



**HAL**  
open science

## Structural Description of the Nipah Virus Phosphoprotein and Its Interaction with STAT1

Malene Ringkjøbing Jensen, Filip Yabukarski, Guillaume Communie, Eric Condamine, Caroline Mas, Valentina Volchkova, Nicolas Tarbouriech, Jean-Marie Bourhis, Viktor Volchkov, Martin Blackledge, et al.

► **To cite this version:**

Malene Ringkjøbing Jensen, Filip Yabukarski, Guillaume Communie, Eric Condamine, Caroline Mas, et al.. Structural Description of the Nipah Virus Phosphoprotein and Its Interaction with STAT1. *Biophysical Journal*, 2020, 118 (10), pp.2470-2488. 10.1016/j.bpj.2020.04.010 . hal-02867772

**HAL Id: hal-02867772**

**<https://hal.science/hal-02867772v1>**

Submitted on 20 May 2022

**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 - NonCommercial 4.0 International License

1 STRUCTURAL DESCRIPTION OF THE NIPAH VIRUS PHOSPHOPROTEIN AND ITS  
2 INTERACTION WITH STAT1

3

4 Malene Ringkjøbing Jensen<sup>1#</sup>, Filip Yabukarski<sup>1#</sup>, Guillaume Communie<sup>1</sup>, Eric Condamine<sup>1</sup>,  
5 Caroline Mas<sup>2</sup>, Valentina Volchkova<sup>3</sup>, Nicolas Tarbouriech<sup>1</sup>, Jean-Marie Bourhis<sup>1</sup>, Viktor  
6 Volchkov<sup>3</sup>, Martin Blackledge<sup>1</sup> and Marc Jamin<sup>1\*</sup>

7

8

9

10 <sup>1</sup>Institut de Biologie Structurale (IBS), Univ. Grenoble Alpes, CEA, CNRS, 71 Avenue des  
11 Martyrs, 38000 Grenoble, France.

12 <sup>2</sup>Integrated Structural Biology Grenoble (ISBG) CNRS, CEA, Univ. Grenoble Alpes, EMBL,  
13 71 avenue des Martyrs, 38000, Grenoble, France.

14 <sup>3</sup>Molecular basis of viral pathogenicity, Centre International de Recherche en  
15 Infectiologie (CIRI), INSERMU1111-CNRS UMR5308, Université Claude Bernard Lyon 1,  
16 ENS de Lyon, 69365 Lyon, France.

17

18 #Co-first authors

19

20 Running title: Structure of Nipah virus P protein

21 Keywords: Nipah virus, phosphoprotein, Mononegavirales, intrinsically disordered protein,  
22 small-angle X-ray scattering, nuclear magnetic resonance, X-ray crystallography

23

24

25

26

27

28 \* Corresponding author: Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale  
29 (IBS), 38000 GRENOBLE.

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

31

32 **STATEMENT OF SIGNIFICANCE**

33 The intrinsically disordered phosphoprotein of Nipah virus orchestrates viral genome  
34 transcription, replication and encapsidation and shares a long N-terminal disordered region  
35 with the non-structural V and W proteins involved in blocking the interferon responses. By  
36 combining multiple biophysical methods, including X-ray crystallography, NMR  
37 spectroscopy and small-angle X-ray scattering, we characterized the structure and flexibility  
38 of this protein, identified several putative short linear motifs of interaction with viral and  
39 cellular partners and investigated the interaction with one of them. The protein appears as a  
40 nanometer-size octopus comprising a central tetrameric core and displaying binding sites on  
41 its flexible arms to recruit multiple viral and cellular partners, thus providing new insights  
42 into how this protein fulfills its multiple functions.

43

44 **SUMMARY**

45 The structural characterization of modular proteins containing long intrinsically disordered  
46 regions (IDRs) intercalated with folded domains is complicated by their conformational  
47 diversity and flexibility and requires the integration of multiple experimental approaches.  
48 Nipah virus (NiV) phosphoprotein, an essential component of the viral RNA  
49 transcription/replication machine and a component of the viral arsenal that hijacks cellular  
50 components and counteracts host immune responses, is a prototypical model for such modular  
51 proteins. Curiously, the phosphoprotein of NiV is significantly longer than the corresponding  
52 protein of other paramyxoviruses. Here, we combine multiple biophysical methods, including  
53 X-ray crystallography, NMR spectroscopy and small-angle X-ray scattering (SAXS), to  
54 characterize the structure of this protein and provide an atomistic representation of the full-  
55 length protein in the form of a conformational ensemble. We show that full-length NiV  
56 phosphoprotein is tetrameric, and we solve the crystal structure of its tetramerization domain.  
57 Using NMR spectroscopy and SAXS, we show that the long N-terminal intrinsically  
58 disordered region and the linker connecting the tetramerization domain to the C-terminal X  
59 domain exchange between multiple conformations while containing short regions of residual  
60 secondary structure. Some of these transient helices are known to interact with partners, while  
61 others represent putative binding sites for yet unidentified proteins. Finally, using NMR  
62 spectroscopy and isothermal titration calorimetry, we map a region of the phosphoprotein,  
63 comprised between residues 110 and 140 and common to the V and W proteins, that binds  
64 with weak affinity to STAT1, and confirm the involvement of key amino acids of the viral  
65 protein in this interaction. This provides new insights into how the phosphoprotein and the  
66 non-structural V and W proteins of NiV perform their multiple functions.

67

68

69

70

71

## 72 INTRODUCTION

73 Nipah virus (NiV) is one of the most pathogenic and deadly viruses infecting humans (1). It is  
74 an emergent zoonotic virus that naturally infects fruit bats in South-East Asia, but that can  
75 also infect other hosts including pigs, dogs, cats, horses and humans, where it becomes highly  
76 pathogenic. Since its discovery in 1998 in Malaysia, NiV has caused regular outbreaks of  
77 severe human encephalitis and respiratory illness with a fatality rate ranging from 40% to  
78 90% mainly in Bangladesh and more recently in Philippines and India (1, 2). With its ability  
79 to transmit from man to man and the lack of vaccine or efficient treatment, NiV has the  
80 potential to cause a pandemic and to be used in biological warfare (3). It is consequently  
81 classified as a biosafety level-4 (BSL-4) pathogen.

82  
83 NiV is an enveloped virus with a nonsegmented negative-sense RNA genome, which belongs  
84 to the order *Mononegavirales* and the family *Paramyxoviridae*, although, several features in  
85 its genome organization and in its protein sequences led to its classification into a new genus,  
86 Henipavirus, with the highly pathogenic Hendra virus and the non-pathogenic Cedar virus (4).  
87 The multiplication of NiV in a suitable host cell, like that of all members of the  
88 *Mononegavirales*, requires both (i) transcription of the single-stranded RNA genome into  
89 monocistronic, capped and polyadenylated messenger RNAs (positive-sense RNA) that can  
90 be used by the cellular ribosomes to generate new viral proteins and (ii) replication of the  
91 negative-sense RNA genome into positive-sense antigenome copies, which are then used to  
92 generate a progeny of negative-sense genomes. The enzymatic machinery that carries out  
93 these activities is conserved among all members of the *Mononegavirales* and is unique in  
94 biology and thus represents an attractive target for the development of much-needed antiviral  
95 compounds. It consists of a megadalton-sized multi-molecular complex located in the core of  
96 the viral particle and timely released in the cytoplasm of the host cell (5, 6). This  
97 transcription/replication complex is composed of the genomic RNA and of three viral  
98 proteins: the nucleoprotein (N), the phosphoprotein (P) and the large RNA-dependent RNA  
99 polymerase (L). With the help of the phosphoprotein acting as a chaperone, the nucleoprotein  
100 assembles into linear homopolymers that coats both negative-sense genomic and positive-  
101 sense antigenomic RNAs, forming long, helical nucleocapsids (7). These nucleocapsids  
102 constitute the biologically active templates used by the polymerase complex, which consists  
103 of the L polymerase and the phosphoprotein acting as a non-catalytic cofactor. Structural and  
104 dynamical studies of the unassembled nucleoprotein (N<sup>0</sup>) in complex with the phosphoprotein  
105 (8–11) and of nucleocapsids (7, 12) have started to unravel the mechanisms of assembly and

106 remodeling of the nucleocapsids during the transcription and replication processes (13). The  
107 multidomain L polymerase catalyzes RNA synthesis, but also the capping, methylation and  
108 polyadenylation of the viral messenger RNAs (14). The underlying mechanisms of the  
109 functioning of this polymerase remain largely to be deciphered (15), but its structure in  
110 complex with its cofactor the phosphoprotein is being revealed by recent cryo-electron  
111 microscopic reconstructions of the complex from various *Mononegavirales* (16–19),  
112 including a first low-resolution reconstruction of the NiV polymerase (14). The third essential  
113 component, the phosphoprotein, is thus a multitasking protein that orchestrates the interplay  
114 between the different components of the transcription/replication complex. The  
115 phosphoprotein takes its name from early observations showing that it is the most heavily  
116 phosphorylated viral protein in cells infected by different members of the *Mononegavirales*  
117 (20). In NiV phosphoprotein, residues Ser224 and Thr239 are the major phosphorylation sites  
118 (21) and a binding site for Polo-like kinase 1 is located near residue 130 (22), but as for other  
119 *Mononegavirales*, the role of these phosphorylations in the viral replication cycle remains  
120 elusive (23).

121  
122 All phosphoproteins of the *Mononegavirales* share the properties of binding the L  
123 polymerase, the unassembled nucleoprotein and the nucleocapsid through independent  
124 structural modules (13). In this way, the phosphoprotein promotes the initiation of RNA  
125 synthesis, confers processivity to the polymerase, maintains the unassembled nucleoprotein in  
126 an RNA-free and soluble form and delivers the latter to the site of RNA synthesis in order to  
127 encapsulate the nascent RNAs (9, 10, 12, 24–28). Despite common functions, similarities in  
128 the phosphoprotein sequence are hardly found beyond the level of the family and in some  
129 cases of the genus (29, 30). However, all phosphoproteins of the *Mononegavirales* are  
130 multimeric and share a common architecture with a long N-terminal intrinsically disordered  
131 region (P<sub>NTR</sub>) and a C-terminal region (P<sub>CTR</sub>) that includes the multimerization domain (P<sub>MD</sub>),  
132 a linker region (P<sub>CT-LINK</sub>) and most often a folded domain at the C-terminal end (P<sub>XD</sub>) (31, 32)  
133 (**Figure 1B**). They also share a rather common localization of functional modules (14, 32–  
134 34), and thus, NiV phosphoprotein exhibits a structural architecture and an organization of the  
135 functional modules involved in binding the nucleoprotein and the polymerase that are typical  
136 of the family *Paramyxoviridae*, although the protein of henipaviruses is appreciably longer  
137 than that of the other genera of *Paramyxoviridae* (25, 33, 35) (**Figure 1B**): (i) The N-terminal  
138 disordered region of more than 400 aa contains a short region (aa 1-38) at the N-terminal end  
139 that acts as a molecular recognition element or MoRE (*i.e.* a region that is partially disordered

140 in isolation but that adopts a stable conformation upon binding to a partner (36)) for recruiting  
141 and chaperoning the RNA-free unassembled nucleoprotein ( $N^0$ ) required for the assembly of  
142 new nucleocapsids during genome replication (9). (ii) The tetramerization domain of NiV (aa  
143 475-578) (37) is involved in the interaction with and in the function of the L polymerase (27,  
144 34, 38). (iii) A folded domain at the C-terminal end of the protein, named the X domain ( $P_{XD}$ )  
145 (aa 655-709) that binds to a MoRE in the C-terminal disordered tail of the nucleoprotein  
146 ( $N_{TAIL}$ ) (39). (iv) A third region of weak interaction with the nucleoprotein that has recently  
147 been identified in measles virus phosphoprotein, appears to be conserved in NiV  
148 phosphoprotein (aa 330-350) (10). Despite all this knowledge, we are still lacking a structural  
149 characterization of the full-length NiV phosphoprotein.

150

151 In addition, besides its roles in the transcription and replication of the viral genome, the  
152 phosphoprotein of many *Mononegavirales* also plays roles in hijacking cellular machineries  
153 and in counteracting the host immune responses (40). As in other *Paramyxoviridae*, the NiV P  
154 gene codes for multiple proteins through the use of alternative open reading frames, and these  
155 additional proteins act as virulence factors, in particular by directly targeting the Signal  
156 Transducer and Activator of Transcription (STAT) proteins and blocking the interferon  
157 signaling pathway (40). The NiV P gene encodes three proteins, the phosphoprotein and the  
158 non-structural V and W proteins. The gene contains an editing site in the part coding for the  
159 N-terminal region of the phosphoprotein, where the polymerase stutters during transcription  
160 and incorporates one, two or more pseudo-templated Gs resulting in shifts of the reading  
161 frame (41). The different messenger RNAs generated through this process lead to the  
162 expression of either the phosphoprotein or of one of the non-structural V (+1 ORF) or W (+2  
163 ORF) proteins, which share their 405 N-terminal residues with the phosphoprotein but have  
164 specific C-terminal moieties of 51 and 44 amino acids, respectively (**Figure 1B**) (41). In  
165 addition, a second gene coding for the C protein is overprinted on the initial region of the P  
166 gene in the same reading frame as the V protein (+1 ORF) (**Figure 1B**) and is expressed from  
167 the three different mRNAs by a leaky ribosome scanning mechanism (42). However, in  
168 contrast to some other *Paramyxoviridae*, NiV phosphoprotein, V and W proteins block  
169 interferon signaling by preventing the phosphorylation of STAT1 and by sequestering STAT1  
170 and STAT2 into non-functional complexes (43). The phosphoprotein and the V protein retain  
171 STAT proteins in the cytoplasm, whereas the W protein sequesters them in the nucleus. The  
172 N-terminal part comprised between residues 100 and 300, shared by the three proteins, is  
173 critical for interferon evasion. The binding site for STAT1 was first localized between

174 residues 100 and 160 (44, 45) and then further mapped down to residues 114 to 140 (6).  
175 Mutation of different residues in this region abrogate interferon inhibition (6, 46). The  
176 interaction with STAT2 requires the entire 100-300 region, in particular a motif between  
177 residues 230 and 237 and the prior binding of STAT1 (44). Furthermore, the region between  
178 residues 174 and 192 contains a CRM1-dependent nuclear export signal (NES), which allows  
179 the V protein to shuttle between the nucleus and cytoplasm and remove STAT1 out of the  
180 nucleus (43, 47–49), whereas the W protein maintains phosphorylated STAT1 in the nucleus  
181 through the presence of a nuclear localization signal (NLS) in its unique C-terminal moiety  
182 (6, 45). Finally, the region around residue 130, overlapping with the STAT1 binding site  
183 interacts with polo-like kinase 1, and this interaction results in the phosphorylation of the V  
184 protein (22).

185

186 Here, we investigate the structure of full-length NiV phosphoprotein ( $P_{FL}$ ). We determine its  
187 oligomeric state by biophysical methods and demonstrate that the multimerization domain is  
188 the only region of the phosphoprotein involved in its tetramerization. We solve the crystal  
189 structure of the tetramerization domain and use NMR spectroscopy to investigate the  
190 structural properties of the long N-terminal region, of the C-terminal linker predicted to be  
191 intrinsically disordered and of the C-terminal X domain. We demonstrate the independence of  
192 the different N-terminal chains in the tetrameric protein and identify several regions with  
193 transiently populated secondary structure elements, which could be molecular recognition  
194 elements (MoRE) for different viral or cellular partners. We generate an ensemble of  
195 conformers of full-length phosphoprotein that incorporates the crystal structure of the  
196 tetramerization domain and a homology model of the X domain and that simultaneously  
197 accounts for NMR and small-angle X-ray scattering data from three different constructs of the  
198 protein. Finally, we investigate the interaction between STAT1 and a fragment of 100 amino  
199 acids derived from the N-terminal region of the phosphoprotein. NMR spectroscopy is used to  
200 validate the binding site in the phosphoprotein, while isothermal titration calorimetry and cell-  
201 based assays are combined with site-directed mutagenesis to map important residues for the  
202 interaction.

203

## 204 **MATERIALS AND METHODS**

### 205 **Disorder meta-prediction**

206 The location of intrinsically disordered regions (IDRs) in NiV phosphoprotein was predicted  
207 by submitting its amino acid sequence to 16 different algorithms accessible through web



208 servers and by calculating a consensus prediction as described previously (32, 50). The  
209 calculated disorder score (D-score) varies between 0 and 1, and generally, regions with a D-  
210 score lower than 0.5 are intrinsically disordered, whereas regions with a score higher than 0.5  
211 are either folded domains (50–52) or potential molecular recognition elements for partner  
212 proteins (53).

213

### 214 **Protein expression and purification**

215 On the basis of our predictions, different fragments of the phosphoprotein (Uniprot: Q9IK91)  
216 were cloned between the NcoI and BamHI restriction sites of a pET28 vector, such that each  
217 construct included a C-terminal His<sub>6</sub>-tag with a Leu-Glu linker. For the constructs lacking the  
218 N-terminal end of the protein, with the exception of the P<sub>655-709</sub> construct which starts at  
219 Met655, a Met residue or a Met-Ala dipeptide (construct P<sub>173-240</sub>) was introduced at the N-  
220 terminal end of the fragment. Three single mutants of P<sub>92-190</sub>: Y116A, G120E, G125E and a  
221 triple mutant: H117A/Q118A/H1119A, were generated by reverse PCR using the  
222 QuickChange mutagenesis protocol (Agilent). All constructs were verified by DNA  
223 sequencing.

224

225 The proteins were expressed in *E. coli* BL21-RIL. Cells were grown in Luria Bertoni medium  
226 at 37°C until the optical density (O.D.) at 600 nm reached a value of 0.6, and protein  
227 expression was then induced by adding 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG)  
228 and cells were allowed to grow for 3 hours. For producing the selenomethionine-substituted  
229 form of the P<sub>471-580</sub> construct, the bacterial cells were grown in minimal medium and SeMet  
230 was added before induction of the heterologous protein expression. In each case, the cells  
231 were harvested by centrifugation, and the pellet was resuspended in 20 mM Tris-HCl buffer  
232 at pH 7.5 containing 150 mM NaCl (buffer A) and EDTA-free protease inhibitor cocktail  
233 (Roche). The cells were broken by sonication, and the soluble fraction was loaded onto a Ni<sup>2+</sup>  
234 column equilibrated in buffer A. After washing the column with buffer A containing 20 mM  
235 imidazole, the protein was eluted with buffer A containing 300 mM imidazole. The protein  
236 was further purified by size exclusion chromatography on a Superdex S75 (GE Healthcare)  
237 equilibrated in buffer A or in 20 mM Bis-Tris buffer at pH 6.0 containing 150 mM NaCl, 50  
238 mM arginine, 50 mM glutamate and 0.5 mM TCEP (buffer B) for NMR experiments or in 20  
239 mM sodium phosphate buffer at pH 7.5 containing 150 mM NaCl (buffer C) for ITC  
240 experiments. Each protein preparation was checked by SDS-PAGE and SEC-MALLS (see  
241 below).

242

243 A fragment of the gene coding for residues 132 to 683 of human STAT1 was cloned in  
244 pET28, was expressed in *E. coli* and was purified as previously described (54). Briefly, after  
245 ammonium sulfate precipitation, the protein was solubilized in a 20 mM Tris-HCl buffer at  
246 pH 8.8 containing 150 mM NaCl, and the solution was loaded onto a SP sepharose column  
247 (GE healthcare) equilibrated in the same buffer. The protein was not retained on the column  
248 and eluted in the flow-through. After concentration, the protein was further purified by  
249 loading the solution containing STAT1<sub>132-683</sub> onto a Superdex S200 (GE Healthcare) size  
250 exclusion column equilibrated in buffer C.

251

### 252 **SEC-MALLS experiments**

253 Size exclusion chromatography (SEC) combined with on-line detection by multi-angle laser  
254 light scattering (MALLS) and refractometry (RI) is a method for measuring the absolute  
255 molecular mass (MM) of a particle in solution that is independent of its dimensions and shape  
256 (55). SEC was performed with TSK G4000 (Tosohbioscience) or S200 Superdex column (GE  
257 Healthcare) equilibrated with 20 mM Tris-HCl buffer containing 150 mM NaCl, and the  
258 columns were calibrated with globular standard proteins of known hydrodynamic radius ( $R_h$ )  
259 (56). Chromatographic separations were performed at 20°C with a flow rate of 0.5 or 0.8  
260 mL.min<sup>-1</sup>, with on-line MALLS detection with a DAWN-HELEOS II detector (Wyatt  
261 Technology Corp.) using a laser emitting at 690 nm, and protein concentration was measured  
262 on-line by the use of differential refractive index measurements using an Optilab T-rEX  
263 detector (Wyatt Technology Corp.) and a refractive index increment,  $dn/dc$ , of 0.185 mL.g<sup>-1</sup>.  
264 The weight-averaged molecular mass was calculated using the ASTRA software (Wyatt  
265 Technology Corp.).

266

### 267 **Small angle X-ray scattering experiments**

268 Small angle X-ray scattering (SAXS) data were collected at the BioSAXS beamline (ID14-3)  
269 of the ESRF (57). The scattering from the buffer alone was measured before and after each  
270 sample measurement and used for background subtraction. All data analyses were performed  
271 using the program PRIMUS from the ATSAS 3.0.0 package (58).

272

### 273 **NMR spectroscopy**

274 The spectral assignment of the different constructs of NiV P protein (P<sub>1-100</sub>, P<sub>92-190</sub>, P<sub>173-240</sub>,  
275 P<sub>223-319</sub>, P<sub>300-401</sub>, P<sub>387-479</sub>, P<sub>588-650</sub> and P<sub>XD</sub> (residues 655-709)) at 25°C in buffer B using BEST-

276 type triple resonance experiments (59). The NMR experiments were acquired at a  $^1\text{H}$   
277 frequency of 600 or 800 MHz (Agilent spectrometers). Six experiments were acquired for  
278 each construct: HNCO, intra-residue HN(CA)CO, HN(CO)CA and intra-residue HNCA,  
279 HN(COCA)CB and intra-residue HN(CA)CB. All spectra were processed in NMRPipe (60),  
280 analyzed in Sparky (61) and automatic assignment of spin systems was done using MARS  
281 (62) followed by manual verification. The assigned chemical shifts have been deposited in the  
282 BMRB database under accession numbers 50098, 50099, 50100, 50101, 50102, 50103,  
283 50105. The completeness of the assignments was above 97% for all constructs with the  
284 unassigned residues mostly belonging to the N-termini that undergo fast solvent exchange.  
285 Only the construct P<sub>387-479</sub> had a lower assignment completeness (92%) with residues S399-  
286 I402 being unassigned. Secondary structure propensities (SSPs) were calculated from the  
287 experimental C $_{\alpha}$  and C $_{\beta}$  chemical shifts according to Marsh et al. (63).

288

### 289 **X-ray crystallography**

290 Initial crystallization conditions for the tetramerization domain (aa 471-580) were identified  
291 at the High Throughput Crystallization platform of the PSB, (<https://htxlab.embl.fr>). The  
292 crystals used for data collection were obtained in 7% PEG 3350, 1 M LiCl, 0.1 M sodium  
293 citrate, 0.1 M arginine, 4% hexanediol, 0.001% NaN<sub>3</sub> at pH 5.0, and were frozen with 15%  
294 glycerol as cryo-protectant. X-ray diffraction data were collected at the ID29 beamline of the  
295 ESRF at a wavelength of 0.9793 Å and at Proximal beamline of SOLEIL at a wavelength of  
296 0.97895 Å and at a temperature of 100 K and were processed with the XDS package (64).  
297 Initial phases were obtained using the anomalous scattering from selenium atoms by the SAD  
298 method with the program HKL2MAP (65). A model was initially constructed with the  
299 Autobuild program (66) from the phenix suite (67) and subsequently refined with the  
300 phenix.refine program (68) and Coot (69). The geometry of the final model was checked with  
301 MolProbity (70). In the model, 98.04% of residues had backbone dihedral angles in the  
302 favored region of the Ramachandran plot, 1.83% fall in the allowed regions and 0.13% were  
303 outliers. Figures were generated with PyMol (71) and Chimera (72).

304

### 305 **Modeling of NiV phosphoprotein as an ensemble of conformers**

306 An ensemble of 10,000 full-length NiV phosphoprotein conformers was generated with the  
307 program Flexible-Meccano (73). These included the high-resolution three-dimensional  
308 structures of the tetramerization domain (P<sub>MD</sub>) and an homology model of the C-terminal X  
309 domain based on the known HeV P<sub>XD</sub> structure (74) by using Modeller (75). To accurately

310 model the disordered N-terminal region, the ASTEROIDS selection algorithm (76) was first  
311 applied to each of the sub-constructs of the phosphoprotein in order to obtain an ensemble of  
312 conformers of these regions in agreement with experimental  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$ , and  $^{13}\text{C}'$   
313 chemical shifts. Briefly, in successive rounds of selection,  $5 \times 200$  pairs of  $(\phi, \psi)$  dihedral  
314 angles were selected for each residue from a pool of 10,000 conformers assembled by  
315 Flexible-Meccano such that the difference between ensemble-averaged  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$   
316 chemical shifts predicted by the program SPARTA (77) and corresponding experimental  
317 values was minimized. The selected dihedral angles were then used in generation of a new set  
318 of 10,000 conformers for the next round of selection. The  $(\phi, \psi)$  databases for each sub-  
319 construct of the phosphoprotein thus obtained after four rounds of selection were combined  
320 and used by Flexible-Meccano to build conformations of full-length protein. Residues in the  
321 linker region between the tetramerization domain and the X domain were stochastically  
322 assigned amino-acid specific backbone dihedral angles in agreement with random coil  
323 statistics. In the disordered regions, side chains were constructed using the program SCCOMP  
324 (78). Scattering curves were calculated with CRY SOL (ATSAS 3.0.0) (79) and ensemble  
325 selections were performed using ASTEROIDS (80) by combining both the experimental  
326 NMR and SAXS data. In order to account for conformational heterogeneity within different  
327 regions of the full-length protein, the ASTEROIDS target function combined SAXS data from  
328 three constructs:  $\text{P}_{\text{FL}}$ ,  $\text{P}_{1-580}$  and  $\text{P}_{471-709}$ , representing the full-length tetramer, and the  
329 tetramerization domain with the N-terminal and C-terminal parts of the protein, respectively.

330

### 331 **Isothermal titration calorimetry**

332 Isothermal titration calorimetry (ITC) experiments for characterizing the interaction between  
333 the five fragments  $\text{P}_{92-190}$  (WT, Y116A, G120E, G125E, H117A/Q118A/H1119A) and  
334  $\text{STAT1}_{132-683}$  were performed with a MicroCal iTC200 system (North Hampton, MA). The  
335  $\text{STAT1}$  solution at a concentration of 100  $\mu\text{M}$  in buffer C was loaded in the calorimetric cell  
336 and titrated at 25°C, typically, by performing 16 injections of 2.5  $\mu\text{l}$  of the wild-type (WT)  
337  $\text{P}_{92-190}$  solution or of one of the variants (at concentrations ranging between 1.5 and 2.5 mM)  
338 in the same buffer. Protein samples and buffers were degassed before data collection. The  
339 heat generated for each injection was obtained by integrating the calorimetric signal. The heat  
340 resulting from the binding reaction between the two proteins was obtained as the difference  
341 between the heat of the reaction and the corresponding heat of dilution. The data were  
342 processed using Origin 7.0 (Malvern Panalytical, Malvern, UK) and fitted to a single-site  
343 binding model.

344

345

### 346 **STAT1 localization in NiV infected cells**

347 Generation of recombinant NiV was described elsewhere (6). NiV-G121E and NiV-G135E  
348 contain a single amino acid change in the P protein at position 121 or 135. To monitor the  
349 effect of NiV infection on endogenous STAT1, Vero E6 cells were grown on polymer  
350 coverslips of the  $\mu$ -Slide VI<sup>0.4</sup> and were mock infected or infected with recombinant NiVs  
351 (WT, G121E or G135E) at a multiplicity of infection (m.o.i.) of 0.5. At 1 h post-infection, the  
352 virus inoculum was replaced with DMEM containing 2.5% fetal bovine serum and the cells  
353 were incubated at 37°C for 28 h. The cells were washed with serum-free medium and then  
354 treated with 1,000 U/mL of human interferon- $\beta$  (PBL) in serum-free DMEM for 40 min at  
355 37°C. For untreated controls, infected cells were incubated with IFN-free medium. Cells were  
356 washed twice with cold phosphate-buffered saline (PBS), fixed with 4% paraformaldehyde-  
357 PBS solution for 20 min at room temperature, and subsequently treated with 0.1 M glycine for  
358 10 min. Cells were permeabilized with 0.2% Triton X-100 in PBS for 10 min and washed  
359 with a blocking solution (2% BSA in PBS). Immunofluorescence staining was performed  
360 with mouse anti-phospho-STAT1 (pY701) (BD Biosciences) (1:250 dilution) antibody or  
361 rabbit anti-NiV matrix protein (1:500) antibody.

362

## 363 **RESULTS**

### 364 **NiV phosphoprotein has a modular architecture**

365 The presence of intrinsically disordered regions (IDR) in a protein can be predicted from its  
366 amino acid sequence by different computational approaches (81, 82). Previously, we devised  
367 a meta-prediction method that calculates a consensus disordered score (D-score) between 16  
368 independent disorder predictors by a simple voting procedure (32) and found that by applying  
369 a threshold at 0.5 to distinguish between ordered and disordered regions we were able to  
370 precisely predicts boundaries between folded domains and IDRs (32, 50, 52, 83) as well as the  
371 existence of MoREs (36).

372

373 The D-score profile calculated from NiV phosphoprotein sequence reveals three disordered  
374 regions of D-score < 0.5 (named IDR1-3) and several regions of D-score > 0.5 predictive of  
375 structural order (**Figures 1A and 1B**) in agreement with the Hydrophobic Cluster Analysis  
376 plot (**Supplementary Figure S1**) (84). Two of these structured regions of more than 100 aa  
377 between residues 475 and 598 (averaged D-score of 0.82) and between residues 658 and 709

378 (averaged D-score of 0.69) coincide with the known localization of the folded tetramerization  
379 ( $P_{MD}$ ) and X domains ( $P_{XD}$ ), respectively (37, 39). A shorter region of D-score  $> 0.5$  at the N-  
380 terminal end (aa 1-29) corresponds to the  $\alpha$ -MoRE known to bind the unassembled  
381 nucleoprotein  $N^0$  (9). However, the D-score also identifies other putatively ordered regions. A  
382 region of D-score  $> 0.5$ , also spanning more than 100 aa between residues 102 and 219,  
383 encompasses the putative binding sites for STAT1, STAT2 and Polo-like kinase 1 as well as  
384 the CRM1-dependent NES (**Figure 1B**). A shorter region of D-score  $> 0.5$  is localized within  
385 the C-terminal linker (aa 607-658). Finally, borderline regions appear between residues 310  
386 and 450 in the form of clearly visible peaks in the D-score even if the score remains under the  
387 threshold of 0.5. One of these peak encompasses the conserved ultra-weak nucleoprotein  
388 binding motif recently discovered in measles virus phosphoprotein (10), while two other ones  
389 overlap with the regions encoding the specific moiety of the V and W proteins (**Figure 1B**).

390

### 391 **Protein production and quality control**

392 We expressed the full-length NiV phosphoprotein and its different fragments as recombinant  
393 proteins in *Escherichia coli* (**Figure 1C**). The full-length protein and the different protein  
394 fragments were soluble and eluted as single peaks from the SEC column. The molecular mass  
395 determined by on-line multi-angle laser light scattering (MALLS) detection was constant  
396 throughout the chromatographic peak (polydispersity index,  $M_w/M_n$  lower than 1.01)  
397 indicating that each sample was monodisperse (**Table 1**).

398

### 399 **NiV full-length phosphoprotein is a tetramer of large size**

400 The molecular mass of full-length NiV phosphoprotein ( $P_{FL}$ ) measured by on-line MALLS is  
401  $306 \pm 30$  kDa, in agreement with the molecular mass calculated for the tetramer ( $4 \times 79.367 =$   
402  $317.468$  kDa) (**Figure 2 and Table 1**). Its hydrodynamic radius ( $R_h$ ) measured by SEC is  $9.3$   
403  $\pm 0.2$  nm, a value that is intermediate between those expected for a compact globular protein  
404 ( $R_{h,theo} = 6.0$  nm) or an unfolded chain ( $R_{h,theo} = 16.8$  nm) of the same molecular mass (**Table**  
405 **1**). Analyses of the different protein fragments by SEC-MALLS indicate that the central  
406 region is solely responsible for the tetramerization of the protein in accordance with cross-  
407 linking experiments (85) and the recent crystal structure of the NiV phosphoprotein  
408 multimerization domain (37). The fragments  $P_{1-580}$  and  $P_{471-709}$ , which both encompass the  
409 tetramerization domain, and the fragment  $P_{471-580}$ , which closely corresponds to this domain,  
410 have molecular masses determined by SEC-MALLS of  $250 \pm 25$  kDa,  $113 \pm 10$  kDa and  $58 \pm$   
411  $6$  kDa (**Figure 2**), respectively, in agreement with the theoretical molecular masses expected

412 for their tetrameric forms (**Table 1**). Similarly to the full-length protein, the hydrodynamic  
413 radius of these fragments is intermediate between those expected for compact globular  
414 proteins and for unfolded chains of the same molecular mass (**Table 1**). By contrast, all other  
415 fragments are monomeric and those from the N-terminal region have a hydrodynamic radius  
416 close to that of an unfolded chain (**Table 1** and see below).

417  
418 SAXS curves were recorded at three different protein concentrations for the full-length  
419 phosphoprotein (**Supplementary Figure S2**) and for both tetrameric constructs: P<sub>1-580</sub>,  
420 (**Supplementary Figure S3**) and P<sub>471-709</sub> (**Supplementary Figure S4**). For each protein, the  
421 scattering profile shows an absence of concentration dependence and the Guinier plots at low  
422 Q values are linear, indicating the absence of aggregation or of significant interchain  
423 interactions (**Supplementary Figures S2, S3 and S4**). We note, however, an unexplained  
424 overestimation of more than 30% in the molecular mass of full-length phosphoprotein and of  
425 P<sub>471-709</sub> calculated by using the calibration with bovine serum albumin (BSA) (Table 2). The  
426 dimensionless Kratky plot and the pair-distance distribution function revealed ambivalent  
427 features for the three proteins, supporting the presence of both structured and disordered parts  
428 (**Supplementary Figures S2, S3 and S4 and Table 2**) (86). A peak slightly above 1.1 is  
429 visible in the Kratky plot at  $Q \cdot R_g \approx 2$  likely reflecting the presence of the folded  
430 tetramerization and X domains. For full-length phosphoprotein and the fragment P<sub>1-580</sub>, the  
431 value evolves towards a plateau at higher  $Q \cdot R_g$  indicating the presence of a large fraction of  
432 disordered regions in agreement with the asymmetric bell-shape curve in the P(r) plots. For  
433 the fragment P<sub>471-709</sub>, a maximum is also present at  $Q \cdot R_g \approx 5$  in the Kratky plot, and two  
434 maxima at  $r \approx 4$  nm and  $r \approx 8.5$  nm, likely reflecting on the long-range contacts between the  
435 tetramerization domain and the four X domains. The radius of gyration determined from the  
436 Guinier plot and the longest pair distance  $D_{\max}$  were almost identical for the full-length  
437 protein and the N-terminal fragment P<sub>1-580</sub> indicating the overall size of these molecules is  
438 mainly determined by the N-terminal disordered region (**Table 1**). The corresponding values  
439 for the C-terminal fragment P<sub>471-709</sub> was clearly smaller in agreement with the largest fraction  
440 of residues involved in structured domains (~65% of folded residues) (**Table 1**).

441

#### 442 **Crystal structure of the tetramerization domain**

443 The fragment P<sub>471-580</sub> crystallized in different conditions and different space groups. We  
444 solved the structure by the SAD method using selenomethionine labeled protein crystallized

445 in space group P12<sub>1</sub>1 (see **Table 3** for details of data collection and statistics) The asymmetric  
446 unit contains two copies of a tetramer (**Supplementary Figures S5A and S5B**), in which  
447 each monomer is composed of two short helices (helices  $\alpha_1$  and  $\alpha_2$ ) and a long C-terminal  
448 helix (helix  $\alpha_3$  - aa 508-570) (**Figures 3A and 3B**). The long helix of each monomer is  
449 composed of eight complete heptad repeats and a stammer at position 540-542 resulting in a  
450 kink near residue Pro544 (**Figure 3C**). Within one tetramer, the four long helices fold into a  
451 parallel coiled-coil forming a central shaft, while the short helices form a crown around the N-  
452 terminal end of this shaft (**Figures 3A and 3B**). The global net charge of the tetrameric  
453 domain is negative at neutral pH, and the calculated electrostatic surface potential shows an  
454 asymmetrical distribution, with a negative head and a patch of negative charges at the C-  
455 terminal end of the coiled coil shaft, whereas the central part of the coiled-coil shaft is rather  
456 neutral (**Supplementary Figure S5C**).

457  
458 The N-terminal part of the tetrameric domain exhibits four-fold symmetry, but in the coiled-  
459 coil shaft, the four-fold symmetry is progressively replaced by a two-fold symmetry when  
460 going from the N-terminal end to the C-terminal end (**Figures 3D and 3E**). The monomers  
461 within a tetramer superimpose by pairs with  $C_\alpha$  r.m.s.d. values around 0.6 (two-fold  
462 symmetrical mate) and 1.0 Å (other two chains). Each protomer  $N_i$  contacts the three other  
463 protomers. The long helix  $\alpha_3$  of  $N_i$  is packed between helix  $\alpha_3$  of the  $N_{i-1}$  and  $N_{i+1}$  protomers,  
464 while helix  $\alpha_2$  contacts the  $N_{i+1}$  protomer and helix  $\alpha_1$  contacts the  $N_{i+2}$  protomer (**Figure 3B**).  
465 In the corresponding tetrameric domain of Sendai virus phosphoprotein, each protomer only  
466 interacts with two other adjacent protomers (38).

467  
468 The tetrameric structure of NiV tetramerization domain is stabilized by multiple interchain  
469 interactions, each protomer forming on average 19 H-bonds and 11 salt bridges with the  
470 others protomers and burying a surface area of 2,200 Å<sup>2</sup> between each pair of adjacent  
471 protomers (for a total of 8,800 Å<sup>2</sup> of accessible surface area buried in the tetrameric  
472 structure). The interior of the coiled-coil shaft is composed of 19 sidechain layers, mainly  
473 hydrophobic but with four layers of polar residues (C512, S515, G519 and A522) at the N-  
474 terminal end and one near the C-terminal end (T560) (**Figure 3C**).

475  
476 The structure is very similar to the crystal structure of NiV tetramerization domain published  
477 by Sapphire and coworkers (37) with, however, two noticeable differences: (1) Whereas the  
478 constructs are similar in length, the coiled-coil region in our structure is six residues shorter



479 than in the previous structure (62 aa instead of 68 aa), with a length intermediate between that  
480 of Sendai virus (64 aa) (38) and that of measles virus (60 aa) (87), suggesting crystal contacts  
481 play a role in the stabilization of this part of the structure. (2) The C-terminal part of the  
482 coiled coil exhibits a slightly different superhelical frequency ( $-3.39^\circ$  versus  $-3.63^\circ$ ) resulting  
483 in a different superhelical pitch (158 Å *versus* 148 Å) (**Supplementary Figure S5D**). The  
484 mapping of crystallographic B-factors onto the structure confirms that the C-terminal end of  
485 the coiled-coil shaft is more flexible than the remaining part of the molecule (**Supplementary**  
486 **Figure S5E**) as shown previously (34, 37).

487

### 488 **The N-terminal region and the C-terminal linker are globally disordered**

489 Several fragments of about 100 amino acids covering the entire N-terminal region but also the  
490 linker in the C-terminal part (P<sub>CT-LINK</sub>) that is predicted to be at least partially disordered  
491 (IDR3 in **Figure 1A**) and the X domain were produced as  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled proteins and  
492 purified (**Figure 1C**). Each fragment from the N-terminal region has a hydrodynamic radius  
493 ( $R_h$ ) significantly larger than that expected for a globular protein of the same molecular mass  
494 and slightly smaller than that expected for a fully unfolded protein in a good solvent (**Table**  
495 **1**). SAXS data recorded for these fragments revealed no concentration dependence. For the  
496 different N-terminal fragments and for the fragment corresponding to the C-terminal linker  
497 (P<sub>588-650</sub>), the experimental radius of gyration ( $R_g$ ) values determined using the Guinier  
498 approximation were close to the  $R_g$  values calculated for a random coil according to Flory's  
499 power-law dependence on chain length using parameters for chemically unfolded proteins  
500 (88) or for intrinsically disordered proteins (89) (**Table 1**), in line with the proposal that these  
501 regions are disordered. By contrast, the radius of gyration for the X domain is significantly  
502 smaller than that of a disordered chain of the same length, as expected for a folded X domain  
503 similar to that of Hendra virus (74).

504

505 To characterize the structure of these fragments in solution, we carried out assignment of the  
506 backbone NMR resonances and calculated their secondary structure propensities (SSP) from  
507 the experimental  $\text{C}\alpha$  and  $\text{C}\beta$  chemical shifts. The  $^1\text{H}$ - $^{15}\text{N}$  two-dimensional heteronuclear  
508 single quantum coherence (HSQC) spectra of the different N-terminal fragments exhibit poor  
509 chemical shift dispersion of the amide  $^1\text{H}$  resonances as is typical for a disordered protein (90)  
510 (**Supplementary Figure S6**). The combined SSP score for the entire N-terminal region  
511 (P<sub>NTR</sub>) reveals the presence of several transient  $\alpha$ -helices (**Figure 4A**). In particular, we  
512 observe two transiently populated helical elements at the very N-terminus that make up the

513 binding region for N<sup>0</sup> (9). In the N<sup>0</sup>-P complex, these helical elements (aa 1-29) bind to the  
514 C-terminal domain of the nucleoprotein mainly through hydrophobic interactions, forming a  
515 stable helix, kinked at residue Q21 (9). In addition, a third helical element with a population  
516 of around 40% is observed at residues 340-355, which coincides with the putative additional  
517 weak binding region for the nucleoprotein recently discovered in measles virus (10). The <sup>1</sup>H-  
518 <sup>15</sup>N HSQC spectrum of the fragment P<sub>588-650</sub> corresponding to the C-terminal linker (P<sub>CT-LINK</sub>)  
519 is also typical of a disordered protein (**Supplementary Figure S6**) and the combined SSP  
520 score also revealed the presence of low populated  $\alpha$ -helices (aa 600-625) (**Figure 4B**). By  
521 contrast, the <sup>1</sup>H-<sup>15</sup>N HSQC spectrum of the isolated X domain, (P<sub>XD</sub>, aa 655-709) confirms  
522 that this domain is folded in solution, and the chemical shift assignments clearly show that a  
523 three-helix bundle structure is predominant in solution (**Figure 4C**) as also observed in the  
524 corresponding domains of measles, mumps, Sendai, and Hendra viruses (74, 91–93) .

525  
526 The HSQC spectrum of the tetrameric full-length phosphoprotein (a tetrameric protein  
527 complex of 2,868 aa with our construct) contains resonances with limited chemical shift  
528 dispersion that superposed well with the resonances from the residues of the N-terminal  
529 disordered part of the protein (**Figure 4D** and **Supplementary Figure S6**). This clearly  
530 indicates that the N-terminal part is flexible with similar conformational behavior in the  
531 tetramer as in the isolated peptides and negligible interchain contacts. No resonances for the  
532 tetrameric domain, C-terminal linker or the X domain could be identified in the spectrum of  
533 full-length phosphoprotein, possibly reflecting slower rotational tumbling caused by a  
534 hydrodynamic drag due to the connection to the tetrameric domain and the presence of the X  
535 domain at the C-terminus, although we cannot rule out the possibility of interactions between  
536 the four C-terminal regions that may also give rise to broadening.

537  
538 Overall, these results show that (i) the N-terminal region of NiV phosphoprotein is  
539 intrinsically disordered and exchanges between multiple conformers on a fast timescale, (ii)  
540 this region contains transient  $\alpha$ -helical elements, (iii) there are no significant interactions  
541 between the N-terminal region of the four protomers within the tetrameric phosphoprotein and  
542 (iv) the C-terminal linker also seem to be disordered and to contain transient  $\alpha$ -helical  
543 elements.

544

#### 545 **Ensemble description of full-length NiV phosphoprotein tetramer**

546 To study the conformational diversity of NiV phosphoprotein, we generated ensembles of all-

547 atom conformations of the full-length protein and of both tetrameric fragments: P<sub>1-580</sub> and  
548 P<sub>471-709</sub>. We used our crystal structure of the tetrameric domain (**Figure 3**), a homology model  
549 of the X domain based on the known structure of the X domain of Hendra virus (74),  
550 heteronuclear backbone chemical shifts (C $\alpha$ , C $\beta$ , C' and <sup>15</sup>N) for the N-terminal fragments  
551 (**Figure 4**) and SAXS data for the three proteins (**Supplementary Figures S2, S3 and S4**) as  
552 experimental constraints to delineate the contours of conformational space sampled by the  
553 full-length protein. We used a combine fit of multiple curves to select conformational sub-  
554 ensembles of 10 conformers for which the three average theoretical scattering curves  
555 reproduced the three experimental curves. The average SAXS curve calculated for this  
556 ensemble is comparable to the experimental SAXS curve measured for the full-length  
557 phosphoprotein, while the average curves calculated with the properly truncated forms of the  
558 same 10 conformers are comparable to the experimental SAXS curves obtained for P<sub>1-580</sub> and  
559 P<sub>471-709</sub> (**Figure 5A**). A global  $\chi^2$  value of 1.53 and the plot of reduced residuals indicate the  
560 quality of the fit (**Figure 5A**). It should be noted that the independent fits performed for each  
561 of the three proteins yielded ensembles of conformers with averaged R<sub>g</sub> values in agreement  
562 with the experimental values (full-length P: 13  $\pm$  3 nm; P<sub>471-709</sub>: 6.3  $\pm$  0.5 nm; P<sub>1-580</sub>: 12  $\pm$  3  
563 nm), but did not improve the overall quality of the fits (**Supplementary figure S7**). Some  
564 discrepancies between the experimental and fitted curves are visible. The downward deviation  
565 observed at low Q values for full-length phosphoprotein and to a lesser extent for the P<sub>1-580</sub>  
566 fragment can be due to intermolecular interactions, whereas the discrepancy observed  
567 between the experimental and calculated curves for the fragment P<sub>471-709</sub> at Q values > 1.0 nm<sup>-1</sup>  
568 <sup>1</sup> (**Figure 5A**) might result from non-optimal [buffer subtraction as already observed for](#)  
569 [proteins such as this fragment, which comprises both folded and unfolded domains that](#)  
570 [may have distinct electron densities \(94, 95\)](#). Alternatively, it can also result from a bias  
571 in the construction of the initial ensemble of the full-length protein. As the initial ensemble of  
572 conformers was built for full-length protein and was then truncated to generate the initial  
573 ensembles for the fragments P<sub>1-580</sub> and P<sub>471-709</sub>, it is possible that the large N-terminal  
574 disordered region has constrained the X domains to position at one pole of the tetramerization  
575 domain, whereas such constrain is absent in the fragment P<sub>471-709</sub>. Finally, additional structural  
576 organization or inter-chain interactions might be present in the C-terminal part of the  
577 phosphoprotein, which has not been taken into account in our models. It is thus remarkable  
578 that the same set of conformers is able to represent the curves obtained for the three  
579 constructions. The R<sub>g</sub> values of the initial ensemble for the full-length protein exhibited broad  
580 Gaussian distribution centered around 12 nm, and the R<sub>g</sub> values of the selected ensembles of

581 10 conformers exhibit a similar broad distribution (**Figure 5B**), suggesting the absence of  
582 constraints on the overall size of the protein. **Figure 5C** shows a typical model of the full-  
583 length phosphoprotein and **Figure 5D** shows the overlays of the 10 conformers of each  
584 construct selected in the same fitting process for reproducing the data shown panel A.

### 585 **The N-terminal region of NiV phosphoprotein binds human STAT1 with low affinity**

586 The region of NiV phosphoprotein that binds STAT1 was previously mapped between  
587 residues 100 and 160 (43, 47), which contains several highly conserved residues among  
588 henipaviruses (**Figure 6A**). Point mutations within residues 81 to 120 of the phosphoprotein  
589 affected the expression of a minireplicon suggesting that this region is involved in the  
590 functioning of the polymerase machine, while mutations within residues 111 and 140  
591 prevented the inhibition of interferon signaling (6).

592

593 To investigate the interaction between NiV phosphoprotein and STAT1, we used the fragment  
594 P<sub>92-190</sub> that contains the entire putative binding site for STAT1, and we expressed and purified  
595 a fragment of human STAT1 that comprises residues 132 to 683 encompassing the coiled-coil  
596 domain, the DNA binding domain, the linker domain and the SH2 domain (96) (**Figure 6B**).  
597 The missing 131 N-terminal residues and 38 C-terminal residues from the STAT1 $\alpha$  isoform  
598 are not required for interaction with the phosphoprotein, V or W proteins of NiV (44). The  
599 protein STAT1<sub>132-683</sub> eluted from the SEC column as a single peak and its molecular mass  
600 determined by SEC-MALLS of  $64 \pm 6$  kDa is in good agreement with the theoretical mass of  
601 64.1 kDa (**Figure 6C**). The fragment P<sub>92-190</sub> alone also eluted as a single peak and its  
602 molecular mass of  $11 \pm 1$  kDa is in agreement with the theoretical mass of the monomer. In a  
603 first approach, we mixed STAT1<sub>132-683</sub> ( $\sim 54 \mu\text{M}$ ) with P<sub>92-190</sub> ( $\sim 540 \mu\text{M}$ ) and loaded the  
604 mixture on a SEC column after 2 hour incubation at 20°C. Both proteins eluted independently  
605 at elution volumes and with molecular masses corresponding to isolated STAT1<sub>132-683</sub> and P<sub>92-190</sub>  
606 (**Figure 6C**). In a second approach, we used NMR spectroscopy and recorded <sup>1</sup>H-<sup>15</sup>N  
607 HSQC spectra of <sup>15</sup>N labeled P<sub>92-190</sub> at a final concentration of  $\sim 22 \mu\text{M}$  in the absence or  
608 presence of unlabeled STAT1<sub>132-683</sub> ( $\sim 44 \mu\text{M}$  final concentration) (**Figure 6D**). In a complex  
609 of the size expected for the heterodimer STAT1<sub>132-683</sub>-P<sub>92-190</sub> ( $\sim 75$  kDa), NMR signals are  
610 strongly broadened in protonated samples, thus precluding their detection. By comparing the  
611 HSQC spectra, we find a partial reduction of the intensity in the presence of STAT1<sub>132-683</sub> for  
612 resonances corresponding to residues 100 to 150 (**Figure 6E**). The most affected resonance  
613 intensities are within the region of residues 110 to 140, which contains several residues of

614 crucial importance for blocking STAT1 signaling in cell-based assays (6). We interpret this  
615 result as evidence that, under these experimental conditions, only a fraction of P<sub>92-190</sub>  
616 molecules bind STAT1<sub>132-683</sub>, and we estimate the dissociation constant for STAT1<sub>132-683</sub> in  
617 the range of 10 to 100 μM. We note that conformational exchange contributions from the  
618 exchange between free and STAT1-bound P<sub>92-190</sub> could also be the source of some of the  
619 observed line broadening in the interaction experiments.

620

621 We used isothermal titration calorimetry (ITC) and site-directed mutagenesis to further  
622 characterize the interaction between the two protein fragments. We used STAT1<sub>132-683</sub> as  
623 receptor at a concentration of 100 μM and performed titration with P<sub>92-190</sub> as ligand at an  
624 initial concentration between 1.5 and 2.5 mM. Despite that complete saturation was not  
625 achieved, the binding isotherm can be fitted to a single binding site model, yielding a  
626 dissociation constant (K<sub>d</sub>) of 100 ± 15 μM ( $\Delta G_{Bind}^0 = + RT \ln K_d = -5.4 \text{ kcal.mol}^{-1}$ ) (from five  
627 independent measurements) (**Figure 7A**), in good agreement with the NMR interaction  
628 experiment (**Figure 6E**). The favorable enthalpy of binding ( $\Delta H_{Bind}^0$ ) of -9.5 kcal.mol<sup>-1</sup> is  
629 partially balanced by an unfavorable entropic contribution of 4.1 kcal.mol<sup>-1</sup> ( $\Delta S_{Bind}^0 = -14$   
630 cal.mol<sup>-1</sup>.K<sup>-1</sup>) (**Table 4**).

631

632 Point mutations Y116A and G125E as well as a triple alanine mutation  
633 H117A/D118A/H119A (<sup>117</sup>AAA<sup>119</sup>) in full-length NiV phosphoprotein were previously  
634 shown to block interferon signaling and to disrupt the interaction with STAT1 (6, 46). By  
635 contrast, the single amino-acid mutation G120E had no effect (6). We expressed and purified  
636 NiV phosphoprotein fragments encompassing residues 92 to 190 (P<sub>92-190</sub>) and harboring these  
637 different mutations (Y116A, G120E, G125E and H117A/Q118A/H119A). Each protein  
638 eluted as a single peak from the size exclusion chromatography column with a molecular  
639 mass calculated from MALLS in agreement with the expected molecular mass for a  
640 monomeric protein. We performed calorimetric titrations of STAT1<sub>132-683</sub> with these different  
641 variants, but we were unable to detect any interaction with the variants Y116A and  
642 H117A/Q118A/H119A. With the G120E variant, the affinity was comparable to that of the  
643 WT protein (**Figure 7B-F and Table 4**). With the G125E variant, the affinity is slightly lower  
644 than that of WT protein, with a K<sub>d</sub> value of 210 ± 90 μM, although the difference in  $\Delta G_{Bind}^0$  of  
645 0.4 kcal/mol is not significant (**Table 4**). These results confirm that the interaction between  
646 the N-terminal region of NiV phosphoprotein and STAT1 is direct but weak.

647

648 In order to further highlight the importance of the STAT1-binding domain (amino acids 110  
649 to 140) that is present in the phosphoprotein but also in the V and W proteins, we also used  
650 recombinant WT or mutant viruses, which were generated by reverse genetics (6) and possess  
651 either intact or a defective P gene. Two mutant viruses containing substitutions at positions  
652 121 and 135 of the phosphoprotein, V and W proteins (NiV-G121E and NiV-G135E) were  
653 compared with NiV WT for their ability to block translocation into the nucleus of endogenous  
654 tyrosine-phosphorylated STAT1 activated by interferon (**Figure 7G**). NiV matrix protein was  
655 used as a marker of infection and a specific anti tyrosine-phosphorylated STAT1 antibody  
656 was used to monitor the localization of NiV P gene specific proteins (**Figure 7G**). Little or no  
657 tyrosine-phosphorylated STAT1 was detectable in the nucleus of cells infected with the WT  
658 NiV, whereas adjacent, uninfected cells exhibited clear nuclear phospho-STAT1 staining. By  
659 contrast, a loss of STAT1 binding ability in NiV G121E resulted in strong and apparent  
660 tyrosine-phosphorylated STAT1 signal present in the nuclei (**Figure 7G**). Interestingly, NiV-  
661 G135E mirrored NiV WT and showed a similar pattern of disruption of normal STAT1  
662 trafficking. Earlier it had been shown that the V protein harboring the same substitution failed  
663 to detectably bind STAT1 in immunoprecipitation studies but inhibited ISG54-driven reporter  
664 induction, albeit less efficiently (6). Our data suggest that this mutant retains sufficient  
665 STAT1-binding activity but that the interaction is too weak to be seen in simple co-expression  
666 co-immunoprecipitation assays. The mechanism for such behavior will need to be further  
667 investigated, in particular with respect to the implication of STAT2.

668

## 669 **DISCUSSION**

670 The highly disordered, tetrameric NiV phosphoprotein is an integral component of the  
671 transcription/replication machinery of this dangerous virus and of all members of the  
672 *Mononegavirales* and a hitherto largely unexplored target for viral inhibition. Due to the size  
673 and flexible nature of the phosphoprotein its conformational behavior has remained  
674 unexplored.

675

676 Here, we present the first atomistic representation of full length NiV phosphoprotein in  
677 solution. We show that, under our experimental conditions, full-length NiV phosphoprotein as  
678 well as two other long fragments containing the tetramerization domain also form tetramers in  
679 solution (**Figure 2**), and we have obtained a second crystal structure of this tetramerization  
680 domain (37, 85). The phosphoprotein of other paramyxoviruses also contains a tetrameric  
681 domain (38, 92, 97, 98), although the existence of a trimeric form in solution has been

682 described (98, 99). NiV tetramerization domain has dimensions and crystal structure similar  
683 to those of Sendai virus tetramerization domain with a parallel tetrameric coiled-coil and a  
684 crown of helices at its N-terminal end (38). By contrast, the tetramerization domains of  
685 measles and mumps viruses lack the crown of helices, and while the structure of measles  
686 domain is also a parallel tetrameric coiled-coil, that of mumps domain is composed of two  
687 pairs of parallel helices packed in opposite orientation (92, 97). Phylogenetic trees based on  
688 the phosphoprotein sequences but also on the whole genome sequences place the genus  
689 Henipavirus in an intermediate position between the genera Morbillivirus (measles virus) and  
690 Respirovirus (Sendai virus), while the genus Rubulavirus (mumps virus) is more distant (100,  
691 101). The crown of helices may thus be an ancestral feature lost by some viruses or a modern  
692 one acquired by others, that stabilizes the central coiled coil shaft or provides additional  
693 function. Up to now, no function has been specifically associated with the N-terminal part of  
694 the tetramerization domain, and mutations that perturb the crystallographic water molecules  
695 inside the coiled-coil shaft at the crown level had no effect on phosphoprotein function in a  
696 minireplicon system (34). Mapping the crystallographic B-factors on the domain structure  
697 shows that the N-terminal part of the central coiled-coil, upstream of the kink, is more rigid  
698 than the downstream part (**Supplementary Figure S5E**). Although the crystallographic  
699 packing can influence B factors, the conservation of this dynamical feature in the  
700 tetramerization domain of the related measles virus and Sendai virus argues for functional  
701 relevance. The presence of inter-helical networks of h-bonds and salt-bridges in the crown  
702 and in the upper part of the coiled-coil preceding the kink region could explain this additional  
703 rigidity of the structure. By contrast, a conserved basic patch in the C-terminal part of the  
704 shaft is known to interact with the polymerase and mutations in this part affect polymerase  
705 activity (34) (**Figures 3C and 3F**). Multiple sequence alignment of Henipavirus  
706 phosphoprotein shows that the crown is not the most conserved part of the tetramerization  
707 domain, while it reveals a highly conserved region following the basic patch and straddling  
708 the C-terminal end of the coiled coil, and the N-terminal part of the P<sub>CT-LINK</sub> (aa 565-600)  
709 (**Figures 3E and 3F**). This region includes the less rigid part of the structure  
710 (**Supplementary Figure S5E**), and in related *Mononegavirales*, it interacts directly with L  
711 polymerase (16, 17). Because the polymerase changes conformation while it operates and  
712 moves along its RNA template, the interaction between the phosphoprotein and the  
713 polymerase is submitted to tensional, shear and/or rotational stresses and thus likely requires  
714 some flexibility in the protein interfaces to maintain the attachment.

715

716 Previous sequence analyses predicted that the N-terminal part of the phosphoprotein  
717 encompassing the region common to the V and W proteins and extending up to the  
718 tetramerization domain, is an intrinsically disordered region except for the N-terminal end  
719 (31, 102), which consists of a MoRE that binds and chaperones unassembled N<sup>0</sup> (9). The D-  
720 score calculated as a consensus of multiple disorder predictions reveals a more complex  
721 picture with intervening putatively ordered regions and undefined regions (**Figure 1A**).  
722 Experimentally, we confirm the overall disorder of the N-terminal region of the  
723 phosphoprotein. The hydrodynamic radius determined by SEC and the radius of gyration  
724 determined by SAXS indicate that each fragment in the N-terminal region (**Table 1**) is rather  
725 extended, in accordance with the behavior of intrinsically disordered polypeptide chains in  
726 water (89). The poor chemical shift dispersion of the NMR <sup>1</sup>H-<sup>15</sup>N HSQC spectra also  
727 supports the high flexibility of the different fragments, but the assignment of the spectra  
728 revealed the presence of transient secondary structure elements (**Figure 4D and**  
729 **Supplementary Figure S6**), and an overlay of the SSP parameter (**Figure 4A**) and D-score  
730 (**Figure 1A**) plots shows that in several instances, short regions with an intermediate D-score  
731 (moderately above or below 0.5) are transiently folded into helical conformation and could  
732 correspond to  $\alpha$ -MoRE involved in recruiting different partners, in particular the polo-like  
733 kinase 1, the exportin or the viral nucleoprotein (**Supplementary Figure S8**). The  
734 comparison of the NMR spectra from the different N-terminal fragments with the full-length  
735 phosphoprotein shows that most resonances overlap (**Figure 4D**), thereby showing that the  
736 structural properties of fragments are the same in the full-length protein. In other words, the  
737 overall disorder of N-terminal region is not a result of our divide-and-conquer approach, but  
738 is an intrinsic property of this region of the protein. This view is also confirmed by the  
739 ensemble modeling, which shows that an ensemble model of full-length phosphoprotein,  
740 where the N-terminal region and the C-terminal linker are highly flexible and adopt random  
741 coil conformation, is in agreement with the SAXS and NMR data (**Figure 5**).

742

743 The ensemble model of full-length phosphoprotein provides new insights about the  
744 conformation of this protein in solution (**Figure 5**). Overall, the molecule appears elongated  
745 in the direction of the tetramerization domain shaft with the four X domains clustering at the  
746 C-terminal end (**Figure 5D**). This spatial arrangement certainly provides an ideal positioning  
747 of the four X domains for interacting with the polymerase or the nucleocapsid, and for  
748 "walking" along the latter, driven by the polymerase. The four N-terminal chains mainly  
749 occupy the other end and the sides of the tetramerization domain, likely creating the steric



750 hindrance that pushes the X domains towards one end. When multiple decoys are overlaid  
751 with respect to the tetramerization domain, thus mimicking the time average spatial  
752 occupancy of the four flexible N-terminal chains and four C-terminal linkers (**Figure 5D**), the  
753 central tetramerization domain including the basic site of putative interaction with the L  
754 polymerase is masked. The recent structure of the polymerase complexes of two related  
755 viruses revealed that one of the chain of the phosphoprotein is mainly involved in the  
756 interaction with the L polymerase, but the three other chains are also binding (16, 17). These  
757 multiple interactions could be required to move the N-terminal chains of the phosphoprotein  
758 away from the multimerization domain and to allow access of the polymerase to the central  
759 tetrameric shaft.

760

761 The tetrameric NiV phosphoprotein is thus a large molecule, with a hydrodynamic radius of  
762  $9.3 \pm 0.3$  nm and a radius of gyration of  $10.9 \pm 0.2$  nm. The  $R_g/R_h$  ratio is a sensitive indicator  
763 of chain compactness; experimental values for this ratio lie between limits of about 0.8 for  
764 globular proteins (103) and of about 1.5 or higher for random coils (104). The value of 1.18  
765 found here for NiV full-length phosphoprotein is in agreement with an elongated molecule  
766 and/or the presence of long intrinsically disordered regions and with the values of 1.0 and  
767 1.14 that we previously reported for the full-length phosphoproteins of vesicular stomatitis  
768 virus and rabies virus (105), respectively. Despite that these phosphoproteins of the family  
769 *Rhabdoviridae* are dimeric and less than half its size, they share a similar architecture with  
770 NiV phosphoprotein (32). Assuming a partial specific volume of  $0.73 \text{ cm}^3 \cdot \text{gr}^{-1}$ , the intrinsic  
771 molecular volume of a protein the size of tetrameric NiV phosphoprotein should be about 390  
772  $\text{nm}^3$ , whereas the volume of the equivalent sphere calculated by using the experimental  $R_h$   
773 value is  $3,400 \text{ nm}^3$ , almost ten-fold larger. Multiple copies of this protein are packaged in the  
774 virion and might occupy an important part of the particle volume if not otherwise compacted.  
775 With respect to the nucleocapsid, the length of the disordered C-terminal linker (70 aa) allows  
776 an extension of the order of 50 nm ( $2 \times 70 \times 0.35$  nm) between two X domains, which would  
777 allow to connect nucleoproteins at least 10 subunits apart along the nucleocapsid, and  
778 probably even more distant if the possible extension of the C-terminal disordered region of  
779 the nucleoprotein ( $N_{\text{TAIL}}$ ) is taken into account. In the virion, the nucleocapsids have a  
780 diameter of 18-19 nm and an average pitch of 5 nm (7), and thus one phosphoprotein could  
781 connect nucleoproteins across the nucleocapsid. The larger size of tetrameric NiV  
782 phosphoprotein may thus be of importance in the packaging of the viral particle and in the

783 functioning of the polymerase machine, catching at long distance and directing partners to  
784 their site of use.

785

786 The architecture of the phosphoprotein is conserved among *Mononegavirales*, including the  
787 localization of functional and structural modules, but the length of the protein varies  
788 considerably between families and even between genera within the same family. As  
789 illustrative examples, VSV in the family *Rhabdoviridae* is 267 aa long, whereas measles virus  
790 and NiV phosphoproteins in the family *Paramyxoviridae* are 507 and 709 aa long,  
791 respectively. The reason why NiV phosphoprotein is so long remains enigmatic, considering  
792 that VSV phosphoprotein, for example, perfectly assumes similar functions in the life cycle of  
793 the virus. In a previous study, we compared the structure of the C-terminal N-binding domain  
794 of the phosphoprotein from different *Mononegavirales* and discovered that the variation in  
795 length of this protein domain correlated with the capacity to perform additional functions,  
796 suggesting a model where the protein evolves by accretion of functional modules (52).  
797 Henipavirus phosphoproteins are significantly longer than the phosphoproteins from other  
798 genera of *Paramyxoviridae* (106). Because the L polymerases of the *Mononegavirales* have a  
799 high error-rate, the genome of these viruses has the capacity to rapidly evolve and trim  
800 superfluous parts of the genome, suggesting that the very existence of additional lengths in  
801 NiV phosphoprotein sequence is a sign that they perform additional functions or are required  
802 for the fitness of the virus. We can therefore speculate that the larger size of the N-terminal  
803 domain correlates with its implication in other functions, and that the transient  $\alpha$ -helical  
804 elements evidenced by NMR spectroscopy and predicted by sequence analysis represent short  
805 linear motifs that recruit different and yet unidentified partners. Some of them clearly  
806 encompass regions of the protein known to bind viral or cellular proteins (**Supplementary**  
807 **Figure S8**). A comparison with measles virus phosphoprotein shows that most functional and  
808 structural regions are conserved but that NiV phosphoprotein contains an additional region in  
809 the N-terminal region between residues 150 and 300 that is common to the V and W proteins,  
810 and is required for binding STAT2 (**Supplementary Figure S9**). The binding site of STAT1  
811 appears to be conserved between NiV and measles virus, being localized between residues  
812 110 and 140, but in measles virus the binding of STAT2 is mediated by the specific C-  
813 terminal region of the V protein (107). We speculate that NiV phosphoprotein, V and W  
814 proteins have thus evolved by acquiring an additional extension within their common N-  
815 terminal region that recruits STAT2. The first 405 N-terminal amino acids of N-terminal  
816 region are common to NiV phosphoprotein, V and W proteins (**Figure 1B**), and these three

817 proteins block interferon signaling by preventing the phosphorylation of STAT1 and by  
818 sequestering both STAT1 and STAT2 proteins in the cytoplasm and the nucleus (6, 43–45).  
819 The primary binding site of STAT1 was localized between residues 110 and 140 of the N-  
820 terminal region by co-immunoprecipitation (6) and this localization was confirmed here by a  
821 direct interaction assay. The short section of only 30 amino acids forming this binding site in  
822 the NiV phosphoprotein suggests that it is a typical example of a short linear motif that folds  
823 upon binding to recruit a protein partner (108) as used by numerous viruses to hijack cellular  
824 machineries (109) by mimicking cellular proteins, such as STAT1 itself, which uses a short  
825 linear motif within its transactivation domain (TAD) to interact with the TAZ1 domain of the  
826 CBP/p300 transcriptional coactivator protein (110).

827

828 The Y116A mutant and the H117A/D118A/H119A triple mutant lost their ability to bind  
829 STAT1. Similarly, the mutant G121E prevents the accumulation of phosphorylated STAT1 in  
830 the nucleus of infected cells. Surprisingly, the G125E mutant, which was unable to inhibit  
831 IFN signaling and failed to co-precipitate STAT1 and STAT2 (6, 46), shows similar binding  
832 affinity for STAT1 as the WT protein by ITC measurement. By contrast, the G135E mutant  
833 conserves the ability to block the accumulation of phosphorylated STAT1 in the nucleus,  
834 while it also fails to co-precipitate STAT1. Because the affinity between the phosphoprotein  
835 and STAT1 is weak, it is possible that the co-precipitation with STAT1 also requires the  
836 presence of STAT2 as found for measles virus (107), and that the mutations of Gly125 and  
837 Gly135 perturb the co-interaction between STAT1 and STAT2 rather than the direct  
838 interaction with STAT1. It is also possible that the difference in affinity, although small, is  
839 sufficient to produce an effect *in vivo* and that interactions with WT and the G125E variant  
840 are influenced differently by intracellular conditions such as molecular crowding.

841

842 In summary, our results revealed that the NiV phosphoprotein behaves as a nanometer-size  
843 octopus-like protein that uses its long flexible arms to expose binding sites for multiple viral  
844 and cellular partners and to orchestrate their participation in virus replication, in hijacking  
845 cellular machines and in interfering with the interferon response.

846

847 **AUTHOR CONTRIBUTIONS**

848 MRJ, FY, JMB, VV, MB and MJ conceived the experiments; MRJ, FY, GC, EC, CM, VVa,  
849 JMB performed the experiments; MRJ, FY, CM, NT, JMB, VV, MB and MJ analyzed and  
850 interpreted the data; MRJ, JMB, MB and MJ wrote the paper.

851

852 **ACCESSION CODES**

853 Coordinates and structure factors have been deposited in the Protein Data Bank under  
854 accession code [4GJW](#). SAXS data have been deposited in the Small Angle Scattering  
855 Biological Data Bank under accession code SASDHT4.

856

857 **ACKNOWLEDGMENTS**

858 We thank Dr. Rob W. Ruigrok (IBS, Grenoble, France) for helpful discussions and Dr. Daniel  
859 Panne (EMBL, Grenoble, France) for the gift of a plasmid coding for human STAT1. F.Y.  
860 was supported by a predoctoral fellowship from the Région Rhône-Alpes. The work was  
861 supported by grants from the French Agence Nationale de la Recherche to M.J. and M.B.  
862 (ANR BSV8-2012—NNViPol) and to M.J. and V.V. (ANR-18-CE11-0014-02), from the  
863 Fond de la Recherche Médicale (FRM) (grant "Equipe 2017" DEQ20170336754) to M.J.  
864 and from the FINOVI Foundation to M.J. and V.V. We acknowledge the European  
865 Synchrotron Radiation Facility for the allocation of beamtime on ID14-3 and ID23-1 and the  
866 Synchrotron SOLEIL for the allocation of beamtime on Proxima 1. We acknowledge the  
867 platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL)  
868 within the Grenoble Partnership for Structural Biology (PSB). Platform access was supported  
869 by FRISBI (ANR-10-INBS-05-02) and GRAL, a project of the University Grenoble Alpes  
870 graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003).  
871 All experiments involving live NiVs were carried out in the INSERM Jean Merieux BSL4  
872 laboratory in Lyon, France. We thank the biosafety team members for their support in  
873 conducting experiments.

874 **REFERENCES**

- 875 1. Chua, K.B., W.J. Bellini, P.A. Rota, B.H. Harcourt, A. Tamin, S.K. Lam, T.G.  
876 Ksiazek, P.E. Rollin, S.R. Zaki, W. Shieh, C.S. Goldsmith, D.J. Gubler, J.T. Roehrig,  
877 B. Eaton, A.R. Gould, J. Olson, H. Field, P. Daniels, A.E. Ling, C.J. Peters, L.J.  
878 Anderson, and B.W. Mahy. 2000. Nipah virus: a recently emergent deadly  
879 paramyxovirus. *Science* 288: 1432–1435.
- 880 2. Arunkumar, G., R. Chandni, D.T. Mourya, S.K. Singh, R. Sadanandan, P. Sudan, and  
881 B. Bhargava. 2019. Outbreak investigation of nipah virus disease in Kerala, India,  
882 2018. *J. Infect. Dis.* 219: 1867–1878.
- 883 3. Luby, S.P. 2013. The pandemic potential of Nipah virus. *Antiviral Res.* 100: 38–43.
- 884 4. Wang, L.F., M. Yu, E. Hansson, L.I. Pritchard, B. Shiell, W.P. Michalski, and B.T.  
885 Eaton. 2000. The exceptionally large genome of Hendra virus: support for creation of a  
886 new genus within the family Paramyxoviridae. *J. Virol.* 74: 9972–9979.
- 887 5. Halpin, K., B. Bankamp, B.H. Harcourt, W.J. Bellini, and P.A. Rota. 2004. Nipah virus  
888 conforms to the rule of six in a minigenome replication assay. *J Gen Virol.* 85: 701–  
889 707.
- 890 6. Ciancanelli, M.J., V.A. Volchkova, M.L. Shaw, V.E. Volchkov, and C.F. Basler. 2009.  
891 Nipah virus sequesters inactive STAT1 in the nucleus via a P gene-encoded  
892 mechanism. *J Virol.* 83: 7828–7841.
- 893 7. Gutsche, I., A. Desfosses, G. Effantin, W.L. Ling, M. Haupt, R.W.H. Ruigrok, C.  
894 Sachse, and G. Schoehn. 2015. Near-atomic cryo-EM structure of the helical measles  
895 virus nucleocapsid. *Science* 348: 704–708.
- 896 8. Leyrat, C., F. Yabukarski, N. Tarbouriech, E.A. Ribeiro Jr., M.R. Jensen, M.  
897 Blackledge, R.W. Ruigrok, and M. Jamin. 2011. Structure of the vesicular stomatitis  
898 virus N-P complex. *PLoS Pathog.* 7: e1002248.
- 899 9. Yabukarski, F., P. Lawrence, N. Tarbouriech, J.M. Bourhis, E. Delaforge, M.R. Jensen,  
900 R.W. Ruigrok, M. Blackledge, V. Volchkov, and M. Jamin. 2014. Structure of Nipah  
901 virus unassembled nucleoprotein in complex with its viral chaperone. *Nat Struct Mol*  
902 *Biol.* 21: 754–759.
- 903 10. Milles, S., M.R. Jensen, C. Lazert, S. Guseva, S. Ivashchenko, G. Communie, D.  
904 Maurin, D. Gerlier, R.W.H. Ruigrok, and M. Blackledge. 2018. An ultraweak  
905 interaction in the intrinsically disordered replication machinery is essential for measles  
906 virus function. *Sci. Adv.* 4: eaat7778.
- 907 11. Aggarwal, M., G.P. Leser, C.A. Kors, and R.A. Lamb. 2017. Structure of the

- 908 paramyxovirus PIV5 nucleoprotein in complex with an amino-terminal peptide of the  
909 phosphoprotein. *J. Virol.* 92: e1304-17.
- 910 12. Milles, S., M.R. Jensen, G. Communie, D. Maurin, G. Schoehn, R.W.H. Ruigrok, and  
911 M. Blackledge. 2016. Self-Assembly of Measles Virus Nucleocapsid-like Particles:  
912 Kinetics and RNA Sequence Dependence. *Angew. Chemie - Int. Ed.* 55: 9356–9360.
- 913 13. Jamin, M., and F. Yabukarski. 2016. Non-segmented negative-sense RNA viruses –  
914 structural data brings new insights into nucleocapsid assembly. *Adv Virus Res.* 97:  
915 143–185.
- 916 14. Jordan, P.C., C. Liu, P. Raynaud, M.K. Lo, C.F. Spiropoulou, J.A. Symons, L.  
917 Beigelman, and J. Deval. 2018. Initiation, extension, and termination of RNA synthesis  
918 by a paramyxovirus polymerase. *PLOS Pathog.* 14: e1006889.
- 919 15. Ogino, T., and T.J. Green. 2019. RNA Synthesis and Capping by Non-segmented  
920 Negative Strand RNA Viral Polymerases: Lessons From a Prototypic Virus. *Front.*  
921 *Microbiol.* 10: 1–28.
- 922 16. Gilman, M.S.A., C. Liu, A. Fung, I. Behera, P. Jordan, P. Rigaux, N. Ysebaert, S.  
923 Tcherniuk, J. Sourimant, J.-F. Eléouët, P. Sutto-Ortiz, E. Decroly, D. Roymans, Z. Jin,  
924 and J.S. McLellan. 2019. Structure of the Respiratory Syncytial Virus Polymerase  
925 Complex. *Cell.* 179: 193-204.e14.
- 926 17. Pan, J., X. Qian, S. Lattmann, A. El Sahili, T.H. Yeo, H. Jia, T. Cressey, B. Ludeke, S.  
927 Noton, M. Kalocsay, R. Fearn, and J. Lescar. 2019. Structure of the human  
928 metapneumovirus polymerase phosphoprotein complex. *Nature.* 577: 275–279.
- 929 18. Jenni, S., L.M. Bloyet, R. Diaz-Avalos, B. Liang, S.P.J. Whelan, N. Grigorieff, and  
930 S.C. Harrison. 2020. Structure of the Vesicular Stomatitis Virus L Protein in Complex  
931 with Its Phosphoprotein Cofactor. *Cell Rep.* 30: 53-60.e5.
- 932 19. Horwitz, J.A., S. Jenni, S.C. Harrison, and S.P.J. Whelan. 2020. Structure of a rabies  
933 virus polymerase complex from electron cryo-microscopy. *Proc. Natl. Acad. Sci. U. S.*  
934 *A.* 117: 2099–2107.
- 935 20. 2013. *Fields Virology* 6th edition. 6th ed. Philadelphia: Lippincott Williams &  
936 Wilkins.
- 937 21. Shiell, B.J., D.R. Gardner, G. Cramer, B.T. Eaton, and W.P. Michalski. 2003. Sites of  
938 phosphorylation of P and V proteins from Hendra and Nipah viruses: Newly emerged  
939 members of Paramyxoviridae. *Virus Res.* 92: 55–65.
- 940 22. Ludlow, L.E., M.K. Lo, J.J. Rodriguez, P.A. Rota, and C.M. Horvath. 2008.  
941 Henipavirus V Protein Association with Polo-Like Kinase Reveals Functional Overlap

- 942 with STAT1 Binding and Interferon Evasion. *J. Virol.* 82: 6259–6271.
- 943 23. Fuentes, S.M., D. Sun, A.P. Schmitt, and B. He. 2010. Phosphorylation of  
944 paramyxovirus phosphoprotein and its role in viral gene expression *Sandra. Futur.*  
945 *Microbiol.* 5: 9–13.
- 946 24. Morin, B., A.A. Rahmeh, and S.P. Whelan. 2012. Mechanism of RNA synthesis  
947 initiation by the vesicular stomatitis virus polymerase. *EMBO J.* 31: 1320–1329.
- 948 25. Ranadheera, C., R. Proulx, M. Chaiyakul, S. Jones, A. Grolla, A. Leung, J. Rutherford,  
949 D. Kobasa, M. Carpenter, and M. Czub. 2018. The interaction between the Nipah virus  
950 nucleocapsid protein and phosphoprotein regulates virus replication. *Sci. Rep.* 8: 1–14.
- 951 26. Bloyet, L.-M., J. Brunel, M. Dosnon, V. Hamon, J. Eroles, A. Gruet, C. Lazert, C.  
952 Bignon, P. Roche, S. Longhi, and D. Gerlier. 2016. Modulation of Re-initiation of  
953 Measles Virus Transcription at Intergenic Regions by PXD to NTAIL Binding  
954 Strength. *PLOS Pathog.* 12: e1006058.
- 955 27. Bloyet, L.-M., A. Schramm, C. Lazert, B. Raynal, M. Hologne, O. Walker, S. Longhi,  
956 and D. Gerlier. 2019. Regulation of measles virus gene expression by P protein coiled-  
957 coil properties. *Sci. Adv.* 5: eaaw3702.
- 958 28. Cox, R.M., S.A. Krumm, V.D. Thakkar, M. Sohn, and R.K. Plemper. 2017. The  
959 structurally disordered paramyxovirus nucleocapsid protein tail domain is a regulator  
960 of the mRNA transcription gradient. *Sci. Adv.* 3: e1602350.
- 961 29. Karlin, D., and R. Belshaw. 2012. Detecting remote sequence homology in disordered  
962 proteins: discovery of conserved motifs in the N-termini of Mononegavirales  
963 phosphoproteins. *PLoS One.* 7: e31719.
- 964 30. McCarthy, A.J., and S.J. Goodman. 2010. Reassessing conflicting evolutionary  
965 histories of the Paramyxoviridae and the origins of respiroviruses with Bayesian  
966 multigene phylogenies. *Infect. Genet. Evol.* 10: 97–107.
- 967 31. Karlin, D., F. Ferron, B. Canard, and S. Longhi. 2003. Structural disorder and modular  
968 organization in Paramyxovirinae N and P. *J Gen Virol.* 84: 3239–3252.
- 969 32. Gerard, F.C.A., E. de A. Ribeiro, C. Leyrat, I. Ivanov, D. Blondel, S. Longhi, R.W.H.  
970 Ruigrok, and M. Jamin. 2009. Modular Organization of Rabies Virus Phosphoprotein.  
971 *J. Mol. Biol.* 388: 978–996.
- 972 33. Chan, Y.P., C.L. Koh, S.K. Lam, and L.F. Wang. 2004. Mapping of domains  
973 responsible for nucleocapsid protein-phosphoprotein interaction of henipaviruses. *J.*  
974 *Gen. Virol.* 85: 1675–1684.
- 975 34. Bruhn, J.F., A.L. Hotard, C.F. Spiropoulou, M.K. Lo, and E.O. Saphire. 2019. A

- 976 Conserved Basic Patch and Central Kink in the Nipah Virus Phosphoprotein  
977 Multimerization Domain Are Essential for Polymerase Function. *Structure*. 27: 1–9.
- 978 35. Omi-Furutani, M., M. Yoneda, K. Fujita, F. Ikeda, and C. Kai. 2010. Novel  
979 phosphoprotein-interacting region in Nipah virus nucleocapsid protein and its  
980 involvement in viral replication. *J Virol*. 84: 9793–9799.
- 981 36. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa. 2004. Preformed structural  
982 elements feature in partner recognition by intrinsically unstructured proteins. *J Mol*  
983 *Biol*. 338: 1015–1026.
- 984 37. Bruhn, J.F., K.C. Barnett, J. Bibby, J.M.H. Thomas, R.M. Keegan, D.J. Rigden, Z.A.  
985 Bornholdt, and E.O. Saphire. 2014. Crystal Structure of the Nipah Virus  
986 Phosphoprotein Tetramerization Domain. *J. Virol*. 88: 758–762.
- 987 38. Tarbouriech, N., J. Curran, R.W.H. Ruigrok, and W.P. Burmeister. 2000. Tetrameric  
988 coiled coil domain of Sendai virus phosphoprotein. *Nat Struct Biol*. 7: 777–781.
- 989 39. Habchi, J., S. Blangy, L. Mamelli, M.R. Jensen, M. Blackledge, H. Darbon, M.  
990 Oglesbee, Y. Shu, and S. Longhi. 2011. Characterization of the interactions between  
991 the nucleoprotein and the phosphoprotein of Henipavirus. *J Biol Chem*. 286: 13583–  
992 13602.
- 993 40. Audsley, M.D., and G.W. Moseley. 2013. Paramyxovirus evasion of innate immunity:  
994 Diverse strategies for common targets. *World J. Virol*. 2: 57–70.
- 995 41. Kulkarni, S., V. Volchkova, C.F. Basler, P. Palese, V.E. Volchkov, and M.L. Shaw.  
996 2009. Nipah virus edits its P gene at high frequency to express the V and W proteins. *J*  
997 *Virol*. 83: 3982–3987.
- 998 42. Lo, M.K., T.M. Søggaard, and D.G. Karlin. 2014. Evolution and Structural Organization  
999 of the C Proteins of Paramyxovirinae. *PLoS One*. 9: e90003.
- 1000 43. Rodriguez, J.J., J.P. Parisien, and C.M. Horvath. 2002. Nipah virus V protein evades  
1001 alpha and gamma interferons by preventing STAT1 and STAT2 activation and nuclear  
1002 accumulation. *J Virol*. 76: 11476–11483.
- 1003 44. Rodriguez, J.J., C.D. Cruz, and C.M. Horvath. 2004. Identification of the nuclear  
1004 export signal and STAT-binding domains of the Nipah virus V protein reveals  
1005 mechanisms underlying interferon evasion. *J Virol*. 78: 5358–5367.
- 1006 45. Shaw, M.L., A. Garcia-Sastre, P. Palese, and C.F. Basler. 2004. Nipah virus V and W  
1007 proteins have a common STAT1-binding domain yet inhibit STAT1 activation from  
1008 the cytoplasmic and nuclear compartments, respectively. *J Virol*. 78: 5633–5641.
- 1009 46. Hagmaier, K., N. Stock, S. Goodbourn, L.F. Wang, and R. Randall. 2006. A single



1010 amino acid substitution in the V protein of Nipah virus alters its ability to block  
1011 interferon signalling in cells from different species. *J Gen Virol.* 87: 3649–3653.

1012 47. Rodriguez, J.J., L.-F. Wang, and C.M. Horvath. 2003. Hendra Virus V Protein Inhibits  
1013 Interferon Signaling by Preventing STAT1 and STAT2 Nuclear Accumulation. *J.*  
1014 *Virol.* 77: 11842–11845.

1015 48. Smith, K.M., S. Tsimbalyuk, M.R. Edwards, E.M. Cross, J. Batra, T.P. Soares da  
1016 Costa, D. Aragão, C.F. Basler, and J.K. Forwood. 2018. Structural basis for importin  
1017 alpha 3 specificity of W proteins in Hendra and Nipah viruses. *Nat. Commun.* 9: 1–13.

1018 49. Atkinson, S.C., M.D. Audsley, K.G. Lieu, G.A. Marsh, D.R. Thomas, S.M. Heaton, J.J.  
1019 Paxman, K.M. Wagstaff, A.M. Buckle, G.W. Moseley, D.A. Jans, and N.A. Borg.  
1020 2018. Recognition by host nuclear transport proteins drives disorder-to-order transition  
1021 in Hendra virus v. *Sci. Rep.* 8: 1–17.

1022 50. Ribeiro, E.A., A. Favier, F.C.A. Gerard, C. Leyrat, B. Brutscher, D. Blondel, R.W.H.  
1023 Ruigrok, M. Blackledge, and M. Jamin. 2008. Solution Structure of the C-Terminal  
1024 Nucleoprotein-RNA Binding Domain of the Vesicular Stomatitis Virus  
1025 Phosphoprotein. *J. Mol. Biol.* 382: 525–538.

1026 51. Ivanov, I., T. Crepin, M. Jamin, and R.W.H. Ruigrok. 2010. Structure of the  
1027 Dimerization Domain of the Rabies Virus Phosphoprotein. *J. Virol.* 84: 3707–3710.

1028 52. Martinez, N., E.A. Ribeiro Jr., C. Leyrat, N. Tarbouriech, R.W.H. Ruigrok, M. Jamin,  
1029 E. a Ribeiro, C. Leyrat, N. Tarbouriech, R.W.H. Ruigrok, and M. Jamin. 2013.  
1030 Structure of the C-terminal domain of lettuce necrotic yellows virus phosphoprotein. *J*  
1031 *Virol.* 87: 9569–9578.

1032 53. Leyrat, C., M.R. Jensen, E.A. Ribeiro, F.C.A. Gérard, R.W.H. Ruigrok, M. Blackledge,  
1033 and M. Jamin. 2011. The N0-binding region of the vesicular stomatitis virus  
1034 phosphoprotein is globally disordered but contains transient  $\alpha$ -helices. *Protein Sci.* 20:  
1035 542–556.

1036 54. Vinkemeier, U., S.L. Cohen, I. Moarefi, B.T. Chait, J. Kuriyan, and J.E. Darnell. 1996.  
1037 DNA binding of in vitro activated Stat1 alpha, Stat1 beta and truncated Stat1:  
1038 interaction between NH2-terminal domains stabilizes binding of two dimers to tandem  
1039 DNA sites. *EMBO J.* 15: 5616–5626.

1040 55. Wyatt, P.J. 1998. Submicrometer Particle Sizing by Multiangle Light Scattering  
1041 following Fractionation. *J Colloid Interface Sci.* 197: 9–20.

1042 56. Uversky, V.N. 1993. Use of fast protein size-exclusion liquid chromatography to study  
1043 the unfolding of proteins which denature through the molten globule. *Biochemistry.* 32:

1044 13288–13298.

1045 57. Pernot, P., P. Theveneau, T. Giraud, R.N. Fernandes, D. Nurizzo, D. Spruce, J. Surr, S.  
1046 McSweeney, a Round, F. Felisaz, L. Foedinger, a Gobbo, J. Huet, C. Villard, and F.  
1047 Cipriani. 2010. New beamline dedicated to solution scattering from biological  
1048 macromolecules at the ESRF. *J. Phys. Conf. Ser.* 247: 12009.

1049 58. Konarev, P., M. Petoukhov, V. Volchkov, and D.I. Svergun. 2006. ATSAS 2.1, a  
1050 program package for small-angle scattering data analysis. *J. Appl. Cryst.* 39: 277–286.

1051 59. Lescop, E., P. Schanda, and B. Brutscher. 2007. A set of BEST triple-resonance  
1052 experiments for time-optimized protein resonance assignment. *J Magn Reson.* 187:  
1053 163–169.

1054 60. Delaglio, F., S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995.  
1055 NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J*  
1056 *Biomol NMR.* 6: 277–293.

1057 61. Goddard, T.D., and D.G. Kneller. 2003. SPARKY 3. .

1058 62. Jung, Y.S., and M. Zweckstetter. 2004. Mars – robust automatic backbone assignment  
1059 of proteins. *J. Biomol. NMR.* 30: 11–23.

1060 63. Marsh, J.A., V.K. Singh, Z. Jia, and J.D. Forman-Kay. 2006. Sensitivity of secondary  
1061 structure propensities to sequence differences between alpha- and gamma-synuclein:  
1062 implications for fibrillation. *Protein Sci.* 15: 2795–2804.

1063 64. Kabsch, W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr.* 66: 125–132.

1064 65. Pape, T., and T.R. Schneider. 2004. HKL2MAP: a graphical user interface for phasing  
1065 with SHELX programs . *J. Appl. Cryst.* 37: 843–844.

1066 66. Terwilliger, T.C., R.W. Grosse-Kunstleve, P. V Afonine, N.W. Moriarty, P.H. Zwart,  
1067 L.W. Hung, R.J. Read, and P.D. Adams. 2008. Iterative model building, structure  
1068 refinement and density modification with the PHENIX AutoBuild wizard. *Acta*  
1069 *Crystallogr D Biol Crystallogr.* 64: 61–69.

1070 67. Adams, P.D., P. V Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J.  
1071 Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty,  
1072 R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, and P.H.  
1073 Zwart. 2010. PHENIX: a comprehensive Python-based system for macromolecular  
1074 structure solution. *Acta Crystallogr D Biol Crystallogr.* 66: 213–221.

1075 68. Afonine, P. V, R.W. Grosse-Kunstleve, N. Echols, J.J. Headd, N.W. Moriarty, M.  
1076 Mustyakimov, T.C. Terwilliger, A. Urzhumtsev, P.H. Zwart, and P.D. Adams. 2012.  
1077 Towards automated crystallographic structure refinement with phenix.refine. *Acta*

- 1078 Crystallogr D Biol Crystallogr. 68: 352–367.
- 1079 69. Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics.  
1080 Acta Crystallogr D Biol Crystallogr. 60: 2126–2132.
- 1081 70. Chen, V.B., W.B. Arendall 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J.  
1082 Kapral, L.W. Murray, J.S. Richardson, and D.C. Richardson. 2010. MolProbity: all-  
1083 atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol  
1084 Crystallogr. 66: 12–21.
- 1085 71. DeLano, W.L. 2002. The PyMOL Molecular Graphics System. .
- 1086 72. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng,  
1087 and T.E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research  
1088 and analysis. J Comput Chem. 25: 1605–1612.
- 1089 73. Bernado, P., L. Blanchard, P. Timmins, D. Marion, R.W. Ruigrok, and M. Blackledge.  
1090 2005. A structural model for unfolded proteins from residual dipolar couplings and  
1091 small-angle x-ray scattering. Proc Natl Acad Sci U S A. 102: 17002–17007.
- 1092 74. Communie, G., J. Habchi, F. Yabukarski, D. Blocquel, R. Schneider, N. Tarbouriech,  
1093 N. Papageorgiou, R.W.H. Ruigrok, M. Jamin, M.R. Jensen, S. Longhi, and M.  
1094 Blackledge. 2013. Atomic Resolution Description of the Interaction between the  
1095 Nucleoprotein and Phosphoprotein of Hendra Virus. PLoS Pathog. 9: e1003631.
- 1096 75. Sali, A., and T.L. Blundell. 1993. Comparative protein modelling by satisfaction of  
1097 spatial restraints. J. Mol. Biol. 234: 779–815.
- 1098 76. Jensen, M.R., L. Salmon, G. Nodet, and M. Blackledge. 2010. Defining conformational  
1099 ensembles of intrinsically disordered and partially folded proteins directly from  
1100 chemical shifts. J Am Chem Soc. 132: 1270–1272.
- 1101 77. Shen, Y., and A. Bax. 2007. Protein backbone chemical shifts predicted from searching  
1102 a database for torsion angle and sequence homology. J Biomol NMR. 38: 289–302.
- 1103 78. Eyal, E., R. Najmanovich, B.J. McConkey, M. Edelman, and V. Sobolev. 2004.  
1104 Importance of solvent accessibility and contact surfaces in modeling side-chain  
1105 conformations in proteins. J Comput Chem. 25: 712–724.
- 1106 79. Svergun, D., C. Barberato, and M.H. Koch. 1995. CRY SOL - a program to evaluate X-  
1107 ray solution scattering of biological macromolecules from atomic coordinates. J. Appl.  
1108 Cryst. 28: 768–773.
- 1109 80. Nodet, G., L. Salmon, V. Ozenne, S. Meier, M.R. Jensen, and M. Blackledge. 2009.  
1110 Quantitative description of backbone conformational sampling of unfolded proteins at  
1111 amino acid resolution from NMR residual dipolar couplings. J Am Chem Soc. 131:

- 1112 17908–17918.
- 1113 81. Ferron, F., S. Longhi, B. Canard, and D. Karlin. 2006. A practical overview of protein  
1114 disorder prediction methods. *Proteins*. 65: 1–14.
- 1115 82. Meng, F., V.N. Uversky, and L. Kurgan. 2017. Comprehensive review of methods for  
1116 prediction of intrinsic disorder and its molecular functions. *Cell. Mol. Life Sci.* 74:  
1117 3069–3090.
- 1118 83. Ivanov, I., T. Crépin, M. Jamin, and R.W.H. Ruigrok. 2010. Structure of the  
1119 dimerization domain of the rabies virus phosphoprotein. *J. Virol.* 84: 3707–3710.
- 1120 84. Callebaut, I., G. Labesse, P. Durand, A. Poupon, B. Canard, J. Chomilier, B. Henrissat,  
1121 and J.P. Mornon. 1997. Deciphering protein sequence information through  
1122 hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci.*  
1123 53: 621–645.
- 1124 85. Salvamani, S., Z.H. Goh, K.L. Ho, B.T. Tey, and W.S. Tan. 2013. Oligomerization  
1125 state of the multimerization domain of Nipah virus phosphoprotein. *Process Biochem.*  
1126 48: 1476–1480.
- 1127 86. Bizien, T., D. Durand, P. Roblin, A. Thureau, P. Vachette, and J. Pérez. 2016. A Brief  
1128 Survey of State-of-the-Art BioSAXS. *Protein*. 23: 217–231.
- 1129 87. Communie, G., T. Crepin, D. Maurin, M.R. Jensen, M. Blackledge, R.W.H. Ruigrok,  
1130 T. Crépin, D. Maurin, M.R. Jensen, M. Blackledge, and R.W.H. Ruigrok. 2013.  
1131 Structure of the tetramerization domain of measles virus phosphoprotein. *J Virol.* 87:  
1132 7166–7169.
- 1133 88. Fitzkee, N.C., and G.D. Rose. 2004. Reassessing random-coil statistics in unfolded  
1134 proteins. *Proc Natl Acad Sci U S A.* 101: 12497–12502.
- 1135 89. Bernadó, P., and M. Blackledge. 2009. A self-consistent description of the  
1136 conformational behavior of chemically denatured proteins from NMR and small angle  
1137 scattering. *Biophys. J.* 97: 2839–2845.
- 1138 90. Zhang, O., and J.D. Forman-Kay. 1995. Structural characterization of folded and  
1139 unfolded states of an SH3 domain in equilibrium in aqueous buffer. *Biochemistry.* 34:  
1140 6784–6794.
- 1141 91. Johansson, K., J.M. Bourhis, V. Campanacci, C. Cambillau, B. Canard, and S. Longhi.  
1142 2003. Crystal structure of the measles virus phosphoprotein domain responsible for the  
1143 induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem.* 278:  
1144 44567–44573.
- 1145 92. Cox, R., T.J. Green, S. Purushotham, C. Deivanayagam, G.J. Bedwell, P.E. Prevelige,

- 1146 and M. Luo. 2013. Structural and functional characterization of the mumps virus  
1147 phosphoprotein. *J Virol.* 87: 7558–7568.
- 1148 93. Blanchard, L., N. Tarbouriech, M. Blackledge, P. Timmins, W.P. Burmeister, R.W.H.  
1149 Ruigrok, and D. Marion. 2004. Structure and dynamics of the nucleocapsid-binding  
1150 domain of the Sendai virus phosphoprotein in solution. *Virology.* 319: 201–211.
- 1151 94. Bernadó, P., and D.I. Svergun. 2012. Structural analysis of intrinsically disordered  
1152 proteins by small-angle X-ray scattering. *Mol. Biosyst.* 8: 151.
- 1153 95. Brennich, M., P. Pernot, and A. Round. 2017. How to analyze and present SAS data for  
1154 publication. *Adv Exp Med Biol.* 1009: 47–64.
- 1155 96. Chen, X., U. Vinkemeier, Y. Zhao, D. Jeruzalmi, J.E. Darnell, and J. Kuriyan. 1998.  
1156 Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell.* 93:  
1157 827–839.
- 1158 97. Communie, G., T. Crépin, D. Maurin, M.R. Jensen, M. Blackledge, and R.W.H.  
1159 Ruigrok. 2013. Structure of the tetramerization domain of measles virus  
1160 phosphoprotein. *J. Virol.* 87: 7166–7169.
- 1161 98. Blocquel, D., J. Habchi, E. Durand, M. Sevajol, F. Ferron, J. Eroles, N. Papageorgiou,  
1162 and S. Longhi. 2014. Coiled-coil deformations in crystal structures: The measles virus  
1163 phosphoprotein multimerization domain as an illustrative example. *Acta Crystallogr.*  
1164 *Sect. D Biol. Crystallogr.* 70: 1589–1603.
- 1165 99. Beltrandi, M., D. Blocquel, J. Eroles, P. Barbier, A. Cavalli, and S. Longhi. 2015.  
1166 Insights into the coiled-coil organization of the Hendra virus phosphoprotein from  
1167 combined biochemical and SAXS studies. *Virology.* 477: 42–55.
- 1168 100. Thibault, P.A., R.E. Watkinson, A. Moreira-Soto, J.F. Drexler, and B. Lee. 2017.  
1169 Zoonotic Potential of Emerging Paramyxoviruses: Knowns and Unknowns. *Adv. Virus*  
1170 *Res.* 98: 1–55.
- 1171 101. Marsh, G. a., C. de Jong, J. a. Barr, M. Tachedjian, C. Smith, D. Middleton, M. Yu, S.  
1172 Todd, A.J. Foord, V. Haring, J. Payne, R. Robinson, I. Broz, G. Cramer, H.E. Field,  
1173 and L.F. Wang. 2012. Cedar Virus: A Novel Henipavirus Isolated from Australian  
1174 Bats. *PLoS Pathog.* 8: e1002836.
- 1175 102. Habchi, J., L. Mamelli, H. Darbon, and S. Longhi. 2010. Structural disorder within  
1176 Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental  
1177 assessment. *PLoS One.* 5: e11684.
- 1178 103. Damaschun, G., H. Damaschun, K. Gast, R. Misselwitz, D. Zirwer, K.H. Guhrs, M.  
1179 Hartmann, B. Schlott, H. Triebel, and D. Behnke. 1993. Physical and conformational

- 1180 properties of staphylokinase in solution. *Biochim Biophys Acta*. 1161: 244–248.
- 1181 104. Tanford, C. 1961. *Physical chemistry of macromolecules*. New York.
- 1182 105. Gerard, F.C.A., E.D.A. Ribeiro, A.A.V. Albertini, I. Gutsche, G. Zaccari, R.W.H.
- 1183 Ruigrok, and M. Jamin. 2007. Unphosphorylated Rhabdoviridae phosphoproteins form
- 1184 elongated dimers in solution. *Biochemistry*. 46: 10328–10338.
- 1185 106. Wang, L.F., W.P. Michalski, M. Yu, L.I. Pritchard, G. Cramer, B. Shiell, and B.T.
- 1186 Eaton. 1998. A novel P/V/C gene in a new member of the Paramyxoviridae family,
- 1187 which causes lethal infection in humans, horses, and other animals. *J Virol*. 72: 1482–
- 1188 1490.
- 1189 107. Ramachandran, A., J.-P. Parisien, and C.M. Horvath. 2008. STAT2 Is a Primary Target
- 1190 for Measles Virus V Protein-Mediated Alpha/Beta Interferon Signaling Inhibition. *J*
- 1191 *Virol*. 82: 8330–8338.
- 1192 108. Wright, P.E., and H.J. Dyson. 1999. Intrinsically unstructured proteins: re-assessing the
- 1193 protein structure-function paradigm. *J Mol Biol*. 293: 321–331.
- 1194 109. Davey, N.E., G. Travé, and T.J. Gibson. 2011. How viruses hijack cell regulation.
- 1195 *Trends Biochem. Sci*. 36: 159–169.
- 1196 110. Wojciak, J.M., M.A. Martinez-Yamout, H.J. Dyson, and P.E. Wright. 2009. Structural
- 1197 basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation
- 1198 domains. *Embo J*. 28: 948–958.
- 1199

**Table 1. Mass and size of the different phosphoprotein constructs**

Protein	N <sub>aa</sub> <sup>1</sup>	pI <sub>calc</sub> <sup>2</sup>	MM <sub>calc</sub> <sup>3</sup> (kDa)	MM <sub>exp</sub> <sup>4</sup> (kDa)	R <sub>h,folded</sub> <sup>5</sup> (nm)	R <sub>h,unfolded</sub> <sup>5</sup> (nm)	R <sub>h,exp</sub> <sup>6</sup> (nm)	R <sub>g,theo</sub> <sup>7</sup> (Bernado) (nm)	R <sub>g,theo</sub> <sup>8</sup> (Fitskee) (nm)	R <sub>g,exp</sub> <sup>9</sup> (nm)	D <sub>max</sub> <sup>10</sup> (nm)
P <sub>1-100</sub>	108	5.1	12.347	13.5 ± 1.5	1.8	3.1	2.6 ± 0.2	2.9	3.4	3.2 ± 0.1	12
P <sub>92-190</sub>	108	4.9	11.875	12.0 ± 0.5	1.8	3.0	2.6 ± 0.2	2.9	3.4	3.1 ± 0.2	12
P <sub>173-240</sub>	78	6.4	8.691	8.0 ± 2	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P <sub>92-250</sub>	168	4.6	18.583	21.0 ± 2.0	n.d.	n.d.	n.d.	3.7	4.5	4.0 ± 0.3	15
P <sub>223-319</sub>	106	4.5	11.536	12.9 ± 1.5	1.8	3.0	3.1 ± 0.2	2.9	3.4	3.1 ± 0.1	11.5
P <sub>300-401</sub>	111	5.9	12.421	11.5 ± 1.5	n.d.	n.d.	n.d.	2.9	3.5	n.d.	n.d.
P <sub>387-479</sub>	102	7.1	11.151	12.4 ± 1.5	1.8	3.0	2.5 ± 0.2	3.3	3.3	3.2 ± 0.1	12.5
P <sub>471-580</sub>	119	5.2	13.757 (55.028)	58 ± 6	3.1	6.7	4.3 ± 0.2	n.r.	n.r.	n.d.	n.d.
P <sub>655-709</sub>	63	5.3	7.364	8 ± 1	n.d.	n.d.	n.d.	2.2	2.5	1.6 ± 0.1	5.5
P <sub>588-650</sub>	72	5.3	8.289	8 ± 2	n.d.	n.d.	n.d.	2.4	2.7	2.6 ± 0.1	10
P <sub>588-709</sub>	131	4.9	15.041	15.3 ± 1	1.6	2.5	1.9 ± 0.2	n.d.	n.d.	n.d.	n.d.
P <sub>1-580</sub>	588	4.7	64.756 (259.024)	250 ± 25	5.5	14.8	9.2 ± 0.3	n.r.	n.r.	10.8 ± 0.3	41
P <sub>471-709</sub>	248	4.9	28.369 (113.476)	113 ± 10	4.1	9.8	7.8 ± 0.3	n.r.	n.r.	6.3 ± 0.1	22
P <sub>FL</sub>	717	4.7	79.367 (317.468)	306 ± 30	6.0	16.8	9.3 ± 0.2	n.r.	n.r.	10.9 ± 0.1	39

<sup>1</sup> Number of amino acids in the used constructs including C-terminal His tag and linker (LEHHHHHH) and in some cases a N-terminal M.

<sup>2</sup> Isoelectric point of the used constructs calculated using the webserver ProtParam. In parentheses, isoelectric point calculated without the C-terminal His tag.

<sup>3</sup> In parentheses, calculated mass for a tetramer.

<sup>4</sup> Measured by SEC-MALLS.

<sup>5</sup> The theoretical  $R_h$  values are calculated from  $R_h$  values of folded and chemically unfolded standard proteins taken from Uversky et al. (1997) and using the experimental molecular mass determined by SEC-MALLS.

<sup>6</sup> Measured from SEC data by calibrating the system with standard globular proteins.

<sup>7</sup> Calculated using the power law and parameters defined by Bernado et al.

<sup>8</sup> Calculated using the power law and parameters defined by Fitskee et al.

<sup>9</sup> Measured from SAXS data at low  $Q$  using the Guinier approximation.

<sup>10</sup> Determined from SAXS data using the program PRIMUS.

n.d. = not determined

n.r. = not relevant

Gray shaded lines show proteins that were found to be tetrameric in solution.



**Table 2. Data collection and structure statistics for small angle X-ray scattering**

	P <sub>FL</sub>	P <sub>1-580</sub>	P <sub>471-709</sub>
<b>Data collection parameters</b>			
Instrument	ID14 with PILATUS detector (ESRF, Grenoble, France)		
Beam geometry (mm)	0.7x0.7		
Sample distance (m)	2.43		
Wavelength (Å)	0.931		
Q range (nm <sup>-1</sup> )	0.0406-5.0000		
Exposure time	10 x 10 s		
Temperature (K)	277		
Conc. range (mg.mL <sup>-1</sup> )	1.0-1.8	0.5-1.0	3.5-5.8
<b>Structural parameters</b>			
R <sub>g</sub> (nm) (Guinier)	10.9 ± 0.1	10.8 ± 0.3	6.3 ± 0.1
R <sub>g</sub> (nm) (P(r))	11.2	11.3	6.4
D <sub>max</sub> (nm)	39	41	22
<b>Molecular mass determination (kDa)</b>			
From BSA calibration	432	240	164
Calculated from seq.	317	259	113
<b>Software employed</b>			
Primary data	BsxCuBE		
reduction			
Data processing	PRIMUS (ATSAS 3.0.0)		
Ensemble modeling	ASTERIODS		
3D representations	PYMOL (1.8.6)		

**Table 3. Crystallographic data collection and refinement statistics**

Parameter	
Resolution (Å)	48.75 - 3.00 (3.06-3.00)
Space group	P1 2 <sub>1</sub> 1
Unit cell	
a, b, c (Å)	59.5, 85.60, 122.6
α, β, γ (°)	90.0, 94.7, 90.0
Total reflections	195091
Unique reflections <sup>1</sup>	47924
Rfree test set	1257
Multiplicity	4.1 (1.9)
Completeness (%)	99.5 (99.9)
I/σ(I)	5.3 (2.27)
Rwork	0.175 (0.253)
Rfree	0.229 (0.286)
RMSD (bond lengths) (Å)	0.013
RMSD (angles) (°)	1.16
Ramachandran favored (%)	98.04
Ramachandran allowed (%)	1.83
Ramachandran outliers (%)	0.13
Average B factor (Å <sup>2</sup> )	55

Values in parentheses are for the highest- resolution shell.

<sup>1</sup> Friedel's pairs unmerged

Table 4. Isothermal titration calorimetry data.

Ligand	n	$K_d$ ( $\mu\text{M}$ )	$\Delta G_{\text{Bind}}^0$ ( $\text{kcal.mol}^{-1}$ )	$\Delta H_{\text{Bind}}^0$ ( $\text{kcal.mol}^{-1}$ )	$-T\Delta S_{\text{Bind}}^0$ ( $\text{kcal.mol}^{-1}$ )	$\Delta S_{\text{Bind}}^0$ ( $\text{cal.mol}^{-1}.\text{K}^{-1}$ )
wt	1.19	$100 \pm 15$	$-5.4 \pm 0.8$	$-9.5 \pm 1.2$	$4.2 \pm 1.2$	$-14 \pm 4.0$
Y116A	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
H117A/D118 A/H119A	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
G120E	1.23	$70 \pm 9$	$-5.7 \pm 0.7$	$-8.4 \pm 0.6$	$2.7 \pm 0.6$	$-8.9 \pm 2.1$
G125E	1.31	$210 \pm 90$	$-5.0 \pm 2.0$	$-7.5 \pm 2.6$	$2.5 \pm 2.7$	$-8.4 \pm 9.2$

T = 298 K

n.d. : not determined

## LEGENDS FOR THE FIGURES

**FIGURE 1. Modular architecture of Nipah virus phosphoprotein.** (A) Consensus disordered prediction. The D-score was calculated as described in (32). The threshold to distinguish between ordered and disordered region is set at 0.5. The shaded areas indicate the known positions of the folded tetramerization domain and X domain. (B) Modular organization of NiV phosphoprotein. The upper part shows the structural organization of the phosphoprotein encoded in the main open reading frame (0 ORF). Boxes indicate the localization of folded domains, undulated lines the localization of predicted MoRE and lines the localization of intrinsically disordered regions (IDR 1-3). The green dots show two identified phosphorylation sites. In the middle part, the lines show the location of the known functional regions. In the lower part, the blue lines show the position of the C gene and of the regions coding for the specific parts of the V and W proteins in the alternative open reading frames (+1 ORF and +2 ORF). (C) Divide-and-conquer strategy. The lines indicate the position of the different fragments of the phosphoprotein used in this study.

**FIGURE 2. Molecular masses and size of full length phosphoprotein ( $P_{FL}$  in black) and of different fragments ( $P_{1-580}$  in red;  $P_{471-709}$  in green;  $P_{471-580}$  in blue) in solution.** Size exclusion chromatography was performed at 20 °C on a TSK G4000 column equilibrated with 20 mM Tris/HCl pH 7.5 containing 150 mM NaCl using a flow rate of 0.8 mL.min<sup>-1</sup>. The elution was monitored on-line by multi-angle laser light scattering and differential refractometry. The lines show the chromatograms monitored by differential refractive index measurements. The crosses indicate the molecular mass calculated from static light scattering across the elution peak and the numbers indicate the weight-averaged molecular mass (kDa) with standard deviations. The hydrodynamic radius ( $R_h$ ) was determined from the elution volume by calibrating the column with standard globular proteins (**Table 1**).

**FIGURE 3. Crystal structure of NiV phosphoprotein tetramerization domain ( $P_{MD}$ ).** (A) Cartoon representation of tetrameric NiV  $P_{MD}$  (4GJW.pdb). The different chains are shown in different colors and the overall dimensions are indicated (black arrows). The N and C-terminal end are indicated for each chain and the three  $\alpha$ -helices,  $\alpha_1$  to  $\alpha_3$ , are shown for chain D (in red). The position of the stammer in the long coiled-coil is shown and the dotted lines indicate the location of the a-layers displayed in panels D and E. (B) Contacts between

the chains within the crown. Cartoon representation of the NiV tetramerization domain showing that each chain contacts the other three chains. (C) Sequence alignment of the tetramerization domain of members of the genus Henipavirus. (HeV: Hendra virus; CeV: Cedar virus; GBHV: Ghanaian bat henipavirus). The letters under the sequences indicate the position of the different heptad repeats (h<sub>1-8</sub>) and of the stammer. A and D positions in the heptad repeats, which correspond to residues pointing inside the coiled-coil, are shown in green to indicate hydrophobic residues or in blue to indicate polar residues. (D and E) Symmetry of layers in the coiled coil. Stick representations of two A-layer (Cys512 and Leu557) showing the transition from four-fold to two-fold symmetry along the coiled-coil (see positions in panel A). (F) Sequence conservation map. Space filling model of NiV tetramerization domain where surface residues are colored according to the degree of conservation (using the multiple sequence alignment of panel C). Figures were drawn with PYMOL.

**FIGURE 4. NMR characterization of NiV phosphoprotein N-terminal region (P<sub>NTR</sub>), C-terminal linker and X domain.** (A) Secondary structure propensities (SSP) of the N-terminal region of the phosphoprotein obtained on the basis of experimental <sup>13</sup>C $\alpha$  and <sup>13</sup>C $\beta$  chemical shifts. The propensities were obtained by combining spectral assignments of individual sub-constructs of this region. (B) Secondary structure propensities of the C-terminal intrinsically disordered linker region of the phosphoprotein (P<sub>CT-LINK</sub>) obtained from experimental <sup>13</sup>C $\alpha$  and <sup>13</sup>C $\beta$  chemical shifts. (C) Secondary structure propensities of the C-terminal X domain of the phosphoprotein (P<sub>XD</sub>) obtained from experimental <sup>13</sup>C $\alpha$  and <sup>13</sup>C $\beta$  chemical shifts. (D) <sup>1</sup>H-<sup>15</sup>N HSQC spectrum of full-length NiV phosphoprotein (red) superimposed with HSQC spectra (blue) of individual sub-constructs of the protein (P<sub>1-100</sub>, P<sub>92-190</sub>, P<sub>173-240</sub>, P<sub>223-319</sub>, P<sub>300-401</sub>, P<sub>387-479</sub>).

**FIGURE 5. Ensemble structure of NiV phosphoprotein.** (A) Multiple-curve fitting of SAXS data. The upper panel shows experimental SAXS curves of the three tetrameric proteins: P<sub>FL</sub> (in red), P<sub>1-580</sub> (in blue) and P<sub>471-709</sub> (in orange), with experimental standard deviations shown as error bars. The black lines show the multiple-curve fit obtained for the three SAXS curves for a selected ensemble of 10 conformers. The reduced  $\chi^2$  value is 1.53 over all three constructs. The lower panels show the reduced residuals ( $\Delta I / \sigma$ ). (B) R<sub>g</sub> distribution for an ensemble of 10 conformers of full-length phosphoprotein. The thin black

bars show the distribution for the initial ensemble of conformers. The large red bars show the  $R_g$  distribution for selected ensembles of 10 conformers. (C) Single model of NiV phosphoprotein structure. The tetramerization domain and X domain are shown as surface representations. The different chains are shown in different colors and the N-terminal end and X domain ( $P_{XD}$ ) are indicated for each chain. (D) Ensemble representations of full-length NiV phosphoprotein ( $P_{FL}$ ) and of the two tetrameric fragments:  $P_{1-580}$  and  $P_{471-709}$ . The 10 different models are aligned by superimposing their tetramerization domain. The tetramerization domain and X domain are shown as surface representations and the different chains are shown in same colors as in panel C. The models of the  $P_{471-709}$  fragment are shown in two orientations.

**FIGURE 6. Interaction of the N-terminal region of NiV phosphoprotein (V and W proteins) with STAT1.** (A) Sequence conservation between residues 85 and 140 of the phosphoprotein (V and W proteins) across different Henipaviruses (see Figure 3C). The labels above the alignment indicate the mutants used in the ITC experiments (in blue) and in the generation and assays of recombinant viruses (in green). (B) Schematic representation of STAT1 architecture. The color boxes show the different domains present in our construct, which encompasses residues 135 to 683 of STAT1, whereas the gray boxes indicate the missing part. (C) SEC-MALLS analysis. The lines show the chromatograms monitored by refractometry and the red crosses show the molecular mass calculated from on-line static light scattering measurements (MALLS). The upper panel shows the results for  $STAT1_{132-683}$ , the middle panel for the fragment  $P_{92-190}$ , and the lower panel for the mixture of the two protein fragments in a 10:1 ratio after 2 hour incubation at 20°C. (D) Overlay of a region of the 2D  $^1H$ - $^{15}N$  HSQC NMR spectra of  $P_{92-190}$  (~540  $\mu$ M) in the absence (in red) and presence of  $STAT1_{132-683}$  (~54  $\mu$ M) (in blue). (E) Intensity profile calculated as the ratio of peak intensities in the  $^1H$ - $^{15}N$  HSQC spectrum of  $P_{92-190}$  in isolation ( $I_0$ ) and in the presence of  $STAT1_{132-683}$  ( $I$ ).

**Figure 7. Effect of mutations on the interaction between the phosphoprotein and STAT1.** Calorimetric titration of  $STAT1_{132-683}$  with  $P_{92-190}$  W.T. (A), G120E (B), G125E (C), Y116A (D) and the triple mutant (H117A/Q118A/H119A) (E). The upper panels show the heat effects associated with the injection of  $P_{92-190}$  (2.5  $\mu$ L per injection of about 2 mM solution) into the calorimetric cell (200  $\mu$ L) containing  $STAT1_{132-683}$  at a concentration of 100  $\mu$ M. All experiments were performed in 20 mM phosphate buffer (pH 7.5) containing 150

mM NaCl at 25 °C. The lower panels show the binding isotherm corresponding to the data in the upper panels and the lines show the best fitted curves for a 1:1 binding model. (F) Average dissociation constant obtained from three experiments for WT P<sub>92-190</sub> and the different variants. AAA = H117A/Q118A/H119A. (G) Impact of NiVs infections (NiV-WT, NiV-G121E, NiV-G135E) on the localization of phosphorylated endogenous STAT1. Vero E6 cells were infected with the indicated viruses and treated with interferon- $\beta$  at 28 h post-infection. The phospho-Y701 form of endogenous STAT1 was visualized in paraformaldehyde fixed cells 40 min after interferon- $\beta$  addition (phospho-STAT1, green staining). NiV matrix (M) protein expression is a marker for infection (red). Arrowheads point to representative non-infected cells and nuclei stained for phospho-STAT1. Arrows indicate cells infected with NiV-G121E mutant defective in STAT1 binding, and again nuclei stained for phospho-STAT1.















