleigh mesuré (c-à-d la quantité de lumière diffusée par l'échantillon relativement à l'intensité du laser) déterminé par le détecteur MALLS à plusieurs angles et moyenné pour l'angle 0, C est la



concentration en protéine déterminée par le détecteur dRI,  $\frac{dn}{dc}$  est l'incrément d'indice de réfraction du soluté et K est une constante optique qui dépend des caractéristiques de l'instrument et de la solution.

$$\text{Avec } K = \frac{4\pi^2 n_0^2}{N_A \lambda_0^4}$$

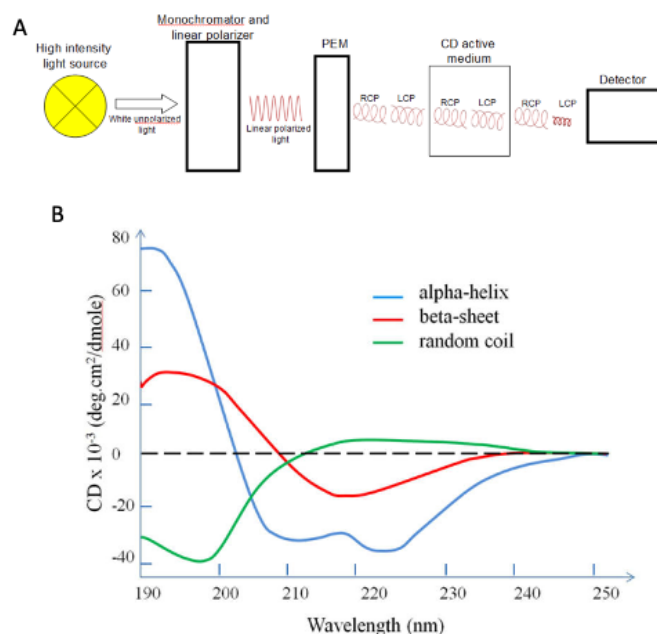
où  $n_0$  est l'indice de réfraction du milieu,  $\lambda_0$  la longueur d'onde du rayonnement et  $N_A$  le nombre d'Avogadro.

Les protéines purifiées au cours de ce travail ont été injectées sur des colonnes Superdex Increase 200 ou 75 10/300 à température ambiante, à 0,5 mL.mn<sup>-1</sup>.

## II. Évaluation de la structure tertiaire d'un échantillon par dichroïsme circulaire (CD).

Le dichroïsme circulaire repose sur la capacité d'un échantillon possédant un chromophore intrinsèquement chiral ou lié par une liaison covalente à un centre chiral de la molécule ou placé dans un environnement asymétrique par la structure 3D de la molécule, d'absorber différemment la lumière polarisée circulairement vers la droite et la lumière polarisée circulairement vers la gauche (**Figure 28A**). L'absorption préférentielle de l'une de ces polarisations pour chaque longueur d'onde résulte en une déviation de la résultante correspondant au spectre dichroïque (spectre CD) (Corrêa and Ramos, 2009).

Dans le cas des protéines, la mesure du dichroïsme circulaire dans l'UV lointain (180 à 260nm) lié à la liaison peptidique fournit des informations sur la présence et la proportion moyenne de structures secondaires (hélices  $\alpha$ , feuillets  $\beta$ , tours et structures désordonnées) pour lesquelles les spectres CD obtenus sont caractéristiques de ces structures secondaires (**Figure 28B**). Le spectre CD d'une protéine repliée en hélices  $\alpha$  aura deux minima négatifs à 222nm et 208nm et un maxima positif à 190nm. Le spectre CD d'une protéine repliée en feuillets  $\beta$  aura en général une bande négative entre 210 et 220nm et une bande positive entre 195 et 200nm. Le spectre CD d'une protéine désordonnée atteindra une grande série de valeurs négatives vers 200nm.



**Figure 28 : Le dichroïsme circulaire. (A)** Représentation schématique d'un spectrophotomètre. La lumière non polarisée est produite depuis une source de lumière et traverse un monochromateur et polariseur permettant de produire de la lumière polarisée. Cette lumière traverse ensuite un modulateur photoélastique (PEM) qui permet de circulariser la lumière polarisée à droite et à gauche avant sa traversée de l'échantillon (CD active medium) et sa détection. **(B)** Spectres CD représentatifs des différentes structures secondaires (hélices  $\alpha$  (bleu), feuillet  $\beta$  (rouge), désordonné (vert)) (Corrêa and Ramos, 2009).

Au-dessus de 260 nm, le signal de dichroïsme circulaire provient des chaînes latérales des acides aminés et dépend de leur position relative, de leur environnement, de leur mobilité et de la quantité de chacun d'entre eux. Ceci nous donne alors une information sur la structure tertiaire de la protéine.

Le contenu moyen en hélice  $\alpha$  a été calculé par la détermination de la valeur de l'ellipticité molaire par résidu  $[\theta]$  expérimentale à 222 nm après soustraction du signal du tampon.  $[\theta]$  est calculée selon l'équation :

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

où  $\delta$  est le signal CD mesuré en mdeg,  $l$  est la longueur de la cuvette,  $N_{aa}$  est le nombre d'acides aminés de la protéine et  $C$  est sa concentration.

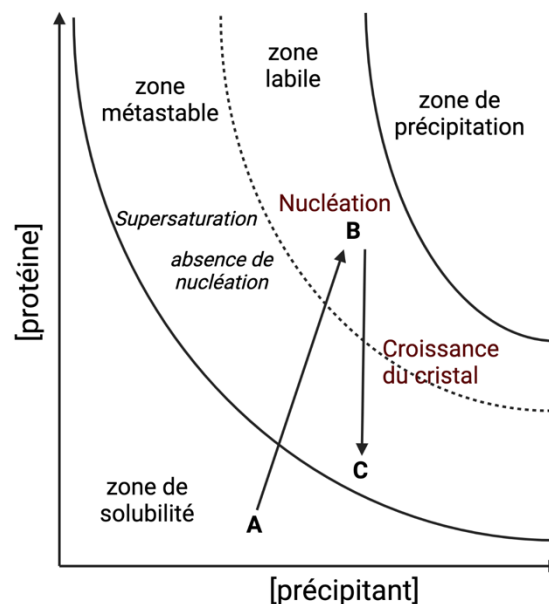
On peut alors comparer cette valeur à la valeur de l'ellipticité molaire par résidu pour une protéine contenant 100% d'hélices  $\alpha$  ( $[\theta]_{helix,222nm} = -35,000 \text{ deg.cm}^2.\text{dmol}^{-1}$ ) et ainsi déterminer le contenu moyen en hélices  $\alpha$  de l'échantillon (Luo and Baldwin, 1997).

Dans le cadre de cette étude, les mesures de dichroïsme circulaire ont été réalisées sur un spectrophotomètre MOS500 (BioLogic) à température ambiante. Les spectres CD dans l'UV lointain ont été enregistrés de 200 nm à 260 nm.

### III. La Cristallographie aux rayons X

#### (A) Cristallogenèse

Dans des conditions de solvant et de concentration protéique favorables, les protéines sont capables de s'assembler en agrégats spécifiques ordonnés et thermodynamiquement stables à même de former un germe cristallin (étape de nucléation). Ces germes cristallins entrent alors dans une phase de croissance cristalline caractérisée par un assemblage en cristaux ordonnés de protéines. La croissance cristalline s'arrête lorsque la concentration en protéines devient trop faible dans la solution.



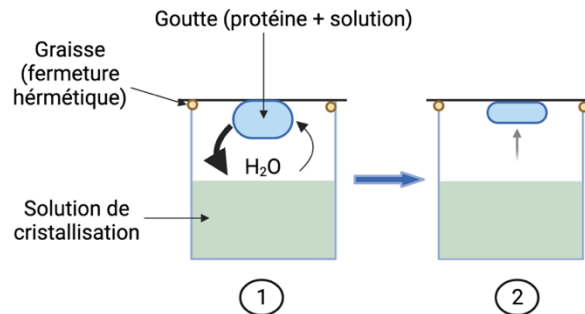
**Figure 29 : Diagramme de phase bidimensionnel de la cristallisation d'une macromolécule en fonction de sa concentration et de la concentration de la solution en agent cristallisant.** *Figure réalisée dans BioRender.*

Le processus de cristallisation peut être visualisé sur un diagramme de phase (**Figure 29**) dont les variables sont la concentration en protéines et la concentration en agents de cristallisation. Ce diagramme distingue 4 zones : la zone de solubilité des protéines, une zone métastable, une zone de nucléation (zone labile) et la zone de précipitation. Afin d'identifier les conditions de nucléation, les concentrations en protéine et en agent de cristallisation sont progressivement augmentées jusqu'à l'entrée dans la zone de nucléation (**A → B**). Après nucléation (**B**), si les conditions sont favorables, la croissance cristalline démarre (**B → C**) et progressivement la concentration en protéines diminue jusqu'à un retour vers la zone de solubilité.

Afin d'identifier les conditions favorables permettant la formation de cristaux de macromolécules, les échantillons de protéines purifiées à différentes concentrations pour les protéines étudiées ont été envoyés à la plateforme HTX (High Throughput Crystallization) de la Plateforme de Biologie Structurale (PSB). Cette plateforme a mis au point une méthode robotisée et automatisée de cristallisation haut-débit par la méthode de la goutte assise et de la

diffusion de vapeur. Une fois les conditions de cristallisation optimales identifiées, un affinement manuel est réalisé afin d'obtenir des cristaux permettant l'obtention de données de diffraction de bonne qualité.

Dans le cas de cette étude, le raffinement manuel de la production de cristaux a été effectué par la méthode de la diffusion en phase vapeur (**Figure 30**). Cette étape consiste à faire varier les conditions de cristallisation identifiées par HTX en faisant varier la concentration en précipitants, le pH ou la concentration en protéines (liste non exhaustive).

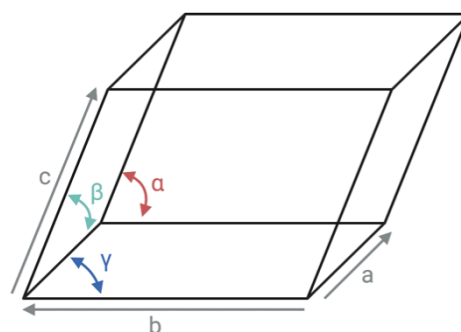


**Figure 30 : La méthode de cristallisation par la diffusion de vapeur pour le dispositif de la goutte suspendue.** *Figure réalisée dans BioRender.*

Cette méthode consiste à déposer une microgoutte sur une lamelle de verre d'un mélange de protéine concentrée et de solution de cristallisation. Cette lamelle est ensuite retournée et déposée sur un réservoir contenant la même solution de cristallisation en grande quantité et scellée par de la silicone. Le réservoir étant plus concentré en cristallisant, de la vapeur d'eau s'échappe alors de la goutte vers le réservoir concentrant ainsi la goutte en agent de cristallisation et en protéine jusqu'à atteindre un équilibre entre les deux compartiments. Cette concentration progressive permet d'entrer dans une zone de sursaturation en protéine favorable à l'apparition de cristaux dans les conditions favorables à l'échantillon étudié.

### **(B) La diffraction et la densité électronique**

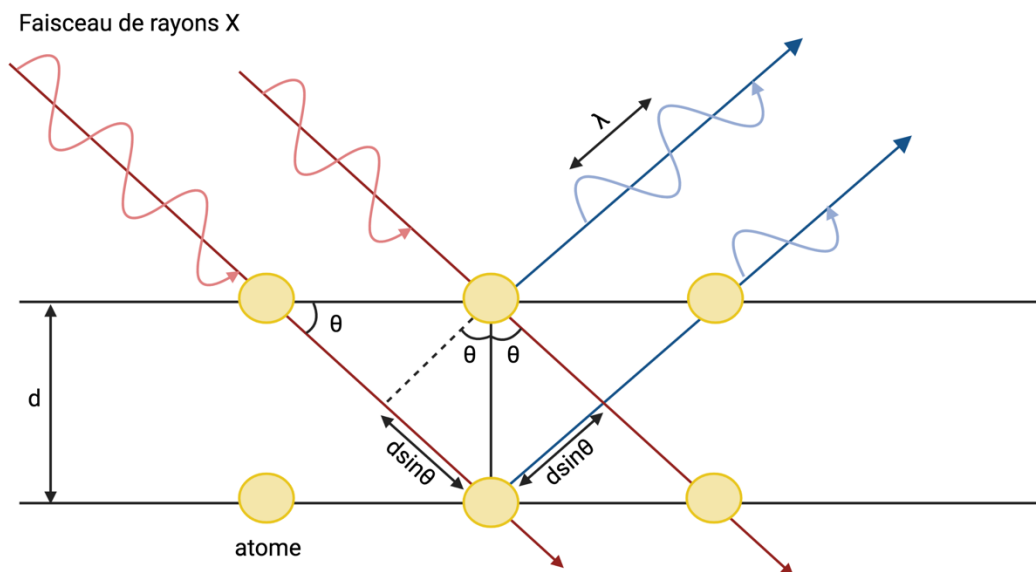
Un cristal est formé par la répétition d'une maille cristalline dans les trois directions. On peut décrire un cristal par les paramètres de maille (trois côtés  $a$ ,  $b$  et  $c$  et les angles  $\alpha$ ,  $\beta$  et  $\gamma$  qui les relient) (**Figure 30**) et les opérations de symétrie qui le constituent. Une maille contient souvent plusieurs fois le même motif et l'on peut passer d'un motif à l'autre en appliquant des opérations de translation et de rotation.



**Figure 31 : Paramètres de la maille cristalline.** *Figure réalisée dans BioRender.*

Les rayons X sont des ondes électromagnétiques comme la lumière visible, situées entre la lumière ultraviolette et les rayons gamma. Ils sont caractérisés à ce titre par leur comportement ondulatoire défini par une longueur d'onde et une fréquence d'ondulation. Leur longueur d'onde varie de 0.01 nm à 10 nm, ce qui est comparable aux distances entre les atomes. Lorsque ces rayons X sont diffusés par un atome dans toutes les directions, il se produit un phénomène d'interférence. Lorsque les amplitudes maximales des ondes diffusées sont en phase, c'est ce qu'on appelle une interférence constructive et l'amplitude de l'onde résultante correspondra à l'addition des maxima. Si le maximum de l'amplitude de l'onde diffusée se superpose au minimum d'une seconde onde diffusée, cela correspond à une interférence destructrice et l'onde résultante sera d'amplitude nulle.

Les rayons X sont diffractés par le cristal d'un angle de réflexion  $\theta$  avec le faisceau incident et on mesure au niveau d'un détecteur l'intensité des tâches de diffraction (appelées réflexions) correspondant aux interférences constructives des ondes diffusées.



**Figure 32 : Illustration de la condition de Bragg : réflexion sélective sur une famille de plan (hkl).**  
*Figure réalisée dans BioRender.*

On obtient alors, pour un cristal donné, un patron de diffraction qui consiste en un réseau régulier de tâches de diffraction caractérisées par des indices de Miller (dépendant de la position  $h,k,l$  des tâches détectées), par une amplitude  $|F|$  enregistrée et une phase  $\varphi$  qu'il faudra résoudre et qui correspond à l'instant où l'onde est venue frapper le détecteur. Grâce à la loi de Bragg, on peut estimer la limite de résolution du cristal étudié avec la formule suivante (**Figure 32**) :

$$n\lambda = 2 d_{max} \sin\theta_{max}$$

où  $n$  est un entier,  $d_{max}$  est la limite de résolution et  $\theta_{max}$  correspond à la limite de diffraction détectée.

Le jeu de données collecté correspond à une succession d'images mises bout à bout qui permettent de reconstituer le réseau tridimensionnel formant le réseau réciproque. Afin

d'analyser ces données dans l'espace réel, il est important de noter que chacun des spots de réflexions résultent de la contribution à la diffraction d'un ensemble d'atomes à des positions  $x$ ,  $y$  et  $z$  dans l'espace réel de la maille élémentaire du cristal. Ces points sont décrits dans l'espace réciproque par des facteurs de structure  $F_{hkl}$ . La transformée de Fourier permet de traduire l'intensité des tâches de diffraction de l'espace réciproque dans l'espace réel par la relation suivante :

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

où  $f_i$  est le facteur de diffusion de l'atome  $i$ , l'exponentielle est une fonction périodique tridimensionnelle appliquée à chaque coordonnée  $x_i, y_i, z_i$  de l'atome  $i$  dans la maille élémentaire.

Chaque rayon diffracté constitue une courbe périodique complexe résultant des contributions diffractives de tous les atomes de la maille élémentaire. Chaque facteur de structure dépend donc de la contribution de tous les éléments volumiques de la maille élémentaire, c'est ce qu'on appelle la densité électronique. La formule précédente peut donc être réécrite de la façon suivante :

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

où  $V$  est le volume de la maille élémentaire et l'intégrale est calculée pour l'ensemble des coordonnées  $x, y, z$  des atomes présents dans la maille élémentaire. Dans l'espace complexe, le facteur de structure peut être décrit par la formule d'Euler de la façon suivante :

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

où  $|F_{hkl}|$  représente l'amplitude du vecteur complexe correspondant au facteur de structure  $F_{hkl}$  et  $\alpha_{hkl}$  la phase correspondante à ce vecteur par rapport à l'origine. L'amplitude du vecteur est directement déterminée par la mesure de l'intensité des tâches de diffraction  $I_{hkl}$ .

En conclusion, on peut calculer les densités électroniques de l'espace réel correspondant aux tâches de diffraction de l'espace réciproque par le calcul suivant :

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

Cette relation permet la génération d'une carte de densité électronique en trois dimensions qui sera utilisée pour reconstruire le modèle tridimensionnel de la macromolécule étudiée. Pour cela, il est indispensable de déterminer la phase  $\alpha_{hkl}$  qui reste inconnue. Cette étape représente le problème central de la cristallographie et plusieurs méthodes pour la déterminer ont été développées.

### ***(C) Collecte et traitement des données***

La collecte de données permet d'enregistrer un grand nombre d'images à différents angles qui seront ensuite analysées grâce au logiciel XDS et l'interface graphique XDSgui (Kabsch, 2010), c'est l'étape d'indexation des données. Durant cette étape, les programmes vont également produire une première estimation des paramètres de la maille cristalline.

L'indexation des images par XDS est constituée de plusieurs étapes. Le programme commence par estimer le bruit de fond moyen de chaque image et constitue un masque de correction permettant de séparer les pixels associés aux tâches de diffraction de ceux associés au bruit de fond. Il recherche et extrait ensuite les tâches de diffraction afin de déterminer les paramètres de la maille cristalline et les symétries du cristal. Le programme réalise ensuite une étape d'auto-indexation permettant d'attribuer pour chaque image de diffraction une liste d'indices hkl correspondant aux positions des tâches de diffraction extraites. Il mesure aussi leur intensité  $I_{hkl}$  et les erreurs associées à ces intensités  $\sigma_{hkl}$ .

Les variations d'intensité des réflexions causées par la rotation du cristal durant la collecte par rapport au faisceau peuvent induire des erreurs qui seront corrigées par une mise à l'échelle des intensités.

Il existe un certain nombre de variables statistiques permettant d'estimer la qualité des données incluant notamment le facteur R, le ratio signal sur bruit ( $I/\sigma$ ), la complétude et la valeur du coefficient de corrélation de Pearson ( $CC_{1/2}$ ) (Dubach and Guskov, 2020). Ces mesures statistiques sont également utilisées pour décider de la limite de résolution des données collectées en décidant jusqu'à quelles valeurs statistiques il est possible d'utiliser les données collectées pour construire un modèle.

Traditionnellement, le ratio signal sur bruit est un bon indicateur pour déterminer les réflexions conservées pour la construction du modèle. Il est souvent conseillé de ne conserver que les réflexions pour lesquelles le ratio est supérieur ou égal à 2 pour obtenir la meilleure résolution. Cependant, dans de nombreux cas, les réflexions sont conservées jusqu'à une valeur de 1.5.

La qualité des données récoltées est également estimée par le facteur  $R_{merge}$  ou  $R_{sym}$  qui permet d'estimer la précision des mesures de chaque réflexion par l'équation suivante :

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^n |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)}$$

où  $I_i(hkl)$  est l'intensité des réflexions  $i$  et  $\bar{I}(hkl)$  correspond à l'intensité moyenne. Ce facteur permet de prendre en compte la symétrie du cristal en quantifiant la différence associée aux mesures correspondant aux mêmes réflexions qui diffèrent en intensité de la moyenne associée à cette réflexion. Plus la valeur du  $R_{merge}$  est élevée, plus les mesures associées aux mêmes réflexions ne sont pas similaires. Il est conseillé de ne garder que les images pour lesquelles le  $R_{merge}$  est inférieur à 40 - 60%. Cependant, le facteur  $R_{merge}$  est biaisé car il dépend de la multiplicité des réflexions et sa valeur augmente avec de multiples mesures des mêmes réflexions. C'est pourquoi il est associé à un second facteur,  $R_{meas}$ , dont la valeur est indépendante de la multiplicité des réflexions par une correction du facteur par  $\sqrt{\frac{n_{hkl}}{n_{hkl}-1}}$  :

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n_{hkl}}{n_{hkl} - 1}} \sum_{i=1}^n |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)}$$

Communément, les valeurs de  $R_{meas}$  sont utilisées en association avec le ratio signal sur bruit pour déterminer la limite de résolution des données à la collecte. Il est considéré que les réflexions ne sont pas de bonne qualité si la valeur du  $R_{meas}$  dépasse 60% pour les réflexions de la dernière enveloppe conservée.

La valeur de complétude permet d'estimer le rapport du nombre de réflexions mesurées expérimentalement en fonction du nombre de réflexions attendues dans une symétrie donnée. Une mesure expérimentale de bonne qualité aura une complétude proche de 100%.

Le  $CC_{1/2}$ , est utilisé pour estimer la qualité des données collectées. Il consiste à diviser les données de manière aléatoire en deux ensembles de réflexions pour lesquelles on calcule la corrélation. Il s'agit d'estimer les valeurs de l'intensité de l'un des deux ensembles à partir de l'autre ensemble. Si le  $CC_{1/2}$  a une valeur de 1, la corrélation entre les deux ensembles est parfaite. Pour une valeur de 0, il n'y a aucune corrélation et les données sont de mauvaise qualité ou anisotropes. Ce facteur est dépendant de la résolution utilisée.

**Dans le cadre de notre étude, nous avons résolu l'ambiguïté de phases évoquée précédemment par deux méthodes : la diffusion anormale et le remplacement moléculaire. La première méthode repose sur la dérivation des cristaux collectés.**

#### ***(D) La dérivation des protéines***

Lorsqu'aucun homologue de la protéine étudiée n'est disponible, il est possible d'utiliser certaines propriétés des protéines pour résoudre l'ambiguïté de phase résiduelle. En effet, les cristaux sont poreux et contiennent près de 50% de solvant présent au niveau de canaux où de petites molécules peuvent diffuser. Ces composés (gaz noble, sel d'atomes lourds, chélates de métaux, halogénures ou métal alcalin) causent des perturbations au niveau des réflexions, phénomène permettant de résoudre les phases initiales. Ils sont introduits soit lors de la cristallisation, soit par trempage lors de la congélation.

#### ***(E) La diffusion anormale***

Dans la cas de la résolution par diffusion anormale, on enregistre un jeu de données anormales à une longueur d'onde spécifique au composé introduit.

En effet, les rayons X produits par un synchrotron peuvent être produits à différentes longueurs d'onde. Or, dans des conditions dites « normales », l'amplitude du signal de diffusion d'un atome ne dépendra que de l'interaction entre l'onde lumineuse et le nombre d'électrons de l'atome. En réalité, les électrons d'un atome subissent l'attraction du noyau. Pour certaines longueurs d'onde, on peut alors observer un phénomène de résonance qui



illustre l'excitation de couches d'électrons profondes induisant alors une variation brusque de l'amplitude du signal de diffusion que l'on appelle le phénomène de diffusion anormale et qui s'exprime par la relation suivante :

$$f = f_o + \delta\lambda$$

avec  $\delta\lambda = f' + if''$ , la correction anormale.  $f$  est appelé le facteur de diffusion atomique,  $f_o$  le facteur de diffusion en conditions normales,  $f'$  et  $f''$  correspondent, respectivement, au terme dispersif et au terme d'absorption.

Ce phénomène permet de déterminer la position des atomes lourds par l'estimation de la relation entre l'amplitude des facteurs de structure dérivés  $|F_{PH}|$  et natifs  $|F_P|$  grâce à la fonction de Patterson suivante :

$$|F_H|^2 = |F_{PH}|^2 - |F_P|^2$$

$F_H$  correspond à la contribution de l'atome lourd à la modification de l'amplitude des facteurs de structure du cristal natif. On peut ainsi déterminer la position de l'atome lourd et, à partir des coordonnées et de l'utilisation des sections de Harker liées aux symétries inhérentes à la maille cristalline (groupe d'espace), on sera en mesure de déterminer la contribution de l'atome lourd à la phase  $\alpha_H$  et donc en définitive on pourra déterminer la phase du cristal natif  $\alpha_P$  (**Figure 33**).

La loi de Friedel indique que les paires de réflexions liées par le centre de symétrie de la maille réciproque ont des intensités identiques par les relations suivantes (**Figure 33A**) :

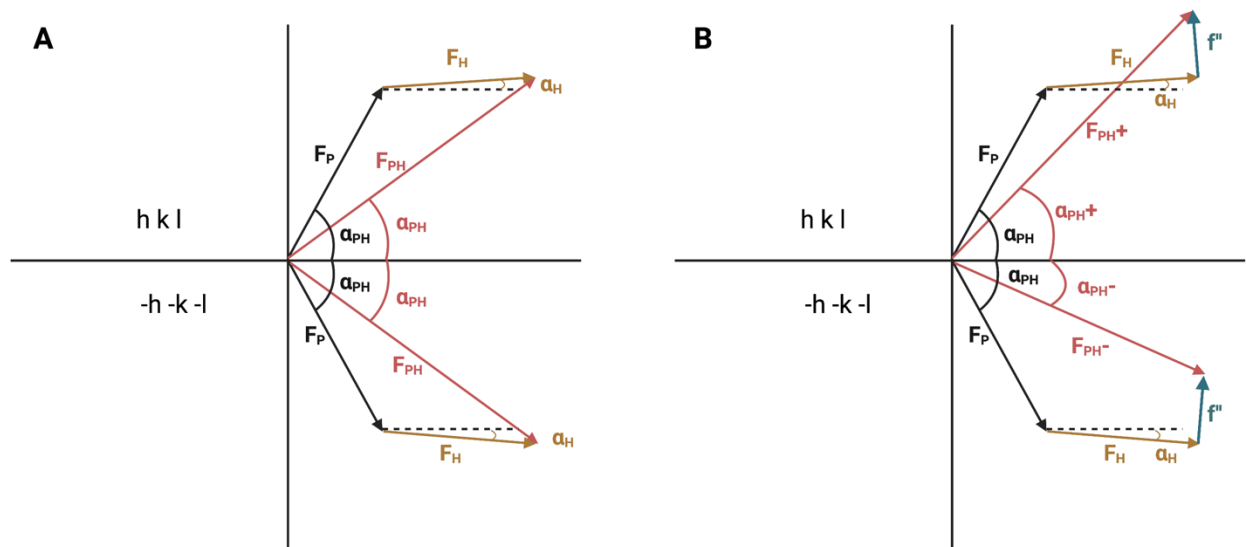
$$I(hkl) = F(hkl) \times F^*(hkl)$$

$$\text{Avec } F^*(hkl) = F(-h, -k, -l)$$

$$\text{Donc } I(hkl) = I(-h, -k, -l)$$

Dans les conditions normales, cette loi est respectée, donc les réflexions aux positions (h,k,l) et aux positions (-h,-k,-l) ont la même amplitude et des phases complémentaires.

Cependant, si l'on se place à la longueur d'onde du pic de Friedel où le signal anomal est maximal, les termes dispersifs et d'absorption  $f'$  et  $f''$  participent également au facteur de diffusion et créent une variation des intensités pour les réflexions aux positions (h,k,l) et (-h,-k,-l) introduisant un décalage entre leurs phases qui ne sont plus équivalentes (**Figure 33B**). On observe une rupture de la loi de Friedel. Cette rupture de la loi est essentielle et s'applique dans la méthode prénommée SAD (Single Wavelength Anomalous Diffraction). Ce phénomène de diffusion anormale est utilisé pour localiser la position des atomes lourds par l'utilisation de fonction de Patterson sur un jeu anomal.

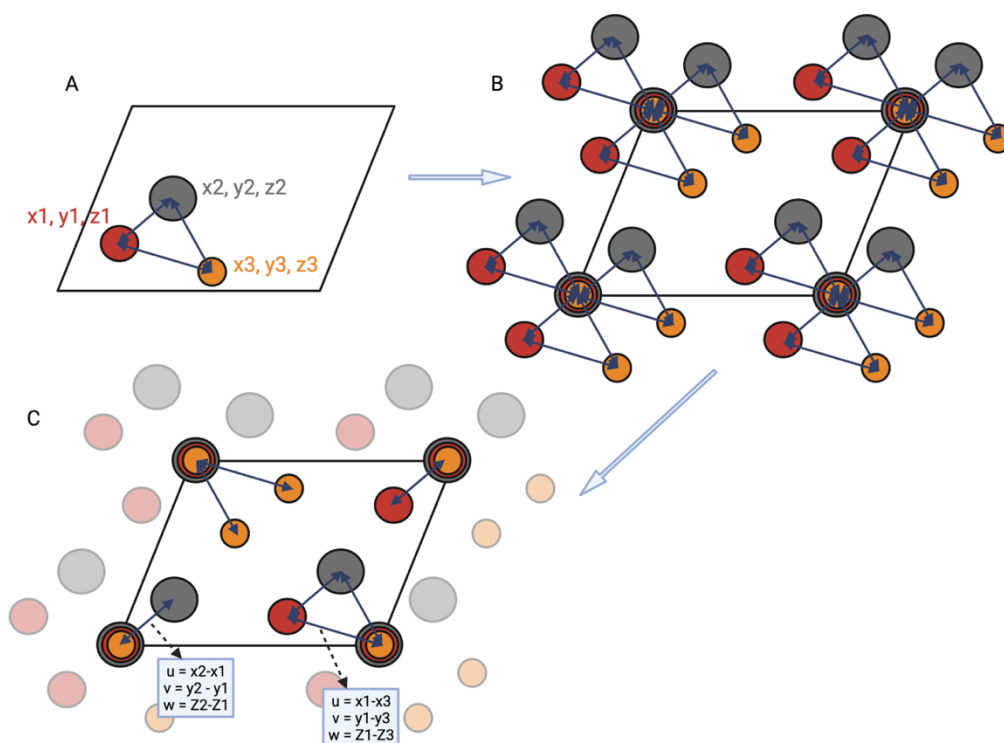


**Figure 33 : Le phasage expérimental par remplacement isomorphe et diffusion anormale. (A)** Représentation du diagramme d'Argand dans le cas d'un remplacement isomorphe pour un atome lourd  $H$  (SIR) (Loi de Friedel applicable).  $F_P$  est l'amplitude pour une réflexion du cristal natif,  $F_{PH}$  pour le cristal dérivé et  $F_H$  est la différence isomorphe correspondant au facteur de structure de l'atome lourd. **(B)** Représentation du diagramme d'Argand dans le cas d'un remplacement isomorphe pour un atome lourd  $H$  à la diffusion anormale (loi de Friedel non applicable).  $F_P$  correspond à l'amplitude en absence de diffusion anormale,  $F_{PH+}$  et  $F_{PH-}$  sont les paires de Friedel lorsqu'une diffusion anormale est présente,  $f''$  représentent la contribution du diffuseur anormal. Figure réalisée dans BioRender.

La recherche de diffuseurs anomaux est implémentée dans la suite de programme SOLVE (Terwilliger and Berendzen, 1999) et SHELXD de la suite SHELX (Sheldrick, 1990). Les programmes RESOLVE, SHELXE et DM permettent de résoudre l'ambiguïté de phase résiduelle de la méthode SAD. Pour cela, ces méthodes modifient progressivement la densité électronique du modèle anormal générant alors de nouvelles phases qui seront comparées au modèle expérimental. Il est possible de résoudre progressivement l'ambiguïté de phase par itération.

### **(F) Le remplacement moléculaire**

La méthode de remplacement moléculaire consiste à résoudre la structure de la molécule étudiée à partir d'un fichier de coordonnées d'une molécule structuralement proche telle qu'une protéine homologue, un morceau de la protéine dont la structure est déjà résolue ou, plus récemment, un modèle tridimensionnel généré par le programme AlphaFold2 (Mirdita *et al.*, 2022). On appellera ces coordonnées « le modèle de phasage ».



**Figure 34 : Projection de la fonction de Patterson dérivée d'un cristal à trois atomes.** Pour obtenir graphiquement cette fonction à partir de (A) la structure connue d'un cristal, (B) tous les vecteurs interatomiques sont tracés et déplacés parallèlement à eux-mêmes jusqu'à l'origine de la maille cristalline formant la carte de Patterson. Les extrémités de ces vecteurs correspondent aux valeurs maximales de la fonction de Patterson, dont les hauteurs sont proportionnelles au produit des numéros atomiques des atomes concernés. (C) Les positions de ces maxima au sein d'une maille cristalline (de coordonnées  $u, v, w$ ) représentent les différences entre les coordonnées de chaque paire d'atomes dans le cristal. À l'origine (aux coins de la cellule de Patterson), il existe un maximum élevé correspondant aux vecteurs interatomiques de chaque atome avec lui-même, c'est-à-dire de coordonnées  $[0, 0, 0]$ . Figure réalisée dans BioRender.

La première étape de cette méthode consiste à calculer la carte de Patterson, à partir des intensités de diffraction, correspondant à la molécule modèle utilisée, puis de replacer les atomes du modèle dans la maille élémentaire de la molécule étudiée grâce à la fonction de Patterson suivante :

$$\rho(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l |I_{hkl}| \cos(2\pi(hu + kv + lw))$$

Pour cela, il faut effectuer une série de rotations successives de la carte de Patterson générée (Figure 34) pour trouver l'orientation du modèle dans la maille élémentaire. Ensuite, les programmes de remplacement moléculaire effectuent un ajustement translationnel de la carte du modèle dans la maille élémentaire. La connaissance du groupe d'espace du cristal de la molécule étudié permettra ensuite de généraliser la position des atomes de la maille élémentaire depuis le modèle au reste de la molécule étudiée. Ces méthodes sont implémentées dans les programmes PHASER (McCoy *et al.*, 2007) et MOLREP (Vagin and Teplyakov, 2010).

### (G) Construction et affinement des modèles

En fonction de la méthode de résolution des données choisies, les étapes de construction du modèle à partir de la carte de densité électronique initialement obtenue peuvent varier mais reposent globalement sur une même logique.

Par le biais de logiciels graphiques de la suite CCP4 (Collaborative Computational Project No.4) tel que Coot (Emsley *et al.*, 2010), nous plaçons au sein de la carte de densité électronique choisie un squelette de la chaîne carbonée que nous allons construire et orienter. La résolution du modèle continue par la construction des chaînes latérales et l'ajout de résidus dans la carte de densité électronique (molécules d'eau, ions, solvant,...). Au cours de ces multiples étapes de construction, on valide le modèle en le confrontant aux données expérimentales.

Le modèle cristallographique sera par la suite affiné au cours d'une multitude d'étapes que l'on appelle des cycles d'affinements. Pour cela il est nécessaire de déterminer une méthode d'optimisation qui repose sur les paramètres du modèle qui décrivent le cristal, tels que les coordonnées des atomes, les facteurs de température B correspondant à la vibration de chaque atome, l'occupation des atomes et du solvant. Ces paramètres sont utilisés pour valider les facteurs de structure du modèle  $F_{calc}$ . Si les nouveaux facteurs de structure calculés associés au modèle construit sont plus proches des facteurs de structure expérimentaux,  $F_{obs}$ , le modèle est validé et on peut générer la carte de densité électronique du nouveau modèle. Cette boucle itérative permet d'améliorer progressivement le modèle. À chaque boucle itérative d'affinement, les paramètres du modèles sont donc altérés afin d'optimiser la fonction cible d'affinement.

L'affinement est validé par des paramètres statistiques dont notamment le  $R_{facteur}$  qui est une valeur permettant de quantifier les différences entre les amplitudes des facteurs de structures expérimentaux  $F_{obs}$  et ceux calculés depuis le modèle construit  $F_{calc}$  :

$$R_{facteur} = \frac{\sum_{hkl} |F_{obs}(hkl)| - |F_{calc}(hkl)|}{\sum_{hkl} |F_{obs}(hkl)|}$$

Ce facteur varie entre 0 et 1 et doit diminuer au fur et à mesure de l'affinement. En effet, cela indique que les modifications apportées au modèle nous rapproche des mesures expérimentales. Cependant, il a été montré que ce facteur peut être artificiellement biaisé et peut diminuer même lorsque l'affinement est mauvais. En effet, son calcul comprend des réflexions qui sont utilisées pour affiner la carte de densité électronique et qui sont donc déjà « corrigées ». Afin d'éviter ce biais, 1 à 5% des réflexions sont exclues du processus d'affinement et du calcul des cartes. Elles sont appelées « réflexions test » à la différence des réflexions utilisées pour l'ajustement des facteurs de structures observés et calculés qui constituent les réflexions de travail. La qualité de l'ajustement est estimée par le facteur  $R_{work}$  selon l'équation ci-dessus. On calcule également un facteur  $R_{free}$  sur l'ensemble test des réflexions conservées. La comparaison des  $R_{free}$  et  $R_{work}$  permet alors de juger de la qualité de l'affinement.

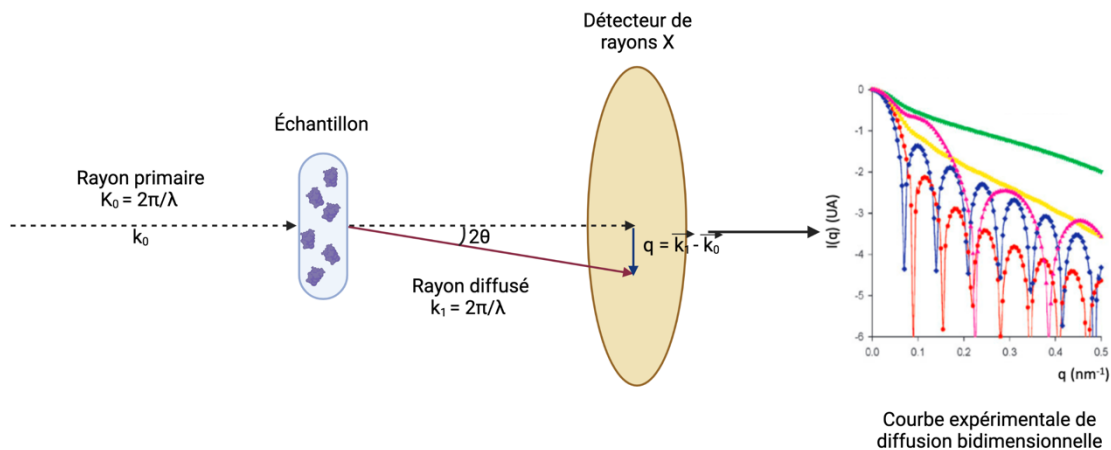
L'affinement et la construction du nouveau modèle reposent également sur un ensemble de propriétés connues des protéines, telles que la structure des acides aminés, la

longueur des liaisons covalentes et des angles de liaison. En fonction de la résolution des données, on peut s'appuyer sur la stéréochimie des macromolécules pour limiter les déplacements autorisés des atomes les moins bien résolus. D'autres outils de validation tels que le diagramme de Ramachandran (Ramachandran, Ramakrishnan and Sasisekharan, 1963) permettent de valider la conformation du squelette polypeptidique de la protéine modélisée.

#### IV. Diffusion aux rayons X à petits angles (SAXS)

La diffusion des rayons X aux petits angles (SAXS) est une méthode d'analyse de la structure des macromolécules en solution. Elle permet d'obtenir des informations sur la forme et le rayon de giration des protéines et/ou de complexes de protéines dans des gammes de poids moléculaires allant de quelques kDa au GDa sous diverses conditions. Le SAXS permet notamment d'obtenir des informations structurales sur des protéines intrinsèquement désordonnées (IDPs) et sur des macromolécules flexibles en solution.

Une expérience de SAXS requiert la génération de rayons X à haute intensité fournis généralement par un synchrotron. Elle consiste ensuite à illuminer un échantillon placé dans un capillaire de quartz par un faisceau monochromatique de rayons X (**Figure 35**). On mesure alors l'intensité de la lumière diffusée aux petits angles par l'échantillon étudié dans toutes les directions au niveau d'un détecteur. On soustrait ensuite à ce signal la diffusion du tampon. En raison de l'orientation aléatoire des particules en solution, le profil de diffusion est isotrope. Il est donc possible de moyennner radialement le schéma de diffusion bidimensionnel enregistré par le détecteur.



**Figure 35 :** Représentation schématique d'une expérience de SAXS. (Figure reproduite de Kikhney and Svergun, 2015 et réalisée dans BioRender)

On mesure l'intensité de la lumière diffusée  $I(q)$  en fonction du vecteur de diffusion  $q$  défini par :

$$\vec{q} = \vec{k}_1 - \vec{k}_0$$

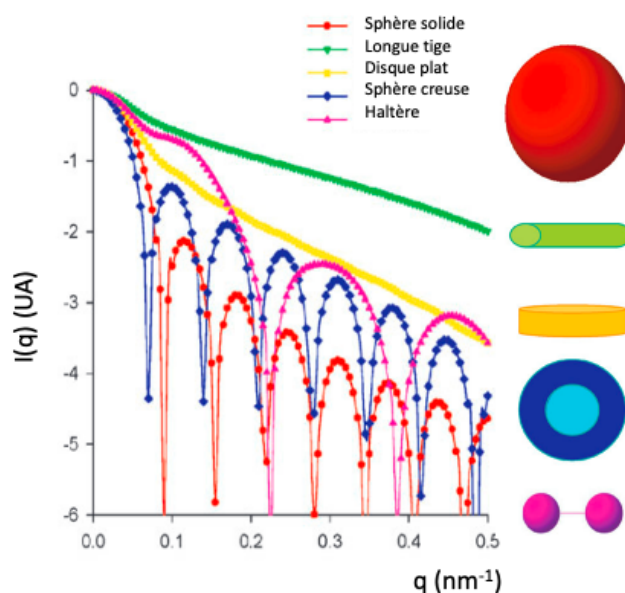
où  $q = \frac{4\pi}{\lambda} \sin \frac{\theta}{2}$  avec  $\lambda$  la longueur d'onde du faisceau et  $\theta$  l'angle de diffusion.

L'intensité de diffusion qui illustre la répartition spatiale des hétérogénéités dans l'échantillon s'exprime de la façon suivante :

$$I(q) = 4\pi \int_0^\infty \Delta\rho(r) \frac{\sin qr}{qr} dr$$

avec le contraste  $\Delta\rho(r) = \rho(r) - \rho(s)$ , où  $\rho(r)$  la densité électronique liée à la particule et  $\rho(s)$  au solvant. La transformée de Fourier de l'équation représente l'amplitude de diffusion d'un ensemble d'atomes.

Plusieurs caractéristiques de l'échantillon sont obtenues directement à partir des données expérimentales de diffusion. En effet, la distribution des intensités diffusées en fonction du vecteur de diffusion  $q$  donne des informations sur la taille, la forme et l'organisation (distance et orientation) des échantillons étudiés (**Figure 36**).



**Figure 36 : Courbe caractéristique des intensités de diffusion en fonction de la forme et de l'organisation des corps géométriques (Figure adaptée de Svergun and Koch, 2003).**

Un échantillon bien concentré permettra de s'affranchir du signal de bruit lié au solvant. Il faut s'assurer, de plus, que l'échantillon étudié se comporte de façon idéale, c'est-à-dire qu'il est monodisperse en solution et qu'il n'y a pas d'interaction inter et intramoléculaire d'attraction ou de répulsion. L'utilisation de la chromatographie d'exclusion de taille couplée à l'analyse SAXS (SEC-SAXS) permet en théorie de s'affranchir de la polydispersité et de la présence d'agrégats de l'échantillon éluant dans le volume mort.

De ces informations d'intensité de diffusion en fonction des angles, on peut construire différents graphiques informatifs sur la structure en solution de l'échantillon étudié à la condition de vérifier que l'échantillon respecte l'approximation de Guinier.

#### ***(A) Approximation de Guinier et rayon de giration***

L'approximation de Guinier établit que pour de très petits angles ( $q < \frac{1}{R_g}$ ) l'intensité de diffusion  $I(q)$  ne dépend que de deux paramètres selon l'équation suivante :

$$I(q) \approx I(0)e^{-q^2 \frac{R_g^2}{3}}$$

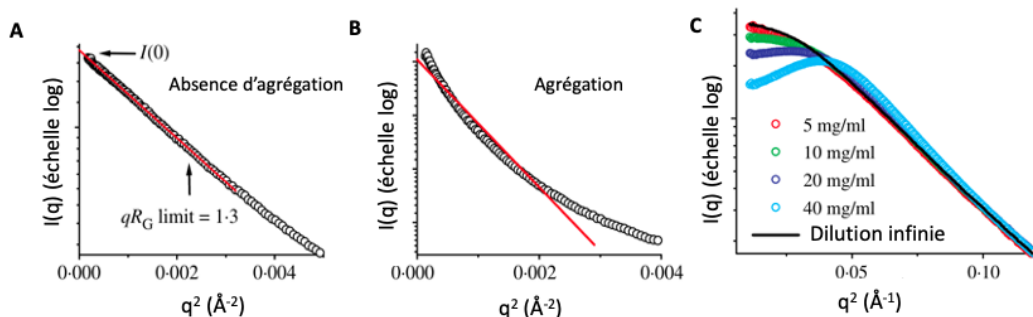
où  $I(0)$  correspond à l'intensité de diffusion à angle nul ( $q=0$ ) et  $R_g$  est le rayon de giration des particules. Cette valeur est une estimation de la taille globale de la macromolécule en solution. Il correspond à la distance quadratique moyenne depuis le centre de gravité de la molécule et peut être déduit, dans le cas de l'approximation de Guinier, à partir de la valeur de la pente du graphique de Guinier de  $\ln(I)$  en fonction de  $q^2$  par l'équation suivante :

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

$$\text{Donc } R_g = \sqrt{-3 \text{pente}}$$

Cela signifie concrètement que, pour de très petits angles  $q$  (zone appelée aussi la région de Guinier), le graphique de Guinier est linéaire ( $I$  en fonction de  $q^2$ ) (**Figure 37A**). À partir de la pente de l'ajustement linéaire, on peut déterminer le  $R_g$  et l'ordonnée à l'origine  $I(0)$ .  $I(0)$  est proportionnelle à la masse moléculaire et à la concentration de l'échantillon analysé. Sa valeur est une extrapolation de l'intensité de diffusion à angle zéro car, à angle nul, le faisceau incident est bloqué par un masque appelé « beam stop » qui protège le détecteur. La région de linéarité du graphique de Guinier est dépendante de la forme et de la flexibilité de la molécule analysée. En effet, l'approximation est valable pour une plus large gamme de  $q$  pour une petite molécule. Traditionnellement, il est estimé que le graphique de Guinier est linéaire pour les protéines globulaires pour des valeurs de  $q$  telles que le produit  $qR_g < 1.3$  et pour des protéines allongées ou désordonnées pour des valeurs telles que  $qR_g < 1.1$ .

Une absence de linéarité dans la région de Guinier est caractéristique d'un échantillon agrégé ou qui aurait subi des dommages liés aux rayonnements (**Figure 37B**), mais également de l'existence d'interactions intermoléculaires propres à la particule (**Figure 37C**) ou encore d'un problème de réduction du solvant. En cas d'interactions intermoléculaires, les molécules étudiées ne peuvent plus être traitées comme des particules isolées, mais présentent une certaine diffusion entre les molécules. Une variation de la valeur de  $R_g$  au travers du pic d'éluion de l'échantillon est également révélatrice de la polydispersité de l'échantillon, de son agrégation ou de l'existence d'interactions.



**Figure 37 : Graphiques de Guinier caractéristiques.** Ces graphiques sont obtenus pour (A) un bon échantillon ne montrant pas d'agrégation, (B) un échantillon agrégé et (C) différentes concentrations du même échantillon soumis à des interactions intermoléculaire à forte concentration (figure adaptée de (Putnam et al., 2007))

Une fois l'approximation de Guinier établie, il est alors possible de tracer plusieurs graphiques afin, notamment, de modéliser la forme de l'échantillon et celle des régions désordonnées.

### ***(B) Fonction de distribution des paires de distances, P(r)***

Une particule en solution peut être décrite par la fonction de distribution P(r). C'est un histogramme de distribution des distances entre paires d'atomes au sein d'une particule. La courbe de diffusion de SAXS est mesurée dans l'espace réciproque (Svergun and Koch, 2003). Afin d'obtenir des informations dans l'espace réel sur la macromolécule, on peut appliquer une transformée de Fourier sur les données d'intensité de diffusion de la particule I(q) décrite par l'équation :

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

La fonction de distribution ne peut pas être directement calculée par cette équation. En effet, l'étendue limitée des mesures réalisées ainsi que l'augmentation du bruit de fond pour des valeurs de q plus élevées vont fortement affecter la transformation directe de Fourier des valeurs de I(q) et déformer la fonction de distribution P(r). C'est pourquoi l'approche traditionnelle consiste à ajuster progressivement la fonction aux données en utilisant l'équation suivante par une transformée indirecte de Fourier :

$$I(q) = 4\pi \int_0^{D_{MAX}} P(r) \frac{\sin(qr)}{qr} dr$$

où  $D_{MAX}$  est la dimension maximale de la molécule.

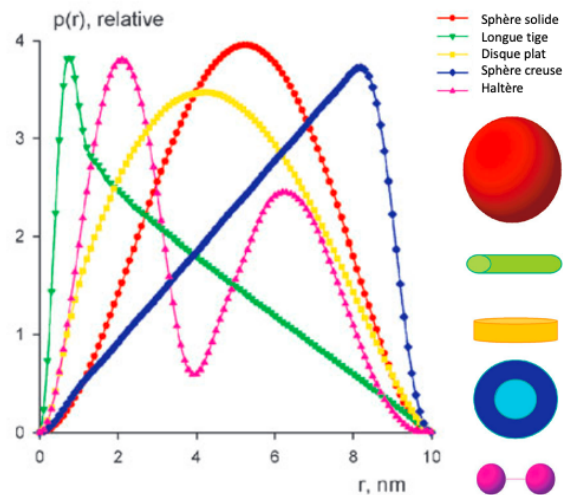
En utilisant cette équation, on génère la meilleure fonction décrivant les données expérimentales et les valeurs de P(r) sont extrapolées pour  $q = 0$  et  $q \rightarrow \infty$ . Elle est égale à 0 pour P(0) et pour les distances supérieures à  $D_{MAX}$  (la distance entre paire d'atomes maximale mesurée). Un certain nombre de programmes permettent d'estimer la fonction de distribution dont parmi eux le programme GNOM de la suite ATSAS (Svergun, 1992) ou la méthode BIFT (Bayesian Indirect Fourier Transform) du logiciel RAW (Hansen, 2000).

La fonction P(r) contient des informations précieuses sur la forme et la taille d'une macromolécule. Tout d'abord, en réalisant la fonction P(r), nous obtenons une estimation de la dimension interatomique maximale de la macromolécule ( $D_{MAX}$ ). Elle fournit également un autre moyen, potentiellement plus précis, de calculer le Rg et I(0). La forme de la fonction P(r) peut également être directement interprétée en termes de forme de la macromolécule, en fournissant des informations sur la forme globale, par exemple globulaire ou en forme de bâtonnet, ou si la macromolécule contient plusieurs domaines (**Figure 38**). Une molécule globulaire aura une P(r) en forme de cloche symétrique dont le maximum est situé à  $\frac{D_{MAX}}{2}$ . Les molécules désordonnées ou allongées présentent une courbe P(r) moins symétrique dont



le maximum est inférieur à  $\frac{D_{MAX}}{2}$ . Les protéines présentant plusieurs domaines sont caractérisées par une succession de maxima.

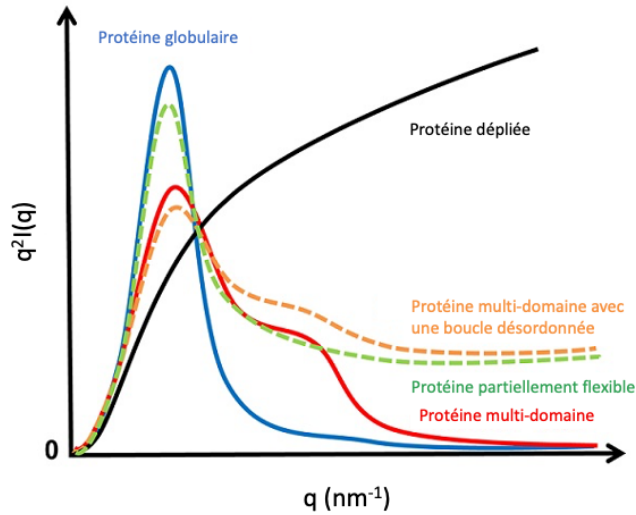
En outre, la fonction  $P(r)$  et les paramètres dérivés tels que la  $D_{MAX}$  sont nécessaires pour de nombreuses techniques d'analyse avancées, notamment les reconstructions ab-initio de la forme.



**Figure 38 :** Courbes caractéristiques des fonctions de distribution des paires de distances. Chaque courbe de couleur correspond à une courbe caractéristique pour différents corps géométriques de la même couleur représentés à droite de l'image (Figure de Svergun and Koch, 2003)

### ***(C) Le graphique sans dimension de Kratky***

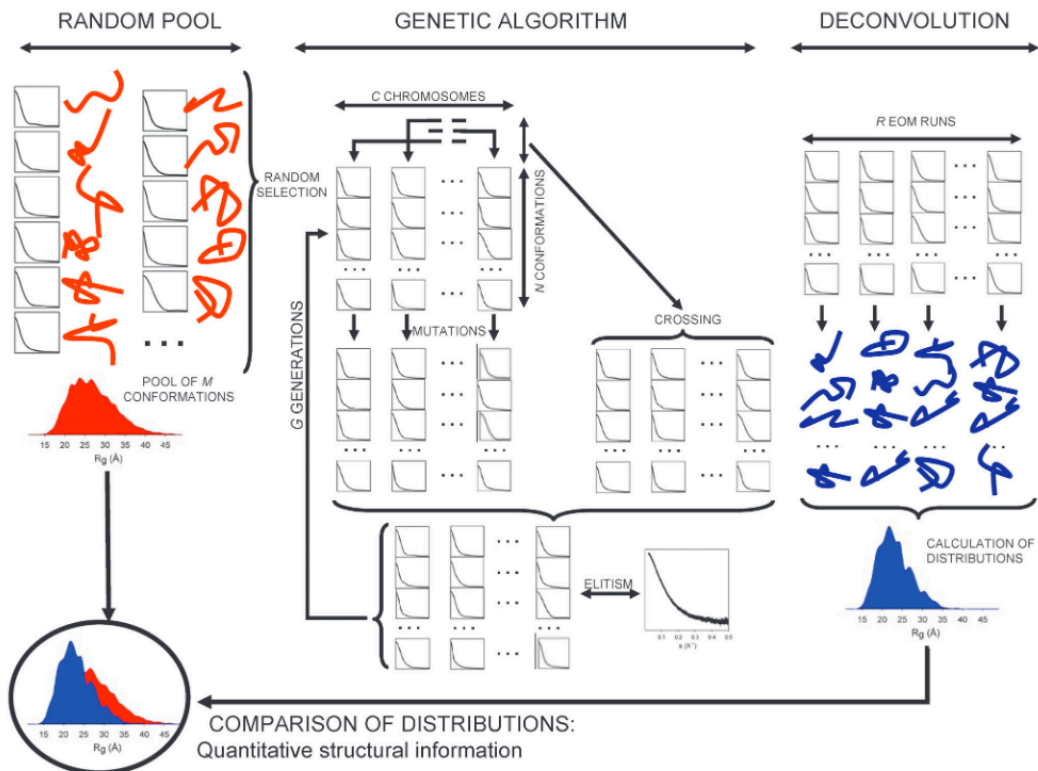
La représentation de Kratky permet d'identifier qualitativement la présence de domaines désordonnés au sein de la molécule étudiée, informe sur l'état de repliement de la particule et sur sa flexibilité en solution (**Figure 39**). Cette représentation de  $q^2I(q)$  en fonction de  $q$  aura une forme en cloche avec un maximum bien défini pour une particule structurée et globulaire tandis qu'une protéine désordonnée ou étendue présentera un plateau pour de grandes valeurs de  $q$  suivi d'une courbe croissante. Une particule partiellement désordonnée présentera un profil intermédiaire. La courbe du graphique de Kratky atteint un plateau qui décroît ensuite et tend vers les abscisses. On peut normaliser ce graphique en traçant  $Rg^2q^2I(q)/I(0)$  en fonction de  $Rg \cdot q$ . Ceci permet de normaliser le graphique par la masse et la concentration de l'échantillon.



**Figure 39 : Représentation schématique des courbes du graphique de Kratky.** Chaque courbe correspond à un échantillon présentant des caractéristiques structurales spécifiques détaillées sur la figure.

**(D) La méthode d'Optimisation d'Ensembles (EOM)**

Dans le cas des protéines intrinsèquement désordonnées (IDP), le développement de modèles décrivant de façon réaliste les données de diffusion de SAXS a longtemps été très limité. Ces protéines présentent de nombreuses conformations en solution autour d'une structure globulaire moyenne.



**Figure 40 : Représentation schématique de la stratégie EOM pour l'analyse de données SAXS dans le cas du  $R_g$ .** Un nombre  $M$  de conformations sont générées à partir de la courbe de données SAXS (partie gauche de la figure) avec une distribution aléatoire. Ces  $M$  conformations sont ensuite

utilisées pour alimenter l'algorithme génétique sur un nombre de générations  $G$ . Le processus complet est répété  $R$  fois indépendamment (déconvolution) et chaque passage fournit un nombre  $N$  de structure par courbe sélectionnées qui correspondent au profil expérimental. L'analyse des conformations résultantes est présentée à droite. On compare en bas à gauche de la figure la distribution des  $R_g$  des conformations sélectionnées par rapport au pool initial (Figure de Bernadó and Svergun, 2012).

La Méthode d'Optimisation d'Ensemble (EOM) (Bernadó *et al.*, 2007; Bernadó and Svergun, 2012; Tria *et al.*, 2015) comporte deux étapes consécutives (**Figure 40**) :

(i) La génération informatique par le script RANCH de dizaines de milliers de conformères aléatoires. La génération des domaines repliés de ces modèles peut s'appuyer sur des structures résolues par cristallographie ou sur des modèles tridimensionnels générés par d'autres méthodes comme AlphaFold2 (Mirdita *et al.*, 2022) afin de générer des conformères décrivant au mieux à priori la structure étudiée. Pour chaque conformère généré, la méthode calcule le profil de diffusion qui lui serait associé en utilisant le programme CRY SOL (Svergun, Barberato and Koch, 1995).

(ii) La sélection par le script GAJOE, parmi cet ensemble, d'un certain nombre de conformères qui reproduisent collectivement le profil expérimental de diffusion SAXS en utilisant l'équation suivante :

$$I(q) = \sum_{k=1}^K v_k I_k(q)$$

où  $v_k$  et  $I_k(q)$  sont, respectivement, les volumes et les intensités de diffraction des composants  $k$ .

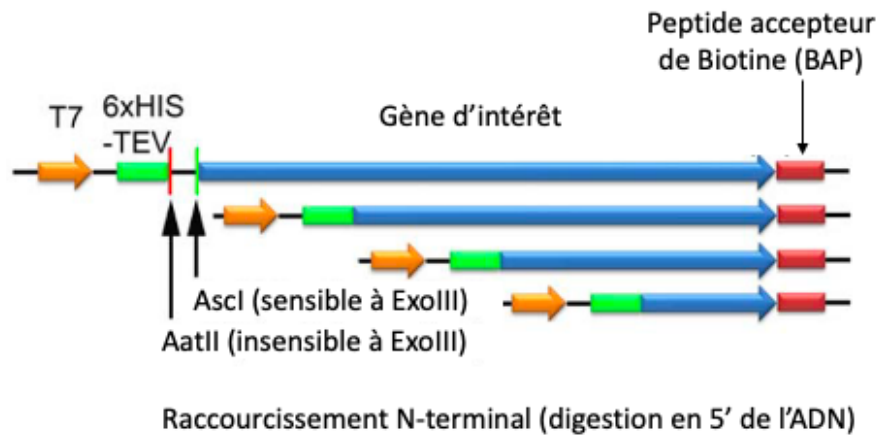
Cette équation permet de décrire au mieux les systèmes présentant de la polydispersité ou de la flexibilité car la mesure de l'intensité de diffusion totale dépend de la moyenne de la mesure de l'intensité de diffusion de chacune des particules présentes dans l'échantillon. Le profil de diffusion SAXS final est une combinaison linéaire représentant la contribution de chacune des particules de l'échantillon étudié. Cette méthode permet donc de sélectionner un sous-ensemble de conformères parmi un ou plusieurs ensembles permettant de décrire au mieux les propriétés de la particule étudiée en solution en termes de  $R_g$  et de  $D_{MAX}$ .

## V. La génération à haut débit de clones de protéines par la méthode ESPRIT

La méthode ESPRIT (Expression of Soluble Proteins by Random Incremental Truncations) permet la génération à haut débit par mutagenèse dirigée d'une librairie de mutants aléatoirement tronqués ou fragmentés pour l'identification de mutants solubles de protéines difficile à exprimer et/ou à solubiliser (An *et al.*, 2011). Cette plateforme est disponible via le Partenariat pour la Biologie Structurale (PSB).

Les plasmides utilisés pour la construction de la librairie sont dérivés du vecteur pET9a qui code pour un gène d'intérêt fusionné, dans le cas d'une mutagenèse unidirectionnelle en N-terminal, avec une étiquette 6HIS hors du cadre de lecture du gène en

5' et une séquence d'interaction avec la biotine (BAP) dans le cadre de lecture en 3'. Deux sites de restriction AatII et AscI sont insérés entre l'étiquette 6HIS et le gène d'intérêt. Le site AatII est insensible à l'action de l'exonucléase III en opposition au site AscI qui permet l'activité nucléase en direction du gène d'intérêt (**Figure 41**).

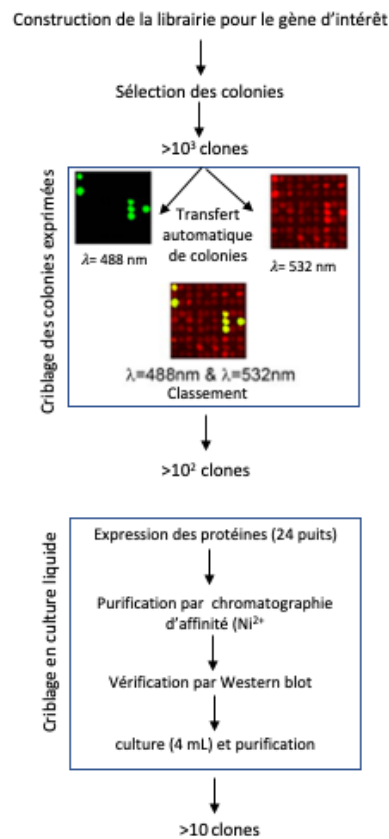


**Figure 41 : ADN plasmidique utilisé par la méthode ESPRIT.** La construction contient un insert de gène PB2 pré-cloné et linéarisé par une paire d'enzymes de restriction : AatII qui produit un site résistant à l'exonucléase III, AscI qui est un substrat pour l'exonucléase III. Une digestion au fil du temps est effectuée pour produire un ensemble de constructions uniformément tronquées. La nucléase de haricot mungo et la polymérase Pfu sont ensuite utilisées pour la digestion dirigée du plasmide (Figure adaptée de An et al., 2011).

Dans un premier temps, une librairie de clones tronqués de la protéine d'intérêt est générée par digestion à l'exonucléase III. Les fragments d'ADN générés sont ensuite reclés grâce à la nucléase d'haricots mungo et la polymérase Pfu dans le vecteur d'expression ESPRIT et transformés en *E. coli* BL21\* (DE3).

Un premier criblage permettant de sélectionner les clones solubles et exprimant les étiquettes de purification est effectué (**Figure 42**). Pour cela, les colonies de la librairie d'expression sont prélevées et réparties sur des membranes de nitrocellulose où l'expression des protéines est induite. Le niveau d'expression et la solubilité des clones sont évalués par la mesure des niveaux d'expression par immunofluorescence des étiquette N- ( $\lambda=532\text{nm}$ ) et C-terminales du site BAP biotinylé ( $\lambda=488\text{nm}$ ).

Les clones présentant les meilleurs niveaux d'expression sont ensuite exprimés en culture liquide de 4mL et purifiés par chromatographie d'affinité en résine Nickel (NiNTA). Les protéines éluées sont ensuite analysées par western blot et SDS-PAGE.



**Figure 42 : Étapes de la méthode ESPRIT** (Figure adaptée de An *et al.*, 2011)

## VI. La production d'anticorps à domaine unique spécifiques de la protéine CΔ28 du virus Nipah

Un anticorps à domaine unique (SdAb) correspond à un fragment d'anticorps produit par les camélidés composé d'un seul domaine variable. Ces SdAb ont un poids moléculaire de seulement 12 à 15 kDa, beaucoup plus petits que les anticorps communs, et sont extrêmement résistants à une diversité de pH, de tampons et de températures. Ces anticorps conservent la capacité de liaison à leurs antigènes originaux au même titre que les anticorps communs.

Les sdAb sont des chaperons reconnus pour la cristallisation de protéines et de complexes difficiles à cristalliser tels que les protéines membranaires, les assemblages de multi-protéines et les protéines intrinsèquement désordonnées. Ils sont également utilisés comme sondes structurales permettant la stabilisation et le repliement des protéines (Pardon *et al.*, 2014).

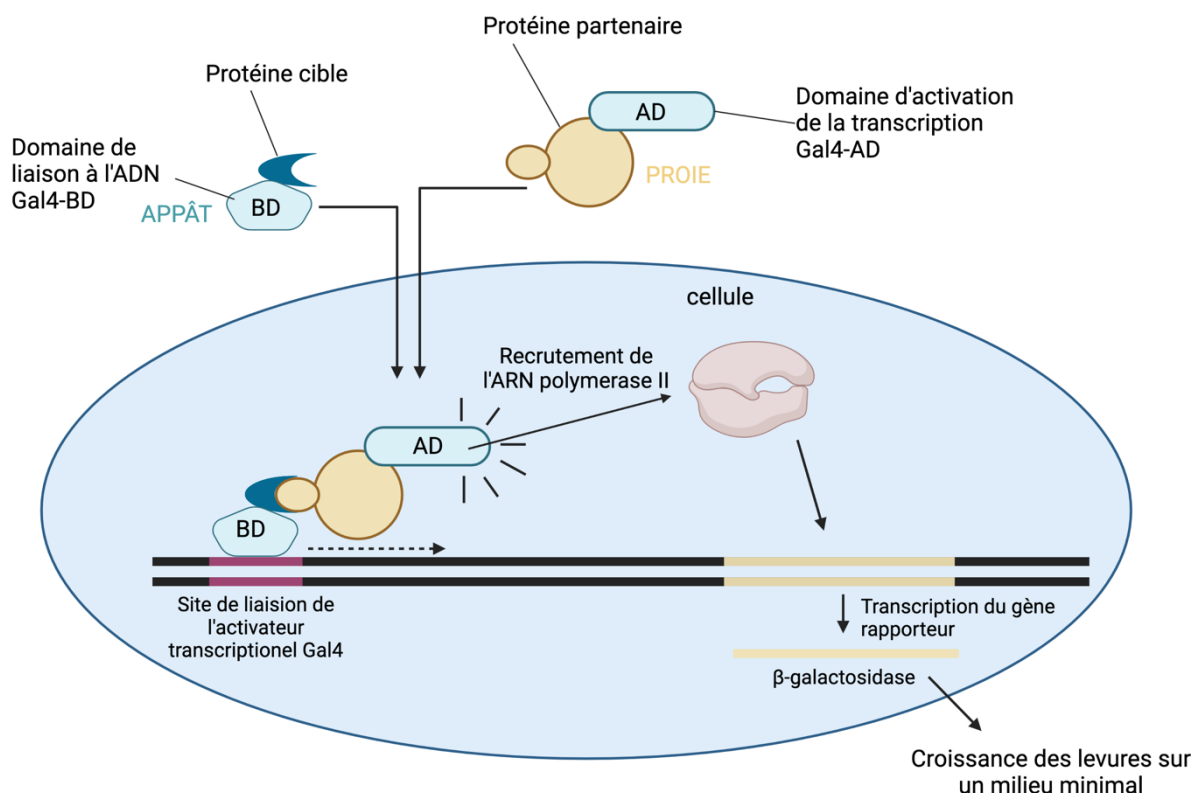
À la suite de l'immunisation, le sang de l'animal immunisé est prélevé. Cet échantillon contient des cellules B exprimant une grande diversité de SdAb maturés présentant une affinité pour les antigènes injectés. L'ADN codant pour les SdAb est isolé et amplifié par PCR à partir des lymphocytes du sang périphérique avant d'être cloné dans un vecteur de présentation sur phage (phage display). La protéine cible est ensuite isolée à la

surface d'un puits de plaque à micro-titrage et mise en contact avec les phages modifiés. En cas d'interaction entre le SdAb exprimé à la surface du phage et la protéine cible, le phage reste fixé sur le puits. Le phage est alors récupéré et utilisé pour infecter des bactéries afin d'être amplifié. L'ADN codant pour le SdAb des phages sélectionnés est ensuite séquencé. Ces nanocorps sont fusionnés avec une étiquette hexahistidine (6HIS) C-terminale.

Ils sont ensuite exprimés dans le périplasma d'*E. coli* WK6 et purifiés par chromatographie d'affinité sur résine Ni<sup>2+</sup>NTA. Par ce protocole, des familles de SdAb sont identifiées. Une famille correspond à un groupe de SdAb ayant plus de 80% d'identité de séquence au niveau de leur domaine CDR3. Ces familles sont dérivées de la même lignée de cellules B et se lient au même épitope.

## VII. Identification de partenaires cellulaires par la méthode du double-hybride en levure

L'étude d'interactome par la méthode du double-hybride en levure (Y2H) *Saccharomyces cerevisiae* permet la détection haut-débit d'interactions de protéines dans un environnement natif (Brückner *et al.*, 2009).



**Figure 43 :** L'étude d'interactome par la méthode du double-hybride en levure (Y2H). Représentation schématique de l'utilisation de la méthode Y2H pour l'identification de protéines partenaires (proie) d'une protéine d'intérêt (cible). La protéine cible est cloné derrière un domaine de liaison fusionnée au domaine N-terminal de liaison à l'ADN de Gal4 appelé BD (Binding Domain). La protéines partenaire est fusionnée au domaine C-terminal d'activation AD de gal4. L'interaction des protéines appât et cible, après transformation en cellule, permet la reconstitution du facteur de transcription Gal4 fonctionnel qui peut alors recruter la polymérase II et transcrire le gène Gal1-

*lacZ*. La production de  $\beta$ -galactosidase permet aux levures de croître sur un milieu minimal (Figure adaptée de <https://www.hybrigenics-services.com/protein-interaction-discovery>).

Cette méthode repose sur l'activation d'un gène rapporteur par le facteur de transcription Gal4 en présence de galactose. La protéine d'intérêt, appelée appât X, est fusionnée au domaine N-terminal de liaison à l'ADN de Gal4 appelé BD (Binding Domain) capable de reconnaître une séquence spécifique de l'ADN promotrice du gène rapporteur. La protéine testée en interaction, appelée proie Y, est fusionnée au domaine C-terminal d'activation AD de Gal4. Il a été montré que l'interaction des protéines X et Y permet la reconstitution d'un facteur de transcription Gal4 fonctionnel capable de recruter l'ARN polymérase II activant ainsi la transcription du gène GAL1-*lacZ*. Ce gène code pour l'enzyme  $\beta$ -galactosidase. Le substrat histochimique bluo-Gal (indolyl- $\beta$ -galactoside halogéné) présent dans le milieu de croissance des levures est alors hydrolysé par la  $\beta$ -galactosidase produisant un précipité bleu colorant la colonie de levure.







# Chapitre 4 : Étude biophysique et structurale de la protéine C du virus de la toupaye

## I. Introduction

Suite à l'étude bio-informatique menée à l'échelle des *Orthoparamyxovirinae*, nous avons souhaité caractériser la structure d'une protéine C appartenant au groupe II identifié dans le but de confirmer l'existence de similarités structurales avec les protéines C des autres groupes et plus largement de la sous-famille. L'étude de Lo *et al.* (Lo, Søggaard and Karlin, 2014) a permis d'identifier les candidats, parmi 20 protéines C de différents virus, les plus prometteurs pour la production en bactérie. La protéine C du TupV en fait partie, ce qui a justifié notre choix.

Dans cette partie, nous présentons les résultats d'expression, de purification et de caractérisation structurale de la protéine C du virus de la toupaye dans sa forme entière  $C_{FL}$  et d'une construction correspondant au domaine structuré C-terminal,  $C_{\Delta 53}$ . Ces fragments ont été construits sur les bases d'une étude bio-informatique minutieuse des prédictions de structure et de désordre de la protéine C du TupV qui a confirmé, pour cette protéine C aussi, la présence d'une région C-terminale riche en hélices  $\alpha$  (résidus 54 à 150) et d'une région N-terminale majoritairement désordonnée, riche en résidus chargés et en prolines.

Ces résultats ont ensuite fait l'objet d'une étude d'homologies structurales avec la protéine C du SeV et ont fait l'objet d'une publication soumise en janvier 2023 au journal *Biomolecules*.

## II. Article : Les protéines C des *Orthoparamyxovirinae* partagent une origine et une organisation structurale communes

Le PDF correspondant à l'article présenté dans cette partie est disponible à l'adresse suivante : <https://www.mdpi.com/2218-273X/13/3/455>.

Article

# Orthoparamyxovirinae C Proteins Have a Common Origin and a Common Structural Organization

Ada Roy <sup>1</sup>, Emeric Chan Mine <sup>2</sup>, Lorenzo Gaifas <sup>1</sup>, Cédric Leyrat <sup>3</sup>, Valentina A. Volchkova <sup>2</sup>, Florence Baudin <sup>4</sup>, Luis Martinez-Gil <sup>5</sup>, Viktor E. Volchkov <sup>2</sup>, David G. Karlin <sup>6,\*</sup>, Jean-Marie Bourhis <sup>1,\*</sup> and Marc Jamin <sup>1,\*</sup>

- <sup>1</sup> Institut de Biologie Structurale, Université Grenoble Alpes, CNRS, CEA, 38000 Grenoble, France
  - <sup>2</sup> Molecular Basis of Viral Pathogenicity, Centre International de Recherche en Infectiologie (CIRI), INSERMU1111-CNRS UMR5308, Université Claude Bernard Lyon 1, ENS de Lyon, 69365 Lyon, France
  - <sup>3</sup> Institut de Génétique Fonctionnelle, Université de Montpellier, CNRS, INSERM, 34094 Montpellier, France
  - <sup>4</sup> Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany
  - <sup>5</sup> Department of Biochemistry and Molecular Biology, Institute for Biotechnology and Biomedicine (BIOTECMED), University of Valencia, 4610 Valencia, Spain
  - <sup>6</sup> Division Phytomedicine, Thae Institute of Agricultural and Horticultural Sciences, Humboldt-Universität zu Berlin, Lentzeallee 55/57, 14195 Berlin, Germany
- \* Correspondence: davidgkarlin@gmail.com (D.G.K.); jean-marie.bourhis@ibs.fr (J.-M.B.); marc.jamin@ibs.fr (M.J.); Tel.: +33-4-57-42-86-36 (J.-M.B.); +33-4-76-20-94-62 (M.J.)



**Citation:** Roy, A.; Chan Mine, E.; Gaifas, L.; Leyrat, C.; Volchkova, V.A.; Baudin, F.; Martinez-Gil, L.; Volchkov, V.E.; Karlin, D.G.; Bourhis, J.-M.; et al. *Orthoparamyxovirinae C Proteins Have a Common Origin and a Common Structural Organization*. *Biomolecules* **2023**, *13*, 455. <https://doi.org/10.3390/biom13030455>

Academic Editors: Vladimir N. Uversky, Elisar Barbar and Nathalie Sibille

Received: 11 January 2023

Revised: 21 February 2023

Accepted: 22 February 2023

Published: 1 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The protein C is a small viral protein encoded in an overlapping frame of the P gene in the subfamily Orthoparamyxovirinae. This protein, expressed by alternative translation initiation, is a virulence factor that regulates viral transcription, replication, and production of defective interfering RNA, interferes with the host-cell innate immunity systems and supports the assembly of viral particles and budding. We expressed and purified full-length and an N-terminally truncated C protein from Tupaia paramyxovirus (TupV) C protein (genus Narmovirus). We solved the crystal structure of the C-terminal part of TupV C protein at a resolution of 2.4 Å and found that it is structurally similar to Sendai virus C protein, suggesting that despite undetectable sequence conservation, these proteins are homologous. We characterized both truncated and full-length proteins by SEC-MALLS and SEC-SAXS and described their solution structures by ensemble models. We established a mini-replicon assay for the related Nipah virus (NiV) and showed that TupV C inhibited the expression of NiV minigenome in a concentration-dependent manner as efficiently as the NiV C protein. A previous study found that the Orthoparamyxovirinae C proteins form two clusters without detectable sequence similarity, raising the question of whether they were homologous or instead had originated independently. Since TupV C and SeV C are representatives of these two clusters, our discovery that they have a similar structure indicates that all Orthoparamyxovirinae C proteins are homologous. Our results also imply that, strikingly, a STAT1-binding site is encoded by exactly the same RNA region of the P/C gene across Paramyxovirinae, but in different reading frames (P or C), depending on which cluster they belong to.

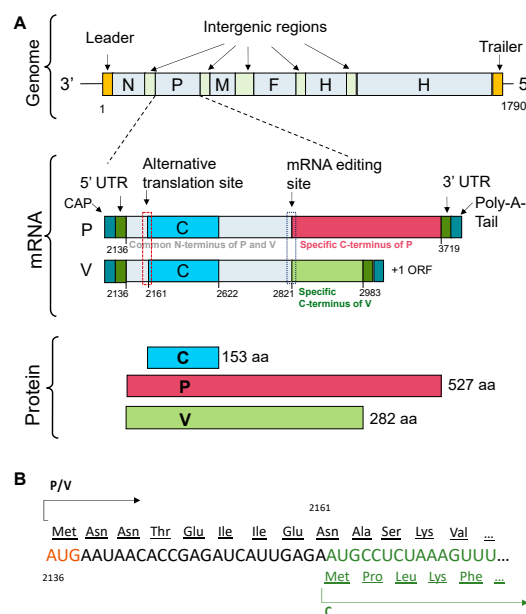
**Keywords:** Paramyxoviridae; virulence factor; overlapping genes; protein structure; viral evolution

## 1. Introduction

The large virus family Paramyxoviridae includes well-known human and livestock pathogenic viruses, such as measles and mumps viruses, parainfluenza viruses or rinderpest and Newcastle disease viruses [1] but also zoonotic viruses such as Hendra and Nipah viruses [2], as well as a large number of understudied viruses infecting mammals, reptiles, and fishes [3,4]. They are enveloped viruses with a negative-sense RNA genome, which are therefore classified in the phylum Negarnaviricota [5]. On the basis of the non-segmented

character and organization of their genome, their replication mechanisms, and multiple sequence alignments of individual genomes or proteins, the family Paramyxoviridae has been classified in the order Mononegavirales along with the families Pneumoviridae (human respiratory syncytial virus, human metapneumovirus), Filoviridae (Ebola and Marburg viruses), and Rhabdoviridae (rabies virus) [6,7].

A hallmark of the family Paramyxoviridae is the architecture of the P gene, which comprises several coding regions in overlap in different reading frames [1]. The V coding region appeared to be an ancestral feature of this viral family [8], and most current paramyxoviruses produce P and V proteins by a co-transcriptional editing mechanism, in which the viral L RNA-dependent RNA polymerase (L-RdRP) stutters at a specific U-rich sequence, leading to the non-templated insertion of one or more guanosines (G) (in TupV insertion occurs at nt 2821) [9–13] (Figure 1).



**Figure 1.** Architecture of the Tupaia paramyxovirus (TupV) genome and expression mechanisms of its P, V, and C proteins. (A) The upper panel shows a schematic representation of TupV genome, containing from the 3' end to the 5' end, the genes coding for the N, P, M, F, H, and L proteins. The middle panel shows the architecture of the mRNAs generated from the P gene by editing at the nucleotide 2821. The mRNA coding for the P protein is unedited, whereas that coding for the V protein has one G inserted at the editing site. Both mRNA contain the 3' and 5' UTR and are processed at both extremities with a CAP structure at the 3' end and a polyA tail at the 5' end. The lower panel shows the proteins expressed from these mRNAs. The P and V proteins are expressed by translation initiation at the first AUG codon of their respective mRNA, whereas the C protein is expressed by alternative translation initiation at the next downstream AUG codon from both mRNAs. (B) Excerpt of the mRNA coding sequence (of both P and V mRNAs) showing the first AUG initiation codon used for the translation of P or V proteins and the second AUG codon, located in the +1 frame, used for the translation for the C protein.

Members of the subfamily Orthoparamyxovirinae, which includes the genera Morbillivirus, Henipavirus, Jeilongvirus, Respirovirus, and Narmovirus, possess an additional C coding region in overlap with the 3' end of the P coding region, and thus, the mRNAs

encoding P, V, and possibly W, are polycistronic (Figure 1). In most viruses, the P coding region starts upstream of the C coding region, but in some viruses, the opposite occurs [1]. The P coding region and the C coding region are expressed by an alternative translation initiation mechanism; the ribosome fails to initiate at the first AUG start codon because of a weak Kozak sequence and initiates at the next downstream AUG codon, which happens to be in the second reading frame, leading to the expression of the other protein (Figure 1).

The C proteins of Orthoparamyxovirinae are virulence factors [14,15], which act through different mechanisms. Most C proteins have at least three functions: (1) they inhibit viral RNA synthesis, thereby controlling the rate of replication of the virus (in some viruses, they also regulate the balance between transcription and replication) and the production of defective interfering (DI) particles (by enhancing polymerase processivity) [16–18]; (2) they counteract antiviral innate immune responses (high degree of attenuation in KO virus) [14,19,20] and (3) they recruit cellular partners in particular, to promote viral particle assembly and budding (by recruiting ESCRT components) [21].

A previous sequence analysis revealed an architecture common to all Orthoparamyxoviridae C proteins, with a predicted N-terminal disordered moiety and a C-terminal helical moiety [22]. This analysis identified three groups of C proteins, represented by the measles virus, the Nipah virus, and the Sendai virus. The C proteins of the measles virus group and Nipah virus group are homologous; in contrast, the sequence and functional properties of the C proteins from the Sendai viruses were different from those of the other two groups, suggesting that the C protein might have arisen independently in the ancestor of the measles and Nipah viruses and in the ancestor of the Sendai virus group [22].

The C protein of Orthoparamyxovirinae, in particular of important pathogens (e.g., measles virus or Nipah virus), has resisted structure determination for over 20 years. The major bottleneck was the production of recombinant C protein, and it is known that in such cases, testing different orthologs increases the chances of solving the structure of at least one of them. Therefore, the same study tested the bacterial expression of the C protein of over 20 Orthoparamyxovirinae [22]. The most abundantly expressed C protein was that of Tupaia paramyxovirus (TupV), a virus that infects the tree shrew (*Tupaia belangeri*). Of interest for our study is that its P gene exhibited typical features of members of the subfamily Orthoparamyxovirinae [13] (Figure 1).

The tree shrew (*Tupaia belangeri*) is a small mammal that lives in a tropical rainforest in Southeast Asia and belongs to the order Scandentia. Recent genome sequencing confirmed that it is genetically closer to primates (Primates) than to rodents (Rodentia) [23], making the tree shrew a promising model for a variety of human diseases such as depression, myopia, hepatitis B and C infections, and hepatocellular carcinoma, as well as the development of therapeutic treatments, because it combines small size, easy breeding, and rapid reproduction [24].

Here, to obtain structural information about a full-length C protein and to further investigate the evolutionary relationships between the Orthoparamyxovirinae C proteins, we expressed two constructs of the TupV C protein, we solved the crystal structure of its C-terminal folded part and characterized both the truncated and full-length proteins in solution, confirming that the N-terminal part was intrinsically disordered. Comparison with the folded domain of Sendai virus C protein revealed a common structural architecture, which suggested a descent from a common ancestor and led us to propose different possible evolution models for explaining the existence of C protein coding region in the current members of the sub-family Orthoparamyxovirinae.

## 2. Materials and Methods

### 2.1. Disorder Metaprediction

The location of intrinsically disordered regions (IDRs) in TupV C protein was predicted by submitting its amino-acid sequence to 16 different predictors accessible through web servers and by calculating a consensus prediction as previously described [25]. The fold index was run with a window size of 51 residues [26]. For the PONDR VLXT, XL1,

CAN\_XT, VL3\_BA, and VSL2 predictors, residues with a score > 0.5 were considered disordered [27–30]. PONDR-FIT was run using the default parameters [28]. For the IUPred2 predictors for long and short-disordered regions, residues with a score > 0.5 were considered disordered [31]. The DisEMBL (Loops/Coils, Hot-loops, and Remark465) server was run using the default parameters [32]. The GLOBPLOT server was run with the default Russel/Linding propensity [33]. The ESpritz predictors (NMR, X-ray, Disprot) were used with the default parameters [34]. HCA (Hydrophobic Cluster Analysis) plots were drawn with DRAWHCA [35].

A simple scoring procedure was then 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 averaging the value from all predictions. Regions with a normalized D-score lower than 0.50 are considered intrinsically disordered, whereas regions with a score higher than 0.5 are either folded domains or potential molecular recognition elements (MoRE) for partner proteins [25,36–39]. A freely accessible GitHub for automatically querying the different servers, processing the information, and calculating the D-score is now accessible at <https://github.com/brisivag/dscore> (accessed on 20 December 2022).

## 2.2. Protein Expression and Purification

A gene encoding full-length Tupaia paramyxovirus C protein (Uniprot: Q9WS38) in fusion with a C-terminal 3C protease cleavage site and an N-terminal 6His tag was obtained in a PopinF vector and was expressed and purified as previously described [22]. Briefly, the protein was expressed in *Escherichia coli* Rosetta™ (DE3) cells (Novagen, Darmstadt, Germany). The bacteria were grown at 25 °C for 20 h in auto-induction ZYM5052 medium [40]. The cells were harvested by centrifugation, and the pellet was resuspended in 50 mM Tris-HCl buffer at pH 7.5 containing 300 mM NaCl, 5 mM imidazole, 50 mM arginine, 50 mM glutamate, 0.2 mM of Tris(2-carboxyethyl)phosphine (TCEP), and EDTA-free protease inhibitor cocktail (Buffer A). The cells were broken by sonication, and the soluble fraction was loaded onto a Ni<sup>2+</sup> column (NiNTA, Qiagen, Hilden, Germany) equilibrated in buffer A. After washing with buffer A, the protein was eluted with 50 mM Tris-HCl buffer at pH 7.5 containing 500 mM NaCl, 250 mM imidazole, 50 mM arginine, 50 mM glutamate, and 0.2 mM of TCEP (Buffer B). The protein solution was incubated overnight at 4 °C with the 3C protease to remove the 6His tag and was concomitantly dialyzed (molecular weight cut-off of 3.5 kDa) against 50 mM Tris-HCl pH 7.5 containing 50 mM NaCl and 0.2 mM TCEP (Buffer C). The protein was then loaded on a Hitrap Q Sepharose Fast Flow column (Cytiva, Freiburg im Breisgau, Germany) equilibrated with buffer C, and then eluted with a gradient of 50 to 300 mM NaCl solution. After concentration with a centrifugal filter (molecular weight cut-off of 10 kDa—Amicon, Burlington, MA, USA), the protein was finally purified by size-exclusion chromatography (SEC) on a Hiloal Superdex 75 16/600 GL (Cytiva) column equilibrated with 50 mM Tris-HCl pH 7.5, containing 150 mM NaCl and 0.2 mM TCEP (Buffer D).

According to our bioinformatics analysis, a construct of the TupV C protein encompassing residues 54 to 153 (C<sub>A53</sub>) was subcloned between NcoI and XhoI restriction sites in a pET-M40 in fusion with TEV protease-cleavable N-terminal 6His tag and maltose binding protein (MBP) tags. The protein was expressed in *E. coli* BL21(DE3). The cells were grown in Luria Bertani (LB) medium at 37 °C until the optical density at 600 nm reached a value of 0.6. Protein expression was then induced by adding 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cells were allowed to grow for 18 h at 18 °C. The cells were harvested by centrifugation, and the pellet was resuspended in buffer A. The cells were broken by sonication, and the soluble fraction was loaded onto a Ni<sup>2+</sup> column (NiNTA, Qiagen) equilibrated in buffer A. After washing with 50 mM Tris-HCl buffer at pH 7.5 containing 1 M NaCl, 15 mM imidazole, 50 mM arginine, 50 mM glutamate, 0.2 mM of TCEP, and EDTA-free protease inhibitor cocktail (Buffer E), the proteins were eluted with 50 mM Tris-HCl buffer at pH 7.5 containing 300 mM NaCl, 250 mM imidazole,

50 mM arginine, 50 mM glutamate, and 0.2 mM of TCEP (Buffer F). The protein sample was incubated overnight at 4 °C with the TEV protease to remove the 6His and MBP tags. To remove all traces of cleaved MBP, the sample was applied on an amylose column equilibrated with buffer F, and the flowthrough was collected and further purified by SEC on a Superdex 75 (Cytiva) equilibrated in buffer B. Purity of the protein samples was routinely checked by SDS-PAGE and protein concentration was determined by UV spectroscopy or by refractometry during an SEC-MALLS analysis (see below).

### 2.3. Size Exclusion Chromatography Coupled with Multiangle Laser Light Scattering Experiments [41]

SEC was performed on a Superdex Increase 75 column (Cytiva) equilibrated with 50 mM Tris-HCl buffer at pH 7.5 containing 150 mM NaCl and 0.2 mM TCEP. The column was calibrated with globular proteins of known hydrodynamic radius ( $R_h$ ) [42] using commercially available calibration kits (Cytiva). Chromatographic separations were performed at room temperature with a flow rate of 0.5 mL·min<sup>-1</sup> with online MALLS detection using a DAWN-HELEOS II detector (Wyatt Technology, Santa Barbara, CA, USA), and protein concentration was measured online by differential refractive index measurements using an Optilab T-rEX detector (Wyatt Technology) and a refractive index increment,  $dn/dc$ , of 0.185 mL·g<sup>-1</sup>. The data were analyzed with the software ASTRA (Wyatt Technology).

### 2.4. Size Exclusion Chromatography Coupled with Small Angle X-ray Scattering Experiments [43]

Small-angle X-ray scattering (SAXS) data were collected at the BioSAXS beamline (SWING) of the Synchrotron Soleil. Samples were analyzed on a Superdex Increase 75 5/150 GL (Cytiva). The scattering from the buffer alone was measured before and after each sample measurement and used for background subtraction. All data analyses were performed using the program PRIMUS from the ATSAS package [44]. Ensembles of conformers were generated with the program RANCH, while ensemble selection was performed with the program GAJOE [45].

### 2.5. X-ray Crystallography

Two datasets were collected on the PROXIMA-1 beamline at the SOLEIL Synchrotron at the wavelength of 1.07169 Å: a native crystal diffracting at 2.4 Å and a platinum derivative obtained by soaking diffracting at 2.7 Å. For both datasets, indexing and integration were performed using the XDS program suite [46]. Initial phases were obtained with the platinum derivative dataset by the single-wavelength anomalous dispersion method using the SHELX suite through the HKL2map program [47]. The alpha carbon trace from SHELXE was then used with the native dataset to build a first model with Phenix [48]. The model was improved and refined against the native dataset with COOT [49] and Refmac5 [50] from the CCP4 program suite [51].

### 2.6. AlphaFold Predictions

An SBGrid consortium installation of AlphaFold version 2.1.2 [52,53] was run on a local server equipped with an NVIDIA Tesla P100 GPU in order to predict the structure of NiV and MeV C proteins. The full databases were used, with `max_template_date = 2022-08-29`. All other parameters were left to their default values.

### 2.7. Cell Culture

HEK293T cell line (ATCC, Manassas, VA, USA) was cultivated in Dulbecco's Modified Eagle's Medium (DMEM) GlutaMAX (Thermo Fisher Scientific, Waltham, MA, USA) containing 50 U/mL penicillin, 50 µg/mL streptomycin (Thermo Fisher Scientific) and 10% Fetal Calf Serum (FCS) (Eurobio, Les Ullis, France) at 37 °C and 5% CO<sub>2</sub> in a humidified atmosphere.

### 2.8. Plasmids Constructs

Nipah virus P, C, N, and L ORFs were amplified by PCR from the plasmids pTM1 NiV P, pTM1 NiV N, and pTPM1 NiV L (kind gifts of C. Basler, Mount Sinai School of Medicine, New York) [54] and cloned into the pHCMV vector. To avoid the expression of Nipah C from the plasmid encoding NiV P through translational leaky scanning mechanism, the weak Kozak sequence of P (CAT CCA **ATG** GAT) was replaced by the corresponding sequence of N (ATC ATC **ATG** GAT), for which no expression of proteins due to leaky scanning has been reported.

The construction of the Nipah virus minigenome was performed as described for the La Crosse virus minigenome [55]. The NiV leader region, the “gene start” signal (GS) and the untranslated region from the NiV N gene were inserted upstream from the mouse RNA polI terminator sequence. The NiV trailer, the “gene stop” signal (GE), and the untranslated region from the NiV L gene were inserted after the RNA polI promoter sequence. The NiV sequence was replaced with the Renilla luciferase gene (pPol I NiV-REN), and four nucleotides (GCAT) were inserted after the stop codon of the Renilla luciferase gene in order to respect the rule of 6 [56,57]. All expression plasmids were verified by sequencing.

### 2.9. Minireplicon Assay

The subconfluent monolayer of HEK293T cells was transfected with 100 ng plasmid pPol I NiV-REN, 125 ng pHCMV NiV N, 25 ng pHCMV NiV P, and 350 ng pHCMV NiV L using jetOPTIMUS transfection reagent (Polyplus, Illkirch-Graffenstaden, France). Additional transfection of pGL4.50 (Promega, Madison, WI, USA) encoding the firefly luciferase was performed for normalization of transfection efficiency. In all experiments, the total amount of transfected DNA was kept constant by including additional empty vector plasmid DNA where appropriate. Reporter activity was measured 24 h post-transfection using the Dual-Glo luciferase assay (Promega) following the instructions of the manufacturer. Details about the NiV mini-replicon assay will be published elsewhere.

### 2.10. TupV C Protein Interactions with the Human Proteome

For the identification of TupV C-human protein–protein interactions, DNAs encoding the C protein fused to a tandem affinity purification (TAP) tag sequence at either the amino or carboxy terminus were transfected into HEK293T cells. Briefly, 40 µg of DNA (combining at equal parts constructs with the tag at the N- and C-terminus) was added to 1.5 mL of serum-free medium (Gibco, Waltham, MA, USA) (solution A). As a control, cells were transfected with a plasmid encoding the EYFP. Separately, 1.5 mL of serum-free DMEM was supplemented with Lipofectamine 2000 (2 µL·µg<sup>-1</sup> DNA; Life Technologies, Carlsbad, CA, USA) (solution B). Solutions A and B were then mixed and incubated at room temperature for at least 10 min. The mixture was then overlaid onto a 15 cm tissue culture dish. Approximately 48 h post-transfection, cells were harvested and washed with phosphate-buffered saline (PBS) (×3). Cells were then lysed with 300 µL of lysis buffer (30 mM Tris-HCl, 150 mM NaCl, 0.5% NP-40, protease inhibitors (cOmplete protease inhibitor cocktail; Roche, Basel, Switzerland), and phosphatase inhibitor (PhosSTOP; Roche), and the lysate was passed through a 30-gauge syringe (×3). The cell extract was clarified by centrifugation (10 min, 10,000× g) and the supernatant was transferred to a new tube. TAP, using the Strep and Flag tags in tandem, was performed as described in reference [1]. Protein digestion and identification (liquid chromatography–tandem mass spectrometry [LC-MS/MS] using an LTQ Orbitrap) was completed by the Proteomic Core Facility at the University of Valencia. Four experimental replicas were done.

The MS data were filtered according to the following selection criteria. All pseudogenes, predicted proteins, and immunoglobulin fragments were discarded. Common contaminants in MS experiments, such as keratins and trypsins, were also eliminated. The proteins purified with the EYFP control were deleted. To further eliminate common artifacts, we utilized the CRAPome database (2) of common protein contaminants. We selected proteins found with two or more peptides and in at least two experimental replicates.



### 3. Results

#### 3.1. TupV C Protein Contains a Predicted Structured Region (aa 54–153)

The sequence of TupV C protein is conserved within the genus Narmovirus, except for a short, highly variable region (aa 23–34) (Figure 2A). Visual inspection of an HCA plot [35] of TupV C revealed an N-terminal part (residues 1–60) rich in charged residues and prolines, suggesting that it is globally disordered and a C-terminal part that contains five hydrophobic patches typical of  $\alpha$ -helices (Figure 2B). In order to precisely localize the boundaries between structured and disordered regions, we predicted the location of secondary structure elements with PSIPRED [58], which predicted six helices named A to F, and combined the results of multiple predictors of intrinsic disorder into a consensus D-score, as previously described [25] (Figure 2C). In agreement with previous predictions and limited proteolysis data [22], these results indicated the presence of (1) a C-terminal region comprising four  $\alpha$ -helices (aa 77–150) with a D-score value continuously above 0.5, and (2) another helical region (aa 54–72), with a lower D-score value (Figure 2C).

#### 3.2. Expression, Purification, and Quality Control of N-Terminally Truncated and Full-Length TupV C Proteins

Using this information, we generated and expressed in *E. coli* a truncated form of the C protein lacking either its 53 N-terminal residues ( $C_{\Delta 53}$ ) in addition to the full-length protein ( $C_{FL}$ ). We purified both proteins in two steps using affinity chromatography and size-exclusion chromatography, and we checked their purity by SDS PAGE. We determined the molecular mass and oligomerization state using SEC-MALLS. Both  $C_{\Delta 53}$  and  $C_{FL}$  eluted from the Superdex 75 column in a single, symmetrical peak with a constant molecular mass across the peak indicating monodisperse species (Figure 3A). The weight-average molecular mass values of  $12.0 \pm 0.3$  kDa for  $C_{\Delta 53}$  and  $17.8 \pm 0.6$  kDa for  $C_{FL}$  were in agreement with the theoretical masses of 11.9 kDa and 17.8 kDa, respectively. The polydispersity index ( $M_w/M_n$ ) of 1.000 and 1.003, respectively, confirmed that both proteins were monodisperse and monomeric in solution. We determined the hydrodynamic radius ( $R_h$ ) values of  $2.1 \pm 0.1$  nm for  $C_{\Delta 53}$  and  $2.4 \pm 0.1$  nm for  $C_{FL}$  by calibrating the SEC column with globular proteins (Figure 3B). These values were much closer to the theoretical radii of globular proteins of the same molecular mass (1.8 and 2.1 nm, respectively) than to those of chemically unfolded proteins (3.1 and 3.8 nm, respectively) (Figure 3B). These results suggest that  $C_{\Delta 53}$  and  $C_{FL}$  comprise short flexible regions but are mostly compact.

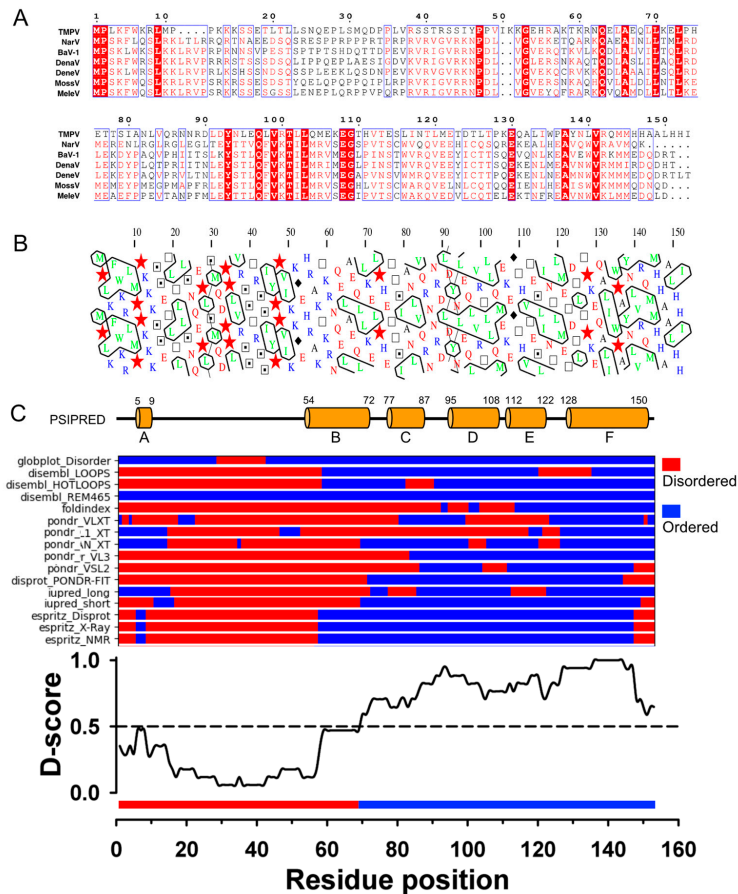
#### 3.3. Crystal Structure of the Folded C-Terminal Domain Revealed Structural Similarity with the Sendai Virus C Protein

The protein  $C_{\Delta 53}$  crystallized readily with six molecules in the asymmetric unit in the P3<sub>2</sub>21 space group and diffracted at 2.4 Å resolution. The crystal structure was solved with the SAD method by soaking crystals in a platinum derivative solution, and the structure was refined to a resolution of 2.4 Å (Table 1 and Figure 3C).

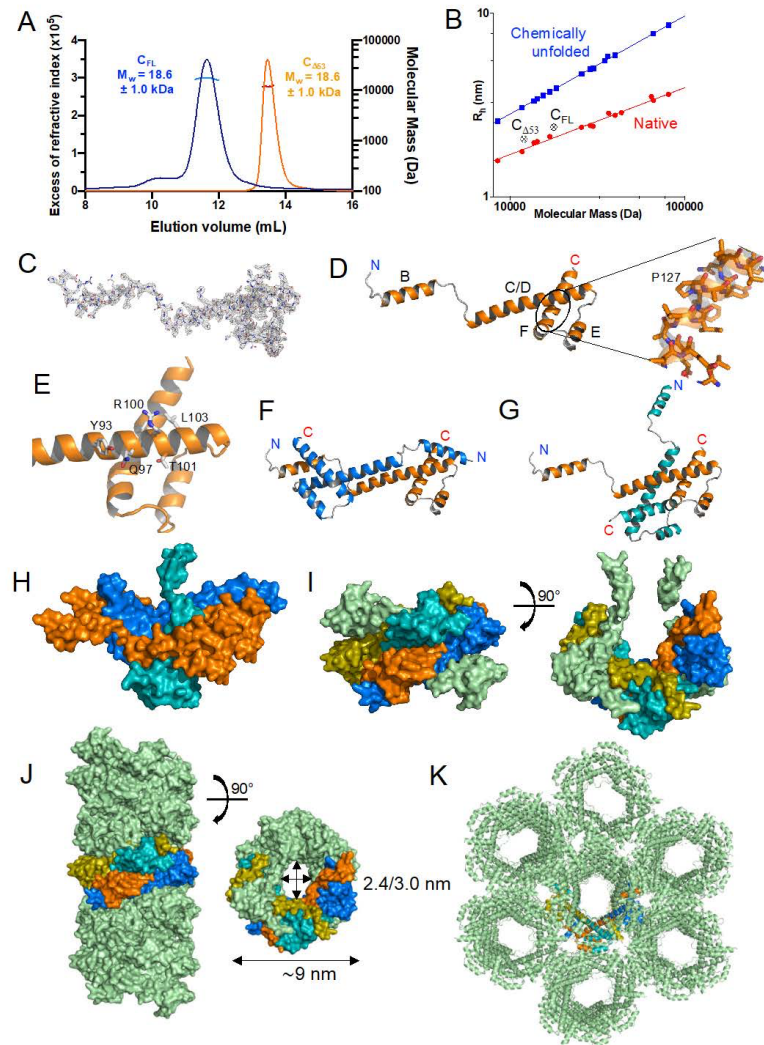
Each polypeptide chain forms four  $\alpha$ -helices (Figure 3D): helices B, C/D (a long helix so named because it encompassed predicted helices C and D), E and F. Helix B and the N-terminal half of helix C/D are stabilized by intermolecular interactions with other C proteins (see below). Helix F is kinked at Pro127 (Figure 3D). The C-terminal part of helix C/D (aa 86–99), is packed onto a hairpin formed by helix E (aa 105–114) and helix F (aa 119–141) (Figure 3D). Helices D, E, and F form a small domain containing a hydrophobic core. Of note, the residues in the C-terminal part of the C/D helix (Y93, Q97, V99, (R/K)100, T101, L102, E108, and G109) that are strictly conserved among Narmoviruses (Figure 2A) are exposed to the solvent and could thus represent a recognition motif (Figure 3E).

In the crystal, the proteins are packed against each other in two different modes. In the first mode (Figure 3F), two molecules of  $C_{\Delta 53}$  assembled head-to-tail, forming 14 interchain H-bonds, 2 salt bridges, and burying  $\sim 3300$  Å<sup>2</sup> of accessible surface area (PISA Complex formation Significance Score of 1.0). Helix B of each monomer binds in the fork formed by helices C/D and F of the other monomer. The interaction involves conserved residues Q63,

L65, A66, L/169, L70, and L73 in helix B, and M144 and M145 in helix F. The N-terminal part of helix C/D of two protomers packs against each other in an antiparallel orientation, mainly forming polar and symmetrical interactions. In the second mode (Figure 3G), two molecules assembled by the orthogonal interlocking of their C-terminal domain, forming 3 H-bonds and 1 salt bridge and burying ~1200 Å<sup>2</sup> of accessible surface area (PISA Complex formation Significance Score of 1.0).



**Figure 2.** TupV C protein sequence analysis. (A) Multiple sequence alignment of Narmovirus C protein. Members of the Narmovirus genus and their UniProt accession number: TupV—Tupaia paramyxovirus—Q9WS38; NarV—Nariva narmovirus—B8XH61; BaV-1—Myodes narmovirus (Bank Vole virus-1)—A0A2H4PJ60; DenaV—Denalis virus—UQM99579; DeneV—Denestis virus—UQM99571.1; MossV—Q6WGM3; MeleV—Meleucus virus—UQM99623. (B) HCA (Hydrophobic Cluster Analysis) plot. (C) Secondary structure and disorder predictions. The upper part shows the secondary structure prediction from PSIPRED (the position of predicted helices is indicated by the orange cylinders with the residue numbers shown above). The middle part shows the predictions from individual predictors as obtained in the output of the D-score script. The lower part shows a consensus disordered prediction (D-score) calculated as described in [25]. The threshold to distinguish between ordered and disordered regions was set at 0.5.



**Figure 3.** Crystal structure of TupV C protein C-terminal domain (C<sub>Δ53</sub>). (A) SEC-MALLS of TupV C<sub>Δ53</sub> and C<sub>FL</sub>. The lines show the chromatograms monitored by differential refractive index measurements. The crosses indicate the molecular mass across each elution peak calculated from static light scattering and refractive index, and the numbers indicate the weight-averaged molecular mass (kDa) with standard deviations (the molecular mass calculated from the aa sequence are 17,807 Da for C<sub>FL</sub> and 11,897 Da for C<sub>Δ53</sub>, respectively). (B) Plot of the hydrodynamic radius measured by SEC as a function of the molecular mass measured by MALLS and RI. The blue and red circles show data taken from [42] for globular proteins in native or chemically unfolded forms, respectively. (C) Electron density map of one protomer in the asymmetric unit contoured at 1.8 Å and stick representation of the TupV C<sub>Δ53</sub> protein in the crystal structure (PDB ID: 8BJW). (D) Cartoon representation of TupV C<sub>Δ53</sub> in the crystal. The N and C-terminal residues of C<sub>Δ53</sub> are indicated and helices are named B to F. (E) Close-up of the C-terminal domain showing residues conserved among Narmoviruses (see also Figure 2A). (F) Cartoon representation of the first mode of interaction between protomers within the

asymmetric crystal unit. (G) Cartoon representation of the second mode of interaction between protomers within the asymmetric crystal unit. The orange protomer is in the same orientation as in panel (F). (H) Surface representation showing the assembly of three protomers within the asymmetric crystal unit by the two modes of interaction. The color is the same as in panel (F,G). (I) Surface representation in two orthogonal orientations of one asymmetric unit containing six protomers. The same three protomers of panel (H) are shown with the same color code, the protomer associated with the one in teal by the first mode of interaction is shown in olive, while the other two are shown in pale green. (J) Surface representation of the tubular assembly seen in the crystal packing. The same four protomers described in panels (F–I) are shown with the same color code, and all the others are shown in pale green. (K) Cartoon representation of the crystal packing showing the side-by-side assembly of the tubes described in panel (J) with the same four protomers using the same color code.

**Table 1.** Crystallographic Data Collection and Refinement Statistics.

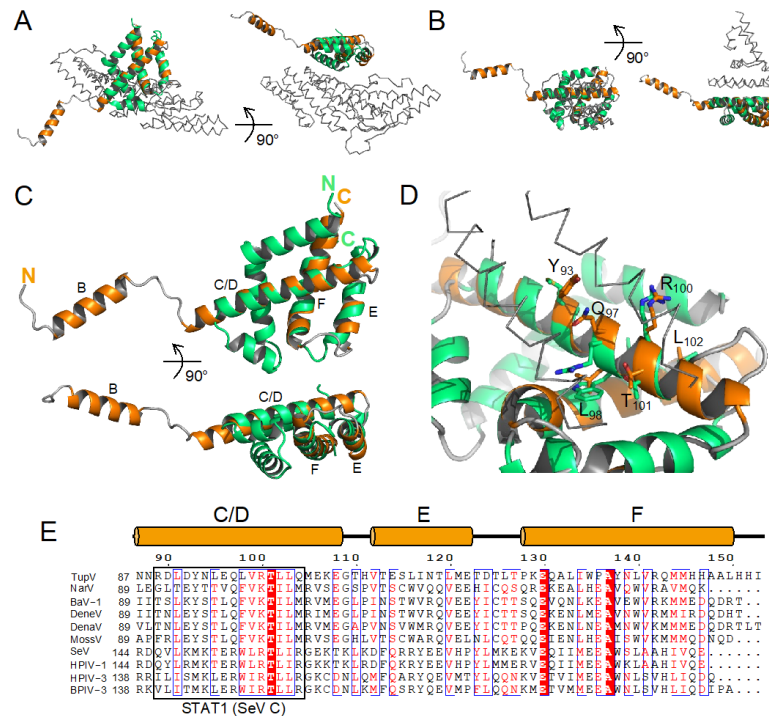
	C <sub>A53</sub> Native	C <sub>A53</sub> Pt Derivative
Data collection		
Beamline	SOLEIL-PX1	SOLEIL-PX1
Wavelength (Å)	1.07169	1.07169
Space group	P3 <sub>1</sub> 21	P2 <sub>1</sub>
a,b,c (Å)	135.7, 135.7, 63.1	135.6, 135.6, 64
a,b,g (°)	90, 90, 120	90, 90, 120
Resolution range (Å) <sup>1</sup>	68.4–2.4	46.5–2.7
R <sub>merge</sub> <sup>1</sup>	0.085–0.661	0.078–2.46
I/sI <sup>1</sup>	18.9–4.4	16.3–0.9
Completeness (%) <sup>1</sup>	100–100	99.5–93.3
Total reflections <sup>1</sup>	481,705–43,842	25,1742–17,210
Unique reflections <sup>1</sup>	27,136–2824	18,927–1316
Multiplicity <sup>1</sup>	17.8–15.5	
Anomalous completeness (%) <sup>1</sup>	n/a	99.3–91.7
Anomalous multiplicity <sup>1</sup>	n/a	6.9–6.7
DelAnom correlation <sup>1</sup>	n/a	0.432–0.042
Refinement		n/a
R <sub>work</sub> /R <sub>free</sub>	0.22 / 0.25	
No. water	49	
R.m.s. deviations		
Bond length (Å)	0.0099	
Bond angle (°)	1.5851	
Ramachadran statistic <sup>2</sup>		n/a
Favoured (%)	98.35	
Allowed (%)	1.47	
Outlier (%)	0.18	
PDB code	8BJW	n/a

<sup>1</sup> Values in for all data set and the high-resolution shell; <sup>2</sup> Values from MolProbity webservice.

The assembly of three protomers by both modes is shown in Figure 3H. The asymmetric unit comprised six molecules and had a horseshoe shape (Figure 3I). The chaining of asymmetric units leads to a helical assembly and to the formation of tubular structures with external diameters of ~9.0 nm and hexagonal internal cavities of 2.4–3.0 nm width (Figure 3J). In the crystal, the tubular helices are packed side-by-side (Figure 3K).

A search for structural similarities with DALI [59] identified the C-terminal folded domain of the Sendai virus C protein either in complex with the N-terminal domain of human STAT1 (PDB ID: 3WWT-Z-score 4.9, 3.3 Å rmsd over 62 CA atoms), a key intermediate in the interferon signaling pathway [60], or in complex with the BRO-1 like the domain of the human protein Alix (PDB ID: 6KP3-Z-score 3.5, 2.2 Å rmsd over 54 Ca atoms), a component of the ESCRT machinery [61].

In both complexes, Sendai virus C protein (in green in Figure 4A,B) has the shape of a flat disk formed by six helices arranged in two superposed layers. In both cases, the three C-terminal helices of SeV C align with the C-terminal part of helix C/D and with helices E and F of TupV C, while additional N-terminal helices are packed on the common helical core (Figure 4C).



**Figure 4.** Superposition with the C-terminal domain of Sendai virus C protein. (A) Superposition in two orthogonal orientations of the C-terminal domain of TupV C protein with the C-terminal domain of SeV C protein complexed with the BRO-1-like domain of the protein Alix (PDB ID: 6KP3). TupV C is shown in orange, SeV C in green and the BRO-1-like domain of Alix in a grey ribbon. (B) Superposition in two orthogonal orientations of the C-terminal domain of TupV C protein with the C-terminal domain of SeV C protein in complex with the N-terminal domain of STAT1 (PDB ID: 3WWT). TupV C is shown in orange, SeV C in green, and STAT1 in a grey ribbon. (C) Superposition in two orthogonal orientations of TupV C (in orange) and SeV C (in 3WWT) taken from its complex with STAT1 (in green). (D) Close up of the superposition of TupV C and SeV C (in 3WWT) showing the orientation of conserved residues in both proteins. The labeled residues are those of TupV C protein. STAT1 is shown in grey ribbon. (E) Sequence alignment of Narmovirus and Respirivirus C proteins. Members of the Narmovirus genus and their UniProt accession number are the same as in Figure 2A, members of the Respirivirus genus are: SeV-Sendai virus-Q38KG9; HPIV1-human parainfluenza virus 1-Q8QT30; HPIV3-human parainfluenza virus 3-Q81077; BPIV1-bovine parainfluenza virus 3-P06164. The orange cylinders above the MSA indicate the location of helices in the crystal structure of TupV C. The black box indicates residues of SeV C that interact with STAT1.

The interaction of SeV C with the N-terminal domain of STAT1 is mediated by residues in the first and fourth helices of C (respectively aa 104–109 and 146–161) [60]. Some of the residues involved in the interaction are conserved in TupV C (aa 98–103, Figure 4D),

indicating that in principle it also has the potential to interact with STAT1 (see Section 4.3). Figure 4E presents a sequence alignment of TupV C and SeV C based on their structural superposition, made using Promals3D [62]. The second region of SeV interacting with STAT1 (aa 146–161) and its equivalent in Narmovirus C proteins are boxed (Figure 4E).

In contrast, the interaction of SeV C protein with the BRO-1 like domain of Alix, mediated by an  $L_{122}XXW_{125}$  motif located in the second  $\alpha$ -helix of SeV C, has no equivalent in the TupV C protein structure. Therefore, TupV C is not expected to bind the BRO-1-like domain of Alix, at least not in the same manner as SeV C.

#### 3.4. The C Proteins of NiV, MeV, and SeV Contain a Core of 3 Helices That Superpose to Helices D-E-F of TupV C

As the C-terminal end of SeV C adopts the same fold as the core made by helices D-E-F in TupV C (Figure 5A), we wanted to verify whether the C protein of other Orthoparamyxovirinae contains the same three core helices. For this, we predicted the structure of NiV and MeV C proteins using AlphaFold2 [52,53]. The C models of NiV and MeV C contained three helices superposable with helices D-E-F of TupV C (Figure 5B,C). Thus, the helical core domain is common to the C proteins of the three groups previously identified: the measles virus group, the Nipah virus group, and the Sendai virus group [22] (Figure 5D).

We note that both NiV (166 aa) and MeV C (186 aa) proteins are longer than that of TupV (153 aa) and have a C-terminal extension, which forms one or two additional helices at the C-terminal end, respectively (Figure 5E).

#### 3.5. TupV $C_{\Delta 53}$ Does Not Multimerize in Solution, Unlike in the Crystal

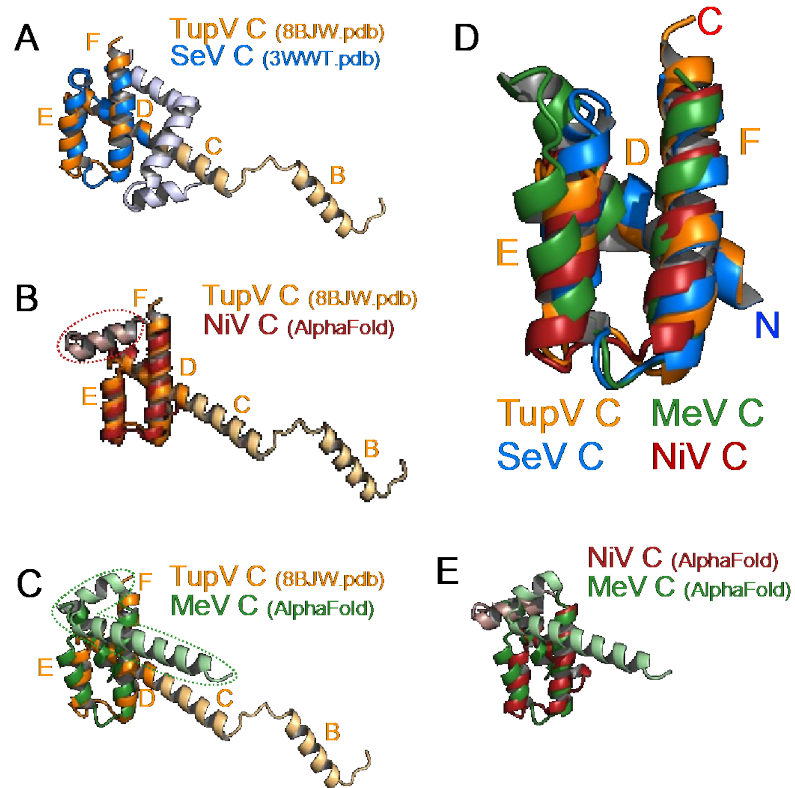
To determine whether TupV  $C_{\Delta 53}$  self-assembled in solution, we first analyzed the dependence of the elution on protein concentration by SEC-MALLS. Protein samples at a concentration ranging from 3.3 mg/mL to 23.5 mg/mL were injected into the Superdex 75 column (Figure 6). At all concentrations, the protein eluted as a single peak; its molecular mass, as determined by light scattering and refractive index, remained constant throughout the chromatographic peak ( $M_w/M_n \leq 1.005$ ); and the weight average molecular mass remained close to the theoretical value calculated for the monomer within the experimental errors. In conclusion, TupV  $C_{\Delta 53}$  is essentially monomeric in solution, in contrast with its multimeric state in the crystal. Only a small variation in the elution volume might suggest the presence of larger species, which would represent only a small fraction (less than 5%).

#### 3.6. In Solution, TupV C Contains Both Disordered and Globular Regions

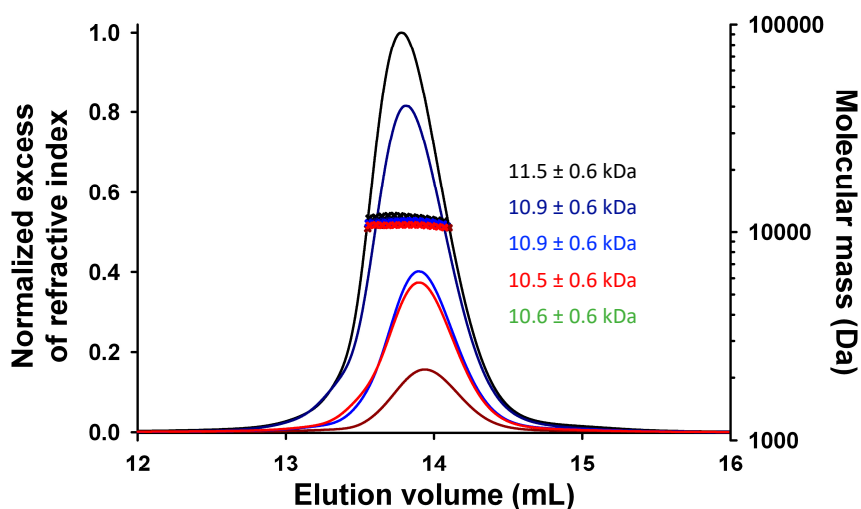
In the crystal structure, helix B and the N-terminal half of helix C/D are stabilized by interactions between different C molecules (Figure 3). In solution, since C is monomeric (see above), we assumed that these helices might be mostly disordered. To experimentally characterize the state of these helices in  $C_{\Delta 53}$  and obtain information about the N-terminal disordered region of the full-length protein in solution, we performed SEC-SAXS experiments with both  $C_{\Delta 53}$  and  $C_{FL}$  (Table 2). We injected samples of each protein on a Superdex 75 Increase column and collected SAXS data at regular intervals along the elution peak over scattering vector ( $q$ ) values ranging from 0.08 to  $5.0 \text{ nm}^{-1}$  (Figure 7A,B).

For both proteins, the Guinier plots at low  $q$  values ( $q \cdot R_g < 1.3$ ) were linear, and the  $R_g$  values calculated by using the Guinier approximation were constant throughout the chromatographic peaks showing no dependence on protein concentration and the absence of aggregation or intermolecular interactions (Figure 7A,B). We obtained representative SAXS curves by averaging 10 frames in the center of the chromatographic peaks (Figure 7C) and determined average  $R_g$  values of  $1.7 \pm 0.1 \text{ nm}$  for  $C_{\Delta 53}$  and  $2.5 \pm 0.1 \text{ nm}$  for  $C_{FL}$  by using the Guinier approximation (Figure 7D) (Table 2). In the dimensionless Kratky plot for  $C_{\Delta 53}$ , the curve reached a maximum near 1.5 for  $q \cdot R_g$  values around 2.6 and rose at  $q \cdot R_g$  values above 4, suggesting the presence of both a globular region and disordered regions (Figure 7E). In the plot for  $C_{FL}$ , the curve reached a less pronounced maximum at 1.8 at  $q \cdot R_g$  values above 2 and decreased slightly at higher  $q \cdot R_g$  values before rising again at

values above 6, which suggested the presence of a longer disordered region (Figure 7E). The pair distance distribution function (PDDF) revealed similar structural conclusions with  $D_{\max} = 6.5$  nm for  $C_{\Delta 53}$  and  $D_{\max} = 10.5$  nm for  $C_{FL}$  (Table 2).



**Figure 5.** Structural alignments. (A) Superposition of TupV C (8BJW.pdb) and SeV C (taken from 3WWT.pdb) crystal structures. The non-superposable parts of TupV and SeV C are, respectively in light blue and light orange. (B) Superposition of the crystal structure of TupV C with an AlphaFold model of NiV C (aa 100–166). The averaged DDT score for this model of NiV C is 92.6, which indicates a highly reliable prediction. The additional C-terminal helix of NiV C is circled with a dotted line. (C) Superposition of the crystal structure of TupV C with an AlphaFold model of MeV C (aa 101–186). The averaged pLDDT score for this model is 86.0, also indicating a reliable prediction. The additional C-terminal helices of MeV C are shown and circled with a dotted line. (D) Superposition of (i) the helical domain of TupV C taken from the crystal structure (8BJW.pdb), (ii) the corresponding regions of SeV C taken from the crystal structure (3WWT.pdb), (iii) an AlphaFold model of MeV C and (iv) an AlphaFold model of NiV C. The N and C-termini and the name of the helices in TupV C (D–F) are labeled. (E) Superposition of the AlphaFold models of the NiV and MeV C protein domains.



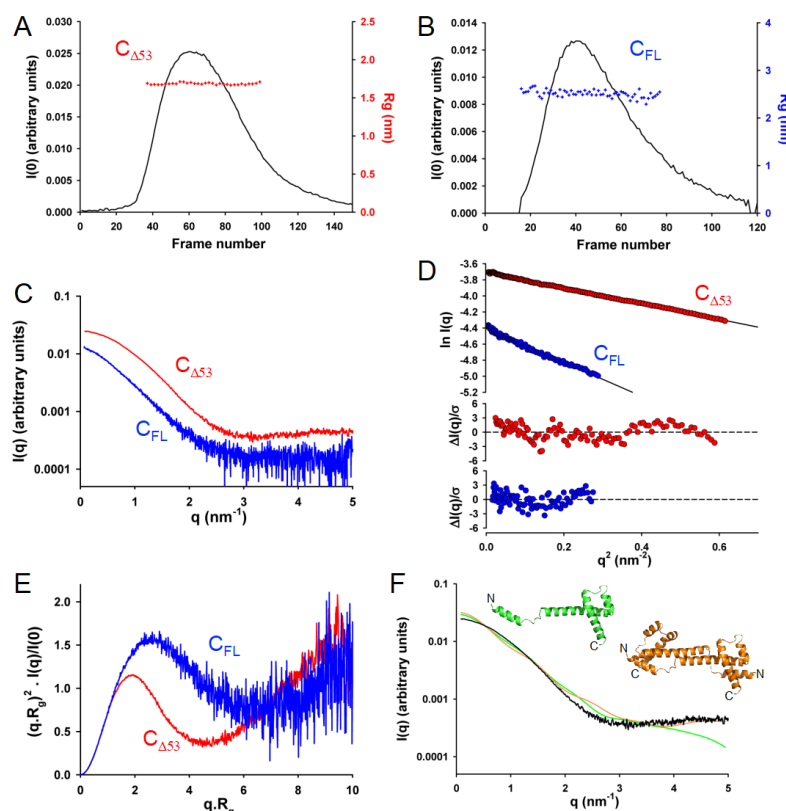
**Figure 6.** SEC MALLS experiments at different TupV C<sub>D53</sub> protein concentrations. 50 mL of protein solution were injected onto a Superdex 75 column at different initial concentrations: 3.3 mg.mL<sup>-1</sup> in dark red, 8.5 mg.mL<sup>-1</sup> in red, 8.8 mg.mL<sup>-1</sup> in blue, 20.0 mg.mL<sup>-1</sup> in dark blue and 23.5 mg.mL<sup>-1</sup> in black. The lines show the chromatograms monitored by refractive index and the crosses show the molecular masses calculated at each time from the light scattering intensity and refractive index. Numbers show the calculated weight average molecular masses ( $M_w$ ).

**Table 2.** Small-angle X-ray Scattering Data Collection and Analysis.

	C <sub>FL</sub>	C <sub>Δ53</sub>
Data collection parameters		
Instrument	SOLEIL-SWING	
Energy (keV)	12.000	
Detector	EigerX-4M	
Detector distance (m)	2	
Exposure (s per image)	1	
q range (Å <sup>-1</sup> )	0.008–0.50	
Column	S75inc 5/150 GL	
Flow rate (mL.min <sup>-1</sup> )	0.3	
Sample concentrations (mg.mL <sup>-1</sup> )		
Injection volume (μL)	50	50
Temperature (K)	293	293
Guinier Analysis		
I(0) (A.U.)	1.2·10 <sup>-2</sup> ± 0.03·10 <sup>-2</sup>	2.4·10 <sup>-2</sup> ± 0.01·10 <sup>-2</sup>
Rg (nm)	2.5 ± 0.1	1.7 ± 0.1
qr <sub>gmin</sub>	0.2951	0.2702
qr <sub>gmax</sub>	1.3070	1.3027
P(R) analysis <sup>1</sup>		
D <sub>max</sub> (nm)	10.5	6.4
Molecular weight <sup>2</sup>		
Theoretical M <sub>w</sub> (kDa) <sup>3</sup>	17,807	11,897
Measured M <sub>w</sub> (kDa)	18.6 ± 0.8	12.3 ± 0.4

<sup>1</sup> Calculated with GNOM [63] and BIFT [64]. <sup>2</sup> Calculated from the volume of correlation. <sup>3</sup> Calculated from the aa sequence.





**Figure 7.** Small-angle X-ray scattering experiments. (A) SEC-SAXS analysis of TupV  $C_{\Delta 53}$  protein. The black line shows the scattering at zero angles ( $I(0)$ ), which is proportional to both molecular mass and concentration, as a function of the frames recorded at regular time intervals. The red crosses show the radius of gyration calculated from the Guinier plots at different time intervals. (B) SEC-SAXS analysis of TupV  $C_{FL}$  protein. The black line and the blue crosses are as in panel (A). (C) Averaged scattering curves. The lines show the scattering profiles obtained for  $C_{\Delta 53}$  (in red) and  $C_{FL}$  (in blue) by averaging the individual profiles recorded throughout the SEC elution peak shown in panels (A,B). (D) Guinier plots. The upper panel shows the Guinier plots for  $C_{\Delta 53}$  (in red) and  $C_{FL}$  (in blue). The lower panels show the plots of the normalized residuals. (E) Normalized Kratky plots. (F) Comparison of TupV  $C_{\Delta 53}$  protein in solution and in crystal. The experimental SAXS curve obtained for TupV  $C_{\Delta 53}$  protein (black line) is compared to the theoretical curve calculated with CRY SOL for one monomer (green line) and one dimer (orange line) extracted from the crystal structure (shown in cartoon representations in green and orange, respectively). The curves were scaled in order to simply compare their shape and not to take into account differences in molecular mass.

Comparing the experimental SAXS curve for  $C_{\Delta 53}$  with the theoretical SAXS curves calculated either for a monomer or for dimers extracted from the crystal structure revealed strong discrepancies ( $c^2 > 100$ ), clearly indicating that the C protein is neither a monomer nor a dimer with helix B and full helix C/D present as in the crystal (Figure 7F). This confirmed our initial intuition that the N-terminal part of the truncated protein is disordered and highly flexible in solution.

### 3.7. In Solution, *TapV* C Is Well Described by Ensembles of Conformers with a Disordered N-Terminal and Fluctuating Helices B and C

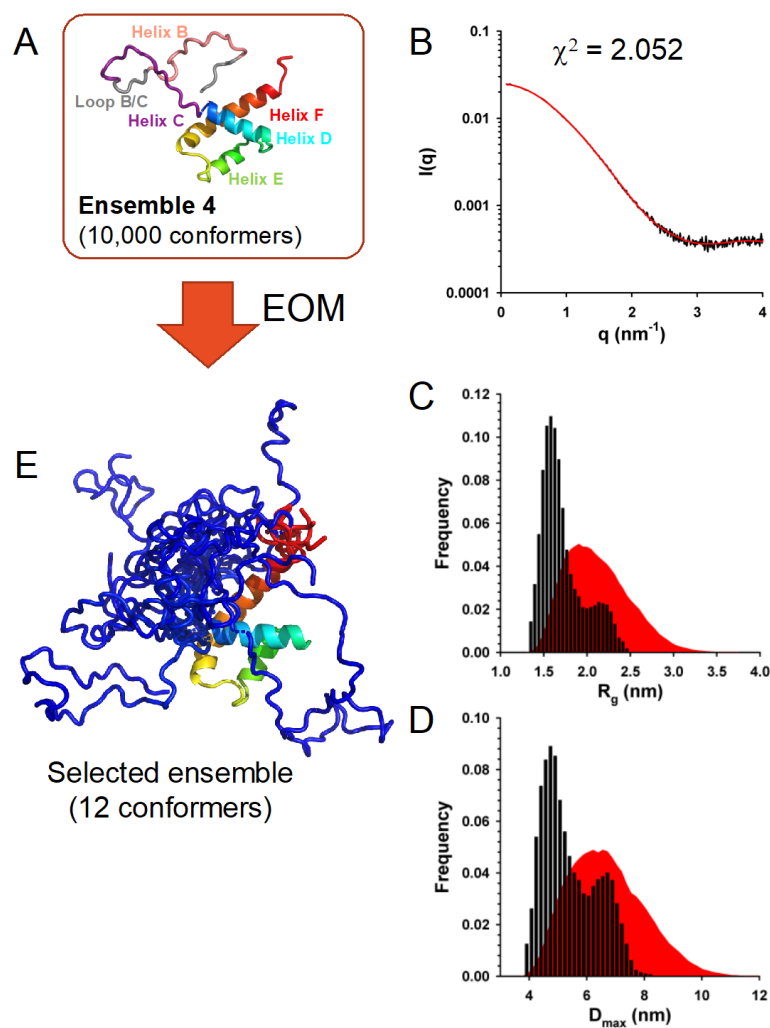
To further explore the conformational diversity of  $C_{\Delta 53}$ , we used the Ensemble Optimization Method (EOM) [45]. This approach consists of generating a large initial pool of independent conformers of the protein (using the software RANCH) and in selecting sub-ensembles of conformers that collectively reproduce the experimental SAXS curve (using the software GAJOE).

We generated four initial ensembles of 10,000 conformers (Figure 8A): (i) **Ensemble 1**, in which folded helix B and helix C/D were conserved but the B/C loop connecting both helices was allowed to adopt random conformations, (ii) **Ensemble 2**, in which helix B and the B/C loop were allowed to adopt random conformations, while helix C/D was conserved in its folded form, (iii) **Ensemble 3**, in which helix B was conserved in its folded form while the B/C loop and helix C were allowed to adopt random conformations and (iv) **Ensemble 4**, in which helix B, the B/C loop, and helix C were allowed to adopt random conformations.

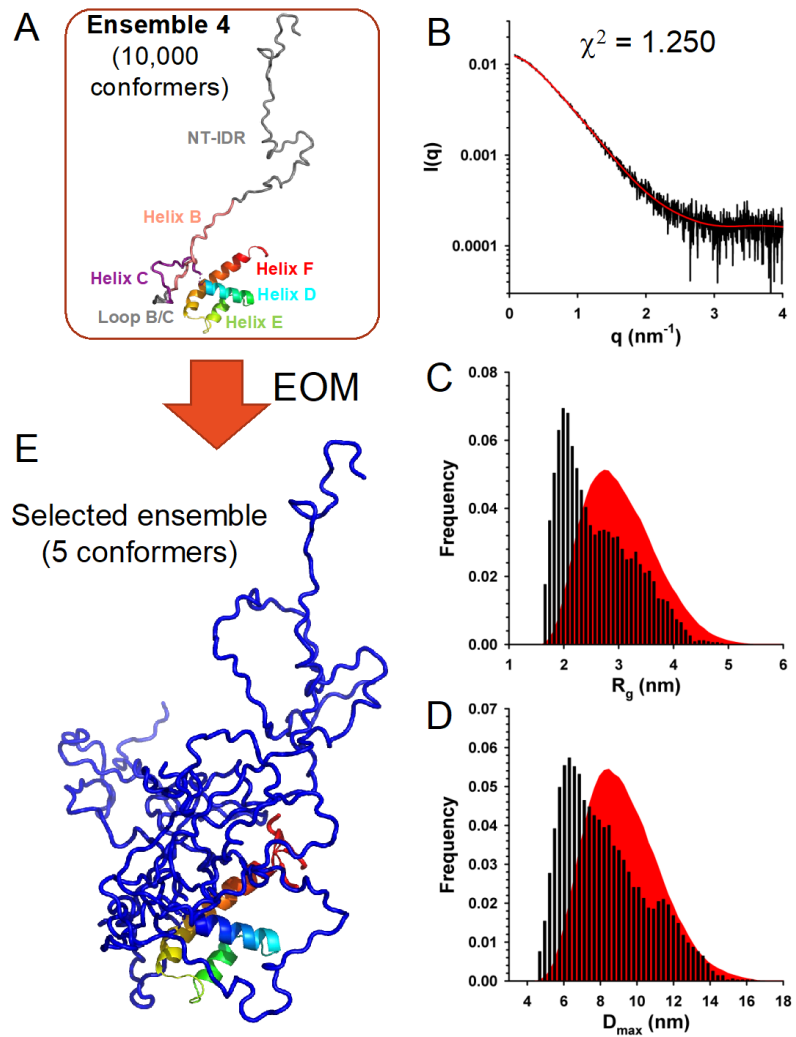
The best fits of the experimental curve obtained with sub-ensembles of conformers taken from ensembles 1 and 2 had  $c^2$  values of 2.64 and 2.85, respectively. Slightly better fits were obtained with sub-ensembles of conformers taken from ensembles 3 and 4 with  $c^2$  values of 2.05 and 2.07, respectively. The fit obtained with a sub-ensemble model of 12 conformers selected from the initial Ensemble 4, which had the lowest  $c^2$  value, is shown in Figure 8B. The initial Ensemble 4 displayed Gaussian distributions of  $R_g$  and  $D_{max}$  values (Figure 8C,D), while selected ensembles of 12 conformers that best reproduced the SAXS curve displayed bimodal  $R_g$  and  $D_{max}$  distributions with a large fraction of compact conformers suggesting that the flexible N-terminal part (helix B and helix C) are somehow packed on the helical core (Figure 8C,D).

A representative ensemble of 12 conformers is shown in Figure 8E. These results support a model in which the N-terminal part of  $C_{\Delta 53}$  is flexible, while the C-terminal domain is folded. However, they do not allow us to discriminate whether helix B was transiently formed or whether helix C transiently formed, extending helix D further than the region in contact with helices E and F.

We used a similar approach to explore the conformational diversity of  $C_{FL}$  with EOM. We generated four initial ensembles of 10,000 conformers with the same features as for  $C_{\Delta 53}$  but with an N-terminal intrinsically disordered region (NT-IDR) allowed to adopt random conformations (Figure 9A). Selections of sub-ensembles from each of these different initial ensembles yielded fits of the experimental curve of similar quality ( $\chi^2 = 1.25$ ) (Figure 9B), supporting a model in which the N-terminal part of the protein was intrinsically disordered and the C-terminal part structured like in the crystal. However, these data were not sufficient to assess the fine structural organization of helices B and C/D, because the contribution of the long N-terminal disordered tail blurred the fine structure of the SAXS curve, thus preventing a more detailed determination of the fold. The initial ensembles displayed Gaussian distributions of  $R_g$  and  $D_{max}$  values (Figure 9C,D), while selected ensembles displayed similar bimodal  $R_g$  and  $D_{max}$  distributions with a large fraction of compact conformers but with a less prominent population of larger conformers than for  $C_{\Delta 53}$ . A representative ensemble of five conformers is shown in Figure 9E.



**Figure 8.** Conformational ensemble modeling of TupV C $_{\Delta 53}$  protein from SAXS data. (A) Description of the initial ensemble, which is comprised of 10,000 conformers in which the regions corresponding to the helix B (in pink), the helix C (in purple), and the loop connecting them (in grey) were allowed to adopt random conformations. (B) Experimental SAXS curve for TupV C $_{\Delta 53}$  (in black) and fitted curve (in red) calculated for one representative ensemble of conformers selected with the program GAJOE (shown in panel (E)). (C)  $R_g$  distribution for an ensemble of 12 conformers of TupV C $_{\Delta 53}$ . The red area shows the distribution for the initial ensemble of conformers. The black bars show the  $R_g$  distribution for selected ensembles of 12 conformers. (D)  $D_{max}$  distribution for an ensemble of 12 conformers of TupV C $_{\Delta 53}$ . The red area shows the distribution for the initial ensemble of conformers. The black bars show the  $D_{max}$  distribution for selected ensembles of 12 conformers. (E) Representative ensemble of 12 conformers that reproduced the experimental SAXS curve. The N-terminal disordered part is shown in blue and the structured domain is shown in color from blue to red going from N to C-terminal.



**Figure 9.** Conformational ensemble modeling of TupV  $C_{FL}$  protein from SAXS data. (A) Description of the initial ensemble, which is comprised of 10,000 conformers in which the regions corresponding to the N-terminal intrinsically disordered region (NT-IDR in grey), the helix B (in pink), the helix C (in purple), and the loop connecting them (in grey) were allowed to adopt random conformations. (B) Experimental SAXS curve for TupV  $C_{FL}$  (in black) and fitted curve (in red) calculated for one representative ensemble of conformers selected with the program GAJOE (shown in panel (E)). (C)  $R_g$  distribution for an ensemble of five conformers of TupV  $C_{FL}$ . The red area shows the distribution for the initial ensemble of conformers. The black bars show the  $R_g$  distribution for selected ensembles of five conformers. (D)  $D_{max}$  distribution for an ensemble of five conformers of TupV  $C_{FL}$ . The red area shows the distribution for the initial ensemble of conformers. The black bars show the  $D_{max}$  distribution for selected ensembles of five conformers. (E) Representative ensemble of five conformers that reproduced the experimental SAXS curve. The N-terminal disordered part is shown in blue and the structured domain is shown in color from blue to red going from N to C-terminal.

Altogether, both the truncated ( $C_{\Delta 53}$ ) and full-length ( $C_{FL}$ ) proteins are well described by ensembles of conformers, in which the C-terminal structured domain is present, the N-terminal region is highly flexible but compact, and helices B and C are frayed or unfolded and/or take different orientations from those in the crystal.

### 3.8. *TupV C Protein Cross Inhibits Nipah Virus Minigenome Expression*

Although the interactions between the C protein and components of the innate responses and of the ESCRT machinery vary from one virus to another, all C proteins that have been tested downregulate the production of viral RNA, in some cases demonstrating cross-species activities [18]. To test whether the TupV C protein can regulate RNA synthesis in related Orthoparamyxovirinae, we assessed whether it inhibits the expression of the minigenome of the Nipah virus.

For this purpose, we constructed a minigenome for the Nipah virus, in which the gene of the Renilla-Luciferase was under the control of 3' and 5' untranslated regions of the Nipah virus (Figure 10A), and we developed a mini-replicon assay (Figure 10B). Cultured 293T cells were transfected with the plasmid encoding the minigenome together with helper plasmids encoding NiV N, P, and L-RdRP and with a plasmid encoding either NiV C protein or TupV C protein, and we monitored the expression of the minigenome by quantifying the Renilla luciferase luminescent signal (using the firefly luciferase luminescent signal as an internal reference) (Figure 10B).

We found that both NiV and TupV C proteins inhibited the expression of Renilla luciferase in a concentration-dependent manner (Figure 10C). As a control, we assessed cell viability by measuring the firefly luminescent signal and found a slight concentration-dependent reduction in the expression of these proteins, which was accounted for in Figure 6B, indicating that, at 24 h after transfection, the reduction in Renilla luciferase expression was neither resulting from a cytopathic effect nor an overall negative effect on the cellular translation of the C protein. These results suggest that the inhibition was specific to the viral RNA synthesis machine.

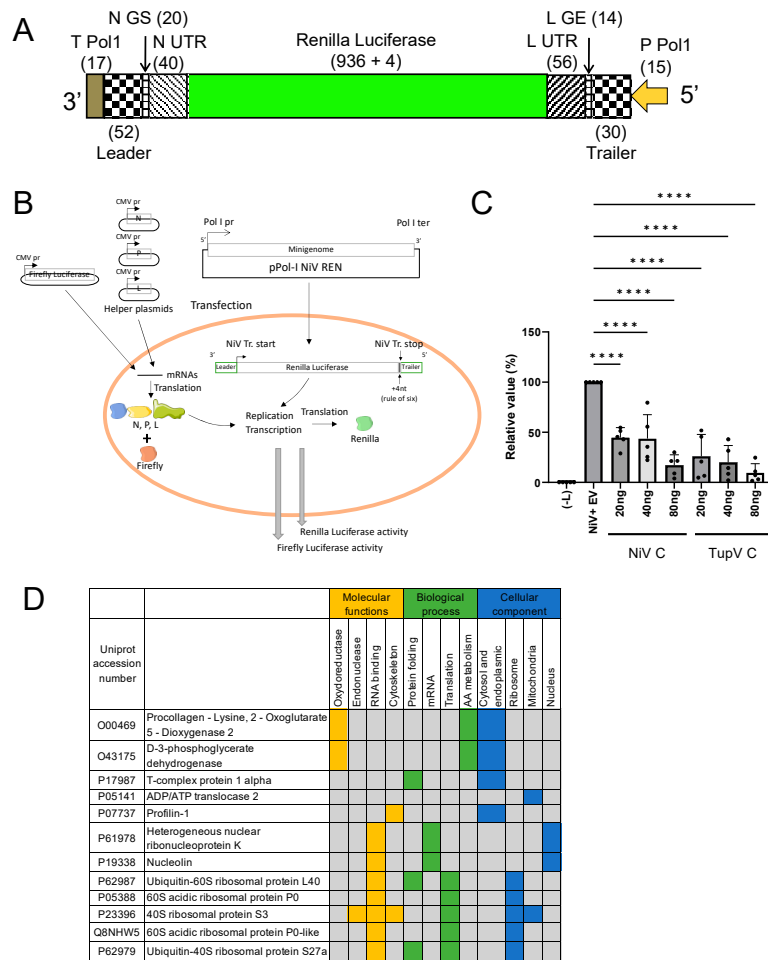
### 3.9. *TupV C Does Not Interact with Human STAT1 Nor with Known Interactors of Measles Virus and Nipah Virus C Proteins*

Because Tupaia belangeri protein sequences are very close to the corresponding human protein sequences (98% sequence identity), we tried identifying human proteins interacting with TupV C by tandem affinity purification (TAP) combined with mass spectrometry (MS) [65]. By applying the filtering previously described [65], we identified 12 protein-protein interactions consistently picked up by this procedure (Figure 10D). Interestingly, none of these proteins were common to those previously identified by the same approach for the Nipah virus C protein [65] or for measles virus C [66].

In particular, no interaction was found between TupV C and human STAT1, despite the conservation between the sequence of TupV C and one of the two regions of SeV C interacting with STAT1 (Figure 4E). In principle, TupV C could only interact with T. belangeri STAT1 instead, but we think this unlikely given its extremely high sequence identity (99%) with human STAT1.

No interaction was found with Alix either, which is coherent with the fact that TupV C has no equivalent for the region of interaction between SeV C and Alix. No interaction was found either with any other component of the ESCRT machinery beyond Alix.

Although we cannot rule out false positives or false negatives in TAP experiments, the divergences between the different virus-host interactomes could reflect that each virus co-evolved with its host, and thus that different C proteins interact with different partners in the infected cells.



**Figure 10.** In vivo studies. (A) Schematic representation of the RNA minigenome. The scheme drawn as a single strand of negative-sense RNA shows the positions of the PolI promoter (P Pol1) and terminator (T Pol1), of the NiV leader, trailer, gene start, and untranslated regions from the N gene (N GS and N UTR), of the gene end and untranslated regions from the L gene (L GE and L UTR) and of the Renilla luciferase, including four bases added to correct the length of the minigenome to be evenly divisible by six. Lengths in nt of the different elements are shown in brackets. (B) Schematic representation of the mini-replicon assay. Initially, the minigenome synthesis is controlled by the polymerase I promoter in the pPol I NiV-REN plasmid. The components of the NiV ribonucleoprotein complex (N, P, L) are expressed from helper plasmids under the control of polymerase II (CMV pr). Successful encapsidation of the minigenome RNA will lead to replication of the minigenome and transcription of the reporter gene under the control of NiV polymerase L. Renilla luciferase expression is then used to quantify NiV polymerase activity. A plasmid expressing firefly luciferase under the CMV promoter (CMV pr) is used as a normalization control. (C) Effect of NiV and TupV C protein on NiV mini-replicon assay. HEK293T cells were transfected with plasmids encoding Nipah viral N, P, and L proteins and one NiV minigenome encoding Renilla luciferase. A plasmid encoding Firefly luciferase for normalization. In the negative control (-L), the plasmid encoding NiV L was omitted.

Increasing plasmid amounts of NiV C or TupV C were co-transfected as indicated, while the amount of transfected plasmid DNA was kept constant by the addition of the pCDNA3 vector. At 24 h after transfection, luciferase activities were measured and results obtained in the absence of C protein (NiV+EV) were set to 100% whereas results of the negative control (-L) were set to 0%. For relative Firefly expression analysis, NiV+EV values were set to 100%. Individual values are represented by dots. Mean and SD from five independent experiments are indicated. \*\*\*\*  $p < 0.0001$ ; ns not significant (One-way ANOVA, Dunnett Test). (D) Identification of interactions between TupV C protein and human proteins by TAP-MS. A total of 12 human proteins were captured with the TupV C bait, representing different molecular functions involved in several biological processes and localized in different cellular components.

#### 4. Discussion

##### 4.1. Structural Comparisons and Predictions

Overall, the structural characterizations of the TupV C protein in the crystal and in solution confirmed its predicted architecture, with a disordered N-terminal part (NT-IDR), a C-terminal helical folded domain, and a central region whose structure depends on the context. We can state with confidence that the formation of helices B and C in the crystal is due to interactions between different C proteins, while in solution, our data suggest that these helices fluctuate, being either partially disordered or at least oriented differently than in the crystal. The comparison of the experimental SAXS curve with the theoretical curves calculated from monomeric and dimeric forms of the protein taken from the crystal clearly confirmed that the structure of the N-terminal part of C<sub>Δ53</sub> in solution is different from that in the crystal. The best fit was obtained with ensembles of conformers, in which helices B and C were disordered (Figure 8).

Unfortunately, the difference in the quality of fit of the SAXS data with ensembles, in which these helices were present but allowed to move relative to the C-terminal domain, was not sufficient for clearly establishing the conformation of this region of the TupV C protein. It is possible that helices B and C are present but frayed at the ends and packed on the folded domain. In both TupV C<sub>Δ53</sub> and C<sub>FL</sub>, the N-terminal disordered region is rather compact as seen in the  $R_g$  and  $D_{max}$  distribution plots of the selected ensembles (Figures 8C,D and 9C,D) and in the cartoon representation of representative ensembles (Figures 8E and 9E). This chain compaction could regulate some activity of the protein by hiding interacting surfaces.

The crystal packing revealed that TupV C protein can assemble into a helical structure forming tubes, but in solution, no evidence was found for the assembly of C into dimers or higher multimers in solution. Thus, this organization could entirely be due to crystal contacts. However, the analysis by the algorithm of the webserver PISA qualifies both types of interactions with other monomers as structurally significant. Because the C proteins of various Orthoparamyxovirinae recruit components of the ESCRT machinery and participate in the budding of new viral particles [21,61], we hypothesize that C multimers form in the presence of these viral or cellular partners, in particular when the latter themselves form multimers.

##### 4.2. The C Protein Originated in the Ancestor of Orthoparamyxovirinae

The Paramyxoviridae P gene is a complex transcription unit, encoding P, V, and C in overlapping reading frames (Figure 1). The appearance and evolution of the V coding region have been recently retraced [8]. In contrast, that of the C coding region has been difficult to trace, since the sequences of both P and C are poorly conserved among Paramyxoviridae, and thus beyond the reach of current phylogeny reconstruction methods. Having solved the 3D structure of TupV C protein now provides a means to retrace the origin of the C proteins since 3D structure enables the distinction of sequence homology, even in the absence of meaningful sequence similarity.

The C coding region is present in almost all members of the subfamily Orthoparamyxovirinae. A previous study [22] showed that Orthoparamyxovirinae C proteins form

two main clusters according to sequence similarity: (1) the SeV group, i.e., the genera Respirivirus and Aquaparamyxovirus; and (2) the measles group and Nipah group, i.e., all other genera (Narmovirus [TupV], Morbillivirus [MeV], Henipavirus [NiV], Jeilongvirus, Salemvirus). The C protein of respiroviruses and aquaparamyxoviruses have no detectable sequence similarity with that of the other Orthoparamyxovirinae; in contrast, all other C proteins have detectable sequence similarity to each other, proving they are homologous. Likewise, the function of the C proteins differs between these two clusters: in Respirivirus, it binds STAT1 [60], whereas in other Orthoparamyxovirinae (Morbillivirus and Henipavirus), STAT1 is bound instead by the overlapping N-terminal intrinsically disordered region that is common to P, V, and W [37,67–69].

In principle, two scenarios could thus account for the present distribution and function of the C proteins:

- (a) The C coding region appeared in the common ancestor of Orthoparamyxovirinae and later diverged in sequence beyond recognition in respiroviruses and aquaparamyxoviruses. In this scenario, all C proteins are homologous.
- (b) The C coding region appeared independently in the ancestor of respiroviruses and aquaparamyxoviruses and in the rest of the Orthoparamyxovirinae. In this scenario, the C protein of respiroviruses and aquaparamyxoviruses would not be homologous to that of other Orthoparamyxovirinae.

Our discovery that the structures of respiroviruses and narmoviruses C proteins are significantly similar, and that they are also similar to AlphaFold2-predicted models of morbilliviruses and henipaviruses C proteins, indicate that the C proteins of all Orthoparamyxovirinae are homologous (scenario 1). This scenario is in complete agreement with previously noted common characteristics of the C coding regions and proteins [22]:

- (a) All Orthoparamyxovirinae C proteins are encoded in the same frame with respect to P (i.e., the +1 frame), and overlap at exactly the same position within the P gene (the very 5' end);
- (b) All Orthoparamyxovirinae C proteins have the same architecture (proven or predicted), composed of an N-terminal disordered region and C-terminal helical domain.

#### 4.3. A STAT1-binding Site Is Encoded by Exactly the Same RNA Region of the P Gene but in Different Reading Frames (P or C Coding Region) in Orthoparamyxovirinae

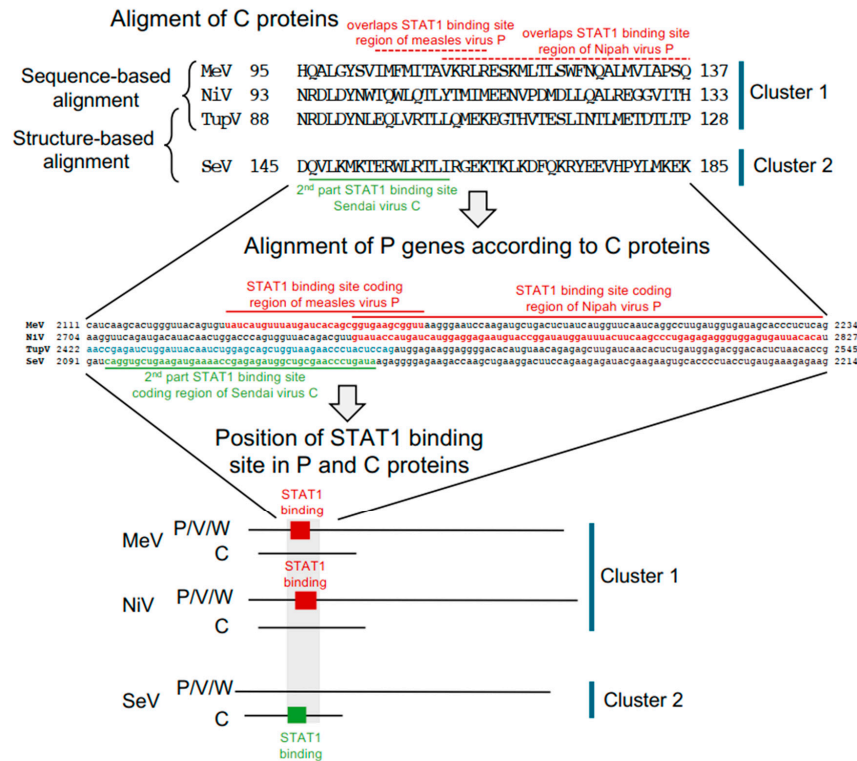
Many Orthoparamyxoviruses use the products of the P/V/W and C reading frames to interfere with interferon signaling, in particular by binding STAT1. In some viruses (NiV and HeV—Henipavirus and MeV—Morbillivirus), the STAT1-binding site is monopartite and located in the disordered N-terminal region of the P/V/W proteins [37,67–69], while in others (SeV—Respirovirus), it is bipartite, with both parts located in the ordered C-terminal domain of the C protein (thus encoded in the +1 frame with respect to P) [60].

Previous sequence alignments showed that the STAT1-binding site in the morbilliviruses (MeV) and henipaviruses (NiV and HeV) P/V/W proteins overlaps the segment corresponding to helix D of TupV C protein (see Figures 5 and 7 in [22]). Here, structure-based alignments show that the C-terminal part of the STAT1-binding site of SeV C is similar to the same helix D of TupV C (see Section 3.3 and Figure 4D,E). Thus, our study reveals that a STAT1-binding site is encoded by exactly the same RNA region in all Orthoparamyxovirinae, but in different frames (either the P or C frame) (Figure 11). This seems too striking to be a coincidence, and suggests that this location may be particularly suitable for encoding a STAT1-binding site.

One reason might be regulatory: both the P and C frames have the potential to be translated at the same time and same cellular location since they are encoded by the same mRNA. If the production of the P/C mRNA becomes optimized by natural selection so that the product of one of its reading frames (for example P) blocks STAT1, the product of the other frame (i.e., C) will also benefit from this regulation. One can speculate that the positioning of the STAT1 binding site in either protein may be a means of improving the



fitness of these viruses and that this possibility is an additional advantage conferred by the overlap of coding regions [70].



**Figure 11.** Position of the STAT1 binding site in *Orthoparamyxovirinae* P or C proteins. The upper panel of the figure shows the alignment of the part of the C proteins encompassing helices D and E of TupV C for members of both clusters. The sequence alignment of MeV, NiV, and TupV C is taken from [22], and SeV C is aligned on the TupV C sequence according to the structural alignment shown in Figure 4D. This alignment is converted into a nucleotide alignment (middle panel), where the regions coding for the STAT1 binding site are highlighted in red (encoded in the reference frame in MeV and NiV) or in green (encoded in the +1 frame in SeV). The lower panel is a schematic representation of the proteins expressed from the P gene, showing the location of the STAT1 binding site either in P/V/W proteins (cluster 1) or in C protein (cluster 2).

**Author Contributions:** A.R., E.C.M., V.A.V., F.B., L.M.-G., V.E.V., D.G.K., J.-M.B. and M.J. conceived the experiments; A.R., E.C.M., L.G., C.L., V.A.V., F.B. and J.-M.B. performed the experiments; A.R., E.C.M., V.A.V., L.M.-G., V.E.V., D.G.K., J.-M.B. and M.J. analyzed and interpreted the data; A.R., D.G.K., J.-M.B. and M.J. wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was supported by grants from the French Agence Nationale de la Recherche to M.J. and V.V. (ANR-18-CE11-0014-02), from the Fond de la Recherche Médicale (FRM) (grant “Equipe 2017” DEQ20170336754) to M.J. and from the FINOVI Foundation to M.J. and V.V. We acknowledge the Synchrotron SOLEIL for the allocation of beamtime on Proxima1 and Swing. This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02)

and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 8BJW.pdb.

**Acknowledgments:** The authors thank A. Desfosses (IBS, Grenoble) for the initial development of the computational tool for calculating the D-score and for helpful discussion regarding the development of the GitHub.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

- Lamb, R.A.; Parks, G.D. *Paramyxoviridae: The Viruses and Their Replication*; Fields, B.N., Knipe, D.M., Howler, P.M., Eds.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2013; Volume 1, pp. 957–995.
- Eaton, B.T.; Broder, C.C.; Middleton, D.; Wang, L. Hendra and Nipah Viruses: Different and Dangerous. *Nat. Rev. Microbiol.* **2006**, *4*, 23–35. [[CrossRef](#)] [[PubMed](#)]
- Vanmechelen, B.; Meurs, S.; Horemans, M.; Loosen, A.; Maes, T.J.; Laenen, L.; Vergote, V.; Koundouno, F.R.; Magassouba, N.; Konde, M.K.; et al. The Characterization of Multiple Novel Paramyxoviruses Highlights the Diverse Nature of the Subfamily Orthoparamyxovirinae. *Virus Evol.* **2022**, *8*, veac061. [[CrossRef](#)] [[PubMed](#)]
- Drexler, J.F.; Corman, V.M.; Muller, M.A.; Maganga, G.D.; Vallo, P.; Binger, T.; Gloza-Rausch, F.; Rasche, A.; Yordanov, S.; Seebens, A.; et al. Bats Host Major Mammalian Paramyxoviruses. *Nat. Commun.* **2012**, *3*, 796. [[CrossRef](#)]
- Kuhn, J.H.; Wolf, Y.I.; Krupovic, M.; Zhang, Y.Z.; Maes, P.; Dolja, V.V.; Koonin, E.V. Classify Viruses—The Gain Is Worth the Pain. *Nature* **2019**, *566*, 318–320. [[CrossRef](#)]
- Kuhn, J.H.; Adkins, S.; Alioto, D.; Alkhovsky, S.V.; Amarasinghe, G.K.; Anthony, S.J.; Avšič-Županc, T.; Ayllón, M.A.; Bahl, J.; Balkema-Buschmann, A.; et al. Taxonomic Update for Phylum Negarnaviricota (Riboviria: Orthornavirae), Including the Large Orders Bunyavirales and Mononegavirales. *Arch. Virol.* **2020**, *165*, 3023–3072. [[CrossRef](#)]
- Pringle, C.R. The Order Mononegavirales—Current Status. *Arch. Virol.* **1997**, *142*, 2321–2326.
- Douglas, J.; Drummond, A.J.; Kingston, R.L. Evolutionary History of Cotranscriptional Editing in the Paramyxoviral Phosphoprotein Gene. *Virus Evol.* **2021**, *7*, veab028. [[CrossRef](#)]
- Hausmann, S.; Garcin, D.; Delenda, C.; Kolakofsky, D. The Versatility of Paramyxovirus RNA Polymerase Stuttering. *J. Virol.* **1999**, *73*, 5568–5576. [[CrossRef](#)]
- Vidal, S.; Curran, J.; Kolakofsky, D. A Stuttering Model for Paramyxovirus P mRNA Editing. *Embo J.* **1990**, *9*, 2017–2022. [[CrossRef](#)] [[PubMed](#)]
- Vidal, S.; Curran, J.; Kolakofsky, D. Editing of the Sendai Virus P/C mRNA by G Insertion Occurs during MRNA Synthesis via a Virus-Encoded Activity. *J. Virol.* **1990**, *64*, 239–246. [[CrossRef](#)]
- Thomas, S.M.; Lamb, R.A.; Paterson, R.G. Two mRNAs That Differ by Two Nontemplated Nucleotides Encode the Amino Coterminal Proteins P and V of the Paramyxovirus SV5. *Cell* **1988**, *54*, 891–902. [[CrossRef](#)]
- Tidona, C.A.; Kurz, H.W.; Gelderblom, H.R.; Darai, G. Isolation and Molecular Characterization of a Novel Cytopathogenic Paramyxovirus from Tree Shrews. *Virology* **1999**, *258*, 425–434. [[CrossRef](#)]
- Mathieu, C.; Guillaume, V.; Volchkova, V.A.; Pohl, C.; Jacquot, F.; Looi, R.Y.; Wong, K.T.; Legras-Lachuer, C.; Volchkov, V.E.; Lachuer, J.; et al. Nonstructural Nipah Virus C Protein Regulates Both the Early Host Proinflammatory Response and Viral Virulence. *J. Virol.* **2012**, *86*, 10766–10775. [[CrossRef](#)]
- Siering, O.; Sawatsky, B.; Pfaller, C.K. C Protein Is Essential for Canine Distemper Virus Virulence and Pathogenicity in Ferrets. *J. Virol.* **2021**, *95*, e01840-20. [[CrossRef](#)]
- Pfaller, C.K.; Radeke, M.J.; Cattaneo, R.; Samuel, C.E. Measles Virus C Protein Impairs Production of Defective Copyback Double-Stranded Viral RNA and Activation of Protein Kinase R. *J. Virol.* **2014**, *88*, 456–468. [[CrossRef](#)]
- Pfaller, C.K.; Bloyet, L.-M.; Donohue, R.C.; Huff, A.L.; Bartemes, W.P.; Yousaf, I.; Urzua, E.; Clavière, M.; Zachary, M.; de Masson d’Autume, V.; et al. The C Protein Is Recruited to Measles Virus Ribonucleocapsids by the Phosphoprotein. *J. Virol.* **2019**, *94*, e01733-19. [[CrossRef](#)]
- Sleeman, K.; Bankamp, B.; Hummel, K.B.; Lo, M.K.; Bellini, W.J.; Rota, P.A. The C, V and W Proteins of Nipah Virus Inhibit Minigenome Replication. *J. Gen. Virol.* **2008**, *89*, 1300–1308. [[CrossRef](#)]
- Park, M.S.; Shaw, M.L.; Munoz-Jordan, J.; Cros, J.F.; Nakaya, T.; Bouvier, N.; Palese, P.; Garcia-Sastre, A.; Basler, C.F. Newcastle Disease Virus (NDV)-Based Assay Demonstrates Interferon-Antagonist Activity for the NDV V Protein and the Nipah Virus V, W, and C Proteins. *J. Virol.* **2003**, *77*, 1501–1511. [[CrossRef](#)] [[PubMed](#)]

20. Sparrer, K.M.J.; Pfaller, C.K.; Conzelmann, K.-K. Measles Virus C Protein Interferes with Beta Interferon Transcription in the Nucleus. *J. Virol.* **2012**, *86*, 796–805. [[CrossRef](#)] [[PubMed](#)]
21. Park, A.; Yun, T.; Vigant, F.; Pernet, O.; Won, S.T.; Dawes, B.E.; Bartkowski, W.; Freiberg, A.N.; Lee, B. Nipah Virus C Protein Recruits Tsg101 to Promote the Efficient Release of Virus in an ESCRT-Dependent Pathway. *PLoS Pathog.* **2016**, *12*, e1005659. [[CrossRef](#)]
22. Lo, M.K.; Sogaard, T.M.; Karlin, D.G. Evolution and Structural Organization of the C Proteins of Paramyxovirinae. *PLoS ONE* **2014**, *9*, e90003. [[CrossRef](#)]
23. Fan, Y.; Huang, Z.Y.; Cao, C.C.; Chen, C.S.; Chen, Y.X.; Fan, D.D.; He, J.; Hou, H.L.; Hu, L.; Hu, X.T.; et al. Genome of the Chinese Tree Shrew. *Nat. Commun.* **2013**, *4*, 1426. [[CrossRef](#)] [[PubMed](#)]
24. Xiao, J.; Liu, R.; Chen, C.S. Tree Shrew (*Tupaia belangeri*) as a Novel Laboratory Disease Animal Model. *Zool. Res.* **2017**, *38*, 127–137. [[CrossRef](#)] [[PubMed](#)]
25. Gerard, F.C.A.; Ribeiro, E.A.; Leyrat, C.; Ivanov, I.; Blondel, D.; Longhi, S.; Ruigrok, R.W.H.; Jamin, M. Modular Organization of Rabies Virus Phosphoprotein. *J. Mol. Biol.* **2009**, *388*, 978–996. [[CrossRef](#)] [[PubMed](#)]
26. Prilusky, J.; Felder, C.E.; Zeev-Ben-Mordehai, T.; Rydberg, E.H.; Man, O.; Beckmann, J.S.; Silman, I.; Sussman, J.L. FoldIndex: A Simple Tool to Predict Whether a given Protein Sequence Is Intrinsically Unfolded. *Bioinformatics* **2005**, *21*, 3435–3438. [[CrossRef](#)]
27. Romero, P.; Obradovic, Z.; Li, X.; Garner, E.C.; Brown, C.J.; Dunker, A.K. Sequence Complexity of Disordered Protein. *Proteins* **2001**, *42*, 38–48. [[CrossRef](#)]
28. Xue, B.; Dunbrack, R.L.; Williams, R.W.; Dunker, A.K.; Uversky, V.N. PONDR-FIT: A Meta-Predictor of Intrinsically Disordered Amino Acids. *Biochim. Biophys. Acta* **2010**, *1804*, 996–1010. [[CrossRef](#)]
29. Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C.J.; Dunker, A.K. Predicting Intrinsic Disorder from Amino Acid Sequence. *Proteins* **2003**, *53* (Suppl. 6), 566–572. [[CrossRef](#)] [[PubMed](#)]
30. Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Dunker, A.K. Exploiting Heterogeneous Sequence Properties Improves Prediction of Protein Disorder. *Proteins* **2005**, *61* (Suppl. 7), 176–182. [[CrossRef](#)]
31. Dosztányi, Z.; Csizmek, V.; Tompa, P.; Simon, I. Structural Bioinformatics IUPred: Web Server for the Prediction of Intrinsically Unstructured Regions of Proteins Based on Estimated Energy Content. *Bioinformatics* **2005**, *21*, 3433–3434. [[CrossRef](#)]
32. Linding, R.; Jensen, L.J.; Diella, F.; Bork, P.; Gibson, T.J.; Russell, R.B. Protein Disorder Prediction: Implications for Structural Proteomics. *Structure* **2003**, *11*, 1453–1459. [[CrossRef](#)] [[PubMed](#)]
33. Linding, R.; Russell, R.B.; Neduva, V.; Gibson, T.J. GlobPlot: Exploring Protein Sequences for Globularity and Disorder. *Nucleic Acids Res.* **2003**, *31*, 3701–3708. [[CrossRef](#)] [[PubMed](#)]
34. Walsh, I.; Martin, A.J.M.; di domenico, T.; Tosatto, S.C.E. Espritz: Accurate and Fast Prediction of Protein Disorder. *Bioinformatics* **2012**, *28*, 503–509. [[CrossRef](#)] [[PubMed](#)]
35. Callebaut, I.; Labesse, G.; Durand, P.; Poupon, A.; Canard, B.; Chomilier, J.; Henrissat, B.; Mornon, J.P. Deciphering Protein Sequence Information through Hydrophobic Cluster Analysis (HCA): Current Status and Perspectives. *Cell. Mol. Life Sci.* **1997**, *53*, 621–645. [[CrossRef](#)]
36. Yabukarski, F.; Lawrence, P.; Tarbouriech, N.; Bourhis, J.M.; Delaforge, E.; Jensen, M.R.; Ruigrok, R.W.; Blackledge, M.; Volchkov, V.; Jamin, M. Structure of Nipah Virus Unassembled Nucleoprotein in Complex with Its Viral Chaperone. *Nat. Struct. Mol. Biol.* **2014**, *21*, 754–759. [[CrossRef](#)]
37. Jensen, M.R.; Yabukarski, F.; Communie, G.; Condamine, E.; Mas, C.; Volchkova, V.; Tarbouriech, N.; Bourhis, J.M.; Volchkov, V.; Blackledge, M.; et al. Structural Description of the Nipah Virus Phosphoprotein and Its Interaction with STAT1. *Biophys. J.* **2020**, *118*, 2470–2488. [[CrossRef](#)]
38. Ribeiro, E.A.; Favier, A.; Gerard, F.C.A.; Leyrat, C.; Brutscher, B.; Blondel, D.; Ruigrok, R.W.H.; 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. [[CrossRef](#)]
39. Martinez, N.; Ribeiro, E.A., Jr.; Leyrat, C.; Tarbouriech, N.; Ruigrok, R.W.H.; Jamin, M.; Ribeiro, E.A.; Leyrat, C.; Tarbouriech, N.; Ruigrok, R.W.H.; et al. Structure of the C-Terminal Domain of Lettuce Necrotic Yellows Virus Phosphoprotein. *J. Virol.* **2013**, *87*, 9569–9578. [[CrossRef](#)]
40. Studier, F.W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr. Purif.* **2005**, *41*, 207–234. [[CrossRef](#)]
41. Wyatt, P.J. Submicrometer Particle Sizing by Multiangle Light Scattering Following Fractionation. *J. Colloid Interface Sci.* **1998**, *197*, 9–20. [[CrossRef](#)]
42. Uversky, V.N. Use of Fast Protein Size-Exclusion Liquid Chromatography to Study the Unfolding of Proteins Which Denature through the Molten Globule. *Biochemistry* **1993**, *32*, 13288–13298. [[CrossRef](#)]
43. Pérez, J.; Thureau, A.; Vachette, P. SEC-SAXS: Experimental Set-up and Software Developments Build up a Powerful Tool. In *Methods in Enzymology*; Academic Press Inc.: Cambridge, MA, USA, 2022; Volume 677, pp. 221–249.
44. Franke, D.; Petoukhov, M.V.; Konarev, P.V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.D.T.; Kikhney, A.G.; Hajizadeh, N.R.; Franklin, J.M.; Jeffries, C.M.; et al. ATSAS 2.8: A Comprehensive Data Analysis Suite for Small-Angle Scattering from Macromolecular Solutions. *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225. [[CrossRef](#)] [[PubMed](#)]
45. Bernadó, P.; Mylonas, E.; Petoukhov, M.V.; Blackledge, M.; Svergun, D.I. Structural Characterization of Flexible Proteins Using Small-Angle X-Ray Scattering. *J. Am. Chem. Soc.* **2007**, *129*, 5656–5664. [[CrossRef](#)]

46. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132. [[CrossRef](#)]
47. Pape, T.; Schneider, T.R. HKL2MAP: A Graphical User Interface for Macromolecular Phasing with SHELX Programs. *J. Appl. Cryst.* **2004**, *37*, 843–844. [[CrossRef](#)]
48. Adams, P.D.; Afonine, P.V.; Bunkóczi, G.; Chen, V.B.; Davis, I.W.; Echols, N.; Headd, J.J.; Hung, L.-W.; Kapral, G.J.; Grosse-Kunstleve, R.W.; et al. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213–221. [[CrossRef](#)]
49. Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126–2132. [[CrossRef](#)] [[PubMed](#)]
50. Murshudov, G.N.; Skubak, P.; Lebedev, A.A.; Pannu, N.S.; Steiner, R.A.; Nicholls, R.A.; Winn, M.D.; Long, F.; Vagin, A.A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355–367. [[CrossRef](#)] [[PubMed](#)]
51. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; et al. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235–242. [[CrossRef](#)]
52. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Židek, A.; Bates, R.; Blackwell, S.; Yim, J.; et al. Protein Complex Prediction with AlphaFold-Multimer. *bioRxiv* **2021**. [[CrossRef](#)]
53. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
54. Ciancanelli, M.J.; Volchkova, V.A.; Shaw, M.L.; Volchkov, V.E.; Basler, C.F. Nipah Virus Sequesters Inactive STAT1 in the Nucleus via a P Gene-Encoded Mechanism. *J. Virol.* **2009**, *83*, 7828–7841. [[CrossRef](#)] [[PubMed](#)]
55. Blakqori, G.; Kochs, G.; Haller, O.; Weber, F. Functional L Polymerase of La Crosse Virus Allows in Vivo Reconstitution of Recombinant Nucleocapsids. *J. Gen. Virol.* **2003**, *84*, 1207–1214. [[CrossRef](#)] [[PubMed](#)]
56. Halpin, K.; Bankamp, B.; Harcourt, B.H.; Bellini, W.J.; Rota, P.A. Nipah Virus Conforms to the Rule of Six in a Minigenome Replication Assay. *J. Gen. Virol.* **2004**, *85*, 701–707. [[CrossRef](#)]
57. Beaty, S.M.; Park, A.; Won, S.T.; Hong, P.; Lyons, M.; Vigant, F.; Freiberg, A.N.; TenOever, B.R.; Duprex, W.P.; Lee, B. Efficient and Robust Paramyxoviridae Reverse Genetics Systems. *mSphere* **2017**, *2*, e00376-16. [[CrossRef](#)]
58. Buchan, D.W.A.; Jones, D.T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Res.* **2019**, *47*, W402–W407. [[CrossRef](#)] [[PubMed](#)]
59. Holm, L. Dali Server: Structural Unification of Protein Families. *Nucleic Acids Res.* **2022**, *50*, W210–W215. [[CrossRef](#)]
60. Oda, K.; Matoba, Y.; Irie, T.; Kawabata, R.; Fukushi, M.; Sugiyama, M.; Sakaguchi, T. Structural Basis of the Inhibition of STAT1 Activity by Sendai Virus C Protein. *J. Virol.* **2015**, *89*, 11487–11499. [[CrossRef](#)]
61. Oda, K.; Matoba, Y.; Sugiyama, M.; Sakaguchi, T. Structural Insight into the Interaction of Sendai Virus C Protein with Alix To Stimulate Viral Budding. *J. Virol.* **2021**, *95*, JV10081521. [[CrossRef](#)]
62. Pei, J.; Kim, B.H.; Grishin, N.V. PROMALS3D: A Tool for Multiple Protein Sequence and Structure Alignments. *Nucleic Acids Res.* **2008**, *36*, 2295–2300. [[CrossRef](#)]
63. Svergun, D.I. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Cryst.* **1992**, *25*, 495–503. [[CrossRef](#)]
64. Hansen, S. Bayesian Estimation of Hyperparameters for Indirect Fourier Transformation in Small-Angle Scattering. *J. Appl. Cryst.* **2000**, *33*, 1415–1421. [[CrossRef](#)]
65. Martinez-Gil, L.; Vera-Velasco, N.M.; Mingarro, I. Exploring the Human-Nipah Virus Protein-Protein Interactions. *J. Virol.* **2017**, *91*, e01461-17. [[CrossRef](#)] [[PubMed](#)]
66. Meignié, A.; Combredet, C.; Santolini, M.; Kovács, I.A.; Douché, T.; Gianetto, Q.G.; Eun, H.; Matondo, M.; Jacob, Y.; Grailhe, R.; et al. Proteomic Analysis Uncovers Measles Virus Protein C Interaction with P65–IASPP Protein Complex. *Mol. Cell. Proteom.* **2021**, *20*, 100049. [[CrossRef](#)]
67. Rodriguez, J.J.; Parisien, J.P.; Horvath, C.M. Nipah Virus V Protein Evades Alpha and Gamma Interferons by Preventing STAT1 and STAT2 Activation and Nuclear Accumulation. *J. Virol.* **2002**, *76*, 11476–11483. [[CrossRef](#)]
68. Rodriguez, J.J.; Wang, L.-F.; Horvath, C.M. Hendra Virus V Protein Inhibits Interferon Signaling by Preventing STAT1 and STAT2 Nuclear Accumulation. *J. Virol.* **2003**, *77*, 11842–11845. [[CrossRef](#)] [[PubMed](#)]
69. Devaux, P.; von Messling, V.; Songsungthong, W.; Springfield, C.; Cattaneo, R. Tyrosine 110 in the Measles Virus Phosphoprotein Is Required to Block STAT1 Phosphorylation. *Virology* **2007**, *360*, 72–83. [[CrossRef](#)] [[PubMed](#)]
70. Wright, B.W.; Molloy, M.P.; Jaschke, P.R. Overlapping genes in natural and engineered genomes. *Nat. Rev. Genet.* **2022**, *23*, 154–168. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

### III. Conclusions

Nous avons résolu la structure cristallographique de la protéine C<sub>Δ53</sub> à une résolution de 2.4 Å. L'unité asymétrique du cristal de C<sub>Δ53</sub> est composée de 6 sous-unités assemblées en dimères dans le groupe d'espace P3<sub>2</sub>21. Chaque monomère est formé d'un repliement de 3 hélices α.

Il existe moins de 20% de similarités de séquence entre les protéines C des *Respirovirus* et des *Narmovirus*. Pourtant, une recherche de similarités structurales par DALI a permis d'aligner la protéine C<sub>Δ53</sub> au domaine replié C-terminal de la protéine C du virus Sendai dont la structure a été résolue par cristallographie en complexe avec le domaine N-terminal de la protéine humaine STAT1 (PDB 3WWT) et avec le domaine BRO-1 de la protéine humaine ALIX (PDB 6KP3) (Oda *et al.*, 2015, 2021). Nous avons également retrouvé des résidus conservés entre les protéines C du SeV et du TupV qui sont impliqués dans l'interaction avec le facteur cellulaire STAT1 pouvant suggérer l'existence d'une fonction partagée entre ces deux protéines. Nos résultats, ainsi que la superposition de la structure de la protéine C du TupV aux modèles AlphaFold2 générés pour plusieurs protéines C des *Orthoparamyxovirinae*, représentent les premiers indices d'ordre structural de l'existence d'une homologie entre les protéines C.

Nos résultats de caractérisation de la protéine C pleine longueur par SEC-SAXS et du domaine désordonné par la méthode d'optimisation des Ensembles nous a permis de modéliser le comportement hautement flexible de la région désordonnée N-terminale et des deux premières hélices N-terminales de la protéine C de TupV en solution.

Enfin, nous avons évalué la capacité de la protéine C du TupV d'inhiber de façon croisée, comme cela a été démontré précédemment (Malur, Hoffman and Banerjee, 2004; Sleeman *et al.*, 2008; Su *et al.*, 2015), la réplication du minigénome du virus Nipah. De la même façon, nous avons observé que les protéines C du NiV et du TupV sont capables d'inhiber de façon dose-dépendante l'expression du gène rapporteur Renilla luciférase du minigénome.

Nous avons alors essayé d'identifier par TAP-MS des protéines cellulaires humaines capables d'interagir avec la protéine C du TupV. Aucune des cibles identifiée n'était présente dans l'interactome de la protéine C du NiV (Martinez-Gil, Vera-Velasco and Mingarro, 2017) ou du virus de la rougeole (Meignié *et al.*, 2021). Bien qu'il soit possible que ces résultats soient en lien avec la spécificité du TupV à se répliquer uniquement en cellules de toupaye, nous avons émis l'hypothèse que cette fonction des protéines C pourrait être en lien avec une interaction directe avec la machinerie de synthèse de l'ARN viral.

**C'est dans cette optique que nous avons décidé d'évaluer l'existence d'une interaction avec la phosphoprotéine P du TupV, cofacteur essentiel de la réplication et objet du chapitre suivant.**





# Chapitre 5 : Étude structurale de la phosphoprotéine du virus de la toupaye

## I. Introduction

Toutes les protéines C des *Orthoparamyxovirinae* sont capables d'inhiber de façon non-spécifique la réplication de la machinerie de synthèse de l'ARN viral. Cependant, les études d'interactome réalisées ne semblent pas identifier de partenaires cellulaires communs à toutes ces protéines. Par ailleurs, de nombreuses études de caractérisation ont mis en lumière l'existence de similarités structurales et fonctionnelles des protéines du complexe de réplication/transcription des paramyxovirus. Il serait donc envisageable que ce mécanisme d'inhibition puisse reposer sur une interaction entre protéines virales de ce complexe et les protéines C.

De la même façon que pour les protéines C, malgré moins de 20% d'identité de séquences entre les phosphoprotéines de la famille, de nombreux domaines fonctionnels partagent une même structure. Ces domaines jouent notamment un rôle clé dans l'interaction avec la polymérase virale et les nucléoprotéines faisant de la phosphoprotéine un cofacteur essentiel de la réplication/transcription virale.

Dans ce contexte, nous avons fait l'hypothèse d'une interaction entre la phosphoprotéine et la protéine C du TupV. Cette hypothèse nous a conduit dans un premier temps à exprimer, purifier et caractériser la phosphoprotéine du TupV afin de la comparer aux phosphoprotéines de la famille. Ces résultats de caractérisation feront l'objet d'une publication en cours de préparation qui sera soumise prochainement. Dans la suite de ce chapitre, nous présentons donc nos premières conclusions sur la caractérisation du domaine d'oligomérisation de la P du TupV sous la forme d'un article. Ce manuscrit sera complété prochainement par la caractérisation structurale du domaine P<sub>XD</sub> C-terminal de la phosphoprotéine et par de plus amples analyses de l'interaction avec la protéine C du TupV.

## II. Article en cours de préparation : Caractérisation structurale de domaines conservés de la phosphoprotéine du virus de la toupaye.



# Structural characterization of conserved domains within Tupaia paramyxovirus phosphoprotein

Ada Roy<sup>1</sup>, Paco Tarillon<sup>1</sup>, Jean-Marie Bourhis<sup>1,\*</sup> and Marc Jamin<sup>1,\*</sup>

<sup>1</sup> Université Grenoble Alpes, CNRS, CEA, Institut de Biologie Structurale, 38000, Grenoble, France

\* Correspondence: Marc Jamin - e-mail: marc.jamin@ibs.fr, phone: + 33 4 76 20 94 62; Jean-Marie Bourhis - e-mail: jean-marie.bourhis@ibs.fr, phone: + 33 4 57 42 86 36;

**Abstract:** The Tupaia paramyxovirus (TupV) phosphoprotein (P) is a modular multimeric protein. In several Orthoparamyxovirus, P proteins play a central cofactor role for the viral polymerase by tethering it to the nucleocapsid. Despite a very low sequence identity shared with other Paramyxoviridae phosphoproteins, several functional domains are conserved throughout the family. In this study, we present the first bioinformatics structural review of a phosphoprotein of the genus *Narmovirus*. We have shown that TupV P in solution has a tetrameric organization by SEC-MALLS. We resolved the crystal structure of the multimerization domain (P<sub>MD</sub>) of TupV P, which comprises a long, parallel, tetrameric coiled coil. Using SEC-SAXS, we further described P<sub>MD</sub> structure in solution as an ensemble conformers comprising N- and C-terminal disordered regions. The comparison with other known structures of P<sub>MD</sub> revealed discrepancies with Sendai virus multimerization domain structure, but strong similarity with measles virus (MeV) oligomerization in agreement with the current phylogenetic tree.

**Keywords:** phosphoprotein; *Paramyxoviridae*; Tupaia Paramyxovirus; protein structure

## 1. Introduction

The Tupaia paramyxovirus (TupV) was isolated and characterized in 1999 in Southeast Asia from degenerating kidney cells collected from a small tree-dwelling mammal, the tree shrew (*Tupaia belangeri*), near Bangkok [1]. The tree shrew, which belongs to the order Scandentia, lives in the tropical forest of Southeast Asia where it feeds on fruits, insects and small vertebrates. Recent studies have shown that it is genetically close to the primate order, making this small, easy-to-breed mammal, a laboratory model of choice for the study of a variety of human diseases and the development of innovative therapies.

The characterisation of the virion morphology and the sequencing of the TupV genome has allowed its classification in the large family *Paramyxoviridae* which includes many pathogens such as measles virus, human parainfluenza viruses and the zoonotic Hendra and Nipah viruses. These enveloped viruses have a single-stranded negative RNA genome and are therefore classified in the phylum *Negarnaviricota* and the order *Mononegavirales*. The re-evaluation of the International Committee on Taxonomy of Viruses (ICTV) phylogeny of paramyxoviruses in 2019 led to the classification of TupV in the newly created genus *Narmovirus*.

The multiplication of paramyxoviruses in their host cell is based on common mechanisms in which the phosphoprotein (P) plays an essential role. In particular, it involves the transcription of the RNA genome into monocistronic messenger RNAs (mRNAs), which are then used as a template by the cellular ribosomes for producing viral proteins. It also involves the replication of the negative genomic RNA, taking place in two steps: (i) the production of full-length positive-sense copies of the genomic RNA, called anti-genomes, and (ii), using these copies as templates, the synthesis of negative genomic RNA copies. The latter are then incorporated into new virus particles.

The machinery performing both transcription and replication of the viral genome, which is conserved within the Mononegavirales, is formed by the genomic RNA associated with three viral proteins: the nucleoprotein (N), the phosphoprotein (P) and the large subunit of the RNA-dependent viral polymerase (L).

The nucleoprotein exists in two forms: an unassembled monomeric form containing no RNA, called  $N^0$ , and a polymeric form that encapsidates both genomic RNAs and antigenomic RNAs, forming long helical complexes named nucleocapsid (NC). These NCs are essential to make genomic and antigenomic RNAs competent to be used by the viral polymerase.

Structural dynamics studies of phosphoprotein interaction with  $N^0$  and the nucleocapsid have provided insight into the mechanisms of nucleocapsid assembly and remodelling during transcription and replication [2], [3] where the phosphoprotein plays a central role in the molecular ballet played by these three proteins. Indeed, the viral polymerase (L) does not have the ability to bind strongly to the NC but the phosphoprotein binds to both the polymerase and the nucleoprotein in complex with the RNA [2], [4]. By allowing the polymerase to attach to the NC, P promotes the initiation of RNA synthesis (positioning the polymerase at the 3' end of the genome) and provides the necessary processivity for the synthesis of long RNAs. P independently has the ability to bind unassembled RNA-free  $N^0$ . It then acts as a molecular chaperone preventing  $N^0$  proteins from assembling and from non-specifically binding cellular RNAs and delivering  $N^0$  molecules to the site of RNA synthesis to encapsulate nascent RNAs [5], [6]. Finally, P is involved in host-virus interactions and in many viruses including NiV and MeV, it is involved in counteracting the interferon (IFN) response by associating with the cellular factor STAT-I [6].

Despite their low sequence identity, the paramyxovirus phosphoproteins share a common architecture [7], [8]. They are composed of a long, intrinsically disordered N-terminal region ( $P_{NTR}$ ) that includes an N-terminal molecular recognition element ( $\alpha$ -MoRE), which consist of an intrinsically disordered region that fold upon binding, that recruits  $N^0$  and a C-terminal region ( $P_{CTR}$ ) that includes the oligomerization domain ( $P_{MD}$ ), a flexible linker region ( $P_{CT-LINK}$ ) and a C-terminally folded X-domain ( $P_{XD}$ ) that mediates binding to nucleocapsids [6], [8]–[11].

The  $P_{XD}$  domain is able to bind to a  $\alpha$ -MoRE motif present in the disordered C-terminal tail of the nucleoprotein ( $N_{TAIL}$ ). The structures of the  $P_{XD}$  domain of different paramyxoviruses have been determined including that of Sendai viruses (SeV) [12], measles virus (MeV) [13], Hendra virus (HeV) [14] and Nipah virus (NiV) [9].

The central domain ( $P_{MD}$ ) is involved in both the oligomerization of the protein and in the interaction with the L polymerase [15], [16]. The crystallographic structures of several paramyxovirus  $P_{MD}$  have been determined including that of SeV [16], MeV [17] and NiV [6], [10]. All form tetrameric coiled coils (CC).

Here, we set up to characterize the structure of TupV phosphoprotein. We used bioinformatic to localize the folded domains and determined the oligomerization state of the full-length protein and of two different fragments encompassing the putative multimerization domain ( $P_{MD}$ ) and the C-terminal X domain ( $P_{XD}$ ). We solved the crystallographic structure of its oligomerization domain  $P_{MD}$  at a resolution of 2.6 Å and characterized its structure in solution by SEC-SAXS. We further studied the structure of TupV phosphoprotein  $P_{XD}$  domain by SEC-SAXS and AlphaFold modelling.

## 2. Materials and Methods

### 2.1. Disorder metaprediction

The location of intrinsically disordered regions (IDRs) in TupV P protein was predicted by submitting its amino-acid sequence to the freely accessible Dscore scoring automatic query available on GitHub at <https://github.com/brisvag/dscore> [5], [6], [18].

In this procedure, 16 different predictors accessible through web servers are used to calculate a consensus prediction as previously described [5]. Foldindex was run with a window size of 51 residues [19]. For the PONDR VLXT, XL1, CAN\_XT, VL3\_BA and VSL2 predictors, residues with a score  $> 0.5$  were considered disordered. PONDR-FIT was run using the default parameters [23]. For the IUPred2 predictors for long and short disordered regions, residues with a score  $> 0.5$  were considered disordered [24]. The DisEMBL (Loops/Coils, Hot-loops and Remark465) server was run using the default parameters [25]. The GLOBPLOT server was run with the default Russel/Linding propensity [26]. The ESpritz predictors (NMR, X-ray, Disprot) were used with the default parameters [27]. HCA (Hydrophobic Cluster Analysis) plots were drawn with DRAWHCA [28].

A simple scoring procedure was then used to define consensus disordered regions. For each prediction, residues that were 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 averaging the value from all predictions. Regions with a normalized D-score lower than 0.50 are considered intrinsically disordered, whereas regions with a score higher than 0.5 are either folded domains or potential molecular recognition elements (MoRE) for partner proteins [3], [6], [28].

The prediction of secondary structures of TupV P<sub>FL</sub> was performed via the RPBS web interface. This server allowed the generation of HCA graphs [28] and a PSIPRED 4.0.1 secondary structure analysis [29].

### 2.2. Protein expression and purification

The gene encoding the Full-Length P protein (P<sub>FL</sub>) of TupV (Uniprot accession number: Q9WS39) in fusion with a TEV protease cleavage site and a C-terminal twin-streptag II tag was cloned between the NcoI/XhoI restriction sites into the PET-M11 vector (Novagen Darmstadt, Germany).

We expressed TupV oligomerization domain, P<sub>MD</sub>, consisting of residues 336 to 403, identified by our bioinformatics predictions of secondary structures and domains with a high propensity for coiled coils bundle formation. This gene was cloned into a pET-28a between the NcoI and XhoI restriction sites in fusion with an N-terminal hexahistidine (6HIS) tag.

Similarly, the C-terminal domain, P<sub>XD</sub>, comprising residues 473 to 527, was cloned between the NcoI and XhoI restriction sites in fusion with a separate N-terminal 6HIS tag followed by a TEV protease cleavage site.

The proteins were expressed in *E. coli* BL21 STAR (DE3). The bacteria were grown at 37°C until an optical density value at 600nm (OD<sub>600</sub>) of 0.6 was reached. The temperature was then lowered to 18°C and the protein expression phase was induced by adding isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. The culture is continued for approximately 18h. The cells are then collected by centrifugation.

Cell pellets expressing  $P_{MD}$  and  $P_{XD}$  are resuspended on ice in 50 mM Tris-HCl buffer pH 8, containing 300 mM NaCl, 5 mM imidazole pH 8, 0.2 mM tris(2-carboxyethyl)phosphine (TCEP) and one tablet of protease inhibitor cocktail (buffer A). The cells are lysed by sonication and centrifuged to collect the soluble protein fraction in the supernatant which is applied to a Ni<sup>2+</sup> NTA resin (Qiagen, Hilden, Germany) previously equilibrated with buffer A. The resin is then washed with a solution of 50 mM Tris-HCl buffer pH 8, containing 1 M NaCl, 10 mM imidazole pH 8 and 0.2 mM TCEP. The purification is followed by a Bradford colorimetric assay. The proteins are then eluted with a solution of 50mM Tris-HCl buffer pH 8, containing 150 mM NaCl, 250 mM imidazole pH 8 and 0.2 mM TCEP (buffer B). An additional 6HIS tag cleavage step by TEV protease is optionally performed for the  $P_{XD}$  protein in parallel with dialysis (3 kDa cut-off) of the sample against an imidazole-free buffer B solution (buffer C). The proteins are then purified by size exclusion chromatography on a Hiload Superdex 200 16/600 column (Cytiva, Freiburg im Breisgau, Germany) for the  $P_{MD}$  protein and on a Hiload Superdex 75 16/600 column (Cytiva, Freiburg im Breisgau, Germany) for the  $P_{XD}$  protein, equilibrated with a 25 mM Tris-HCl buffer solution at pH 8, containing 150 mM NaCl and 0.2 mM TCEP (buffer D).

Cell pellets expressing the  $P_{FL}$  protein are resuspended on ice in 100 mM Tris-HCl buffer pH 8, containing 150 mM NaCl and 5 mM  $\beta$ -mercaptoethanol and one tablet of protease inhibitor cocktail (buffer E). The cells are lysed by sonication and centrifuged to collect the soluble protein fraction in the supernatant, which is applied to Strep-Tactin® Sepharose resin (IBA Lifescience, Göttingen, Germany) previously equilibrated with buffer E. The resin is then washed with buffer E. The purification is followed by a Bradford colorimetric assay. The protein is then eluted with a solution of buffer E supplemented with 50 mM Biotin. The eluted protein is concentrated using a Millipore concentrator (Darmstadt, Germany) (cut-off 100 kDa). The protein is then purified by size exclusion chromatography on a Hiload Superdex 200 16/600 column equilibrated with buffer D.

After checking the purity of the proteins under the elution peaks by SDS-PAGE in a 4-20% acrylamide gradient gel (Biorad, Hercules, California), they are then concentrated using Millipore concentrators (3 kDa threshold).

### 2.3. Circular dichroism spectrum

Circular dichroism measurements for the  $P_{MD}$  protein were performed on a MOS500 spectrophotometer (BioLogic) at room temperature in 20 mM Tris-HCl buffer at pH 8 containing 150 mM NaCl and 0.2 mM TCEP. The far UV CD spectra were recorded from 190 nm to 260 nm in a 1 mm cell at a protein concentration of about 30  $\mu$ M.

### 2.4. Size exclusion chromatography coupled with multiangle laser light scattering experiments [41]

SEC was performed on a Superdex Increase 200 10/300 GL column (Cytiva) equilibrated with 20 mM Tris-HCl buffer at pH 8 containing 150 mM NaCl and 0.2 mM TCEP. Chromatographic separations were performed at room temperature with a flow rate of 0.5 mL.min<sup>-1</sup> with on-line MALLS detection using a DAWN-HELEOS II detector (Wyatt Technology, Santa Barbara, CA, United States), and protein concentration was measured on-line by differential refractive index measurements using an Optilab T-rEX detector (Wyatt Technology) and a refractive index increment,  $dn/dc$ , of 0.185 mL.g<sup>-1</sup>. The data were analyzed with the software ASTRA (Wyatt Technology).

### 2.5. X-ray crystallography

Initial crystallisation conditions for the TupV P<sub>MD</sub> domain were identified via the Partnership for Structural Biology (PSB) High Throughput Crystallization (HTX) platform (<https://htxlab.embl.fr>). Crystals used for data collection were obtained by the hanging drop method and/or the HTX platform in 0.1M phosphate citrate buffer pH 5.2, containing 5% PEG 1000, 20% ethanol and 3% 2-methyl-2,4-pentanediol and were frozen without cryoprotectant.

X-ray diffraction data were collected at the synchrotron SOLEIL PROXIMA-1 beamline (Saint-Aubin, France) at a wavelength of 1.0088 Å and a temperature of 100K using a Dectris Eiger-X 16M detector. They were indexed and integrated with the XDS suite [30]. Then, the data were processed using STARANISO [31] which applies non-elliptical anisotropic bounds based on a locally averaged  $I/\sigma(I)$  threshold, performs Bayesian estimation of structure amplitudes and applies an anisotropic correction to the data. A three-dimensional model of P<sub>MD</sub> tetramer was constructed by the ColabFold collaborative suite (AlphaFold2 using MMseqs2) [32]. Molecular replacements were obtained with Phaser [33] from the CCP4 suite [34] using the AlphaFold2 model and subsequently refined with Refmac5 [35] and COOT software [36].

### 2.6. Size exclusion chromatography coupled with small angle x-ray scattering experiments [43]

Small angle x-ray scattering (SAXS) data were collected at the BioSAXS beamline (SWING) of the Synchrotron Soleil (Saint-Aubin, France). Purified protein samples were analyzed on a Superdex Increase 200 5/150 GL (Cytiva, Freiburg im Breisgau, Germany) in a mobile phase of 25 mM Tris-HCl buffer pH 8, containing 150 mM NaCl and 1 mM TCEP at a temperature of 20°C and a flow rate of 0.3 mL.min<sup>-1</sup>. The scattering from the buffer alone was measured before and after each sample measurement and used for background subtraction. Data were integrated and subtracted with Foxtrot [37] providing the first estimate of  $I(0)$  and  $R_g$  for each SEC profile image. This allowed the identification along the elution peak of frames with constant  $R_g$  values. The corresponding images were merged into an average curve used for further analysis. Final  $R_g$  values were determined by ATSAS [38] and RAW [39].  $D_{MAX}$  were determined by GNOM [40] in the ATSAS suite and compared to the values obtained with BIFT [41]. The molecular weights of the proteins were estimated using the  $Q_r$  ratio derived from the correlation invariant volume [42].

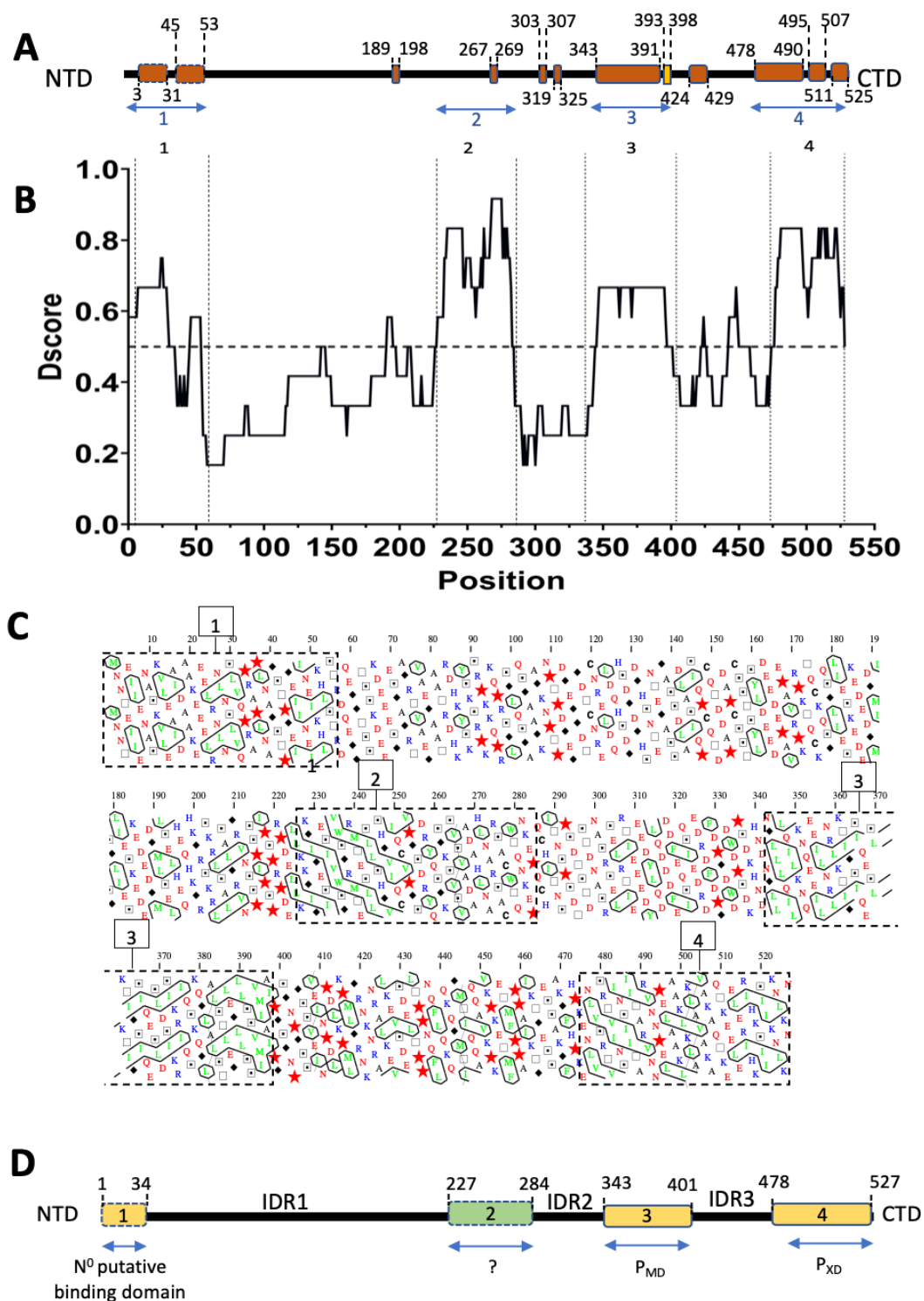
Ensembles of conformers were generated with the program RANCH, while ensemble selection was performed with the program GAJOE [43].

The ab-initio three-dimensional reconstruction of the electron density of P<sub>MD</sub> from the experimental SAXS data was performed by DENSS algorithm [44].

## 3. Results

### 3.1. *Tupaia paramyxovirus phosphoprotein has a modular architecture that resembles that of other phosphoproteins*

In a first approach, we performed a bioinformatic analysis of the TupV P amino acid sequence (Uniport accession number: Q9WS39) using previously described disorder prediction methods (Figure 1A) [5], [6] and secondary structure prediction by PSIPRED programme (Figure 1B&C).



**Figure 1. Bioinformatics study of the TupV phosphoprotein from the aa sequence.** (A) Predictions of secondary structures by the PSIPRED method. Orange rectangles correspond to regions predicted to fold into  $\alpha$ -helices and yellow rectangles into  $\beta$ -sheets. (B) Dscore plot: regions with a calculated Dscore below 0.5 are considered disordered and those with a score higher than 0.5 as structured. 4 identified domains are highlighted by the dotted lines superimposed on the graph. (C) HCA plot: red stars = prolines, black diamonds = glycine, square = threonine, dashed square = serine, joined contours represent hydrophobic clusters. 4 hydrophobic domains are identified and highlighted by black rectangle contoured by dotted lines. The first 4 superimposed with the domains identified in (B). (D) Schematic representations of identified regions by the bioinformatics analysis. IDR stands for Intrinsically Disordered Region.

255

256

257

258

259

260

261

262

263

264

265

The D-score identified four regions (summarized in **Figure 1D**), predicted to be folded shown by the boxes 1 to 4 in Figure 1A for which we further analysed sequence conservation (**Figure 2**) by multiple sequence alignment within the *Narmovirus* genus. These folded domains alternate with three intrinsically disordered regions (named IDR1 to IDR3 in figure 1D) shown in **Figure 1D**.

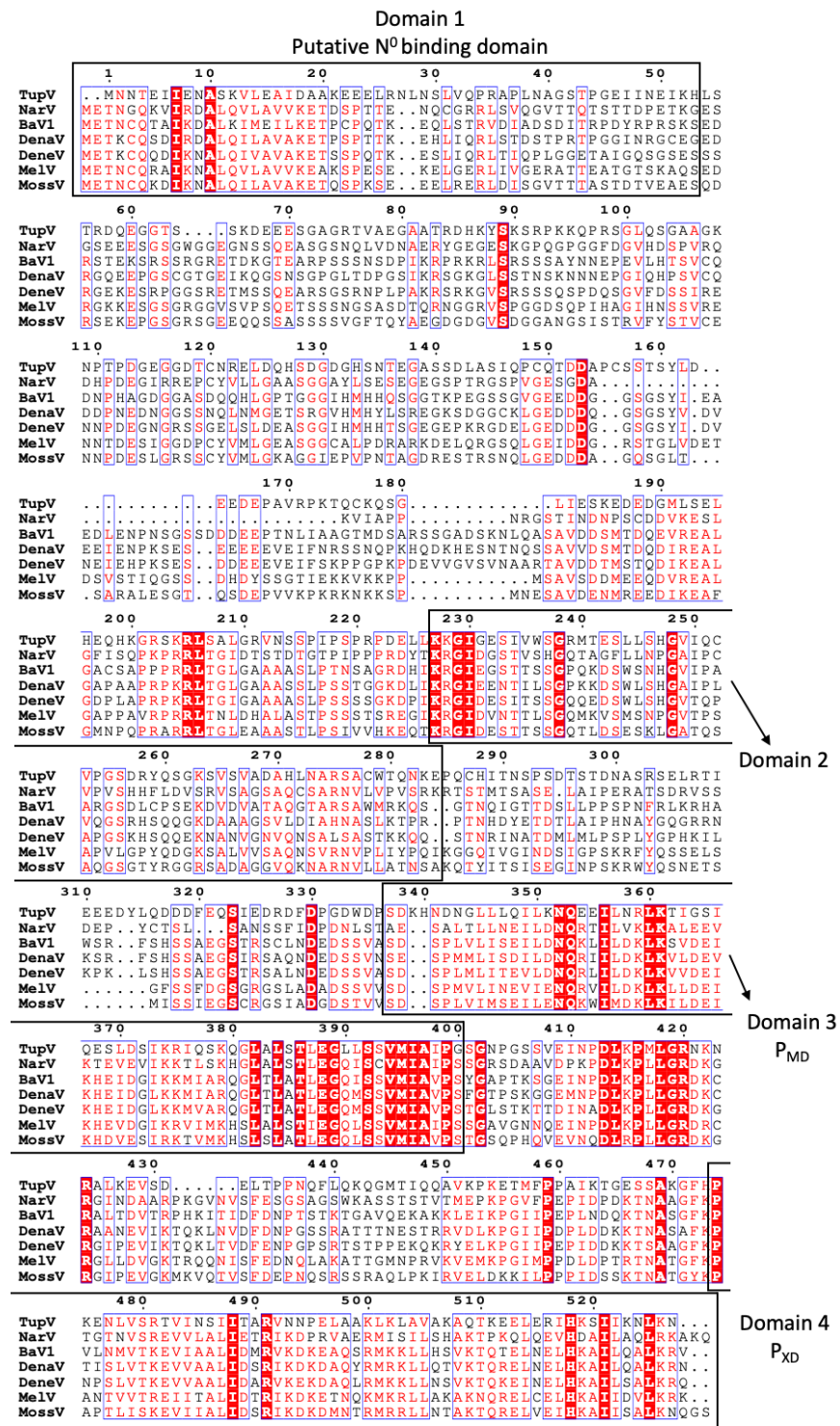
The N-terminal end of the protein contains a short structured region (Box 1: aa 1 to 34) with 2 predicted helical segments that could act as a  $\alpha$ -MoRE. Comparison with the P of other paramyxoviruses suggests that this  $\alpha$ -MoRE could chaperone the N<sup>0</sup> nucleoproteins [12]. This region is well conserved pattern within the genus (residues 1 to 25) with two strictly conserved residues: I<sub>7</sub> and A<sub>10</sub> (**Figure 2, domain 1**).

The first identified IDR1 includes residues 34 to 226. Within this region, peaks in the Dscore analysis are visible between aa 125 and 225, but do not correspond to predicted secondary structure.

The region between residues 227 and 284 (Box 2) contains a hydrophobic cluster which has a Dscore > 0.5, but almost no predicted secondary structure. A K(K/R)GI motif between 228 and 231 is conserved in the entire genus (**Figure 2, domain 2**). AlphaFold2 predicted no folded region between residues 227 and 284 (**Supplementary data 1**). We hypothesized that this region is mostly disordered and could be a binding motif to a cellular protein for an unknown function. Of note, an equivalent region was found in the NiV P protein, which correspond to the binding sites for STAT1 and STAT2 [6]. However, the IDR1 is shorter in NiV and thus the motif is located 100 residues upstream, closer to the N<sup>0</sup> binding  $\alpha$ -MoRE [6].

The third ordered domain of Dscore > 0.5 between residues 343 to 401 (Box 3 corresponds to the putative oligomerization domain, P<sub>MD</sub>). This region is rich in hydrophobic residues and is predicted to be folded in  $\alpha$ -helices. This region has the highest number of strictly conserved residues with other narmovirus P proteins with 16 strictly conserved positions (**Figure 2, domain 3**).

At the C-terminal end, a predicted folded, helical domain (Box 4: residues 478 to 527) correspond to the putative X domain of TupV P protein, P<sub>XD</sub>. This domain also shows strong sequence conservation between residues 450 to 527 with 8 strictly conserved residues (**Figure 2D**). 34% of the residues in this region are hydrophobic and there are two strictly conserved prolines that interrupt predicted  $\alpha$ -helices.



**Figure 2. Multiple sequence alignment (CLUSTAL method) of *Narmovirus* phosphoproteins.** Accession number of the phosphoprotein aa sequence used: Tupaia paramyxovirus (TupV) - Q9WS39; Nariva virus (NarV) - B8XH59; Mossman virus (MossV) - Q6WGM5; Bank vole virus 1 (BaV1) - A0A2H4PJ46; Meleucus virus (MelV); Denestis virus (DeneV); Denalis virus (DenaV). Strictly conserved positions are highlighted in red and highly conserved regions are framed in blue. The boxes correspond to the domains identified by bioinformatics analysis: Domain 1 N<sup>0</sup> binding domain (residues 1 to 33); Domain 2 unknown motif (residues 227 to 284); Domain 3 PMD (residues 337 to 402); Domain 4 PXD (residues 474 to 527).

300

301

302

303

304

305

306

307

308

309



### 3.2. Biophysical characterization of TupV P<sub>FL</sub> and its autonomous folding domains P<sub>MD</sub> and P<sub>XD</sub>

#### 3.2.1 Stoichiometry of TupV P<sub>FL</sub> and its autonomous folding domains P<sub>MD</sub> and P<sub>XD</sub>

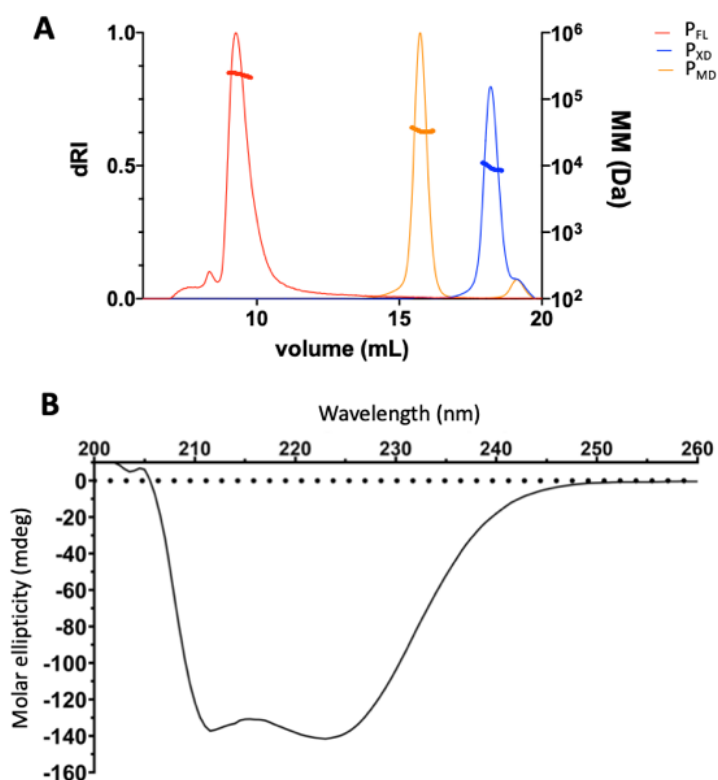
Based on these predictions, we localized the putative oligomerization domain P<sub>MD</sub> (residues 336 to 403) and the X domain P<sub>XD</sub> (residues 473 to 527). The synthetic full-length gene (P<sub>FL</sub>) was commercially obtained and P<sub>FL</sub>, P<sub>MD</sub> and P<sub>XD</sub> were subcloned into bacterial expression vectors. The proteins were expressed in *E. coli* BL21 STAR (DE3) and purified in two successive steps by affinity chromatography on a Ni<sup>2+</sup>-NTA column (P<sub>MD</sub> and P<sub>XD</sub>) or on a Streptactin column (P<sub>FL</sub>) and by a size exclusion chromatography step. All 3 proteins eluted under a single elution peak in size exclusion chromatography.

We determined the molecular mass (MM) and oligomerization status by SEC-MALLS (**Table 1**). For each protein, the MM of the proteins was constant throughout the elution peak with a polydispersity index  $M_w / M_n < 1.005$  indicating that the samples were monodisperse (**Figure 3A**).

**Table 1.** Average molecular weight predicted from the amino acid sequences and measured by SEC-MALLS-RI for TupV P<sub>FL</sub> protein and P<sub>MD</sub> and P<sub>XD</sub> domains.

Construct	P <sub>FL</sub>	P <sub>MD</sub>	P <sub>XD</sub>
MM <sub>predicted</sub> (kDa)	60.5	8.8	9.6
MM <sub>SEC-MALLS</sub> (kDa)	245 ± 3.00	34.0 ± 0.80	9.5 ± 0.4

The MM of the P<sub>FL</sub> protein determined from multi-angle light scattering and excess refractive index data was 245 ± 10 kDa indicating that the protein exists in solution as tetramers (theoretical MM calculated from the gene = 60.5 kDa) (**Figure 3A**, red curve). Furthermore, the MM of the P<sub>MD</sub> fragment of 34.0 ± 0.8 kDa (MM calculated from the gene = 8,8 kDa) confirms that this part of the protein also forms tetramers in solution (**Figure 3A**, orange curve). Finally, we observed that the P<sub>XD</sub> protein was monomeric in solution with a MM at 9.5 ± 0.4 kDa (MM calculated from the gene = 9,6 kDa) (**Figure 3A**, blue curve).

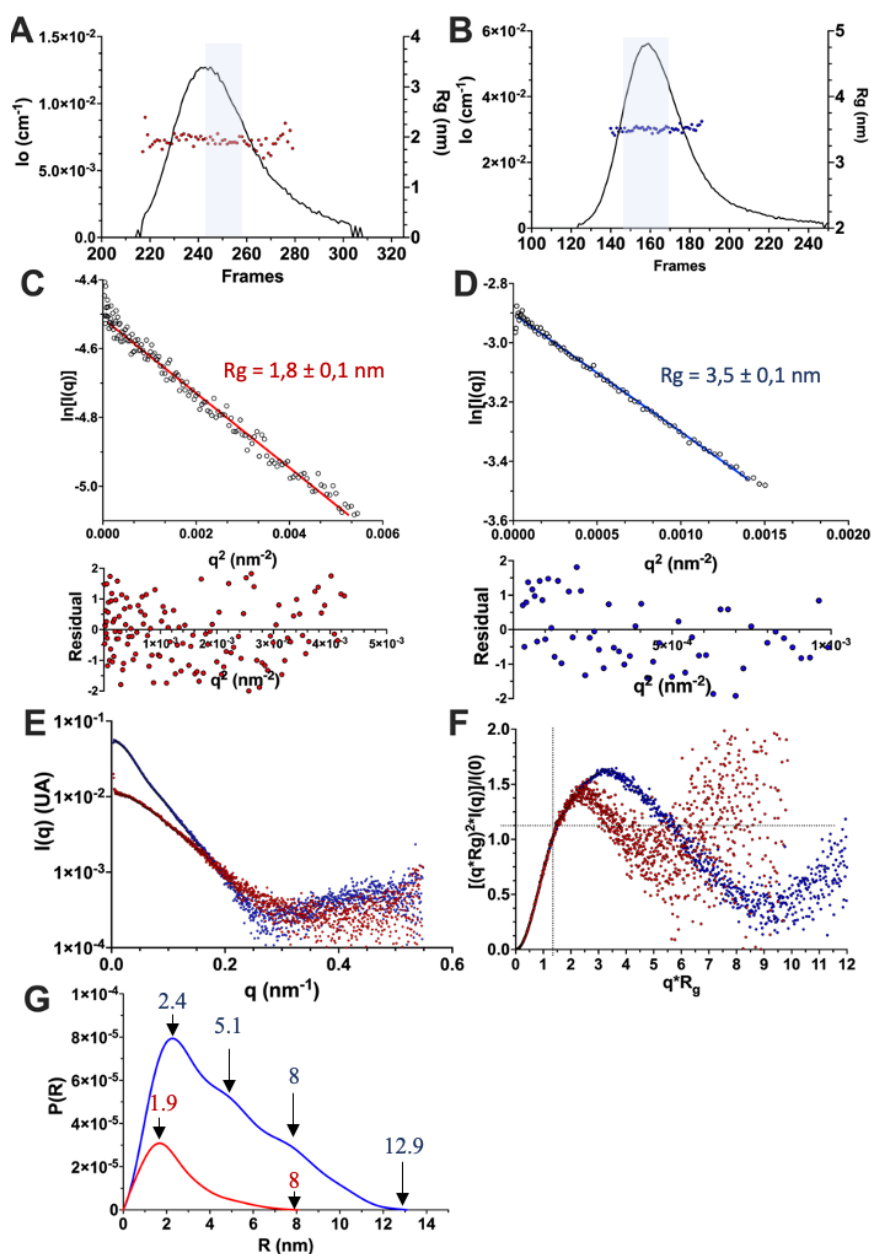


**Figure 3. Characterisation of the TupV  $P_{FL}$ ,  $P_{MD}$  and  $P_{XD}$  oligomerisation status and secondary structures study  $P_{MD}$  domain:** (A) SEC-MALLS of  $P_{FL}$  (red curve),  $P_{MD}$  (orange curve) and  $P_{XD}$  (blue curve). The lines show the chromatograms monitored by differential refractive index (DRI) measurements. The molecular mass were measured across each elution peak calculated from static light scattering and refractive index. (B) Circular dichroism spectrum of  $P_{MD}$  domain measured between 200 to 260 nm.

The far UV circular dichroism (CD) spectrum of the  $P_{MD}$  protein shows negative bands of comparable intensities near 208 and 222 nm characteristic of the presence of  $\alpha$ -helices (Figure 3B).

### 3.2.2. TupV oligomerization and X domains characterization by SEC-SAXS

$P_{XD}$  and  $P_{MD}$  proteins were analysed by SEC-SAXS (respectively, Figure 4A and Figure 4B). Their Guinier plots at small  $q$  ( $q \cdot R_g < 1.3$ ) are linear (respectively, Figure 4C and Figure 4D) and the slope does not change with concentration, indicating the absence of aggregation or intermolecular interactions of  $P_{XD}$  protein (Figure 4C, bottom panel), but a rise in the signal at very small  $q$  may suggest slight aggregation of  $P_{MD}$  protein (Figure 4D, bottom panel). The radii of gyration ( $R_g$ ) were calculated using the Guinier approximation allowing us to estimate values of  $1.8 \pm 0.1$  nm and  $3.5 \pm 0.1$  nm for  $P_{XD}$  and  $P_{MD}$  respectively. These values are in agreement with the values obtained by the pair distribution function ( $1.8 \pm 0.2$  nm and  $3.7 \pm 0.1$  nm for  $P_{XD}$  and  $P_{MD}$  respectively). The MM of the proteins is extrapolated from the scattering intensity at zero angle  $I(0)$  and, from the Guinier estimate and the volume correlation method, we calculated a MM of  $35.7 \pm 1.8$  kDa for  $P_{MD}$  ( $MM_{theoric} = 8,8$  kDa) and  $8.8 \pm 0.4$  kDa for  $P_{XD}$  ( $MM_{theoric} = 9,6$  kDa), which confirms that  $P_{MD}$  is likely tetrameric and  $P_{XD}$  monomeric in solution. These data reinforce the hypothesis about the functions of the domains studied.



**Figure 4. Small-angle X-ray scattering experiment for P<sub>XD</sub> and P<sub>MD</sub> domains.** (A) Evolution of  $I(0)$  and  $R_g$  values along the SEC elution profile for the P<sub>XD</sub> domain. The black line shows the scattering at zero angle ( $I(0)$ ), which is proportional to both molecular mass and concentration, as a function of the frames recorded at regular time intervals. The red dots show the radius of gyration calculated from the Guinier plots at the different time intervals. (B) Evolution of  $I(0)$  and  $R_g$  values along the SEC elution profile for the P<sub>MD</sub> domain. The black line and the blue dots are as in panel A. (C) Guinier plots. The upper panel shows the Guinier plots for P<sub>XD</sub>. The lower panels show the plots of the normalized residuals. (D) Guinier plots. The upper panel shows the Guinier plots for P<sub>MD</sub>. The lower panels show the plots of the normalized residuals. (E) Averaged scattering curves. The dots show the scattering profiles obtained for P<sub>XD</sub> (in red) and P<sub>MD</sub> (in blue) by averaging the individual profiles recorded throughout the SEC elution peak shown in panels A and B. (F) Normalized Kratky plots for P<sub>XD</sub> (red) and P<sub>MD</sub> (blue). (G) Distance distribution function for the domains P<sub>XD</sub> (red) and P<sub>MD</sub> (blue).

The dimensionless Kratky plot of P<sub>MD</sub> protein (**Figure 4F**, blue curve) reaches a maximum at a value of  $q \cdot R_g = 3.3$  which is well above the position of the maximum of  $\sim 1.7$  for a globular protein. Furthermore, the ordinate value of 1.6 is also above the typical maximum for a globular protein which is  $\sim 1.1$  (Wiltkins et al, 1999). This strong diver-

364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378

379  
380  
381  
382

gence between the experimental values and those expected for a globular protein highlights that the overall structure of  $P_{MD}$  is elongated. Furthermore, the curve decreases without reaching the x-axis until  $q.Rg \pm 10$ , suggesting the presence of disordered regions in the protein.

A similar interpretation is derived from the analysis of the dimensionless Kratky plot of the  $P_{XD}$  protein (**Figure 4F**, red curve). However, the maximum is reached for a  $q.Rg$  value of 1.2 for an abscissa value of 2.2, so it is lower and shifted to smaller abscissas than for  $P_{MD}$  protein. This indicates that  $P_{XD}$  has a more globular and compact structure compared to  $P_{MD}$ . It is also observed that the plot decreases to  $q.Rg$  values of 5 without reaching the abscissa and then increases again indicating the presence of a disordered region.

The pair distribution function for the  $P_{MD}$  protein reaches a maximum at 2.4 nm followed by a series of less pronounced maxima at 3.9 and 5.8 nm and the profile ends with a long tail with a  $D_{MAX}$  of  $12.9 \pm 0.8$  nm indicative of an elongated object (**Figure 4G**, blue curve). The shape of the distribution function suggests that  $P_{MD}$  is an elongated and oligomeric protein.

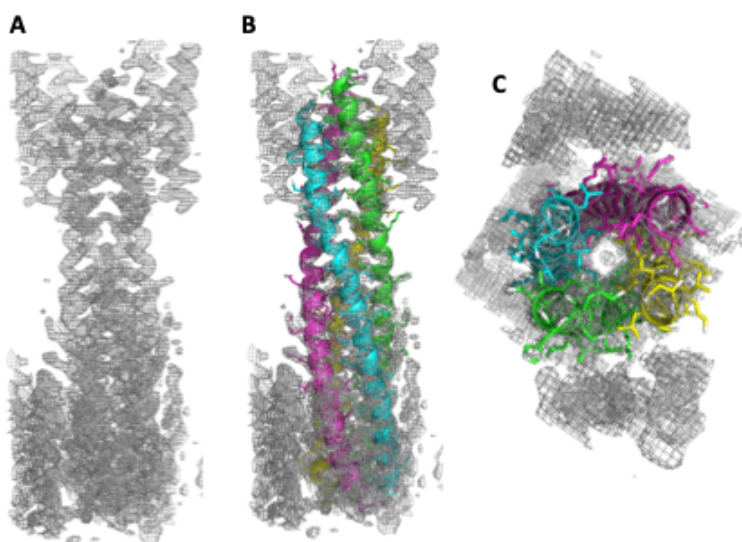
In the case of  $P_{XD}$ , the  $P(R)$  function reaches a peak at 1.9 nm and then exhibits an extended tail that ends at 8.0 nm also suggesting that this protein has an elongated shape, which is surprising in view of the structures of  $P_{XD}$  already solved and could suggest the existence of a disordered region in this construct (**Figure 4G**, red curve).

### 3.3. Crystal structure of the oligomerization ( $P_{MD}$ ) domain revealed structural similarity with other Paramyxoviridae phosphoproteins.

#### 3.3.1. Crystal structure of the oligomerization domain of TupV

TupV  $P_{MD}$  protein was analysed by X-ray crystallography. Initial conditions for crystallization were identified at the High Throughput Crystallisation (HTX) platform (<https://htxlab.embl.fr/>) of the Partnership for Structural Biology (PSB). Crystals used for data collection were obtained in 0.1M citrate/phosphate buffer at pH 5.2, containing 5% PEG 1000 (v/v), 20% ethanol and 3% 2-methyl-2,4-pentanediol and were frozen without cryoprotectant. Two major crystal shapes were obtained: triangular prism crystals (and hexagonal pyramid crystals (**Supplementary figure 2**).

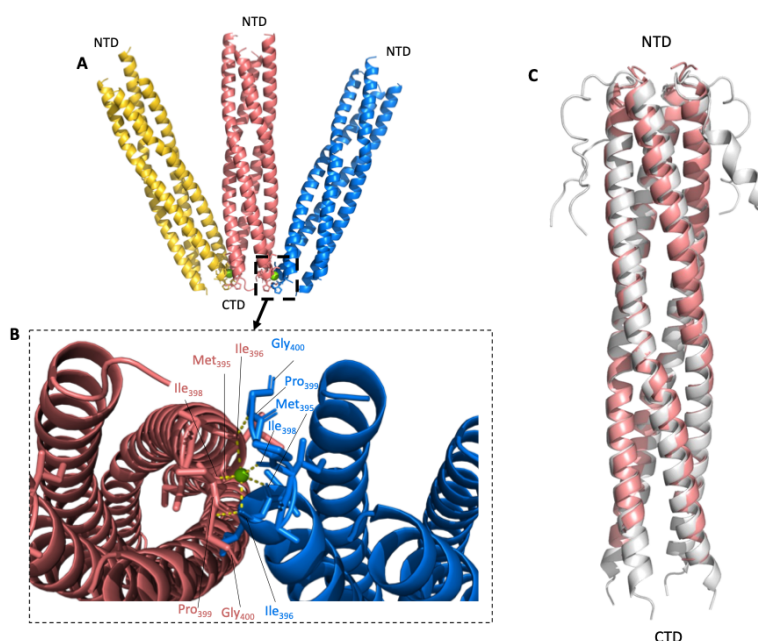
The diffraction pattern of the collected data showed strong anisotropy. After indexing with the XDS suite [30], we could observe that the crystal diffracted in one direction at a resolution of 2.4 Å. The diffraction data were therefore initially corrected using the STARANISO [31] from Global Phasing Ltd. The initial phases were obtained using the molecular replacement method with a tetrameric model obtained with AlphaFold2 [32], [45] and the molecular replacement program PHASER from the CCP4 suite [34], [35], [46]. Multiple refinements were then performed manually using Coot [36] and Refmac5 [46] (**Figure 5A, B, C**). The structure was finally obtained at a resolution of 2.6 Å (**Table 2**).



**Figure 5. Experimental phasing of TupV P<sub>MD</sub> diffraction data.** (A) Electron density map (1 $\sigma$  contour) obtained experimentally by the molecular replacement method for the central tetramer of the asymmetric unit (B) The refined model superimposed on (A). (C) View from above of (B)

The asymmetric unit contains 12 monomers assembled in three tetramers (**Figure 6A**). Residues 339-399 are visible in the electron density map for most of the monomers, an average of 58 residues. Analysis of the model with the PISA webserver [47] suggests that the tetramer association is thermodynamically stable with an average of 10,437 Å<sup>2</sup> of buried surface and a dissociation energy of 75.6 kcal.mol<sup>-1</sup>. These values are highly similar to PISA model analysis of previously resolved *Paramyxoviridae* oligomerization domains (dissociation energy of 198 kcal.mol<sup>-1</sup> for NiV oligomerization domain – PDB 4N5B) with the exception of lower values for MeV oligomerization domain (dissociation energy of 55.8 kcal.mol<sup>-1</sup> - PDB 3ZDO). Remarkably, TupV tetramers are stabilized and interact with each other through an interaction with an Mg<sup>2+</sup> ion at their C-terminal ends (**Figure 6B**). The interaction is mediated by the residues Met<sup>395</sup>, Ile<sup>396</sup>, Ile<sup>398</sup>, Pro<sup>399</sup> and Gly<sup>400</sup> belonging to two neighbouring tetramer.

The root-mean-square deviation (RMSD) between the final P<sub>MD</sub> structure and the model generated by AlphaFold2 that was used for phase resolution is 0.8 Å (**Figure 6C**).



**Figure 6. Description of the asymmetric unit and the coiled-coil association of TupV PMD structure.** (A) The asymmetric unit (AU) of TupV PMD crystal. Each colour corresponds to one of tetramer in the AU (gold and marine for tetramers on the edges and deepsalmon for the central tetramer). (B) Close caption of the C-terminal organization of the AU. The tetramers are stabilised at their C-terminus by 2 Mg<sup>2+</sup> ions (green circle) in contact with one monomer of each of the neighbouring tetramers at the residues Met<sub>394</sub>, Ile<sub>396</sub>, Ile<sub>398</sub>, Pro<sub>399</sub> and Gly<sub>400</sub> (yellow dotted line). (C) Superposition of the resolved central PMD tetramer (deepsalmon) with the AlphaFold2 model (grey).

**Table 2. Crystallographic Data Collection and Refinement Statistics.**

Data collection	PMD		
<b>Beamline</b>	PROXIMA 1 – Soleil		
Wavelength (Å)	1.0088		
Space groupe	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>		
a,b,c (Å)	44.07 ; 114.76 ; 245.31		
$\alpha,\beta,\gamma$ (°)	90 ; 90 ; 90		
Resolution range (Å)	46.97 – 2.52		
<b>Merging statistics (After STARANISO)</b>	Overall	InnerShell	OuterShell
R <sub>merge</sub> <sup>1</sup>	0.072	0.034	0.548
R <sub>meas</sub>	0.083	0.039	0.683
I/ $\sigma$ I <sup>1</sup>	7.3	22.1	1.8
Completeness (%) <sup>1</sup>	90.3	96.4	77.8
Total reflections <sup>1</sup>	101 212	5 201	3 216
Unique reflections <sup>1</sup>	24 393	1 218	1 220
Multiplicity <sup>1</sup>	4.1	4.3	2.6
<b>Refinement</b>	0.2407/0.2870		
R <sub>work</sub> /R <sub>free</sub>	0.2407/0.2870		
No. water	128		
<b>R.m.s.deviations</b>			
Bond length (Å)	0.0067		
Bond angle (°)	1.1904		
Favoured (%)	99.11		
Allowed (%)	0.99		

442  
443  
444  
445  
446  
447  
448  
449  
450

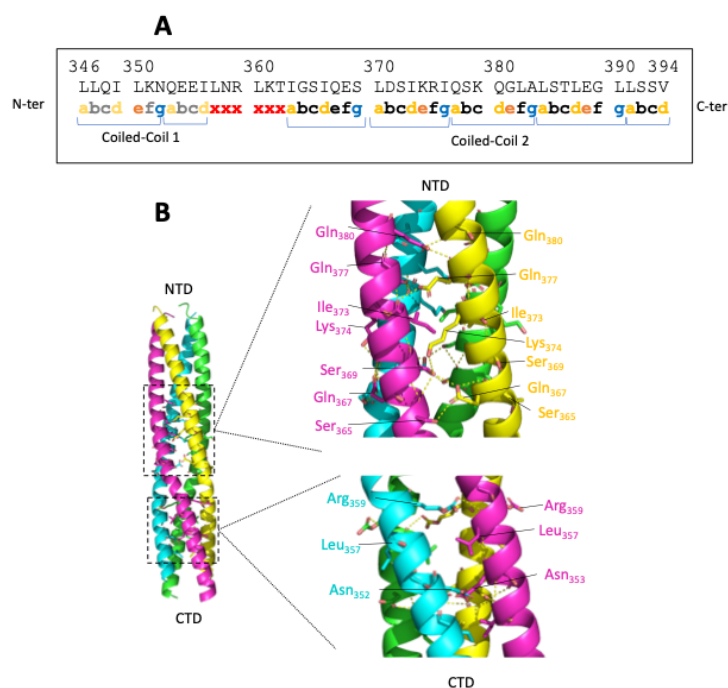
Outlier (%)	0.9
-------------	-----

<sup>1</sup> Values in for all data set and the high-resolution shell.

<sup>2</sup> Values from MolProbity webserver

### 3.3.2. The oligomerization domain coiled-coil canonical organization

We have analysed the organization and folding of the helices forming the coiled coil in P<sub>MD</sub> with SOCKET2 [48], [49] and TWISTER [50]. The canonical primary structure of coiled coils shows the repetition of a motif of seven hydrophobic (h) or polar (p) residues along the  $\alpha$ -helices forming the coiled coil according to the following pattern hpphppp [51]. The aa in this motif are designated by the letters a, b, c, d, e, f and g (**Figure 7A**) [51]. The hydrophobic aa are retained at positions a and d forming a continuous hydrophobic surface along each of the helices in the centre of the coiled coil. These residues are flanked by polar and charged residues.



**Figure 7. Heptad repeats in the coiled coil of TupV P<sub>MD</sub> domain.** (A) The residues involved in the canonical heptad organization are designated by the letters abcdefg along the coiled-coil sequence. The residues forming the hydrophobic core of the tetramer are represented by the letters a and d in yellow. The residues g (in blue) and e (in orange) are systematically on the edge of the hydrophobic core and the polar residues b, f and c (in grey) are outside of the coiled coil. (B) Schematic representation of the heptad organization using the colour code from panel (A) adapted to the aa sequence of TupV P<sub>MD</sub> domain. TupV coiled coil is organised by 5 whole heptads interspersed with a break separating the coiled coil into two parts: a N-terminal one and a C-terminal one. (B) Conserved hydrophobic residues stabilizing the tetrameric structure of the P<sub>MD</sub> domain. The yellow dashed lines represent polar interactions within a distance inferior to 3.5 Å.

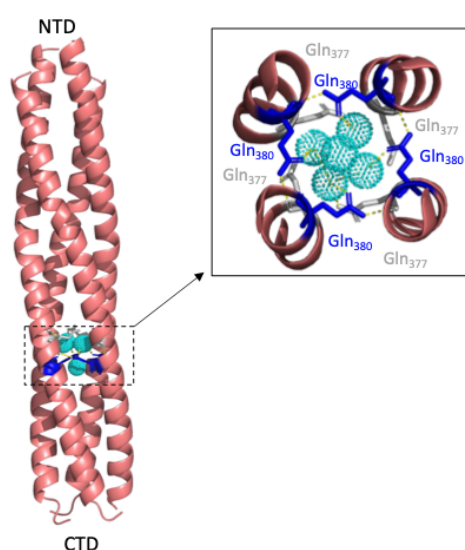
The tetrameric assembly of coiled coils is linked in many cases to the presence of leucine (L) residues in the a-position and isoleucine (I) residues in the d-position. In the case of TupV P<sub>MD</sub>, the heptad motif is present along the helices forming the assembly with a large majority of hydrophobic residues at position a and d (isoleucine, leucine and valine) depicted in **Figure 7B**.

However, at position 353 one can notice the presence of a glutamine (Q) at position a of the heptad which interrupts the heptad organization after two repeats from the N-

terminus at residues 357 to 363 by a small sequence (-LNRLKT-) (**Figure 7A**). This interruption of the heptad repeat does not imply, in the case of TupV P<sub>MD</sub>, a shift of the helix winding whose parameters remain within the ideal Crick values for CC [52]. A major difference between the AlphaFold2 model and the experimental structure of TupV P<sub>MD</sub> lies in the length of the motif interrupting the two heptad repeats, which is only two residues long in the AlphaFold2 model.

Each monomer is composed of a single long  $\alpha$ -helix. The C $\alpha$  of the 12 protomers of the asymmetric unit overlap with an RMSD of less than 0.55 Å and the C $\alpha$  of residues 339 to 399 overlap with an RMSD of less than 0.36 Å.

### 3.3.3. A water-filled pocket within the oligomerization domain structure



Five water molecules are present within a pocket of the central tetramer at the C-terminus formed by the residues Gln<sub>377</sub> and Gln<sub>380</sub> (**Figure 8**).

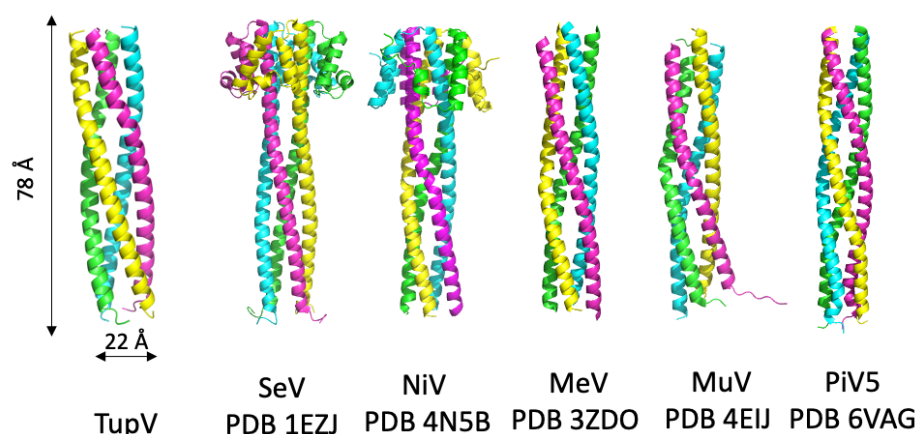
Water molecules are also found in the same proportion within the tetramers formed by the oligomerization domains of MeV [17] and NiV [10], [53]. In contrast, the SeV tetramer encloses 23 water molecules and a Ca<sup>2+</sup> ion and thus appears to be distinct from other structures [16].

**Figure 8. Water-filled pocket within TupV P<sub>MD</sub> domain.** The central tetramer is coloured in deep salmon as in Figure 2 except for the Gln<sub>380</sub> (blue) and Gln<sub>377</sub> (grey) residues whose interactions with 4 of the 5 water molecules enclosed within the structure are detailed. The yellow dotted lines represent polar interactions within a distance of less than 3.5 Å.

### 3.3.4. Comparison with the known *Paramyxoviridae* oligomerization domain

We compared the structure of TupV oligomerisation domain to previously resolved structures in the family Paramyxoviridae. As a first observation, it appears that, as for the other structures of the *Orthoparamyxovirinae* subfamily (MeV, NiV, PiV5 and SeV), the coiled coil of TupV is formed of four parallel  $\alpha$ -helices. In the case of mumps virus (MuV) oligomerization domain in the *Rubulavirinae* subfamily, the coiled coil is formed by an antiparallel dimer of dimer of  $\alpha$ -helices. Furthermore, similarly to MeV, PiV5 and MuV, the coiled coil of TupV is only formed by a long  $\alpha$ -helix and does not have, as is the case for SeV and NiV an additional crown of small helices at the N-terminal end (**Figure 9**).



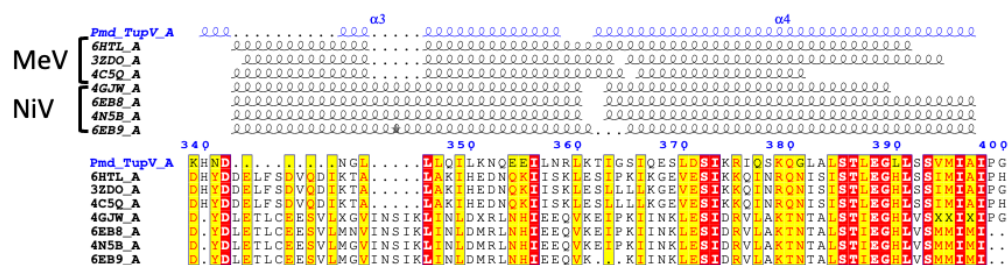


**Figure 9. Structure comparison of the oligomerization domains of different members of the *Paramyxoviridae*.** P<sub>MD</sub> cartoon representation (from left to right) of : TupV, SeV [16], NiV [6], MeV [17], MuV [54], PiV5 [2]. The different chains in each tetramer are coloured in pink, blue, green and yellow.

Note that although only NiV and SeV have this N-terminal crown, they have very different helix parameters. Indeed, in the case of NiV, the crown is made of only two helices and canonical coiled coil residues [10]. For SeV, the crown is made of three helices form this N-terminus and exhibit an assembly of canonical coiled coil residues and less classical polar and charged residues [16]. Finally, the NiV coiled coil contains a proline residue at position 544 causing a break in the ideal Crick parameters for coiled coils inducing a kink in the centre of the motif. It has been shown to be involved in the polymerase function [53] This kink is also present for the MeV and MuV structures but does not induce a variation in the helix parameters [53].

We compared the helical parameters of known oligomerization domain structures via TWISTER [50]. As for NiV, PiV5 and MeV, the average inner radius of the super helix forming the structure is smaller than that measured for SeV (7.5 Å for TupV, 7.3 to 7.5 Å for NiV, PiV5 and MeV against 9 Å for SeV) [2], [6], [10], [16], [17]. Similarly, the frequency of residues per helix turn for TupV (3.5 residues per helix turn) is very similar to NiV, MeV and PiV5 (between 2.8 and 3.5 residues per helix turn) compared to SeV which has a lower frequency (1.7 residues per helix turn).

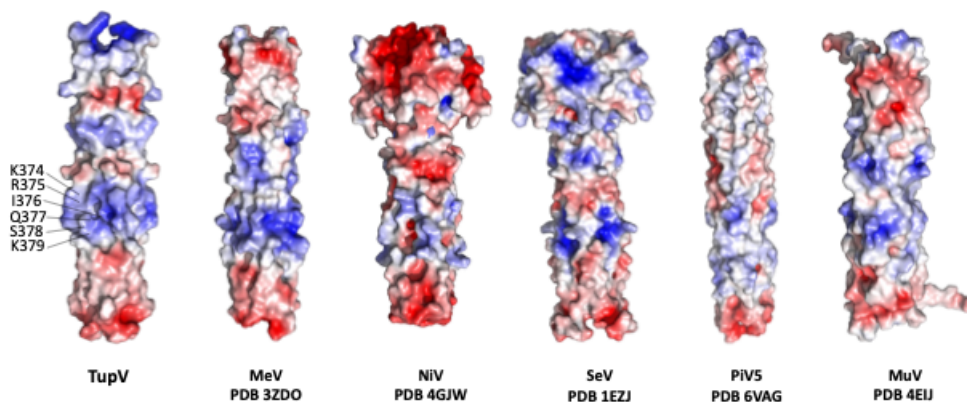
A structural similarity search using the DALI server [55] with TupV P<sub>MD</sub> structure identified several paramyxovirus phosphoprotein oligomerization domains including the MeV and NiV structure. This result was confirmed by ENDscript [56], which enables homologous structures to be identified in the PDB from a protein structure. These servers detected clear homologies between the TupV structure and those of MeV and NiV (not shown). The multiple sequence alignment generated by ENDscript allowed to visualize the conservation between the oligomerization domains of TupV, NiV and MeV which presents 13 strictly conserved positions mainly distributed in the C-terminal domain (**Figure 10**). The oligomerization domain of SeV was not identified by these structural homology searches, thus indicating its outsider position. Moreover, the alignment was centered on the coiled coil C-terminal part and did not include NiV  $\alpha$ -helix cap.



**Figure 10. Multiple sequence alignment generated by ENDSCRIPT [56] base on TupV P<sub>MD</sub> structure.** The upper part of the ENDScript result correspond to secondary structures (helix) for each homologous structures identified in the PDB. The second part displays the corresponding aa sequences. The 3 first PDB identified corresponds to several MeV P<sub>MD</sub> structures (chain A). The 4 following PDB identified corresponds to several NiV P<sub>MD</sub> structures (chain A).

### 3.3.5. A conservation of a central basic patch in *Paramyxoviridae* oligomerization domain

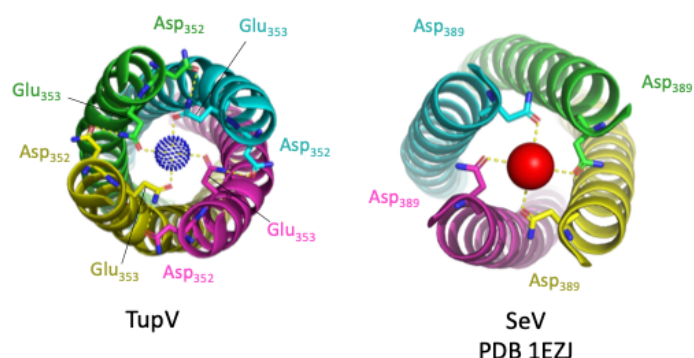
We generated and compared the electrostatic potential surface of known tetramer structures with TupV using the Adaptive Poisson-Boltzmann Solver (APBS) program [57], [58] (Figure 11). We found the presence of a basic motif between the residues Lys<sub>373</sub> and Lys<sub>379</sub> of TupV P, which could correspond to a polymerase L binding site, as previously described for SeV P<sub>MD</sub> by mutagenesis studies [59]. This motif is present for all paramyxovirus P<sub>MD</sub> domain [10], [53]. In particular, the residues Lys<sub>374</sub>, Arg<sub>375</sub> and Lys<sub>329</sub> are directed towards the outside of the helix and could therefore interact with the polymerase. However, this observation is currently challenged by the cryo-EM resolution of PiV5 polymerase complex which identify the C-terminal residue Met<sub>264</sub> of the oligomerization domain as the only residue of this domain interacting with PiV5 polymerase L [2]. Nonetheless, this structure raises numerous questions concerning the multimerization domain 3D position in regard to the polymerase complex functions.



**Figure 11. Electrostatic potential surface comparison from blue (+5KbT) to red (-5KbT) of oligomerisation domain for several members of the *Paramyxoviridae*.** P<sub>MD</sub> electrostatic potential surface representation (from left to right) of: TupV, MeV [17], NiV [6], SeV [16], PiV5 [2] and MuV [54].

### 3.3.6. Potential ion binding site in TupV oligomerization domain

Interestingly, P<sub>MD</sub> of TupV appears to be stabilised by a tetra-coordination involving the Asp<sub>359</sub> and Glu<sub>360</sub> residues of each monomer (Figure 12) which form, as already observed for the structure of the SeV oligomerization domain [16], a potential ion binding site.



**Figure 12. Comparison of the coordination sites of TupV and SeV oligomerisation domains.** TupV potential ion binding site is shown on the left with a blue sphere in dots. SeV ion binding site is shown on the right with the bound  $\text{Ca}^{2+}$  coloured as a red sphere. Equivalent monomer in each structure are coloured with the same colour.

We tried to interpret the presence of electron density at this site by modelling different ions which did not improve the data quality.

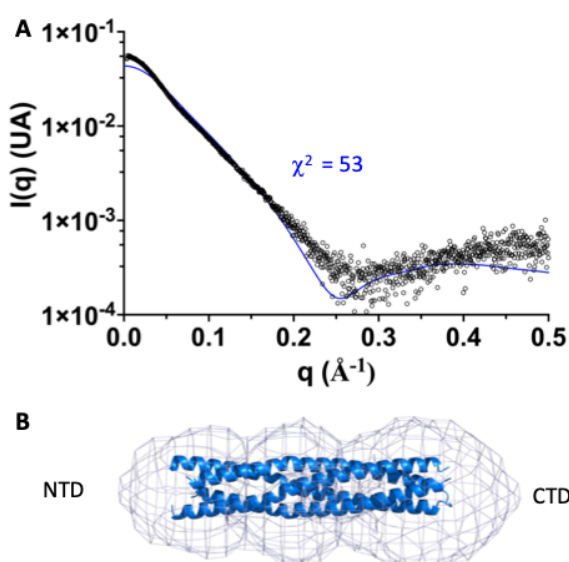
In order to characterize more precisely the potential interaction between  $P_{\text{MD}}$  and the ions  $\text{Ni}^{2+}$  and  $\text{Ca}^{2+}$ , we used isothermal titration calorimetry (ITC). To do this, we inserted 350  $\mu\text{L}$  of overnight dialyzed  $P_{\text{MD}}$  protein (against a 30  $\mu\text{M}$  EDTA solution) in 20mM Tris-HCl buffer at pH8, containing 150mM NaCl and 0.2mM of TCEP into the titration cell as a first step. A 300  $\mu\text{M}$   $\text{CaCl}_2$  solution was then injected into the cell in successive 2.5  $\mu\text{L}$  injections. We observed no interaction between  $\text{Ca}^{2+}$  ions and  $P_{\text{MD}}$ . Similarly,  $P_{\text{MD}}$  in 20mM citric acid buffer at pH 5.5 (preventing the association of  $\text{Ni}^{2+}$  ions with the 6HIS tag), containing 150mM NaCl and 0.2mM of TCEP was then placed in the ITC cell and we injected a 600  $\mu\text{M}$   $\text{NiSO}_4$  solution into the cell in 2.5  $\mu\text{L}$  volumes. No interaction was observed for  $\text{Ni}^{2+}$  ions.

In conclusion,  $P_{\text{MD}}$  does not appear to be able to bind  $\text{Ca}^{2+}$  nor  $\text{Ni}^{2+}$  ions and we found no evidence to support metal ion binding at this site. The electron density can also be explained by the presence of a water molecule at the centre of the tetramer between residues  $\text{Asp}_{352}$  and  $\text{Glu}_{353}$ .

### 3.4. TupV oligomerization and X domains structure in solution

#### 3.4.1. TupV oligomerization domain has disordered N- and C-terminal regions

Comparison of the experimental SAXS curve obtained for the  $P_{\text{MD}}$  protein with the theoretical SAXS curve (Figure 13A, black circles) calculated for the crystal structure of the tetramer using CRYSOLO [60] (Figure 13A, blue curve) showed large discrepancies ( $\chi^2 = 53$ ) clearly indicating that the TupV oligomerization domain structure in solution differed from the crystal structure. Many residues were not visible in the electron density map and could be responsible for the discrepancy observed between the two curves.

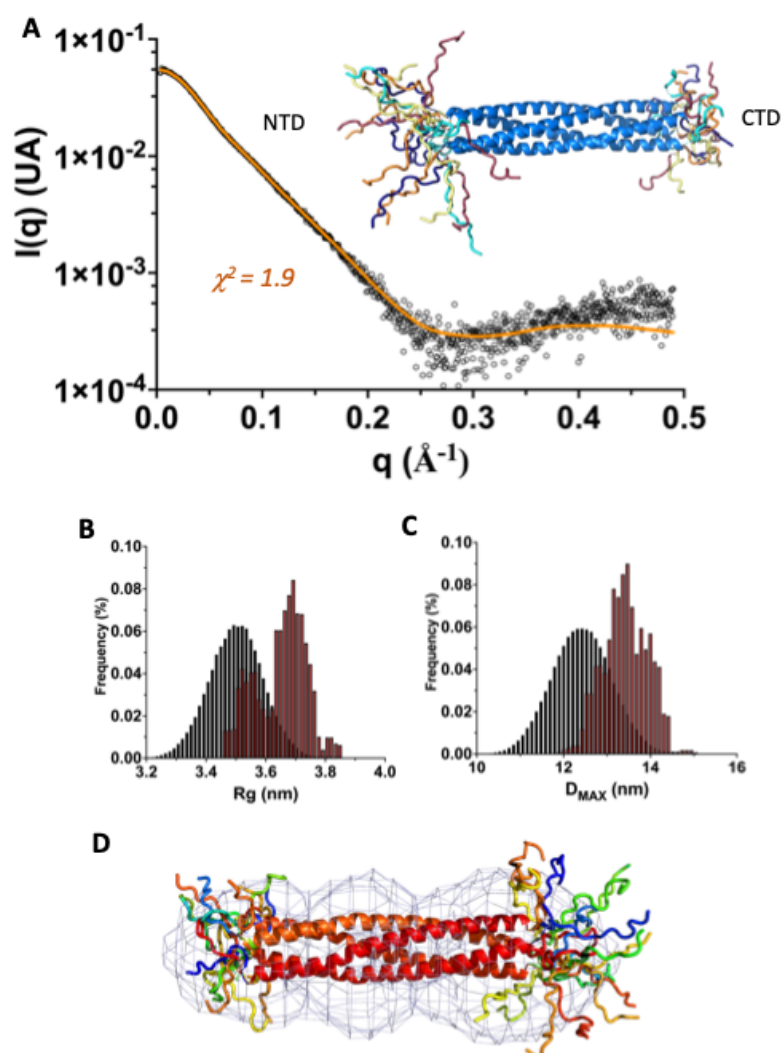


**Figure 13. Comparison of TupV P<sub>MD</sub> experimental SAXS data with the theoretical scattering curve obtained from the crystal structure.** (A) Superimposition of the Crysol theoretical SAXS profile calculated for P<sub>MD</sub> crystal structure (blue curve) with the experimental SAXS data (black circle). (B) Calculation of an ab-initio model for P<sub>MD</sub> experimental SAXS data by the DENSS program superimposed with the crystal structure of P<sub>MD</sub> central tetramer of the asymmetric crystal unit.

Independently of the crystallographic structure, ab initio models were constructed with the DENSS algorithm on the basis of SAXS data. The generated electron density envelope showed an elongated shape characteristic of coiled coils into which the crystallographic structure could be inserted (**Figure 12B**). However, we observed an excess of density on both sides of the coiled coil.

In order to further explore the conformation of the missing residues in the electron density of the crystal, we used the ensemble optimization method (EOM) [43], [61]. This approach generates a large set of conformers of the molecule of interest with RANCH and then selects with GAJOE different subsets of conformers that collectively reproduce the experimental SAXS data. In this case, we generated 10,000 conformers of the P<sub>MD</sub> molecule based on the resolved crystallographic structure and allowed the missing residues to adopt random coil conformations. The program then selected subsets of 5 conformers ( $\chi^2 = 1.94$ ) reproducing the experimental data (**Figure 13A**). The presence of disordered and highly flexible regions in solution at the N- and C-terminal ends of the selected conformers can be clearly observed.

637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662



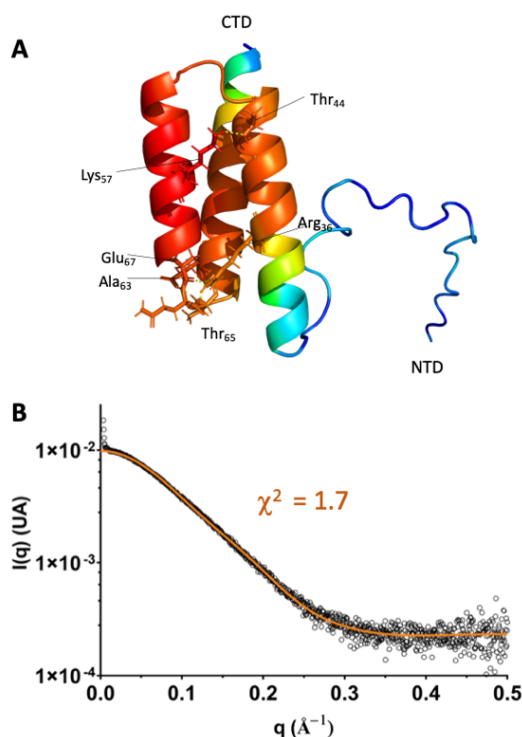
**Figure 14. Modelling of P<sub>MD</sub> N- and C-terminal disordered region flexibility.** (A) We generated an initial ensemble of 10,000 full-length P<sub>MD</sub> conformers based on the resolved tetrameric crystal structure to model N- and C-terminal residues not visible in the electron density. The Ensemble Optimisation Method (EOM) then selected a subset of 5 conformers describing the experimental SAXS data (black dots). The corresponding SAXS scattering curve is depicted in orange and fitted the data with a ( $\chi^2 = 1.94$ ). The selected models are shown above the scattering plot. The blue  $\alpha$ -helices correspond to the domain common to all profiles. The N- and C-terminal disordered segments are differentiated by colour and correspond to the 5 different conformers selected by the method. (B) R<sub>g</sub> distribution for an ensemble of 5 conformers of P<sub>MD</sub>. The black bars show the distribution for the initial ensemble of conformers. The red bars show the R<sub>g</sub> distribution for a selected ensemble of 5 conformers. (C) D<sub>MAX</sub> distribution for an ensemble of 5 conformers of P<sub>MD</sub>. The black bars show the distribution for the initial ensemble of conformers. The red bars show the D<sub>MAX</sub> distribution for a selected ensemble of 5 conformers. Frequency distribution of (D) Ab-initio model calculated by the DENSS program for P<sub>MD</sub> experimental SAXS data superimposed with the selected subset of 5 conformers obtained by EOM.

The initial sets showed Gaussian distributions of R<sub>g</sub> and D<sub>max</sub> values (**Figure 14B** and **Figure 14C**) while the selected sets showed similar bimodal distributions of R<sub>g</sub> and D<sub>max</sub> with a large fraction of more extended conformers and a smaller population of conformers in the average of the initial sets. This suggests that the flexible N- and C-terminal domains of P<sub>MD</sub> are more extended than a random population.

### 3.4.2. TupV X domain can be modeled with a structure of 3 $\alpha$ -helices in the up-down-up conformation

Several initial conditions for crystallization of TupV P<sub>XD</sub> domain were screened by the HTX platform and small crystals of poor quality were obtained in particular for buffer containing 1 to 1.5 M of sodium malonate at pH 4 to 6 as well as for buffer containing 0.1 M of citric acid at pH 5 and 2.4 M of sodium formate at pH 4. We are currently reproducing these crystals manually by the hanging drop method in order to obtain good quality crystals and solve the crystallographic structure of the domain.

Similarly to the P<sub>MD</sub> domain, we planned to solve the domain structure by molecular replacement. For this purpose, we have also generated a three-dimensional model of the protein using AlphaFold2 on the Colab platform (Figure 15A) [32]. This model is composed, as for other known structures of paramyxoviruses X domains, of three  $\alpha$ -helices in the up-down-up conformation stabilized by a hydrophobic core and polar interactions between side chains. An N-terminal disordered region, which is predicted to be structured by the previously performed Dscore calculation, can be observed. This N-terminal tail is linked in part to the N-terminal 6HIS tag but also to an intrinsic disorder in the aa sequence of the protein and could explain the difficulty encountered in the crystallization of the P<sub>XD</sub> protein. It would be useful, for crystallization purposes, to remove these residues.



Crysol.

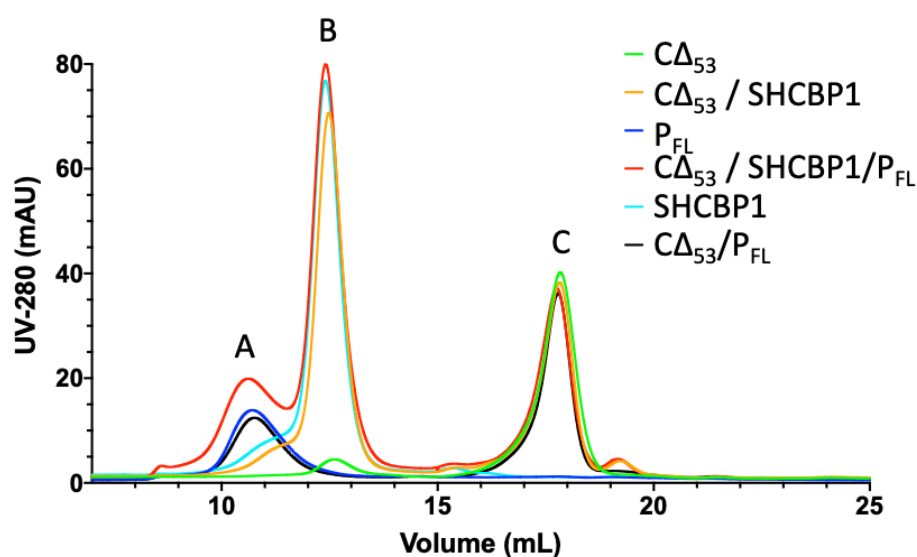
**Figure 15. Crysol fit of P<sub>XD</sub> SAXS experimental scattering curve with an AlphaFold2 model.** (A) Modelling of the three-dimensional structure of the P<sub>XD</sub> domain by AlphaFold2. Residues involved in the folding of the three  $\alpha$ -helices constituting the structure are detailed and the yellow dashed lines represent polar interactions with a distance of less than 3.5 Å. The helices are coloured according to the pLDDT confidence index. The confidence index is > 90% for the C-terminal domain of P<sub>XD</sub> (residues coloured in red). In contrast, this index is lower (< 70%) in the N-terminal domain (residues coloured in blue), indicating a high degree of uncertainty about the disordered region and the first residues composing the most N-terminal  $\alpha$ -helix. (B) Comparison of the theoretical SAXS profile obtained for the AlphaFold2 model (orange curve) with the experimental scattering curve (black circle) with

We assessed the quality of this model by comparing the associated theoretical SAXS curve (Figure 15B, orange curve) generated by CRYSOLOG [60] to the experimental SAXS data (Figure 15B, black circles). We can observe a high similarity ( $\chi^2 = 1.7$ ) of the two curves indicating that the model seems to be a good representation of the domain structure in solution and could probably be used for crystallographic structure resolution by molecular replacement.

3.5. TupV transcription/replication is not regulated through a similar mechanism with MeV

We found structural similarities between the oligomerization domains phosphoproteins structures of MeV and TupV. MeV C protein is able to regulate viral RNA synthesis through an interaction with the cellular factor SHCBP1, thereby disrupting its association with MeV phosphoprotein [62]. We recently determined the ability of TupV C protein to cross-inhibit NiV minigenome replication as previously shown for MeV C protein [63], [64]. This led us to assess whether the ability to inhibit viral replication of TupV C protein was based on similar cellular mechanisms as MeV. We therefore determined whether the proteins P<sub>FL</sub>, C<sub>Δ53</sub> (residues 54 to 157 of TupV C protein) and the human cellular factor SHCBP1 could interact in the same way.

TupV P<sub>FL</sub> and C<sub>Δ53</sub> proteins were purified as described previously [64] and injected alone onto a Superdex 200 Increase column. Then, the proteins were mixed overnight in a small volume at 4°C with an equimolar ratio. All proteins eluted at the same volume alone or after interaction (Figure 16), which indicates, at first glance, that under these conditions the proteins do not appear to interact.



**Figure 16. Preliminary analysis of C<sub>Δ53</sub>, P<sub>FL</sub> and SHCBP1 potential protein complex.** Elution profile of P<sub>FL</sub> (blue curve), SHCBP1 (cyan curve), C<sub>Δ53</sub> (green curve), P<sub>FL</sub> + C<sub>Δ53</sub> (black curve), C<sub>Δ53</sub> + SHCBP1 (orange curve) and C<sub>Δ53</sub> + SHCBP1 + P<sub>FL</sub> (red curve). Proteins were injected alone onto a Superdex 200 10/300 GL increase column before being mixed overnight at 4°C with an equimolar ratio. No shift in elution peaks is observed, indicating a probable lack of interaction under these conditions.

#### 4. Discussion

By interacting with all proteins of the replication machinery and recruiting numerous cellular partners, the phosphoproteins P of paramyxoviruses can be considered as the keystone of viral proliferation. Their particular architecture organized by alternating long disordered regions and structured domains is largely correlated with all these functions.

In this study, we confirmed that TupV P<sub>FL</sub> protein shares with phosphoproteins of paramyxoviruses, the same modular organization with, on the one hand, a mostly dis-

ordered N-terminal region except for an N-terminal domain (residues 1 to 34) probably folded into  $\alpha$ -helices upon binding with partners such as N<sup>0</sup> nucleoproteins and, on the other hand, a C-terminal region composed of two structured domains folded into  $\alpha$ -helices : the oligomerisation domain, P<sub>MD</sub>, (residues 336 to 403) and the X domain, P<sub>XD</sub> (located between residues 473 to 527). Another highly conserved domain from residues 227 to 284 is found on TupV phosphoprotein, which is predicted to be ordered with several hydrophobic residues. However, no secondary structure is predicted by the server PSIPRED and we hypothesized that it could correspond to a specific binding motif mostly unstructured, as it is observed 100 aa further for NiV phosphoprotein with the STAT1 binding motif [6].

We confirmed by SEC-MALLS that this phosphoprotein associates as a tetramer in solution and that this interaction is mediated by the oligomerization domain P<sub>MD</sub>. We solved the first *Narmovirus* phosphoprotein crystallographic structure, the P<sub>MD</sub> domain, at a resolution of 2.6 Å. We have extensively compared our result to previously resolved structures of paramyxoviruses oligomerization domains and observe great similarities with measles virus [17] and, with the exception of a N-terminal cap formed by two  $\alpha$ -helices folding on the coiled-coil, the Nipah virus [6], [10] oligomerization domains. Surprisingly, adding a new structure to the previously resolved phosphoprotein oligomerisation domains seems to reinforce the marginal position of the SeV P<sub>MD</sub> structure (Tarbouriech et al., 2000). This result is somewhat contradictory with the high similarity we found when comparing the crystallographic structure of SeV and TupV C proteins, a product of an overlapping frame within P gene [64].

Our SAXS results highlighted the presence of small disordered regions at the N- and C-terminus of the P<sub>MD</sub> coiled-coil not visible in the electron density. Using the ensemble optimization method, we selected five conformers that better describe the conformation of these flexible regions in solution.

Despite unsuccessful attempts to crystallize the P<sub>XD</sub> domain, the SEC-SAXS results allowed us to validate the quality of an AlphaFold2 model generated for the protein. P<sub>XD</sub> is probably, as previously observed for other paramyxoviruses, formed from the folding of three short  $\alpha$ -helices in the up-down-up conformation stabilized by a hydrophobic core and polar interactions between side chains.

TupV C is, as previously observed for *Paramyxoviridae* C proteins, able to inhibit NiV minigenome replication [64]. We, therefore, investigated the mechanisms of this inhibition. We were unable to validate an interaction between the TupV P and C proteins with the SHCBP1 cell factor, as it has been observed for MeV [62]. This result indicates either the existence of major structural differences with the TupV SHCBP1 protein, despite nearly 95% sequence homology, or differences from MeV in the mechanism of regulation of viral RNA synthesis by MeV C protein. Interestingly, SHCBP1 cell factor has not been found to interact with MeV C protein by a recent interactome study [65]. Furthermore, MeV N and P proteins mixing in vitro triggers liquid-liquid phase separations (LLPS), a microenvironments thought to provide optimal conditions to enhance specific molecular interactions within the condensate while maintaining immiscibility of other cellular molecules [66]. In this regard, it will be interesting to examine whether an interaction between TupV P and C proteins could interact with SHCBP1 in a LLPS context.



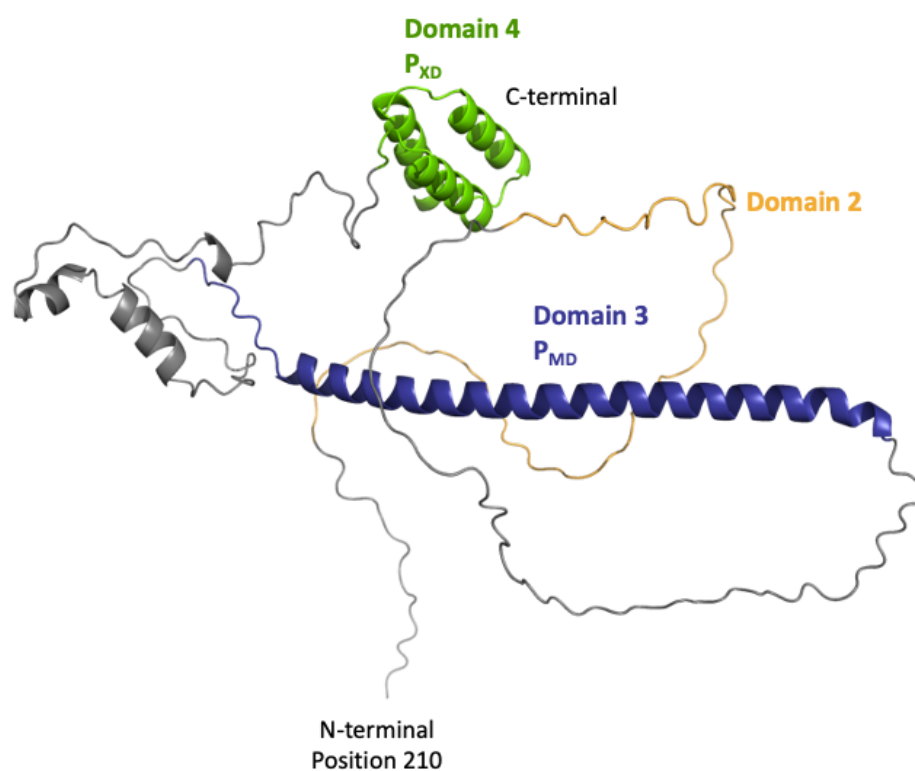
## References

- [1] C. A. Tidona, H. W. Kurz, H. R. Gelderblom, et G. Darai, « Isolation and Molecular Characterization of a Novel Cytopathogenic Paramyxovirus from Tree Shrews », *Virology*, vol. 258, n° 2, p. 425-434, juin 1999, doi: 10.1006/viro.1999.9693.
- [2] R. Abdella, M. Aggarwal, T. Okura, R. A. Lamb, et Y. He, « Structure of a paramyxovirus polymerase complex reveals a unique methyltransferase-CTD conformation », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 117, n° 9, p. 4931-4941, mars 2020, doi: 10.1073/pnas.1919837117.
- [3] F. Yabukarski *et al.*, « Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone », *Nat Struct Mol Biol*, vol. 21, n° 9, p. 754-759, sept. 2014, doi: 10.1038/nsmb.2868.
- [4] M. Jamin et F. Yabukarski, « Chapter Four - Nonsegmented Negative-Sense RNA Viruses—Structural Data Bring New Insights Into Nucleocapsid Assembly », M. Kielian, T. C. Mettenleiter, et M. J. Roossinck, Éd., in *Advances in Virus Research*, vol. 97. Academic Press, 2017, p. 143-185. doi: <https://doi.org/10.1016/bs.aivir.2016.09.001>.
- [5] F. C. A. Gerard *et al.*, « Modular Organization of Rabies Virus Phosphoprotein », *Journal of Molecular Biology*, vol. 388, n° 5, p. 978-996, mai 2009, doi: 10.1016/j.jmb.2009.03.061.
- [6] M. R. Jensen *et al.*, « Structural Description of the Nipah Virus Phosphoprotein and Its Interaction with STAT1 », *Biophysical Journal*, vol. 118, n° 10, p. 2470-2488, mai 2020, doi: 10.1016/j.bpj.2020.04.010.
- [7] A. J. McCarthy et S. J. Goodman, « Reassessing conflicting evolutionary histories of the Paramyxoviridae and the origins of respiroviruses with Bayesian multigene phylogenies », *Infection, Genetics and Evolution*, vol. 10, n° 1, p. 97-107, janv. 2010, doi: 10.1016/j.meegid.2009.11.002.
- [8] D. Karlin et R. Belshaw, « Detecting Remote Sequence Homology in Disordered Proteins: Discovery of Conserved Motifs in the N-Termini of Mononegavirales phosphoproteins », *PLoS ONE*, vol. 7, n° 3, p. e31719, mars 2012, doi: 10.1371/journal.pone.0031719.
- [9] J.-M. Bourhis *et al.*, « Structural Dynamics of the C-terminal X Domain of Nipah and Hendra Viruses Controls the Attachment to the C-terminal Tail of the Nucleocapsid Protein », *Journal of Molecular Biology*, vol. 434, n° 10, p. 167551, mai 2022, doi: 10.1016/j.jmb.2022.167551.
- [10] J. F. Bruhn *et al.*, « Crystal Structure of the Nipah Virus Phosphoprotein Tetramerization Domain », *J Virol*, vol. 88, n° 1, p. 758-762, janv. 2014, doi: 10.1128/JVI.02294-13.
- [11] D. Karlin, F. Ferron, B. Canard, et S. Longhi, « Structural disorder and modular organization in Paramyxovirinae N and P », *Journal of General Virology*, vol. 84, n° 12, p. 3239-3252, déc. 2003, doi: 10.1099/vir.0.19451-0.
- [12] L. Blanchard *et al.*, « Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution », *Virology*, vol. 319, n° 2, p. 201-211, févr. 2004, doi: 10.1016/j.virol.2003.10.029.
- [13] K. Johansson, J.-M. Bourhis, V. Campanacci, C. Cambillau, B. Canard, et S. Longhi, « Crystal Structure of the Measles Virus Phosphoprotein Domain Responsible for the Induced Folding of the C-terminal Domain of the Nucleoprotein », *Journal of Biological Chemistry*, vol. 278, n° 45, p. 44567-44573, nov. 2003, doi: 10.1074/jbc.M308745200.
- [14] G. Communie *et al.*, « Atomic Resolution Description of the Interaction between the Nucleoprotein and Phosphoprotein of Hendra Virus », *PLoS Pathog*, vol. 9, n° 9, p. e1003631, sept. 2013, doi: 10.1371/journal.ppat.1003631.
- [15] D. E. Holmes et S. A. Moyer, « The Phosphoprotein (P) Binding Site Resides in the N Terminus of the L Polymerase Subunit of Sendai Virus », *Journal of Virology*, vol. 76, n° 6, p. 3078-3083, mars 2002, doi: 10.1128/JVI.76.6.3078-3083.2002.
- [16] N. Tarbouriech, J. Curran, R. W. H. Ruigrok, et W. P. Burmeister, « Tetrameric coiled coil domain of Sendai virus phosphoprotein », 2000.
- [17] D. Blocquel *et al.*, « Coiled-coil deformations in crystal structures: the measles virus phosphoprotein multimerization domain as an illustrative example », *Acta Crystallogr D Biol Crystallogr*, vol. 70, n° 6, p. 1589-1603, juin 2014, doi: 10.1107/S139900471400234X.

- [18] A. Roy *et al.*, « Orthoparamyxovirinae C Proteins Have a Common Origin and a Common Structural Organization », *Biomolecules*, vol. 13, n° 3, p. 455, mars 2023, doi: 10.3390/biom13030455. 860  
861
- [19] J. Prilusky *et al.*, « FoldIndex(C): a simple tool to predict whether a given protein sequence is intrinsically unfolded », *Bioinformatics*, vol. 21, n° 16, p. 3435-3438, août 2005, doi: 10.1093/bioinformatics/bti537. 862  
863
- [20] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, et A. K. Dunker, « Exploiting heterogeneous sequence properties improves prediction of protein disorder », *Proteins*, vol. 61, n° S7, p. 176-182, 2005, doi: 10.1002/prot.20735. 864  
865
- [21] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, et A. K. Dunker, « Predicting intrinsic disorder from amino acid sequence », *Proteins*, vol. 53, n° S6, p. 566-572, 2003, doi: 10.1002/prot.10532. 866  
867
- [22] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, et A. K. Dunker, « Sequence complexity of disordered protein », p. 11, 2001. 868  
869
- [23] B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, et V. N. Uversky, « PONDR-FIT: A meta-predictor of intrinsically disordered amino acids », *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1804, n° 4, p. 996-1010, avr. 2010, doi: 10.1016/j.bbapap.2010.01.011. 870  
871  
872
- [24] Z. Dosztanyi, V. Csizmok, P. Tompa, et I. Simon, « IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content », *Bioinformatics*, vol. 21, n° 16, p. 3433-3434, août 2005, doi: 10.1093/bioinformatics/bti541. 873  
874  
875
- [25] R. Linding, L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, et R. B. Russell, « Protein Disorder Prediction », *Structure*, vol. 11, n° 11, p. 1453-1459, nov. 2003, doi: 10.1016/j.str.2003.10.002. 876  
877
- [26] R. Linding, « GlobPlot: exploring protein sequences for globularity and disorder », *Nucleic Acids Research*, vol. 31, n° 13, p. 3701-3708, juill. 2003, doi: 10.1093/nar/gkg519. 878  
879
- [27] I. Walsh, A. J. M. Martin, T. Di Domenico, A. Vullo, G. Pollastri, et S. C. E. Tosatto, « CSpritz: accurate prediction of protein disorder segments with annotation for homology, secondary structure and linear motifs », *Nucleic Acids Research*, vol. 39, n° suppl\_2, p. W190-W196, juill. 2011, doi: 10.1093/nar/gkr411. 880  
881  
882
- [28] I. Callebaut *et al.*, « Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives », p. 25, 1997. 883  
884
- [29] L. J. McGuffin, K. Bryson, et D. T. Jones, « The PSIPRED protein structure prediction server », p. 2, 1999. 885
- [30] W. Kabsch, « XDS », *Acta Crystallogr D Biol Crystallogr*, vol. 66, n° 2, p. 125-132, févr. 2010, doi: 10.1107/S0907444909047337. 886  
887
- [31] I. Tickle *et al.*, « STARANISO », 2018. <http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi> 888
- [32] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, et M. Steinegger, « ColabFold: making protein folding accessible to all », *Nat Methods*, vol. 19, n° 6, p. 679-682, juin 2022, doi: 10.1038/s41592-022-01488-1. 889  
890
- [33] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, et R. J. Read, « Phaser crystallographic software », *J Appl Crystallogr*, vol. 40, n° 4, p. 658-674, août 2007, doi: 10.1107/S0021889807021206. 891  
892
- [34] E. Potterton, P. Briggs, M. Turkenburg, et E. Dodson, « A graphical user interface to the CCP 4 program suite », *Acta Crystallogr D Biol Crystallogr*, vol. 59, n° 7, p. 1131-1137, juill. 2003, doi: 10.1107/S0907444903008126. 893  
894
- [35] M. D. Winn *et al.*, « Overview of the CCP 4 suite and current developments », *Acta Crystallogr D Biol Crystallogr*, vol. 67, n° 4, p. 235-242, avr. 2011, doi: 10.1107/S0907444910045749. 895  
896
- [36] P. Emsley, B. Lohkamp, W. G. Scott, et K. Cowtan, « Features and development of Coot », *Acta Crystallogr D Biol Crystallogr*, vol. 66, n° 4, p. 486-501, avr. 2010, doi: 10.1107/S0907444910007493. 897  
898
- [37] A. Thureau, P. Roblin, et J. Pérez, « BioSAXS on the SWING beamline at Synchrotron SOLEIL », *J Appl Crystallogr*, vol. 54, n° 6, p. 1698-1710, déc. 2021, doi: 10.1107/S1600576721008736. 899  
900
- [38] K. Manalastas-Cantos *et al.*, « ATASAS 3.0: expanded functionality and new tools for small-angle scattering data analysis », 901

- J Appl Crystallogr*, vol. 54, n° 1, p. 343-355, févr. 2021, doi: 10.1107/S1600576720013412. 902
- [39] J. B. Hopkins, R. E. Gillilan, et S. Skou, « *BioXTAS RAW*: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis », *J Appl Crystallogr*, vol. 50, n° 5, p. 1545-1553, oct. 2017, doi: 10.1107/S1600576717011438. 903-905
- [40] D. I. Svergun, « Determination of the regularization parameter in indirect-transform methods using perceptual criteria », *J Appl Crystallogr*, vol. 25, n° 4, p. 495-503, août 1992, doi: 10.1107/S0021889892001663. 906-907
- [41] S. Hansen, « Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering », *J Appl Crystallogr*, vol. 33, n° 6, p. 1415-1421, déc. 2000, doi: 10.1107/S0021889800012930. 908-909
- [42] R. P. Rambo et J. A. Tainer, « Accurate assessment of mass, models and resolution by small-angle scattering », *Nature*, vol. 496, n° 7446, p. 477-481, avr. 2013, doi: 10.1038/nature12070. 910-911
- [43] P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, et D. I. Svergun, « Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering », *J. Am. Chem. Soc.*, vol. 129, n° 17, p. 5656-5664, mai 2007, doi: 10.1021/ja069124n. 912-914
- [44] T. D. Grant, « AB Initio Electron Density Determination Directly from Solution Scattering Data », *Biophysical Journal*, vol. 118, n° 3, p. 487a, févr. 2020, doi: 10.1016/j.bpj.2019.11.2693. 915-916
- [45] J. Jumper *et al.*, « Highly accurate protein structure prediction with AlphaFold », *Nature*, vol. 596, n° 7873, p. 583-589, août 2021, doi: 10.1038/s41586-021-03819-2. 917-918
- [46] M. D. Winn, G. N. Murshudov, et M. Z. Papiz, « Macromolecular TLS Refinement in REFMAC at Moderate Resolutions », in *Methods in Enzymology*, Elsevier, 2003, p. 300-321. doi: 10.1016/S0076-6879(03)74014-2. 919-920
- [47] E. Krissinel et K. Henrick, « Inference of Macromolecular Assemblies from Crystalline State », *Journal of Molecular Biology*, vol. 372, n° 3, p. 774-797, sept. 2007, doi: 10.1016/j.jmb.2007.05.022. 921-922
- [48] P. Kumar et D. N. Woolfson, « *Socket2*: a program for locating, visualizing and analyzing coiled-coil interfaces in protein structures », *Bioinformatics*, vol. 37, n° 23, p. 4575-4577, déc. 2021, doi: 10.1093/bioinformatics/btab631. 923-924
- [49] J. Walshaw et D. N. Woolfson, « SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures » Edited by J. Thornton », *Journal of Molecular Biology*, vol. 307, n° 5, p. 1427-1450, avr. 2001, doi: 10.1006/jmbi.2001.4545. 925-927
- [50] S. V. Strelkov et P. Burkhard, « Analysis of  $\alpha$ -Helical Coiled Coils with the Program TWISTER Reveals a Structural Mechanism for Stutter Compensation », *Journal of Structural Biology*, vol. 137, n° 1-2, p. 54-64, janv. 2002, doi: 10.1006/jsbi.2002.4454. 928-930
- [51] L. Truebestein et T. A. Leonard, « Coiled-coils: The long and short of it », *BioEssays*, vol. 38, n° 9, p. 903-916, sept. 2016, doi: 10.1002/bies.201600062. 931-932
- [52] G. Grigoryan et W. F. DeGrado, « Probing Designability via a Generalized Model of Helical Bundle Geometry », *Journal of Molecular Biology*, vol. 405, n° 4, p. 1079-1100, janv. 2011, doi: 10.1016/j.jmb.2010.08.058. 933-934
- [53] J. F. Bruhn, A. L. Hotard, C. F. Spiropoulou, M. K. Lo, et E. O. Saphire, « A Conserved Basic Patch and Central Kink in the Nipah Virus Phosphoprotein Multimerization Domain Are Essential for Polymerase Function », *Structure*, vol. 27, n° 4, p. 660-668.e4, avr. 2019, doi: 10.1016/j.str.2019.01.012. 935-937
- [54] R. M. Cox et R. K. Plemper, « Structure and organization of paramyxovirus particles », *Current Opinion in Virology*, vol. 24, p. 105-114, juin 2017, doi: 10.1016/j.coviro.2017.05.004. 938-939
- [55] L. Holm, « Dali server: structural unification of protein families », *Nucleic Acids Research*, vol. 50, n° W1, p. W210-W215, juill. 2022, doi: 10.1093/nar/gkac387. 940-941
- [56] X. Robert et P. Gouet, « Deciphering key features in protein structures with the new ENDscript server », *Nucleic Acids Research*, vol. 42, n° W1, p. W320-W324, juill. 2014, doi: 10.1093/nar/gku316. 942-943

- [57] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, et N. A. Baker, « PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations », *Nucleic Acids Research*, vol. 32, n° Web Server, p. W665-W667, juill. 2004, doi: 10.1093/nar/gkh381.
- [58] E. Jurrus *et al.*, « Improvements to the APBS biomolecular solvation software suite », *Protein Science*, vol. 27, n° 1, p. 112-128, janv. 2018, doi: 10.1002/pro.3280.
- [59] M. C. Bowman, S. Smallwood, et S. A. Moyer, « Dissection of Individual Functions of the Sendai Virus Phosphoprotein in Transcription », *J Virol*, vol. 73, n° 8, p. 6474-6483, août 1999, doi: 10.1128/JVI.73.8.6474-6483.1999.
- [60] D. Svergun, C. Barberato, et M. H. J. Koch, « CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates », *J Appl Crystallogr*, vol. 28, n° 6, p. 768-773, déc. 1995, doi: 10.1107/S0021889895007047.
- [61] G. Tria, H. D. T. Mertens, M. Kachala, et D. I. Svergun, « Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering », *IUCrJ*, vol. 2, n° 2, p. 207-217, mars 2015, doi: 10.1107/S205225251500202X.
- [62] M. Ito, M. Iwasaki, M. Takeda, T. Nakamura, Y. Yanagi, et S. Ohno, « Measles Virus Nonstructural C Protein Modulates Viral RNA Polymerase Activity by Interacting with Host Protein SHCBP1 », *Journal of Virology*, vol. 87, n° 17, p. 9633-9642, sept. 2013, doi: 10.1128/JVI.00714-13.
- [63] K. Sleeman, B. Bankamp, K. B. Hummel, M. K. Lo, W. J. Bellini, et P. A. Rota, « The C, V and W proteins of Nipah virus inhibit minigenome replication », *Journal of General Virology*, vol. 89, n° 5, p. 1300-1308, mai 2008, doi: 10.1099/vir.0.83582-0.
- [64] A. Roy *et al.*, « Orthoparamyxovirinae C Proteins Have a Common Origin and a Common Structural Organization », *LIFE SCIENCES*, preprint, janv. 2023. doi: 10.20944/preprints202301.0217.v1.
- [65] A. Meignié *et al.*, « Proteomic Analysis Uncovers Measles Virus Protein C Interaction With p65–iASPP Protein Complex », *Molecular & Cellular Proteomics*, vol. 20, p. 100049, 2021, doi: 10.1016/j.mcpro.2021.100049.
- [66] S. Guseva *et al.*, « Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly », *Sci. Adv.*, vol. 6, n° 14, p. eaaz7095, avr. 2020, doi: 10.1126/sciadv.aaz7095.



970

971

**Supplementary 1. AlphaFold2 modelization of the N-terminal region of TupV phosphoprotein.**

Residues 210 to 527 of TupV phosphoprotein were modeled using AlphaFold 2. We analyzed the structure of domain 1 to 3 identified by secondary structure prediction and disorder analysis. Putative P<sub>MD</sub> domain 3 is coloured in blue. The unidentified domain 2 is shown in orange. The putative P<sub>XD</sub> domain is coloured in green. The domain 2 is predicted disordered in contradiction with disordered result which calculate a Dscore > 0.5.

972

973

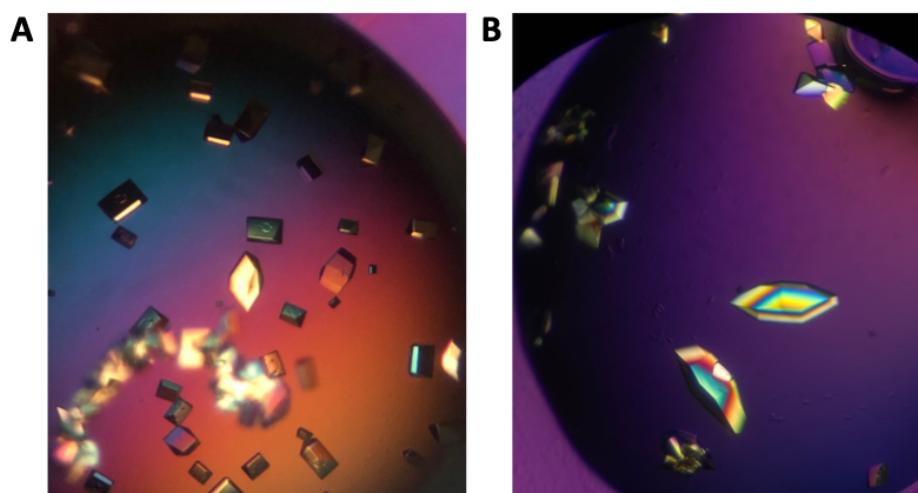
974

975

976

977

978



979

**Supplementary 2. P<sub>MD</sub> crystals obtained by the hanging drop method.** Two characteristic crystal shapes are found in 0.1M phosphate/citrate buffer at pH 5.2, containing 5% PEG 1000 (v/v), 20% EtOH and 3% 2-methyl-2,4-pentanediol: (A) an oblong triangular prism shape and (B) a hexagonal based pyramid shape.

980

981

982

983

984

### III. Conclusions

Au travers de l'étude bio-informatique de la structure de la phosphoprotéine (P) du virus de la toupe, nous avons mis en évidence la conservation, pour le genre *Narmovirus*, d'une organisation modulaire commune avec les phosphoprotéines des *Paramyxoviridae* avec, d'une part, une région N-terminale majoritairement désordonnée à l'exception d'un domaine N-terminal structuré (résidus 1 à 34) et, d'autre part, une région C-terminale composée de deux domaines structurés repliés en hélices  $\alpha$  : le domaine d'oligomérisation, P<sub>MD</sub>, (résidus 336 à 403) et le domaine X, P<sub>XD</sub> (situé entre les résidus 473 à 527). Nous avons également déterminé la présence d'un motif K(K/R)GI conservé au sein du genre localisé entre les résidus 227 et 284 pour lesquelles aucune structure secondaire n'est prédite malgré un Dscore supérieur à 0.5. La fonction de ce domaine est inconnue mais, comme précédemment observé pour la phosphoprotéine du virus Nipah, il pourrait correspondre à un motif d'interaction avec des protéines cellulaires (Jensen *et al.*, 2020).

Nous avons confirmé que la P du TupV forme, comme les autres membres de la famille, des tétramères en solution grâce son domaine d'oligomérisation. Nous avons résolu la structure cristallographique de ce domaine à une résolution de 2.6 Å par remplacement moléculaire sur la base d'un modèle généré par AlphaFold2. Cette structure a été comparée avec les structures précédemment résolues des domaines d'oligomérisation de paramyxovirus et nous avons observé de nombreuses similarités avec le domaine P<sub>MD</sub> du NiV et du MeV. De manière surprenante, l'ajout d'une nouvelle structure aux domaines d'oligomérisation des phosphoprotéines précédemment résolues semble renforcer la position marginale de la structure P<sub>MD</sub> de SeV (Tarbouriech *et al.*, 2000) allant à l'inverse de nos résultats sur la similarité des structures des protéines C du SeV et du TupV.

Nous souhaitons prolonger cette étude par la résolution de la structure cristallographique du domaine P<sub>XD</sub> pour lequel nous avons identifié des conditions initiales de cristallisation qui devraient nous permettre d'obtenir prochainement des cristaux de bonne qualité. Nous avons, de plus, généré un modèle tridimensionnel par AlphaFold2 qui semble décrire avec précision la structure de la protéine en solution et qui servira de modèle pour la résolution de la structure cristallographique par remplacement moléculaire.

Enfin, nos résultats préliminaires sur l'étude d'un complexe entre les protéines C et P du TupV et le facteur cellulaire SHCBP1 pour comprendre le mécanisme d'inhibition de la réplication du minigénome du NiV observé dans le chapitre précédent n'ont pas été concluants. Dans le cas du MeV, cette interaction a été démontré en cellule (Ito *et al.*, 2013). Or, il est aujourd'hui reconnu que les phosphoprotéines des paramyxovirus sont capables d'induire la formation de « liquid-liquide phase separation » (LLPS) en cellule dont la formation joue un rôle essentiel pour la multiplication virale (Guseva *et al.*, 2020). Dans ce contexte, il est possible que l'interaction ne puisse se produire que dans certaines conditions favorisant la formation de ces structures.

**En parallèle des essais de caractérisation structurale de la protéine C du virus de la Toupaye, nous avons cherché à déterminer la structure d'une protéine C du groupe I identifié par l'étude bio-informatique précédente, la protéine C du virus Nipah. Nos résultats sont présentés dans le chapitre 6.**





# Chapitre 6 : Étude biophysique et structurale de la protéine C du virus Nipah

## I. Stratégies élaborées pour étudier cette protéine difficile

L'importance de la protéine C dans la pathogénicité, la réplication virale et l'antagonisme du système immunitaire a été démontrée à plusieurs reprises. Cependant, les mécanismes moléculaires sous-jacents des fonctions biologiques de la protéine C du NiV restent largement inconnus et la protéine C résiste depuis près de 20 ans à la caractérisation structurale en grande part limitée par l'incapacité à produire des protéines recombinantes purifiées en grande quantité. Dans le cadre de ce projet, nous avons mis en place un certain nombre de stratégies pour résoudre ce problème.

### *(A) Approche préliminaire et analyse bio-informatique*

Dans un premier temps, notre travail sur la protéine C pleine longueur ( $C_{FL}$ ) nous a permis de constater l'insolubilité et l'instabilité de la protéine C du NiV. Nous avons donc développé différentes approches pour essayer de résoudre ces problèmes.

Dans le but de définir des constructions de fragments solubles et structurés, nous avons réalisé l'étude bio-informatique de la protéine C du NiV :

- Nous avons déterminé la localisation des régions intrinsèquement désordonnées (IDR) par le méthode du Dscore.
- Nous avons obtenu des prédictions de structure secondaire avec PSIPRED
- Nous avons réalisé une analyse de la présence de groupements hydrophobes en utilisant le graphe HCA (Hydrophobic Cluster Analysis).

De ces résultats, nous avons construit deux fragments :

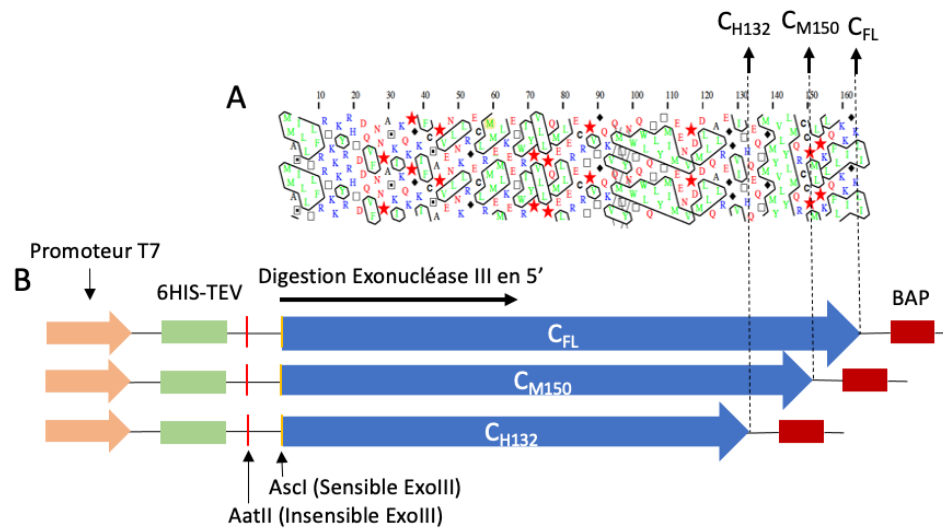
- Un fragment incluant les résidus 44 à 166 formant le domaine C-terminal structuré prédit de la protéine C pleine longueur, la protéine  $C_{\Delta 45}$ .
- Un fragment incluant les résidus 90 à 166, fragment pour lequel nous avons retiré en plus un cluster hydrophobe qui pourrait être à l'origine de l'insolubilité de la protéine pleine longueur, la protéine  $C_{\Delta 91}$ .

### *(B) Approche de mutagenèse haut-débit*

Nous avons ensuite appliqué une méthode de criblage haut débit, la méthode ESPRIT, permettant la sélection de clones solubles par troncation aléatoire en 5' des gènes codant pour trois constructions de la protéine C du NiV.

Ces trois constructions ont été définies sur la base de l'analyse HCA de la séquence en aa de la protéine C pleine longueur qui révèle la présence de deux larges clusters hydrophobes C-terminaux (**Figure 44A**). Ces constructions sont (**Figure 44B**) :

- (i) La protéine pleine longueur  $C_{FL}$
- (ii)  $C_{M150}$  : retrait de la dernière hélice C-terminale.
- (iii)  $C_{H132}$  : retrait des deux hélices C-terminales



**Figure 44 : Représentation schématique des constructions  $C_{FL}$ ,  $C_{H132}$  et  $C_{M150}$ .** (A) Analyse HCA ayant permis de définir les constructions initiales pour la méthode ESPRIT :  $C_{FL}$ ,  $C_{H132}$  et  $C_{M150}$  de la protéine C de NiV. (B) Représentation du clonage des ADN plasmidiques codant pour les constructions  $C_{FL}$ ,  $C_{H132}$  et  $C_{M150}$  de la méthode ESPRIT. Les gènes sont clonés en aval d'un promoteur T7 entre une étiquette hexahistidine (6HIS) en fusion à un site de clivage à la protéase TEV en N-terminal et une protéine liant la biotine (BAP) en C-terminal. La mutagénèse dirigée par l'exonucléase III est permise grâce à deux sites de restriction en 5' du gène d'intérêt : AatII qui produit un site résistant à l'exonucléase III qui permet la conservation de l'étiquette 6HIS après ligation des ADN digérés et AscI qui produit un site sensible à l'exonucléase III. La digestion par l'exonucléase III au fil du temps permet de produire un ensemble de constructions tronquées en N-terminal (Figure adaptée de An et al., 2011).

À la suite de cette méthode, nous avons travaillé sur 4 constructions tronquées aléatoirement en N-terminal de la protéine, représentatives des trois constructions C-terminales initiales :  $C_{\Delta 28}$ ,  $C_{\Delta 23}$ ,  $C_{\Delta 12}$  et  $C_{\Delta 35}$  (constructions détaillées dans la partie Résultat).

### (C) Reconstitution de complexes pour stabiliser la protéine C du NiV

Nous avons également cherché à stabiliser le fragment  $C_{\Delta 28}$  (résidus 27 à 166), issu de la méthode ESPRIT, par la reconstitution de complexes avec des protéines partenaires cellulaires ou virales identifiées dans la littérature.

Notamment, comme présenté dans le chapitre 2, la protéine C de MeV inhibe la synthèse de l'ARN viral par une interaction avec le facteur cellulaire de croissance SHCBP1, qui lui-même interagit avec la phosphoprotéine de MeV (Ito et al., 2013). L'interaction a été cartographiée sur une région de 12 aa de la protéine C de MeV (résidus 127 à 138, motif -QALMVIAPSQEET-). Cette région présente trois positions bien conservées, dont une strictement (E) (représentées en gras -LREGGVITHQE-) avec le genre *Henipavirus* (Lo, Sogaard and Karlin, 2014). Récemment, il a également été démontré que la protéine C du MeV colocalise avec les LLPS générée par les protéines N, P et L et interagit physiquement

avec les ribonucléocapsides virales *via* la phosphoprotéine suggérant que la protéine C est capable d'interagir avec le complexe de réplication (Pfaller *et al.*, 2019).

Nous avons supposé qu'un mécanisme identique pourrait exister dans le cas de la protéine C du virus Nipah. Nous avons donc étudié l'interaction de la protéine C<sub>Δ28</sub> avec la phosphoprotéine P du virus Nipah et le facteur cellulaire SHCBP1 humain.

Nous avons en parallèle produit des anticorps à domaine unique (SdAb) spécifiques de la protéine C<sub>Δ28</sub> du virus Nipah qui pourraient, en interagissant avec la protéine, aider à la solubilisation et la stabilisation au cours de la purification et de la cristallisation.

#### ***(D) Co-expression en complexes pour solubiliser la protéine C du NiV***

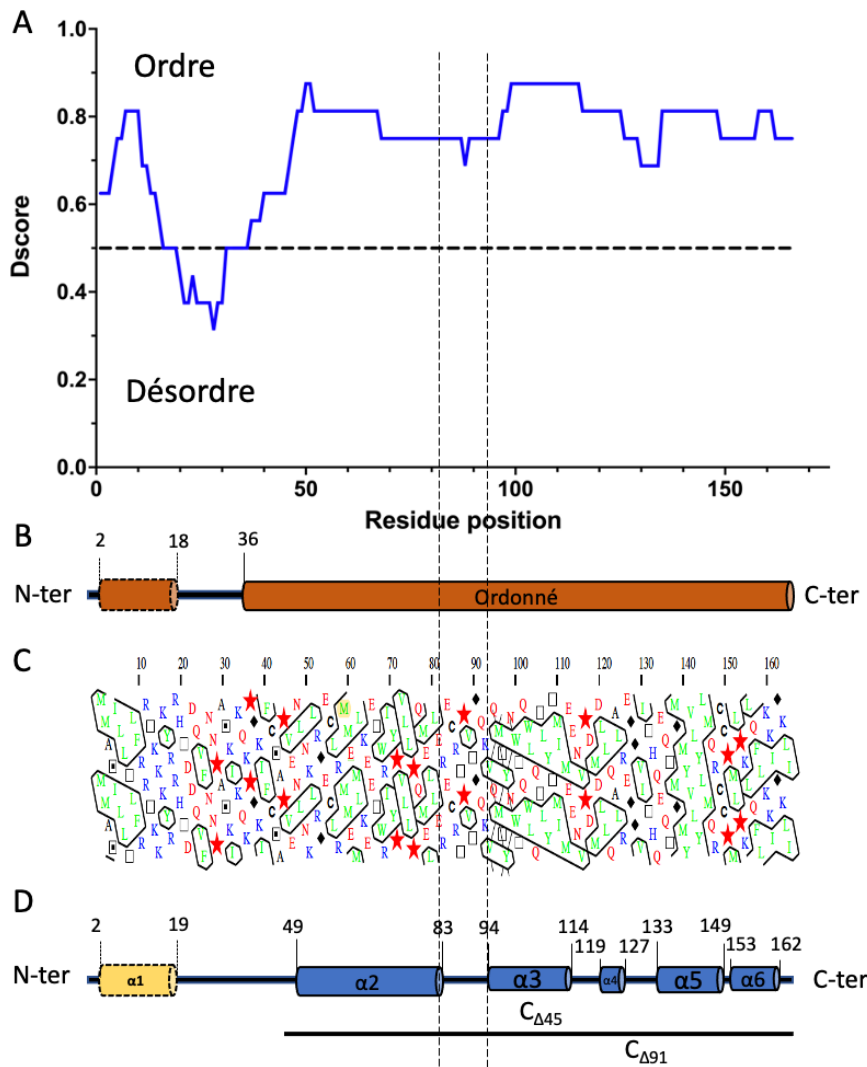
Nous avons également identifié de nouveaux partenaires cellulaires des protéines C<sub>FL</sub> et C<sub>Δ28</sub> par la méthode du double hybride en levure avec l'équipe du Dr. Nicolas Lapaque (INRAE). L'ensemble de l'ORFeome 1.1 humain contenant 8076 ORFs a été criblé.

La protéine C<sub>Δ28</sub> de NiV a par la suite été coexprimée avec l'un des partenaires identifiés.

## **II. Résultats**

#### ***(A) Étude bio-informatique de la protéine C du virus Nipah***

Nous avons effectué, en première approche rationnelle, l'étude bio-informatique de la séquence de la protéine C du virus Nipah (NiV) (numéro d'accèsion Uniprot : Q997F1) grâce à des méthodes de prédiction de structure décrites précédemment (Gerard *et al.*, 2009 ; Jensen *et al.*, 2020 ; Roy *et al.*, 2023) (**Figure 45**).



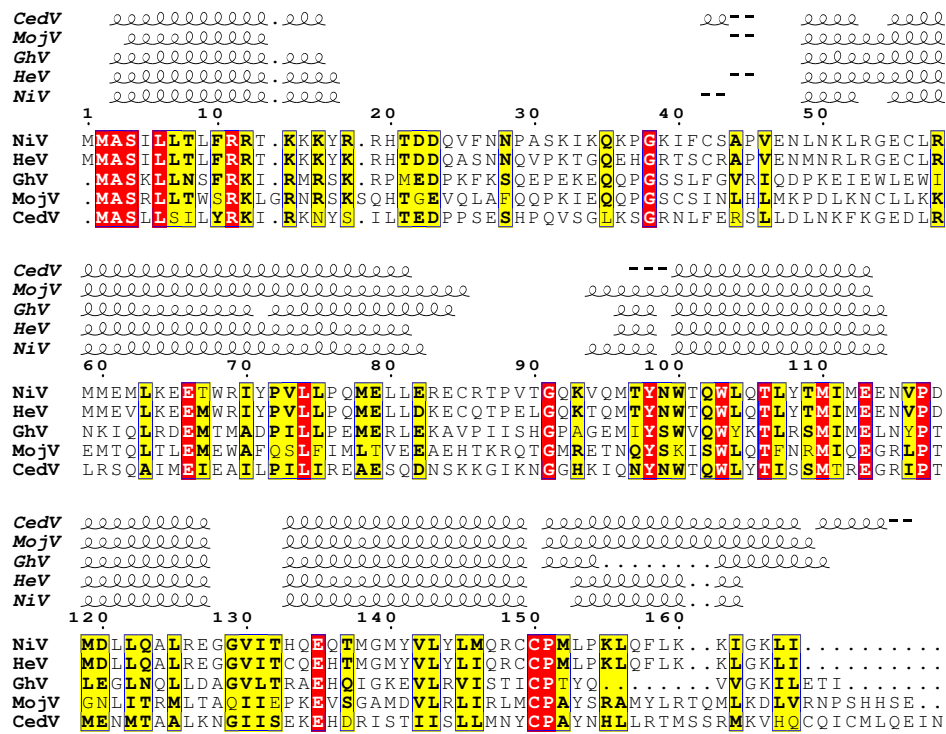
**Figure 45 : Analyse bio-informatique des structures secondaires et de désordre de la protéine C de NiV. (A) Prédiction consensuelle de désordre (Dscore). Le seuil pour distinguer la région ordonnée de la région désordonnée a été arbitrairement fixé à 0,5. (B) Les limites de l'analyse de désordre sont précisées dans la représentation schématique. (C) Tracé HCA. (D) Représentation schématique des prédictions de structures secondaires de PSIPRED de la protéine C de NiV. Les hélices C-terminales sont représentées par des cylindres bleu et le supposé  $\alpha$ -MoRE N-terminal par un cylindre jaune. Les limites du clone  $C_{\Delta 45}$  sont superposées au schéma.**

Ces analyses révèlent que la protéine présente :

- (a) une région N-terminale composée d'un domaine prédit structuré (résidus 1 à 19) suivi d'une partie désordonnée (résidus 20 à 33) (**Figure 45A/B**). Le domaine structuré N-terminal replié présente un regroupement de résidus hydrophobes bien conservés au sein du genre formant une courte hélice- $\alpha$  N-terminale (**Figure 45D** et **Figure 46**), probable  $\alpha$ -MoRE.
- (b) un long domaine C-terminal structuré dont le Dscore est supérieur à 0,5 de manière continue entre les résidus 36 et 166 (**Figure 45A/B**) et qui est composé de l'enchaînement de 5 hélices- $\alpha$  ( $\alpha 2$  à  $\alpha 6$ ) (**Figure 45D**) comme en témoigne la présence de plusieurs clusters de résidus hydrophobes caractéristiques de la formation d'hélices amphipatiques.

(c) une région prédite ordonnée avec des valeurs de Dscore supérieure à 0.5 entre les résidus 83 et 94 dont le profil HCA ne correspond pas et pour laquelle aucune structure secondaire n'est prédite (Figure 45A/C/D).

L'alignement multiple de séquences obtenu pour les protéines C du genre *Henipavirus* (Figure 46) indique une relative conservation de séquence au sein du genre avec notamment 18 positions strictement conservées, majoritairement restreintes à la région C-terminale structurée et au motif  $\alpha$ -MoRE identifié précédemment. Nous avons superposé à cet alignement les prédictions de structures secondaires du serveur PSIPRED pour les différentes protéines C.



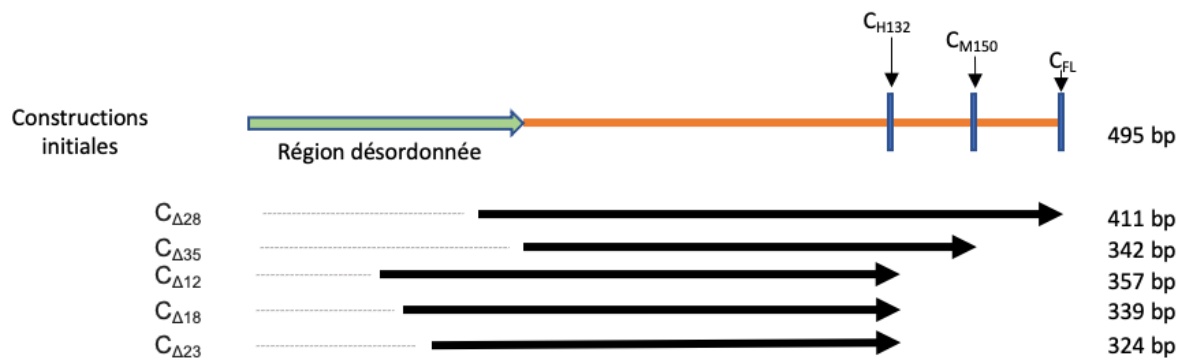
**Figure 46 : Alignement multiple des séquences des protéines C du genre *Henipavirus*. Les numéros d'accèsion Uniprot des séquences sont : le virus Cédar (CedV) - J7H4I1, le virus Mojiang (MojV) - W8SUI, le virus de la chauve-souris du Ghana (GhV) - I0E090, le virus Hendra (HeV) - O55779 et le virus Nipah (NiV) - Q997F1.**

En analysant ces résultats, nous avons choisi d'étudier trois constructions : la forme entière de la protéine C du virus Nipah ( $C_{FL}$ ), le domaine C-terminal structuré composé des résidus 45 à 166 ( $C_{\Delta 45}$ ) et une construction composée des résidus 90 à 166 ( $C_{\Delta 91}$ ) délestée des positions 83 et 94.

**(B) Approche de mutagenèse haut-débit pour déterminer des constructions solubles de la protéine C du virus Nipah : le clone  $C_{\Delta 28}$**

Par l'approche de troncations aléatoire à haut-débit ESPRIT, une librairie de 4500 clones a été générée lors de la première étape de digestion unidirectionnelle en 5'. 95 clones, présentant la meilleure expression et solubilité, ont ensuite été sélectionnés au cours de la

seconde étape. 22 des 95 clones (valeur statistiquement pertinente pour couvrir la diversité présente des 95 clones) ont ensuite été exprimés en format liquide à petite échelle (4 ml), lysés et purifiés.



**Figure 47 : Clones de la protéine C du NiV sélectionnés par la méthode ESPRIT.** Représentation schématique des 6 constructions résultant de l'application de la méthode ESPRIT à la protéine C du NiV après séquençage superposés à l'organisation prédite pour la forme entière de la protéine. La nomenclature utilisée pour les mutants indique la position du premier résidu présent en N-terminal.

**Tableau 1 : MM moyenne des troncations de la protéine C du NiV par la méthode ESPRIT.**

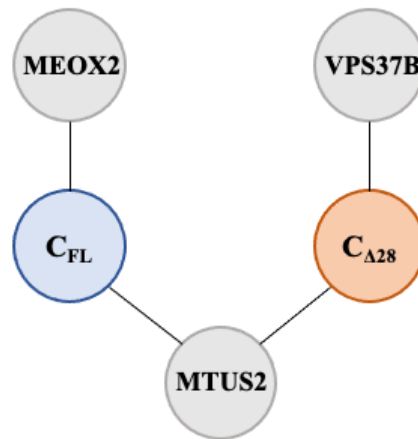
Protéine + 6HIS +BAP	MM (kDa)
$C_{\Delta 28}$	21.3
$C_{\Delta 35}$	18.7
$C_{\Delta 12}$	19.3
$C_{\Delta 18}$	18.5
$C_{\Delta 23}$	18

5 clones parmi les 22 clones ont finalement été choisis en fonction des niveaux d'expression en culture de 4mL observés par SDS-PAGE puis séquencés :  $C_{\Delta 28}$ ,  $C_{\Delta 35}$ ,  $C_{\Delta 12}$ ,  $C_{\Delta 18}$  et  $C_{\Delta 23}$  (Figure 47). Nous avons de cette façon sélectionné au moins un clone de chacune des constructions de départ  $C_{FL}$ ,  $C_{M150}$  (résidus 1 à 150) et  $C_{H132}$  (résidus 1 à 132) où 0, 1 ou 2 hélices  $\alpha$  C-terminales ont été enlevées. Ces clones présentent des portions du domaine N-terminal désordonné variable. Les

masses moléculaires (MM) prédites à partir de la séquence en aa pour les 5 clones sélectionnés sont détaillées dans le **tableau 1**.

### ***(C) Identification de partenaires des protéines $C_{FL}$ et $C_{\Delta 28}$ par double-hybride***

La recherche d'interactants cellulaires par la méthode du double-hybride en levure a permis d'identifier 3 interactants des protéines  $C_{FL}$  et/ou  $C_{\Delta 28}$  (Figure 48) :



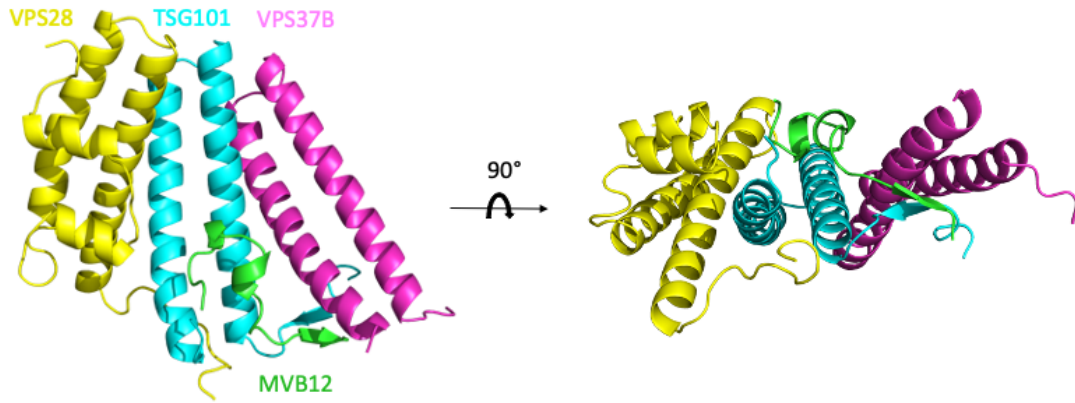
**Figure 48 : Résultats du double-hybride en levure pour C<sub>FL</sub> et C<sub>Δ28</sub>.**

1. **Vacuolar Protein Sorting-Associated Protein 37B (VPS37b)** : une sous-unité du complexe ESCRT-I (Endosomal Sorting Complexes Required for Transport I), trouvée seulement avec C<sub>Δ28</sub>. Ce complexe est un régulateur des processus de trafic vésiculaire souvent impliqué dans le bourgeonnement viral.
2. **Microtubule associated scaffold protein 2 (MTUS2)** : une protéine de liaison aux microtubules impliquée dans la facilitation des mouvements précis des chromosomes et l'orchestration de la dynamique cellulaire et dans la stabilité des chromosomes pendant la mitose, trouvée avec les deux protéines cibles.
3. **Mesenchyme Homeobox (MEOX2)** : un facteur de transcription mésodermique qui joue un rôle clé dans la somitogénèse et la différenciation des muscles des membres, trouvé uniquement avec C<sub>FL</sub>. Cette protéine est également connue pour réguler l'activité du facteur nucléaire NFκB dans les cellules endothéliales vasculaires (Chen *et al.*, 2010). Son expression est également altérée chez les patients souffrant de la maladie d'Alzheimer (Wu *et al.*, 2005).

En comparant ces résultats avec ceux d'un précédent interactome de Nipah (Martinez-Gil, Vera-Velasco and Mingarro, 2017) réalisé par la méthode TAP-MS (Tandem Affinity Purification and Mass Spectrometry), nous avons constaté que VPS37B avait déjà été identifié comme un interactant potentiel de la protéine C<sub>FL</sub> du virus Nipah.

Il a également été suggéré que la protéine C<sub>FL</sub> était capable de recruter la protéine TSG101 du complexe ESCRT-I (Park *et al.*, 2016) afin de promouvoir le bourgeonnement des particules virales de NiV par une voie dépendante des complexes ESCRT. Cependant, aucune interaction directe avec TSG101 n'a été révélée par notre interactome ou les interactomes précédents (Martinez-Gil, Vera-Velasco and Mingarro, 2017).

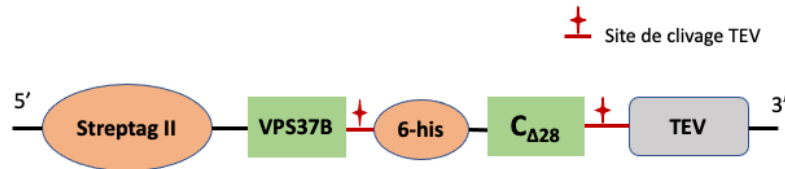
La structure de la région structurée du complexe ESCRT-I humain comprenant les protéines TSG101 (résidus 308 à 188), VPS37B (résidus 97 à 167), VPS28 (résidus 1 à 122) et MVB12A (résidus 206 à 268) a été résolue par cristallographie (Flower *et al.*, 2020) démontrant une interaction entre TSG101 et VPS37B (**Figure 49**).



**Figure 49 : Structure cristalline de la tête globale du complexe ESCRT-I.** Structure en ruban des protéines du complexe ESCRT-I : MVB12A (206-228, vert), TSG101 (308-388, cyan), VPS37B (97-167, rose), VPS28 (1-122, jaune) (Flower et al., 2020).

Nous avons alors émis l'hypothèse que la protéine C de NiV serait capable d'interagir avec le complexe ESCRT-I dans son ensemble au niveau de la zone d'interaction des protéines TSG101 et VPS37B (**Figure 49**). Afin d'étudier cette interaction, deux approches ont été envisagées :

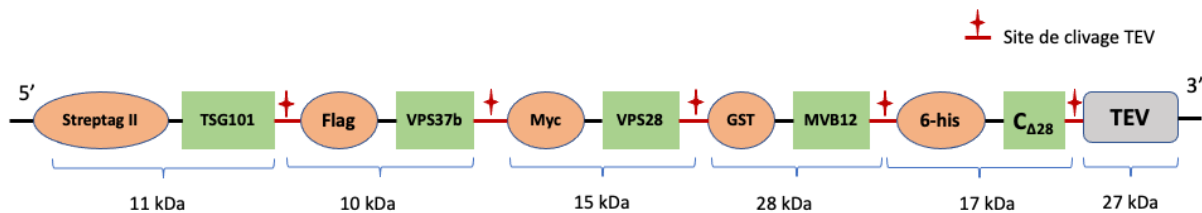
(i) co-exprimer  $C_{\Delta 28}$  avec VPS37B en complexe (**Figure 50**) ( $C_{\Delta 28}$ -VPS37B) :



**Figure 50 : Construction et expression du complexe  $C_{\Delta 28}$ -VPS37B.** Les résidus 97 à 167 de la protéine VPS37B et la protéine  $C_{\Delta 28}$  ont été sous-clonés sous le même promoteur avec, respectivement, des étiquettes streptag II et 6HIS N-terminales. Une séquence codant pour la protéase TEV entre chaque cassette génétique permettant de cliver les différentes protéines du complexe (ligne rouge).

Les protéines et leurs étiquettes respectives sont coexprimées avec la protéase TEV sous le même promoteur. Chaque cassette d'expression est séparée par une séquence de clivage TEV.

(ii) co-exprimer  $C_{\Delta 28}$  avec un complexe multimoléculaire correspondant à la région repliée du complexe ESCRT-I identifiée par Flower et al (Flower et al., 2020) (**Figure 51**) ( $C_{\Delta 28}$ -ESCRT-I) :



**Figure 51 : Construction et expression du complexe TSG101-VPS37B-VPS28-MVB12- $C_{\Delta 28}$ .** TSG101 308-388, VPS37B 97-167, VPS28 1-122, MVB12A 206-268,  $C_{\Delta 28}$  et la protéase TEV ont été sous-clonés séquentiellement dans le vecteur pet28a et ont été marqués, respectivement, avec les étiquettes streptag II, flag, Myc, GST, hexahistidine N-terminales. Une séquence codant pour une protéase TEV a été ajoutée en 5' complétée par une séquence de clivage TEV entre chaque cassette génétique (ligne rouge).



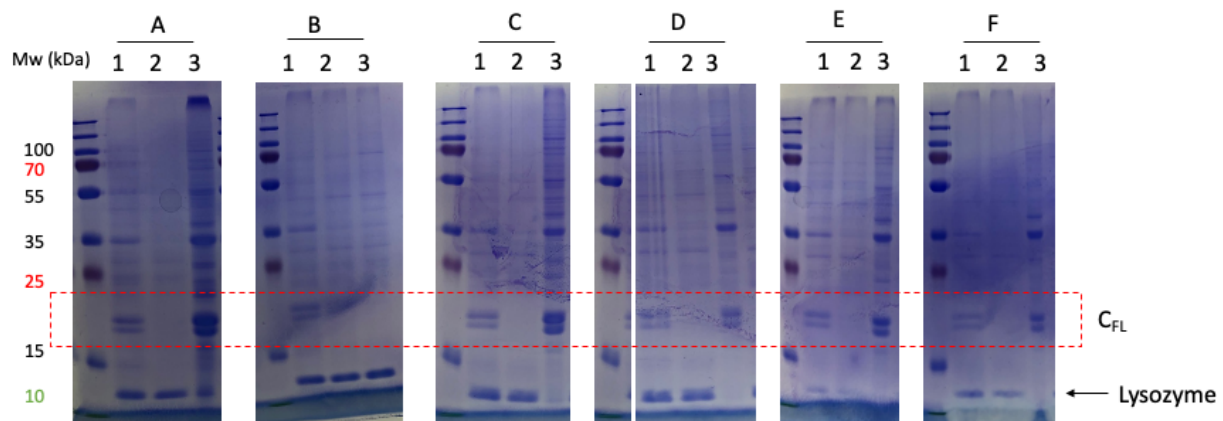
## (D) Expression et purification des protéines C, P et SHCBP1

### (1) Purification de la protéine pleine longueur C<sub>FL</sub>

Les résultats d'une purification de la protéine C<sub>FL</sub> sont présentés dans la **Figure 51**. La masse moléculaire de C<sub>FL</sub> prédite par la séquence est de 19 kDa.

On remarque une bande correspondant à la protéine entre les marqueurs de poids moléculaire de 15 kDa et 25 kDa dans la fraction totale correspondant aux protéines présentes dans le culot cellulaire (**Figure 52A**, échantillon 1) et dans la fraction correspondant au culot après lyse cellulaire et centrifugation (protéines non solubles) (**Figure 52A**, échantillon 3). Aucune bande correspondant à la protéine n'est visible dans le surnageant après lyse cellulaire et centrifugation (protéines solubles) (**Figure 52A**, échantillon 2). Nous avons vérifié par Western Blot et marquage par anticorps anti-6HIS-HRP que cette bande correspondait bien à la protéine C<sub>FL</sub> (données non montrées).

Nous avons cherché à solubiliser la protéine par l'addition, dans le tampon de lyse durant la sonication, de différents détergents : le NDSB 1M (**Figure 52B**),  $\beta$ -octylglucopyraoside ( $\beta$ OG) 0,8% (**Figure 52C**), Triton X-100 0,1% (**Figure 52D**), nonyl-phénoxy polyéthoxyléthanol (NP-40) 1% (**Figure 52E**) et Polyéthylène glycol hexadécyl éther (Brij-58) 0,5% (**Figure 52F**).



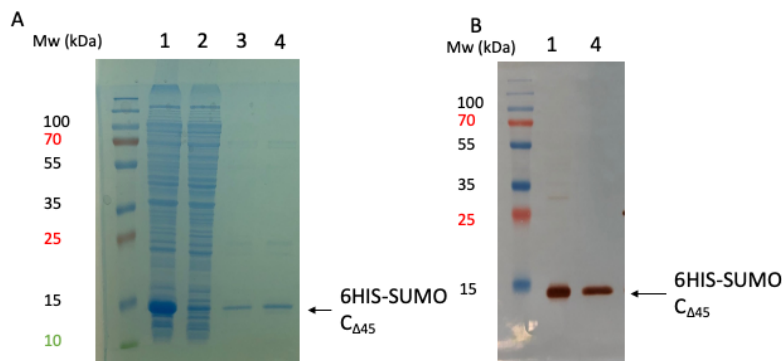
**Figure 52 : Résultats des purifications de la protéine C<sub>FL</sub> de NiV en présence de différents détergents :** (A) sans détergents, (B) NDSB 1M, (C)  $\beta$ OG 0.8%, (D) Triton X-100 0.1%, (E) NP-40 1% et (F) Brij-58 0,5%. Les fractions déposées dans les puits sont (1) Fraction totale, (2) Surnageant (protéines solubles), (3) Culot (protéines non solubles).

Les résultats sont identiques à ceux présentés pour la solubilisation de C<sub>FL</sub> sans détergent et aucun détergent testé ne permet de solubiliser la protéine.

### (2) Purification de la protéine plein longueur C <sub>$\Delta$ 45</sub>

La protéine C <sub>$\Delta$ 45</sub> est solubilisée en présence du détergent Triton X-100 0.1% dans l'ensemble des solutions de purification comme le montrent les résultats de purification présentés dans la **figure 52A/B**. La protéine est bien éluée comme le confirme le Western Blot (WB) révélé par un anticorps anti-His-HRP (**Figure 53A/B**, échantillon 4).

Cependant, nous observons que la protéine  $C_{\Delta 45}$  est instable (précipite lors du clivage de l'étiquette 6HIS-SUMO) et agrégée (élue dans le volume mort en chromatographie d'exclusion de taille).



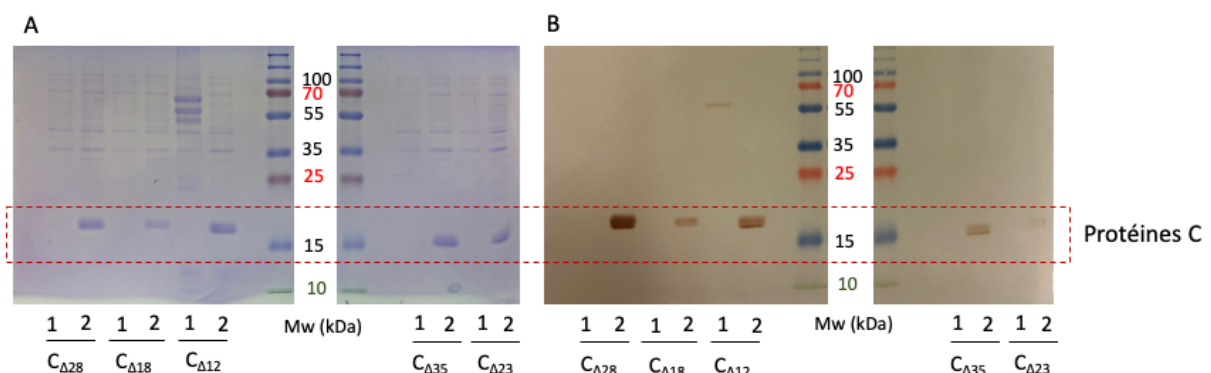
**Figure 53 : Résultats des purifications de la protéine  $C_{\Delta 45}$  de NiV en Triton X-100 0.1% :** (A) SDS-Page 4-15% d'une purification de la protéine  $C_{\Delta 45}$ . Les fractions déposées dans les puits sont (1) Fraction totale, (2) Surnageant (protéines solubles), (3) Flowthrough, (4) Éluion et (B) Western Blot immunomarcqué par des anticorps HRP spécifiques à l'étiquette 6HIS.

### (3) Purification de la protéine plein longueur $C_{\Delta 91}$

Afin de vérifier que l'instabilité et l'insolubilité des deux constructions précédentes ne soient pas liées au groupement de résidus hydrophobes observé entre les résidus 83 à 94, nous avons voulu exprimer la construction  $C_{\Delta 91}$ . Nos premiers essais d'expression ont été infructueux et seront reproduits.

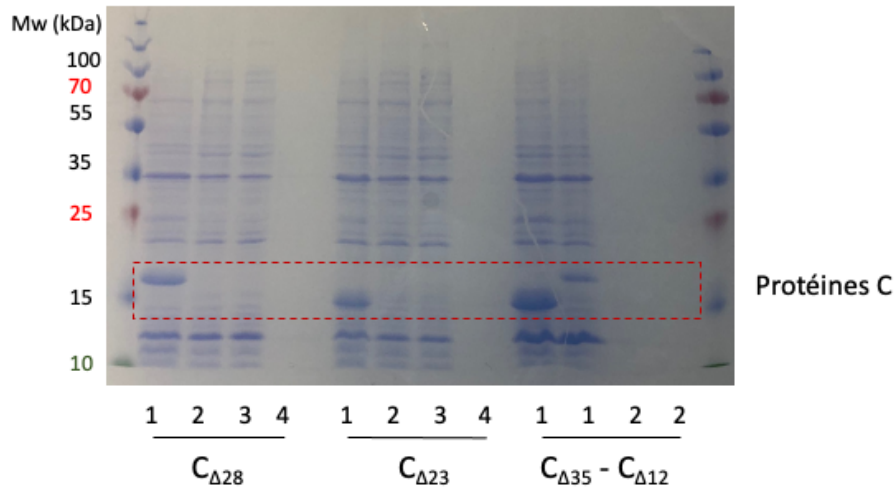
### (4) Purification des clones sélectionnés par la méthode ESPRIT

Les niveaux d'expression après induction en culture bactérienne à grande échelle ont été vérifiés pour les 6 clones de la méthode ESPRIT ( $C_{\Delta 28}$ ,  $C_{\Delta 18}$ ,  $C_{\Delta 12}$ ,  $C_{\Delta 35}$  et  $C_{\Delta 23}$ ) et sont présentés en **Figure 54** (**Figure 54A**, échantillon 2). Nous observons clairement la présence de bandes sur-exprimées entre les marqueurs 15 kDa et 25 kDa révélées par un WB marqué avec un anticorps anti-HIS-HRP (**Figure 54B**, échantillon 2).



**Figure 54 : Test d'expression des mutants  $C_{\Delta 28}$ ,  $C_{\Delta 18}$ ,  $C_{\Delta 12}$ ,  $C_{\Delta 35}$  et  $C_{\Delta 23}$ .** (A) SDS-Page 4-15% et (B) Western Blot correspondant (révélé par des anticorps spécifiques à l'étiquette 6HIS). Les bandes encadrées correspondent aux mutants  $C_{\Delta 28}$ ,  $C_{\Delta 18}$ ,  $C_{\Delta 12}$ ,  $C_{\Delta 35}$  et  $C_{\Delta 23}$  issus de la méthode ESPRIT après expression en culture bactérienne de 1L et induction par addition de 1mM d'IPTG. Les fractions déposées dans les puits correspondent à un échantillon (1) de culture avant induction, (2) de culture après induction.

Les résultats des purifications préliminaires pour les protéines  $C_{\Delta 28}$ ,  $C_{\Delta 23}$ ,  $C_{\Delta 12}$  et  $C_{\Delta 35}$  sont présentés dans la **figure 55**. Nous observons les protéines C dans la fraction totale (**Figure 55**, échantillon 1) mais aucune bande n'est observée dans les étapes suivantes de la purification indiquant que ces protéines ne sont toujours pas solubles en l'absence de détergent (**Figure 55**, échantillon 2,3 et 4).

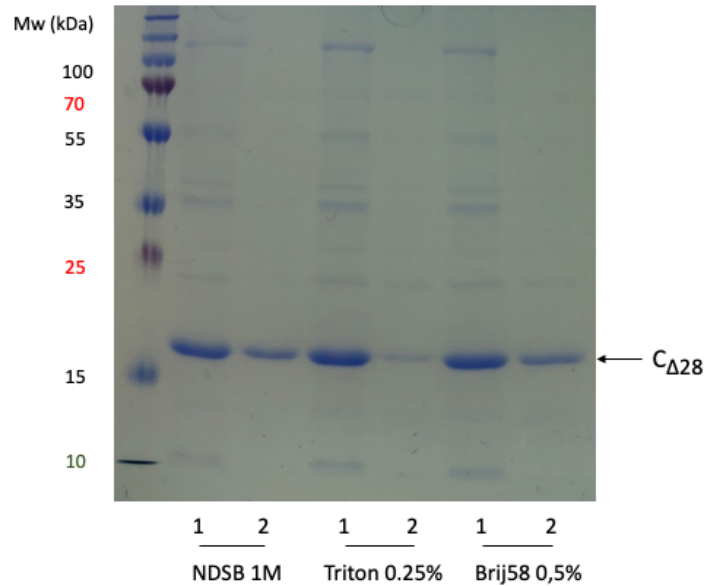


**Figure 55 : Résultats des purification des protéines  $C_{\Delta 28}$ ,  $C_{\Delta 23}$ ,  $C_{\Delta 35}$  et  $C_{\Delta 12}$  de NiV : SDS-Page 4-15% résumant les purifications des protéines  $C_{\Delta 28}$ ,  $C_{\Delta 23}$ ,  $C_{\Delta 35}$  et  $C_{\Delta 12}$  de NiV par chromatographie d'affinité  $Ni^{2+}$ . Les fractions déposées dans les puits sont (1) Fraction totale, (2) Surnageant (protéines solubles), (3) Flowthrough, (4) Éluion**

Nous avons solubilisé les protéines précédentes par l'utilisation de 2 additifs utilisés classiquement pour la purification de protéines insolubles ou membranaires : le NDSB-201 et le Triton X-100. Seul le NDSB est dialysable.

De plus, nous avons également vérifié la solubilité des échantillons après addition de Brij-58, détergent non-ionique et non dialysable utilisé lors du criblage ESPRIT. Nous avons constaté que les trois constructions  $C_{\Delta 28}$ ,  $C_{\Delta 35}$  et  $C_{\Delta 12}$  sont solubilisées par ces 3 additifs. Une fois de plus, l'élimination par dialyse du NDSB entraîne la précipitation des protéines C témoignant à nouveau d'une grande instabilité de la protéine C en solution.

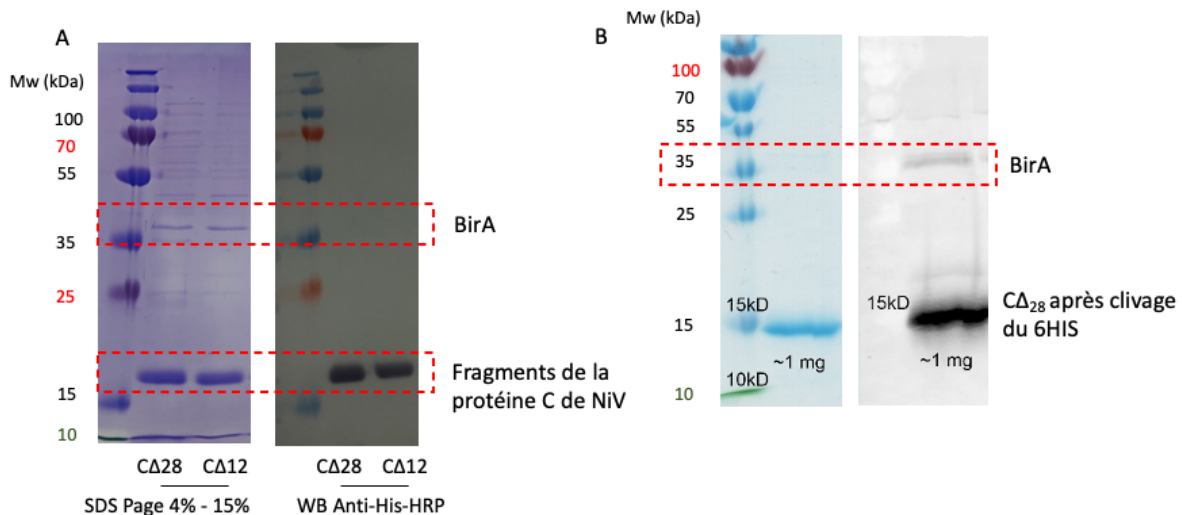
Nous présentons nos résultats dans le cas de la construction  $C_{\Delta 28}$  en **Figure 56**. Nous observons une bande entre les marqueurs 15 kDa et 25 kDa correspondant à la protéine dans la fraction totale après sonication (**Figure 56**, échantillon 1) et dans le surnageant après centrifugation correspondant aux protéines solubles (**Figure 56**, échantillon 2).



**Figure 56 : Solubilisation de la protéine  $C_{\Delta 28}$  de NiV par addition de détergent.** *SDS-Page 4-15% résumant la solubilisation de la protéine  $C_{\Delta 28}$  de NiV par l'addition de NDSB 1M, Triton 0.25 % et de Brij-58 0.5% dans le tampon de lyse durant la sonication. Les fractions déposées dans les puits sont (1) Fraction totale, (2) Surnageant (protéines solubles).  $C_{\Delta 28}$  est solubilisée par l'addition de détergents.*

Nous avons mis au point un protocole de purification permettant l'élimination de la majorité du Brij-58 utilisé pour la solubilisation par lavage sur la résine d'affinité après la fixation des protéines.

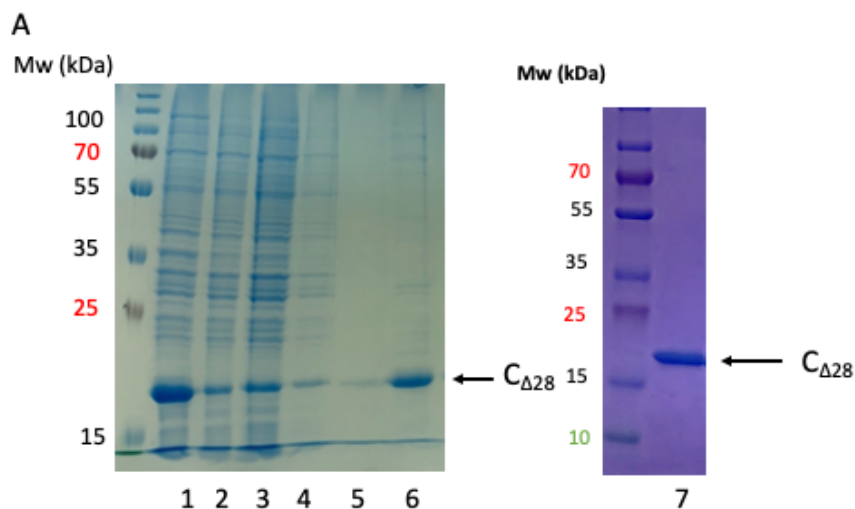
De cette façon, les protéines  $C_{\Delta 28}$  et  $C_{\Delta 12}$  ont été purifiées (**Figure 57A**). Cependant, nous avons constaté la surexpression constitutive des systèmes d'expression bactérien de la protéine BirA (36 kDa) formant un complexe avec l'étiquette BAP C-terminale. BirA catalyse la biotinylation post-traductionnelle, en s'associant à la protéine BAP. On peut constater en exprimant  $C_{\Delta 28}$  en milieu de culture riche en biotine le marquage clair par WB de BirA par un anticorps anti-Streptavidin (se fixe à la Biotine). BirA est indissociable des fragments de protéine NiV C (**Figure 57B**).



**Figure 57 : Présence de la protéine constitutive BirA lors de la solubilisation et purification des protéines  $C_{\Delta 28}$  et  $C_{\Delta 12}$  de NiV :** (A) *SDS-Page 4-15% (gauche) et le Western Blot correspondant*

(droite) révélé par des anticorps spécifiques à l'étiquette 6HIS. Ces analyses présentent les éluions de purification des protéines  $C_{\Delta 28}$  et  $C_{\Delta 12}$  de NiV solubilisées par addition de Brij-58 0.5% dans le tampon de lyse durant la sonication. On constate la présence d'une protéine supplémentaire autour de 35 kDa qui n'est pas révélée par les anticorps anti-6HIS. **(B)** SDS-Page 4-15% (gauche) et le Western Blot correspondant (droite) révélé par des anticorps spécifiques à l'étiquette streptavidine. Ces analyses présentent la protéine  $C_{\Delta 28}$  après clivage de l'étiquette 6HIS concentrée à  $1\text{mg.mL}^{-1}$ . La protéine a été cultivée en milieu enrichi en biotine. Ceci nous a permis de vérifier la contamination de l'échantillon par la protéine BirA, une protéine constitutive des *E. coli* ayant une forte affinité pour la biotine. La bande supplémentaire présente autour de 30 kDa correspond à BirA qui est présente dans l'échantillon purifié.

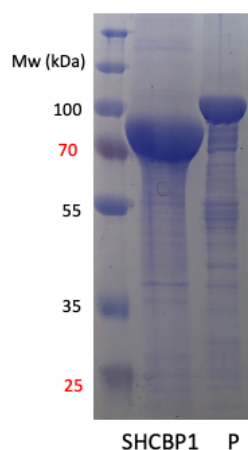
C'est pourquoi, nous avons finalement choisi d'exprimer les protéines C en cellule d'insecte par la génération de baculovirus recombinants. La figure suivante présente les résultats de la purification selon le protocole précédent de la protéine  $C_{\Delta 28}$  produite en cellule d'insecte, Hi5 (**Figure 58**).  $C_{\Delta 28}$  est éluée et concentrée (**Figure 58**, échantillon 7).



**Figure 58 : Purification de la protéine  $C_{\Delta 28}$  exprimée en cellule d'insecte Hi5 en présence de 0.5% de Brij-58 :** SDS-Page 4-15% résumant la purification de la protéine  $C_{\Delta 28}$  de NiV par chromatographie d'affinité  $\text{Ni}^{2+}$ . Les fractions déposées dans les puits sont (1) Fraction totale, (2) Surnageant, (3) Flowthrough, (4) Lavage 1, (5) Lavage 2, (6) Éluion, (7)  $C_{\Delta 28}$  purifiée.

### (5) Purification des protéines P du NiV, SHCBP1

Nous avons exprimé et purifié les protéines SHCBP1 et la protéine P du virus Nipah et le facteur cellulaire SHCBP1 (**Figure 59**).



Du fait de la présence de nombreux acides aminés chargés négativement, la protéine P migre difficilement à travers le gel SDS-Page et présente donc une bande de masse apparente vers 100 kDa (**Figure 59**) supérieure à la MM prédite de 78 kDa depuis la séquence en aa.

**Figure 59 : Purification des protéines P et SHCBP1.** La protéine SHCBP1 (gauche) est visible entre les marqueurs de poids moléculaire 70 et 100 kDa (MM prédite de 76 kDa). La protéine P du NiV (droite) est visible entre les marqueurs de poids moléculaire 100 et 130kDa (MM prédite de 78 kDa).

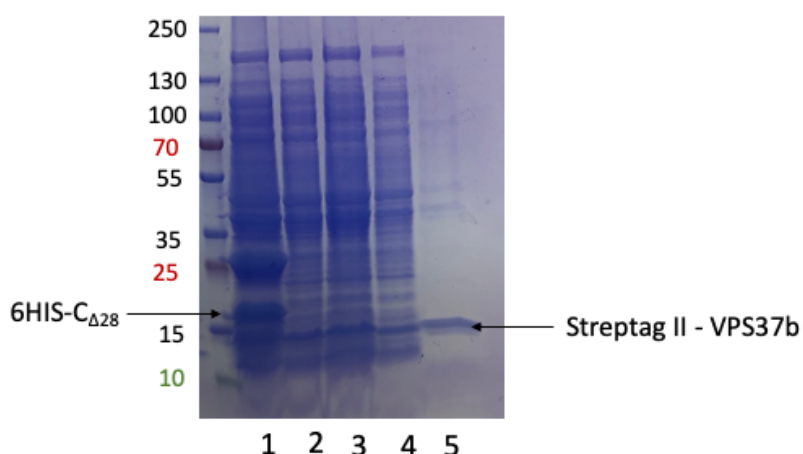
## (6) Purification de la protéine C $\Delta$ 28 coexprimées avec VPS37B ou le complexe ESCRT-I

### • C $\Delta$ 28- VPS37B

Les résultats de l'élution du complexe C $\Delta$ 28/VPS37B par chromatographie d'affinité sur résine streptactine sont présentés en **figure 60**.

Les deux protéines sont bien surexprimées et l'on observe une bande en-dessous de 15 kDa correspondant à la protéine VPS37B et une bande au-dessus de 15 kDa correspondant à la protéine C $\Delta$ 28 dans la fraction totale après sonication (**Figure 60**, colonne 1).

Cependant, on ne retrouve, après élution, que la bande correspondant à VPS37B indiquant clairement que la présence en co-expression de la protéine VPS37B 97-167 seule ne suffit pas à solubiliser la protéine C $\Delta$ 28 (**Figure 60**, colonne 5).



**Figure 60 : Construction et expression du complexe C $\Delta$ 28-VPS37B.** SDS-Page 4-15 % représentant la purification par affinité sur résine Ni<sup>2+</sup> de C $\Delta$ 28-VPS37B co-exprimé en complexe. Les échantillons déposés dans les puits sont : 1. Fraction totale, 2. Surnageant, 3. Flowthrough, 4. Lavage 1, 5. Élution.

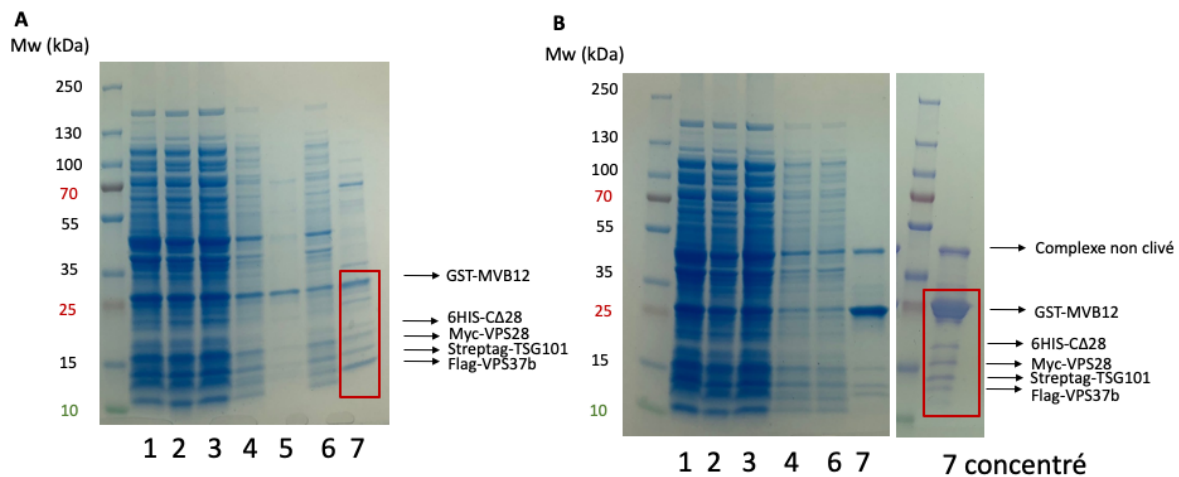
### • C $\Delta$ 28 – ESCRT-I

Le complexe C $\Delta$ 28/ESCRT-I a été purifié à la fois par chromatographie d'affinité sur résine Ni<sup>2+</sup>NTA (accrochage de 6HIS-C $\Delta$ 28) (**Figure 61A**) et sur résine glutathion (accrochage de GST-MVB12A) (**Figure 61B**).

Sur les deux **figures 61A et 61B**, la colonne 1 représente la fraction totale après sonication. Les bandes 2 à 6 correspondent à l'accrochage des protéines solubles sur la résine et le lavage de la résine.

Enfin, nous pouvons observer, dans les deux cas, dans la colonne 7 correspondant à l'élution des purifications, les mêmes bandes entre les marqueurs moléculaires 25 kDa et 10 kDa après accrochage de 6HIS-C $\Delta$ 28 (**Figure 61A**) ou de Strepatg II-MVB12A après concentration de l'élution (**Figure 61B**). Seule une bande fait exception entre les marqueurs moléculaires 35 kDa et 55 kDa dans l'élution de la purification après accrochage de MVB12A (**Figure 61B**, colonne numéro 7) qui pourrait correspondre aux protéines avant clivage co-traductionnel à la TEV.

L'identité des protéines présentes dans l'éluat a été analysée par WB par la révélation des étiquettes 6HIS, GST, Streptag-II et par Spectrométrie de Masse (MS). Nous avons confirmé que toutes les protéines étaient présentes dans l'éluat. La protéine C<sub>Δ28</sub> est donc solubilisée par la présence de la tête du complexe ESCRT-I.



**Figure 61 : Construction et expression du complexe TSG101-VPS37B-VPS28-MVB12-C<sub>Δ28</sub>.** (A&B) SDS-Page 4-20 % représentant la purification par affinité sur résine Ni<sup>2+</sup> (A) et sur résine Glutathion (B) du complexe ESCRT-I/ C<sub>Δ28</sub>. Les échantillons déposés dans les puits sont : 1. Fraction totale, 2. Surnageant, 3. Flowthrough, 4. Lavage 1, 5. Lavage 2, 6. Lavage 3, 7. Éluat.

De nos résultats de purifications, nous avons tiré plusieurs conclusions :

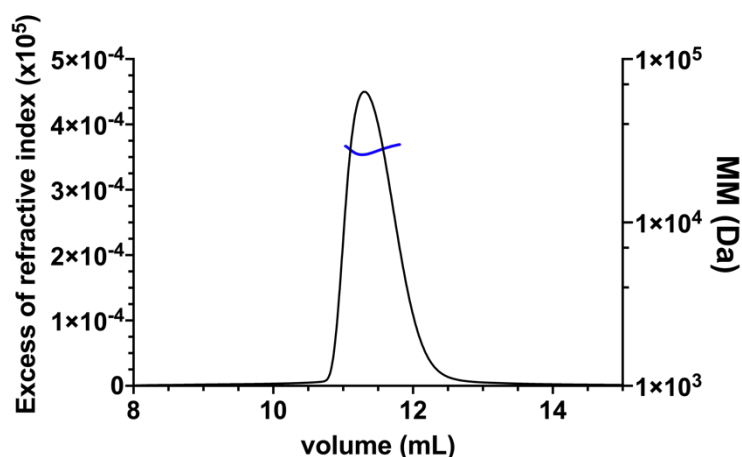
- (1) Toutes les protéines sélectionnées sont insolubles en l'absence de détergents.
- (2) Les protéines C<sub>Δ45</sub>, C<sub>Δ28</sub>, C<sub>Δ35</sub>, C<sub>Δ12</sub> et C<sub>Δ23</sub> sont, à l'inverse de la forme entière, solubles en présence de détergents. Toutes les protéines sélectionnées redeviennent instables lors de la dialyse du détergent NDSB.
- (3) Les protéines issues de la méthode ESPRIT interagissent avec la protéine BirA constitutive en bactérie et nécessitent donc une expression en cellules d'insecte.
- (4) Nous avons réussi à purifier C<sub>Δ28</sub> sans détergent par co-expression avec la région structurée du complexe ESCRT-I. Ce résultat est prometteur et souligne, à nouveau, une possible fonction de la protéine C du virus Nipah dans la machinerie ESCRT. Par la suite, nous souhaitons explorer plus amplement ce résultat par cristallographie et par SAXS.

### (E) Caractérisation biophysique de la protéine C<sub>Δ28</sub> du virus Nipah

#### (1) Évaluation du repliement tridimensionnel et de l'oligomérisation de C<sub>Δ28</sub> en solution en présence de faible quantité de détergent

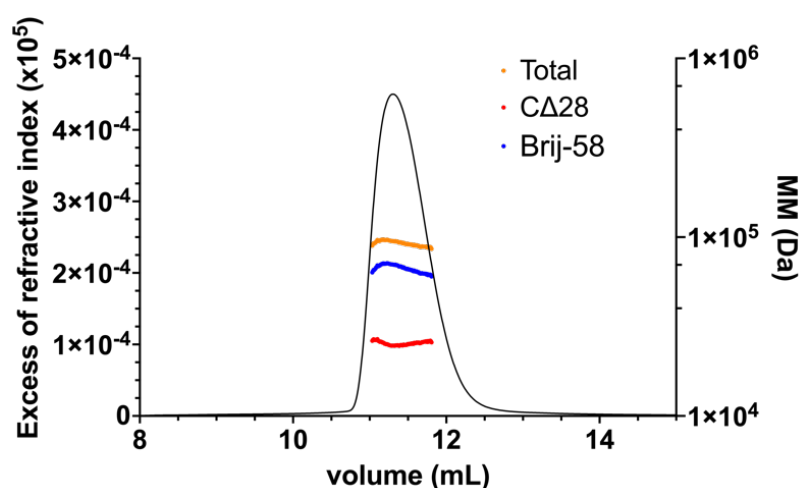
La protéine C<sub>Δ28</sub> purifiée a été analysée par SEC-MALLS-RI (**Figure 62**). La masse moléculaire prédite à partir de la séquence en acide aminé de la protéine monomérique non clivée, qui comprend à la fois le tag BAP et le tag hexahistidine, est de 21 kDa.

De manière surprenante, notre première analyse SEC-MALLS-RI indique une masse moléculaire moyenne (MM) de  $71 \pm 3.5$  kDa et donc potentiellement une forme multimérique en solution.



**Figure 62 : Analyse SEC-MALLS de la protéine  $C_{\Delta 28}$ .** La SEC a été réalisée sur une colonne Superdex Increase 200 équilibrée avec une solution tampon de 50 mM Tris-HCl à pH 7.5 contenant 150 mM NaCl à 20°C avec un débit de 0.5 mL.min<sup>-1</sup>. Le profil d'éluion (courbe noire) a été suivi en ligne par MALLS et réfractométrie différentielle. La masse moléculaire calculée à partir de la diffusion statique de la lumière à travers le pic d'éluion est représentée en bleue.

Nous avons alors appliqué une correction sur l'indice de réfraction (dRI) en tenant compte de l'incrément d'indice de réfraction spécifique du Brij-35 (très proche du Brij-58) ( $\frac{dn}{dc}$  de  $0,134 \pm 0,013$ ), nous avons observé que près de 70% du signal dRI mesuré précédemment (**Figure 63**, orange) était dû à la présence résiduelle de Brij-58 dans la solution (**Figure 63**, bleu). La MM calculée à partir du signal provenant uniquement de la protéine a été mesurée à  $25,0 \pm 1,2$  kDa (**Figure 63**, rouge). Cette valeur est proche de la MM prédite à partir de la séquence en aa égale à 21,3 kDa.

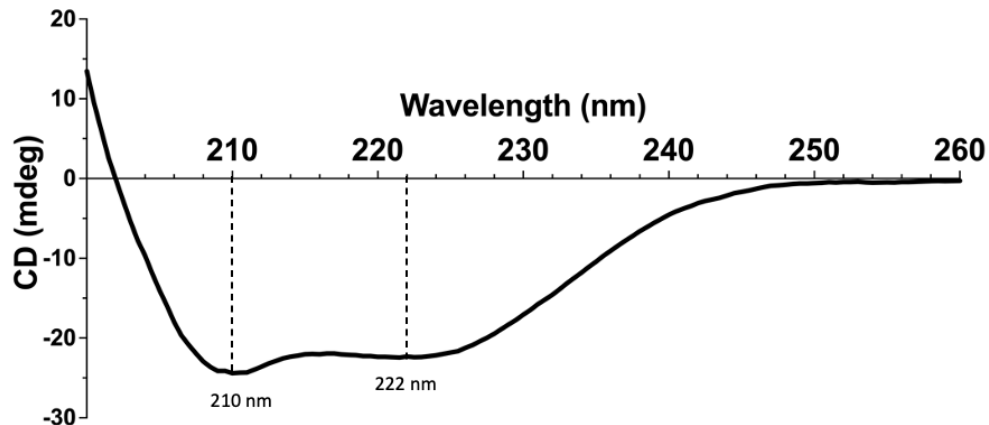


**Figure 63 : Deuxième analyse de  $C_{\Delta 28}$  par SEC-MALLS.** Voir légende précédente. La masse moléculaire calculée à partir de la diffusion statique de la lumière à travers le pic d'éluion est représentée soit pour la protéine  $C_{\Delta 28}$  seule en tenant compte de l'indice de réfraction spécifique du Brij-58 (rouge), soit pour le Brij-58 seul (bleu), soit pour la protéine et le Brij-58 combiné (orange).

Nous avons analysé le contenu de la structure secondaire de  $C_{\Delta 28}$  par dichroïsme circulaire (CD). Le spectre CD de  $C_{\Delta 28}$  est caractéristique d'une protéine repliée en hélices  $\alpha$  conformément aux prédictions de structures secondaires (**Figure 64**). Étant donné que les feuillets  $\beta$  et les pelotes aléatoires participent peu au signal CD à 222nm, nous avons estimé la proportion d'hélices- $\alpha$  dans l'échantillon, par l'utilisation de la valeur d'ellipticité molaire à 222 nm de 35 000 deg.cm<sup>2</sup>.dmol<sup>-1</sup>, à 69%. Cette valeur est proche de la valeur de 63% estimée à partir de la séquence en acides aminés. Nous avons également vérifié que le Brij-58



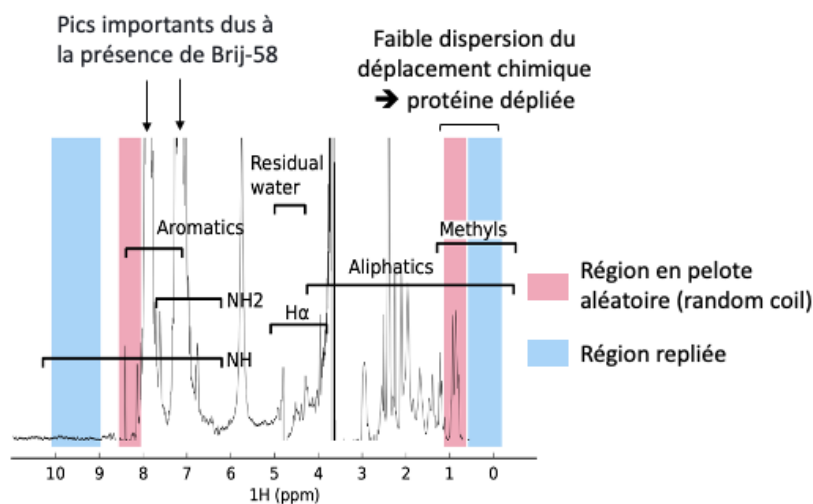
résiduel de l'échantillon n'était pas à l'origine du signal observé. Nous pouvons donc conclure qu'en présence de Brij-58 en solution,  $C_{\Delta 28}$  est bien repliée.



**Figure 64 : Spectre de dichroïsme circulaire de la protéine  $C_{\Delta 28}$ .** Le spectre de  $C_{\Delta 28}$  a été obtenu pour une concentration de  $30\mu\text{M}$  dans une cuve de 1mm dans une solution tampon de Tris HCl 20 mM contenant 150 mM de NaCl et 0,2 mM de TCEP à une température de  $20^\circ\text{C}$ . Il est caractéristique d'une protéine contenant majoritairement des hélices  $\alpha$ .

Nous avons voulu confirmer les résultats de SEC-MALLS-RI sur la présence de détergents dans notre échantillon par une analyse 1D standard par résonance magnétique nucléaire (**Figure 65**). Le spectre obtenu montre une très faible dispersion chimique dans la région des protons méthylés (0 à 1 ppm) caractéristique d'une protéine dépliée ou non structurée. Étant donné que notre analyse CD montre que la protéine  $C_{\Delta 28}$  est bien repliée en solution, nous émettons l'hypothèse que ces pics pourraient correspondre à la région N-terminale désordonnée de  $C_{\Delta 28}$ .

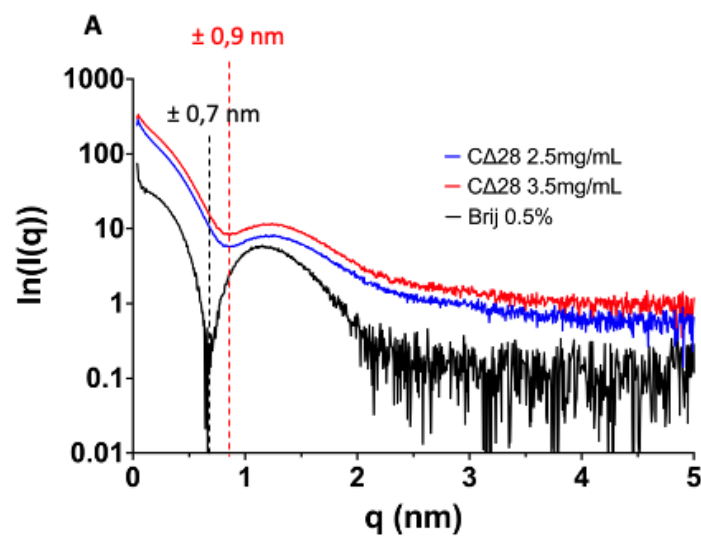
Nous observons deux larges pics inhabituels dans la région des protons amide. En comparant ces pics avec le spectre 1D du même tampon contenant du détergent Brij-58 à 0,5 %, nous avons retrouvé les mêmes pics dans la région amide confirmant les conclusions tirées de l'analyse des courbes de SEC-MALLS-RI précédentes sur la présence résiduelle de détergent dans notre échantillon.



**Figure 65 : Spectre 1D standard par RMN pour la protéine  $C_{\Delta 28}$ .** Les dispersions dans la région correspondant aux cases rouges sont représentatives d'une protéine dépliée ou désordonnée et les cases bleues d'une protéine repliée.

$C_{\Delta 28}$  a été analysée par diffraction aux rayons X aux petits angles (SAXS) en direct sur la ligne de faisceau ID29 du Synchrotron ESRF de Grenoble à différentes concentrations :  $2,5 \text{ mg.mL}^{-1}$  (**Figure 66**, courbe bleue) et  $3.5 \text{ mg.mL}^{-1}$  (**Figure 66**, courbe rouge). Les courbes de diffusion expérimentale obtenues se sont avérées caractéristiques d'une protéine globulaire et non d'une protéine intrinsèquement désordonnée. Nous avons comparé le profil SAXS expérimental de la protéine  $C_{\Delta 28}$  à celui d'une solution de tampon contenant 0,5% de Brij-58 (**Figure 66**, courbe noire). Nous avons constaté que les deux profils étaient similaires. Le minimum atteint par la courbe de diffusion de la protéine  $C_{\Delta 28}$  aux deux concentrations est significativement différent ( $0.9 \text{ nm}$ ) de celui de l'échantillon de Brij-58 à 0,5% ( $0.7 \text{ nm}$ ). Cette différence souligne que notre échantillon est présent dans la solution mais pourrait interagir avec des molécules de Brij-58. La protéine  $C_{\Delta 28}$  semble être stabilisée par une interaction forte avec le détergent.

**Figure 66 : Analyse SAXS de  $C_{\Delta 28}$ .** Courbe de diffusion SAXS de la protéine  $C_{\Delta 28}$  à  $2.5 \text{ mg/mL}$  (bleu) et  $3.5 \text{ mg/mL}$  (rouge) dans du Tris-HCl (pH7.5) contenant  $150 \text{ mM NaCl}$  et  $0.2 \text{ mM TCEP}$  superposée à la courbe de diffusion du même tampon contenant 0.5% de Brij-58.



En conclusion, ces résultats montrent que la protéine  $C_{\Delta 28}$  du virus Nipah et plus généralement les différentes constructions pour la protéine C nécessitent l'addition de détergent pour leur solubilisation et leur stabilisation en solution.

La concentration finale de notre solution protéique en Brij-58 est difficile à estimer. Au vu de la faible CMC ( $0,077 \text{ mM}$  ( $0,0086\% \text{ } \rho/v$ )) du Brij-58, il est probable que la concentration de la solution protéique soit supérieure à la CMC, ce qui pourrait expliquer les courbes de diffusions SAXS obtenues. La protéine pourrait donc être stabilisée par la formation de micelles. Néanmoins, la protéine étant à priori bien repliée en détergent, nous avons cherché à la cristalliser.

## (2) Essais de cristallographie de $C_{\Delta 28}$ en présence de détergent

Nous avons procédé à plusieurs essais de cristallisation de la protéine  $C_{\Delta 28}$  via la plateforme HTX de la Plateforme de Biologie Structurale (PSB) de l'ISBG. Nous avons observé des microcristaux pour certaines conditions, présentées dans la **tableau 2**, pour une concentration de  $10 \text{ mg.mL}^{-1}$ .

Au vu de la difficile maîtrise de la quantité résiduelle de Brij-58 de notre échantillon, ces essais sont difficiles à reproduire. Suite aux résultats obtenus précédemment, nous pensons que, la présence de détergents étant indispensables à la stabilisation de cette protéine, il serait intéressant de cristalliser cette protéine dans des conditions spécifiques des protéines membranaires.

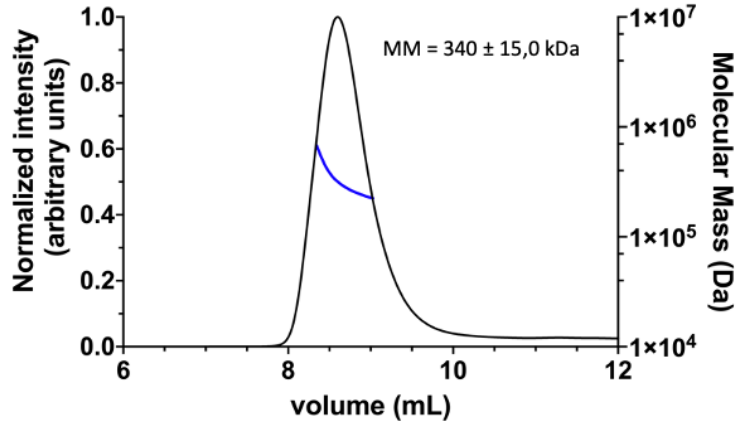
**Tableau 2 : Résultats des premiers essais de cristallisation de la protéine C<sub>Δ28</sub>.**

Plaques	Concentration	Additif	Conditions	Résultat			
Salt-Grid hampton	10	NA	0.1M Acide Citrique 1.6M Sulfate d'ammonium	Micro-cristaux et oursins			
			0.1M Acide Citrique 2.4M Sulfate d'ammonium				
			1.5M et 1.9 M Malonate de sodium				
			0.1M Hepes pH7 1.6M Sulfate d'ammonium				
			0.1M Bicine pH9 1.6M Sulfate d'ammonium				
			0.21M Dihydrogénophosphate de sodium 1.19M Hydrogénatophosphate de potassium pH final = 7.25				
			0.27M Dihydrogénophosphate de sodium 1.53M Hydrogénatophosphate de potassium pH final = 7.25				
			0.1M Tris-HCl pH7 1.6M Sulfate d'ammonium				
			Salt-Grid hampton		10	Trypsine	0.1M Acid Citrique 1.6M Sulfate d'ammonium
							0.1M bicine pH9 1.6M Sulfate d'ammonium
1.9 M Malonate de sodium							
0.2M Chlorure de lithium 0.1M Tris-HCl pH8 20% w/v PEG6000							
PACT_MD	10	NA					
Classics Suite Qiagen	5	NA	NA	Précipitation			

**(F) Etude de l'interaction de C<sub>Δ28</sub> avec les protéines P du NiV et SHCBP1**

**(1) Étude du statut d'oligomérisation de la protéine P du NiV par SEC-MALLS-RI**

Nous avons vérifié l'état d'oligomérisation et la pureté de la protéine P du NiV par SEC-MALLS-RI. La MM prédite à partir de la séquence pour la protéine P monomérique en fusion avec un tag 6HIS est de 78 kDa. La MM déterminée par l'analyse SEC-MALLS pour la protéine P est de 340 ± 15,0 kDa et confirme que la protéine P forme un tétramère en solution comme établi précédemment (**Figure 67**) (Jensen *et al.*, 2020).

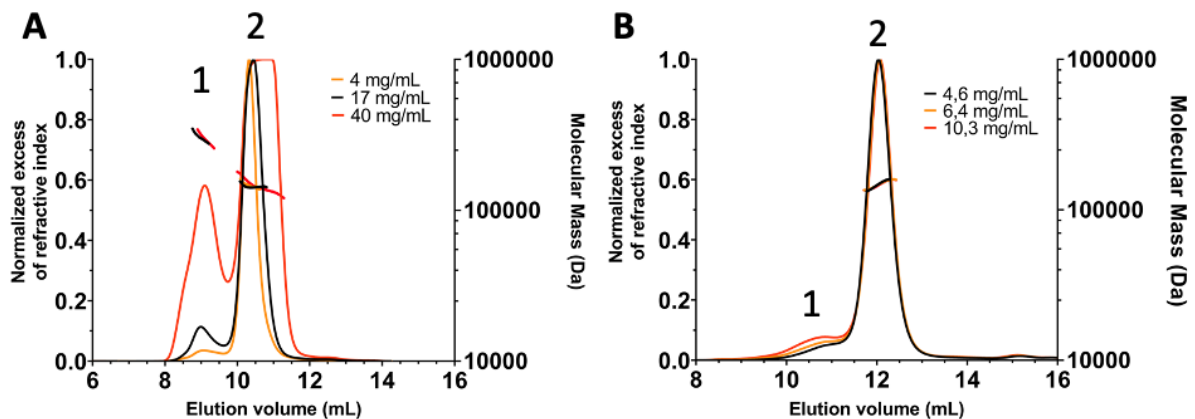


**Figure 67 : Analyse SEC-MALLS-RI de la protéine P du virus Nipah.** La SEC a été réalisée sur une colonne Superdex Increase 200 équilibrée avec une solution tampon de 20 mM Tris-HCl à pH 8 contenant 150 mM NaCl à 20°C avec un débit de 0.5 mL.min<sup>-1</sup>. Le profil d'éluion (courbe noire) a été suivi en ligne par MALLS et réfractométrie différentielle. La masse moléculaire calculée à partir de la diffusion statique de la lumière à travers le pic d'éluion est représentée en bleue.

## (2) Étude du statut d'oligomérisation de la protéine SHCBP1 par SEC-MALLS-RI

De la même façon, nous avons analysé la protéine SHCBP1 dont la MM prédite à partir de la séquence est évaluée à 76 kDa pour le monomère fusionné à l'étiquette 6HIS.

Nous avons injecté la protéine SHCBP1 à différentes concentrations (variant de 4 à 40 mg.mL<sup>-1</sup>) par SEC-MALLS-RI (**Figure 68A/B**). Nous observons que pour des concentrations inférieures à 10 mg.mL<sup>-1</sup>, une forme majoritaire prédomine, indiquée par le pic d'éluion 2 sur la figure 67, pour laquelle la MM moyenne calculée est de 140 ± 2,40 kDa. Ce résultat indique que SHCBP1 forme un dimère en solution dès 0,2 mg.mL<sup>-1</sup>. À plus haute concentration, nous observons l'apparition d'un oligomère, indiqué par le pic d'éluion 1 sur la figure 67, pour laquelle la MM moyenne calculée est de 301 ± 3,00 kDa correspondant à un tétramère.



**Figure 68 : Étude de l'oligomérisation de la protéine SHCBP1 par SEC-MALLS.** (A) La SEC a été réalisée à 20°C avec un débit de 0.5 mL.min<sup>-1</sup> sur une colonne Superdex Increase 200 équilibrée avec une solution tampon de 20 mM Tris-HCl à pH 7.5 contenant 150 mM NaCl. Les profils d'éluion ont été suivis en ligne par MALLS et réfractométrie différentielle à différentes concentrations (courbe rouge à 40 mg.mL<sup>-1</sup>, courbe noire à 17 mg.mL<sup>-1</sup> et courbe orange à 4 mg.mL<sup>-1</sup>). La masse moléculaire

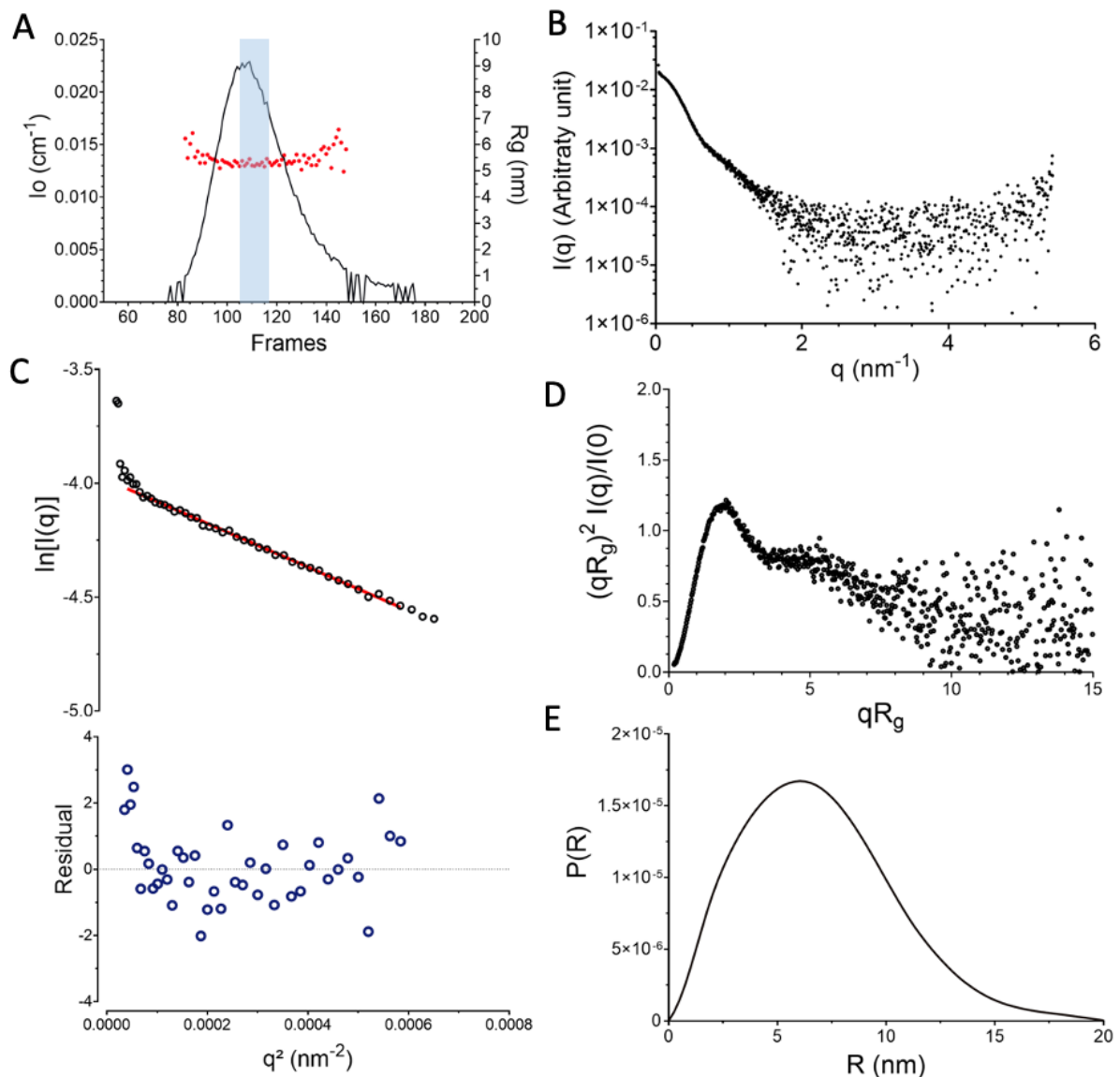
calculée à partir de la diffusion statique de la lumière à travers le pic d'élution est représentée dans les mêmes couleurs. L'échantillon de SHCBP1 est dilué tout au long des mesures. La MM mesurée pour le pic d'élution 1 est de  $301 \pm 3,00$  kDa correspondant à la forme tétramérique de la protéine. La MM mesurée pour le pic d'élution 2 est de  $140 \pm 2,40$  kDa correspondant à la forme dimérique de la protéine. **(B)** Analyse SEC-MALLS de 3 échantillons de la protéine cellulaire SHCBP1 concentrés au cours de l'expérience (courbe noire à  $4.5 \text{ mg.mL}^{-1}$ , courbe orange à  $6.4 \text{ mg.mL}^{-1}$  et courbe rouge à  $10.3 \text{ mg.mL}^{-1}$ ). Les pics 1 et 2 correspondant à la forme tétramérique et à la forme dimérique de la protéine sont également indiqués et observés.

L'état d'oligomérisation de SHCBP1 semble donc dépendant de la concentration et réversible. La **figure 68A** présente les résultats d'injections séquentielles d'un même échantillon dilué au cours de l'expérience ( $40 \text{ mg.mL}^{-1}$  en rouge,  $17 \text{ mg.mL}^{-1}$  en noir et  $4 \text{ mg.mL}^{-1}$  en orange). Tandis que la **figure 68B** présente les résultats d'injections séquentielles d'un même échantillon concentré au cours de l'expérience ( $4.6 \text{ mg.mL}^{-1}$  en noir,  $6.4 \text{ mg.mL}^{-1}$  en orange et  $10.3 \text{ mg.mL}^{-1}$  en rouge). On observe que, pour un même échantillon, le pic d'élution correspondant au tétramère (entre 8 et 10 mL) varie en fonction de la concentration de l'échantillon.

### (3) Caractérisation structurale de la protéine SCHBP1 par SEC-SAXS

La protéine SHCBP1 a été par la suite étudiée par SEC-SAXS sur la ligne de lumière de BioSAXS SWING du synchrotron Soleil (**Figure 69**).

Au vu de la relation entre la concentration de SHCBP1 et sa capacité à former des oligomères, nous avons choisi d'analyser par SEC-SAXS uniquement le centre du pic d'élution de SHCBP1 sans concentration et après injection sur une colonne de chromatographie d'exclusion de taille. De cette façon, nous espérons limiter la formation de tétramères qui pourrait perturber les analyses.

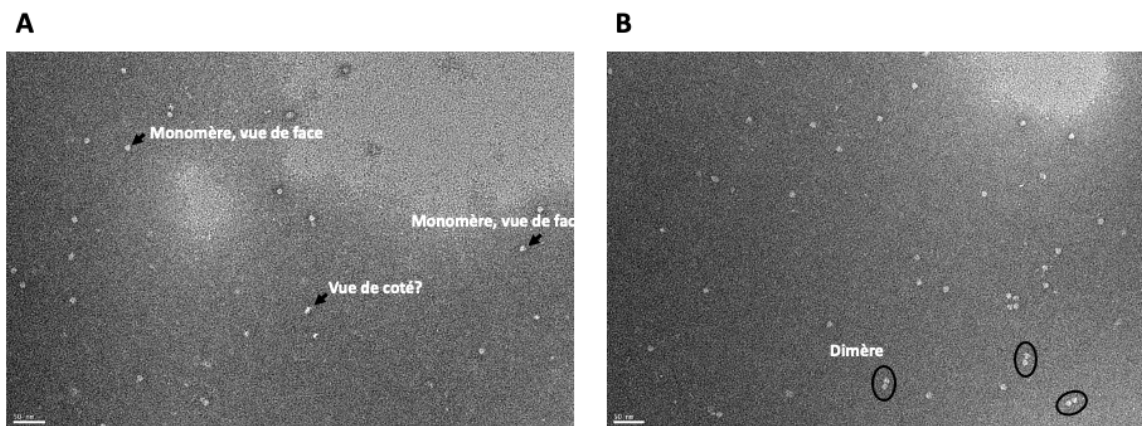


**Figure 69 : Étude de la protéine SHCBP1 par SEC-SAXS.** (A) Évolution des valeurs  $I(0)$  et  $R_g$  le long de l'élué par SEC. La ligne noire montre la diffusion à angle nul ( $I(0)$ ), qui est proportionnelle à la fois à la masse moléculaire et à la concentration, en fonction des trames enregistrées à intervalles de temps réguliers. Les ronds rouges montrent le rayon de giration calculé à partir des tracés de Guinier aux différents intervalles de temps. Les analyses suivantes ont été limitées à 12 images à droite du pic chromatographique (rectangle bleu). (B) Courbe de diffusion moyennée. La courbe montre le profil de diffusion obtenu pour SHCBP1 en faisant la moyenne des profils individuels enregistrés tout au long du pic d'élué SEC montré dans le panneau A. (C) Graphique de Guinier. Le panneau supérieur montre le diagramme de Guinier dans la région de Guinier ( $q \cdot R_g < 1.3$ ). Le panneau inférieur montre les tracés des résidus normalisés. (D) Diagramme de Kratky normalisé. (E) Fonction de distribution des paires de distance  $P(r)$  obtenue par une transformation indirecte de Fourier donnant accès à la distance maximale  $D_{MAX}$ .

Le tracé de Guinier aux faibles valeurs de  $q$  ( $q \cdot R_g < 1.3$ ) est linéaire et les valeurs de  $R_g$ , calculées en utilisant l'approximation de Guinier, sont constantes le long du pic chromatographique (Figure 69A/C) indiquant la monodispersité, à priori, de ces échantillons. On peut observer une faible agrégation de l'échantillon (Figure 69B) par la remontée des valeurs à faible  $q$ . Nous avons obtenu une courbe SAXS représentative par le calcul de la moyenne de 12 images au centre du pic chromatographique (Figure 69B). Les données à plus

grand  $q$  ( $>2$ ) sont très bruitées comme le montrent les **Figures 69C** et **69D** de par la faible concentration de l'échantillon ( $0,8 \text{ mg.mL}^{-1}$ ) et sont donc difficiles à interpréter mais semblent suggérer que cette protéine est multi-domaine (deuxième maxima du graphique de Kratky à  $q.R_g = 5$  (**Figure 69D**)). La  $D_{MAX}$  est estimée par la fonction de distribution des paires de distance à  $20,1 \text{ nm}$  (**Figure 69E**). Nous avons déterminé des valeurs moyennes de  $R_g$  de  $5,3 \pm 2,6 \text{ nm}$  en utilisant l'approximation de Guinier (**Figure 69C**). La méthode du volume de corrélation nous a permis d'estimer la MM moyenne de l'échantillon à  $185 \text{ 153 Da}$ , soit une valeur intermédiaire entre un dimère (mesuré à  $140 \text{ kDa}$  par SEC-MALLS-RI) et un trimère ( $MM = 228 \text{ kDa}$  d'après les prédictions depuis la séquence). Cette estimation ne correspond pas aux mesures correspondantes en SEC-MALLS-RI qui suggèrent que, la protéine est dimérique. Il semblerait que même à faible concentration, la protéine SHCBP1 en solution forme des homomultimères impossibles à séparer par chromatographie d'exclusion de taille et qu'il coexiste dynamiquement en solution des dimères et trimères.

Des images réalisées en microscopie électronique ont permis d'observer la coexistence de monomères et dimères de la protéine (**Figure 70**).



**Figure 70 : Image de microscopie électronique à transmission de la protéine SHCBP1.** Les clichés sont colorés négativement et permettent d'observer la coexistence de monomères (flèches (A)) et de dimères (cercles noirs (B)) dans l'échantillon.

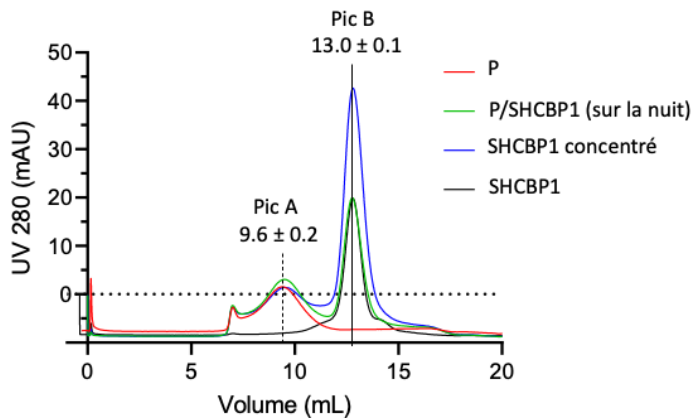
En parallèle de ces analyses, nous avons cherché à confirmer ou infirmer l'hypothèse suggérée par l'étude de Lo *et al* (Lo, Søggaard and Karlin, 2014) et Ito *et al* (Ito *et al.*, 2013), supposant l'existence d'une possible interaction entre la protéine C du NiV et la protéine SHCBP1 régulant les mécanismes d'actions de la protéine P du NiV.

#### (4) interactions entre les protéines C et P du virus Nipah

Pour cela, les protéines purifiées indépendamment ont été injectées séparément et après interaction durant la nuit à  $4^\circ\text{C}$  en colonne de chromatographie d'exclusion de taille. Une interaction et la formation d'un complexe se traduiraient par une décalage des pics d'élution vers de plus petits volumes et donc vers des poids moléculaires plus grands.

- **P et SHCBP1**

Les protéines P et SHCBP1 ont été injectées séparément dans des proportions équimolaires et nous avons suivi le profil d'éluion à 280nm (**Figure 71**, courbe rouge (P) et noire (SHCBP1)). Les protéines ont ensuite été mélangées sur la nuit à 4°C en ratio équimolaire (**Figure 71**, courbe verte) ou avec un excès de SHCBP1 (**Figure 71**, courbe bleue). Les profils d'éluion ont été superposés dans la **figure 71**. Aucune variation des volumes d'éluion n'est observée après interaction des protéines sur la nuit. La protéine P de NiV élue au niveau du pic A ( $9.6 \pm 0.2$  mL) et la protéine SHCBP1 élue au niveau du pic B ( $13.0 \pm 0.1$  mL).



**Figure 71 : Comparaison des profils d'éluion des protéines SHCBP1 et P de NiV.** Les courbes rouge et noire correspondent respectivement aux profils d'éluion de la protéine P de NiV à 22.6  $\mu$ M et de la protéine SHCBP1 à 18.8  $\mu$ M injectées seules dans un tampon de Tris-HCl 50mM pH8, NaCl 150mM et TCEP 0.2mM. Les protéines ont ensuite été mélangées à 4°C sur la nuit et injectées sur la même colonne soit en ratio équimolaire 1:1 (courbe verte) ou en ratio 1:2 (P :SHCBP1) (courbe bleue).

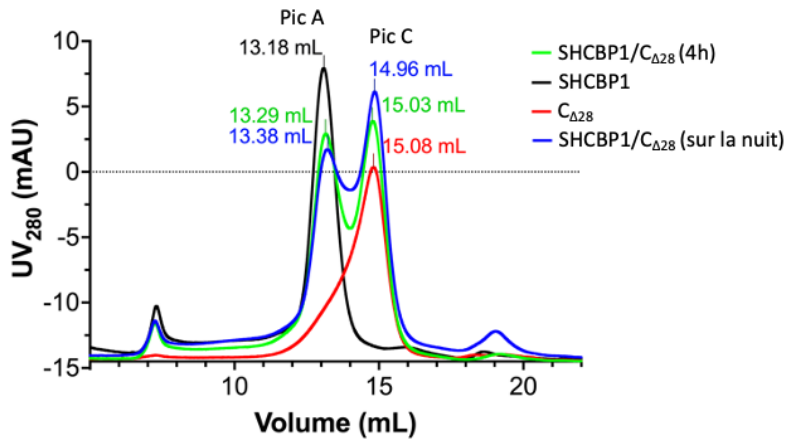
Ce résultat semble indiquer que la protéine P du NiV n'est pas capable, dans ces conditions, d'interagir avec la protéine SHCBP1.

- **C<sub>Δ28</sub> et SHCBP1**

De la même façon que précédemment, les protéines SHCBP1 et C<sub>Δ28</sub> ont été injectées séparément et en ratio équimolaire (**Figure 72**, courbe noire (SHCBP1) et courbe rouge (C<sub>Δ28</sub>)). Ces protéines ont ensuite été mises en contact à 4°C pendant 4h (**Figure 72**, courbe verte) et sur la nuit (**Figure 72**, courbe bleue). À première vue, les volumes d'éluion ne varient pas significativement lorsque les protéines sont mises en contact. La protéine SHCBP1 élue au niveau du pic A ( $13.28 \pm 0.11$  mL) et la protéine C<sub>Δ28</sub> élue au niveau du pic C ( $14.02 \pm 0.06$  mL).

Cependant, on constate une diminution de l'absorbance des UV à 280nm du pic d'éluion correspondant à SHCBP1 en faveur du pic d'éluion correspondant à C<sub>Δ28</sub>. Cette inversion des absorbances est d'autant plus marquée que le temps d'interaction est long. L'interaction des deux protéines pourrait favoriser la dissociation des oligomères de SHCBP1 au profit d'une forme monomérique (MM prédite à 76 kDa) qui éluerait en même temps que C<sub>Δ28</sub> (MM mesurée à 75 kDa en SEC-MALLS-RI en présence de Brij-58). Il n'est pas exclu que l'oligomérisation de SHCBP1 soit également affectée par la présence résiduelle de Brij-58 de la solution de C<sub>Δ28</sub>.





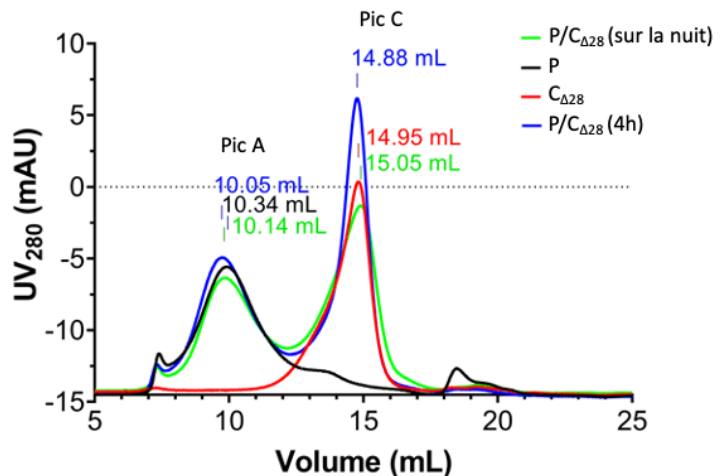
**Figure 72 : Comparaison des profils d'éluion des protéines SHCBP1 et C $\Delta$ 28 de NiV.** Les courbes rouge et noire correspondent respectivement aux profils d'éluion de la protéine C $\Delta$ 28 de NiV à 4,8  $\mu$ M et de la protéine SHCBP1 à 18.8  $\mu$ M injectées seules dans un tampon de Tris-HCl 50mM à pH8, NaCl 150mM et TCEP 0.2mM. Les protéines ont ensuite été mélangées en ratio équimolaire à 4°C pendant 4 heures (courbe verte) et sur la nuit (courbe bleue) et injectées sur la même colonne.

heures (courbe verte) et sur la nuit (courbe bleue) et injectées sur la même colonne.

- C $\Delta$ 28 et P

Nous constatons que les volumes d'éluion pour les protéines P et C $\Delta$ 28 sont presque identiques que les protéines soient injectées seules (**Figure 73**, C $\Delta$ 28 seule (courbe rouge) et P seule (courbe noire)) ou après une interaction sur la nuit (**Figure 73**, 4h d'interaction à 4°C (courbe bleue) et interaction à 4°C sur la nuit (courbe verte)). La protéine P élue au niveau du pic A (10.17  $\pm$  0.13 mL) et la protéine C $\Delta$ 28 élue au niveau du pic C (14.96  $\pm$  0.09 mL). Il semble donc que les protéines C $\Delta$ 28 et P du NiV n'interagissent pas.

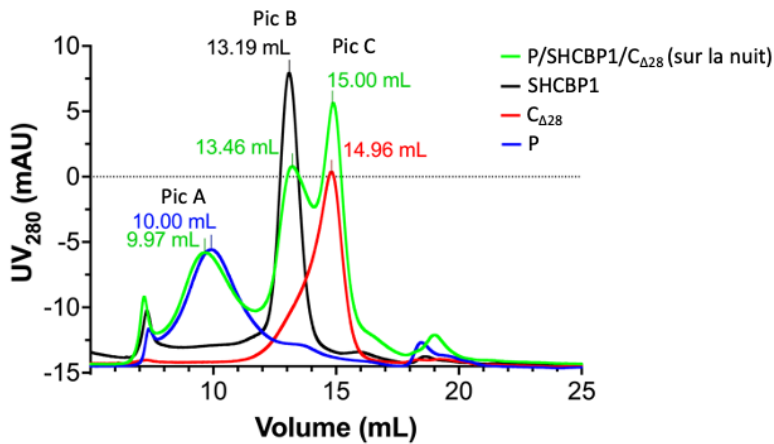
**Figure 73 : Comparaison des profils d'éluion des protéines P et C $\Delta$ 28 de NiV.** Les courbes rouge et noire correspondent respectivement aux profils d'éluion de la protéine C $\Delta$ 28 de NiV à 4,8  $\mu$ M et de la protéine P de NiV à 22,6  $\mu$ M injectées seules dans un tampon de Tris-HCl 50mM à pH8, NaCl 150mM et TCEP 0.2mM. Les protéines ont ensuite été mélangées à 4°C 4 heures (courbe bleue) et sur la nuit (courbe verte) en ratio équimolaire et injectées sur la même colonne.



- C $\Delta$ 28, P et SHCBP1

De la même façon, nous avons étudié la formation d'un complexe lorsque les trois protéines sont injectées ensemble après une nuit d'interaction à 4°C (**Figure 74**, courbe verte). Nous observons une diminution de l'absorbance des UV à 280nm correspondant au pic B (13.3  $\pm$  0.13 mL) d'éluion de la protéine SHCBP1 (**Figure 74**, courbe noire). Cette diminution est compensée par une augmentation de l'absorbance au niveau du pic d'éluion C 14.98  $\pm$  0.02 mL correspondant à l'éluion de la protéine C $\Delta$ 28 (**Figure 74**, courbe rouge). Les volumes d'éluion, pour autant, ne varie pas significativement que les protéines soient

injectées seules ou en interaction. Ce résultat est similaire aux courbes obtenues lors de l'étude de l'interaction des protéines SHCBP1 et  $C_{\Delta 28}$ . La protéine P du NiV élue au niveau du pic A ( $9.99 \pm 0.02$  mL). Aucune variation des volumes d'élution n'est observée pour la protéine P du NiV seule ou en interaction et l'intensité de l'absorbance pour le pic A ne varie pas (**Figure 74**, courbe bleue).



**Figure 74 : Comparaison des profils d'élution des protéines P,  $C_{\Delta 28}$  et SHCBP1.** Les courbes rouge, bleue et noire correspondent respectivement aux profils d'élution de la protéine  $C_{\Delta 28}$  de NiV à  $4,8 \mu\text{M}$ , de la protéine P de NiV à  $22,6 \mu\text{M}$  et de la protéine SHCBP1 à  $18,8 \mu\text{M}$  injectées seules dans un tampon de Tris-HCl  $50\text{mM}$  pH8, NaCl  $150\text{mM}$  et TCEP  $0,2\text{mM}$ . Les protéines ont ensuite été

mélangées à  $4^\circ\text{C}$  sur la nuit (courbe verte) en ratio équimolaire et injectées sur la même colonne.

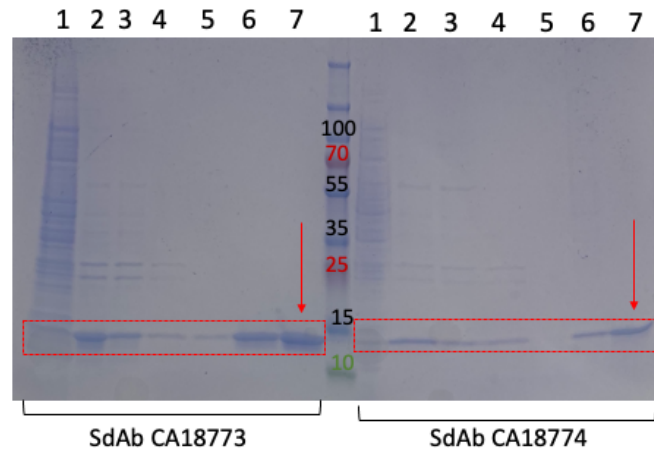
Des analyses supplémentaires sont nécessaires pour comprendre ces observations. À première vue, aucune interaction des protéines P,  $C_{\Delta 28}$  du NiV et de SHCBP1 n'est clairement observée. Cependant, il serait intéressant d'approfondir le rôle de  $C_{\Delta 28}$  dans l'oligomérisation de SHCBP1 en présence de la protéine P du NiV.

### ***(G) Production d'anticorps à domaine unique (SdAb) anti- $C_{\Delta 28}$ : vers de futurs cristaux pour la protéine C du virus Nipah ?***

6 immunisations consécutives d'un lama par la protéine  $C_{\Delta 28}$  ont été réalisées. Deux familles de SdAb<sup>1</sup> ont été isolées comprenant pour la première 3 SdAb et pour la deuxième 1 SdAb.

Nous avons purifié deux des 4 SdAb produits, CA18773 (**Figure 75, flèche rouge gauche**) et CA18774 (**Figure 75, flèche rouge droite**) qui seront prochainement utilisés pour la solubilisation, la stabilisation en solution et la cristallisation de  $C_{\Delta 28}$  et d'autres fragments de la protéine C du NiV.

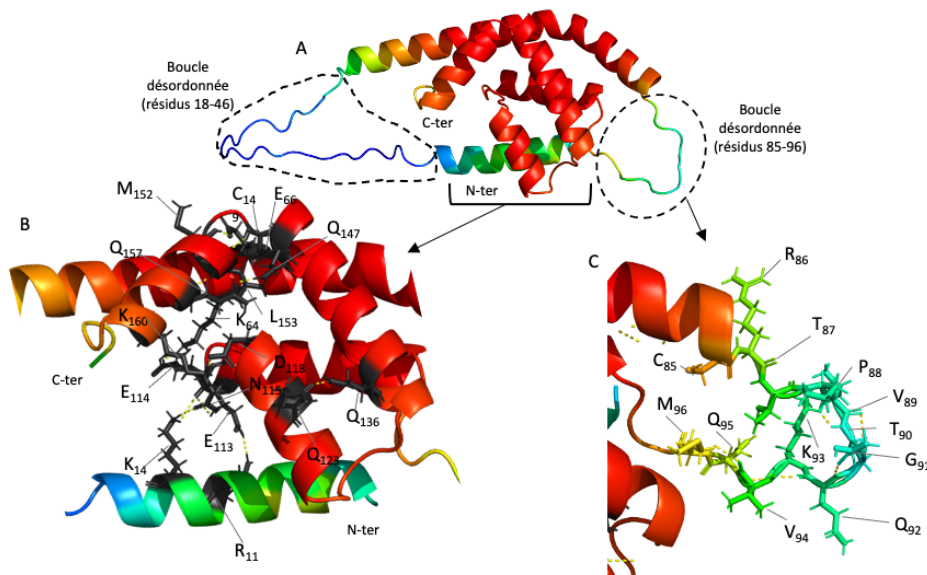
<sup>1</sup> Une famille de SdAb correspond à un groupe d'anticorps ayant plus de 80% d'identité de séquence au niveau de leur domaine CDR3. Ils dérivent de la même lignée de cellules B et se lient au même épitope.



**Figure 75 : Purification des SdAb CA18773 et CA18774.** Les deux SdAb ont été produits et purifiés selon un protocole décrit précédemment (Pardon *et al.*, 2014). La position des SdAb est encadrée en rouge et la fraction correspondant à l'éluion est indiquée par une flèche rouge. Les fractions déposées correspondent à 1. Fraction totale, 2. Surnageant, 3. Flowthrough, 4. Lavage 1, 5. Lavage 2, 6. Résine NiNTA avant éluion, 7. Éluion.

### (H) Modèle AlphaFold2 de la protéine C du virus Nipah

Nous avons généré un modèle AlphaFold2 (Mirdita *et al.*, 2022) de la protéine C pleine longueur (Figure 76A).



**Figure 76 : Modèle AlphaFold2 de la structure tridimensionnelle pour la protéine C du NiV.** Les résidus sont colorés du rouge au bleu selon leur valeur d'indice de confiance pLDDT : les prédictions les plus fiables sont représentées par la couleur rouge tandis que les moins fiables par la couleur bleue. **(A)** Prédiction AlphaFold2 pour la protéine C pleine longueur du NiV. La région C-terminale est composée de 5 hélices  $\alpha$  repliées, en rouge, dont l'architecture est interrompue par une boucle désordonnée (entourée à droite) entre les résidus 85 et 96. La région N-terminale est composée de la boucle désordonnée en bleue (entourée à gauche) entre les résidus 18 et 46 suivie d'une courte hélice repliée sur la région C-terminale colorée du bleu au vert. **(B)** Détail des interactions polaires impliquées dans le repliement de la région N-terminale et stabilisant l'hélice N-terminale sur la région C-terminale. **(C)** Détail des résidus hydrophobes formant la boucle désordonnée interrompant l'architecture de la région repliée N-terminale.

Ce modèle prédit une organisation similaire à nos résultats bio-informatiques avec :

(i) Une petite hélice  $\alpha$  N-terminale qui vient se replier sur la région C-terminale par le biais d'une interaction entre les résidus Arg<sub>11</sub>, Lys<sub>14</sub>, Glu<sub>113</sub> et Asn<sub>115</sub>. Cette région est suivie d'une région majoritairement désordonnée présentant une boucle désordonnée (résidus 18 à 46).

(ii) Une région C-terminale repliée formée de 5 hélices  $\alpha$  (résidus 47 à 166) (**Figure 76B**). 17 résidus, indiqués sur la figure 75B, participent au repliement de ces hélices par des interactions polaires. On observe une interruption de l'architecture en hélices entre les résidus 85 et 96. Ces résidus correspondent en partie au groupement présentant un profil HCA et une prédiction PSIPRED sans structures secondaires contredisant le résultat de Dscore. La structure de ce motif est modélisée par une boucle désordonnée où plusieurs résidus sont exposés au solvant (**Figure 76C**). La position de cette boucle désordonnée n'est pas certaine. Il est possible que ces résidus soient en interaction avec les détergents utilisés pour solubiliser la protéine. Ils pourraient être à l'origine du mauvais repliement de la protéine entraînant sa précipitation lors de leur dialyse. Il serait intéressant d'étudier si le remplacement de ce motif par celui d'une autre protéine C peut jouer sur la stabilité de la protéine C du virus Nipah.

### III. Conclusions et perspectives

Caractériser la structure de la protéine C du NiV est une étape essentielle à la compréhension des mécanismes à la source de la pathogénicité du virus. Malgré l'application de différentes méthodes pour identifier des fragments solubles de cette protéine, elle semble encore aujourd'hui résister à sa caractérisation.

Notre étude bio-informatique a permis d'affiner les prédictions de structures secondaires de la protéine C du NiV. Cette protéine, de la même façon que les autres protéines C des *Orthoparamyxovirinae*, est formée, en prédiction, d'un domaine N-terminal désordonné, à l'exception d'une courte région N-terminale structurée en hélice  $\alpha$  (résidus 1 à 19) correspondant probablement à une région  $\alpha$ -MoRE, et d'une longue région C-terminale ordonnée, composée de l'enchaînement de 5 hélices  $\alpha$ .

Nous avons notamment pu observer que les fragments de protéines C auxquels nous avons retiré les 18 premiers (ou plus) résidus N-terminaux sont solubilisés par l'addition de détergents variés à la différence de la forme entière qui reste, quant à elle, insoluble dans toutes les conditions testées.

Les fragments de protéines C pour lesquelles les premiers résidus N-terminaux ont été retirés sont solubles en détergents mais restent grandement instables. Nous avons cependant confirmé par dichroïsme circulaire que le fragment C <sub>$\Delta$ 28</sub> (tronqué des 27 premiers résidus N-terminaux) de la protéine C du NiV est replié, pour près de 70% de la structure, en hélices  $\alpha$ . Nous avons observé l'existence de fortes interactions entre les molécules de Brij-58, détergent utilisé pour la solubilisation de la protéine, et C <sub>$\Delta$ 28</sub>. En prenant en compte la présence de Brij-58 dans la solution, C <sub>$\Delta$ 28</sub> apparaît monomérique en solution. Une première approche

consistant à stabiliser la protéine en présence de potentiels partenaires viraux et cellulaires semble indiquer un rôle de la protéine C<sub>Δ28</sub> sur l'oligomérisation de SHCBP1 qui devra être confirmé. Nous n'observons, cependant, pas de formation de complexe claire entre ces protéines.

Nous avons, en collaboration avec l'Institut de Biologie Structurale de Bruxelles, produit 4 SdAb spécifiquement dirigés contre le fragment C<sub>Δ28</sub> qui seront utilisés pour stabiliser et cristalliser la protéine. Ces SdAb seront d'une aide précieuse pour résoudre la structure cristallographique. En effet, l'absence de bon homologue dans la base de donnée PDB nous empêchera d'utiliser la méthode du remplacement moléculaire. Enfin, l'utilisation de SdAb marqués par des fluorophores pourrait nous permettre de suivre la protéine C en cellules vivantes afin d'étudier sa localisation durant une infection. Ces SdAb pourront également être testés dans l'altération des fonctions de la protéine C du NiV en cellule durant une infection.

Enfin, nous avons identifié par double hybride en levure, trois partenaires de la protéine C du NiV : VPS37B, MEOX2 et MTUS2. Nous avons coexprimé l'ensemble de la région repliée du complexe ESCRT-I comprenant VPS37B. Ceci nous a permis de solubiliser, pour la première fois, le fragment C<sub>Δ28</sub> sans détergents. De multiples optimisations seront nécessaires afin de permettre la caractérisation du complexe (dont notamment l'élimination de l'étiquette GST) mais ce résultat est prometteur et pourrait indiquer la capacité de la protéine C du virus Nipah à interagir avec le complexe ESCRT.

Afin de caractériser cette protéine en détergent, il serait également possible d'appliquer la méthode de diffusion aux neutrons aux petits angles (SANS - Small Angle Neutron Scattering). Cette méthode permet d'obtenir un contraste nul, par l'utilisation d'isotope de l'hydrogène comme le deutérium (D<sub>2</sub>O), entre la densité moyenne d'une particule et celle du solvant (détergent) de sorte à ne conserver que l'amplitude de diffusion correspondant à la protéine d'intérêt. Deux détergents communs, Octyl-β-D-glucopyranoside (β-OG) et n-dodecyl-β-D-maltopyranoside (DDM), utilisés pour la solubilisation de protéines membranaires ont été synthétisés dans des versions substituées par des isotopes afin de rendre les micelles de détergent « invisible » en D<sub>2</sub>O (Midtgaard *et al.*, 2018). L'addition de β-OG durant la purification de la protéine C<sub>Δ28</sub> a été testée et permet sa solubilisation. Il serait donc intéressant d'envisager d'appliquer la méthode SANS à C<sub>Δ28</sub>.

## IV. Matériels appliqués à cette partie

### (A) Prédiction de désordre

L'emplacement des régions intrinsèquement désordonnées (IDR) dans les protéines de l'étude a été prédit en soumettant leurs séquences d'acides aminés à 16 algorithmes différents accessibles via des serveurs web et en calculant une prédiction consensuelle comme décrit précédemment (Gerard *et al.*, 2009; Jensen *et al.*, 2020; Roy, Chan Mine, Gaifas, Leyrat, V. Volchkova, *et al.*, 2023).

## (B) Clonage

### (1) Clones d'expression de la protéine C du virus Nipah

Les différentes constructions testées dans ce projet pour la protéine C et leur vecteur d'expression sont présentés dans le tableau suivant :

**Tableau 3 : Présentation des constructions de C utilisées dans l'étude.**

Construction	Résidus	Vecteur	Système d'expression	Étiquette N-terminale	Étiquette C-terminale	Site de clivage N-terminal
C <sub>FL</sub>	1-166	pET-28a	Bactérien	6HIS	NA	TEV
C <sub>Δ45</sub>	44-166					
C <sub>Δ91</sub>	90-166					
C <sub>Δ28</sub>	27-166					
C <sub>Δ35</sub>	34-150					
C <sub>Δ12</sub>	11-132					
C <sub>Δ23</sub>	22-132					
C <sub>Δ28</sub>	27-166	pFastBacHTA	Hi5			

### (2) Clones d'expression des protéines P du virus Nipah et du facteur cellulaire humain SHCBP1

Le plasmide codant pour la phosphoprotéine P du virus Nipah correspond à une construction développée précédemment (Jensen *et al.*, 2020). Brièvement, le gène synthétique de la protéine P (Uniprot : Q9IK91) est cloné dans le vecteur pET28 en fusion avec une étiquette 6HIS C-terminale.

Le gène codant pour la protéine humaine SHCBP1 (Uniprot : Q8NEM2) a été synthétisé et optimisé pour l'expression en *E. coli*. Il a été ensuite cloné dans le vecteur pET-M11 entre les sites de restriction NcoI et XhoI en fusion avec une étiquette 6HIS N-terminale suivie par un site de clivage à la protéase TEV.

### (3) Clones de co-expression de C<sub>Δ28</sub> avec les protéines du complexe ESCRT-I

Les gènes synthétiques optimisés codant pour la protéine C<sub>Δ28</sub> et les résidus 97 à 167 de la protéine VPS37B (Flower *et al.*, 2020) ont été sous-clonés séquentiellement dans le vecteur pET-28a. Le gène codant pour la protéase TEV a été ajouté en 3' permettant le clivage co-translationnel des protéines VPS37B et C<sub>Δ28</sub>. Les séquences codantes de VPS37B, de C<sub>Δ28</sub> et de la TEV sont séparées par un site de clivage à la protéase TEV. Les protéines VPS37B et C<sub>Δ28</sub> sont fusionnées, respectivement, avec une étiquette Streptag II et une étiquette hexahistidine en N-terminal.

Les gènes synthétiques optimisés codant pour les résidus 308 à 388 de la protéine TSG101, 97 à 167 de VPS37B, 1 à 122 de VPS28, 206 à 268 de MVB12A (Flower *et al.*, 2020) et C<sub>Δ28</sub> ont été clonés séquentiellement dans le vecteur PET-28a suivis de la séquence

codant pour la protéase TEV en 3'. Les protéines TSG101, VPS37b, VPS28, MVB12A et CA28 ont été fusionnées, respectivement, à une étiquette streptag-II, flag, Myc, GST, hexahistidine en N-terminal. Un site de clivage TEV a été ajouté entre chaque gène. Le complexe TSG101, VPS37B, VPS28 et MVB12 constitue la tête globulaire structurée du sous-complexe ESCRT-I de la machinerie ESCRT et sera désigné ESCRT-I dans la suite du texte.

### ***(C) Expression et purification***

#### **(1) Expression et purification des protéines C de NiV en système bactérien**

Les plasmides codant pour les protéines C de NiV ont été exprimés dans une variété de souches *E. coli* (BL21 STAR (DE3), ROSETTA, BL21 RIL) par addition de 0,5 mM d'IPTG (isopropyl  $\beta$ -D-1-thiogalactopyranoside). Les souches et conditions testées sont présentées en **annexe 4**. Nous ne détaillons ici que le matériel des résultats présentés.

Après expression des protéines C, les cellules sont collectées par centrifugation et les culots contenant les protéines sont conservés à -20°C. Les culots cellulaires sont remis en suspension dans une solution tampon de 50 mM de Tris-HCL à pH 7.5 contenant 300 mM de NaCl, 5 mM d'imidazole, 0.2 mM de tris(2-carboxyéthyl)phosphine (TCEP) et un cocktail d'inhibiteurs de protéases sans EDTA (**tampon A**).

Les protéines C ont été solubilisées par addition de divers détergents dans le tampon A dont notamment 0.5% de Brij-58. Le Brij-58 est un détergent non-ionique de CMC faible égale à 0,077 mM (0,0086%  $\rho/v$ ). Ces détergents sont communément utilisés pour obtenir des protéines en état natif. Ils peuvent perturber les interactions protéines-lipides mais n'interfèrent normalement pas dans les interactions protéines-protéines.

Les cellules sont lysées par sonication et la fraction soluble est appliquée à une colonne Ni<sup>2+</sup> (NiNTA, Qiagen) équilibrée dans le **tampon A**. Après lavage avec un tampon de 50 mM de Tris-HCL à pH 7.5 contenant 1 M de NaCl, 10 mM d'imidazole et 0.2 mM de TCEP (**tampon B**), les protéines sont éluées dans une solution tampon de 50 mM de Tris-HCL à pH 7.5 contenant 150mM de NaCl, 250 mM d'imidazole, 0.2 mM de TCEP (**tampon C**). Les protéines éluées sont incubées pendant la nuit à 4°C avec la protéase TEV pour éliminer l'étiquette 6HIS et dialysées de façon concomitante contre une solution tampon de 20 mM de Tris-HCl pH 7.5 contenant 150 mM de NaCl et 0.2 mM de TCEP (**tampon D**). Les membranes de dialyse (Spectrum<sup>TM</sup>) utilisées possèdent un seuil de poids moléculaire de 10 kDa.

Les protéines sont ensuite purifiées par chromatographie d'exclusion de taille (SEC) sur une colonne Hiload S200 (GE Healthcare) équilibrée avec le **tampon D**. Les fractions contenant la protéine pure sont ensuite concentrées dans un concentrateur Millipore (seuil de poids moléculaire de 10 kDa).

## (2) Expression et purification de la protéine C<sub>Δ28</sub> de NiV en cellules d'insecte

La protéine C<sub>Δ28</sub> est exprimée en baculovirus recombinants utilisés pour infecter des cellules d'insecte, Hi5. Les cellules d'insectes sont collectées 48 à 72h post-infection par centrifugation.

Elles sont ensuite lysées par sonication douce sur glace dans le **tampon A** contenant 0.5% de Brij-58. Le surnageant contenant les protéines solubles est ensuite collecté par centrifugation et appliqué sur une résine Ni<sup>2+</sup>NTA. Cette résine est ensuite intensément lavée par le **tampon B** ne contenant pas de détergent et l'élimination du Brij-58 est suivie par la méthode Bradford. Enfin la protéine d'intérêt est éluée dans le **tampon C**.

C<sub>Δ28</sub> est ensuite injectée sur une colonne Hiload Superdex 200 et purifiée par chromatographie d'exclusion de taille. Les fractions contenant la protéine pure sont ensuite concentrées dans un concentrateur Millipore (seuil de poids moléculaire de 10 kDa).

## (3) Expression et purification de la protéine P de NiV et SHCBP1

La protéine P de NiV a été purifiée selon le protocole précédemment décrit (Jensen *et al.*, 2020).

La protéine SHCBP1 a été purifiée de la même façon que les protéines C de NiV exprimées en système bactérien sans addition de Brij-58 0,5%. Les concentrateurs Millipore utilisés pour concentrer la protéine dans ce cas possèdent un seuil de poids moléculaire de 30 kDa.

## (4) Co-expression et purification de la protéine C<sub>Δ28</sub> avec les protéines du complexe ESCRT-I

Les plasmides coexprimant les protéines des complexes ESCRT-I/C<sub>Δ28</sub> ou VPS37B/C<sub>Δ28</sub> ont été transformés dans la souche *E. coli* BL21 STAR (DE3) et l'expression a été induite par 1 mM d'IPTG pendant la nuit à 18 °C.

Dans le cas de la co-expression de C<sub>Δ28</sub> et VPS37B, le complexe a été purifié par affinité sur résine Ni<sup>2+</sup>NTA (accrochage de 6HIS-C<sub>Δ28</sub>) ou par affinité sur résine streptactine (accrochage de Streptag II-VPS37B). Les cellules sont collectées par centrifugation et conservées à -20°C.

Dans le cas de la purification par affinité streptactine, les cellules sont ensuite lysées par sonication dans une solution tampon de 100 mM Tris-HCl à pH 8.0 contenant 300 mM de NaCl et 5 mM β-mercaptoéthanol (**tampon E**). Les protéines sont centrifugées et le surnageant contenant les protéines solubles est appliqué sur une résine streptactine équilibrée dans le **tampon E**. La résine est ensuite lavée avec le **tampon E**. Après lavage, les protéines



sont éluées par une solution tampon de 100 mM Tris-HCl à pH 8 contenant 150 mM de NaCl, 50 mM de Biotine et 0,2 mM de TCEP (**tampon F**).

Dans le cas de la co-expression de C<sub>Δ28</sub> et de l'ensemble des protéines du complexe ESCRT-I, le complexe est purifié par affinité sur résine Ni<sup>2+</sup>NTA (accrochage de 6HIS-C<sub>Δ28</sub>) ou par affinité sur résine glutathion (accrochage de GST-MVB12A).

Dans le cas de la purification par affinité glutathion, les cellules sont lysées par sonication dans une solution tampon de 50 mM Tris-HCl à pH 8.0 contenant 300 mM de NaCl et 5 mM de β-mercaptoéthanol (**tampon G**). Les protéines sont centrifugées et le surnageant contenant les protéines solubles est appliqué sur une résine GST équilibrée dans le **tampon G**. La résine est ensuite lavée avec le **tampon G**. Après lavage, les protéines sont éluées par une solution tampon de 50 mM Tris-HCl à pH 8 contenant 150 mM de NaCl et 50 mM de Glutathion réduit.

De même que pour SHCBP1, les complexes sont purifiés comme les protéines C de NiV exprimées en système bactérien sans addition de Brij-58 et avec une solution tampon dont le pH est de 8.



# Chapitre 7 : Conclusions et perspectives

## I. Comment expliquer l'insolubilité de la protéine C du virus Nipah ?

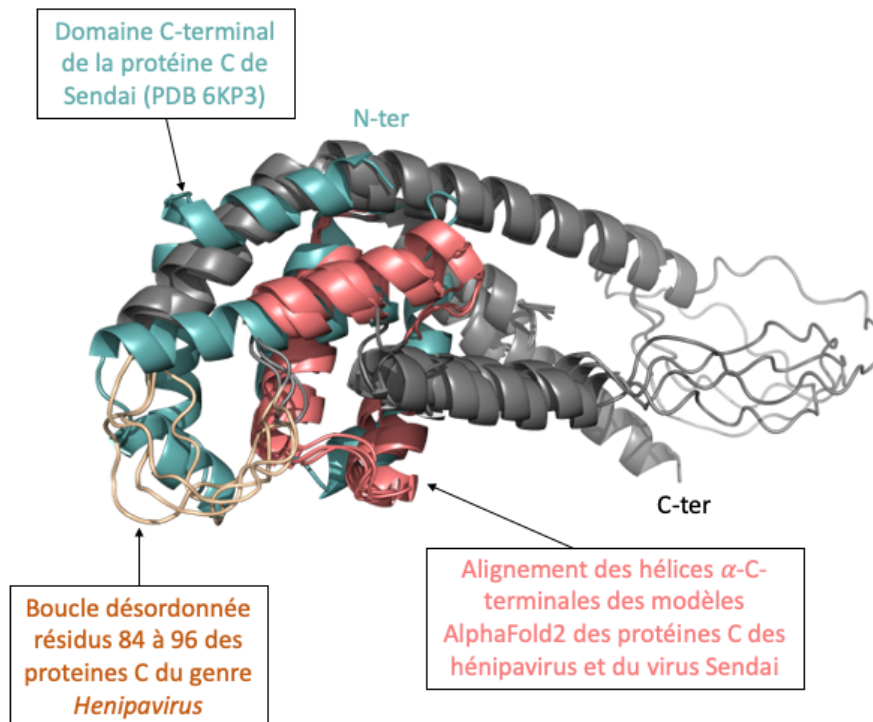
Malgré l'application de différentes méthodes pour identifier des fragments solubles de la protéine C du virus Nipah, elle semble encore aujourd'hui résister à sa caractérisation.

Nous avons constaté par une étude minutieuse des prédictions de désordre et de structures secondaires depuis la séquence en acides aminés que cette protéine présentait une organisation structurale partagée avec les autres protéines C des *Orthoparamyxovirinae* avec une région N-terminale désordonnée comportant un motif  $\alpha$ -MoRE suivi d'une région C-terminale repliée en hélices  $\alpha$ . Sur les constats de cette étude et sans détection de motifs hydrophobes en dehors du domaine structuré C-terminal, il semblait possible d'exprimer, de purifier et de caractériser la protéine C du virus Nipah.

Au cours de notre étude, nous avons pu observer que les fragments de protéines C auxquels nous avons retiré les 18 premiers (ou plus) résidus N-terminaux sont solubilisés par l'addition de détergents variés à la différence de la forme entière qui reste, quant à elle, insoluble dans toutes les conditions testées. De plus, la forme courte C $_{\Delta 28}$  de la protéine C de NiV interagit avec le domaine structuré du complexe ESCRT-I à la différence de la forme entière.

Il a été observé précédemment que la protéine C du SeV se lie aux membranes cellulaires par le biais d'une séquence d'adressage formée des 23 premiers aa N-terminaux, majoritairement hydrophobes et basiques (Marq *et al.*, 2007). Ce processus a été relié avec la capacité de la protéine C de SeV à favoriser la production de particules virales (VLP) (Marq *et al.*, 2007; Irie *et al.*, 2008b). La comparaison de cette séquence avec les 19 premiers résidus N-terminaux de la protéine C de NiV révèle que ces deux régions sont basiques et hydrophobes et prédites comme potentielles régions  $\alpha$ -MoRE.

Quel est le rôle des résidus N-terminaux de la protéine C du NiV dans le comportement membranaire de cette protéine ? Existe-t-il des mécanismes d'adressage ou d'interaction médiés par les résidus N-terminaux de cette protéine qui inhiberait sa fonction d'interaction avec la machinerie ESCRT et pourrait favoriser une forme insoluble et/ou instable de la protéine ?



**Figure 77 : Superposition des domaines C-terminaux de la protéine C du virus Sendai et des modèles AlphaFold2 des protéines C du genre *Henipavirus*.** Le cœur structural commun d'hélices  $\alpha$  des modèles AlphaFold2 des protéines C du genre *Henipavirus* est représenté en rose. La structure cristallographique du domaine C-terminal de la protéine C du virus Sendai (PDB 6KP3) superposée au cœur structural évoqué précédemment est représentée en bleu. Les hélices non superposées des modèles AlphaFold2 sont représentées en gris. La boucle désordonnée, comprise entre les résidus 84 à 96, des protéines C des *henipavirus* est représentée en orange.

Nous avons également constaté que la délétion des résidus N-terminaux de la protéine C et la délétion d'une ou deux hélices  $\alpha$  C-terminales, fortement hydrophobes, ne permet pas de solubiliser la protéine C de NiV en solution sans détergents et la protéine reste hautement instable en solution lors de leur élimination. Nous avons observé l'existence d'une boucle désordonnée entre les résidus 84 et 96, résidus majoritairement exposés au solvant pour la protéine C du NiV. Cette boucle n'est présente, d'après l'alignement des hélices  $\alpha$  C-terminales des modèles AlphaFold2 générés pour les protéines C des *Orthoparamyxovirinae* avec la structure du domaine C-terminal de la protéine C du virus Sendai (**Figure 77**), que pour les protéines C du genre *Henipavirus*. Nous émettons l'hypothèse que ces résidus pourraient être à l'origine de la haute instabilité de la protéine en solution et nous souhaitons étudier prochainement le comportement de la protéine  $C_{\Delta 91}$  délestée de ces résidus en solution. Il serait également intéressant de modifier les résidus en question afin d'évaluer leur impact sur la stabilité de la protéine.

## II. Caractérisation structurale des protéines P et C du virus de la toupeye

Il a été établi précédemment, par l'étude de motifs bien conservés chez les protéines P, que les phosphoprotéines des paramyxovirus sont homologues. Ces relations entre les protéines P sont difficiles à établir avec fiabilité à partir de la séquence en aa (Lo, Søggaard

and Karlin, 2014). Cependant, l'homologie a été confirmée par la résolution de plusieurs domaines conservés des phosphoprotéines de paramyxovirus. L'étude structurale de la protéine P du virus de la toupe (TupV) réalisée dans notre étude représente la première caractérisation d'une phosphoprotéine d'un *Narmovirus*. Cette étude nous a permis de renforcer les constats précédents en soulignant à nouveau l'existence d'une homologie détectée à partir de la structure avec les autres P des *Orthoparamyxovirinae*.

Nous avons de plus découvert que les structures tridimensionnelles des protéines C du virus Sendai (SeV) du genre *Respirovirus* et du TupV sont significativement similaires. Ces structures partagent un cœur structural commun formé par le repliement des 3 hélices C-terminales. Nos résultats confirment les prédictions bio-informatiques de la structure de la C du TupV dont l'organisation est identique à celle déterminée pour la protéine C du NiV.

À travers nos alignements de structure, nous avons constaté qu'un site de liaison au facteur cellulaire STAT1 est codé par exactement la même région d'ARN du gène P mais dans des cadres de lecture différents (région codante pour P ou C) chez les *Orthoparamyxovirinae*. Pour certains virus (NiV, HeV et MeV), ce site est monopartite et situé dans la région N-terminale désordonnée des protéines P/V/W (Rodriguez, Parisien and Horvath, 2002; Rodriguez, Wang and Horvath, 2003; Rodriguez, Cruz and Horvath, 2004; Devaux *et al.*, 2007; Jensen *et al.*, 2020), tandis que chez d'autres (SeV), il est bipartite, les deux parties étant situées dans le domaine C-terminal ordonné de la protéine C (donc codée dans le cadre +1 par rapport à P) (Oda *et al.*, 2017). D'un point de vue évolutif, ce constat est très intéressant. En effet, la conservation de ce motif sur le même emplacement, depuis l'un ou l'autre des produits du gène P, souligne probablement l'importance que l'inhibition de STAT1 se produise au même moment et au même endroit dans la cellule de telle sorte que cette inhibition puisse être avantageuse pour le second produit du gène.

De manière intéressante, il suffit d'une simple insertion d'un nucléotide en amont de la séquence codant pour le motif de liaison à STAT1 et d'une délétion compensatoire pour faire basculer cette séquence d'un cadre de lecture à l'autre. Il est alors possible d'imaginer que ce motif a basculé de l'une à l'autre des séquences codantes au cours de l'évolution expliquant les différences constatées entre les genres.

### III. Quelles sont les liens de phylogénie partagés par les protéines C des *Orthoparamyxovirinae* ?

#### ***(A) La difficulté de retracer la phylogénie récente des Orthoparamyxovirinae***

De nouveaux paramyxovirus sont identifiés chaque année dans le monde et cela entraîne la réévaluation continue de la classification de la sous-famille des *Orthoparamyxovirinae* par le Comité International de la Taxonomie des Virus (ICTV) avec la création régulière de nouveaux genres, dont les plus récents, en 2018, *Jeilongvirus*, *Narmovirus*, *Salemvirus*. Pour exemple, la découverte récente de paramyxovirus infectant des chauves-souris et des rongeurs en Amérique du Nord et du Sud (Rima *et al.*, 2018; Wells *et al.*, 2022) a entraîné la proposition en 2018 d'établir un nouveau genre dénommé *Shaanvirus*,

proposition qui n'a pas reçu l'approbation de l'ICTV. Ces virus ont finalement été classés dans le genre *Jeilongvirus* (Amarasinghe *et al.*, 2019).

Historiquement, la difficulté de l'établissement d'une phylogénie rigoureuse sur des virus extrêmement divergents, tels que les paramyxovirus, a été mise en lumière (McCarthy and Goodman, 2010). Cette observation reposait sur le constat des nombreuses discordances des phylogénies basées sur la séquence en nucléotides des gènes M, F et L avec ceux basés sur les gènes H/HN/G ou P (McCarthy and Goodman, 2010; Douglas, Drummond and Kingston, 2021; Larsen *et al.*, 2022). Les premiers permettent d'établir que les *Respirovirus* et *Aquaparamyxovirus* forment des clades basaux par rapport aux autres genres de la sous-famille bien distincts de la lignée regroupant les avulavirus, rubulavirus et metaparamyxovirus. Ces relations ne sont pas retrouvées dans le cas du gène P qui conserve le genre *Respirovirus* en position basale mais ne permet plus de distinguer les autres genres des lignées correspondant aux autres sous-familles.

Ces discordances illustrent les limitations des outils classiques de phylogénie lors de l'étude de l'histoire évolutive récente de virus à ARN (-) qui mutent rapidement. En ce sens, les gènes chevauchants, tels que ceux qui codent pour les protéines P et C des *Orthoparamyxovirinae*, peuvent présenter un intérêt particulier pour retracer une histoire évolutive récente. En effet, ces gènes expriment plusieurs protéines depuis une unique séquence d'ARN par décalage du cadre de lecture au cours de la traduction et, dans le cas de la protéine C des *Orthoparamyxovirinae*, les protéines sur-imprimées auraient émergé plus récemment que le gène ancestral P permettant donc de retracer l'histoire évolutive récente de la sous-famille. De plus, il est pensé que ces protéines pourraient être soumises à de fortes contraintes évolutives car limitant les mutations délétères qui affecteraient à la fois C et P. Cette hypothèse est encore discutée aujourd'hui (Simon-Loriere, Holmes and Pagán, 2013).

L'étude de l'histoire évolutive de ces protéines pourraient permettre d'étayer certaines théories évolutives proposées pour l'émergence des *Orthoparamyxovirinae*. Pour exemple, la phylogénie des protéines V a été récemment étudiée (Douglas, Drummond and Kingston, 2021) et montre une histoire évolutive complexe du gène P dont le mode d'expression serait passé plusieurs fois d'un mode d'expression P à un mode d'expression V. L'histoire évolutive des gènes P et C au sein de la famille des *Paramyxoviridae* est intéressante à retracer pour mieux comprendre comment ces virus sont corrélés les uns aux autres.

### ***(B) Les scénarios évolutifs de l'émergence de la région codante de la protéine C chez les Orthoparamyxovirinae***

De manière intéressante, nous avons observé que la protéine C est présente pour tous les virus de la sous-famille *Orthoparamyxovirinae*, à l'exception du genre *Ferlavirus* pour lequel on observe cependant la présence d'une protéine supplémentaire dont le gène est situé en amont de l'unité de transcription codant pour la phosphoprotéine, la protéine U.

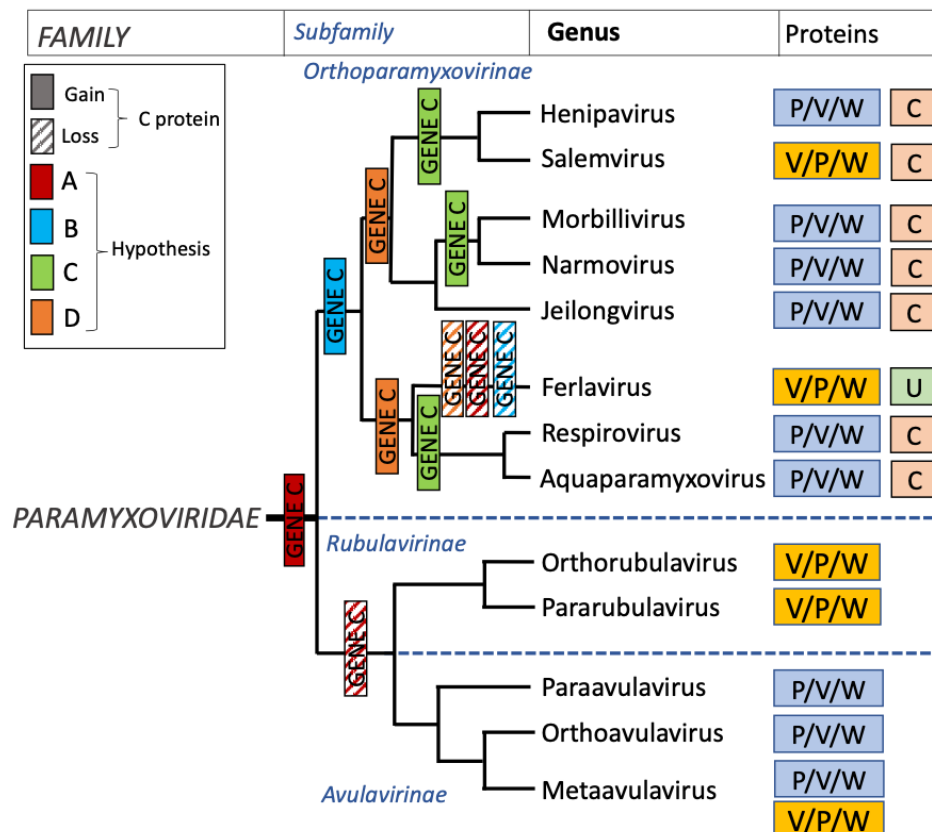
Au travers de l'analyse des séquences en aa et des prédictions de structures secondaires depuis les séquences en aa des protéines C des *Orthoparamyxovirinae*, nous avons confirmé l'existence d'une organisation structurale partagée entre les protéines C de la sous-famille qui comprennent toutes :

- (i) une partie N-terminale désordonnée présentant éventuellement une région complètement N-terminale organisée en  $\alpha$ -MoRE.
- (ii) un domaine C-terminal structuré comprenant un nombre variable d'hélices  $\alpha$ .

Toutes les protéines C sont codées dans le même cadre de lecture par rapport au gène P (+1 ORF). L'analyse de la séquence des protéines permet le classement des protéines C en trois groupes :

- le groupe I qui comprend les genres *Henipavirus* et *Jeilongvirus*,
- le groupe II qui comprend les genres *Narmovirus*, *Salemvirus* et *Morbillivirus*,
- le groupe III qui regroupe les genres *Respirovirus* et *Aquaparamyxovirus*.

Les protéines C du groupe III n'ont aucune similarité de séquences détectables avec les protéines C des groupes I et II, contrairement aux groupes I et II pour lesquels une homologie est détectée à partir de la séquence en aa.



**Figure 78 : Hypothèses évolutives de l'acquisition du cadre de lecture codant pour la protéine C chez les *Paramyxoviridae*.** Trois scénarios sont présentés sur un arbre phylogénétique construits sur le gène codant pour la protéine L. Ces scénarios sont représentés par les couleurs rouge, bleue, verte et orange. Le scénario rouge (A) représente l'hypothèse de l'acquisition du cadre de lecture codant pour la protéine C chez un ancêtre commun à tous les *Paramyxoviridae* qui aurait ensuite été perdu

dans la lignée des *Rubulavirinae* et *Avulavirinae*. Le scénario bleu (B) représente l'hypothèse de l'acquisition du cadre de lecture codant pour la protéine C chez un ancêtre commun à tous les *Orthoparamyxovirinae* qui aurait ensuite été perdu par les *Ferlavirus*. Le scénario vert (C) représente l'hypothèse de multiples apparitions de la région codante pour la C au sein des différents groupes I, II et III. Enfin, le scénario orange (D) représente l'hypothèse de l'acquisition de la région codante pour la protéine C dans l'ancêtre commun aux groupes I et II indépendamment au groupe III.

Dans l'état actuel des connaissances, quatre scénarios évolutifs peuvent être proposés illustrant l'émergence de la région codante de la protéine C (**Figure 78**) :

(a) La région codante de la protéine C est apparue par surimpression depuis le gène P dans l'ancêtre commun des *Paramyxoviridae* et a ensuite été perdue dans la lignée des *Rubulavirinae* et *Avulavirinae*. Ces protéines C ont ensuite divergé en séquence au-delà de la reconnaissance entre le groupe III et les groupes I et II. Dans ce scénario, toutes les protéines C sont homologues.

(b) La région codante de la protéine C est apparue par surimpression depuis le gène P dans l'ancêtre commun des *Orthoparamyxovirinae* et a ensuite divergé en séquence au-delà de la reconnaissance entre le groupe III et les groupes I et II. Dans ce scénario, toutes les protéines C sont homologues.

(c) La région codante C est apparue indépendamment chez l'ancêtre des groupes I, II et III. Dans ce scénario, les protéines C des groupes I, II et III ne sont pas homologues.

(d) La région codante C est apparue indépendamment chez l'ancêtre du groupe III et chez l'ancêtre des autres *Orthoparamyxovirinae*. Dans ce scénario, les protéines C du groupe III ne sont pas homologues à celle des autres *Orthoparamyxovirinae*.

### **(C) Les scénarios évolutifs a et b confirmés par nos résultats**

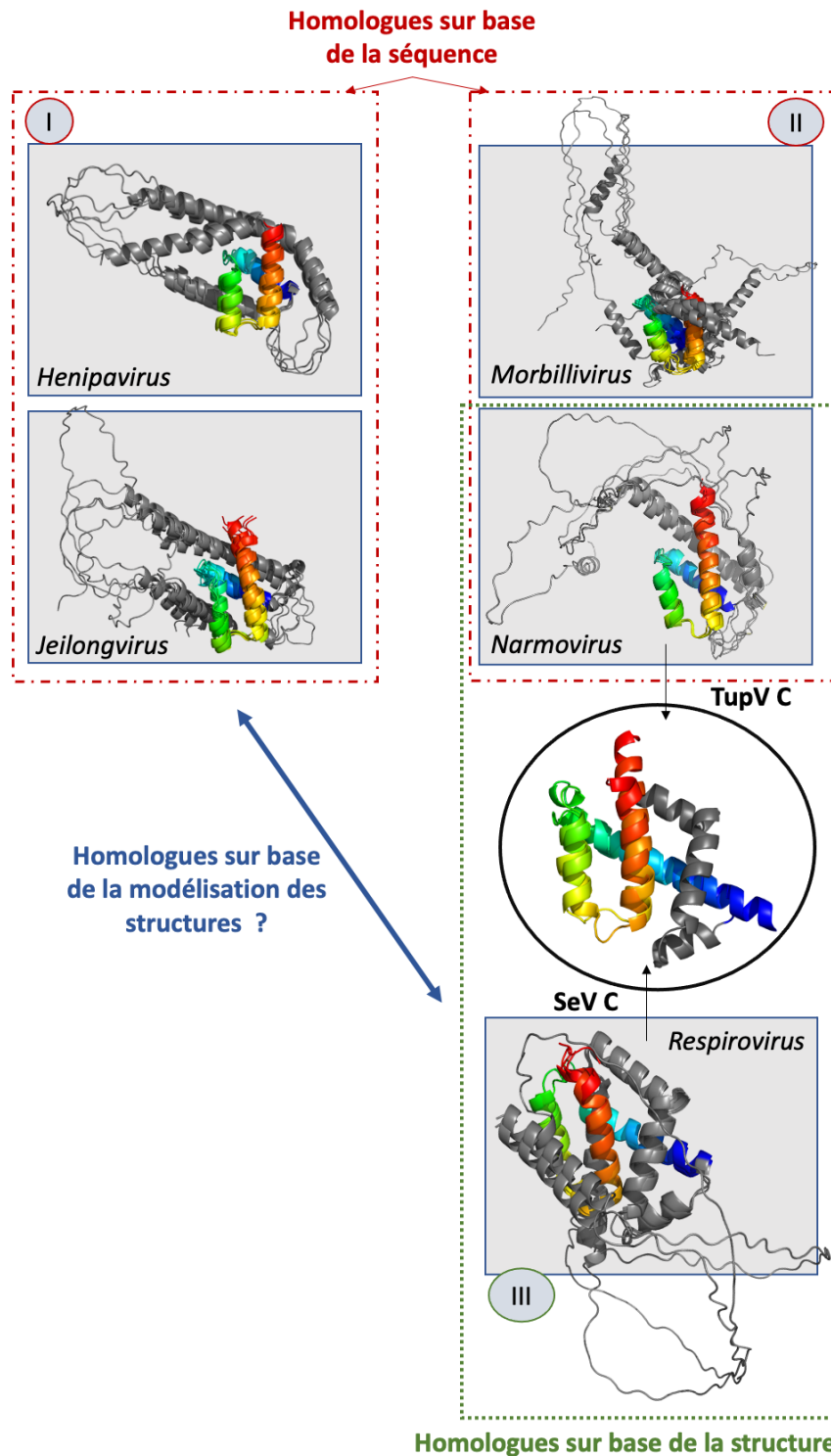
Dans ce contexte, nos résultats de caractérisation de la structure de la protéine C du virus de la toupe (TupV) ont permis de confirmer, d'une part, les prédictions bio-informatiques et, d'autre part, d'observer l'existence d'une homologie à partir de la structure avec la protéine C du virus Sendai. Les protéines C du groupe III et du groupe II sont donc, d'après nos résultats, bien homologues (**Figure 79**).

Ce repliement est également partagé avec les modèles tridimensionnels des protéines C des *Orthoparamyxovirinae* que nous avons établis par AlphaFold2 (**Figure 79**). Le cœur structural des trois hélices  $\alpha$  C-terminales est retrouvé pour tous les modèles (**Figure 79, rubans en couleur**) et est très conservé au sein des genres à la différence de la région N-terminale désordonnée très flexible (**Figure 79, gris**). Enfin, ce cœur structural s'aligne de façon significativement similaire avec les structures cristallographiques des domaines C-terminaux des protéines C du SeV et du TupV par comparaison *via* l'algorithme FATCAT (Li, Ye and Godzik, 2006) indiquant que ce cœur structural est, sur base de modélisation, bien conservé au sein de la sous-famille.

Nous émettons, en conséquence, l'hypothèse que les modèles prédits par AlphaFold2, bien que très incertains pour le domaine N-terminal de par sa grande flexibilité, suggèrent une



homologie sur base de la structure, pour les protéines C partagées par la sous-famille qui n'est pas retrouvée par leurs séquences.



**Figure 79 : Les protéines C des *Orthoparamyxovirinae* sont homologues par modélisation.** Nous avons superposé au sein de chaque genre (encadré gris) les régions C-terminales des modèles AlphaFold2 générés pour les protéines C des *Orthoparamyxovirinae*. L'étude des alignements multiples de séquences en aa des protéines C a permis de les classer en trois groupes partageant les plus grandes identités de séquences : le groupe I comprend les genres *Henipavirus*, *Jeilongvirus* et *Salemvirus* ; le groupe II comprend les genres *Morbillivirus* et *Narmovirus* et le groupe III les genres *Aquaparamyxovirus* et *Respirovirus*. Pour plus de clarté, nous n'avons pas inclus les modèles produits

*pour les salemvirus et aquaparaxyovirus dont les modèles sont prédits avec un indice de confiance faible ( nombre restreint de séquences disponibles). Les encadrés rouges regroupent les groupes I et II pour lesquelles une homologie est détectée par la méthode HHalign basée sur la séquence. Les genres Narmovirus et Respirovirus, non homologues à partir de la séquence, sont homologues en structure (cœur structural C-terminal commun formé de trois hélices  $\alpha$  colorées du rouge au bleu). Ce cœur structural est conservé pour tous les modèles. Ces modèles témoignent d'une homologie conservée à l'échelle de la sous-famille des Orthoparamyxovirinae.*

Cette étude représente la première évidence structurale de l'homologie suspectée précédemment. Ces résultats fournissent des indices d'ordre structural et de modélisation permettant de confirmer l'hypothèse évolutive a.

Nous espérons, notamment grâce à la production d'anticorps à domaine unique spécifiques à la protéine C du virus Nipah (NiV), confirmer prochainement ce scénario évolutif par la caractérisation de la structure d'une protéine C du groupe I.



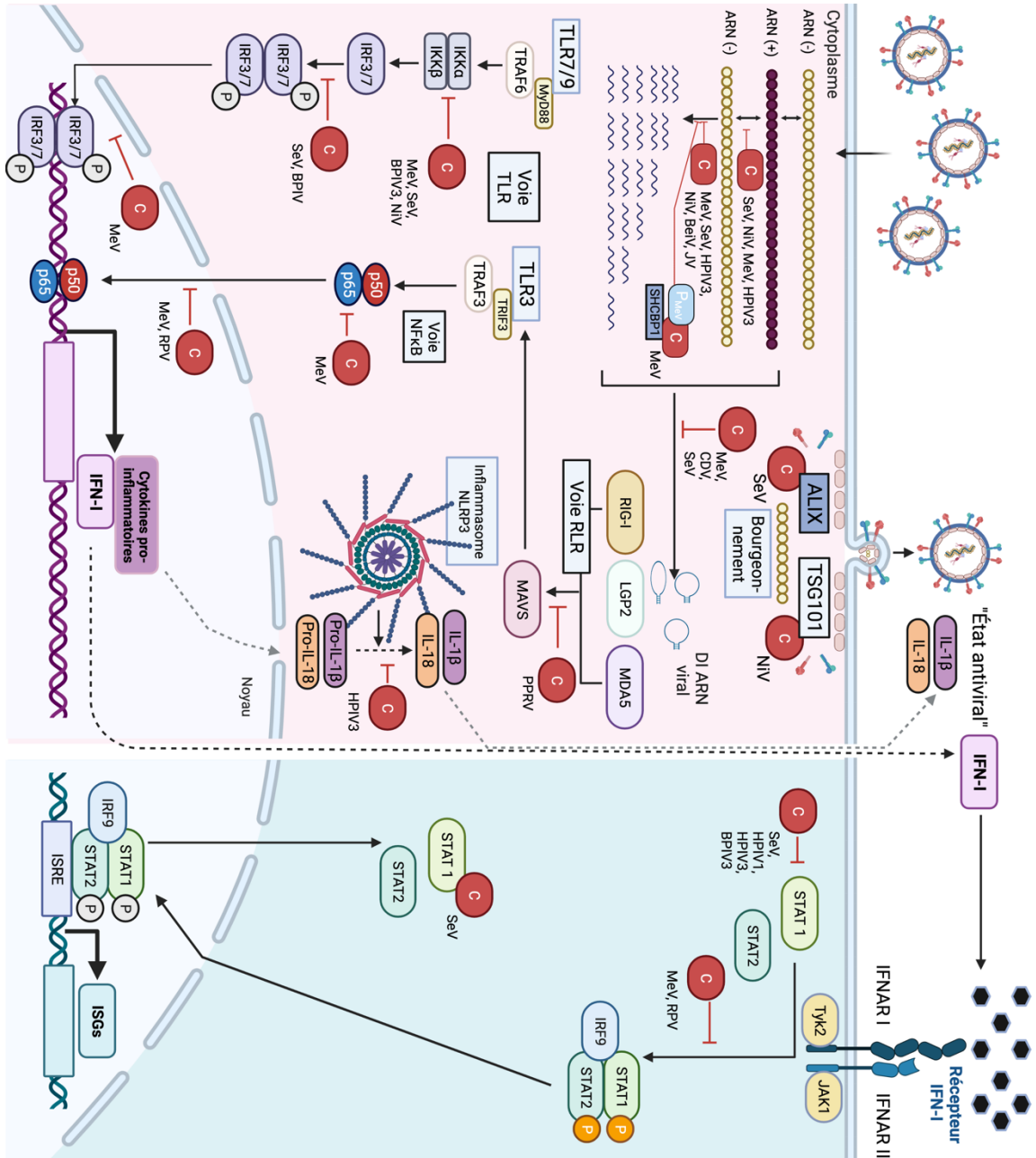


# ANNEXE 1

Tableau 4 : Abréviations des virus cités dans l'étude et numéros d'accèsions des protéines L et C citées.

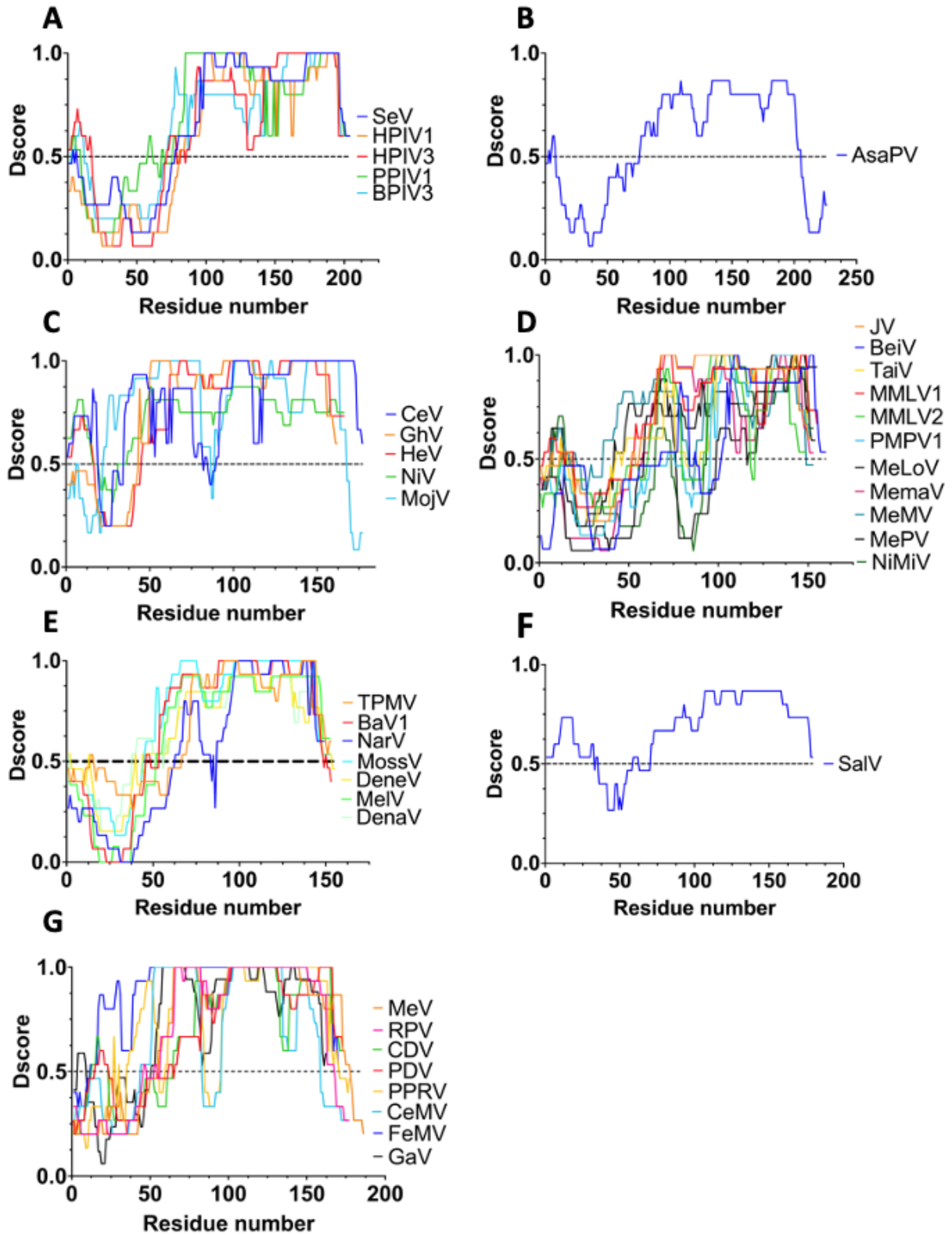
Protéines	Abbreviations	Virus	Genre	Genbank Accession Number	
C	SeV	Sendai virus	Respirovirus	ANQ45723.1	
L				CAA27273.1	
C	HPIV1	human Parainfluenza virus 1		NP_604434.1	
L				NP_604442.1	
C	PPIV1	porcine parainfluenza virus 1		AGR39559.1	
L				AGR39563.1	
C	HPIV3	human Parainfluenza virus 3		NP_599251.1	
L				NP_067153.2	
C	BPIV3	bovine parainfluenza virus 3		CAA68295.1	
L				UZC34578.1	
C	AsaPV	Atlantic salmon paramyxovirus		Aquaparamyxovirus	ABW38050.1
L				ABW38056.1	
C	CedV	Cedar virus	Henipavirus	AFP87276.1	
L				AFP87280.1	
C	MojV	Möjiāng virus		YP_009094092.1	
L				YP_009094096.1	
C	GhV2	Ghana virus 2		AFH96008.1	
L				AET43339.1	
C	HeV	Hendra virus		AAC04242.1	
L				NP_047113.3	
C	NIV	Nipah virus		AAV43913.1	
L				AAV43917.1	
C	LayV	Langya virus		UUV47238.1	
L				UUV47242.1	
C	PMPV1	Pohorje Myodes Paramyxovirus 1	Jeilongvirus	AVM86042.1	
L				AVM86048.1	
C	MMLV2	Mount Mavu Lophuromys virus 2		YP_009666850.1	
L				YP_009666855.1	
C	TaiV	Tailam virus		AEU08859.1	
L				AEU08865.1	
C	BeiV	Beilong virus		YP_512248.1	
L				YP_512254.1	
C	JV	J virus		YP_338079.1	
L				YP_009666842	
C	MMLV1	Mount Mavu Lophuromys virus 1		YP_009666842	
L				YP_009666846.1	
C	NIMiV	Ninove microtus virus	UQM99532.1		
L			UQM99539.1		
C	MeMV	Meliandou mastomys virus	UQM99587.1		
L			UQM99593.1		
C	MePV	Meliandou praomys virus	UQM99596.1		
L			UQM99602.1		
C	MeUV	Meliandou uranomys virus	UQM99505.1		
L			UQM99511.1		
C	MeLoV	Meliandou lophuromys virus	UQM99614.1		
L			UQM99620.1		
C	MemaV	Memana virus	UQM99605.1		
L			UQM99611.1		
C	TupV	Tuapia paramyxovirus	AAD28697.1		
L			AAF63393.1		
C	NarV	Nariva virus	YP_006347585.1		
L			YP_006347589.1		
C	BaV1	bank vole virus 1	ATW63186.1		
L			ATW63190.1		
C	DeneV	denestis virus	UQM99571.1		
L			UQM99576.1		
C	DenaV	denalis virus	UQM99579.1		
L			UQM99584.1		
C	MossV	Mossman virus	NP_958051.1		
L			NP_958055.1		
C	MeiV	meleucus virus	UQM99623.1		
L			UQM99628.1		
C	SalV	Salem virus	Salemvirus	AAF63743.1	
L			AFM97198.1		
C	RPV	rinderpest virus	Morbillivirus	ADF32062.1	
L				ADF32066.1	
C	MeV	measles virus		NP_056920.1	
L				BAA35122.1	
C	PDV	phocine distemper virus		BAA01208.1	
L				UJ92732.1	
C	CDV	canine distemper virus		NP_047203.1	
L				NP_047207.1	
C	CeMV	cetacean morbillivirus		NP_945026.1	
L				AYR16893.1	
C	PPRV	peste-des-petits-ruminants virus		YP_133824.1	
L				YP_133828.1	
C	FeMV	feline morbillivirus	AFH55514.1		
L			AFH55518.1		
C	GaV	Gierle apodemus virus	UQM99542.1		
L			UQM99547.1		

## ANNEXE 2



**Annexe 2 : Fonctions des protéines C des Orthoparamyxoviridae : dans la régulation et l'inhibition de la réplication et transcription virale, dans le contrôle du ratio génome/antigénome, dans l'antagonisme du système immunitaire et dans le bourgeonnement viral (figure adaptée de (Siering, Cattaneo and Pfaller, 2022) et (Pelissier, Iampietro and Horvat, 2019))**

## ANNEXE 3



Annexe 3 : Dscore calculée à partir de la séquence en acides aminés des protéines C des *Orthoparamyxovirinae* répartis par genre : (A) *Respirovirus* ; (B) *Aquaparamyxovirus* ; (C) *Henipavirus* ; (D) *Jeilongvirus* ; (E) *Narmovirus* ; (F) *Salemvirus* ; (G) *Morbillivirus*

## ANNEXE 4

**Tableau 5 : Résumé des conditions d'expression et de purification testées pour les constructions de C de NiV de l'étude.**

Construction	Vecteur	Expression	Induction	Étiquettes	Conditions de purification	Résultats
C <sub>FL</sub>	pET-M11	<i>E. coli</i> (BL21 STAR - Rosetta - Artic)	IPTG Auto-induction	6HIS NTD MBP	CB <sup>1</sup> Triton X-100 0.1% NP-40 1% NDSB-201 1M Brij-58 0.5% β-OG	Protéine insoluble
C <sub>FL</sub>	pET-M11	<i>E. coli</i> (BL21 STAR)	ITPG	6HIS-NTD	Urée 4M	Pas de repliement après élimination de l'urée
C <sub>FL</sub>	pFastBacHTA	Sf21 - Hi5	NA	6HIS NTD	CB <sup>1</sup> Brij-58 0.5%	Protéine insoluble
C <sub>Δ45</sub>	pET-M40	<i>E. coli</i> (BL21 STAR - Rosetta)	ITPG Auto-induction	MBP NTD	CB <sup>1</sup>	Protéine soluble mais précipite après clivage de la MBP
C <sub>Δ45</sub>	pET-28a	<i>E. coli</i> (BL21 STAR - Rosetta)	IPTG	6HIS-SUMO NTD	CB <sup>1</sup>	Protéine soluble mais précipite après clivage du 6HIS-SUMO
C <sub>Δ45</sub>	pET-M11	<i>E. coli</i> (BL21 STAR - Rosetta - Artic)	IPTG Auto-induction	6HIS NTD	CB <sup>1</sup> Triton X-100 0.1% NP-40 1% NDSB-201 1M Brij-58 0.5% β-OG	Protéines solubles en détergent mais précipitent lors de son élimination
C <sub>Δ27</sub> C <sub>Δ28</sub> C <sub>Δ35</sub> C <sub>Δ23</sub> C <sub>Δ18</sub> C <sub>Δ12</sub>	pET-28a	<i>E. coli</i> (BL21 STAR - Rosetta)	IPTG Auto-induction	6HIS NTD - BAP CTD	CB <sup>1</sup> NDSB-201 1M Triton X-100 0.1% Brij-58 0.5 %	Protéines solubles en détergent mais précipitent lors de son élimination. Purifiées en brij-58 pour C <sub>Δ28</sub> . Brij-58 résiduel dans la solution protéique. Contamination de l'échantillon par BirA.
C <sub>Δ28</sub> C <sub>Δ12</sub> C <sub>Δ35</sub>	pFastBacHTA	Sf21 - Hi5	NA	6HIS NTD - BAP CTD	CB <sup>1</sup> Brij-58 0.5%	Protéines solubles en détergent mais précipitent lors de son élimination. Purifiées en brij-58 pour C <sub>Δ28</sub> . Brij-58 résiduel dans la solution protéique.
C <sub>Δ91</sub>	pET-28a	<i>E. coli</i> (BL21 STAR)	IPTG	6HIS-NTD	CB <sup>1</sup>	Aucune expression



## Références bibliographiques

Aguilar, H.C. *et al.* (2010) ‘A Quantitative and Kinetic Fusion Protein-Triggering Assay Can Discern Distinct Steps in the Nipah Virus Membrane Fusion Cascade’, *Journal of Virology*, 84(16), pp. 8033–8041. Available at: <https://doi.org/10.1128/JVI.00469-10>.

Aguilar, H.C. *et al.* (2016) ‘Paramyxovirus Glycoproteins and the Membrane Fusion Process’, *Current Clinical Microbiology Reports*, 3(3), pp. 142–154. Available at: <https://doi.org/10.1007/s40588-016-0040-8>.

Ahmad, F. *et al.* (2022) ‘Discovery of Potential Antiviral Compounds against Hendra Virus by Targeting Its Receptor-Binding Protein (G) Using Computational Approaches’, *Molecules*, 27(2), p. 554. Available at: <https://doi.org/10.3390/molecules27020554>.

Alam, A.M. (2022) ‘Nipah virus, an emerging zoonotic disease causing fatal encephalitis’, *Clinical Medicine*, 22(4), pp. 348–352. Available at: <https://doi.org/10.7861/clinmed.2022-0166>.

Amarasinghe, G.K. *et al.* (2019) ‘Taxonomy of the order Mononegavirales: update 2019’, *Archives of Virology*, 164(7), pp. 1967–1980. Available at: <https://doi.org/10.1007/s00705-019-04247-4>.

An, Y. *et al.* (2011) ‘CoESPRIT: A Library-Based Construct Screening Method for Identification and Expression of Soluble Protein Complexes’, *PLoS ONE* e16261. Edited by B. Kobe, 6(2). Available at: <https://doi.org/10.1371/journal.pone.0016261>.

Andrejeva, J. *et al.* (2004) ‘The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN- promoter’, *Proceedings of the National Academy of Sciences*, 101(49), pp. 17264–17269. Available at: <https://doi.org/10.1073/pnas.0407639101>.

Ang, B.S.P., Lim, T.C.C. and Wang, L. (2018) ‘Nipah Virus Infection’, *Journal of Clinical Microbiology*. Edited by C.S. Kraft, 56(6), pp. e01875-17. Available at: <https://doi.org/10.1128/JCM.01875-17>.

Anthony, S.J. *et al.* (2013) ‘A Strategy To Estimate Unknown Viral Diversity in Mammals’, *mBio*. Edited by A. Moscona, 4(5), pp. e00598-13. Available at: <https://doi.org/10.1128/mBio.00598-13>.

Arunkumar, G. *et al.* (2019) ‘Outbreak Investigation of Nipah Virus Disease in Kerala, India, 2018’, *The Journal of Infectious Diseases*, 219(12), pp. 1867–1878. Available at: <https://doi.org/10.1093/infdis/jiy612>.

Audsley, M.D., Jans, D.A. and Moseley, G.W. (2016a) ‘Nucleocytoplasmic trafficking of Nipah virus W protein involves multiple discrete interactions with the nuclear import and export machinery’, *Biochemical and Biophysical Research Communications*, 479(3), pp. 429–433. Available at: <https://doi.org/10.1016/j.bbrc.2016.09.043>.

Audsley, M.D., Jans, D.A. and Moseley, G.W. (2016b) ‘Roles of nuclear trafficking in

infection by cytoplasmic negative-strand RNA viruses: paramyxoviruses and beyond', *Journal of General Virology*, 97(10), pp. 2463–2481. Available at: <https://doi.org/10.1099/jgv.0.000575>.

Baker, M.L., Schountz, T. and Wang, L.-F. (2013) 'Antiviral Immune Responses of Bats: A Review: Antiviral Immune Responses of Bats', *Zoonoses and Public Health*, 60(1), pp. 104–116. Available at: <https://doi.org/10.1111/j.1863-2378.2012.01528.x>.

Bankamp, B. *et al.* (2005) 'Identification of naturally occurring amino acid variations that affect the ability of the measles virus C protein to regulate genome replication and transcription', *Virology*, 336(1), pp. 120–129. Available at: <https://doi.org/10.1016/j.virol.2005.03.009>.

Bankamp, B. *et al.* (2008) 'Adaptation to cell culture induces functional differences in measles virus proteins', *Virology Journal*, 5(1), p. 129. Available at: <https://doi.org/10.1186/1743-422X-5-129>.

Baron, M.D. and Barrett, T. (2000) 'Rinderpest Viruses Lacking the C and V Proteins Show Specific Defects in Growth and Transcription of Viral RNAs', *Journal of Virology*, 74(6), pp. 2603–2611. Available at: <https://doi.org/10.1128/JVI.74.6.2603-2611.2000>.

Bartlett, E.J. *et al.* (2010) 'A novel human parainfluenza virus type 1 (HPIV1) with separated P and C genes is useful for generating C gene mutants for evaluation as live-attenuated virus vaccine candidates', *Vaccine*, 28(3), pp. 767–779. Available at: <https://doi.org/10.1016/j.vaccine.2009.10.069>.

Bernadó, P. *et al.* (2007) 'Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering', *Journal of the American Chemical Society*, 129(17), pp. 5656–5664. Available at: <https://doi.org/10.1021/ja069124n>.

Bernadó, P. and Svergun, D.I. (2012) 'Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering', *Mol. BioSyst.*, 8(1), pp. 151–167. Available at: <https://doi.org/10.1039/C1MB05275F>.

Bloyet, L.-M. *et al.* (2016) 'Modulation of Re-initiation of Measles Virus Transcription at Intergenic Regions by PXD to NTAIL Binding Strength', *PLoS Pathogens*. Edited by R. Cattaneo, 12(12), p. e1006058. Available at: <https://doi.org/10.1371/journal.ppat.1006058>.

Bloyet, L.-M. (2021) 'The Nucleocapsid of Paramyxoviruses: Structure and Function of an Encapsidated Template', *Viruses*, 13(12), p. 2465. Available at: <https://doi.org/10.3390/v13122465>.

Boonyaratanakornkit, J. *et al.* (2013) 'Alix Serves as an Adaptor That Allows Human Parainfluenza Virus Type 1 to Interact with the Host Cell ESCRT System', *PLoS ONE*. Edited by D.X. Liu, 8(3), p. e59462. Available at: <https://doi.org/10.1371/journal.pone.0059462>.

Bossart, K.N. *et al.* (2009) 'A Neutralizing Human Monoclonal Antibody Protects against Lethal Disease in a New Ferret Model of Acute Nipah Virus Infection', *PLoS Pathogens*. Edited by C.F. Basler, 5(10), p. e1000642. Available at: <https://doi.org/10.1371/journal.ppat.1000642>.

<https://doi.org/10.1371/journal.ppat.1000642>.

Bossart, K.N. *et al.* (2012) ‘A Hendra Virus G Glycoprotein Subunit Vaccine Protects African Green Monkeys from Nipah Virus Challenge’, *Science Translational Medicine*, 4(146). Available at: <https://doi.org/10.1126/scitranslmed.3004241>.

Bourhis, J.-M. *et al.* (2022) ‘Structural Dynamics of the C-terminal X Domain of Nipah and Hendra Viruses Controls the Attachment to the C-terminal Tail of the Nucleocapsid Protein’, *Journal of Molecular Biology*, 434(10), p. 167551. Available at: <https://doi.org/10.1016/j.jmb.2022.167551>.

Bowden, T.A. *et al.* (2008) ‘Crystal Structure and Carbohydrate Analysis of Nipah Virus Attachment Glycoprotein: a Template for Antiviral and Vaccine Design’, *Journal of Virology*, 82(23), pp. 11628–11636. Available at: <https://doi.org/10.1128/JVI.01344-08>.

Bowden, T.A., Jones, E.Y. and Stuart, D.I. (2011) ‘Cells under siege: Viral glycoprotein interactions at the cell surface’, *Journal of Structural Biology*, 175(2), pp. 120–126. Available at: <https://doi.org/10.1016/j.jsb.2011.03.016>.

Boxer, E.L., Nanda, S.K. and Baron, M.D. (2009) ‘The rinderpest virus non-structural C protein blocks the induction of type 1 interferon’, *Virology*, 385(1), pp. 134–142. Available at: <https://doi.org/10.1016/j.virol.2008.11.022>.

Brende, B. *et al.* (2017) ‘CEPI—a new global R&D organisation for epidemic preparedness and response’, *The Lancet*, 389(10066), pp. 233–235. Available at: [https://doi.org/10.1016/S0140-6736\(17\)30131-9](https://doi.org/10.1016/S0140-6736(17)30131-9).

Broder, C.C., Weir, D.L. and Reid, P.A. (2016) ‘Hendra virus and Nipah virus animal vaccines’, *Vaccine*, 34(30), pp. 3525–3534. Available at: <https://doi.org/10.1016/j.vaccine.2016.03.075>.

Broz, P. and Dixit, V.M. (2016) ‘Inflammasomes: mechanism of assembly, regulation and signalling’, *Nature Reviews Immunology*, 16(7), pp. 407–420. Available at: <https://doi.org/10.1038/nri.2016.58>.

Brückner, A. *et al.* (2009) ‘Yeast Two-Hybrid, a Powerful Tool for Systems Biology’, *International Journal of Molecular Sciences*, 10(6), pp. 2763–2788. Available at: <https://doi.org/10.3390/ijms10062763>.

Bruhn, J.F. *et al.* (2014) ‘Crystal Structure of the Nipah Virus Phosphoprotein Tetramerization Domain’, *Journal of Virology*, 88(1), pp. 758–762. Available at: <https://doi.org/10.1128/JVI.02294-13>.

Caignard, G., Vidalain, P.-O. and Tangy, F. (2012) ‘Résistance des Paramyxoviridae aux interférons de type I : mécanismes d’échappement et interactions virus-hôte’, 16, p. 13.

Carty, M., Guy, C. and Bowie, A.G. (2021) ‘Detection of Viral Infections by Innate Immunity’, *Biochemical Pharmacology*, 183, p. 114316. Available at: <https://doi.org/10.1016/j.bcp.2020.114316>.

- Chadha, M.S. *et al.* (2006) ‘Nipah Virus-associated Encephalitis Outbreak, Siliguri, India’, *Emerging Infectious Diseases*, 12(2), pp. 235–240. Available at: <https://doi.org/10.3201/eid1202.051247>.
- Chan, Y.P. *et al.* (2001) ‘Complete nucleotide sequences of Nipah virus isolates from Malaysia’, *Journal of General Virology*, 82(9), pp. 2151–2155. Available at: <https://doi.org/10.1099/0022-1317-82-9-2151>.
- Chang, A. and Dutch, R.E. (2012) ‘Paramyxovirus Fusion and Entry: Multiple Paths to a Common End’, *Viruses*, 4(4), pp. 613–636. Available at: <https://doi.org/10.3390/v4040613>.
- Chen, S. *et al.* (2010) ‘Heterocellular induction of interferon by negative-sense RNA viruses’, *Virology*, 407(2), pp. 247–255. Available at: <https://doi.org/10.1016/j.virol.2010.08.008>.
- Choudhary, O.P. *et al.* (2022) ‘Spillover zoonotic “Langya virus”: is it a matter of concern?’, *Veterinary Quarterly*, 42(1), pp. 172–174. Available at: <https://doi.org/10.1080/01652176.2022.2117874>.
- Chua, K.B. (2003) ‘Nipah virus outbreak in Malaysia’, *Journal of Clinical Virology*, 26(3), pp. 265–275. Available at: [https://doi.org/10.1016/S1386-6532\(02\)00268-8](https://doi.org/10.1016/S1386-6532(02)00268-8).
- Ciancanelli, M.J. *et al.* (2009) ‘Nipah Virus Sequesters Inactive STAT1 in the Nucleus via a P Gene-Encoded Mechanism’, *Journal of Virology*, 83(16), pp. 7828–7841. Available at: <https://doi.org/10.1128/JVI.02610-08>.
- Ciancanelli, M.J. and Basler, C.F. (2006) ‘Mutation of YMYL in the Nipah Virus Matrix Protein Abrogates Budding and Alters Subcellular Localization’, *Journal of Virology*, 80(24), pp. 12070–12078. Available at: <https://doi.org/10.1128/JVI.01743-06>.
- Cleaveland, S., Laurenson, M.K. and Taylor, L.H. (2001) ‘Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence’, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. Edited by M.E.J. Woolhouse and C. Dye, 356(1411), pp. 991–999. Available at: <https://doi.org/10.1098/rstb.2001.0889>.
- Communie, G. *et al.* (2013) ‘Atomic Resolution Description of the Interaction between the Nucleoprotein and Phosphoprotein of Hendra Virus’, *PLoS Pathogens*. Edited by F.A. Rey, 9(9), p. e1003631. Available at: <https://doi.org/10.1371/journal.ppat.1003631>.
- Corrêa, D.H.A. and Ramos, C.H.I. (2009) ‘The use of circular dichroism spectroscopy to study protein folding, form and function’, p. 11.
- Cunningham, A.A., Daszak, P. and Wood, J.L.N. (2017) ‘One Health, emerging infectious diseases and wildlife: two decades of progress?’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1725), p. 20160167. Available at: <https://doi.org/10.1098/rstb.2016.0167>.
- Curran, J. (1998) ‘A Role for the Sendai Virus P Protein Trimer in RNA Synthesis’, *Journal of Virology*, 72(5), pp. 4274–4280. Available at: <https://doi.org/10.1128/JVI.72.5.4274-4280.1998>.

Curran, J.A., Richardson, C. and Kolakofsky, D. (1986) ‘Ribosomal initiation at alternate AUGs on the Sendai virus P/C mRNA’, *Journal of Virology*, 57(2), pp. 684–687. Available at: <https://doi.org/10.1128/jvi.57.2.684-687.1986>.

Daszak, P. *et al.* (2011) ‘Pteropid Bats are Confirmed as the Reservoir Hosts of Henipaviruses: A Comprehensive Experimental Study of Virus Transmission’, *The American Journal of Tropical Medicine and Hygiene*, 85(5), pp. 946–951. Available at: <https://doi.org/10.4269/ajtmh.2011.10-0567>.

Dawes, B.E. *et al.* (2018) ‘Favipiravir (T-705) protects against Nipah virus infection in the hamster model’, *Scientific Reports*, 8(1), p. 7604. Available at: <https://doi.org/10.1038/s41598-018-25780-3>.

Devaux, P. *et al.* (2007) ‘Tyrosine 110 in the measles virus phosphoprotein is required to block STAT1 phosphorylation’, *Virology*, 360(1), pp. 72–83. Available at: <https://doi.org/10.1016/j.virol.2006.09.049>.

Diederich, S., Thiel, L. and Maisner, A. (2008) ‘Role of endocytosis and cathepsin-mediated activation in Nipah virus entry’, *Virology*, 375(2), pp. 391–400. Available at: <https://doi.org/10.1016/j.virol.2008.02.019>.

Donnelly, C.M. *et al.* (2021) ‘The Structural Features of *Henipavirus* Matrix Protein Driving Intracellular Trafficking’, *Viral Immunology*, 34(1), pp. 27–40. Available at: <https://doi.org/10.1089/vim.2020.0056>.

Douglas, J., Drummond, A.J. and Kingston, R.L. (2021) ‘Evolutionary history of cotranscriptional editing in the paramyxoviral phosphoprotein gene’, *Virus Evolution*, 7(1), p. veab028. Available at: <https://doi.org/10.1093/ve/veab028>.

Du Pont, V., Jiang, Y. and Plemper, R.K. (2019) ‘Bipartite interface of the measles virus phosphoprotein X domain with the large polymerase protein regulates viral polymerase dynamics’, *PLoS Pathogens*. Edited by S.P.J. Whelan, 15(8), p. e1007995. Available at: <https://doi.org/10.1371/journal.ppat.1007995>.

Dubach, V.R.A. and Guskov, A. (2020) ‘The Resolution in X-ray Crystallography and Single-Particle Cryogenic Electron Microscopy’, *Crystals*, 10(7), p. 580. Available at: <https://doi.org/10.3390/cryst10070580>.

Durbin, A.P. *et al.* (1999) ‘Mutations in the C, D, and V Open Reading Frames of Human Parainfluenza Virus Type 3 Attenuate Replication in Rodents and Primates’, *Virology*, 261(2), pp. 319–330. Available at: <https://doi.org/10.1006/viro.1999.9878>.

Eby, P. *et al.* (2022) ‘Pathogen spillover driven by rapid changes in bat ecology’, *Nature* [Preprint]. Available at: <https://doi.org/10.1038/s41586-022-05506-2>.

El Najjar, F., Schmitt, A. and Dutch, R. (2014) ‘Paramyxovirus Glycoprotein Incorporation, Assembly and Budding: A Three Way Dance for Infectious Particle Production’, *Viruses*, 6(8), pp. 3019–3054. Available at: <https://doi.org/10.3390/v6083019>.

Emsley, P. *et al.* (2010) 'Features and development of *Coot*', *Acta Crystallographica Section D Biological Crystallography*, 66(4), pp. 486–501. Available at: <https://doi.org/10.1107/S0907444910007493>.

Enchéry, F. *et al.* (2021) 'Nipah virus W protein harnesses nuclear 14-3-3 to inhibit NF- $\kappa$ B-induced proinflammatory response', *Communications Biology*, 4(1), p. 1292. Available at: <https://doi.org/10.1038/s42003-021-02797-5>.

Engeland, C.E. *et al.* (2017) 'A Tupaia paramyxovirus vector system for targeting and transgene expression', *Journal of General Virology*, 98(9), pp. 2248–2257. Available at: <https://doi.org/10.1099/jgv.0.000887>.

Fitzgerald, K.A. and Kagan, J.C. (2020) 'Toll-like Receptors and the Control of Immunity', *Cell*, 180(6), pp. 1044–1066. Available at: <https://doi.org/10.1016/j.cell.2020.02.041>.

Floden, E.W. *et al.* (2016) 'PSI/TM-Coffee: a web server for fast and accurate multiple sequence alignments of regular and transmembrane proteins using homology extension on reduced databases', *Nucleic Acids Research*, 44(W1), pp. W339–W343. Available at: <https://doi.org/10.1093/nar/gkw300>.

Flower, T.G. *et al.* (2020) 'A helical assembly of human ESCRT-I scaffolds reverse-topology membrane scission', *Nature Structural & Molecular Biology*, 27, pp. 570–580. Available at: <https://doi.org/10.1038/s41594-020-0426-4>.

Garcin, D. *et al.* (2001) 'Longer and Shorter Forms of Sendai Virus C Proteins Play Different Roles in Modulating the Cellular Antiviral Response', *Journal of Virology*, 75(15), pp. 6800–6807. Available at: <https://doi.org/10.1128/JVI.75.15.6800-6807.2001>.

Garcin, D. *et al.* (2002) 'All Four Sendai Virus C Proteins Bind Stat1, but only the Larger Forms also Induce Its Mono-ubiquitination and Degradation', *Virology*, 295(2), pp. 256–265. Available at: <https://doi.org/10.1006/viro.2001.1342>.

Garcin, D. *et al.* (2004) 'A Short Peptide at the Amino Terminus of the Sendai Virus C Protein Acts as an Independent Element That Induces STAT1 Instability', *Journal of Virology*, 78(16), pp. 8799–8811. Available at: <https://doi.org/10.1128/JVI.78.16.8799-8811.2004>.

Garcin, D., Curran, J. and Kolakofsky, D. (2000) 'Sendai Virus C Proteins Must Interact Directly with Cellular Components To Interfere with Interferon Action', *Journal of Virology*, 74(19), pp. 8823–8830. Available at: <https://doi.org/10.1128/JVI.74.19.8823-8830.2000>.

Garcin, D., Itoh, M. and Kolakofsky, D. (1997) 'A Point Mutation in the Sendai Virus Accessory C Proteins Attenuates Virulence for Mice, but Not Virus Growth in Cell Culture', *Virology*, 238(2), pp. 424–431. Available at: <https://doi.org/10.1006/viro.1997.8836>.

Gerard, F.C.A. *et al.* (2009) 'Modular Organization of Rabies Virus Phosphoprotein', *Journal of Molecular Biology*, 388(5), pp. 978–996. Available at: <https://doi.org/10.1016/j.jmb.2009.03.061>.

Gerlier, D. and Lyles, D.S. (2011) 'Interplay between Innate Immunity and Negative-Strand

RNA Viruses: towards a Rational Model', *Microbiology and Molecular Biology Reviews*, 75(3), pp. 468–490. Available at: <https://doi.org/10.1128/MMBR.00007-11>.

Goldsmith, C.S. *et al.* (2003) 'Elucidation of Nipah virus morphogenesis and replication using ultrastructural and molecular approaches', *Virus Research*, 92(1), pp. 89–98. Available at: [https://doi.org/10.1016/S0168-1702\(02\)00323-4](https://doi.org/10.1016/S0168-1702(02)00323-4).

Gotoh, B. *et al.* (1999) 'Knockout of the Sendai virus C gene eliminates the viral ability to prevent the interferon- $\alpha/\beta$ -mediated responses', *FEBS Letters*, 459(2), pp. 205–210. Available at: [https://doi.org/10.1016/S0014-5793\(99\)01241-7](https://doi.org/10.1016/S0014-5793(99)01241-7).

Gotoh, B. *et al.* (2003) 'The C-terminal half-fragment of the Sendai virus C protein prevents the gamma-activated factor from binding to a gamma-activated sequence site', *Virology*, 316(1), pp. 29–40. Available at: [https://doi.org/10.1016/S0042-6822\(03\)00590-7](https://doi.org/10.1016/S0042-6822(03)00590-7).

Gotoh, B., Takeuchi, K. and Komatsu, T. (2004) 'Inhibition of the gamma interferon response by a Sendai virus C protein mutant with no STAT1-binding ability', *FEBS Letters*, 567(2–3), pp. 291–296. Available at: <https://doi.org/10.1016/j.febslet.2004.05.001>.

Grogan, C.C. and Moyer, S.A. (2001) 'Sendai Virus Wild-Type and Mutant C Proteins Show a Direct Correlation between L Polymerase Binding and Inhibition of Viral RNA Synthesis', *Virology*, 288(1), pp. 96–108. Available at: <https://doi.org/10.1006/viro.2001.1068>.

Guseva, S. *et al.* (2020) 'Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly', *Science Advances*, 6(14), p. eaaz7095. Available at: <https://doi.org/10.1126/sciadv.aaz7095>.

Gutsche, I. *et al.* (2015) 'Near-atomic cryo-EM structure of the helical measles virus nucleocapsid', *Science*, 348(6235), pp. 704–707. Available at: <https://doi.org/10.1126/science.aaa5137>.

Harcourt, B.H. *et al.* (2005) 'Genetic Characterization of Nipah Virus, Bangladesh, 2004', *Emerging Infectious Diseases*, 11(10), pp. 1594–1597. Available at: <https://doi.org/10.3201/eid1110.050513>.

Harrison, M.S., Sakaguchi, T. and Schmitt, A.P. (2010) 'Paramyxovirus assembly and budding: Building particles that transmit infections', *The International Journal of Biochemistry & Cell Biology*, 42(9), pp. 1416–1429. Available at: <https://doi.org/10.1016/j.biocel.2010.04.005>.

Hausmann, S. *et al.* (1999) 'Two Nucleotides Immediately Upstream of the Essential A<sub>6</sub>G<sub>3</sub> Slippery Sequence Modulate the Pattern of G Insertions during Sendai Virus mRNA Editing', *Journal of Virology*, 73(1), pp. 343–351. Available at: <https://doi.org/10.1128/JVI.73.1.343-351.1999>.

Hayman, D.T.S. *et al.* (2008) 'Evidence of Henipavirus Infection in West African Fruit Bats', *PLoS ONE*. Edited by J.M. Montgomery, 3(7), p. e2739. Available at: <https://doi.org/10.1371/journal.pone.0002739>.

Horie, R. *et al.* (2016) 'Region of Nipah virus C protein responsible for shuttling between the

cytoplasm and nucleus', *Virology*, 497, pp. 294–304. Available at: <https://doi.org/10.1016/j.virol.2016.07.013>.

Horikami, S.M. *et al.* (1997) 'The Sendai Virus C Protein Binds the L Polymerase Protein to Inhibit Viral RNA Synthesis', *Virology*, 235(2), pp. 261–270. Available at: <https://doi.org/10.1006/viro.1997.8702>.

Hudacek, A.W., Navaratnarajah, C.K. and Cattaneo, R. (2013) 'Development of measles virus-based shielded oncolytic vectors: suitability of other paramyxovirus glycoproteins', *Cancer Gene Therapy*, 20(2), pp. 109–116. Available at: <https://doi.org/10.1038/cgt.2012.92>.

Iehlé, C. *et al.* (2007) 'Henipavirus and Tioman Virus Antibodies in Pteropodid Bats, Madagascar', *Emerging Infectious Diseases*, 13(1), pp. 159–161. Available at: <https://doi.org/10.3201/eid1301.060791>.

Iida, Y. and Masuda, T. (1996) 'Strength of translation initiation signal sequence of mRNA as studied by quantification method: effect of nucleotide substitutions upon translation efficiency in rat preproinsulin mRNA', *Nucleic Acids Research*, 24(17), pp. 3313–3316. Available at: <https://doi.org/10.1093/nar/24.17.3313>.

Irie, T. *et al.* (2008a) 'Paramyxovirus Sendai virus C proteins are essential for maintenance of negative-sense RNA genome in virus particles', *Virology*, 374(2), pp. 495–505. Available at: <https://doi.org/10.1016/j.virol.2008.01.004>.

Irie, T. *et al.* (2008b) 'Recruitment of Alix/AIP1 to the plasma membrane by Sendai virus C protein facilitates budding of virus-like particles', *Virology*, 371(1), pp. 108–120. Available at: <https://doi.org/10.1016/j.virol.2007.09.020>.

Irie, T., Yoshida, A. and Sakaguchi, T. (2013) 'Clustered Basic Amino Acids of the Small Sendai Virus C Protein Y1 Are Critical to Its Ran GTPase-Mediated Nuclear Localization', *PLoS ONE*. Edited by H.L. Ploegh, 8(8), p. e73740. Available at: <https://doi.org/10.1371/journal.pone.0073740>.

Iseni, F. *et al.* (2002) 'Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids', *RNA*, 8(8), pp. 1056–1067. Available at: <https://doi.org/10.1017/S1355838202029977>.

Ito, M. *et al.* (2013) 'Measles Virus Nonstructural C Protein Modulates Viral RNA Polymerase Activity by Interacting with Host Protein SHCBP1', *Journal of Virology*, 87(17), pp. 9633–9642. Available at: <https://doi.org/10.1128/JVI.00714-13>.

Jensen, M.R. *et al.* (2020) 'Structural Description of the Nipah Virus Phosphoprotein and Its Interaction with STAT1', *Biophysical Journal*, 118(10), pp. 2470–2488. Available at: <https://doi.org/10.1016/j.bpj.2020.04.010>.

Jones, D.T. (1999) 'Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices', *Journal of Molecular Biology*, 292(2), pp. 195–202. Available at: <https://doi.org/10.1006/jmbi.1999.3091>.

Jordan, I.K., Sutter IV, B.A. and McClure, M.A. (2000) 'Molecular Evolution of the



Paramyxoviridae and Rhabdoviridae Multiple-Protein-Encoding P Gene', *Molecular Biology and Evolution*, 17(1), pp. 75–086. Available at: <https://doi.org/10.1093/oxfordjournals.molbev.a026240>.

Kabsch, W. (2010) 'XDS', *Acta Crystallographica Section D Biological Crystallography*, 66(2), pp. 125–132. Available at: <https://doi.org/10.1107/S0907444909047337>.

Kalodimou, G. *et al.* (2019) 'A Soluble Version of Nipah Virus Glycoprotein G Delivered by Vaccinia Virus MVA Activates Specific CD8 and CD4 T Cells in Mice', *Viruses*, 12(1), p. 26. Available at: <https://doi.org/10.3390/v12010026>.

Karsunke, J. *et al.* (2019) 'W protein expression by Newcastle disease virus', *Virus Research*, 263, pp. 207–216. Available at: <https://doi.org/10.1016/j.virusres.2019.02.003>.

Kato, A. *et al.* (2004) 'Characterization of the Amino Acid Residues of Sendai Virus C Protein That Are Critically Involved in Its Interferon Antagonism and RNA Synthesis Down-Regulation', *Journal of Virology*, 78(14), pp. 7443–7454. Available at: <https://doi.org/10.1128/JVI.78.14.7443-7454.2004>.

Kato, A. *et al.* (2007) 'Importance of the Anti-Interferon Capacity of Sendai Virus C Protein for Pathogenicity in Mice', *Journal of Virology*, 81(7), pp. 3264–3271. Available at: <https://doi.org/10.1128/JVI.02590-06>.

Ker, D.-S. *et al.* (2021) 'CryoEM structure of the Nipah virus nucleocapsid assembly', *PLOS Pathogens*. Edited by F.A. Rey, 17(7), p. e1009740. Available at: <https://doi.org/10.1371/journal.ppat.1009740>.

Kessler, M.K. *et al.* (2018) 'Changing resource landscapes and spillover of henipaviruses', *Annals of the New York Academy of Sciences*, 1429(1), pp. 78–99. Available at: <https://doi.org/10.1111/nyas.13910>.

Kikhney, A.G. and Svergun, D.I. (2015) 'A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins', *FEBS Letters*, 589(19PartA), pp. 2570–2577. Available at: <https://doi.org/10.1016/j.febslet.2015.08.027>.

Kitagawa, Y. *et al.* (2011) 'A Tryptophan-Rich Motif in the Human Parainfluenza Virus Type 2 V Protein Is Critical for the Blockade of Toll-Like Receptor 7 (TLR7)- and TLR9-Dependent Signaling', *Journal of Virology*, 85(9), pp. 4606–4611. Available at: <https://doi.org/10.1128/JVI.02012-10>.

Knipe, D.M. and Howley, P.M. (eds) (2013) *Fields virology*. 6th ed. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins Health.

Kolakofsky, D. *et al.* (2004) 'Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis', *Virology*, 318(2), pp. 463–473. Available at: <https://doi.org/10.1016/j.virol.2003.10.031>.

Kolakofsky, D. *et al.* (2005) 'Paramyxovirus mRNA editing, the “rule of six” and error catastrophe: a hypothesis', *Journal of General Virology*, 86(7), pp. 1869–1877. Available at: <https://doi.org/10.1099/vir.0.80986-0>.

Komatsu, T. *et al.* (2004) ‘C and V proteins of Sendai virus target signaling pathways leading to IRF-3 activation for the negative regulation of interferon- $\beta$  production’, *Virology*, 325(1), pp. 137–148. Available at: <https://doi.org/10.1016/j.virol.2004.04.025>.

Komatsu, T., Takeuchi, K. and Gotoh, B. (2007) ‘Bovine parainfluenza virus type 3 accessory proteins that suppress beta interferon production’, *Microbes and Infection*, 9(8), pp. 954–962. Available at: <https://doi.org/10.1016/j.micinf.2007.03.014>.

Koyama, A. *et al.* (2003) ‘Virus multiplication and induction of apoptosis by Sendai virus: role of the C proteins’, *Microbes and Infection*, 5(5), pp. 373–378. Available at: [https://doi.org/10.1016/S1286-4579\(03\)00043-1](https://doi.org/10.1016/S1286-4579(03)00043-1).

Kozak, M. (1984) ‘Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo’, *Nature*, 6(308), pp. 241–246. Available at: <https://doi.org/10.1038/308241a0>.

Ksiazek, T.G., Rota, P.A. and Rollin, P.E. (2011) ‘A review of Nipah and Hendra viruses with an historical aside’, *Virus Research*, 162(1–2), pp. 173–183. Available at: <https://doi.org/10.1016/j.virusres.2011.09.026>.

Kuhn, J.H. *et al.* (2019) ‘Classify viruses — the gain is worth the pain’, *Nature*, 566(7744), pp. 318–320. Available at: <https://doi.org/10.1038/d41586-019-00599-8>.

Kulkarni, S. *et al.* (2009) ‘Nipah Virus Edits Its P Gene at High Frequency To Express the V and W Proteins’, *Journal of Virology*, 83(8), pp. 3982–3987. Available at: <https://doi.org/10.1128/JVI.02599-08>.

Kumar, S. and Hedges, S.B. (1998) ‘A molecular timescale for vertebrate evolution’, *Nature*, 392(6679), pp. 917–920. Available at: <https://doi.org/10.1038/31927>.

Kurotani, A. *et al.* (1998) ‘Sendai virus C proteins are categorically nonessential gene products but silencing their expression severely impairs viral replication and pathogenesis’, *Genes to Cells*, 3(2), pp. 111–124. Available at: <https://doi.org/10.1046/j.1365-2443.1998.00170.x>.

Lam, S.K. and Chua, K.B. (2002) ‘Nipah Virus Encephalitis Outbreak in Malaysia’, *Clinical Infectious Diseases*, 34(Supplement\_2), pp. S48–S51. Available at: <https://doi.org/10.1086/338818>.

Larsen, B.B. *et al.* (2022) ‘Evolution and Diversity of Bat and Rodent Paramyxoviruses from North America’, *Journal of Virology*, 96(3).

Latorre, P. *et al.* (1998) ‘The Various Sendai Virus C Proteins Are Not Functionally Equivalent and Exert both Positive and Negative Effects on Viral RNA Accumulation during the Course of Infection’, *Journal of Virology*, 72(7), pp. 5984–5993. Available at: <https://doi.org/10.1128/JVI.72.7.5984-5993.1998>.

Lawrence, P. and Escudero-Pérez, B. (2022) ‘Henipavirus Immune Evasion and Pathogenesis Mechanisms: Lessons Learnt from Natural Infection and Animal Models’, *Viruses*, 14(5), p.

936. Available at: <https://doi.org/10.3390/v14050936>.

Lee, B. and Ataman, Z.A. (2011) 'Modes of paramyxovirus fusion: a Henipavirus perspective', *Trends in Microbiology*, 19(8), pp. 389–399. Available at: <https://doi.org/10.1016/j.tim.2011.03.005>.

Li, Z., Ye, Y. and Godzik, A. (2006) 'Flexible Structural Neighborhood--a database of protein structural similarities and alignments', *Nucleic Acids Research*, 34(90001), pp. D277–D280. Available at: <https://doi.org/10.1093/nar/gkj124>.

Liu, Y.C., Grusovin, J. and Adams, T.E. (2018) 'Electrostatic Interactions between Hendra Virus Matrix Proteins Are Required for Efficient Virus-Like-Particle Assembly', *Journal of Virology*. Edited by R.E. Dutch, 92(13), pp. e00143-18. Available at: <https://doi.org/10.1128/JVI.00143-18>.

Lo, M.K. *et al.* (2009) 'Determination of the henipavirus phosphoprotein gene mRNA editing frequencies and detection of the C, V and W proteins of Nipah virus in virus-infected cells', *Journal of General Virology*, 90(2), pp. 398–404. Available at: <https://doi.org/10.1099/vir.0.007294-0>.

Lo, M.K., Sogaard, T.M. and Karlin, D.G. (2014) 'Evolution and Structural Organization of the C Proteins of Paramyxovirinae', *PLoS ONE*. Edited by D.P. Martin, 9(2), p. e90003. Available at: <https://doi.org/10.1371/journal.pone.0090003>.

Longhi, S. *et al.* (2017) 'How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication', *Cellular and Molecular Life Sciences*, 74(17), pp. 3091–3118. Available at: <https://doi.org/10.1007/s00018-017-2556-3>.

Luby, S. *et al.* (2006) 'Foodborne Transmission of Nipah Virus, Bangladesh', *Emerging Infectious Diseases*, 12(12), pp. 1888–1894. Available at: <https://doi.org/10.3201/eid1212.060732>.

Luby, S.P. *et al.* (2009) 'Recurrent Zoonotic Transmission of Nipah Virus into Humans, Bangladesh, 2001–2007', *Emerging Infectious Diseases*, 15(8), pp. 1229–1235. Available at: <https://doi.org/10.3201/eid1508.081237>.

Luby, S.P. (2013) 'The pandemic potential of Nipah virus', *Antiviral Research*, 100(1), pp. 38–43. Available at: <https://doi.org/10.1016/j.antiviral.2013.07.011>.

Luo, P. and Baldwin, R.L. (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(27), pp. 8413–8421. Available at: <https://doi.org/10.1021/bi9707133>.

Lynch, S. and Kolakofsky, D. (1978) 'Ends of the RNA within Sendai virus defective interfering nucleocapsids are not free', *Journal of Virology*, 28(2), pp. 584–589. Available at: <https://doi.org/10.1128/jvi.28.2.584-589.1978>.

Magoffin, D.E., Mackenzie, J.S. and Wang, L.-F. (2007) 'Genetic analysis of J-virus and

Beilong virus using minireplicons', *Virology*, 364(1), pp. 103–111. Available at: <https://doi.org/10.1016/j.virol.2007.01.045>.

Malur, A.G. *et al.* (2005) 'Inhibition of STAT 1 Phosphorylation by Human Parainfluenza Virus Type 3 C Protein', *Journal of Virology*, 79(12), pp. 7877–7882. Available at: <https://doi.org/10.1128/JVI.79.12.7877-7882.2005>.

Malur, A.G. *et al.* (2009) 'Evidence for phosphorylation of human parainfluenza virus type 3 C protein: Mutant C proteins exhibit variable inhibitory activities in vitro', *Virus Research*, 144(1–2), pp. 180–187. Available at: <https://doi.org/10.1016/j.virusres.2009.04.022>.

Malur, A.G., Hoffman, M.A. and Banerjee, A.K. (2004) 'The human parainfluenza virus type 3 (HPIV 3) C protein inhibits viral transcription', *Virus Research*, 99(2), pp. 199–204. Available at: <https://doi.org/10.1016/j.virusres.2003.11.009>.

Mao, H., Chattopadhyay, S. and Banerjee, A.K. (2009) 'N-terminally truncated C protein, CNΔ25, of human parainfluenza virus type 3 is a potent inhibitor of viral replication', *Virology*, 394(1), pp. 143–148. Available at: <https://doi.org/10.1016/j.virol.2009.08.026>.

Marq, J.-B. *et al.* (2007) 'Targeting of the Sendai Virus C Protein to the Plasma Membrane via a Peptide-Only Membrane Anchor', *Journal of Virology*, 81(7), pp. 3187–3197. Available at: <https://doi.org/10.1128/JVI.02465-06>.

Martinez-Gil, L., Vera-Velasco, N.M. and Mingarro, I. (2017) 'Exploring the Human-Nipah Virus Protein-Protein Interactome', *Journal of Virology*. Edited by R.E. Dutch, 91(23), pp. e01461-17, [/jvi/91/23/e01461-17.atom](https://doi.org/10.1128/JVI.01461-17). Available at: <https://doi.org/10.1128/JVI.01461-17>.

Mathieu, C. *et al.* (2011) 'Nipah Virus Uses Leukocytes for Efficient Dissemination within a Host', *Journal of Virology*, 85(15), pp. 7863–7871. Available at: <https://doi.org/10.1128/JVI.00549-11>.

Mathieu, C. *et al.* (2012) 'Nonstructural Nipah Virus C Protein Regulates both the Early Host Proinflammatory Response and Viral Virulence', *Journal of Virology*, 86(19), pp. 10766–10775. Available at: <https://doi.org/10.1128/JVI.01203-12>.

Mathieu, C. *et al.* (2015) 'Heparan Sulfate-Dependent Enhancement of Henipavirus Infection', *mBio*. Edited by J.R. Bennink, 6(2), pp. e02427-14. Available at: <https://doi.org/10.1128/mBio.02427-14>.

McCarthy, A.J. and Goodman, S.J. (2010) 'Reassessing conflicting evolutionary histories of the Paramyxoviridae and the origins of respiroviruses with Bayesian multigene phylogenies', *Infection, Genetics and Evolution*, 10(1), pp. 97–107. Available at: <https://doi.org/10.1016/j.meegid.2009.11.002>.

McCoy, A.J. *et al.* (2007) 'Phaser crystallographic software', *Journal of Applied Crystallography*, 40(4), pp. 658–674. Available at: <https://doi.org/10.1107/S0021889807021206>.

Mebatsion, T. *et al.* (2001) 'A Recombinant Newcastle Disease Virus with Low-Level V Protein Expression Is Immunogenic and Lacks Pathogenicity for Chicken Embryos', *Journal*

of *Virology*, 75(1), pp. 420–428. Available at: <https://doi.org/10.1128/JVI.75.1.420-428.2001>.

Meignié, A. *et al.* (2021) ‘Proteomic Analysis Uncovers Measles Virus Protein C Interaction With p65–iASPP Protein Complex’, *Molecular & Cellular Proteomics*, 20, p. 100049. Available at: <https://doi.org/10.1016/j.mcpro.2021.100049>.

le Mercier, P. and Kolakofsky, D. (2019) ‘Bipartite promoters and RNA editing of paramyxoviruses and filoviruses’, *RNA*, 25(3), pp. 279–285. Available at: <https://doi.org/10.1261/rna.068825.118>.

Midtgaard, S.R. *et al.* (2018) ‘Invisible detergents for structure determination of membrane proteins by small-angle neutron scattering’, *The FEBS Journal*, 285(2), pp. 357–371. Available at: <https://doi.org/10.1111/febs.14345>.

Mirdita, M. *et al.* (2022) ‘ColabFold: making protein folding accessible to all’, *Nature Methods*, 19(6), pp. 679–682. Available at: <https://doi.org/10.1038/s41592-022-01488-1>.

Mire, C.E. *et al.* (2013) ‘Single injection recombinant vesicular stomatitis virus vaccines protect ferrets against lethal Nipah virus disease’, *Virology Journal*, 10(1), p. 353. Available at: <https://doi.org/10.1186/1743-422X-10-353>.

Mire, C.E. *et al.* (2016) ‘Pathogenic Differences between Nipah Virus Bangladesh and Malaysia Strains in Primates: Implications for Antibody Therapy’, *Scientific Reports*, 6(1), p. 30916. Available at: <https://doi.org/10.1038/srep30916>.

Mura, M. *et al.* (2017) ‘Nonencapsidated 5' Copy-Back Defective Interfering Genomes Produced by Recombinant Measles Viruses Are Recognized by RIG-I and LGP2 but Not MDA5’, *Journal of Virology*. Edited by A. García-Sastre, 91(20), pp. e00643-17. Available at: <https://doi.org/10.1128/JVI.00643-17>.

Nahar, N. *et al.* (2010) ‘Date Palm Sap Collection: Exploring Opportunities to Prevent Nipah Transmission’, *EcoHealth*, 7(2), pp. 196–203. Available at: <https://doi.org/10.1007/s10393-010-0320-3>.

Nakatsu, Y. *et al.* (2009) ‘A Highly Attenuated Measles Virus Vaccine Strain Encodes a Fully Functional C Protein’, *Journal of Virology*, 83(22), pp. 11996–12001. Available at: <https://doi.org/10.1128/JVI.00791-09>.

Navaratnarajah, C.K. *et al.* (2020) ‘Receptor-mediated cell entry of paramyxoviruses: Mechanisms, and consequences for tropism and pathogenesis’, *Journal of Biological Chemistry*, 295(9), pp. 2771–2786. Available at: <https://doi.org/10.1074/jbc.REV119.009961>.

Negrete, O.A. *et al.* (2005) ‘EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus’, *Nature*, 436(7049), pp. 401–405. Available at: <https://doi.org/10.1038/nature03838>.

Nikolay, B. *et al.* (2019) ‘Transmission of Nipah Virus — 14 Years of Investigations in Bangladesh’, *New England Journal of Medicine*, 380(19), pp. 1804–1814. Available at: <https://doi.org/10.1056/NEJMoA1805376>.

Nikolic, J. *et al.* (2017) 'Negri bodies are viral factories with properties of liquid organelles', *Nature Communications*, 8(1), p. 58. Available at: <https://doi.org/10.1038/s41467-017-00102-9>.

Nilsson-Payant, B.E. *et al.* (2021) 'Reduced Nucleoprotein Availability Impairs Negative-Sense RNA Virus Replication and Promotes Host Recognition', *Journal of Virology*. Edited by S. Schultz-Cherry, 95(9), pp. e02274-20. Available at: <https://doi.org/10.1128/JVI.02274-20>.

Nishie, T. and Nagata, K. (2015) 'Measles virus C protein facilitates transcription by the control of N protein-viral genomic RNA interaction in early phases of infection', *Biochemical and Biophysical Research Communications*, 463(4), pp. 1262–1266. Available at: <https://doi.org/10.1016/j.bbrc.2015.06.099>.

Nishie, T., Nagata, K. and Takeuchi, K. (2007) 'The C protein of wild-type measles virus has the ability to shuttle between the nucleus and the cytoplasm', *Microbes and Infection*, 9(3), pp. 344–354. Available at: <https://doi.org/10.1016/j.micinf.2006.12.008>.

Nor, M.N.M., Gan, C.H. and Ong, B.L. (2000) 'Nipah virus infection of pigs in peninsular Malaysia', *Rev Sci Tech.*, 19(1), pp. 160–165. Available at: <https://doi.org/10.20506/rst.19.1.1202>.

Norris, M.J. *et al.* (2022) 'Measles and Nipah virus assembly: Specific lipid binding drives matrix polymerization', *Science Advances*, 8(29), p. eabn1440. Available at: <https://doi.org/10.1126/sciadv.abn1440>.

Noton, S.L. and Fearn, R. (2015) 'Initiation and regulation of paramyxovirus transcription and replication', *Virology*, 479–480, pp. 545–554. Available at: <https://doi.org/10.1016/j.virol.2015.01.014>.

Oda, K. *et al.* (2015) 'Structural Basis of the Inhibition of STAT1 Activity by Sendai Virus C Protein', *Journal of Virology*. Edited by D.S. Lyles, 89(22), pp. 11487–11499. Available at: <https://doi.org/10.1128/JVI.01887-15>.

Oda, K. *et al.* (2017) 'Structural analysis of the STAT1:STAT2 heterodimer revealed the mechanism of Sendai virus C protein-mediated blockade of type 1 interferon signaling', *Journal of Biological Chemistry*, 292(48), pp. 19752–19766. Available at: <https://doi.org/10.1074/jbc.M117.786285>.

Oda, K. *et al.* (2021) 'Structural Insight into the Interaction of Sendai Virus C Protein with Alix To Stimulate Viral Budding', *Journal of Virology*. Edited by C.R. Parrish, 95(19), pp. e00815-21. Available at: <https://doi.org/10.1128/JVI.00815-21>.

Ojha, R. *et al.* (2019) 'Strategic Development of a Next-Generation Multi-Epitope Vaccine To Prevent Nipah Virus Zoonotic Infection', *ACS Omega*, 4(8), pp. 13069–13079. Available at: <https://doi.org/10.1021/acsomega.9b00944>.

Pager, C.T. *et al.* (2006) 'A mature and fusogenic form of the Nipah virus fusion protein requires proteolytic processing by cathepsin L', *Virology*, 346(2), pp. 251–257. Available at: <https://doi.org/10.1016/j.virol.2006.01.007>.

Pardon, E. *et al.* (2014) ‘A general protocol for the generation of Nanobodies for structural biology’, *Nature Protocols*, 9(3), pp. 674–693. Available at: <https://doi.org/10.1038/nprot.2014.039>.

Park, A. *et al.* (2016) ‘Nipah Virus C Protein Recruits Tsg101 to Promote the Efficient Release of Virus in an ESCRT-Dependent Pathway’, *PLoS Pathogens*. Edited by R.A. Lamb, 12(5), p. e1005659. Available at: <https://doi.org/10.1371/journal.ppat.1005659>.

Patch, J.R. *et al.* (2007) ‘Quantitative analysis of Nipah virus proteins released as virus-like particles reveals central role for the matrix protein’, *Virology Journal*, 4(1), p. 1. Available at: <https://doi.org/10.1186/1743-422X-4-1>.

Pelissier, R., Iampietro, M. and Horvat, B. (2019) ‘Recent advances in the understanding of Nipah virus immunopathogenesis and anti-viral approaches’, *F1000Research*, 8, p. 1763. Available at: <https://doi.org/10.12688/f1000research.19975.1>.

Pfaller, C.K. *et al.* (2014) ‘Measles Virus C Protein Impairs Production of Defective Copyback Double-Stranded Viral RNA and Activation of Protein Kinase R’, *Journal of Virology*, 88(1), pp. 456–468. Available at: <https://doi.org/10.1128/JVI.02572-13>.

Pfaller, C.K. *et al.* (2015) ‘Measles Virus Defective Interfering RNAs Are Generated Frequently and Early in the Absence of C Protein and Can Be Destabilized by Adenosine Deaminase Acting on RNA-1-Like Hypermutations’, *Journal of Virology*. Edited by T.S. Dermody, 89(15), pp. 7735–7747. Available at: <https://doi.org/10.1128/JVI.01017-15>.

Pfaller, C.K. *et al.* (2019) ‘The C Protein Is Recruited to Measles Virus Ribonucleocapsids by the Phosphoprotein’, *Journal of Virology*. Edited by R.E. Dutch, 94(4), pp. e01733-19, [/jvi/94/4/JVI.01733-19.atom](https://doi.org/10.1128/JVI.01733-19). Available at: <https://doi.org/10.1128/JVI.01733-19>.

Poirot, O., O’Toole, E. and Notredame, C. (2003) ‘Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments’, *Nucleic Acids Research*, 31(13), pp. 3503–3506. Available at: <https://doi.org/10.1093/nar/gkg522>.

Porotto, M. *et al.* (2010) ‘Inhibition of Nipah Virus Infection In Vivo: Targeting an Early Stage of Paramyxovirus Fusion Activation during Viral Entry’, *PLoS Pathogens*. Edited by F.A. Rey, 6(10), p. e1001168. Available at: <https://doi.org/10.1371/journal.ppat.1001168>.

Prescott, J. *et al.* (2015) ‘Single-dose live-attenuated vesicular stomatitis virus-based vaccine protects African green monkeys from Nipah virus disease’, *Vaccine*, 33(24), pp. 2823–2829. Available at: <https://doi.org/10.1016/j.vaccine.2015.03.089>.

Putnam, C.D. *et al.* (2007) ‘X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution’, *Quarterly Reviews of Biophysics*, 40(3), pp. 191–285. Available at: <https://doi.org/10.1017/S0033583507004635>.

Rahman, M. and Chakraborty, A. (2012) ‘Nipah virus outbreaks in Bangladesh: a deadly infectious disease’, *WHO South-East Asia Journal of Public Health*, 1(2), p. 208. Available at: <https://doi.org/10.4103/2224-3151.206933>.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) 'Stereochemistry of polypeptide chain configurations', *Journal of Molecular Biology*, 7(1), pp. 95–99. Available at: [https://doi.org/10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6).

Rehwinkel, J. and Gack, M.U. (2020) 'RIG-I-like receptors: their regulation and roles in RNA sensing', *Nature Reviews Immunology*, 20(9), pp. 537–551. Available at: <https://doi.org/10.1038/s41577-020-0288-3>.

Reutter, G.L. *et al.* (2001) 'Mutations in the Measles Virus C Protein That Up Regulate Viral RNA Synthesis', *Virology*, 285(1), pp. 100–109. Available at: <https://doi.org/10.1006/viro.2001.0962>.

Reynes, J.-M. *et al.* (2005) 'Nipah Virus in Lyle's Flying Foxes, Cambodia', *Emerging Infectious Diseases*, 11(7), pp. 1042–1047. Available at: <https://doi.org/10.3201/eid1107.041350>.

Rima, B. *et al.* (2018) 'Problems of classification in the family Paramyxoviridae', *Archives of Virology*, 163(5), pp. 1395–1404. Available at: <https://doi.org/10.1007/s00705-018-3720-2>.

Rodriguez, J.J., Cruz, C.D. and Horvath, C.M. (2004) 'Identification of the Nuclear Export Signal and STAT-Binding Domains of the Nipah Virus V Protein Reveals Mechanisms Underlying Interferon Evasion', *Journal of Virology*, 78(10), pp. 5358–5367. Available at: <https://doi.org/10.1128/JVI.78.10.5358-5367.2004>.

Rodriguez, J.J., Parisien, J.-P. and Horvath, C.M. (2002) 'Nipah Virus V Protein Evades Alpha and Gamma Interferons by Preventing STAT1 and STAT2 Activation and Nuclear Accumulation', *Journal of Virology*, 76(22), pp. 11476–11483. Available at: <https://doi.org/10.1128/JVI.76.22.11476-11483.2002>.

Rodriguez, J.J., Wang, L.-F. and Horvath, C.M. (2003) 'Hendra Virus V Protein Inhibits Interferon Signaling by Preventing STAT1 and STAT2 Nuclear Accumulation', *Journal of Virology*, 77(21), pp. 11842–11845. Available at: <https://doi.org/10.1128/jvi.77.21.11842-11845.2003>.

Roth, J.P. *et al.* (2013) 'Deletion of the D domain of the human parainfluenza virus type 3 (HPIV3) PD protein results in decreased viral RNA synthesis and beta interferon (IFN- $\beta$ ) expression', *Virus Genes*, 47(1), pp. 10–19. Available at: <https://doi.org/10.1007/s11262-013-0919-x>.

Roy, A., Chan Mine, E., Gaifas, L., Leyrat, C., Volchkova, V.A., *et al.* (2023) 'Orthoparamyxovirinae C Proteins Have a Common Origin and a Common Structural Organization', *Biomolecules*, 13(3), p. 455. Available at: <https://doi.org/10.3390/biom13030455>.

Ruigrok, R.W., Crépin, T. and Kolakofsky, D. (2011) 'Nucleoproteins and nucleocapsids of negative-strand RNA viruses', *Current Opinion in Microbiology*, 14(4), pp. 504–510. Available at: <https://doi.org/10.1016/j.mib.2011.07.011>.

Runkler, N. *et al.* (2007) 'Measles virus nucleocapsid transport to the plasma membrane requires stable expression and surface accumulation of the viral matrix protein', *Cellular*



*Microbiology*, 9(5), pp. 1203–1214. Available at: <https://doi.org/10.1111/j.1462-5822.2006.00860.x>.

Sabath, N., Wagner, A. and Karlin, D. (2012) ‘Evolution of Viral Proteins Originated De Novo by Overprinting’, *Molecular Biology and Evolution*, 29(12), pp. 3767–3780. Available at: <https://doi.org/10.1093/molbev/mss179>.

Saha, C.K. *et al.* (2017) ‘In silico identification and characterization of common epitope-based peptide vaccine for Nipah and Hendra viruses’, *Asian Pacific Journal of Tropical Medicine*, 10(6), pp. 529–538. Available at: <https://doi.org/10.1016/j.apjtm.2017.06.016>.

Sakaguchi, T. *et al.* (2005) ‘AIP1/Alix Is a Binding Partner of Sendai Virus C Protein and Facilitates Virus Budding’, *Journal of Virology*, 79(14), pp. 8933–8941. Available at: <https://doi.org/10.1128/JVI.79.14.8933-8941.2005>.

Salah Uddin Khan, M. *et al.* (2010) ‘Use of Infrared Camera to Understand Bats’ Access to Date Palm Sap: Implications for Preventing Nipah Virus Transmission’, *EcoHealth*, 7(4), pp. 517–525. Available at: <https://doi.org/10.1007/s10393-010-0366-2>.

Sánchez-Aparicio, M.T. *et al.* (2018) ‘Paramyxovirus V Proteins Interact with the RIG-I/TRIM25 Regulatory Complex and Inhibit RIG-I Signaling’, *Journal of Virology*. Edited by T.S. Dermody, 92(6), pp. e01960-17, [/jvi/92/6/e01960-17.atom](https://doi.org/10.1128/JVI.01960-17). Available at: <https://doi.org/10.1128/JVI.01960-17>.

Sanz Bernardo, B., Goodbourn, S. and Baron, M.D. (2017) ‘Control of the induction of type I interferon by Peste des petits ruminants virus’, *PLOS ONE*. Edited by D.-Y. Jin, 12(5), p. e0177300. Available at: <https://doi.org/10.1371/journal.pone.0177300>.

Satterfield, B.A. *et al.* (2016) ‘Nipah Virus C and W Proteins Contribute to Respiratory Disease in Ferrets’, *Journal of Virology*. Edited by D.S. Lyles, 90(14), pp. 6326–6343. Available at: <https://doi.org/10.1128/JVI.00215-16>.

Satterfield, B.A., Dawes, B.E. and Milligan, G.N. (2016) ‘Status of vaccine research and development of vaccines for Nipah virus’, *Vaccine*, 34(26), pp. 2971–2975. Available at: <https://doi.org/10.1016/j.vaccine.2015.12.075>.

Satterfield, B.A., Geisbert, T.W. and Mire, C.E. (2016) ‘Inhibition of the host antiviral response by Nipah virus: current understanding and future perspectives’, *Future Virology*, 11(5), pp. 331–344. Available at: <https://doi.org/10.2217/fvl-2016-0027>.

Schlub, T.E. and Holmes, E.C. (2020) ‘Properties and abundance of overlapping genes in viruses’, *Virus Evolution*, 6(1), p. veaa009. Available at: <https://doi.org/10.1093/ve/veaa009>.

Schountz, T. *et al.* (2017) ‘Immunological Control of Viral Infections in Bats and the Emergence of Viruses Highly Pathogenic to Humans’, *Frontiers in Immunology*, 8, p. 1098. Available at: <https://doi.org/10.3389/fimmu.2017.01098>.

Shaw, M.L. *et al.* (2005) ‘Nuclear Localization of the Nipah Virus W Protein Allows for Inhibition of both Virus- and Toll-Like Receptor 3-Triggered Signaling Pathways’, *Journal of Virology*, 79(10), pp. 6078–6088. Available at: <https://doi.org/10.1128/JVI.79.10.6078->

6088.2005.

Sheldrick, G.M. (1990) 'Phase annealing in SHELX-90: direct methods for larger structures', *Acta Crystallographica Section A Foundations and Advances*, A46, pp. 467–473.

Shil, N.K. *et al.* (2018) 'Inflammasome Antagonism by Human Parainfluenza Virus Type 3 C Protein', *Journal of Virology*. Edited by J.U. Jung, 92(4), pp. e01776-17. Available at: <https://doi.org/10.1128/JVI.01776-17>.

Siering, O., Cattaneo, R. and Pfaller, C.K. (2022) 'C Proteins: Controllers of Orderly Paramyxovirus Replication and of the Innate Immune Response', *Viruses*, 14(1), p. 137. Available at: <https://doi.org/10.3390/v14010137>.

Siering, O., Sawatsky, B. and Pfaller, C.K. (2021) 'C Protein Is Essential for Canine Distemper Virus Virulence and Pathogenicity in Ferrets', *Journal of Virology*. Edited by R.E. Dutch, 95(4), pp. e01840-20. Available at: <https://doi.org/10.1128/JVI.01840-20>.

Simon-Loriere, E., Holmes, E.C. and Pagán, I. (2013) 'The Effect of Gene Overlapping on the Rate of RNA Virus Evolution', *Molecular Biology and Evolution*, 30(8), pp. 1916–1928. Available at: <https://doi.org/10.1093/molbev/mst094>.

Singh, R.K. *et al.* (2019) 'Nipah virus: epidemiology, pathology, immunobiology and advances in diagnosis, vaccine designing and control strategies – a comprehensive review', *Veterinary Quarterly*, 39(1), pp. 26–55. Available at: <https://doi.org/10.1080/01652176.2019.1580827>.

Skowron, K. *et al.* (2022) 'Nipah Virus—Another Threat From the World of Zoonotic Viruses', *Frontiers in Microbiology*, 12, p. 811157. Available at: <https://doi.org/10.3389/fmicb.2021.811157>.

Sleeman, K. *et al.* (2008) 'The C, V and W proteins of Nipah virus inhibit minigenome replication', *Journal of General Virology*, 89(5), pp. 1300–1308. Available at: <https://doi.org/10.1099/vir.0.83582-0>.

Smallwood, S. and Moyer, S.A. (2004) 'The L polymerase protein of parainfluenza virus 3 forms an oligomer and can interact with the heterologous Sendai virus L, P and C proteins', *Virology*, 318(1), pp. 439–450. Available at: <https://doi.org/10.1016/j.virol.2003.09.045>.

Smith, K.M. *et al.* (2018) 'Structural basis for importin alpha 3 specificity of W proteins in Hendra and Nipah viruses', *Nature Communications*, 9(1), p. 3703. Available at: <https://doi.org/10.1038/s41467-018-05928-5>.

Soding, J. (2005) 'Protein homology detection by HMM-HMM comparison', *Bioinformatics*, 21(7), pp. 951–960. Available at: <https://doi.org/10.1093/bioinformatics/bti125>.

Soman Pillai, V., Krishna, G. and Valiya Veetil, M. (2020) 'Nipah Virus: Past Outbreaks and Future Containment', *Viruses*, 12(4), p. 465. Available at: <https://doi.org/10.3390/v12040465>.

Sparrer, K.M.J., Pfaller, C.K. and Conzelmann, K.-K. (2012) 'Measles Virus C Protein Interferes with Beta Interferon Transcription in the Nucleus', *Journal of Virology*, 86(2), pp. 796–805. Available at: <https://doi.org/10.1128/JVI.05899-11>.

Springfield, C. *et al.* (2005) 'Envelope Targeting: Hemagglutinin Attachment Specificity Rather than Fusion Protein Cleavage-Activation Restricts *Tupaia Paramyxovirus* Tropism', *Journal of Virology*, 79(16), pp. 10155–10163. Available at: <https://doi.org/10.1128/JVI.79.16.10155-10163.2005>.

Strahle, L. *et al.* (2007) 'Activation of the Beta Interferon Promoter by Unnatural Sendai Virus Infection Requires RIG-I and Is Inhibited by Viral C Proteins', *Journal of Virology*, 81(22), pp. 12227–12237. Available at: <https://doi.org/10.1128/JVI.01300-07>.

Su, H. *et al.* (2023) 'Discovery and characterization of novel paramyxoviruses from bat samples in China', *Virologica Sinica* [Preprint]. Available at: <https://doi.org/10.1016/j.virs.2023.01.002>.

Su, J. *et al.* (2015) 'Application of minigenome technology in virology research of the Paramyxoviridae family', *Journal of Microbiology, Immunology and Infection*, 48(2), pp. 123–129. Available at: <https://doi.org/10.1016/j.jmii.2014.02.008>.

Sugahara, F. *et al.* (2004) 'Paramyxovirus Sendai virus-like particle formation by expression of multiple viral proteins and acceleration of its release by C protein', *Virology*, 325(1), pp. 1–10. Available at: <https://doi.org/10.1016/j.virol.2004.04.019>.

Sugai, A. *et al.* (2017) 'Gene end-like sequences within the 3' non-coding region of the Nipah virus genome attenuate viral gene transcription', *Virology*, 508, pp. 36–44. Available at: <https://doi.org/10.1016/j.virol.2017.05.004>.

Svergun, D., Barberato, C. and Koch, M.H.J. (1995) 'CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates', *Journal of Applied Crystallography*, 28(6), pp. 768–773. Available at: <https://doi.org/10.1107/S0021889895007047>.

Svergun, D.I. and Koch, M.H.J. (2003) 'Small-angle scattering studies of biological macromolecules in solution', *Reports on Progress in Physics*, 66(10), pp. 1735–1782. Available at: <https://doi.org/10.1088/0034-4885/66/10/R05>.

Takeuchi, K. *et al.* (2001) 'Sendai virus C protein physically associates with Stat1: Interaction of Sendai virus C protein with Stat1', *Genes to Cells*, 6(6), pp. 545–557. Available at: <https://doi.org/10.1046/j.1365-2443.2001.00442.x>.

Takeuchi, K. *et al.* (2005) 'Stringent Requirement for the C Protein of Wild-Type Measles Virus for Growth both In Vitro and in Macaques', *Journal of Virology*, 79(12), pp. 7838–7844. Available at: <https://doi.org/10.1128/JVI.79.12.7838-7844.2005>.

Takimoto, T. *et al.* (2001) 'Role of Matrix and Fusion Proteins in Budding of Sendai Virus', *Journal of Virology*, 75(23), pp. 11384–11391. Available at: <https://doi.org/10.1128/JVI.75.23.11384-11391.2001>.

Takimoto, T. and Portner, A. (2004) 'Molecular mechanism of paramyxovirus budding', *Virus Research*, 106(2), pp. 133–145. Available at: <https://doi.org/10.1016/j.virusres.2004.08.010>.

- Tapparel, C. *et al.* (1997) ‘Inhibition of Sendai virus genome replication due to promoter-increased selectivity: a possible role for the accessory C proteins’, *Journal of Virology*, 71(12), pp. 9588–9599. Available at: <https://doi.org/10.1128/jvi.71.12.9588-9599.1997>.
- Tarbouriech, N. *et al.* (2000) ‘Tetrameric coiled coil domain of Sendai virus phosphoprotein’, *Nature Structural Biology*, 7(9), pp. 777–781. Available at: <https://doi.org/10.1038/79013>.
- Terwilliger, T.C. and Berendzen, J. (1999) ‘Automated MAD and MIR structure solution’, *Acta Crystallographica Section D Biological Crystallography*, 55(Pt 4), pp. 849–861. Available at: <https://doi.org/10.1107/S0907444999000839>.
- Thakur, V., Thakur, P. and Ratho, R.K. (2022) ‘Nipah Outbreak: Is it the beginning of another pandemic in the era of COVID-19 and Zika’, *Brain, Behavior, and Immunity*, 99, pp. 25–26. Available at: <https://doi.org/10.1016/j.bbi.2021.09.015>.
- Tidona, C.A. *et al.* (1999) ‘Isolation and Molecular Characterization of a Novel Cytopathogenic Paramyxovirus from Tree Shrews’, *Virology*, 258(2), pp. 425–434. Available at: <https://doi.org/10.1006/viro.1999.9693>.
- Tria, G. *et al.* (2015) ‘Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering’, *IUCrJ*, 2(2), pp. 207–217. Available at: <https://doi.org/10.1107/S205225251500202X>.
- Tsukiyama-Kohara, K. and Kohara, M. (2014) ‘Tupaia Belangeri as an Experimental Animal Model for Viral Infection’, 63(4), pp. 367–374. Available at: <https://doi.org/10.1538/expanim.63.367>.
- Uchida, S. *et al.* (2018) ‘Possible role of the Nipah virus V protein in the regulation of the interferon beta induction by interacting with UBX domain-containing protein1’, *Scientific Reports*, 8(1), p. 7682. Available at: <https://doi.org/10.1038/s41598-018-25815-9>.
- Vagin, A. and Teplyakov, A. (2010) ‘Molecular replacement with *MOLREP*’, *Acta Crystallographica Section D Biological Crystallography*, 66(1), pp. 22–25. Available at: <https://doi.org/10.1107/S0907444909042589>.
- Vera-Velasco, N.M. *et al.* (2018) ‘Proteomic composition of Nipah virus-like particles’, *Journal of Proteomics*, 172, pp. 190–200. Available at: <https://doi.org/10.1016/j.jprot.2017.10.012>.
- Vigant, F. and Lee, B. (2011) ‘Hendra and Nipah Infection: Pathology, Models and Potential Therapies’, *Infectious Disorders - Drug Targets*, 11(3), pp. 315–336. Available at: <https://doi.org/10.2174/187152611795768097>.
- Wacharapluesadee, S. *et al.* (2005) ‘Bat Nipah Virus, Thailand’, *Emerging Infectious Diseases*, 11(12), pp. 1949–1951. Available at: <https://doi.org/10.3201/eid1112.050613>.
- Walpita, P. *et al.* (2017) ‘A VLP-based vaccine provides complete protection against Nipah virus challenge following multiple-dose or single-dose vaccination schedules in a hamster model’, *npj Vaccines*, 2(1), p. 21. Available at: <https://doi.org/10.1038/s41541-017-0023-7>.

Wang, C. *et al.* (2022) ‘Evasion of Host Antiviral Innate Immunity by Paramyxovirus Accessory Proteins’, *Frontiers in Microbiology*, 12, p. 790191. Available at: <https://doi.org/10.3389/fmicb.2021.790191>.

Wellehan Jr., J.F.X. *et al.* (2009) ‘Detection of specific antibody responses to vaccination in variable flying foxes (*Pteropus hypomelanus*)’, *Comparative Immunology, Microbiology and Infectious Diseases*, 32(5), pp. 379–394. Available at: <https://doi.org/10.1016/j.cimid.2007.11.002>.

Wells, G., Addington-Hall, M. and Malur, A.G. (2012) ‘Mutations within the human parainfluenza virus type 3 (HPIV 3) C protein affect viral replication and host interferon induction’, *Virus Research*, 167(2), pp. 385–390. Available at: <https://doi.org/10.1016/j.virusres.2012.05.009>.

Wells, H.L. *et al.* (2022) ‘Classification of new morbillivirus and jeilongvirus sequences from bats sampled in Brazil and Malaysia’, *Archives of Virology*, 167(10), pp. 1977–1987. Available at: <https://doi.org/10.1007/s00705-022-05500-z>.

Wong, K.T. *et al.* (2002) ‘Nipah virus infection, an emerging paramyxoviral zoonosis’, *Springer Seminars in Immunopathology*, 24(2), pp. 215–228. Available at: <https://doi.org/10.1007/s00281-002-0106-y>.

Wu, Z. *et al.* (2005) ‘Role of the MEOX2 homeobox gene in neurovascular dysfunction in Alzheimer disease’, *Nature Medicine*, 11(9), pp. 959–965. Available at: <https://doi.org/10.1038/nm1287>.

Xu, K. *et al.* (2008) ‘Host cell recognition by the henipaviruses: Crystal structures of the Nipah G attachment glycoprotein and its complex with ephrin-B3’, *Proceedings of the National Academy of Sciences*, 105(29), pp. 9953–9958. Available at: <https://doi.org/10.1073/pnas.0804797105>.

Xu, K., Broder, C.C. and Nikolov, D.B. (2012) ‘Ephrin-B2 and ephrin-B3 as functional henipavirus receptors’, *Seminars in Cell & Developmental Biology*, 23(1), pp. 116–123. Available at: <https://doi.org/10.1016/j.semcdb.2011.12.005>.

Yabukarski, F. *et al.* (2014) ‘Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone’, *Nature Structural & Molecular Biology*, 21(9), pp. 754–759. Available at: <https://doi.org/10.1038/nsmb.2868>.

Yamaguchi, M. *et al.* (2014) ‘An anti-interferon activity shared by paramyxovirus C proteins: Inhibition of Toll-like receptor 7/9-dependent alpha interferon induction’, *FEBS Letters*, 588(1), pp. 28–34. Available at: <https://doi.org/10.1016/j.febslet.2013.11.015>.

Yoneda, M. *et al.* (2010) ‘The Nonstructural Proteins of Nipah Virus Play a Key Role in Pathogenicity in Experimentally Infected Animals’, *PLoS ONE*. Edited by M.G. Masucci, 5(9), p. e12709. Available at: <https://doi.org/10.1371/journal.pone.0012709>.

Yoneda, M. *et al.* (2013) ‘Recombinant Measles Virus Vaccine Expressing the Nipah Virus Glycoprotein Protects against Lethal Nipah Virus Challenge’, *PLoS ONE*. Edited by M.G.

Masucci, 8(3), p. e58414. Available at: <https://doi.org/10.1371/journal.pone.0058414>.