



**HAL**  
open science

# The structure of the nucleoprotein of Influenza D shows that all Orthomyxoviridae nucleoproteins have a similar NPCORE, with or without a NPTAIL for nuclear transport

Amélie Donchet, Justine Oliva, Alice Labaronne, Laura Tengo, Myriam Miloudi, Francine Gérard, Caroline Mas, Guy Schoehn, Rob W H Ruigrok, Mariette Ducatez, et al.

## ► To cite this version:

Amélie Donchet, Justine Oliva, Alice Labaronne, Laura Tengo, Myriam Miloudi, et al.. The structure of the nucleoprotein of Influenza D shows that all Orthomyxoviridae nucleoproteins have a similar NPCORE, with or without a NPTAIL for nuclear transport. *Scientific Reports*, 2019, 9, pp.600. 10.1038/s41598-018-37306-y . hal-01997471v2

HAL Id: hal-01997471

<https://hal.univ-grenoble-alpes.fr/hal-01997471v2>

Submitted on 8 Jun 2021

**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.



Distributed under a Creative Commons Attribution 4.0 International License

# SCIENTIFIC REPORTS

OPEN

## The structure of the nucleoprotein of Influenza D shows that all *Orthomyxoviridae* nucleoproteins have a similar NP<sub>CORE</sub>, with or without a NP<sub>TAIL</sub> for nuclear transport

Amélie Donchet<sup>1</sup>, Justine Oliva<sup>2</sup>, Alice Labaronne<sup>1</sup>, Laura Tengo<sup>1</sup>, Myriam Miloudi<sup>1</sup>, Francine C.A. Gerard<sup>1</sup>, Caroline Mas<sup>3</sup>, Guy Schoehn<sup>1</sup>, Rob W.H. Ruigrok<sup>1</sup>, Mariette Ducatez<sup>2</sup> & Thibaut Crépin<sup>1</sup>

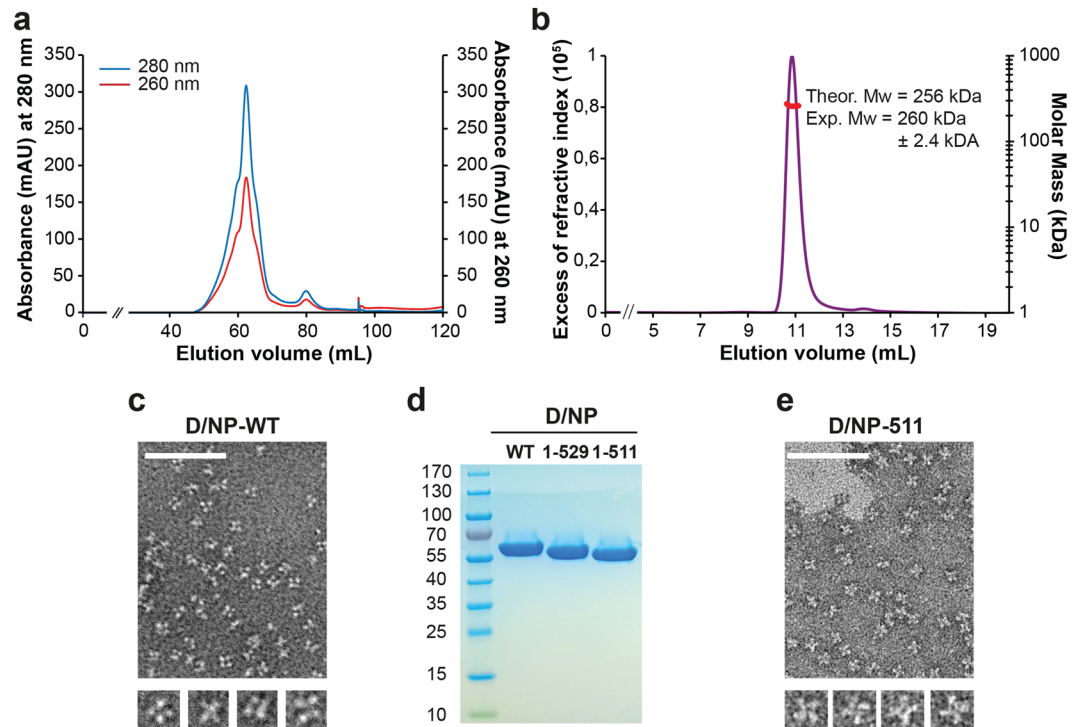
This paper focuses on the nucleoprotein (NP) of the newly identified member of the *Orthomyxoviridae* family, Influenza D virus. To date several X-ray structures of NP of Influenza A (A/NP) and B (B/NP) viruses and of infectious salmon anemia (ISA/NP) virus have been solved. Here we purified, characterized and solved the X-ray structure of the tetrameric D/NP at 2.4 Å resolution. The crystal structure of its core is similar to NP of other Influenza viruses. However, unlike A/NP and B/NP which possess a flexible amino-terminal tail containing nuclear localization signals (NLS) for their nuclear import, D/NP possesses a carboxy-terminal tail (D/NP<sub>TAIL</sub>). We show that D/NP<sub>TAIL</sub> harbors a bipartite NLS and designed C-terminal truncated mutants to demonstrate the role of D/NP<sub>TAIL</sub> for nuclear transport.

In 2011, a virus was isolated from a pig with Influenza-like symptoms in Oklahoma (USA). Electron microscopy showed features of an Influenza virus particle and real-time RT-PCR revealed that it was neither an Influenza A virus (IAV) nor an Influenza B virus (IBV). Next-generation sequencing analyses allowed the identification of RNA segments with around 50% identity to human Influenza C virus (ICV). Further serological analyses showed that antibodies against this new virus failed to cross-react with IAV, IBV or ICV (1). All these results suggested it was a new genus of the *Orthomyxoviridae*, temporarily named C/swine/Oklahoma/1334/2011 (C/swine/OK) and then Influenza D virus (IDV). IDV is widely distributed around the world; it was detected in North America<sup>1-4</sup>, Asia<sup>5-7</sup>, Europe<sup>8-10</sup> and Africa<sup>11</sup>. A serological study in Nebraska (USA) found a seroprevalence of 80% in cattle serum from 2003<sup>3</sup>.

The *Orthomyxoviridae* family includes different *genera*, Influenza A, B, C and D viruses, infectious salmon anemia (ISA) virus, Thogoto virus, Quarantavirus and others. These viruses are segmented negative strand RNA viruses. Their genomes are made of a set of RNA segments coated with multiple copies of the nucleoprotein (NP) and associated to the viral heterotrimeric polymerase. The number of vRNA segments is specific to each type of Influenza viruses and related to the number of glycoproteins at the surface of the viral particle, 8 segments for IAV and IBV, 7 for IVC and IVD and 6 for Thogoto virus. These ribonucleoproteins (RNPs) are competent for both transcription and replication. To date, several X-ray structures of NP have been published, each of them without RNA. There are three structures of Influenza A nucleoprotein (A/NP): two with NP assembled as a trimer<sup>12,13</sup> and one for the monomeric R416A mutant<sup>14</sup>. The X-ray structure of the tetrameric Influenza B nucleoprotein (B/NP) is also known<sup>15</sup>, whereas the structure of Isavirus NP (ISA/NP) was solved as a dimer<sup>16</sup>. The overall folds of A/NP and B/NP are very similar with a root-mean-square deviation (rmsd) of 1.6 Å.

<sup>1</sup>Institut de Biologie Structurale (IBS), Univ. Grenoble Alpes, CEA, CNRS, 38044, Grenoble, France. <sup>2</sup>IHAP, Université de Toulouse, INRA, ENVT, Toulouse, France. <sup>3</sup>Integrated Structural Biology Grenoble (ISBG) - UMS 3518 (CNRS-CEA-UJF-EMBL), 38044, Grenoble, France. Correspondence and requests for materials should be addressed to T.C. (email: [thibaut.crepin@ibs.fr](mailto:thibaut.crepin@ibs.fr))

Received: 19 June 2018  
Accepted: 30 November 2018  
Published online: 24 January 2019



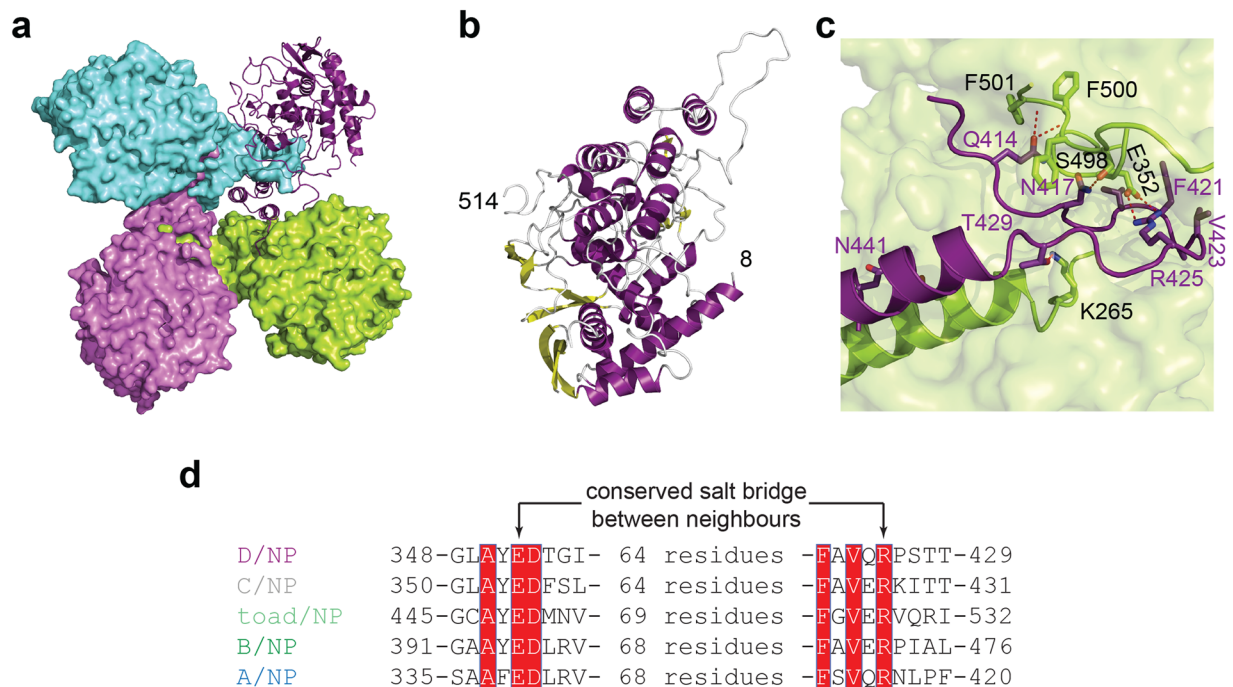
**Figure 1.** Purification and characterized of Influenza D nucleoprotein. **(a)** Size exclusion chromatography profile of wild-type D/NP. The sample was loaded on a Hiload™ 16/600 S200 column equilibrated with the running buffer 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 5 mM  $\beta$ -mercaptoethanol. **(b)** SEC-MALLS-RI analysis of D/NP. SEC was performed with a Superdex™ 200 increase 10/300 GL column equilibrated with 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 5 mM  $\beta$ -ME. The panel shows the theoretical Mw and the measured Mw. **(c)** and **(e)** Electron microscopy images of the elution peak of D/NP and D/NP-511. Samples show different oligomeric states although most oligomers are tetramers. The scale bar corresponds to 100 nm. **(d)** Coomassie blue-stained SDS-PAGE (4–20% gradient polyacrylamide) showing the purified wild-type D/NP and the two C-terminal truncated mutants (D/NP-529 and D/NP-511).

Influenza viruses transcribe and replicate in the nucleus of the infected cells and the NPs and the polymerase subunits need to interact with the nuclear transport system of the cell. For Influenza A and B viruses, several studies have shown that the nuclear localization signals (NLS) of NP recognized by the cellular importins- $\alpha$ , are located within the flexible N-terminal tail<sup>17–24</sup>. Recently, crystal structures of the two NLSs of A/NP bound to importin- $\alpha$  have been solved<sup>25,26</sup>.

In this paper, we characterized D/NP and solved the X-ray structure of its tetramer. The C-terminal D/NP<sub>TAIL</sub> harbouring a classical bipartite nuclear localization signal (NLS) was not visible in the structure. We designed two C-terminal truncated mutants (D/NP-511 and D/NP-529) to study the interaction of D/NP with importin- $\alpha$ . Our biochemical experiments demonstrate that D/NP<sub>TAIL</sub> is involved in the interaction with importins- $\alpha$  and immunofluorescence showed that the wild-type D/NP goes into the nucleus whereas the mutants stay in the cytoplasm.

## Results

**Recombinant D/NP forms a tetramer in solution.** The DNA coding sequence of D/NP was cloned in a bacterial expression plasmid and over-expressed in *Escherichia coli* as a C-terminal His-tagged recombinant protein. D/NP was purified with a nickel affinity chromatography followed by a heparin column and a final gel filtration. Figure 1a shows a typical gel filtration elution profile using the absorbance signals at 260 and 280 nm (ratio 280/260 > 1.75). The protein could then be concentrated at 2 to 6 mg.mL<sup>-1</sup>. Polyacrylamide gels and SEC-MALLS-RI experiments have confirmed the homogeneity and the molecular weight of the recombinant tetrameric D/NP (Fig. 1b,d). By electron microscopy (negative staining), we showed that D/NP forms mainly tetramers in solution (Fig. 1c). Previously, it was shown that the oligomerization of recombinant A/NP and B/NP can be modulated by the NaCl concentration<sup>14,18,27,28</sup>. Starting from purified and stable oligomeric samples (trimers for A/NP and tetramers for B/NP), monomeric proteins can be obtained by decreasing stepwise the NaCl concentration. After purification at 300 mM NaCl followed by dialysis at 150 mM NaCl, D/NP was eluted from gel filtration (with a 150 mM NaCl running buffer) in the same volume, meaning that a smooth reduction of the salt concentration did not change its oligomeric state. However, a decrease to 50 mM NaCl induced an irreversible and total precipitation of D/NP, even with a stepwise reduction at 150 mM NaCl. Therefore, the experiments on D/NP and its mutants were carried out at 300 mM NaCl.



**Figure 2.** Structure of Influenza D nucleoprotein. **(a)** Structure of the tetrameric D/NP, with 3 protomers shown in surface (respectively in green, pink and cyan) and the fourth in cartoon (deep purple). **(b)** Cartoon representation of one monomer of D/NP with the  $\alpha$ -helices in deep purple and  $\beta$ -strands in yellow. **(c)** Detail of the interactions between two protomers of the tetrameric D/NP as shown in **(a)**. The conserved R425 of one protomer (shown as deep purple cartoon) stabilized the position of the oligomerization loop at the surface of the neighbouring protomer, through its conserved E352. **(d)** Sequence alignment of the salt bridges of the oligomerization-loop of one protomer to the NP<sub>CORE</sub> of the neighbour protomers. For the sequences see Table 1.

	A/NP	B/NP	toad/NP	C/NP	D/NP	ISA/NP
A/NP		1.6 Å (383 C $\alpha$ )	—	—	2.1 Å (356 C $\alpha$ )	2.9 Å (340 C $\alpha$ )
B/NP	38%		—	—	1.9 Å (365 C $\alpha$ )	2.8 Å (340 C $\alpha$ )
toad/NP	25%	30%			—	—
C/NP	22%	25%	24%		—	—
D/NP	24%	25%	23%	38%		2.9 Å (330 C $\alpha$ )
ISA/NP	20%	18%	18%	16%	20%	
Tho/NP	18%	21%	20%	18%	17%	20%
WfB/NP	20%	19%	19%	20%	18%	17%

**Table 1.** Nucleoproteins of Orthomyxoviruses. A/NP, Influenza A virus nucleoprotein, strain A/WSN/1933(H1N1) (Uniprot accession number B4URF1, PDB entry 2IQH); B/NP, Influenza B virus nucleoprotein, strain B/Managua/4577.01/2008 (Uniprot accession number C4LQ26, PDB entry 3TJ0); Toad/NP, Wuhan asiatic toad Influenza virus nucleoprotein (GenBank accession number AVM87634); C/NP, Influenza C virus nucleoprotein, strain C/AnnArbor/1/1950 (Uniprot accession number Q6I7C0); D/NP, Influenza D virus nucleoprotein, strain D/bovine/France/2986/2012 (Uniprot accession number A0A0E3VZU8, PDB entry 5N2U); ISA/NP, infectious salmon anemia virus nucleoprotein, isolate salmon/Norway/810/9/99 (Uniprot accession number Q8V3T7, PDB entry 4EWC); Tho/NP, Thogoto virus nucleoprotein (Uniprot accession number A0A0B6VKB5); WfB/NP, Wellfleet Bay virus nucleoprotein (Uniprot accession number A0A0A1E9N5). Sequence identities were obtained using MUSCLE<sup>77</sup> and structure comparisons were calculated using PDBeFold<sup>78</sup>.

Like the nucleoproteins of A, B and ISA, D/NP bound single-stranded RNA. We measured the  $K_d$  of 14 nM with a fluorescence anisotropy assay with an RNA of 24 nucleotides of polyUC labeled in 3' with 6-fluorescein amidite (FAM)<sup>27</sup>. The  $K_d$  of 14 nM can be compared with the  $K_d$ s of 7 nM of the trimer of A/NP, 31 nM of the tetramer of B/NP<sup>29</sup> and 24 nM for the dimer of ISA/NP<sup>16</sup>.

**Structure of D/NP.** Full length D/NP was crystallized in sodium malonate as small fine needles that diffract X-rays up to 2.4 Å resolution (Supplementary Table 1). The structure was solved by molecular replacement using a starting model of the monomer of the A/NP R416A mutant<sup>14</sup>. The automatic search gave the position for three

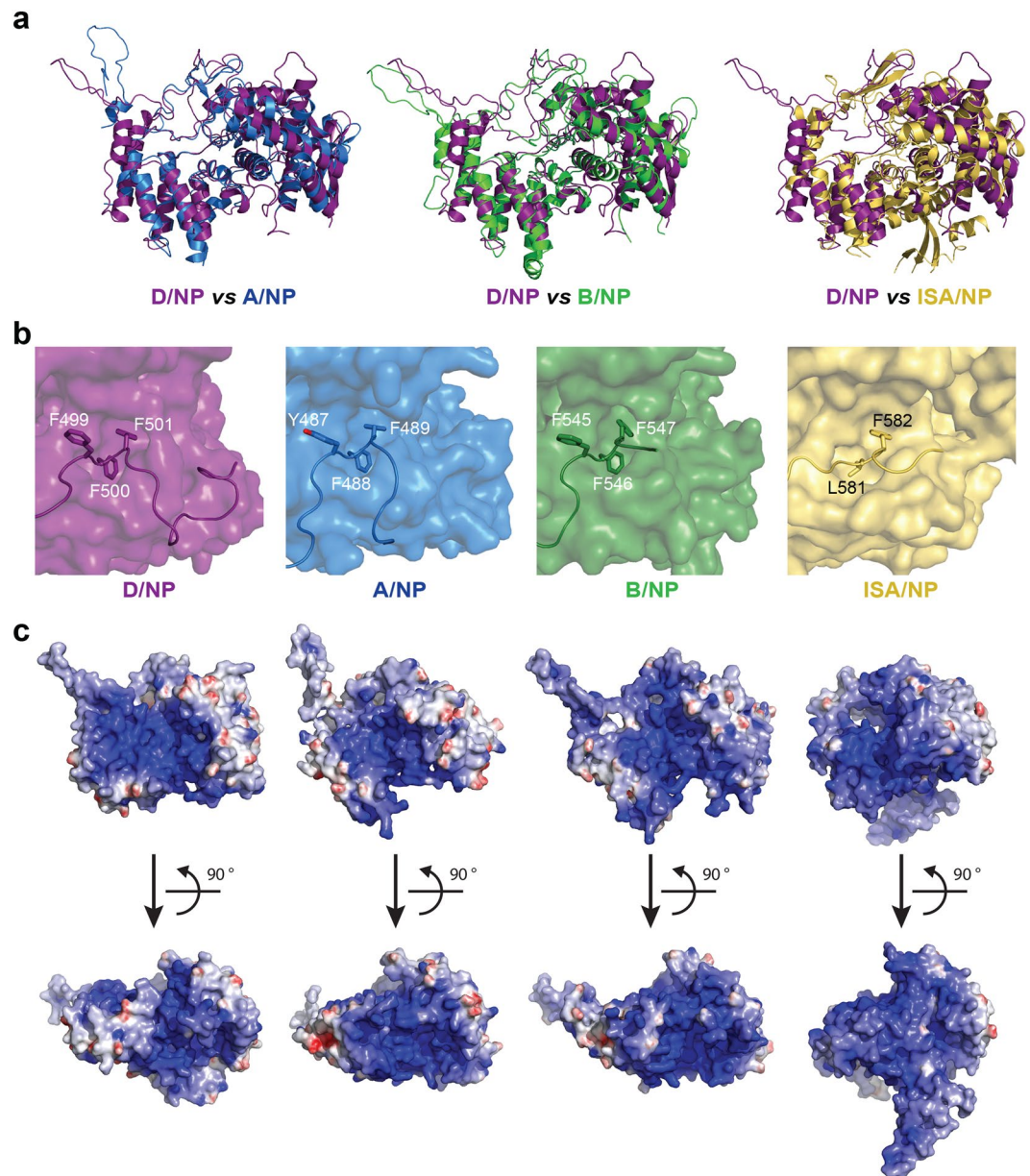
molecules and a fourth protomer was fitted manually after the analysis of the electron density. The molecular replacement using the tetrameric B/NP model did not give any acceptable solutions. The asymmetric unit contained four molecules arranged as a tetramer (Fig. 2a and refinement statistics in Supplementary Table 1), consistent with the observations made by electron microscopy. Based on the electron density, 87% of the model could be built, from residues 8 to 514 (Fig. 2b). Several internal loops were missing and the C-terminal 50 residues were disordered. The oligomerization-loop of one protomer plugged into a cavity of its neighbour (Fig. 2c). The N-terminal part of the loop started with a strictly conserved glutamate residue (Gln-414) that interacted with the backbone of the strictly conserved consecutive aromatic residues  $_{499}\text{FFF}_{501}$ . The loop was then stabilized by several kind of contacts; mainly hydrophobic (by the side chain of the strictly conserved Phe-421 and Val-423) and of several aromatic residues lying in the pocket of the neighbouring monomer but also the salt bridge between an arginine from one monomer (Arg-425) interacting with a glutamate of the neighbour (Glu-352; Fig. 2c,d). These two residues, also strictly conserved in all Influenza NPs (Fig. 2d and Supplementary Fig. 1), are well known for modulating their oligomeric state. Once mutated in alanine, the corresponding single mutants formed monomers, unable to self-assemble<sup>13,14,28,30–33</sup>. Phosphorylation state of A/NP has been shown to also be a key factor of NP/NP interfaces, especially for positions Ser-165 and Ser-407<sup>34,35</sup>. Sequence alignment showed that D/NP contains a threonine (Thr-161) and a serine (Ser-416) respectively, suggesting a conserved mechanism for the regulation of the oligomerization (Supplementary Fig. 1).

**Core structure of all nucleoproteins (NP<sub>CORE</sub>) with tails (NP<sub>TAIL</sub>).** The overall folds of A/NP, B/NP, ISA/NP and D/NP were very similar (Fig. 3a) with root-mean-square deviations (rmsd) between 1.6 Å (for 383 C $\alpha$  for the comparison A/NP with B/NP) and 2.9 Å (for 340 C $\alpha$  for the comparison D/NP with ISA/NP, Table 1). Including ISA/NP, a common architecture for all Influenza-like NPs can be defined based on the X-ray structures with the NP<sub>CORE</sub> starting with the first  $\alpha$ -helix for A/NP, B/NP and D/NP and finishing with three hydrophobic residues, anchored into the surface of the protein (Supplementary Fig. 1). Figure 3b shows the X-ray structures of A/NP, B/NP, D/NP and ISA/NP with a zoom on the C-terminal regions, where a superimposed patch made by three consecutive aromatic residues is found in Influenza NPs ( $_{487}\text{YFF}_{489}$  for A/NP,  $_{545}\text{FFF}_{547}$  for B/NP and  $_{499}\text{FFF}_{501}$  for D/NP). A similar patch is also present in the ISA/NP model ( $_{580}\text{GLF}_{582}$ ). The NP<sub>CORE</sub> contains the large and shallow positively charged surface which might bind RNA (Fig. 3c), without sequence specificity (see above)<sup>12,13,36,37</sup>. Twenty or 71 residues without any structure are found N-terminally before the cores of A/ and B/NP whereas for D/NP, the core starts with a very short version of an N-terminal tail (only 7 residues), which seems to be compensated by the presence of a carboxy-terminal NP<sub>TAIL</sub> of 51 residues (Supplementary Fig. 1). The length of the NP<sub>TAILs</sub> is variable and its location seems to be specific for each genus. Because D/NP differs from A/ and B/NP in its global organization, we decided to further analyse the C-terminal D/NP<sub>TAIL</sub>.

**D/NP<sub>TAIL</sub> behaves as an intrinsically disordered protein.** Based on the crystal structure and a disorder prediction (Fig. 4a), we designed a construct (from residues 505 to 552) for the expression in *E. coli* as an N-terminal His-tagged recombinant version. Considering that the core of D/NP ends after residue Phe-501, we have chosen to start the construct at Gly-505, in order to avoid the hydrophobicity of Phe-503 (Fig. 4a,b). D/NP<sub>TAIL</sub> was eluted from gel filtration as a 40 kDa-protein (data not shown) but with a normal migration pattern on SDS-PAGE (Fig. 4b). As intrinsically disordered proteins are known to be eluted from gel filtration with an aberrant volume<sup>38,39</sup>, a SEC-MALLS-RI experiment (Fig. 4c) confirmed the monodispersity of the purified sample with a Mw of 8.2 kDa for D/NP<sub>TAIL</sub> (with the His-tag). A circular dichroism analysis confirmed that D/NP<sub>TAIL</sub> did not contain any significant stable secondary structure in solution (Fig. 4d), even in presence of TMAO (Supplementary Fig. 2), a chemical known to force unfolded proteins to fold to native-like species with significant functional activity<sup>40–42</sup>.

**D/NP<sub>TAIL</sub> interacts with importin- $\alpha$ .** The N-terminal tails of A/NP and B/NP are known to be involved in the nuclear import by interacting with importins- $\alpha$ <sup>17–19,21,23,25,43</sup>. The sequence analysis of D/NP showed no import signals in its N-terminal part but suggested the presence of a bipartite NLS within D/NP<sub>TAIL</sub> ( $_{514}\text{KRR-X}_{14}\text{-KKRGR}_{535}$ ; Fig. 4a). We thus tested the bipartite NLS with different constructs. First, we showed that the human importin- $\alpha 7$  is co-eluted with the His-tagged D/NP<sub>TAIL</sub> from gel filtration (Fig. 5a). The interaction was confirmed by thermal-shift experiments, where importin- $\alpha 7$  appeared more stable in presence of D/NP<sub>TAIL</sub> (Fig. 5b). We also confirmed the interaction between the two partners using the full-length D/NP. Using surface plasmon resonance, we measured a *Kd* of 100 nM between immobilized D/NP<sub>TAIL</sub> and importin- $\alpha 7$  (Fig. 5c), ten-fold higher than the affinities measured for the N-terminally A/ and B/NP<sub>TAILs</sub><sup>18,25</sup>. By gel filtration, we showed that the D/NP:importin- $\alpha 7$  complex could be eluted as a single peak. A molecular weight of 474 kDa has been measured by SEC-MALLS-RI for the complex (Fig. 5d), corresponding to the interaction of four molecules of importins- $\alpha 7$  (Mw = 4 × 55 kDa) per tetramer of D/NP (Mw = 260 kDa). To confirm the role of D/NP<sub>TAIL</sub> in the interaction with importin- $\alpha$ , we designed two C-terminal truncated constructs of D/NP, D/NP-511 and D/NP-529, ending respectively at residues 511 and 529. D/NP-511 is deprived of the whole bipartite NLS whereas D/NP-529 contains only the first part (Fig. 4a). They both behave as the wild-type protein during the purification (Fig. 1d,e). Using pull-down experiments on nickel-affinity resin, we showed that D/NP-529 still slightly retained importin- $\alpha 7$  whereas the interaction between D/NP-511 and importin- $\alpha 7$  was totally abolished (Fig. 5e). This suggests that importin- $\alpha 7$  binds to the NP<sub>TAIL</sub> of D/NP.

**Nuclear transport of D/NP and its mutants.** First, human HEK 293T cells were infected with D/bovine/France/5920/2014 (*moi* of 5) and indirect immunofluorescence was used to localize D/NP within the cells. After 6 h, most of D/NP was observed in the nucleus (Fig. 6a). As a control, HEK 293T cells were infected with Influenza A/PR/8/34 and we found very similar results (data not shown). We then transfected HEK 293T

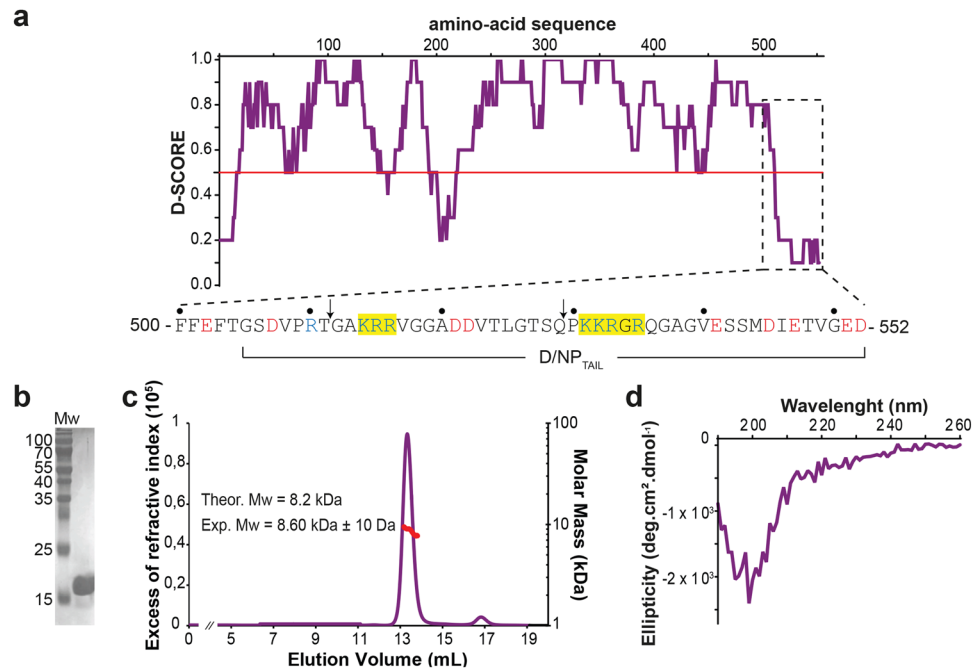


**Figure 3.** Comparison of D/NP with other segmented negative sRNA virus nucleoproteins. **(a)** The structure of one protomer of D/NP (deep purple) has been superimposed from left to right, with one protomer of A/NP (blue; PDB id: 2IQH), B/NP (forest; PDB id: 3TJ0) and ISA/NP (light orange; PDB id: 4EWC). The rmsd values are given in Table 1. **(b)** Anchoring of the C-terminus on NP<sub>CORE</sub> by a patch of 3 consecutive aromatic residues. The panel corresponds to a zoom of the superimposed structures shown in panel A with NP<sub>CORE</sub> represented in surface. **(c)** Electrostatic surface potentials of one protomer of D/NP, A/NP, B/NP and ISA/NP. The electrostatic surfaces were calculated from the crystal structures using DelPhi<sup>75</sup>. The potential scales range from  $-10.0$  kT/e (red) to  $10.0$  kT/e (blue).

cells with a plasmid containing wild-type D/NP and we found that D/NP can be localized in the cytoplasm but mainly in the nucleus (Fig. 6b). With this transfection strategy, D/NP-529 was located mainly in the cytoplasm with a little staining observed in the nucleus, whereas, D/NP-511 was located exclusively in the cytoplasm of the transfected cells (Fig. 6b), clearly showing that the cell used the bipartite NLS between residues 513 and 535 for the nuclear import of D/NP.

## Discussion

In 2011, a new virus has been isolated from pigs exhibiting Influenza-like illness. Subsequent studies identified an *Orthomyxovirus* with seven RNA segments sharing 50% overall amino acids identity with the human Influenza C virus<sup>2</sup>. First considered as a subtype of the Influenza C virus, it has since been officially named Influenza D virus (IDV) by the International Committee of Taxonomy of Viruses in 2016. Since its first isolation, IDV appears to

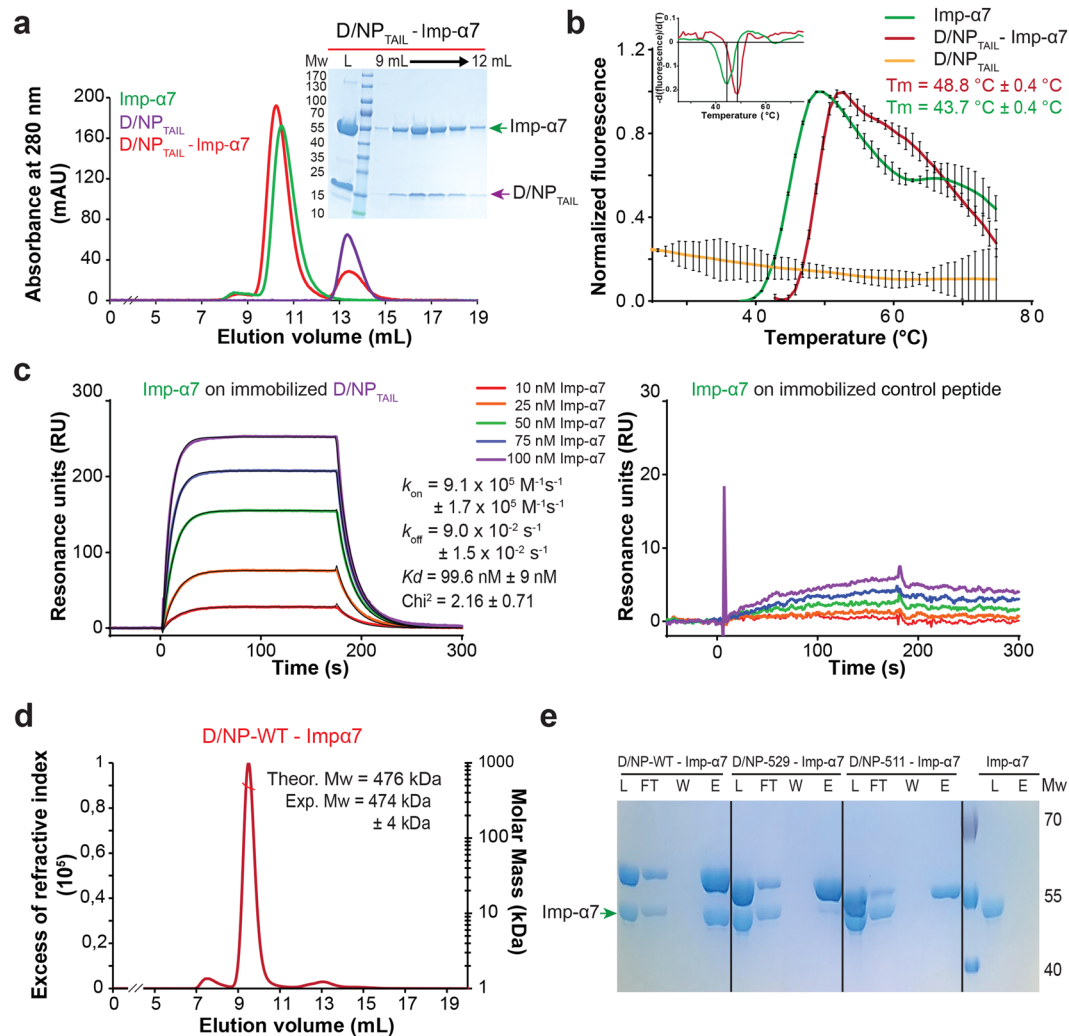


**Figure 4.** Biophysical characterization of D/NP<sub>TAIL</sub>. **(a)** D-score (score for disorder as a function of residue) of D/NP with a zoom (below the graph) on the last 50 residues. The prediction is based on 22 predictor web servers and the D-score was calculated by adding the values for each residue and dividing by the number of used algorithms. We arbitrarily defined a threshold level at 0.50; residues with a D-score < 0.50 were assigned as disordered<sup>38</sup>. The yellow boxes on the sequence are to highlight the putative NLS motifs. The arrows indicate where the sequences were cut for making D/NP-529 and D/NP-511. **(b)** Coomassie blue-stained SDS-PAGE (Tris-Tricine; 15% polyacrylamide) of the purified D/NP<sub>TAIL</sub>. It migrates at a higher molecular weight (17 kDa approximately) than expected (8 kDa). **(c)** SEC-MALLS-RI analysis of D/NP<sub>TAIL</sub> loaded on S75 10/300 GL column. For this experiment, we have chosen to keep the His-tag encoded with the pETM11 plasmid, for an optimal detection of D/NP<sub>TAIL</sub> with UV. The experimental molecular weight is consistent with the expected mass. **(d)** Circular dichroism of D/NP<sub>TAIL</sub>. CD is a biophysical method based on the polarization of light, used for a fast determination of the secondary structures within the proteins in solution.  $\alpha$ -Helical proteins show negative bands at 222 nm and 208 nm and a positive band at 193 nm, proteins with well-defined antiparallel  $\beta$ -sheets have negative bands at 218 nm and positive bands at 195 nm and disordered proteins have very low ellipticity above 210 nm and negative bands near 195 nm.

circulate all around the world in many mammals, with cattle as a possible reservoir, and with serological data suggesting it may transmit to humans<sup>44</sup>. The recent structure of IDV Hemagglutinin-Esterase-Fusion glycoprotein (HEF) has described an open receptor-binding cavity capable of accommodating diverse extended glycan moieties that could be one reason for its broad cell tropism<sup>45</sup>.

With this paper, we provide a detailed characterization of IDV nucleoprotein. We determined the structure of the tetramer of D/NP and we failed to find the biochemical conditions to stabilize the monomer. For A and B/NP we had found that the monomer stays in solution at low salt, 50 mM NaCl<sup>14,27,28</sup> but the D/NP protein makes aggregates in salt conditions lower than 150 mM. Post-translation changes like phosphorylation or ubiquitination may be able to stabilise the monomer, as it has been shown for A/NP<sup>14,34,46,47</sup>. The phosphorylation on A and B/NP are summarized in Supplementary Fig. 1.

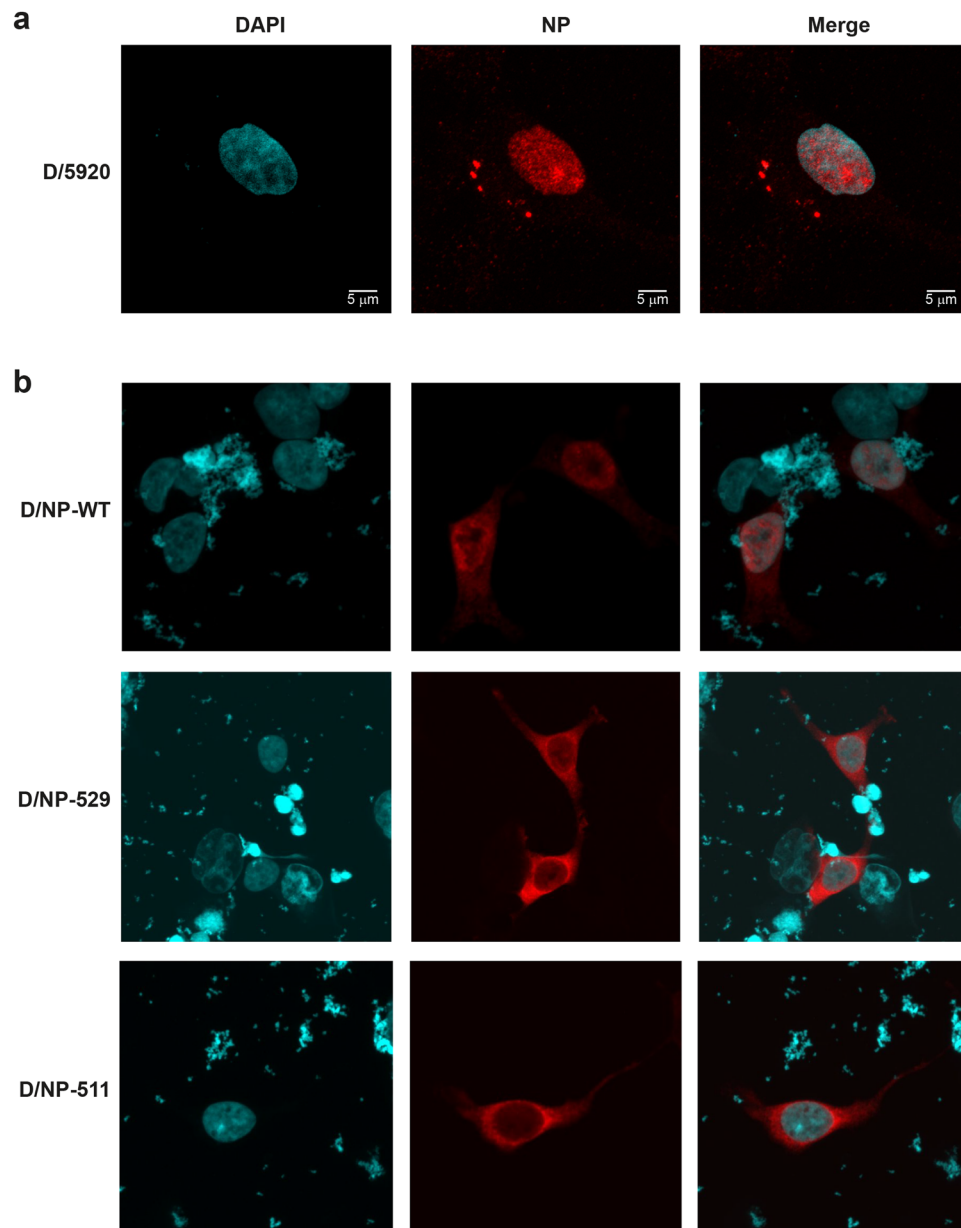
The comparison of the structures of the nucleoproteins of Influenza A, B, D and of infectious salmon anemia virus (ISAV) shows that of all these proteins share a common structural core (NP<sub>CORE</sub>) that start with the first  $\alpha$ -helix in the structure of A, B and D, up to three consecutive aromatic residues (Fig. 3 and Supplementary Fig. 1). Looking at the sequences of all known *Orthomyxoviruses* nucleoproteins shows that such hydrophobic patch is present in C/NP (501FFF<sub>503</sub>) and could be present in Tho/NP (451YLF<sub>453</sub>) and Wfb/NP (Wellfleet bay virus nucleoprotein; 532VIY<sub>534</sub>). The only structure that differs is the ISAV/NP that has a folded domain upstream of the first  $\alpha$ -helix of the NP<sub>CORE</sub> (Fig. 7). One could also see that Tho/NP is constituted only by the NP<sub>CORE</sub> without any other appended sequences. NPs of IAV and IBV have an N-terminal NP<sub>TAIL</sub> whereas the NPs of ICV and IDV have a C-terminal NP<sub>TAIL</sub>. The NP<sub>TAIL</sub> of A is only 20 residues whereas the tails of the other Influenza viruses NPs are much longer; 71 residues for IBV, 62 for ICV, 51 for IDV. Recently, new large-scale methods for finding RNA viruses in vertebrates other than mammals and birds, identified up to 240 “new” viruses<sup>48</sup>. These authors found new Influenza viruses close to IBV, a Wuhan spiny eel Influenza virus and a Wuhan Asiatic toad Influenza virus. The NP for the toad virus (toad/NP) is very close to B/NP (identity of 30%, Table 1) with a very long N-terminal tail of 126 residues (Fig. 7 and Supplementary Fig. 1). Interestingly, toad/NP possess an alanine residue (Ala-278) aligned with A/NP Ser-165 and D/NP Thr-161 and a phenylalanine (Phe-519) aligned with A/NP Ser-407 and D/NP Ser-416.



**Figure 5.** Interaction of D/NP and D/NP<sub>TAIL</sub> with importin- $\alpha$ 7. **(a)** Size exclusion chromatography profile of a mixture between human importin- $\alpha$ 7 and D/NP<sub>TAIL</sub>. The mixture (molar ratio 1 importin- $\alpha$ 7 for 2 D/NP<sub>TAIL</sub>) was incubated 1 hour at room temperature and then loaded on a Superdex<sup>TM</sup>75 10/300GL column equilibrated with the running buffer 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM  $\beta$ -mercaptoethanol. **(b)** Thermal stability assay of importin- $\alpha$ 7 in absence (green) or in presence (red) of D/NP<sub>TAIL</sub> using ThermoFluor<sup>76</sup>. In presence of D/NP<sub>TAIL</sub>, the melting temperature of importin- $\alpha$ 7 is 5 °C higher. D/NP<sub>TAIL</sub> alone using ThermoFluor did not give any denaturation signal (yellow curve). The upper insert corresponds to the derivative of the fluorescence signal for a precise measure of the melting temperature. **(c)** Affinity of importin- $\alpha$ 7 for D/NP<sub>TAIL</sub> by measured by surface plasmon resonance (SPR). Biotinylated D/NP<sub>TAIL</sub> (left) and control peptide (right) were captured on a streptavidin-coated sensor chip surface before injections of several importin- $\alpha$ 7 concentrations (10 nM in red, 25 nM in orange, 50 nM in green, 75 nM in blue and 100 nM in purple). The sensorgrams of the interaction between D/NP<sub>TAIL</sub> and importin- $\alpha$ 7 were fitted under a Langmuir 1:1 binding model with mass-transfer (black line). **(d)** SEC-MALLS analysis of D/NP in complex with importin- $\alpha$ 7. The mixture (molar ratio 1 D/NP for 1.2 importin- $\alpha$ 7) was incubated 1 hour at room temperature and then loaded on a Superdex<sup>TM</sup> 200 increase 10/300 GL. The experimental molecular weight is consistent with the expected mass of four importins- $\alpha$ 7 bound per D/NP tetramer. **(e)** Pull-down assays of human importin- $\alpha$ 7 by D/NP and the two C-terminal truncated mutants (D/NP-529 and D/NP-511). The his-tags are on D/NP. The mixtures (molar ratio 1 D/NP for 1.2 importin- $\alpha$ 7) were incubated 1 hour and the experience was done as described in panel (a). The figure shows the coomassie blue-stained SDS-PAGE (12% polyacrylamide) with the Load, FlowThrough, Wash and the second fractions (E2).

We demonstrated that the C-terminal D/NP<sub>TAIL</sub> presents all the characteristics of an intrinsically disordered protein and that is involved in the nuclear import through its interaction with cellular importins- $\alpha$ . This role is not a big surprise since the presence of such NLS motifs was identified within the flexible NP<sub>TAILS</sub> of viruses A and B located at the N-terminus of these proteins<sup>17,18,21,23,25,43,49</sup>. Tho/NP has been shown to be imported in the nucleus<sup>50</sup> but this nucleoprotein does not possess any additional intrinsically disordered protein-like peptide with an NLS. Recently, a second NLS on A/NP (strain A/X-31) at the rim of the RNA binding channel was shown. A crystal structure shows that this internal NLS (198Lys-Arg-INDRNFWRGENC-Arg-Arg-T-Arg<sub>216</sub>) binds to mouse

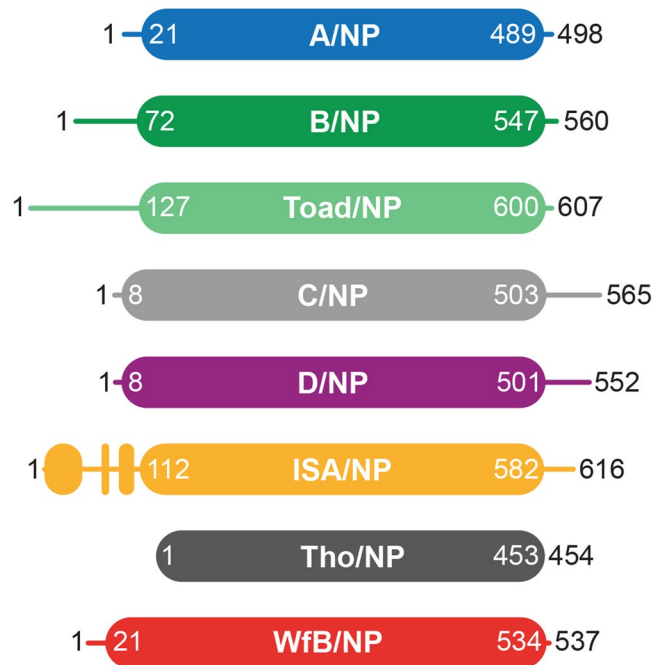




**Figure 6.** Nuclear transport of D/NP in HEK 293T cells. Microscopy pictures of HEK 293T cells with DAPI, NP and merged DAPI + NP staining (left, middle and right panels, respectively). **(a)** D/NP after infection of HEK 293T cells at 6 hours post infection with a *moi* of 5. **(b)** Cells were transfected with wt D/NP, D/NP-529 and D/NP-511. After 24 h cells were fixed, permeabilized and analyzed by indirect immunofluorescence. NP localization was observed using an in house rabbit hyperimmune NP-IDV serum and anti-rabbit IgG labeled Rhodamine RX (in red). Cells were mounted with DAPI-Vectashield and observed with a Leica Zeiss 710 (magnification: x63). In red: NP protein; in blue: nucleus. The bar in (A) represents 5  $\mu$ m.

importin- $\alpha$ 1<sup>26</sup>. Likewise, for the NLSs of Tho/NP and ISA/NP, it has been suggested that the rims on the RNA binding surfaces can be used as an NLS for nuclear import for viral replication<sup>16,50,51</sup>.

Finally, we have shown that the NP-TAILS of these nucleoproteins are intrinsically disordered proteins. The preservation of such intrinsically disordered protein-like peptide with a conserved function highlights its importance. The name for N<sub>TAIL</sub> was first used for the C-terminal disordered tail of measles virus nucleoproteins<sup>52</sup>, which is about 125 residues long. This tail is disordered but can adopt a short  $\alpha$ -helical structure (residue D484-A502) to promote the interaction with the phosphoproteins of these viruses which bind themselves to the corresponding RNA polymerase<sup>53,54</sup>. For these viruses, the tails keep the nucleocapsids in a flexible conformation. When cleaved off with trypsin, the helical nucleocapsids become rigid<sup>55,56</sup>. The tails do not change the affinity of the nucleoprotein for RNA but change the kinetics for the assembly of the nucleocapsids<sup>57</sup>. The RNPs of Influenza A and B viruses are also very flexible<sup>58–61</sup> which makes it difficult to handle for high resolution structures. The EM structure of the measles nucleocapsid with a resolution of 3.6 Å<sup>62</sup> was performed on the nucleocapsids without



**Figure 7.** Schemas for nucleoproteins of *Orthomyxoviruses*. Schematic representation of the nucleoprotein based on the amino acid sequence identity and structure analysis of the protein from representative members of the *Orthomyxoviridae* family. The schema respects the size of the proteins. The protein accession numbers are the same as Table 1. The flexible tails are represented with simple lines whereas the folded parts and the cores are represented with filled boxes.

the C-terminal  $N_{TAIL}$ . If we could generate Influenza RNPs with NP without  $NP_{TAIL}$  to rigidify the RNPs, it would make it easier solving a high resolution structure to observe the RNA on the nucleoproteins.

## Methods

**Molecular biology for expression of nucleoprotein of Influenza D.** The DNA coding sequence of D/bovine/France/2986/2012 NP was bought at GeneArt (ThermoFisher Scientific) optimized for bacterial expression. The different constructs have been PCR amplified and cloned following the supplier procedures (New England Biolabs). D/NP, D/NP-511 and D/NP-529 DNA coding sequences were cloned in pETM13 (EMBL) to express C-terminal His-tagged proteins whereas D/ $NP_{TAIL}$  (505-D/NP-552) DNA coding sequence was cloned in pETM11 (EMBL) to express an N-terminal His-tagged fragment. The DNA coding sequence of the human importin- $\alpha 7$  (KPNA6; Uniprot accession number O60684) was cloned without its IBB domain (amino-acids 58 to 536) in pET9a (Novagen) to express an N-terminal His-tagged protein<sup>63</sup>. Sequencing was performed by Eurofins.

**Expression and purification of proteins.** *Escherichia coli* BL21 RIL (DE3) cells (Life Technology) were transformed with the resulting plasmids. Cultures were induced 12 h by adding 0.3 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) at 18 °C and collected by centrifugation. The pellets were resuspended and sonicated in 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 M NDSB-201 (Sigma), 2 mM  $\beta$ -mercaptoethanol ( $\beta$ -ME) and complete protease inhibitor cocktail (Roche) for D/NP constructs and 50 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM  $\beta$ -ME and complete protease inhibitor cocktail for importin- $\alpha 7$ . All purifications were performed at room temperature. All D/NP constructs were purified by nickel affinity chromatography (Ni-NTA, Qiagen) followed, in the case of D/NP, D/NP-511 and D/NP-529, by a HiTrap<sup>TM</sup> Heparin HP column (GE-Healthcare) on a NGC system (BioRad). Elution fractions of D/NP, D/NP-511, D/NP-529 and D/ $NP_{TAIL}$  were dialyzed against 20 mM Tris-HCl pH 7.5 at 50 mM or 300 mM NaCl, 5 mM  $\beta$ -ME. The last purification step was a size-exclusion chromatography using a Hiload<sup>TM</sup> 16/600 S200 column (GE-Healthcare) for D/NP, D/NP-511 and D/NP-529 or a S75 10/300 GL column (GE-Healthcare) for D/ $NP_{TAIL}$ . Importin- $\alpha 7$  was purified by Nickel affinity chromatography and elution fraction were dialyzed with TEV (1/100) against 20 mM Tris-HCl pH 7.5 at 150 mM NaCl, 5 mM  $\beta$ -ME and 20 mM imidazol. Finally, a size-exclusion chromatography using Superdex<sup>TM</sup> 200 increase 10/300 GL column (GE-Healthcare) was performed in 20 mM Tris-HCl pH 7.5 at 150 mM NaCl, 5 mM  $\beta$ -ME. Peak fractions were concentrated using a 10 kDa Amicon concentrator. Protein concentrations were determined using the extinction coefficients at 280 nm,  $\epsilon = 44\,537\text{ M}^{-1}\cdot\text{cm}^{-1}$  for D/NP,  $\epsilon = 41\,558\text{ M}^{-1}\cdot\text{cm}^{-1}$  for D/NP-529 and D/NP-511,  $\epsilon = 2980\text{ M}^{-1}\cdot\text{cm}^{-1}$  for D/ $NP_{TAIL}$ , and  $\epsilon = 46\,785\text{ M}^{-1}\cdot\text{cm}^{-1}$  for importin- $\alpha 7$ .

**SEC-MALLS-RI analysis.** Multi-angle laser light scattering (MALLS) coupled with size exclusion chromatography (SEC) and refractometry (RI) is a method for measuring the absolute molecular mass of a particle in solution that is independent of its dimensions and shape<sup>64</sup>. SEC was performed with a column (Superdex<sup>TM</sup> 200

increase 10/300 GL or Superdex 75 10/300 GL) equilibrated with 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 5 mM  $\beta$ -ME. Analytical runs were performed at 20 °C with a flow rate of 0.5 mL.min<sup>-1</sup>. MALLS detection was performed with a DAWN-HELEOS II detector (Wyatt Technology) using a laser emitting at 690 nm and protein concentration was measured on-line with the use of differential refractive-index measurements, with an Optilab T-rEX detector (Wyatt Technology) and a refractive-index increment, dn/dc, of 0.185 mL.g<sup>-1</sup>. Weight-average molar masses (M<sub>w</sub>) were calculated with ASTRA (Wyatt Technology) as previously described<sup>38</sup>.

**Electron microscopy.** Samples (concentrations around 0.05 mg.mL<sup>-1</sup>) were applied between a carbon and a mica layer. The carbon was then floated on the top of a 2% (w/v) sodium silicotungstate, pH 7.0 solution. The carbon film was covered by a copper grid. Both were fished using a small piece of journal paper and air dried before insertion in the electron microscope<sup>65</sup>. Charge-coupled Device (CCD) frames were taken with a FEI T12 microscope operating at 120 kV and a nominal magnification of 45 000 times. The dilutions for EM were performed with the size-exclusion buffer right before preparing the grid.

**Crystallisation and structure determination.** D/NP was crystallised by vapor diffusion using the sitting drop method. The crystals were obtained in 2.4 M sodium malonate pH 4 with a protein concentration of 1.5 mg.mL<sup>-1</sup>. The crystals were directly flash-frozen without any cryoprotectant. Data were collected at the ESRF (beamline ID30A-3) and processed with the XDS package<sup>66</sup>. The structure was solved by molecular replacement using the Influenza A virus R416A monomeric nucleoprotein structure (PDB ID code 3ZDP) without the oligomerization loop (from residues 391 to 439) as a model. Model building and refinement were performed using CCP4i suite program for crystallography (PHASER, ARP/wARP, REFMAC5, COOT)<sup>67-71</sup>. The final refinement was done using BUSTER<sup>72</sup>. The coordinates have been deposited in the Protein Data Bank under PDB ID code 5N2U. The protein structure figures were drawn using PyMOL<sup>73</sup>.

**Thermal shift assays.** Thermal shift assays were performed following the established protocols of thermofluor experiments<sup>74</sup> using a fluorescent probe (SYPRO Orange). SYPRO Orange dye binds non-specifically to hydrophobic surfaces and water strongly quenches its fluorescence. When a protein unfolds with the increase of the temperature, the exposed hydrophobic surfaces bind the dye, resulting in an increase of the fluorescence by excluding water. Sample were diluted at 0.25 mg.mL<sup>-1</sup> in 50 mM Tris-HCl pH 7.5, 300 mM NaCl and 5 mM  $\beta$ -ME and 5X SYPRO Orange dye (Invitrogen) in a final volume of 40  $\mu$ L. The thermal stability was measured using a real time PCR machine (Mx3005P Q-PCR, Stratagene). The dye was excited at 488 nm and the emission light was recorded at 585 nm while the temperature was increased by increments of 1 °C per minute from 25 to 75 °C. The relative fluorescence emission was then plotted against its corresponding temperature to produce the thermal shift profile curve. The melting temperatures were estimated from the derivative curves.

**Circular dichroism.** Circular dichroism (CD) spectroscopy is an efficient tool for rapid determination of the secondary structure and folding properties of a protein by measuring the absorption difference between left and right circularly polarized light. A JASCO J-810 CD spectropolarimeter equipped with a temperature-controller (Peltier system) was used to record the far-UV CD spectrum of D/NP<sub>TAIL</sub> at 6.4  $\mu$ M in presence or in absence of trimethylamine N-oxide (TMAO). TMAO is a stabilizing chemical agent known to promote protein folding. It can be used with CD spectroscopy to assess putative folding of unfolded peptides and proteins. In addition to D/NP<sub>TAIL</sub> without TMAO, final concentrations of 0.5, 1, 2 and 3 M TMAO were used in phosphate buffer (10 mM, pH 7.2) to validate the unfolded nature of D/NP<sub>TAIL</sub>. Due to the high absorbance of TMAO around 200 nm, spectra were recorded up to the point before the HT rose above 700 V (190 nm without TMAO, 205 nm with 0.5 M TMAO, 207 nm with 1 M TMAO, 210 nm with 2 M TMAO and 212 nm with 3 M TMAO). Spectra were collected at 20 °C with fifteen runs using a cuvette with a path-length of 1 mm. After blank subtraction, the CD signal (in mdeg) was converted to mean molar residue ellipticity (in deg.cm<sup>2</sup>.dmol<sup>-1</sup>).

**Surface plasmon resonance (SPR) measurements and analysis.** The D/NP<sub>TAIL</sub> sequence was fused to the sequence encoding a biotinylation motif (NGSGGGLNDIFEA-QKIEWHE) and cloned in pETM11 (EMBL). *E. coli* BL21 RIL (DE3) were transformed and grown up to an OD of 0.6–0.8 in a medium supplemented with biotin (12.5  $\mu$ g.mL<sup>-1</sup>) in order to biotinylate the peptide *in vivo* during the expression (12 h, 18 °C, 0.3 mM IPTG). Purification was performed at room temperature in two steps, a Nickel affinity chromatography, followed by TEV cleavage and size exclusion chromatography (S75 10/300 GL, GE-Healthcare). A control biotinylated peptide (LEEMK-KGHLERECMEETCSYEEAREVFEDESEKTNEFWNK-biotin) was also used as negative control.

SPR experiments were carried out on a Biacore 3000 (GE Healthcare). Streptavidin was first immobilized on a CM5 sensor chip (GE Healthcare) surface through amine coupling at a flow rate of 5  $\mu$ L.min<sup>-1</sup>. The biotinylated NP<sub>TAIL</sub> (10  $\mu$ g.mL<sup>-1</sup>) or the control biotinylated peptide (10  $\mu$ g.mL<sup>-1</sup>) were diluted in surfactant-supplemented HBS-N running buffer (10 mM Hepes pH 7.4, 150 mM NaCl, 0.05% Tween 20) (GE Healthcare) and the attachment was carried out through biotin:streptavidin interaction, at a flow rate of 5  $\mu$ L.min<sup>-1</sup> up to 400–500 RU. A flow-cell containing immobilized streptavidin only was used as reference while the flow-cell with the streptavidin:biotinylated NP<sub>TAIL</sub> complex was used as active flow-cell. For kinetic measurements, the importin- $\alpha$ 7 analyte was serially diluted in surfactant-supplemented HBS-N running buffer to final concentrations of 10, 25, 50, 75 and 100 nM. Each concentration was injected in triplicate in both the reference and active flow-cells. Analyte injection and following buffer injection times were set at 180 s and 150 s respectively at a flow-rate of 15  $\mu$ L.min<sup>-1</sup>. As the analyte spontaneously and completely dissociated, no regeneration was required and complete dissociation was achieved in buffer for 300 s. All sensorgrams were reference-subtracted and injection points were aligned to be processed with the analysis. Kinetic data were analysed using the Biacore 3000 Evaluation software (GE Healthcare) under a Langmuir 1:1 binding model with mass-transfer. Chi<sup>2</sup> values for the sensorgram fits used to

determine the kinetic parameters were kept below 3, due to the obtained high signal, and T-value for rate parameters above at least 30. The residuals for the fitting were kept between 1 and  $-1$ .

**Nuclear import of WT and NP mutants in HEK 293T cells.** *Plasmids.* The D/NP, D/NP-511 and D/NP-529 genes were amplified by PCR, cloned into the multiple cloning site of pSC-A-amp/kan vector (Stratagene blunt kit, Agilent) and expressed into competent bacterial cells, according to the manufacturer's instructions. The NP genes in pSC-A-amp/kan vector were cloned into KpnI and BamHI restriction sites of an eukaryotic expression vector, pCDNA3.1 (+), driven by a CMV promoter. The constructs sequences were confirmed by Sanger sequencing (GATC Biotech platform, Germany).

*Rabbit antibodies.* To generate rabbit hyperimmune serum anti-D/NP, three 7 to 8 weeks old New Zealand white rabbits (Pôle Experimental Cunicole de Toulouse, Castanet-Tolosan) were immunized at day 0 and day 35 with 100–200  $\mu$ g D/NP in incomplete Freund adjuvant. Rabbits were anesthetized and bled two weeks post boost (50 mL blood collected intra-cardiac before humane euthanasia).

*Infection.* Human embryonic kidney cells (HEK 293T; ATCC) were infected with D/bovine /France/5920/2014 at a multiplicity of infection of 5. Briefly, cells were washed with PBS, the inoculum was added and the cells were incubated 1 hour at 37 °C + 5% CO<sub>2</sub>. After incubation, infection media (OPTI-MEM) was added and the cells were incubated until the fixation.

*Transfection.* HEK 293T cells were grown on glass coverslips in Dulbecco's modified Eagle's medium (DMEM) complemented with 10% fetal bovine serum (FBS) and penicillin-streptomycin (Dominique DUTSCHER SAS) and incubated at 37 °C + 5% CO<sub>2</sub> for 24 hours. Two hours before transfection, DMEM media was removed and replaced by OPTI-MEM media (Dominique DUTSCHER SAS). Cells were transfected with 2.5  $\mu$ g of the respective plasmids using Mirus-TransIT reagent (Mirus) following the manufacturers' instructions. Cells were then incubated for 24 h at 37 °C + 5% CO<sub>2</sub>.

*Immunofluorescence.* Cells were pre-fixed using 4% paraformaldehyde (Bio-Rad) for 10 minutes at room temperature and then, fixed and permeabilized using ice cold ethanol:acetone for 10 minutes at room temperature. After a blocking step using PBS complemented with 5% horse serum, cells were incubated with a rabbit hyperimmune D/NP serum at a 1:100 dilution for 2 hours at room temperature. After three washing steps with PBS-triton 0.05%, cells were incubated with anti-rabbit IgG labeled Rhodamine RX (Jackson Immuno; Research, 711.296.152) at 1:200 dilution for 1 hour in the dark and at room temperature. Then, cells were washed with PBS-triton 0.05% and coverslips were mounted on glass slides using Vectashield-DAPI mounting medium (Vector Laboratories) and analyzed with a Leica Zeiss 710 at the Cell Imagery Platform in Purpan (Toulouse).

**Ethics Statement.** Experimentations were conducted in accordance with European and French legislations on Laboratory Animal Care and Use (French Decree 2001-464 and European Directive CEE86/609) and the animal protocol was approved by the Ethics Committee "Sciences et santé animale" number 115.

## References

1. Ferguson, L. *et al.* Influenza D virus infection in Mississippi beef cattle. *Virology* **486**, 28–34, <https://doi.org/10.1016/j.virol.2015.08.030> (2015).
2. Hause, B. M. *et al.* Isolation of a novel swine influenza virus from Oklahoma in 2011 which is distantly related to human influenza C viruses. *PLoS Pathog* **9**, e1003176, <https://doi.org/10.1371/journal.ppat.1003176> (2013).
3. Luo, J. *et al.* Serological evidence for high prevalence of Influenza D Viruses in Cattle, Nebraska, United States, 2003–2004. *Virology* **501**, 88–91, <https://doi.org/10.1016/j.virol.2016.11.004> (2017).
4. Mitra, N., Cernicchiaro, N., Torres, S., Li, F. & Hause, B. M. Metagenomic characterization of the virome associated with bovine respiratory disease in feedlot cattle identified novel viruses and suggests an etiologic role for influenza D virus. *J Gen Virol* **97**, 1771–1784, <https://doi.org/10.1099/jgv.0.000492> (2016).
5. Horimoto, T. *et al.* Nationwide Distribution of Bovine Influenza D Virus Infection in Japan. *PLoS One* **11**, e0163828, <https://doi.org/10.1371/journal.pone.0163828> (2016).
6. Murakami, S. *et al.* Influenza D Virus Infection in Herd of Cattle, Japan. *Emerg Infect Dis* **22**, 1517–1519, <https://doi.org/10.3201/eid2208.160362> (2016).
7. Zhai, S. L. *et al.* Influenza D Virus in Animal Species in Guangdong Province, Southern China. *Emerg Infect Dis* **23**, 1392–1396, <https://doi.org/10.3201/eid2308.170059> (2017).
8. Chiapponi, C. *et al.* Detection of Influenza D Virus among Swine and Cattle, Italy. *Emerg Infect Dis* **22**, 352–354, <https://doi.org/10.3201/eid2202.151439> (2016).
9. Ducatez, M. F., Pelletier, C. & Meyer, G. Influenza D virus in cattle, France, 2011–2014. *Emerg Infect Dis* **21**, 368–371, <https://doi.org/10.3201/eid2102.141449> (2015).
10. Foni, E. *et al.* Influenza D in Italy: towards a better understanding of an emerging viral infection in swine. *Sci Rep* **7**, 11660, <https://doi.org/10.1038/s41598-017-12012-3> (2017).
11. Salem, E. *et al.* Serologic Evidence for Influenza C and D Virus among Ruminants and Camelids, Africa, 1991–2015. *Emerg Infect Dis* **23**, 1556–1559, <https://doi.org/10.3201/eid2309.170342> (2017).
12. Ng, A. K. *et al.* Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J* **22**, 3638–3647, <https://doi.org/10.1096/fj.08-112110> (2008).
13. Ye, Q., Krug, R. M. & Tao, Y. J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444**, 1078–1082, <https://doi.org/10.1038/nature05379> (2006).
14. Chenavas, S. *et al.* Monomeric nucleoprotein of influenza A virus. *PLoS Pathog* **9**, e1003275, <https://doi.org/10.1371/journal.ppat.1003275> (2013).
15. Ng, A. K. *et al.* Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein. *J Virol* **86**, 6758–6767, <https://doi.org/10.1128/JVI.00073-12> (2012).

16. Zheng, W., Olson, J., Vakharia, V. & Tao, Y. J. The crystal structure and RNA-binding of an orthomyxovirus nucleoprotein. *PLoS Pathog* **9**, e1003624, <https://doi.org/10.1371/journal.ppat.1003624> (2013).
17. Cros, J. F., Garcia-Sastre, A. & Palese, P. An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. *Traffic* **6**, 205–213, <https://doi.org/10.1111/j.1600-0854.2005.00263.x> (2005).
18. Labaronne, A. *et al.* Structural analysis of the complex between influenza B nucleoprotein and human importin- $\alpha$ . *Sci Rep* **7**, 17164, <https://doi.org/10.1038/s41598-017-17458-z> (2017).
19. Liu, M. *et al.* The Functional Study of the N-Terminal Region of Influenza B Virus Nucleoprotein. *PLoS One* **10**, e0137802, <https://doi.org/10.1371/journal.pone.0137802> (2015).
20. Ozawa, M. *et al.* Contributions of two nuclear localization signals of influenza A virus nucleoprotein to viral replication. *J Virol* **81**, 30–41, <https://doi.org/10.1128/JVI.01434-06> (2007).
21. Sherry, L., Smith, M., Davidson, S. & Jackson, D. The N terminus of the influenza B virus nucleoprotein is essential for virus viability, nuclear localization, and optimal transcription and replication of the viral genome. *J Virol* **88**, 12326–12338, <https://doi.org/10.1128/JVI.01542-14> (2014).
22. Wang, P., Palese, P. & O'Neill, R. E. The NPI-1/NPI-3 (karyopherin  $\alpha$ ) binding site on the influenza A virus nucleoprotein NP is a nonconventional nuclear localization signal. *J Virol* **71**, 1850–1856 (1997).
23. Wanitchang, A., Narkpuk, J. & Jongkaewwattana, A. Nuclear import of influenza B virus nucleoprotein: involvement of an N-terminal nuclear localization signal and a cleavage-protection motif. *Virology* **443**, 59–68, <https://doi.org/10.1016/j.virol.2013.04.025> (2013).
24. Wu, W. W., Sun, Y. H. & Pante, N. Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein. *Virol J* **4**, 49, <https://doi.org/10.1186/1743-422X-4-49> (2007).
25. Nakada, R., Hirano, H. & Matsuura, Y. Structure of importin- $\alpha$  bound to a non-classical nuclear localization signal of the influenza A virus nucleoprotein. *Sci Rep* **5**, 15055, <https://doi.org/10.1038/srep15055> (2015).
26. Wu, W. *et al.* Synergy of two low-affinity NLSs determines the high avidity of influenza A virus nucleoprotein NP for human importin  $\alpha$  isoforms. *Sci Rep* **7**, 11381, <https://doi.org/10.1038/s41598-017-11018-1> (2017).
27. Labaronne, A. *et al.* Binding of RNA by the Nucleoproteins of Influenza Viruses A and B. *Viruses* **8**, <https://doi.org/10.3390/v8090247> (2016).
28. Tarus, B. *et al.* Oligomerization paths of the nucleoprotein of influenza A virus. *Biochimie* **94**, 776–785, <https://doi.org/10.1016/j.biochi.2011.11.009> (2012).
29. Leung, D. W. *et al.* An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. *Cell Rep* **11**, 376–389, <https://doi.org/10.1016/j.celrep.2015.03.034> (2015).
30. Chan, W. H. *et al.* Functional analysis of the influenza virus H5N1 nucleoprotein tail loop reveals amino acids that are crucial for oligomerization and ribonucleoprotein activities. *J Virol* **84**, 7337–7345, <https://doi.org/10.1128/JVI.02474-09> (2010).
31. Elton, D., Medcalf, E., Bishop, K. & Digard, P. Oligomerization of the influenza virus nucleoprotein: identification of positive and negative sequence elements. *Virology* **260**, 190–200, <https://doi.org/10.1006/viro.1999.9818> (1999).
32. Shen, Y. F. *et al.* E339...R416 salt bridge of nucleoprotein as a feasible target for influenza virus inhibitors. *Proc Natl Acad Sci USA* **108**, 16515–16520, <https://doi.org/10.1073/pnas.1113107108> (2011).
33. Turrell, L., Lyall, J. W., Tiley, L. S., Fodor, E. & Vreede, F. T. The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. *Nat Commun* **4**, 1591, <https://doi.org/10.1038/ncomms2589> (2013).
34. Mondal, A., Potts, G. K., Dawson, A. R., Coon, J. J. & Mehle, A. Phosphorylation at the homotypic interface regulates nucleoprotein oligomerization and assembly of the influenza virus replication machinery. *PLoS Pathog* **11**, e1004826, <https://doi.org/10.1371/journal.ppat.1004826> (2015).
35. Mondal, A. *et al.* Influenza virus recruits host protein kinase C to control assembly and activity of its replication machinery. *Elife* **6**, <https://doi.org/10.7554/eLife.26910> (2017).
36. Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. W. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J* **13**, 3158–3165 (1994).
37. Elton, D., Medcalf, L., Bishop, K., Harrison, D. & Digard, P. Identification of amino acid residues of influenza virus nucleoprotein essential for RNA binding. *J Virol* **73**, 7357–7367 (1999).
38. Gerard, F. C. *et al.* Modular organization of rabies virus phosphoprotein. *J Mol Biol* **388**, 978–996, <https://doi.org/10.1016/j.jmb.2009.03.061> (2009).
39. Uversky, V. N. Size-exclusion chromatography in structural analysis of intrinsically disordered proteins. *Methods Mol Biol* **896**, 179–194, [https://doi.org/10.1007/978-1-4614-3704-8\\_11](https://doi.org/10.1007/978-1-4614-3704-8_11) (2012).
40. Baskakov, I. & Bolen, D. W. Forcing thermodynamically unfolded proteins to fold. *J Biol Chem* **273**, 4831–4834 (1998).
41. Baskakov, I., Wang, A. & Bolen, D. W. Trimethylamine-N-oxide counteracts urea effects on rabbit muscle lactate dehydrogenase function: a test of the counteraction hypothesis. *Biophys J* **74**, 2666–2673, [https://doi.org/10.1016/S0006-3495\(98\)77972-X](https://doi.org/10.1016/S0006-3495(98)77972-X) (1998).
42. Baskakov, I. V. *et al.* Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor. *J Biol Chem* **274**, 10693–10696 (1999).
43. Stevens, M. P. & Barclay, W. S. The N-terminal extension of the influenza B virus nucleoprotein is not required for nuclear accumulation or the expression and replication of a model RNA. *J Virol* **72**, 5307–5312 (1998).
44. White, S. K., Ma, W., McDaniel, C. J., Gray, G. C. & Lednicky, J. A. Serologic evidence of exposure to influenza D virus among persons with occupational contact with cattle. *J Clin Virol* **81**, 31–33, <https://doi.org/10.1016/j.jcv.2016.05.017> (2016).
45. Song, H. *et al.* An Open Receptor-Binding Cavity of Hemagglutinin-Esterase-Fusion Glycoprotein from Newly-Identified Influenza D Virus: Basis for Its Broad Cell Tropism. *PLoS Pathog* **12**, e1005411, <https://doi.org/10.1371/journal.ppat.1005411> (2016).
46. Liao, T. L., Wu, C. Y., Su, W. C., Jeng, K. S. & Lai, M. M. Ubiquitination and deubiquitination of NP protein regulates influenza A virus RNA replication. *EMBO J* **29**, 3879–3890, <https://doi.org/10.1038/emboj.2010.250> (2010).
47. Turrell, L., Hutchinson, E. C., Vreede, F. T. & Fodor, E. Regulation of influenza A virus nucleoprotein oligomerization by phosphorylation. *J Virol* **89**, 1452–1455, <https://doi.org/10.1128/JVI.02332-14> (2015).
48. Shi, M. *et al.* The evolutionary history of vertebrate RNA viruses. *Nature* **556**, 197–202, <https://doi.org/10.1038/s41586-018-0012-7> (2018).
49. Hause, B. M. *et al.* Characterization of a novel influenza virus in cattle and Swine: proposal for a new genus in the Orthomyxoviridae family. *MBio* **5**, e00031–00014, <https://doi.org/10.1128/mBio.00031-14> (2014).
50. Weber, F., Kochs, G., Gruber, S. & Haller, O. A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins. *Virology* **250**, 9–18, <https://doi.org/10.1006/viro.1998.9329> (1998).
51. Aspehaug, V. *et al.* Infectious salmon anemia virus (ISAV) genomic segment 3 encodes the viral nucleoprotein (NP), an RNA-binding protein with two monopartite nuclear localization signals (NLS). *Virus Res* **106**, 51–60, <https://doi.org/10.1016/j.virusres.2004.06.001> (2004).
52. Longhi, S. *et al.* The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J Biol Chem* **278**, 18638–18648, <https://doi.org/10.1074/jbc.M300518200> (2003).
53. Bloyet, L. M. *et al.* Modulation of Re-initiation of Measles Virus Transcription at Intergenic Regions by PXD to NTAIL Binding Strength. *PLoS Pathog* **12**, e1006058, <https://doi.org/10.1371/journal.ppat.1006058> (2016).

54. Communie, G. *et al.* Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. *PLoS Pathog* **9**, e1003631, <https://doi.org/10.1371/journal.ppat.1003631> (2013).
55. Miller, M. R. *et al.* Intrinsic disorder in measles virus nucleocapsids. *Proc Natl Acad Sci USA* **108**, 9839–9844, <https://doi.org/10.1073/pnas.1103270108> (2011).
56. Schoehn, G. *et al.* The 12 Å structure of trypsin-treated measles virus N-RNA. *J Mol Biol* **339**, 301–312, <https://doi.org/10.1016/j.jmb.2004.03.073> (2004).
57. Milles, S. *et al.* Self-Assembly of Measles Virus Nucleocapsid-like Particles: Kinetics and RNA Sequence Dependence. *Angew Chem Int Ed Engl* **55**, 9356–9360, <https://doi.org/10.1002/anie.201602619> (2016).
58. Arranz, R. *et al.* The structure of native influenza virion ribonucleoproteins. *Science* **338**, 1634–1637, <https://doi.org/10.1126/science.1228172> (2012).
59. Klumpp, K., Ruigrok, R. W. & Baudin, F. Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *EMBO J* **16**, 1248–1257, <https://doi.org/10.1093/emboj/16.6.1248> (1997).
60. Moeller, A., Kirchoerfer, R. N., Potter, C. S., Carragher, B. & Wilson, I. A. Organization of the influenza virus replication machinery. *Science* **338**, 1631–1634, <https://doi.org/10.1126/science.1227270> (2012).
61. Gallagher, J. R., Torian, U., McCraw, D. M. & Harris, A. K. Structural studies of influenza virus RNPs by electron microscopy indicate molecular contortions within NP supra-structures. *J Struct Biol*, <https://doi.org/10.1016/j.jsb.2016.12.007> (2016).
62. Gutschke, I. *et al.* Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* **348**, 704–707, <https://doi.org/10.1126/science.aaa5137> (2015).
63. Boivin, S. & Hart, D. J. Interaction of the influenza A virus polymerase PB2 C-terminal region with importin alpha isoforms provides insights into host adaptation and polymerase assembly. *J Biol Chem* **286**, 10439–10448, <https://doi.org/10.1074/jbc.M110.182964> (2011).
64. Wyatt, P. J. Submicrometer Particle Sizing by Multiangle Light Scattering following Fractionation. *J Colloid Interface Sci* **197**, 9–20 (1998).
65. Gaudin, Y., Ruigrok, R. W., Tuffereau, C., Knossow, M. & Flamand, A. Rabies virus glycoprotein is a trimer. *Virology* **187**, 627–632 (1992).
66. Kabsch, W. X. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132, <https://doi.org/10.1107/S0907444909047337> (2010).
67. Cohen, S. X. *et al.* ARP/wARP and molecular replacement: the next generation. *Acta Crystallogr D Biol Crystallogr* **64**, 49–60, <https://doi.org/10.1107/S0907444907047580> (2008).
68. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501, <https://doi.org/10.1107/S0907444910007493> (2010).
69. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674, <https://doi.org/10.1107/S0021889807021206> (2007).
70. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255, <https://doi.org/10.1107/S0907444996012255> (1997).
71. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242, <https://doi.org/10.1107/S0907444910045749> (2011).
72. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr D Biol Crystallogr* **68**, 368–380, <https://doi.org/10.1107/S0907444911056058> (2012).
73. Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* (2015).
74. Pantoliano, M. W. *et al.* High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **6**, 429–440, <https://doi.org/10.1089/108705701753364922> (2001).
75. Rocchia, W. *et al.* Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J Comput Chem* **23**, 128–137, <https://doi.org/10.1002/jcc.1161> (2002).
76. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289–298, <https://doi.org/10.1016/j.ab.2006.07.027> (2006).
77. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797, <https://doi.org/10.1093/nar/gkh340> (2004).
78. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256–2268, <https://doi.org/10.1107/S0907444904026460> (2004).

## Acknowledgements

We are highly grateful to Jean-Marie Bourhis, Daphna Fenel and Emmanuelle Neumann for their help. We thank Darren J Hart for the plasmid to express the human importin- $\alpha$ 7. We thank Martin Blackledge, Anny Slama-Schwok and Bernard Delmas for discussion. We thank Stéphane Bertagnoli for his help with the rabbit immunizations. We thank the staff of the ESRF-EMBL Joint Structural Biology Group for access to ESRF beamlines. AD, AL and MM were funded through the Labex GRAL (ANR-10-LABX-49-01), LT by the ANR RNAP-IAV (ANR-14-CE09-0017) and AD and JO are supported by a PhD scholarship of the French Ministry of Higher Education and Research. This work was supported by the Fondation de la Recherche Médicale (FRM; Equipe DEQ. 20170336754) and used the electron microscopy and the biophysical platforms of the Grenoble Instruct-ERIC Center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL, within the Grenoble Partnership for Structural Biology (PSB), as well as the Imagery platform of Centre de Physiopathologie de Toulouse Purpan. The IBS Electron Microscope facility is supported by the Rhône-Alpes Region, the Fonds Feder, the FRM and GIS-IBISA.

## Author Contributions

T.C. and M.D. conceived the experiments. A.D., J.O., A.L., L.T., M.M., F.C.A.G., C.M., G.S., M.D. and T.C. performed the experiments. A.D., J.O., A.L., F.C.A.G., G.S., R.W.H.R., M.D. and T.C. analysed the data. R.W.H.R., M.D. and T.C. wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-37306-y>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019