

Article

Not peer-reviewed version

---

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

---

Ada Roy , Emeric Chan Mine , Lorenzo Gaifas , [Cédric Leyrat](#) , [Valentina Volchkova](#) , Florence Baudin , Luis Martinez-Gil , [Viktor Volchkov](#) , [David G. Karlin](#) <sup>\*</sup> , Jean-Marie Bourhis , [Marc Jamin](#) <sup>\*</sup>

Posted Date: 12 January 2023

doi: 10.20944/preprints202301.0217.v1

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



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 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 Volchkova <sup>2</sup>, Florence Baudin <sup>4</sup>, Luis Martinez-Gil <sup>5</sup>, Viktor Volchkov <sup>2</sup>, David Karlin <sup>6,\*</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

<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 Génomique Fonctionnelle, Université de Montpellier, CNRS, INSERM, Montpellier, France

<sup>4</sup> Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

<sup>5</sup> Department of Biochemistry and Molecular Biology, Institute for Biotechnology and Biomedicine (BIOTECMED), University of Valencia, Valencia, Spain

<sup>6</sup> Independent Researcher, Marseille, France

\* Correspondence: marc.jamin@ibs.fr; jean-marie.bourhis@ibs.fr; davidgkarlin@gmail.com (D.K.); Tel.: + 33 4 76 20 94 62 (M.J.); Tel.: + 33 4 57 42 86 36 (J.-M.B.)

**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 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 minireplicon 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 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 representative of these two clusters, our discovery that they have a similar structure indicates that all *Orthoparamyxovirine* 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.

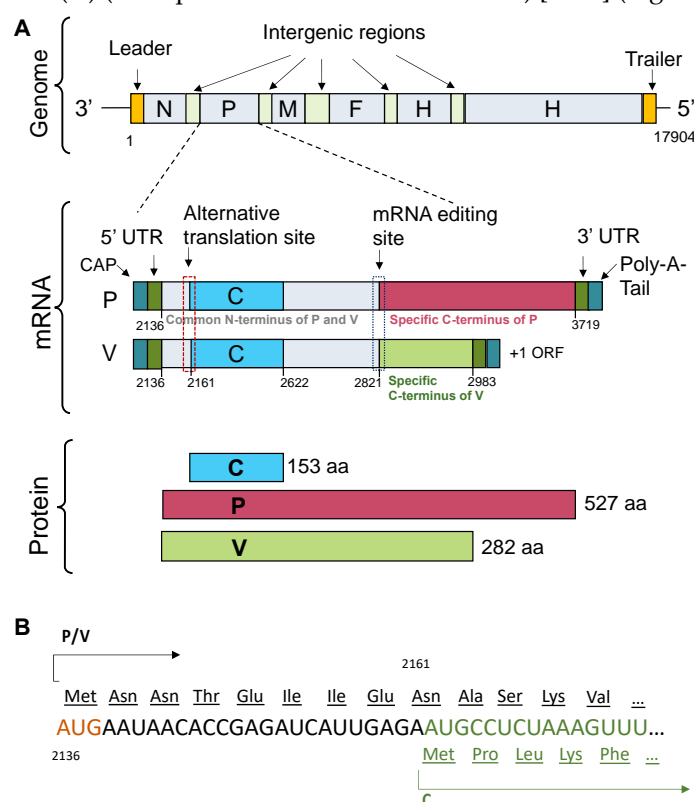
**Keywords:** *Paramyxoviridae*; virulence factor; overlapping genes; protein structure; viral evolution

## 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 guanines (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 measles virus, Nipah virus, and 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].

Since the C protein has proven resistant to recombinant production for over 20 years, the same study tested the bacterial production of the C protein of over 20 viruses and found the most suitable candidate to be the TupV C protein [22].

The tree shrew (*Tupaia belangeri*) is a small mammal that lives in 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 tree shrew a promising model for a variety of human diseases such as depression, myopia, hepatitis B and C infections and hepatocellular carcinoma and for development of therapeutic treatments, because it combines small size, easy breeding and rapid reproduction [24]. Tupaia paramyxovirus (TupV) could be isolated from the supernatant of the degenerated *T. belangeri* cell cultures and propagated in different types of *Tupaia* cells, but in no other mammalian cell lines [13]. TupV belongs to the family *Paramyxoviridae*, and of interest for our study, its P gene exhibited typical features of members of the subfamily *Orthoparamyxovirinae* [13] (Figure 1).

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]. Foldindex 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 query the different servers, process the information and to calculate the D-score is now accessible at <https://github.com/brisvag/dscore>.

## 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 a 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 hours 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, 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, United States), the protein was finally purified by size-exclusion chromatography (SEC) on a Hiload 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 constructs 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 hours 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 a 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 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).

#### 2.4. 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 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

A SBGrid consortium installation of AlphaFold version 2.1.2 [52,53] was run on a local server equipped with a 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, United States) was cultivated in Dulbecco's Modified Eagle's Medium (DMEM) GlutaMAX (Thermo Fisher Scientific, Waltham, MA, USA) containing 50 U.mL<sup>-1</sup> penicillin, 50 µg.mL<sup>-1</sup> 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 pCMV vector. The expression of NiV C from the P gene was reduced by replacing the Kozak sequence of the P protein by that of the N protein (a detailed description of this construct will be published elsewhere).

The construction of Nipah virus minigenome was performed as described for La Crosse virus minigenome [55]. The NiV leader region, the "gene start" signal (GS) and the untranslated region from NiV N gene were inserted upstream from the mouse RNA pol1 terminator sequence. The NiV trailer, the "gene stop" signal (GE) and the untranslated region from NiV L gene were inserted after the RNA pol1 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

Subconfluent monolayer of HEK293T cells were transfected with 100 ng plasmid pPol I NiV-REN, 125 ng phCMV NiV N, 25 ng phCMV NiV P, 350 ng phCMV NiV L using jetOPTIMUS transfection reagent (Polyplus, Illkirch-Graffenstaden, France). Additional transfection of pGL4.50 (Promega, Madison, WI, United States) 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 minireplicon assay will be published elsewhere.

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

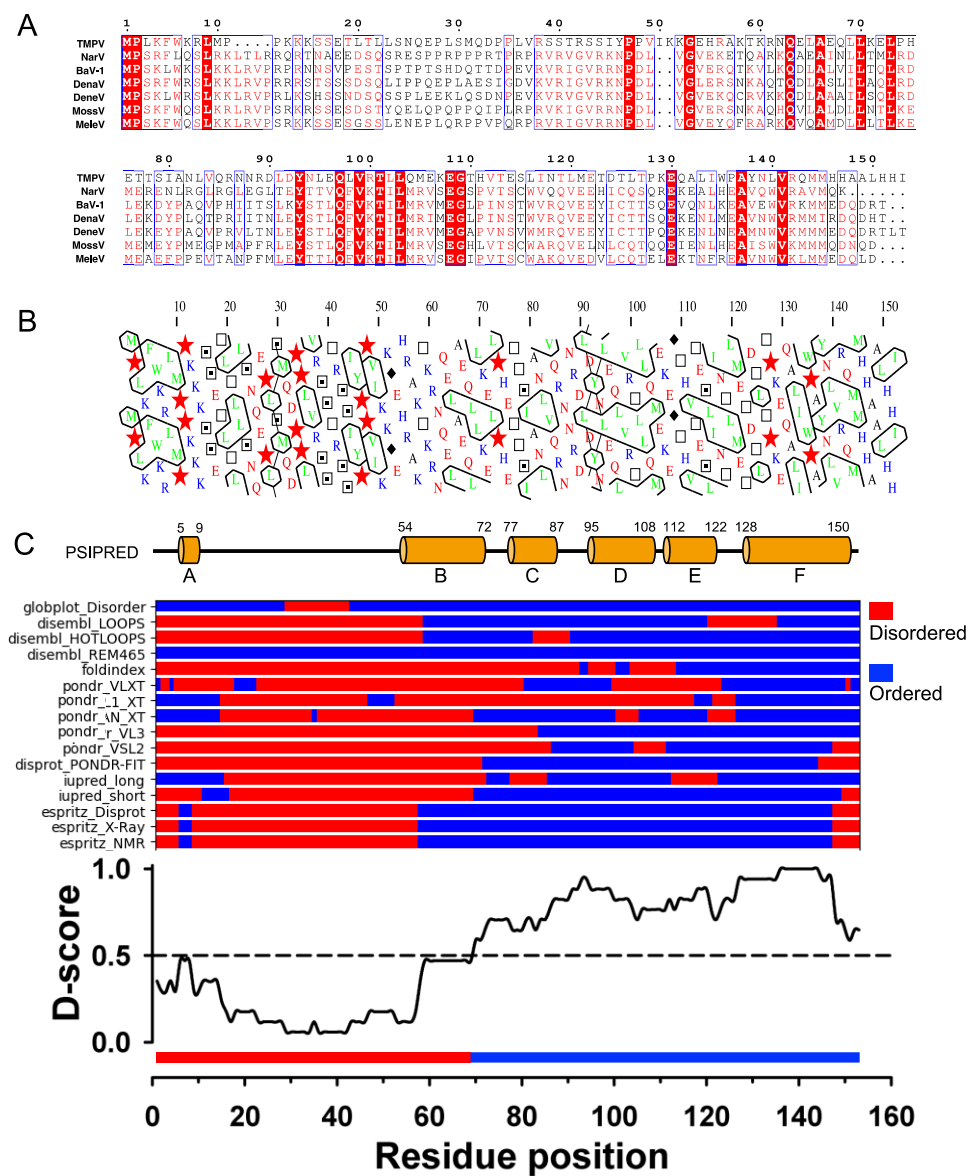
For the identification of TupV C-human protein-protein interactions, DNAs encoding the C protein fused to a tandem affinity purification (TAP) tag sequence at either the amino or carboxy terminus were transfected into HEK293T cells. Briefly, 40 µg of DNA (combining at equal parts constructs with the tag at the N- and C-terminus) was added to 1.5 mL of serum free medium (Gibco, Waltham, MA, United States) (solution A). As 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, United States) (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) (x3). 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 (x3). The cell extract was clarified by centrifugation (10 min, 10,000 x g) and the supernatant 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 a 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 6 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).



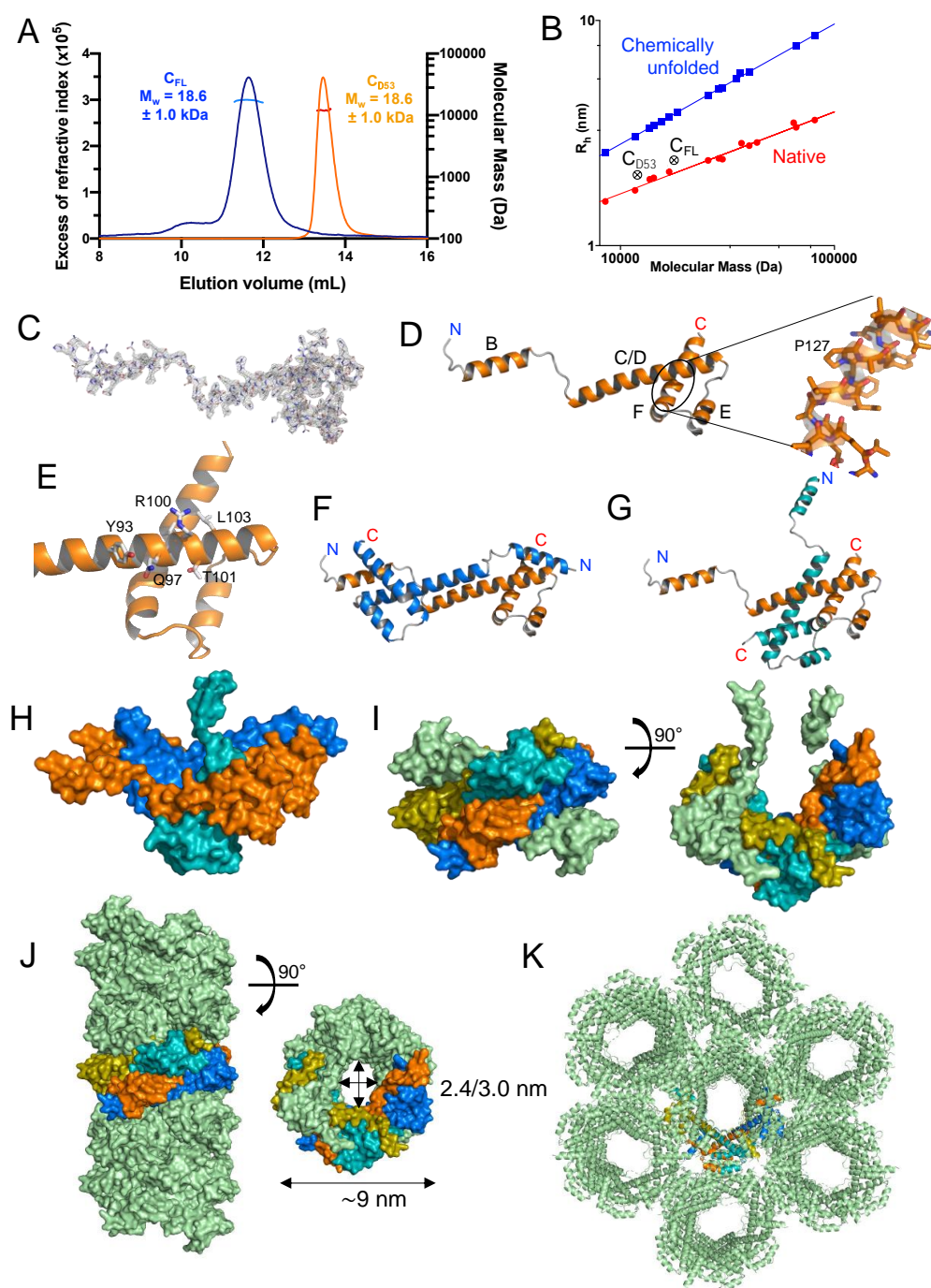
**Figure 2.** TupV C protein sequence analysis. (A) Multiple sequence alignment of Narmovirus C protein. Members of the *Narmovirus* genus and their UniProt accession number: TupV - Tupaia paramyxovirus - Q9WS38; NarV - Nariva narmovirus - B8XH61; BaV-1 -Myodes narmovirus (Bank Vole virus-1) - A0A2H4PJ60; 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.

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 of 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.



**Figure 3.** Crystal structure of TupV C protein C-terminal domain ( $C_{D53}$ ). (A) SEC-MALLS of TupV  $C_{\Delta 53}$  and  $C_{FL}$ . 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 17807 Da for C<sub>FL</sub> and 11897 Da for C<sub>D53</sub>, respectively). (B) Plot of the hydrodynamic radius measured by SEC as a function of the 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  $\sigma$  and stick representation of the TupV C<sub>D53</sub> protein in the crystal structure (PDB ID: 8BJW). (D) Cartoon representation of TupV C<sub>D53</sub> in the crystal. The N and C-terminal residues of C<sub>D53</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 are the same as in panel F and G. (I) Surface representation in two orthogonal orientations of one asymmetric unit containing 6 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 if the tubular assembly seen in the crystal packing. The same four protomers described in panels F to 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.

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

The protein C<sub>A53</sub> crystallized readily with six molecules in the asymmetric unit in the P3<sub>2</sub>21 space group and diffracted at 2.4  $\text{\AA}$  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  $\text{\AA}$  (Table 1 and Figure 3C).

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

|  | C <sub>A54</sub><br>native | C <sub>A54</sub><br>Pt derivative |
|--|----------------------------|-----------------------------------|
| Data collection                                |                            |                                   |
| Beamline                                       | SOLEIL - PX1               | SOLEIL - PX1                      |
| Wavelength ( $\text{\AA}$ )                    | 1.07169                    | 1.07169                           |
| Space group                                    | P3 <sub>2</sub> 21         | P2 <sub>1</sub>                   |
| a,b,c ( $\text{\AA}$ )                         | 135.7, 135.7, 63.1         | 135.6, 135.6, 64                  |
| $\alpha,\beta,\gamma$ ( $^\circ$ )             | 90, 90, 120                | 90, 90, 120                       |
| Resolution range ( $\text{\AA}$ ) <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/ $\sigma$ I <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 - 17210                   |
| Unique reflections <sup>1</sup>                | 27 136 - 2 824             | 18927 - 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 ( $\text{\AA}$ )                   | 0.0099                     |                                   |
| Bond angle ( $^\circ$ )                        | 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 webserver.

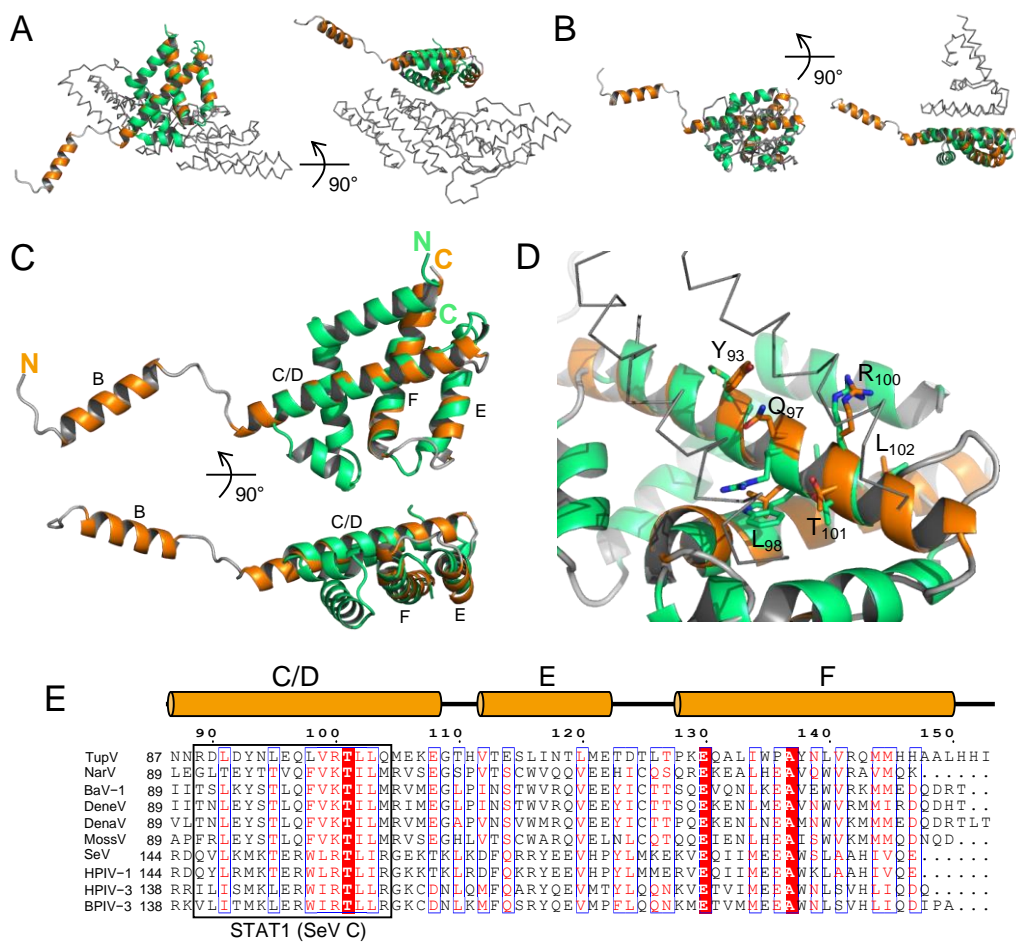
Each polypeptide chain forms four  $\alpha$ -helices (Figure 3D): helices B, C/D (a long helix so named because it encompassed predicted helices C and D), E and F. Helix B and the N-terminal half of helix C/D are stabilized by intermolecular interactions with other C proteins (see below). Helix F is 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 an hydrophobic core. Of note, the residues in the C-terminal part of 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 packed against each other in two different modes. In the first mode (Figure 3F), two molecules of C<sub>A53</sub> assembled head-to-tail, forming 14 interchain H-bonds, 2 salt bridges and burying  $\sim 3,300 \text{ \AA}^2$  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/I69, 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  $\sim 1200 \text{ \AA}^2$  of accessible surface area (PISA Complex formation Significance Score of 1.0).

The assembly of three protomers by both modes is shown in Figure 3H. The asymmetric unit comprised 6 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 an external diameters of  $\sim 9.0 \text{ nm}$  and an hexagonal internal cavity of  $2.4\text{--}3.0 \text{ nm}$  width (Figure 3J). In the crystal, the tubular helices 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,  $3.3 \text{ \AA}$  rmsd over 62 CA atoms), a key intermediate in the interferon signaling pathway [60], or in complex with the BRO-1 like domain of the human protein Alix (PDB ID: 6KP3 - Z-score 3.5,  $2.2 \text{ \AA}$  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 6 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 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 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 *Respirovirus* C proteins. Members of the *Narmovirus* genus and their UniProt accession number are the same as in Figure 2A, members of the *Respirovirus* 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).

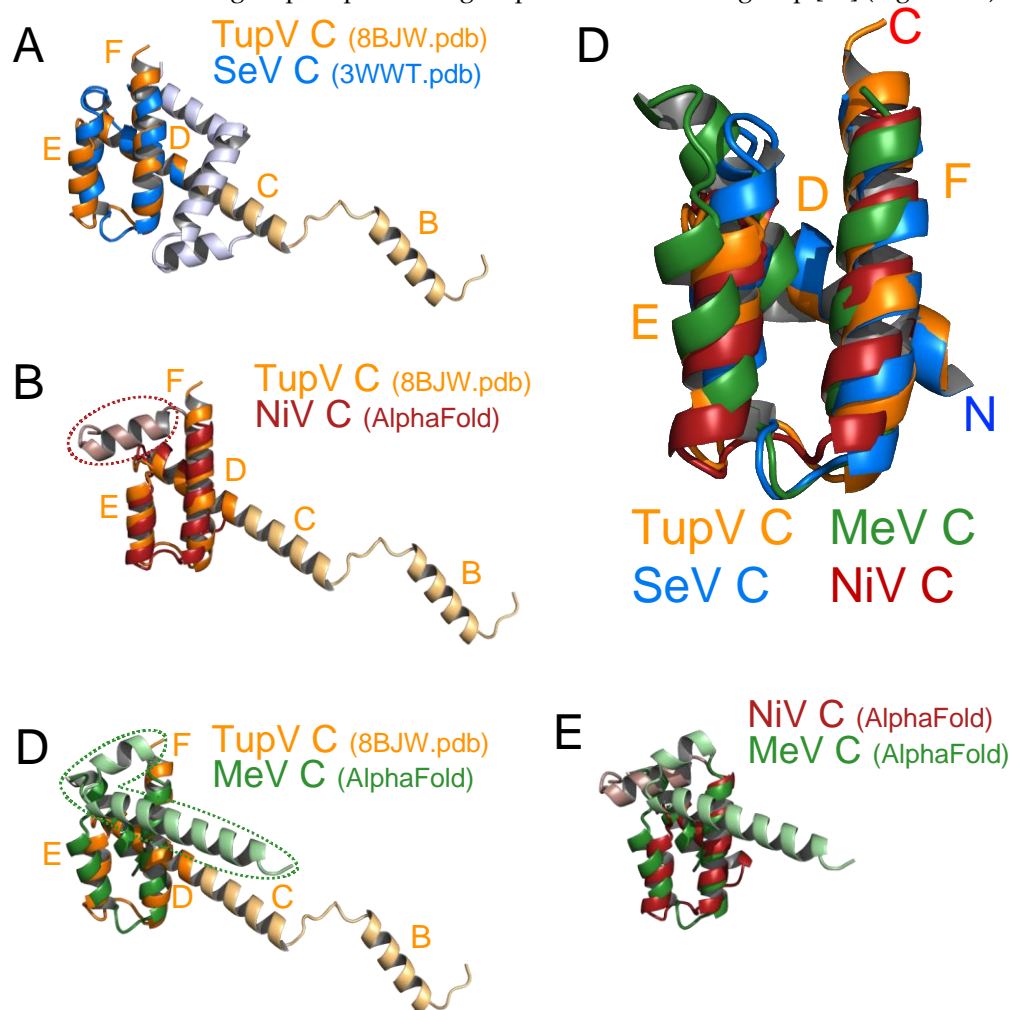
In contrast, the interaction of SeV C protein with the BRO-1 like domain of Alix, mediated by a L122XXW125 motif located in the second  $\alpha$ -helix of SeV C, has no equivalent in the TupV C protein



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* contain the same 3 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 3 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, Nipah virus group and Sendai virus group [22] (Figure 5D).

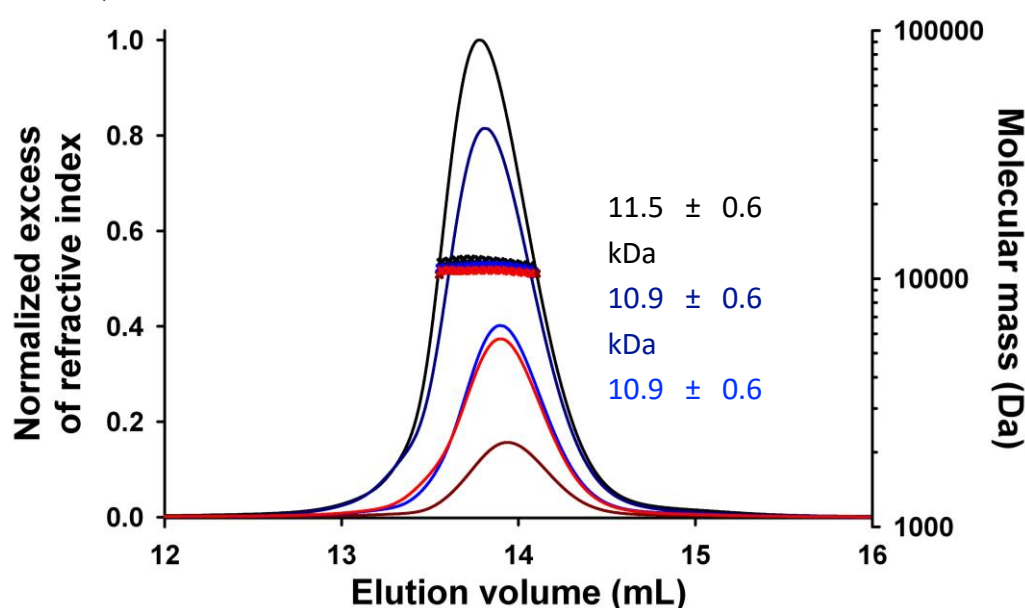


**Figure 5.** Structural alignments. (A) Superposition of TupV C (8BJW.pdb) and SeV C (taken from 3WWT.pdb) crystal structures. The non- superposable parts of TupV and SeV C are respectively in light blue and light orange. (B) Superposition of the crystal structure of TupV C with an AlphaFold model of NiV C (aa 100-166). The averaged pLDDT 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.

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<sub>Δ53</sub> does not Multimerize in Solution, Unlike in the Crystal*

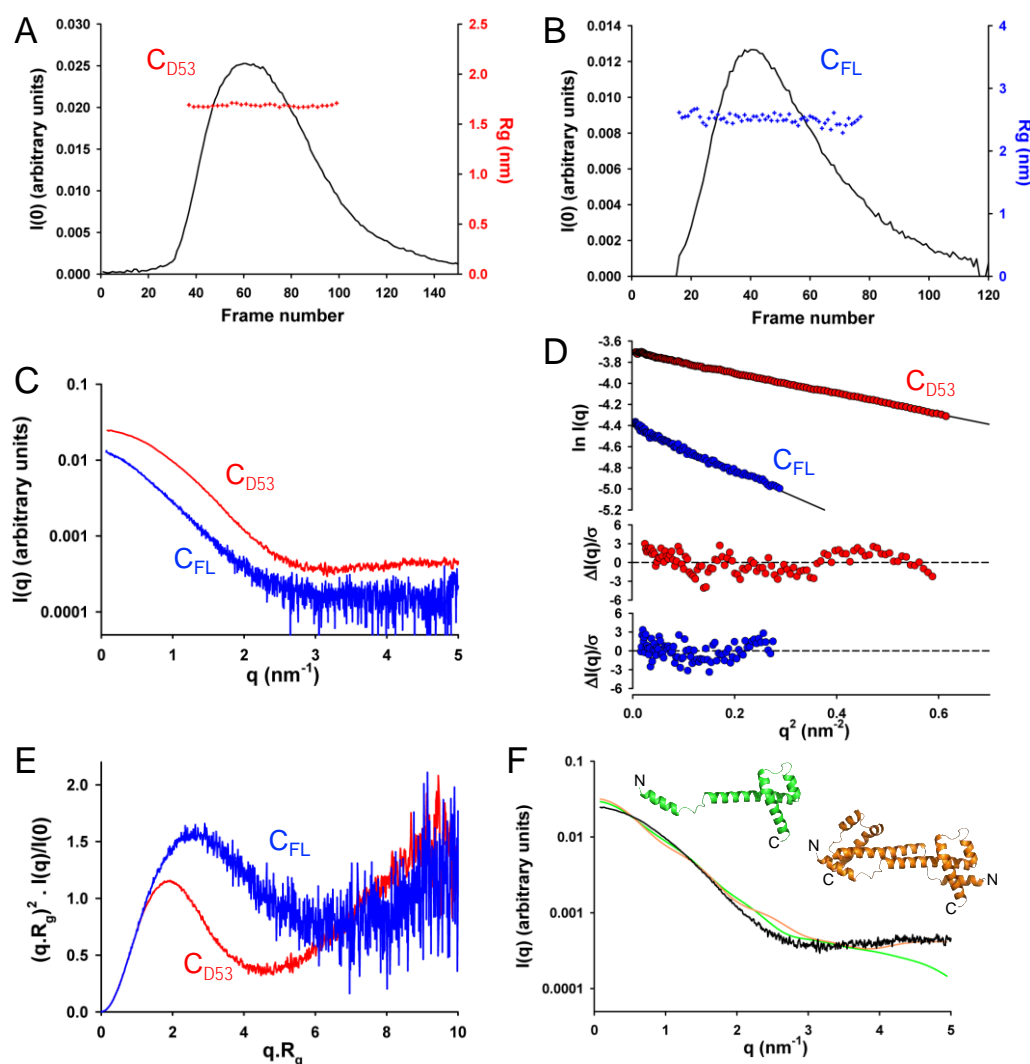
To determine whether TupV C<sub>Δ53</sub> self-assembled in solution, we first analyzed the dependence of the elution on protein concentration by SEC-MALLS. Protein samples at concentration ranging from 3.3 mg.mL<sup>-1</sup> to 23.5 mg.mL<sup>-1</sup> were injected on the Superdex 75 column (Figure 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<sub>Δ53</sub> 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%).



**Figure 6.** SEC MALLS experiments at different TupV C<sub>Δ53</sub> protein concentrations. 50 mL of protein solution were injected onto a Superdex 75 column at different initial concentration: 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 ( $M_w$ ).

### 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<sub>Δ53</sub> and obtain information about the N-terminal disordered region of the full-length protein in solution, we performed SEC-SAXS experiments with both C<sub>Δ53</sub> and C<sub>FL</sub>. 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 nm<sup>-1</sup> (Figure 7A,B).



**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 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 crosses show the radius of gyration calculated from the Guinier plots at the different time intervals. **(B)** SEC-SAXS analysis of TupV CFL 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 CFL (in blue) by averaging the individual profiles recorded throughout the SEC elution peak shown in panels A and B. **(D)** Guinier plots. The upper panel shows the Guinier plots for C $\Delta$ 53 (in red) and CFL (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 CRYSOLO 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.

For both proteins, the Guinier plots at low  $q$  values ( $q.R_g < 1.3$ ) 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  $1.7 \pm 0.1$  nm for C $\Delta$ 53 and  $2.5 \pm 0.1$  nm for CFL by using the Guinier approximation (Figure 7D). In the dimensionless Kratky plot for C $\Delta$ 53, the curve reached a maximum near 1.5 for  $q.R_g$  values around 2.6 and rose at  $q.R_g$  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  $q.R_g$  values above 2 and decreased slightly at higher  $q.R_g$  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.** Small-angle X-ray Scattering Data Collection and Analysis.

|  | C-Full length            | C- $\Delta 53$                |
|--|--------------------------|-------------------------------|
| Data collection parameters                         |                          |                               |
| <i>Instrument</i>                                  | SOLEIL - SWING           |                               |
| <i>Energy (keV)</i>                                | 12.000                   |                               |
| <i>Detector</i>                                    | EigerX-4M                |                               |
| <i>Detector distance (m)</i>                       | 2                        |                               |
| <i>Exposure (s per image)</i>                      | 1                        |                               |
| <i>q range (<math>\text{\AA}^{-1}</math>)</i>      | 0.008 - 0.50             |                               |
| <i>Column</i>                                      | S75inc 5/150 GL          |                               |
| <i>Flow rate (mL.min<sup>-1</sup>)</i>             | 0.3                      |                               |
| <i>Sample concentrations (mg.mL<sup>-1</sup>)</i>  |                          |                               |
| <i>Injection volume (<math>\mu\text{L}</math>)</i> | 50                       | 50                            |
| <i>Temperature (K)</i>                             | 293                      | 293                           |
| Guinier Analysis                                   |                          |                               |
| <i>I(0) (A.U.)</i>                                 | $1,2.10^{-2} \pm 3,0.10$ | $2,4.10^{-2} \pm 1,4.10^{-5}$ |
| <i>R<sub>g</sub> (nm)</i>                          | $2.48 \pm 0.11$          | $1.70 \pm 0.02$               |
| <i>q<sub>rgmin</sub></i>                           | 0.2951                   | 0.2702                        |
| <i>q<sub>rgmax</sub></i>                           | 1.3070                   | 1.3027                        |
| P(R) analysis <sup>1</sup>                         | $26.98 \pm 0.20$         | $17.60 \pm 0.02$              |
| <i>D<sub>max</sub> (nm)</i>                        | $10.5 \pm 0.6$           | $6.4 \pm 0.3$                 |
| Molecular weight <sup>2</sup>                      |                          |                               |
| <i>Theoretical Mw (kDa) <sup>3</sup></i>           | 17 807                   | 11 897                        |
| <i>Measured Mw (kDa)</i>                           | 18,6                     | 12,3                          |
| <i>Deviation (%)</i>                               | 4.3                      | 3.3                           |
| EOM <sup>4,5</sup>                                 |                          |                               |
| <i>Nbr of unique models</i>                        |                          |                               |
| $\chi^2$   |                          |                               |
| SASBDB code  |                          |                               |

<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. <sup>4</sup> Using default parameters <sup>5</sup> Using 10,000 full atoms models.

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.

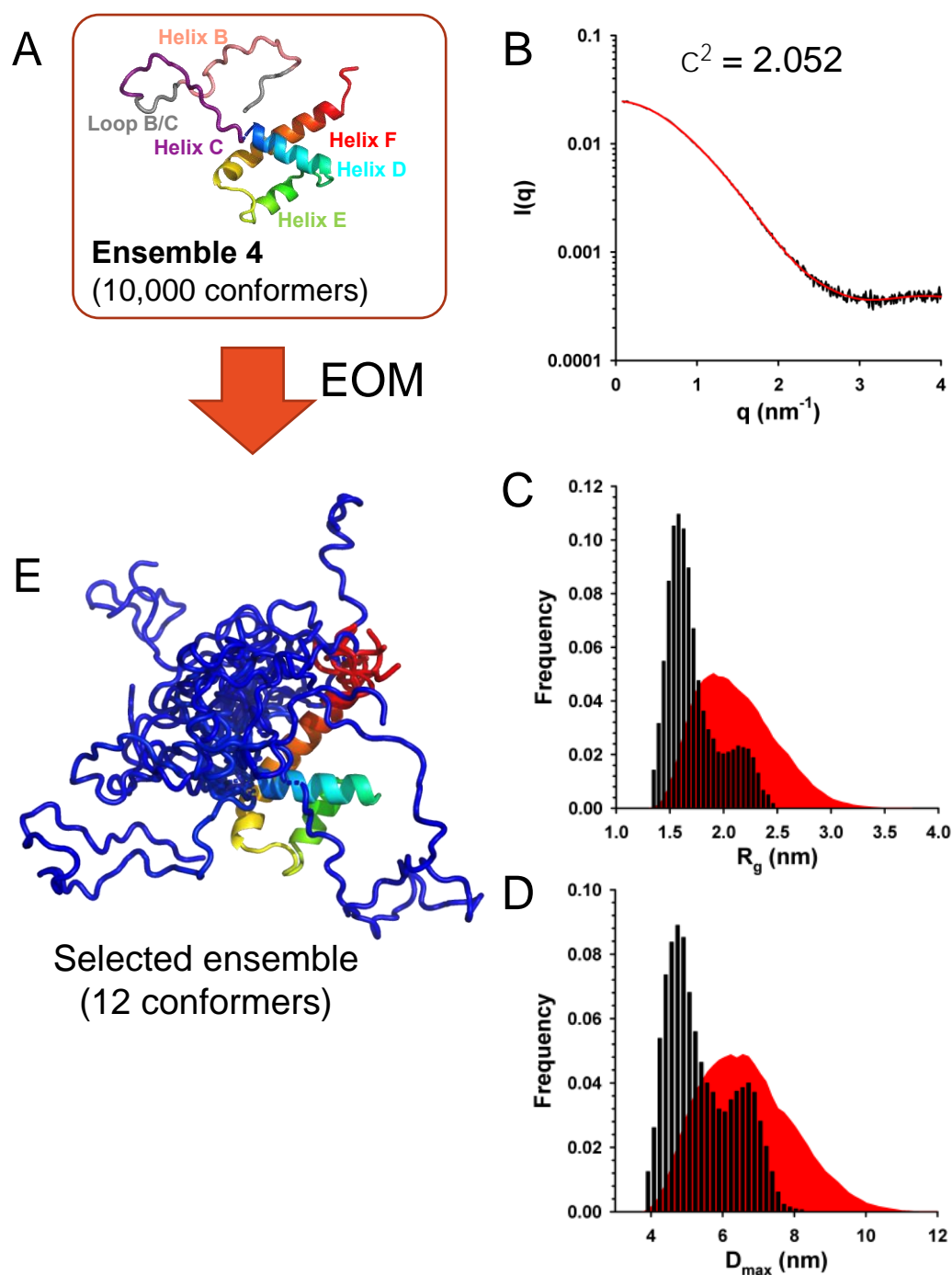
### 3.7. In Solution, TupV C is well Described by Ensembles of Conformers with a Disordered N-Terminal and Fluctuating Helices B and C

To further explore the conformational diversity of  $C_{\Delta 53}$ , we used the Ensemble Optimization Method (EOM) [45]. This approach consists in 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.



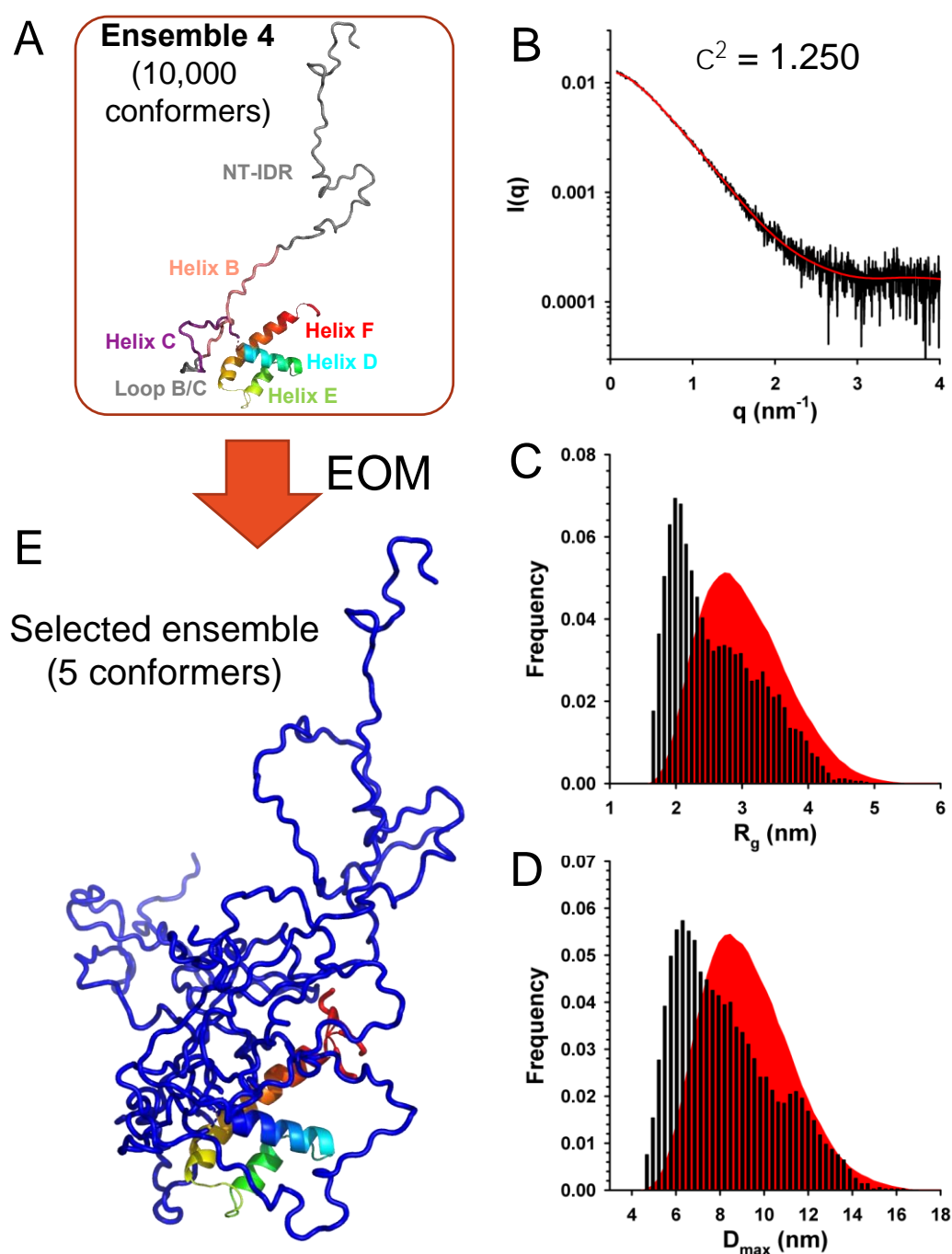
**Figure 8.** Conformational ensemble modeling of TupV C $\Delta$ 53 protein from SAXS data. (A) Description of the initial ensemble, which is comprised of 10,000 conformers in which the regions corresponding to the helix B (in pink), the helix C (in purple) and the loop connecting them (in grey) were 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.

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 (Figures 8C and 8D).

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 forms, 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 than for  $C_{\Delta 53}$  but with 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 5 conformers is shown in Figure 9E.



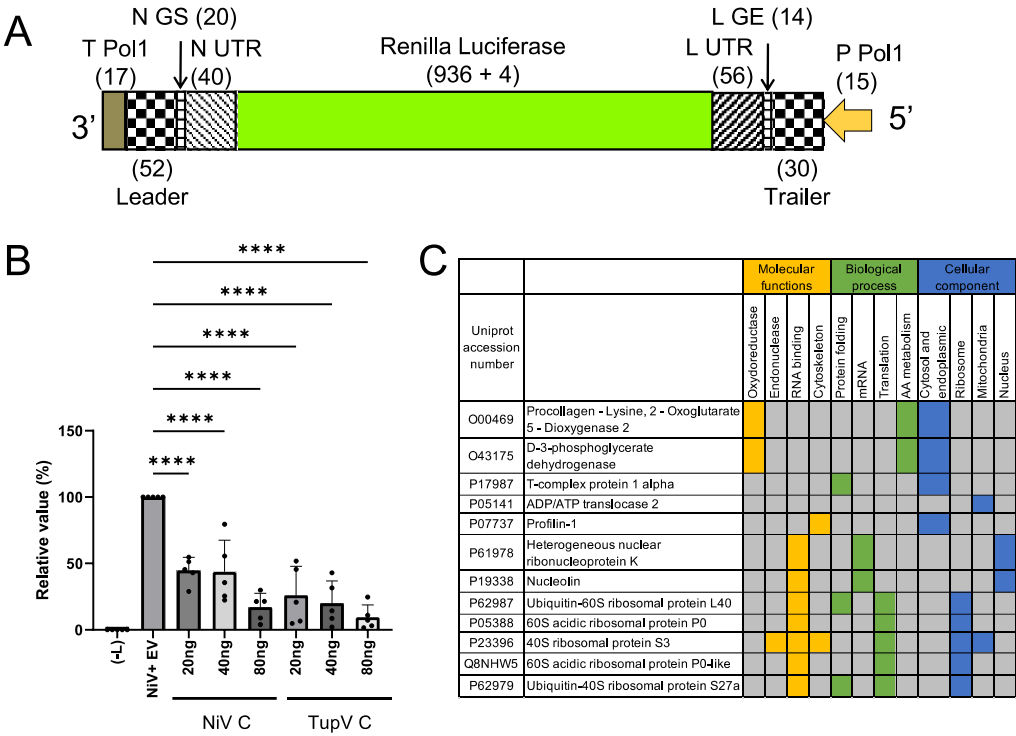
**Figure 9.** Conformational ensemble modeling of TupV C<sub>FL</sub> protein from SAXS data. (A) Description of the initial ensemble, which is comprised of 10,000 conformers in which the regions corresponding to the N-terminal intrinsically disordered region (NT-IDR in grey), the helix B (in pink), 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<sub>FL</sub> (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 5 conformers of TupV C<sub>FL</sub>. The red area shows the distribution for the initial ensemble of conformers. The black bars show the  $R_g$  distribution for selected ensembles of 5 conformers. (D)  $D_{max}$  distribution for an ensemble of 5 conformers of TupV C<sub>FL</sub>. The red area shows the distribution for the initial ensemble of conformers. The black bars show the  $D_{max}$  distribution for selected ensembles of 5 conformers. (E) Representative ensemble of 5 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<sub>Δ53</sub>) and full-length (C<sub>FL</sub>) 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 Nipah virus.

For this purpose, we constructed a minigenome for Nipah virus, in which the gene of the Renilla Luciferase was under the control of 3' and 5' untranslated regions of Nipah virus (Figure 10A), and we developed a minireplicon assay. 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 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 Pol1 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) Effect of NiV and TupV C protein on NiV minireplicon 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 pCDNA3 vector. At 24 hour after transfection, luciferase activities were measured, 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 % (A).



For relative Firefly expression analysis, NiV+EV values were set to 100%. Individual values are represented by dots. Mean and SD from 5 independent experiments are indicated. \*\*\*\* $p < 0.0001$ ; ns not significant (One-way ANOVA, Dunnett Test). (C) 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 function involved in several biological processes and localized in different cellular components.

We found that both NiV and TupV C proteins inhibited the expression of Renilla luciferase in a concentration dependent manner (Figure 10B). 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 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 10C). Interestingly, none of these proteins were common to those previously identified by the same approach for 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.

## 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 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>D53</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 surface.

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 has 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 *Respirovirus* 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 *Respirovirus*, 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:

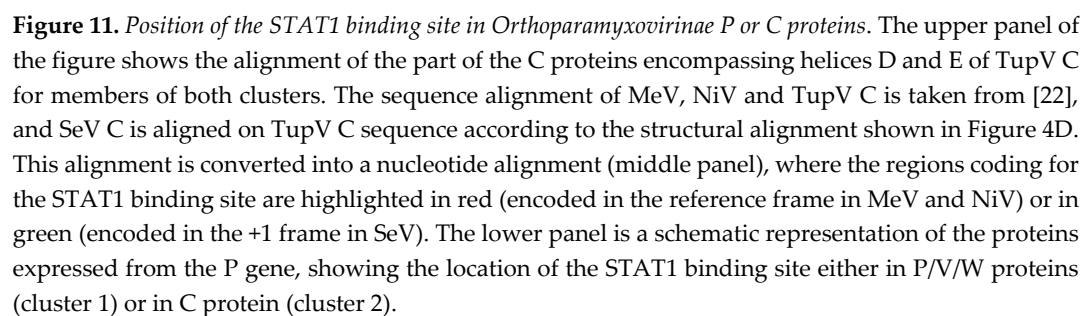
- (a) The C coding region appeared in the common ancestor of *Orthoparamyxovirinae* and later diverged in sequence beyond recognition in respiroviruses and aquaparamyxoviruses. In this scenario, all C proteins are homologous.
- (b) The C coding region appeared independently in the ancestor of respiroviruses 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 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



**Author Contributions:** A.R., E.C.M., V.V., F.B., L.M.G., V.V., D.K., J.M.B. and M.J. conceived the experiments; A.R., E.C.M., L.G., C.L., V.V., F.B. and J.M.B. performed the experiments; A.R., E.C.M., V.V., L.M.G., V.V., D.K., J.M.B. and M.J. analyzed and interpreted the data; A.R., D.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).

**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

1. Lamb, R.A.; Parks, G.D. *Paramyxoviridae*: The viruses and their replication. In B.N. Fields, D.M. Knipe and P.M. Howler (Eds.); Lippincott Williams & Wilkins. *Fields Virology sixth Edition* **2013**, *1*, 957–995.
2. Eaton, B.T.; Broder, C.C.; Middleton, D.; Wang, L. Hendra and Nipah Viruses: Different and Dangerous. *Nat Rev Microbiol* **2006**, *4*, 23–35. <https://doi.org/10.1038/nrmicro1323>.
3. 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*, 1–12.
4. 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. <https://doi.org/10.1038/ncomms1796>.
5. 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. <https://doi.org/10.1038/d41586-019-00599-8>.
6. Kuhn, J.H.; Adkins, S.; Alioto, D.; Alkhovsky, S.V.; Amarasinghe, G.K.; et al. Taxonomic Update for Phylum Negarnaviricota (Riboviria: Orthornavirae), Including the Large Orders Bunyavirales and Mononegavirales. *Arch Virol* **2020**, *165*, 3023–3072. <https://doi.org/10.1007/s00705-020-04731-2>.
7. Pringle, C.R. The Order Mononegavirales—Current Status. *Arch Virol* **1997**, *142*, 2321–2326.
8. Douglas, J.; Drummond, A.J.; Kingston, R.L. Evolutionary History of Cotranscriptional Editing in the Paramyxoviral Phosphoprotein Gene. *Virus Evol* **2021**, *7*, 1–34. <https://doi.org/10.1093/ve/veab028>.
9. Hausmann, S.; Garcin, D.; Delenda, C.; Kolakofsky, D. The Versatility of Paramyxovirus RNA Polymerase Stuttering. *J Virol* **1999**, *73*, 5568–5576.
10. Vidal, S.; Curran, J.; Kolakofsky, D. A Stuttering Model for Paramyxovirus P mRNA Editing. *Embo J* **1990**, *9*, 2017–2022.
11. 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.
12. 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.
13. 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. <https://doi.org/10.1006/viro.1999.9693>.
14. 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. <https://doi.org/10.1128/JVI.01203-12>.
15. Siering, O.; Sawatsky, B.; Pfaller, C.K. C Protein Is Essential for Canine Distemper Virus Virulence and Pathogenicity in Ferrets. *J Virol* **2021**, *95*. <https://doi.org/10.1128/jvi.01840-20>.
16. 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. <https://doi.org/10.1128/jvi.02572-13>.



17. 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*, 1–15. <https://doi.org/10.1128/jvi.01733-19>.
18. 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.
19. 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.
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. <https://doi.org/10.1128/JVI.05899-11>.
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*, 1–22. <https://doi.org/10.1371/journal.ppat.1005659>.
22. Lo, M.K.; Søgaard, T.M.; Karlin, D.G. Evolution and Structural Organization of the C Proteins of Paramyxovirinae. *PLoS ONE* **2014**, *9*, e90003. <https://doi.org/10.1371/journal.pone.0090003>.
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*. <https://doi.org/10.1038/ncomms2416>.
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. <https://doi.org/10.24272/j.issn.2095-8137.2017.033>.
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. <https://doi.org/10.1016/j.jmb.2009.03.061>.
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. <https://doi.org/10.1093/bioinformatics/bti537>.
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. [https://doi.org/10.1002/1097-0134\(20010101\)42](https://doi.org/10.1002/1097-0134(20010101)42).
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. [https://doi.org/S1570-9639\(10\)00013-0](https://doi.org/S1570-9639(10)00013-0).
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. <https://doi.org/10.1002/prot.10532>.
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. <https://doi.org/10.1002/prot.20735>.
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. <https://doi.org/10.1093/bioinformatics/bti541>.
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. <https://doi.org/10.1016/j.str.2003.10.002>.
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. <https://doi.org/10.1093/nar/gkg519>.
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. <https://doi.org/10.1093/bioinformatics/btr682>.
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. <https://doi.org/10.1007/s000180050082>.
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. <https://doi.org/10.1038/nsmb.2868>.
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**, 1–19. <https://doi.org/10.1016/j.bpj.2020.04.010>.
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. <https://doi.org/10.1016/j.jmb.2008.07.028>.
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. <https://doi.org/10.1128/JVI.00999-13>.

40. Studier, F.W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr Purif* **2005**, *41*, 207–234. <https://doi.org/10.1016/j.pep.2005.01.016>.
41. Wyatt, P.J. Submicrometer Particle Sizing by Multiangle Light Scattering Following Fractionation. *J Colloid Interface Sci* **1998**, *197*, 9–20. <https://doi.org/10.1006/jcis.1997.5215>.
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. <https://doi.org/10.1021/bi00211a042>.
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., 2022; Vol. 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. <https://doi.org/10.1107/S1600576717007786>.
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. <https://doi.org/10.1021/ja069124n>.
46. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125–132. <https://doi.org/10.1107/S0907444909047337> [pii].
47. Pape, T.; Schneider, T.R. HKL2MAP: A Graphical User Interface for Macromolecular Phasing with SHELX Programs. *J Appl Cryst.* **2004**, *37*, 843–844. <https://doi.org/10.1107/S0021889804018047>.
48. Adams, P.D.; Afonine, P.v.; Bunkoczi, 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. <https://doi.org/10.1107/S0907444909052925>.
49. Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 2126–2132. <https://doi.org/10.1107/S0907444904019158>.
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. <https://doi.org/10.1107/S0907444911001314>.
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. <https://doi.org/10.1107/S0907444910045749>.
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**. <https://doi.org/10.1101/2021.10.04.463034>.
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. <https://doi.org/10.1038/s41586-021-03819-2>.
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, doi:JVI.02610-08.
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. <https://doi.org/10.1099/vir.0.18876-0>.
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. <https://doi.org/10.1099/vir.0.19685-0>.
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. <https://doi.org/10.1128/msphere.00376-16>.
58. Buchan, D.W.A.; Jones, D.T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Res* **2019**, *47*, W402–W407. <https://doi.org/10.1093/nar/gkz297>.
59. Holm, L. Dali Server: Structural Unification of Protein Families. *Nucleic Acids Res* **2022**, *50*, W210–W215. <https://doi.org/10.1093/nar/gkac387>.
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. <https://doi.org/10.1128/JVI.01887-15>.
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*. <https://doi.org/10.1128/jvi.00815-21>.
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. <https://doi.org/10.1093/nar/gkn072>.
63. Svergun, D.I. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Cryst.* **1992**, *25*, 495–503. <https://doi.org/10.1107/S0021889892001663>.

64. Hansen, S. Bayesian Estimation of Hyperparameters for Indirect Fourier Transformation in Small-Angle Scattering. *J. Appl. Cryst* **2000**.
65. Martinez-Gil, L.; Vera-Velasco, N.M.; Mingarro, I. Exploring the Human-Nipah Virus Protein-Protein Interactions. *J Virol* **2017**, *91*, e01461-17. <https://doi.org/10.1128/JVI.01461-17>.
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. *Molecular and Cellular Proteomics* **2021**, *20*, 100049. <https://doi.org/10.1016/J.MCPRO.2021.100049>.
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.
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. <https://doi.org/10.1128/jvi.77.21.11842-11845.2003>.
69. Devaux, P.; von Messling, V.; Songsunthong, W.; Springfield, C.; Cattaneo, R. Tyrosine 110 in the Measles Virus Phosphoprotein Is Required to Block STAT1 Phosphorylation. *Virology* **2007**, *360*, 72–83. <https://doi.org/10.1016/j.virol.2006.09.049>.
70. Wright, B.W.; Molloy, M.P.; Jaschke, P.R. Overlapping genes in natural and engineered genomes. *Nat Rev Genet.* **2022**, *23*, 154–168. <https://doi.org/10.1038/s41576-021-00417-w>.

**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.