



HAL
open science

Characterization of the Interaction Domains between the Phosphoprotein and the Nucleoprotein of Human Metapneumovirus

Hortense Decool, Benjamin Bardiaux, Luis Checa Ruano, Olivier Sperandio, Jenna Fix, Irina Gutsche, Charles-Adrien Richard, Monika Bajorek, Jean-François Eléouët, Marie Galloux

► **To cite this version:**

Hortense Decool, Benjamin Bardiaux, Luis Checa Ruano, Olivier Sperandio, Jenna Fix, et al.. Characterization of the Interaction Domains between the Phosphoprotein and the Nucleoprotein of Human Metapneumovirus. *Journal of Virology*, 2022, 96 (2), pp.e0090921. 10.1128/JVI.00909-21 . hal-03579501v2

HAL Id: hal-03579501

<https://hal.science/hal-03579501v2>

Submitted on 16 Nov 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.

1 **Characterization of the interaction domains between the phosphoprotein and the nucleoprotein**
2 **of human Metapneumovirus**

3
4 **Running title: Metapneumovirus N-P interaction**

5
6 Hortense Decool¹, Benjamin Bardiaux², Luis Checa Ruano², Olivier Sperandio^{2,3}, Jenna Fix¹, Irina
7 Gutsche⁴, Charles-Adrien Richard¹, Monika Bajorek¹, Jean-François Eléouët^{1#}, Marie Galloux^{1#}.

8
9 ¹Université Paris-Saclay, INRAE, UVSQ, VIM, 78350, Jouy-en-Josas, France.

10 ² Institut Pasteur, Université de Paris, CNRS UMR3528, Unité de Bioinformatique Structurale, 28 rue
11 du Dr Roux, 75015, Paris, France.

12 ³ Inserm U973 MTi, 25 rue Hélène Brion 75013, Paris, France.

13 ⁴ University of Grenoble Alpes, CNRS, CEA, CNRS, IBS, Grenoble, France.

14
15 #Correspondance: jean-francois.eleouet@inrae.fr; marie.galloux@inrae.fr

16
17 **Keywords** : HMPV, phosphoprotein, nucleoprotein, protein-protein interaction, structural modeling

18

19

20 **ABSTRACT**

21 Human metapneumovirus (HMPV) causes severe respiratory diseases in young children. The HMPV
22 RNA genome is encapsidated by the viral nucleoprotein (N), forming an RNA-N complex (N^{Nuc}), which
23 serves as template for genome replication and mRNA transcription by the RNA-dependent RNA
24 polymerase (RdRp). The RdRp is formed by the association of the large polymerase subunit (L), which
25 has RNA polymerase, capping and methyltransferase activities, and the tetrameric phosphoprotein
26 (P). P plays a central role in the RdRp complex by binding to N^{Nuc} and L, allowing the attachment of
27 the L polymerase to the N^{Nuc} template. During infection these proteins concentrate in cytoplasmic
28 inclusion bodies (IBs) where viral RNA synthesis occurs. By analogy to the closely related
29 pneumovirus respiratory syncytial virus (RSV), it is likely that the formation of IBs depends on the
30 interaction between HMPV P and N^{Nuc} , which has not been demonstrated yet. Here, we finely
31 characterized the binding P- N^{Nuc} interaction domains by using recombinant proteins, combined with a
32 functional assay for the polymerase complex activity, and the study of the recruitment of these
33 proteins to IBs by immunofluorescence. We show that the last 6 C-terminal residues of HMPV P are
34 necessary and sufficient for binding to N^{Nuc} , that P binds to the N-terminal domain of N (N_{NTD}), and
35 identified conserved N residues critical for the interaction. Our results allowed to propose a structural
36 model for the HMPV P- N^{Nuc} interaction.

37

38 **IMPORTANCE**

39 Human metapneumovirus (HMPV) is a leading cause of severe respiratory infections in children but
40 also affects human populations of all ages worldwide. Nowadays, no vaccine or efficient antiviral
41 treatments are available for this pneumovirus. A better understanding of the molecular mechanisms
42 involved in viral replication could help the design or discovery of specific antiviral compounds. In this
43 work we have investigated the interaction between two major viral proteins involved in HMPV RNA
44 synthesis, the N and P proteins. We finely characterized their domains of interaction, and identified a
45 pocket on the surface of the N protein, a potential target of choice for the design of compounds
46 interfering with N-P complexes and inhibiting viral replication.

47

48 **INTRODUCTION**

49 Pneumonia is the leading cause of death among children younger than 5 years worldwide, and severe
50 pneumonia is more frequently caused by viruses than bacteria (1). After the closely related respiratory
51 syncytial virus (RSV), human metapneumovirus (HMPV) is recognized to be one of the most important
52 cause of viral bronchiolitis and pneumonia in young children, causing 7 to 19% of all cases of acute
53 respiratory tract infections (1, 2). HMPV infects mainly newborn, children, elderly and
54 immunocompromised individuals worldwide. This virus was first reported in 2001 from Dutch children
55 with acute lower respiratory tract illness, and serological studies have revealed that virtually every
56 child has been exposed to HMPV by the age of 5 years (3). The clinical features of HMPV infection
57 display as mid-to-upper respiratory tract infection, and can be severe enough to cause life-threatening
58 bronchiolitis and pneumonia. As yet, there is no effective treatment or licensed vaccine for HMPV.
59 Together with respiratory syncytial virus (RSV), HMPV belongs to the *Pneumoviridae* family in the
60 order *Mononegavirales* (4). HMPV is an enveloped virus that forms pleomorphic or filamentous
61 virions. The virus genome is composed of a negative-sense single-stranded RNA of approximately
62 13.3 kb in size, which encodes eight genes in the following order: 3'-N-P-M-F-M2-SH-G-L-5' (5, 6).
63 The M2 gene of HMPV contains two overlapping open reading frames (ORFs), encoding for M2-1 and
64 M2-2 proteins which precise functions during HMPV viral cycle remain unclear. The HMPV genome is
65 encapsidated by multiple copies of the nucleoprotein (N) forming helical nucleocapsids (N^{Nuc}). This
66 N^{Nuc} serves as template for genome replication and mRNA transcription by the viral polymerase
67 complex formed by the large polymerase subunit (L) and its main cofactor the phosphoprotein (P).
68 After virus binding to the cell surface and virus-cell membrane fusion mediated by the fusion protein F,
69 nucleocapsids are released into the cytoplasm. Replication and transcription of the viral genome take
70 place within viro-induced cytoplasmic inclusions named inclusion bodies (IBs) (7). These structures
71 can be observed upon expression of P and N proteins alone (8), and it was recently shown for RSV
72 that the interaction between P and N^{Nuc} is critical for the formation of IBs (9).
73 Among the components of the polymerase complex, P plays a pivotal role as a cofactor of the L
74 polymerase, but also as a molecular hub between viral partners. HMPV P, 294 amino acid residues in
75 length, forms homo-tetramers. The atomic structure of the coiled-coil oligomerization domain (residues
76 171-194) was resolved by crystallography (10). Small angle X-ray scattering (SAXS) studies indicated
77 that the flanking N-terminal (residues 1-170) and C-terminal (residues 195-294) regions (named PNT

78 and PCT respectively) are intrinsically disordered, some of them, such as residues 195-237, having α -
79 helical propensity (10, 11). More recently, the structure of the LP complex was resolved by cryo-EM
80 (12). It revealed a tentacular arrangement of P, with each of the four protomers adopting a distinct
81 conformation, demonstrating a “folding-upon-partner-binding” mechanism. Depending on the
82 protomer, the L-binding region involved regions encompassing residues 171-236, 172-217, 170-231 or
83 169-264. On the other hand, by binding to N, P mediates the attachment of the L protein to the N^{Nuc}
84 template for viral RNA synthesis.

85 By analogy to RSV (13), the most C-terminal region of HMPV P is also expected to be the N^{Nuc} binding
86 domain, but this has not been shown yet. It is noteworthy that the C-terminal extremity of P supposed
87 to bind to N^{Nuc} was not visible in the cryo-EM structure of the LP complex, indicating that this region is
88 disordered in the absence of N^{Nuc}. Furthermore, the encapsidation of neosynthesized genome or
89 antigenome necessitates a pool of monomeric, RNA-free N, termed N⁰, which is kept in an
90 unassembled state through an interaction with P, which plays the role of molecular chaperone, until
91 delivery to the sites of viral RNA synthesis. The crystal structures of recombinant HMPV N protein
92 (394 residues) expressed in *E. coli* and purified either as an RNA-free monomeric N in complex with
93 the N-terminal residues of P, or as rings of 10 N protomers complexed to RNA were resolved (14). In
94 both states, the structures show that N presents two globular domains (N_{NTD} and N_{CTD}) separated by a
95 flexible linker that forms the RNA groove, and N- and C-arms. In the oligomeric state, the N- and C-
96 arms play a key role in the interaction between N protomers and oligomerization, the N-arm binding to
97 the flank of the N_{*i-1*} protomer and the C-arm binding atop the N_{*i+1*} protomer (*i* corresponding to the
98 middle subunit of three adjacent N protomers). In the N⁰ state, the important conformational changes
99 consist in packing of the C-arm of N in the RNA groove, impairing RNA binding.

100 In this work, based on the structural data of HMPV P and N proteins and their strong homologies with
101 RSV N and P proteins, we finely characterized the binding domains involved in HMPV P-N interaction.
102 By combining biochemical and functional cellular assays, coupled with a rational mutational approach,
103 we identified residues of P and N critical for this interaction. Our data show that the last C-terminal
104 residues of P bind to the N_{NTD}. These results allowed to establish a structural model of the interaction
105 that could be used for the rational design of antivirals targeting the N^{Nuc}-P interaction of HMPV.

106

107

108 **RESULTS**

109

110 **The HMPV PCT domain allows to purify N-RNA rings expressed in bacteria**

111 HMPV and RSV P proteins present a conserved structural organization, with a central tetramerization
112 domain flanked by N- and C-terminal highly disordered regions (Fig. 1A). Co-expression of N with the
113 C-terminal residues 161-241 of RSV P fused to GST in bacteria allowed the purification of ring shape
114 N-RNA oligomers, showing that the C-terminal disordered region of RSV P (PCT) is involved in N^{Nuc}
115 binding (13, 15). Similarly, when expressed alone in *E. coli*, HMPV N also forms decameric rings
116 containing RNA (14). In a first attempt to characterize the N binding site of HMPV P, we thus co-
117 expressed recombinant HMPV GST-PCT (residues 200-294) and N proteins (from CAN 97-83 strain)
118 in *E. coli*. The purified complexes were analyzed by SDS-PAGE stained with Coomassie blue. As
119 shown on figure.1B, N was co-purified with GST-PCT to > 95% homogeneity. The PCT was then
120 separated from GST by thrombin cleavage (Fig. 1B) and the solubilized complex was analyzed by size
121 exclusion chromatography, following optical density (OD) at 220, 260 and 280 nm. The elution profile
122 showed a major peak (P1) with a OD_{260nm}/OD_{280nm} ratio > 1 and an apparent mass of ~ 500 kDa, in
123 agreement with the expected size of N-RNA decamers (Fig. 1C). A second peak of lower intensity
124 (P2), with an apparent mass of ~ 10 kDa was detected only at 220 nm. This peak should correspond
125 to PCT, which does not present aromatic residues and a higher apparent molecular weight than
126 predicted, likely due to the elongated shape of this fragment. The fractions of P1 peak were pooled,
127 and sample analysis by SDS-PAGE stained with Coomassie blue showed the presence of a unique
128 band corresponding to N (Fig. 1C). The separation of PCT and N upon gel filtration reveals the
129 relatively low affinity of monomeric PCT for N. We then further analyzed the sample collected from the
130 P1 peak by combining dynamic light scattering (DLS) and electron microscopy (EM) approaches. The
131 DLS measurement profile confirmed the homogeneity of the sample, with a single peak at 18 nm (Fig
132 1D), while the observation of purified N by EM confirmed the presence N-rings (Fig 1E). These N-RNA
133 rings are thus similar to those purified by Renner *et al.*, who determined the 3D structure of N-RNA
134 rings with a diameter close to 16 nm (14).

135 Altogether, these results show that the HMPV GST-PCT fusion protein is sufficient to interact with
136 nucleocapsid-like N-RNA rings.

137

138

139 **The last 6 residues of HMPV P constitute the minimum domain for N^{Nuc} binding**

140 We then investigated the minimal domain of HMPV P involved in N-RNA rings binding. The 9 C-
141 terminal residues of RSV P were previously shown to be sufficient to interact with N (13). Sequence
142 alignment of the C-terminal extremity of pneumoviruses P proteins reveals the presence of two distinct
143 motifs, separated by a highly acidic stretch in the case of Metapneumovirus P proteins (Fig. 2A). Of
144 particular interest, the last 9 C-terminal residues of P proteins form a partially conserved motif. To
145 investigate the role of these C-terminal residues of HMPV P in N binding, a series of GST fused
146 peptides of 9 to 1 amino acid long derived from the C-terminal of P was generated. These constructs
147 were co-expressed with N in *E.coli*, and their capacity to purify N was analyzed by SDS-PAGE and
148 Coomassie blue staining. As shown on figure. 2B, the minimal sequence required for N purification
149 corresponds to the 6 last C-terminal residues of P. In parallel, alanine scanning was performed with
150 the GST-P[285-294] construct to characterize the residues of P involved in the interaction with N-RNA
151 rings. Again, GST-P constructs and N were co-expressed in bacteria, followed by purification by
152 Glutathione-Sepharose beads affinity, and co-purification of N was analyzed by SDS-PAGE and
153 Coomassie blue staining. Only the four P substitutions I289A, Y290A, L292A and M294A abrogated
154 the interaction with N (Fig. 2C). In order to further investigate the potential role of P residues Q291 and
155 I293 in the interaction with N, a double substitution Q291A/I293A was inserted in GST-PCT. These
156 two mutations did not affect N binding, confirming that these residues may not be directly involved in
157 the interaction (Fig. 2D). These results confirm that the last residues of P are directly involved in the
158 interaction with N, and suggest that HMPV P-N interaction mainly involves hydrophobic interactions.
159 Of particular interest, the last 7 C-terminal residues of the P proteins of HMPV strains NL 99/1 (B1
160 subgroup), CAN 98-75 (B2 subgroup) and CAN 97-83 (A2 subgroup) are conserved, thus confirming
161 their critical role. As both hydrophobic and acidic residues of RSV P were previously shown to be
162 critical for the interaction with N (Fig. 2E), these results suggest that the binding of HMPV P on N
163 differs from RSV and involves specific interactions. We thus tested the capacity of each PCT
164 fragments to pulldown heterologous N proteins. Figure 2F shows that HMPV and RSV GST-PCT
165 constructs did not allow to purify heterologous N. This last result confirms that HMPV and RSV N and
166 P proteins cannot cross-interact, and that pneumoviruses P-N interactions are specific to each genus.
167

168

169 **Search for the P binding site on the surface of HMPV N**

170 Although HMPV and RSV P-N interactions have their own specificity and cannot cross-react, the
171 strong homologies between N proteins suggest that the PCT binding site on N protein could be
172 partially conserved. For RSV, the P binding site on N^{Nuc} is located at the surface of the helical
173 nucleocapsid, on the N_{NTD} (16). The residues of RSV N_{NTD} involved in the interaction with P, which are
174 partially conserved between HMPV and RSV (Fig. 3A), were shown to form a well-defined pocket
175 constituted of hydrophobic and positively charged residues (16, 17). Corresponding residues of HMPV
176 N_{NTD} also form a pocket exposed at the surface of N-RNA rings (Fig. 3B). In a first attempt to
177 characterize the HMPV P binding domain on N, we thus tested the capacity of P to interact with the
178 N_{NTD}. The N_{NTD} was expressed in *E. coli* and purified using a 6xHis tag, showing that this domain is
179 soluble when expressed alone (Fig. 4A). In parallel, the GST-P or GST-P_{ΔM294} (deletion of the last C-
180 terminal residue of P) proteins were expressed in bacteria, either alone and purified using the GST
181 tag, or co-expressed with N_{NTD} followed by purification by 6xHis tag. Analysis of the samples by SDS-
182 PAGE showed that N_{NTD} allowed to co-purify the wild type P, and that deletion of the last residue of P
183 was sufficient to impair the interaction (Fig. 4A). Of note, the presence of numerous bands was
184 observed in the samples containing GST-P alone or incubated in the presence of N_{NTD}. These bands
185 migrated with a lower apparent mass weight compared to GST-P protein, suggesting a high propensity
186 of recombinant GST-P to be cleaved and degraded. The HMPV N_{NTD} residues identified by analogy
187 with RSV, i.e. residues L46, L47, E50, D128, R132, M135, R151, P152, and S153 (Fig. 4B), were then
188 substituted by alanine. We first assessed the impact of these mutations on the expression and
189 solubility of N proteins fused to a C-terminal poly-His tag expressed in *E. coli*. Analysis of purified N
190 proteins by SDS -PAGE showed that none of these mutations induced a defect of production or
191 solubility of N (Fig. 4C). The capacity of GST-P to co-purify mutated N proteins was then tested.
192 Whereas substitutions L46A, L47A, and P152A did not disrupt N-P interaction P, E50A, D128A,
193 R132A, M135A, R151A and S153A substitutions strongly affected or abrogated N-P interaction (Fig.
194 4D, upper panel). Of note, expression of N mutants in bacteria lysates was assessed by western blot
195 and confirmed that all the N were expressed (Fig 4D, lower panel). These results thus revealed a
196 critical role of hydrophobic residues, but also of negatively and positively charged residues of N_{NTD} in

197 P binding. Overall, these data confirm that the P binding site on N presents strong homologies
198 between HMPV and RSV.

199

200 **Validation of P-N^{Nuc} binding domains in eukaryotic cells**

201 As P-N^{Nuc} interaction is required for viral polymerase activity, we then studied the impact of mutations
202 on P and N on HMPV replication/transcription using a bicistronic minigenome, pGaussia/Firefly. This
203 construct contains the Gaussia and Firefly luciferase genes under the control of HMPV gene start and
204 gene stop sequences, as well as Leader and Trailer sequences of HMPV genome (Fig. S1). Briefly,
205 BSRT7 cells were transfected with plasmids coding for N (WT or mutants), P (WT or mutants), L, and
206 M2-1 proteins, the plasmid pGaussia/Firefly minigenome, and pSV- β -Gal (used to normalize
207 transfection efficiencies). In this system, the expression of luciferase depends on the HMPV
208 polymerase complex activity which can thus be quantified by luminescence measurement. As shown
209 on figure 5A, mutations I289A, Y290A, L292A and M294A of P abrogated the polymerase activity, and
210 only a residual activity was detected for the mutant Q291A. On the contrary, the P mutation I293A
211 induced only a slight decrease of the polymerase activity. On the other hand, nearly all the mutations
212 of on N induced a decrease of $\geq 50\%$ of the polymerase activity compared to wild type N, with only
213 mutations D128A and R132A totally impairing the activity (Fig 5A, right panel). For all these P and N
214 mutants, the levels of protein expression were similar to those of wild type proteins as assessed by
215 western blot (Fig.5B).

216 We then assessed if the impact of N and P mutations on polymerase activity could correlate with a
217 defect in IBs morphogenesis. Knowing that co-expression of N and P is sufficient to induce the
218 formation of pseudo-IBs (8), we used an approach based on fluorescent proteins expression. First, in
219 order to investigate the impact of P mutations, cells were co-transfected with plasmids encoding P
220 (WT or mutants) together with plasmids encoding N and GFP-N (at a ratio 2:1). A similar strategy used
221 for RSV showed that, although GFP-N fusion protein had a dominant negative effect on RSV RNA
222 synthesis, co-expression of GFP-N with wild type N partially restored polymerase activity (18). In our
223 HMPV minigenome assay, co-expression of N with GFP-N at a ratio 2:1 allowed to rescue almost 70%
224 of the polymerase activity (Fig. 5C). Cells transfected with P and N: GFP-N were fixed 24 h post-
225 transfection and labelled with anti-P antibody. As shown on figure 6, although P labelling revealed a
226 diffuse cytoplasmic localization of the protein, wild type P was found to co-localize with GFP-N into

227 cytoplasmic inclusions similar to IBs observed upon HMPV infection (7, 8). No impact on IBs
228 morphology was detected upon alanine substitution of P residue I293. On the contrary, Q291A
229 substitution seemed to induce a defect in IBs formation compared to wild type condition, as smaller
230 IBs were observed, and a total loss of IBs formation was observed when mutating residues I289,
231 Y290, L292 and M294 of P (Fig. 6B and Table 1).

232 In parallel, the impact of N mutations on IBs morphogenesis was investigated by co-transfecting N
233 (WT or mutants) with P carrying BFP inserted between residues S61 and S62 (NL99-1 strain), in a
234 region poorly conserved among HMPV strains and predicted as naturally intrinsically disordered. This
235 P-BFP construct was functional since 65% of the HMPV polymerase activity was recovered when
236 tested in our minigenome assay (Fig. 5D). Transfected cells were fixed before immunolabelling with
237 anti-N antibody. Pseudo IBs were detected upon co-expression of P-BFP and WT N, but also upon
238 alanine substitution of most of the residues (i.e. L46, L47, E50, M135, P252, and S153) (Fig. 7 and
239 Table 1). As previously observed for the mutant Q291A of P, the mutation R151A of N seemed to
240 induce a defect of IBs formation compared to the wild type condition, and no IBs were observed when
241 expressing N mutants D128A and R132A (Fig. 7 and Table 1). Altogether, these results reveal a good
242 correlation between the polymerase functioning and the capacity of N and P to interact and form IBs.
243 Besides, these data confirm that the residues of P and N previously identified as critical for N^{Nuc}-P
244 interaction *in vitro* are also critical for polymerase activity and the formation of IBs and in cells.

245

246 **Molecular modeling of PCT-N_{NTD} interaction**

247 To gain insight into the possible binding mode of P with N_{NTD}, we built models of the last 6 residues of
248 P (peptide I289-YQLI-M294 named P6) bound to N, by analogy with the PCT-N_{NTD} complex of RSV.
249 For RSV, only the last 2 residues of P (D240-F241) were resolved in the PCT-N_{NTD} X-ray structures
250 (16), and are bound to a hydrophobic pocket of N similar to the one we characterized at the surface of
251 HMPV N. Using comparative modeling and refinement with HADDOCK (19), we obtained three
252 clusters of HMPV P6/N_{NTD} complex structures representing possible binding modes of P6 (Table S1).
253 The first two clusters in terms of HADDOCK-score were the most populated and displayed similar
254 Root-mean-square deviation (RMSD) among their members (considering only the last three residues
255 of P, i.e. residues L292-I293-M294, called P3). In those two clusters, the C-terminal carboxyl group of
256 M294 is involved in an ionic interaction with the guanidinium of either R131 or R152 side-chains of N.

257 Additionally, only L292 and M294 of P3 consistently display hydrophobic contacts with N. This result is
258 consistent with our mutation data showing that I293 of P is not required for binding to N, contrary to
259 L292 and M294. The 10 best-scoring structures from cluster 1 and 2 are shown in [figure 9](#),
260 superimposed with the RSV P2/N_{NTD} structure (residues D240-F241 of P). Owing to the best
261 HADDOCK-score and non-bonded energies of cluster 2, we considered this cluster as most
262 representative for the potential binding mode of HMPV P3 to N_{NTD}. Interestingly, more than half of the
263 structures from this cluster display a hydrogen-bond between the backbone oxygen of I293 of P3 with
264 the side-chain of R151 of N, while the side-chain of R132 of N forms a salt-bridge with the C-terminal
265 carboxyl group of M294 of P. Thus, the interactions determined by molecular simulation correlate with
266 the experimental data showing the importance of R132 and R151 of N and of M294 of P for PCT-N
267 interaction.

268 The interactability propensity of each modeled N_{NTD} conformation was further profiled with our in-
269 house tool InDeep (20) in order to select a subset of N_{NTD} models being more likely to interact with the
270 P3 peptide. On this subset of N_{NTD} conformations, the hydrophobic channel of InDeep was also used
271 to locate regions where hydrophobic moieties of P3 are expected to interact at the N_{NTD} surface. Three
272 hydrophobic patches are detected on the N_{NTD} surface, that present a good match with hydrophobic
273 moieties of P3 structures belonging to cluster 2 of HADDOCK: i) one occupied by L292 of P3; ii)
274 another occupied by M294 of P3; and iii) the last one located in an unoccupied region of the N pocket
275 ([Fig.8A](#)). Based on experimental and modeling information, the structure shown in [figure 8](#) represents
276 a consistent model of P3/N_{NTD} complex. The N_{NTD} conformation of this structure is predicted to be
277 favorable to protein-protein interaction by InDeep, and the P3 conformation, belonging to cluster 2 of
278 HADDOCK, binds to N_{NTD} with a combined effect of electrostatic/polar interactions and hydrophobic
279 contacts in agreement with the mutational experiments. In addition, a procedure of virtual alanine
280 scanning using the same InDeep approach confirmed some of the mutagenesis results, such as the
281 importance of R132, M135, and R151 of N for the interaction with P. It also suggests that residue L139
282 of N may be key for the N_{NTD} patch for partners' binding.

283 These analyses reveal that despite the different physico-chemical nature of residues involved in the
284 interaction, PCT binding at the N surface is structurally similar between HMPV and RSV.

285

286 **DISCUSSION**

287 Beyond its interest in better understanding the molecular mechanisms of virus replication, the study of
288 protein-protein interactions required for Pneumoviruses' polymerase complex activity also presents
289 interest for the development of specific antiviral compounds. The activity of this complex depends on
290 multiple highly specific protein-protein interactions that have no cellular counterparts. Among those,
291 N^{Nuc}-P interaction which is critical for viral polymerase functioning constitutes a singular potential
292 target for antivirals. We have previously described RSV N^{Nuc}-P interaction at the molecular level,
293 showing that the C-terminus of P binds to a well-defined pocket at the surface of N (9, 13, 17). We
294 also made the proof of concept that small chemical compounds can target this pocket of N and inhibit
295 viral replication in cells (16). More recently, Sá et al. showed that flavonoids derived molecules could
296 also specifically bind to N pocket and block its interaction with P (21). Given the strong homologies
297 between RSV and HMPV N and P proteins, it could be expected that a similar strategy could be used
298 to block HMPV replication. It is noteworthy that ideally, the discovery of compounds that could inhibit
299 replication of these two closely related viruses at the same time would be much more advantageous.

300 In the present study, we investigated the domains involved in HMPV P-N^{Nuc} interaction using pulldown
301 assays between recombinant proteins (either truncated or mutated proteins) expressed in *E. coli*. As
302 for RSV, the HMPV N protein was purified as nanorings formed by N and bacterial RNA, which can be
303 considered as N^{Nuc}-like complexes (Fig 1). Our results revealed that the mechanisms of binding
304 between P and N-RNA rings are similar between RSV and HMPV, the C-terminus of P interacting with
305 a pocket located on the surface of the N N-terminal globular domain N_{NTD}. However, it appeared that P
306 and N amino acid residues involved in these interactions have different contributions/roles between
307 RSV and HMPV, and that the P-N^{Nuc} interaction is specific of each virus. Our results correlate with
308 these data showing that pneumoviruses' polymerase complexes activity necessitates genus-specific
309 combination of L, P, N, and M2-1 (22, 23). More specifically, we showed that the minimal domain of
310 HMPV P required for N binding is constituted by the last 6 C-terminal residues of P which are mostly
311 hydrophobic residues (IYQLIM), and determined that only Q291 and I293 are not critical for the
312 interaction *in vitro* (Fig. 2). Analysis of the impact of mutations on polymerase activity and IBs'
313 formation confirmed these results, but also revealed that residue Q291 of P is critical for the
314 polymerase activity. This P fragment is thus shorter than for RSV for which N binding minimal domain
315 involves the 9 C-terminal residues of P and is composed of both hydrophobic and acidic residues (13,
316 17). Interestingly, sequence analysis of the C-terminal residues of Orthopneumoviruses and

317 Metapneumoviruses P proteins highlights the conservation of residues critical for N-P interaction (Sup
318 10A), confirming a strong conservation of the interaction mode for each virus. We also identified 6
319 residues of N involved in P binding, *i.e.* residues E50, D128, R132, M135A, R151A and S153A, which
320 form a well-defined pocket at the N_{NTD} surface (Fig 4). Using combined approaches, we revealed a
321 correlation between the impact of the mutations on N-P interaction, and the level of polymerase
322 activity (Table 1). However, only mutations of residues D128 and R132 abrogated IBs' formation and
323 polymerase functioning. Although residues of N proteins involved in P binding are poorly conserved
324 between RSV and HMPV, sequence analysis of N proteins of Orthopneumoviruses and
325 Metapneumoviruses shows that 3 essential residues for N-P interaction present the same physico-
326 chemical properties (132R/R, 135Y/M and 151R/R, Fig 10B). Of most interest, residues R132 and
327 R150 of RSV N were also shown to play a key role in P-N^{Nuc} interaction, the aliphatic part of the R132
328 side chain being involved in the stacking of the aromatic ring of the residue F241 of P. In order to
329 better understand the binding mode of HMPV P to N, molecular modeling was performed and the best
330 structural models support that R132 and R151 of N and of M294 of P are critical for PCT-N interaction
331 (Fig. 7). Nevertheless, molecular modeling was mainly targeted towards the binding of the last 3 C-
332 terminal residues of P, and did not allow to reliably predict the interaction of all the 6 critical residues of
333 P on N. As also observed in the crystal structures of RSV PCT/N_{NTD} complexes (16), the last C-
334 terminal residue of P seems to drive the binding to a well-defined N_{NTD} pocket and this primary binding
335 could allow transient contacts with upstream P residues outside of the N pocket. Remarkably, the
336 acidic stretch in RSV PCT upstream of the main N binding domain was shown to contribute also to N
337 binding (13, 24). Compared to RSV P, HMPV PCT is characterized by the presence of an 18 residues
338 long sequence upstream the last 10 C-terminal residues, mainly composed of acidic residues (Fig.
339 2A), which could, as for RSV, contribute to HMPV P-N interaction. This sequence, located between L
340 and N binding sites of P could also be involved in the modulation of the interaction of P with L and N.
341 In conclusion, our data confirm the strong structural homologies between HMPV and RSV P-N
342 complexes but also highlight some singularities. Altogether, our study suggests that the HMPV P
343 binding pocket at the N_{NTD} surface could represent a new target for the rational design of antivirals.

344

345

346 **MATERIALS AND METHODS**

347 **Plasmid constructions**

348 The N and P sequences used for bacterial expression were derived from HMPV CAN97-83 strain. The
349 sequence of N or N_{NTD} (residues 30-255 of N) were cloned into pET-28a(+) vector (Novagen) at
350 NcoI/XhoI or BamHI/XhoI restriction sites respectively, to allow bacterial expression of full-length N or
351 N_{NTD} domain with a C-terminal or N-terminal 6xHis tag. The sequence of P or PCT (residues 200-294
352 of P), were cloned into pGEX-4T3 vector, at BamHI and SmaI sites to express in bacteria GST-P and
353 GST-PCT fusion proteins. For the construction GST-P[285-294], complementary antiparallel
354 oligonucleotides were hybridized and inserted at the BamHI/SmaI sites in the pGEX-4T-3 vector.

355 A HMPV minigenome plasmid containing Gaussia/Firefly luciferases was designed and synthesized
356 by Genscript, cloned in pUC57 vector, and containing the trailer, leader, gene start (GS) and gene end
357 (GE) sequences derived from HMPV CAN97-83 strain (see Fig S1). The first ORF of this
358 pGaussia/Firefly minigenome codes for the Firefly luciferase and the second one codes for the
359 Gaussia luciferase. The plasmids pP, pN, pL, and pM2-1 corresponding to the sequences of NL99-1
360 HMPV strain cloned into pCite vector. For expression of GFP-N fusion protein, the mGFP gene was
361 amplified by PCR and cloned in frame at the 3' end of N at the EcoRI site of pN, using In-fusion HD
362 cloning kit (Takara Bio). For expression of P-BFP fusion protein, a BamHI cleavage site was inserted
363 between position S61 and S62 of P by site-directed mutagenesis. The BFP gene was then amplified
364 by PCR and cloned at the BamHI site of modified pP, using In-fusion HD cloning kit (Takara Bio).
365 Point mutations were introduced in the P and N sequences by site-directed mutagenesis using the Q5
366 site-directed mutagenesis Kit (New England Biolabs). Sequence analysis was carried out to check the
367 integrity of all constructs. All the oligonucleotides sequences are available on request.

368

369 **Antibodies**

370 Antisera used in this study included polyclonal rabbit antisera raised against recombinant HMPV N
371 and P expressed in bacteria. A mouse monoclonal anti- β -tubulin (Sigma) and secondary antibodies
372 directed against mouse and rabbit Ig G coupled to HRP (P.A.R.I.S) were also used for
373 immunoblotting. Secondary antibodies directed against rabbit Ig G coupled to Alexafluor-594 or to
374 Alexafluor-488 (Invitrogen) were used for immunofluorescence experiments.

375

376 **Cell culture and transfections**

377 BHK-21 cells (clone BSRT7/5) constitutively expressing the T7 RNA polymerase (25) were grown in
378 Dulbecco Modified Essential Medium (Lonza) supplemented with 10% fetal calf serum (FCS), 2 mM
379 glutamine, and antibiotics. The cells were grown at 37°C in 5% CO₂, and transfected using
380 Lipofectamine 2000 (Invitrogen) as described by the manufacturer.

381

382 ***Minigenome replication assay***

383 BSRT7/5 cells at 90% confluence in 48-well dishes were transfected using Lipofectamine 2000
384 (Invitrogen) with a plasmid mixture containing 0.125 µg of pGaussia/Firefly minigenome, 0.125 µg of
385 pN, 0.125 µg of pP (WT and mutants), 0.06 µg of pL, 0.03 µg of pM2-1, as well as 0.03 µg of pSV-β-
386 Gal (Promega) to normalize transfection efficiencies. Cells were harvested at 24 h post-transfection
387 and lysed in Firefly lysis buffer (30 mM Tris [pH 7.9], 10 mM MgCl₂, 1 mM dithiothreitol (DTT), 1%
388 [vol/vol] Triton X-100, and 15% [vol/vol] glycerol). The Firefly luciferase activity was determined for
389 each cell lysate with an Infinite 200 Pro (Tecan, Männedorf, Switzerland) and normalized based on β-
390 galactosidase (β-Gal) expression. Transfections were done in triplicate, and each independent
391 transfection experiment was performed three times. The significance of the differences between wild
392 type condition and mutants was determined using Student's t-test (GraphPad Prism software).
393 Significance is represented: ***p < 0.001; ****p < 0.0001; ns, not significant. For proteins expression
394 analysis, cells were lysed in Laemmli buffer and analyzed by Western blotting (WB) using anti-N, anti-
395 P, and anti-tubulin antibodies according to standard protocols.

396

397 ***Fluorescence microscopy***

398 Immunofluorescence microscopy was performed with cells grown on coverslips and co-transfected
399 with either pP (WT or mutant) and pN and pGFP-N (ratio pN:pGFP-N of 2:1), or pP-BFP and pN (WT
400 or mutants). At 24 h post transfection, cells were fixed with 4% paraformaldehyde (PFA) for 25 min,
401 made permeable, and blocked for 30 min with PBS containing 0.1% Triton X-100 and 3% bovine
402 serum albumin (BSA). Cells were then successively incubated for 1 h at room temperature with
403 primary (rabbit anti-N or rabbit anti-P) and secondary antibodies mixtures diluted in PBS containing
404 0.3% BSA. For nucleus labeling, cells were exposed to Hoechst 33342 stain (Invitrogen) during
405 incubation with secondary antibodies. Coverslips were mounted with ProLong Gold antifade reagent
406 (Invitrogen) and observed with a Nikon TE200 microscope equipped with a CoolSNAP ES2

407 (Photometrics, Tucson, USA) camera. Images were processed with ZEN software (Zeiss) and ImageJ
408 software.

409

410 ***Expression and purification of recombinant proteins***

411 *E. coli* BL21 bacteria (DE3) (Novagen) transformed with pGEX-P plasmids (WT or mutant) or pET-
412 N_{NTD} were grown at 37°C for 8 hours in 100 ml of Luria Bertani (LB) medium containing 100µg/ml
413 ampicillin or 50µg/ml kanamycin, respectively. Bacteria transformed with pET-N-derived plasmids
414 together with pGEX-P derived plasmids were grown in LB medium containing ampicillin (100 µg/ml)
415 and kanamycin (50 µg/ml). The same volume of LB was then added and protein expression was
416 induced by adding 80µg/ml isopropyl-β-D-thio-galactoside (IPTG) to the medium. The bacteria were
417 incubated for 15 hours at 28°C and then harvested by centrifugation. For purification using the GST-
418 tag, bacteria were re-suspended in lysis buffer (50 mM Tris-HCl pH 7.8, 60 mM NaCl, 1 mM EDTA, 2
419 mM DTT, 0.2% Triton X-100, 1 mg/ml lysozyme) supplemented with complete protease inhibitor
420 cocktail (Roche), incubated for 1 hour on ice, sonicated, and centrifuged at 4°C for 30 min at 10,000g.
421 Glutathione-Sepharose 4B beads (GE Healthcare) were added to clarified supernatants and incubated
422 at 4°C for 3 hours. Beads were then washed two times in lysis buffer and three times in PBS 1X, then
423 stored at 4°C in an equal volume of PBS. To isolate GST-free P fragments or N rings, beads
424 containing bound proteins were incubated with thrombin (Novagen) overnight at 20°C. Purified
425 recombinant N proteins were loaded onto a Superdex 200 16/30 column (GE Healthcare) and eluted
426 in 20 mM Tris/HCl pH 8.5, 150 mM NaCl.

427 For purification of 6xHis-N and N_{NTD}-6xHis fusion proteins purification, bacterial pellets were re-
428 suspended in lysis buffer (20 mM Tris-HCl pH8, 500 mM NaCl, 0.1% TritonX-100, 10 mM imidazole, 1
429 mg/ml lysozyme) supplemented with complete protease inhibitor cocktail (Roche). After sonication and
430 centrifugation, lysates were incubated 30 min with chelating Sepharose Fast Flow beads charged with
431 Ni²⁺ (GE Healthcare). Beads were then successively washed in the washing buffer (20 mM Tris-HCl,
432 pH 8, 500 mM NaCl) containing increasing concentration of imidazole (25, 50, and 100 mM). For SDS-
433 PAGE analysis of proteins on beads, samples were prepared in Laemmli buffer, denatured 5 min at 95
434 °C, separated on 10% polyacrylamide gel and detected by Coomassie brilliant blue.

435

436 ***Dynamic light scattering (DLS)***

437 Size measurement of purified N oligomers was performed at 20 °C using a helium-neon laser
438 wavelength of 633 nm and detection angle of 173° with a Zetasizer Nano (Malvern). Ten
439 measurements were made, with an acquisition time of 10 s for each measurement. Hydrodynamic
440 diameters (D_h) were calculated using the DLS software provided by the instrument manufacturer. The
441 results were presented as size distribution (nm).

442

443 ***Negative stain electron microscopy observations of recombinant nucleoproteins***

444 Three microliters of sample were applied to the clean side of carbon on a carbon–mica interface and
445 stained with 2% sodium silicotungstate. Micrographs were recorded on a FEI Tecnai T12 microscope
446 operated at 120 kV with a Gatan Orius 1000 camera. Images were recorded at a nominal
447 magnification of 23 000 × resulting in a pixel size of 2.8 Å.

448

449 ***Molecular modeling***

450 In the X-ray structures of HMPV N-RNA rings (PDB 5fvc) or N⁰-P (PDB 5fvd), the N_{NTD} residues L110
451 to M113, corresponding to a long beta-hairpin close to the PCT binding site, are not modeled. We
452 used Modeller (26) to first construct a complete model HMPV N_{NTD} (residues 32 to 251) using as
453 templates the X-ray structures of HMPV N-RNA rings (residues 32 to 109 and 114 to 251, PDB) and
454 RSV N_{NTD} (PDB 4uc6) for the missing beta-hairpin (residues 110 to 113). Next, using Modeller again,
455 we generated 100 models of HMPV N_{NTD} complexed with the last six residues of P (peptide I289-
456 YQLI-M294 called P6) using the RSV N_{NTD} structure in complex with the last two residues of RSV P
457 (PDB 4uc9), *i.e.* D240-F241 (called P2), as template for HMPV P6 by aligning the last 2 residues of
458 HMPV-P and RSV-P. To compensate for possible alternate conformations of exposed residues in apo
459 and holo-form of N_{NTD}, no template was used for R132 and R151 side-chains. Next, P6-N_{NTD} models
460 where M294 side-chain of P is pointing towards the N_{NTD} hydrophobic pocket (by analogy with the
461 binding of F241 of P on N RSV) were selected, to obtain an ensemble 41 P6-N_{NTD} models. Finally, the
462 models were submitted to a water refinement procedure using the refinement interface of the
463 HADDOCK2.2 server (27). Each input P6-N_{NTD} complex model was refined 20 times to obtain 820
464 refined models. The best 400 models with the lowest HADDOCK-score were clustered based on
465 similarity of the L292-I293-M294 residues of P, called P3 (I289-Y290-Q291 were not considered for
466 clustering since they appear very floppy in the refined models and had no structural templates during

467 homology modeling). Models were first superposed on their N_{NTD} backbone atoms and the RMSD was
468 computed on the backbone atoms of L292-I293-M294 of P. Clustering was performed using the
469 HADDOCK tool *cluster_struc* (28) with a 2Å cut-off and a minimal cluster size of 10.

470 In addition, the 820 P6- N_{NTD} conformations generated by HADDOCK were profiled using InDeep
471 software (20) in order to select the most suitable N_{NTD} conformations to bind P3. InDeep is a deep-
472 learning model based on a FCN (Fully Convolutional Network) trained from known structures of
473 protein complexes capable of predicting protein-protein interaction interfaces
474 (<https://ippidb.pasteur.fr/targetcentric/>) (20). Each N_{NTD} pocket conformation is placed within a 3D grid
475 composed of 1 Å³ voxels. The model computes a probability value for each voxel representing its
476 interaction propensity. The predictions were performed on the putative P6 binding site on N_{NTD}
477 surface. An “interactability-score” of the pocket is computed by taking the mean of the 150 best voxels
478 probabilities. This value represents the binding propensity of a given N_{NTD} pocket conformation. Then,
479 on the 200-top conformations with the highest “interactability-score” the hydrophobic channel of
480 InDeep was applied in order to locate the most probable hydrophobic patches on the N_{NTD} pocket
481 where P3 hydrophobic side chains could interact. In addition, once a set most promising
482 conformations was identified, a virtual alanine scanning was performed using the same tool but by
483 mutating one by one the residues surrounding the patch of interaction of P3. The results are provided
484 as a set of correlations, one for each residue, highlighting the impact of the simulated mutation on the
485 interactability of the N_{NTD} .

486

487

488

489 **ACKNOWLEDGMENTS**

490 We thank the immunologist's team of Seppic laboratory (Maisons-Alfort, France) and the animal
491 facilities of Anses, Maisons-Alfort for rabbit immunization and collect of anti-N and anti-P antisera, P.
492 Collins (NIH, Boston) and S. Biacchesi (INRAE, Jouy-en-Josas) for the sequences of HMPV CAN97-
493 83 N and P genes, B. van den Hoogen (Erasmus MC, Rotterdam) for the plasmids of HMPV NL99-1
494 strain.

495 This work was carried out with the financial support of the French Agence Nationale de la Recherche,
496 specific program ANR DecRisP n° ANR-19-CE11-0017.

497

498 We declare that we have no conflicts of interest with the contents of this article.

499 H.D., M.G. and J.-F.E. designed experiments. H.D., M.B. performed molecular, and cellular assays.

500 B.B., L.C.R. and O.S. performed molecular simulations. I.G performed electron microscopy

501 observations. C.-A.R. performed gel filtration and *in vitro* study. M.G. wrote the paper with

502 contributions from all authors, and J.-F.E. and M.G. edited the manuscript. All authors commented on

503 the manuscript.

504

505 **References**506 1. Pneumonia Etiology Research for Child Health Study G. 2019. Causes of severe
507 pneumonia requiring hospital admission in children without HIV infection from Africa
508 and Asia: the PERCH multi-country case-control study. *Lancet* 394:757-779.509 2. Schildgen V, van den Hoogen B, Fouchier R, Tripp RA, Alvarez R, Manoha C,
510 Williams J, Schildgen O. 2011. Human Metapneumovirus: lessons learned over the
511 first decade. *Clin Microbiol Rev* 24:734-54.512 3. van den Hoogen BG, de Jong JC, Groen J, Kuiken T, de Groot R, Fouchier RA,
513 Osterhaus AD. 2001. A newly discovered human pneumovirus isolated from young
514 children with respiratory tract disease. *Nat Med* 7:719-24.515 4. Afonso CL, Amarasinghe GK, Banyai K, Bao Y, Basler CF, Bavari S, Bejerman N,
516 Blasdel KR, Briand FX, Briesse T, Bukreyev A, Calisher CH, Chandran K, Cheng J,
517 Clawson AN, Collins PL, Dietzgen RG, Dolnik O, Domier LL, Durrwald R, Dye JM,
518 Easton AJ, Ebihara H, Farkas SL, Freitas-Astua J, Formenty P, Fouchier RA, Fu Y,
519 Ghedin E, Goodin MM, Hewson R, Horie M, Hyndman TH, Jiang D, Kitajima EW,
520 Kobinger GP, Kondo H, Kurath G, Lamb RA, Lenardon S, Leroy EM, Li CX, Lin XD,
521 Liu L, Longdon B, Marton S, Maisner A, Muhlberger E, Netesov SV, Nowotny N, et al.
522 2016. Taxonomy of the order Mononegavirales: update 2016. *Arch Virol* 161:2351-60.523 5. van den Hoogen BG, Bestebroer TM, Osterhaus AD, Fouchier RA. 2002. Analysis of
524 the genomic sequence of a human metapneumovirus. *Virology* 295:119-32.525 6. Herfst S, de Graaf M, Schickli JH, Tang RS, Kaur J, Yang CF, Spaete RR, Haller AA,
526 van den Hoogen BG, Osterhaus AD, Fouchier RA. 2004. Recovery of human
527 metapneumovirus genetic lineages a and B from cloned cDNA. *J Virol* 78:8264-70.528 7. Cifuentes-Munoz N, Brantje J, Slaughter KB, Dutch RE. 2017. Human
529 Metapneumovirus Induces Formation of Inclusion Bodies for Efficient Genome
530 Replication and Transcription. *J Virol* 91:e01282-17531 8. Derdowski A, Peters TR, Glover N, Qian R, Utley TJ, Burnett A, Williams JV,
532 Spearman P, Crowe JE. 2008. Human metapneumovirus nucleoprotein and
533 phosphoprotein interact and provide the minimal requirements for inclusion body
534 formation. *J Gen Virol* 89:2698-2708.535 9. Galloux M, Risso-Ballester J, Richard CA, Fix J, Rameix-Welti MA, Eleouet JF. 2020.
536 Minimal Elements Required for the Formation of Respiratory Syncytial Virus
537 Cytoplasmic Inclusion Bodies In Vivo and In Vitro. *mBio* 11:e01202-20..538 10. Leyrat C, Renner M, Harlos K, Grimes JM. 2013. Solution and crystallographic
539 structures of the central region of the phosphoprotein from human metapneumovirus.
540 *PLoS One* 8:e80371.541 11. Renner M, Paesen GC, Grison CM, Granier S, Grimes JM, Leyrat C. 2017. Structural
542 dissection of human metapneumovirus phosphoprotein using small angle x-ray
543 scattering. *Sci Rep* 7:14865.

- 544 12. Pan J, Qian X, Lattmann S, El Sahili A, Yeo TH, Jia H, Cressey T, Ludeke B, Noton
545 S, Kalocsay M, Fearn R, Lescar J. 2020. Structure of the human metapneumovirus
546 polymerase phosphoprotein complex. *Nature* 577:275-279.
- 547 13. Tran TL, Castagne N, Bhella D, Varela PF, Bernard J, Chilmonczyk S, Berkenkamp
548 S, Benhamo V, Grznarova K, Grosclaude J, Nespoulos C, Rey FA, Eleouet JF. 2007.
549 The nine C-terminal amino acids of the respiratory syncytial virus protein P are
550 necessary and sufficient for binding to ribonucleoprotein complexes in which six
551 ribonucleotides are contacted per N protein protomer. *J Gen Virol* 88:196-206.
- 552 14. Renner M, Bertinelli M, Leyrat C, Paesen GC, Saraiva de Oliveira LF, Huiskonen JT,
553 Grimes JM. 2016. Nucleocapsid assembly in pneumoviruses is regulated by
554 conformational switching of the N protein. *eLife* 5:e12627
- 555 15. Tawar RG, Duquerroy S, Vonrhein C, Varela PF, Damier-Piolle L, Castagne N,
556 MacLellan K, Bedouelle H, Bricogne G, Bhella D, Eleouet JF, Rey FA. 2009. Crystal
557 structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial
558 virus. *Science* 326:1279-83.
- 559 16. Ouizougoun-Oubari M, Pereira N, Tarus B, Galloux M, Lassoued S, Fix J, Tortorici MA,
560 Hoos S, Baron B, England P, Desmaele D, Couvreur P, Bontems F, Rey FA, Eleouet
561 JF, Sizun C, Slama-Schwok A, Duquerroy S. 2015. A Druggable Pocket at the
562 Nucleocapsid/Phosphoprotein Interaction Site of Human Respiratory Syncytial Virus.
563 *J Virol* 89:11129-43.
- 564 17. Galloux M, Tarus B, Blazevic I, Fix J, Duquerroy S, Eleouet JF. 2012.
565 Characterization of a viral phosphoprotein binding site on the surface of the
566 respiratory syncytial nucleoprotein. *J Virol* 86:8375-87.
- 567 18. Rincheval V, Lelek M, Gault E, Bouillier C, Sitterlin D, Blouquit-Laye S, Galloux M,
568 Zimmer C, Eleouet JF, Rameix-Welti MA. 2017. Functional organization of
569 cytoplasmic inclusion bodies in cells infected by respiratory syncytial virus. *Nature*
570 *Communications* 8:563.
- 571 19. Dominguez C, Boelens R, Bonvin AM. 2003. HADDOCK: a protein-protein docking
572 approach based on biochemical or biophysical information. *J Am Chem Soc*
573 125:1731-7.
- 574 20. Vincent Mallet, Luis Checa Ruano, Alexandra Moine Franel, Michael Nilges, Karen
575 Druart, Guillaume Bouvier, Olivier Sperandio. 2021. InDeep : 3D fully convolutional
576 neural networks to assist in silico drug design on protein-protein interactions. *bioRxiv*
577 doi: <https://doi.org/10.1101/2021.07.28.453974>
- 578 21. Sa JM, Piloto JV, Cilli EM, Tasic L, Fossey MA, Almeida FCL, Souza FP, Caruso IP.
579 2020. Hesperetin targets the hydrophobic pocket of the nucleoprotein/phosphoprotein
580 binding site of human respiratory syncytial virus. *J Biomol Struct Dyn* 20:1-13.
- 581 22. de Graaf M, Herfst S, Schrauwen EJA, Choi Y, van den Hoogen BG, Osterhaus A,
582 Fouchier RAM. 2008. Specificity and functional interaction of the polymerase complex
583 proteins of human and avian metapneumoviruses. *J Gen Virol* 89:975-983.
- 584 23. Marriott AC, Smith JM, Easton AJ. 2001. Fidelity of leader and trailer sequence usage
585 by the respiratory syncytial virus and avian pneumovirus replication complexes. *J*
586 *Virol* 75:6265-72.
- 587 24. Shapiro AB, Gao N, O'Connell N, Hu J, Thresher J, Gu RF, Overman R, Hardern IM,
588 Sproat GG. 2014. Quantitative investigation of the affinity of human respiratory
589 syncytial virus phosphoprotein C-terminus binding to nucleocapsid protein. *Virol J*
590 11:191.
- 591 25. Buchholz UJ, Finke S, Conzelmann KK. 1999. Generation of bovine respiratory
592 syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in
593 tissue culture, and the human RSV leader region acts as a functional BRSV genome
594 promoter. *J Virol* 73:251-9.
- 595 26. Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial
596 restraints. *J Mol Biol* 234:779-815.
- 597 27. van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastiris PL, Karaca E,
598 Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin A. 2016. The HADDOCK2.2 Web

- 599 Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol*
600 428:720-725.
- 601 28. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. 1999.
602 Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie*
603 *International Edition* 38: 236-240.
- 604 29. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo
605 generator. *Genome Res* 14:1188-90).
- 606
- 607
- 608

609 **Figure legends**

610

611 **Figure 1:** Purification of HMPV N-RNA rings using GST-PCT. A/ Schematic representation of HMPV
612 and RSV P proteins. The oligomerization domain is represented by a rectangle and disordered N- and
613 C-terminal domains (PNT and PCT respectively) by lines. PCT is indicated. Residues flanking the
614 structural domains are indicated by numbers. B/ GST-PCT and N proteins were co-expressed in *E.*
615 *coli*, followed by purification using the GST tag. The product of purification was analyzed by SDS-
616 PAGE and Coomassie blue staining (left lane). The sample was then incubated in the presence of
617 thrombin in order to cleave GST (remaining on beads) and isolate PCT-N in the supernatant (right
618 lane), C/ Gel filtration profile of purified PCT-N complex. The curves corresponding to OD spectra at
619 220 nm, 260 nm, and 280 nm are presented. P1 and P2 indicate the two peaks detected. The
620 fractions corresponding to P1 were pooled and the sample was analyzed by SDS-PAGE colored with
621 Coomassie blue, D/ Dynamic light scattering (DLS) analysis of the purified N protein, showing a
622 homogenous peak at 18 nm, corresponding to N oligomers. E/ Image of purified HMPV N-RNA rings
623 as observed by negative stain electron microscopy. Bar, 100 nm.

624

625 **Figure 2:** Identification of the residues of P involved in N binding. A/ Sequence alignment of
626 *Pneumoviridae* P C-terminus (PCT), from the Orthopneumovirus and Metapneumovirus genus: hRSV-
627 A/-B (human RSV strains A and B, Uniprot entries PHOSP_HRSVA and PHOSP_HRSVB,
628 respectively), oRSV (ovine RSV, Uniprot entry PHOSP_ORSVW), bRSV (bovine RSV, Uniprot entry
629 PHOSP_BRSPA), mPV (murine pneumonia virus, Uniprot entry PHOSP_MPV15), aMPV-A/-C (avian
630 metapneumovirus type A and C, Uniprot entries Q65032 and PHOSP_AMPV1, respectively) and
631 hMPV (human metapneumovirus strain CAN97-83, Uniprot entry PHOSP_HMPVC). Sequence
632 numbering and secondary structures are indicated on top for hRSV and at the bottom for hMPV. B/
633 GST-P fragments and C/ GST-P[285-294] (WT or mutants) (sequences indicated on the left) were co-
634 expressed with N in bacteria, followed by purification using the GST tag. The products of purification
635 were analyzed by SDS-PAGE and Coomassie blue staining (right). D/ GST-PCT or GST-PCT_{ALAM}
636 corresponding to the double substitution of residues Q291 and I293 residues of P by alanine were co-
637 expressed with N. The products of purification by GST were analyzed by SDS-PAGE and Coomassie
638 blue staining. E/ Sequence alignment of the 9 last C-terminal residues of HMPV and RSV P proteins.

639 The residues critical for the interaction with the N protein are indicated in red bold. F/ HMPV and RSV
640 GST-PCT (residues P[200-294] for HMPV and P[161-241] for RSV) constructs were co-expressed
641 with either HMPV or RSV N proteins. Copurification of N proteins by GST-PCT fragment was analyzed
642 by SDS-PAGE and Coomassie blue staining

643

644 **Figure 3:** Search for the PCT binding site on HMPV N. A/ Amino acid sequence alignments between
645 the N-terminal domains of N (residues 1-252/253) proteins of HMPV (strain CAN 97-83) and human
646 RSV (HRSV, strain Long VR-26). Invariant residues are highlighted in white font on a red background.
647 The secondary structure elements observed in the crystal structure of HMPV N protein (11) are
648 indicated above the sequence. Green asterisks below the sequence indicate the residues constituting
649 the P binding pocket of RSV N. Uniprot accession codes: NCAP_HRSVA (Human RSV);
650 A1DZS3_9MONO (Human MPV). Sequences were aligned with Clustal W and treated with ESPrpt 3.
651 B/ Side view (left) and bottom view (right) of the HMPV N-RNA rings (PDB accession number 5fvc)
652 showing one N monomer coloured according to domains: N-arm in blue, C-arm in green, and N- and
653 C-terminal domains (N_{NTD} and N_{CTD} , respectively) are in yellow and red. The residues identified by
654 analogy with RSV as involved in P binding on N_{NTD} are indicated by black spheres.

655

656 **Figure 4:** Identification of the residues of N_{NTD} involved in P binding. A/ GST-P or GST- $P_{\Delta M294}$ and
657 N_{NTD} -6xHis proteins were co-expressed in *E. coli*, followed by purification using the 6xHis tag. GST-P
658 or GST- $P_{\Delta M294}$ alone were purified using the GST-tag. The products of purification were analyzed by
659 SDS-PAGE and Coomassie blue staining, B/ Surface representation of HMPV N_{NTD} (from Gly30 to Gly
660 253, indicated in green) showing the potential P binding pocket, with acidic amino acids colored in red,
661 basic residues in blue, and hydrophobic residues in orange, C/ The impact of N mutations on the
662 expression and solubility of the protein was assessed by expressing 6xHis-N protein (WT or mutants)
663 in *E. coli*, followed by purification and SDS-PAGE analysis. D/ GST-PCT and mutant N proteins were
664 co-expressed in *E. coli*, followed by purification using the GST tag. The products of purification were
665 analyzed by SDS-PAGE and Coomassie blue staining (upper panel), and the expression of N proteins
666 in bacteria was validated by western blot (lower panel).

667

668 **Figure 5:** The residues involved in P-N-RNA rings interaction are critical for the polymerase activity. A/
669 HMPV polymerase activity in the presence of P and N mutants. BSRT7/5 cells were transfected with

670 plasmids encoding the P (WT or mutants), N (WT, or mutants), L, and M2-1 proteins, the
671 pGaussia/Firefly minigenome, together with pCMV- β Gal for transfection standardization. Viral RNA
672 synthesis was quantified by measuring the Gaussia luciferase activity after cell lysis 24 h after
673 transfection. Each luciferase minigenome activity value was normalized based on β -galactosidase
674 expression and is the average of three independent experiments performed in triplicate. Error bars
675 represent standard deviations calculated based on two independent experiments made in
676 quadruplicate. B/ Western blot showing the expression of P and N proteins (WT and mutants) in
677 BSRT7/5 cells. C and D/ HMPV polymerase activity in the presence of pN:pGFP-N at a ratio 2:1, or P-
678 BFP, respectively. Error bars represent standard deviations calculated based on two independent
679 experiments made in triplicate. The statistical significance of differences was calculated using
680 Student's *t* test. *** $p < 0.001$; **** $p < 0.001$; ns, not significant.

681

682 **Figure 6:** Impact of P mutations on the formation of HMPV cytoplasmic IBs. BSRT7/5 cells transfected
683 with plasmids encoding for P (WT or mutants) and N + GFP-N (ratio 2:1) proteins were fixed 24 h
684 post-transfection, labeled with anti-P antibody, and the distribution of viral proteins was observed by
685 fluorescence microscopy. Nuclei were stained with Hoechst 33342. Scale bars, 10 μ m.

686

687 **Figure 7:** Impact of N mutations on the formation of HMPV cytoplasmic IBs. BSRT7/5 cells
688 transfected with plasmids encoding for P-BFP and N (WT or mutants) proteins were fixed 24 h post-
689 transfection, labeled with anti-N antibody, and the distribution of viral proteins was observed by
690 fluorescence microscopy. Scale bars, 10 μ m.

691

692 **Figure 8:** Model of HMPV P3/N_{NTD} complex from HADDOCK refinement. A/ The N_{NTD} protein is
693 colored by electrostatic surface and the P3 peptide (residues L292-I293-M294 of P) is shown as
694 sticks. The residues of N_{NTD} critical for the interaction with P are indicated, with residues for which
695 mutation to alanine abrogate P binding are labeled in bold. Yellow Surface shows InDeep hydrophobic
696 channel prediction where hydrophobic contacts are expected to occur at the N_{NTD} surface. B/
697 Interaction diagram between P3 peptide (Leu292-Ile293-Met294) and N_{NTD}. Legend indicated below
698 the diagram.

699

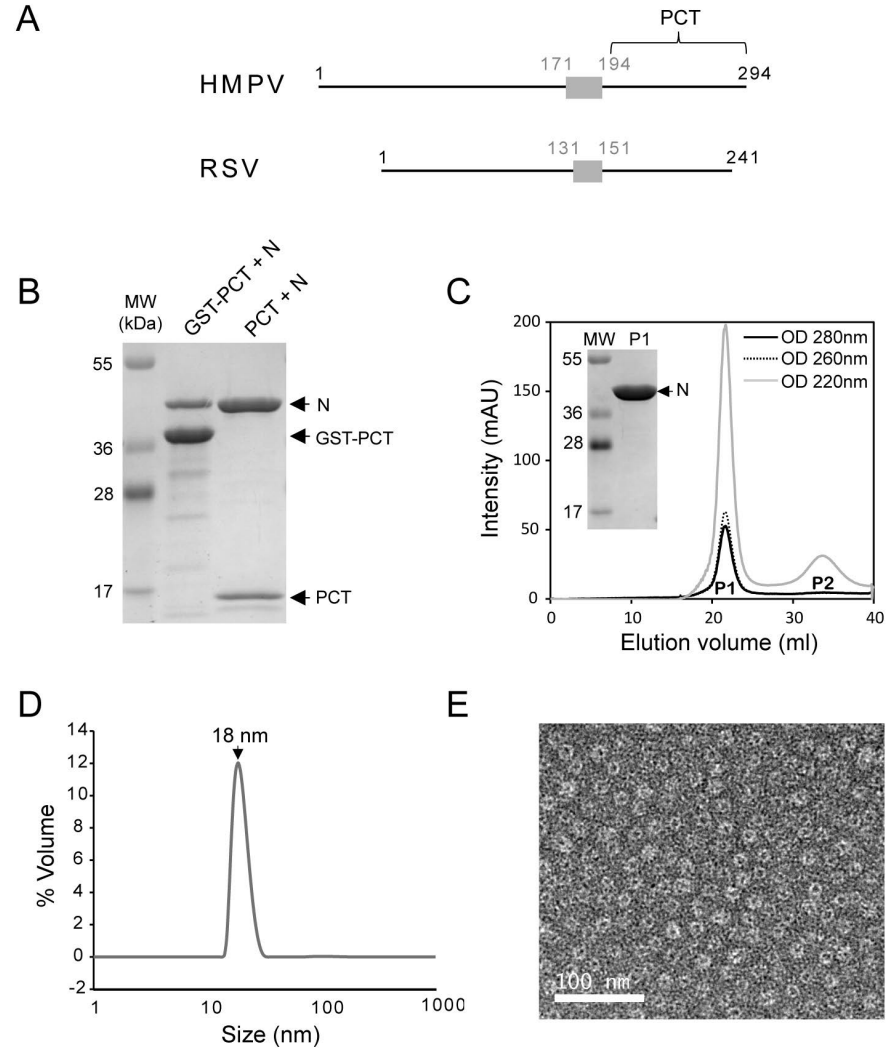
700 **Figure 9.** 10 best-scoring structures of clusters 1 (A) and 2 (B) from HADDOCK refinement of HMPV
701 P3 (residues L292-I293-M294 of P)/N_{NTD} complex superimposed with RSV P2/N_{NTD}. N_{NTD} is colored in
702 beige (HMPV) or yellow (RSV), and HMPV P3 peptide in cyan (HMPV) or RSV P2 peptide
703 (corresponding to residues D240-F241 of P) in pink. Side-chains of residues in contact with the P
704 peptide in HADDOCK models are also shown, along with their RSV equivalent.

705

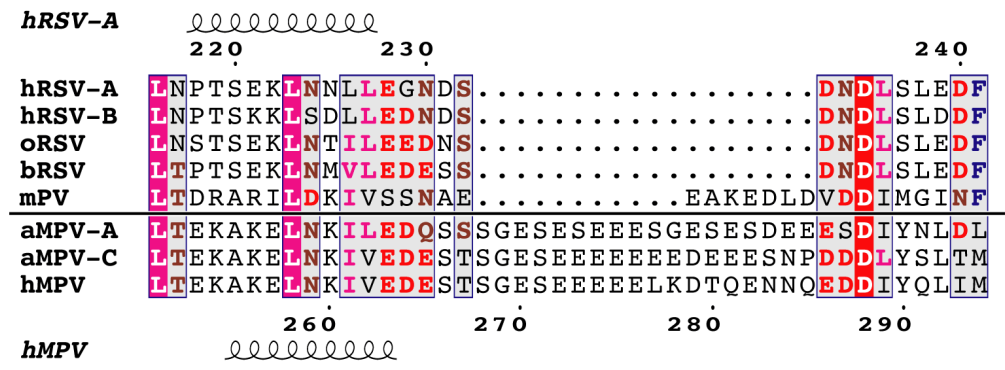
706 **Figure 10.** Conservation of pneumoviruses N and P residues. A/ Sequence logo of the last 9 C-
707 terminal residues of P of Metapneumovirus (left) and Orthopneumovirus (right) sequences. P
708 sequences were collected from the Interpro entry IPR003487 (464 unique sequences) and logo were
709 created from non-redundant peptides sequences of the last 9 amino acids. B/ Sequence logo of the
710 residues forming the P-binding site on N of the Metapneumovirus (left) and Orthopneumovirus (right)
711 genus. N sequences were collected from the Interpro entry IPR004930 (359 unique sequences) and
712 aligned. Classification as Ortho- and Metapneumovirus was based on Uniprot taxonomy (taxon ids
713 1868215 for OrthoPV and 162387 for MetaPV). Residue numbers are shown at the bottom, and grey
714 markers indicate residues identified as essential for P-N interaction. Logo were generated with
715 WebLogo (29).

716

717

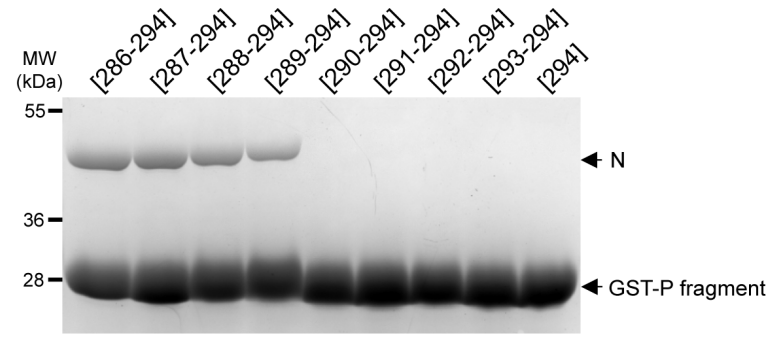


A



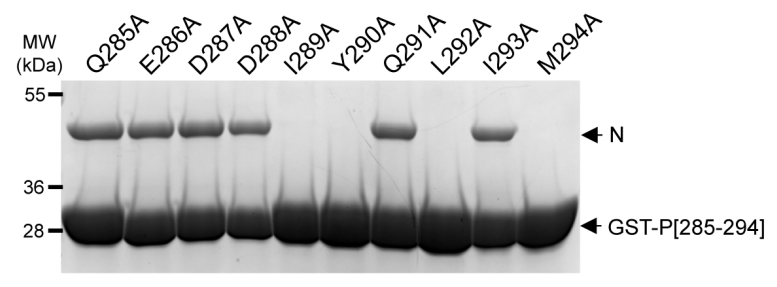
B

P segment	Sequence	Interaction with N
P [286-294]	EDDIYQLIM	+
P [287-294]	DDIYQLIM	+
P [288-294]	DIYQLIM	+
P [289-294]	IYQLIM	+
P [290-294]	YQLIM	-
P [291-294]	QLIM	-
P [292-294]	LIM	-
P [293-294]	IM	-
P [294]	M	-

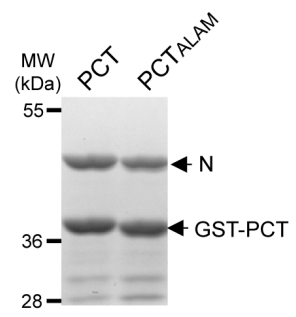


C

P segment	Sequence	Interaction with N
P[285-294]	QEDDIYQLIM	
Q285A	A - - - - -	+
E286A	- A - - - - -	+
D287A	-- A - - - -	+
D288A	--- A - - - -	+
I289A	---- A - - - -	-
Y290A	----- A - - -	-
Q291A	----- A - - -	+
L292A	----- A - -	-
I293A	----- A -	+
M294A	----- A	-



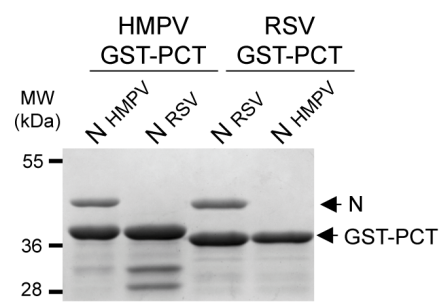
D



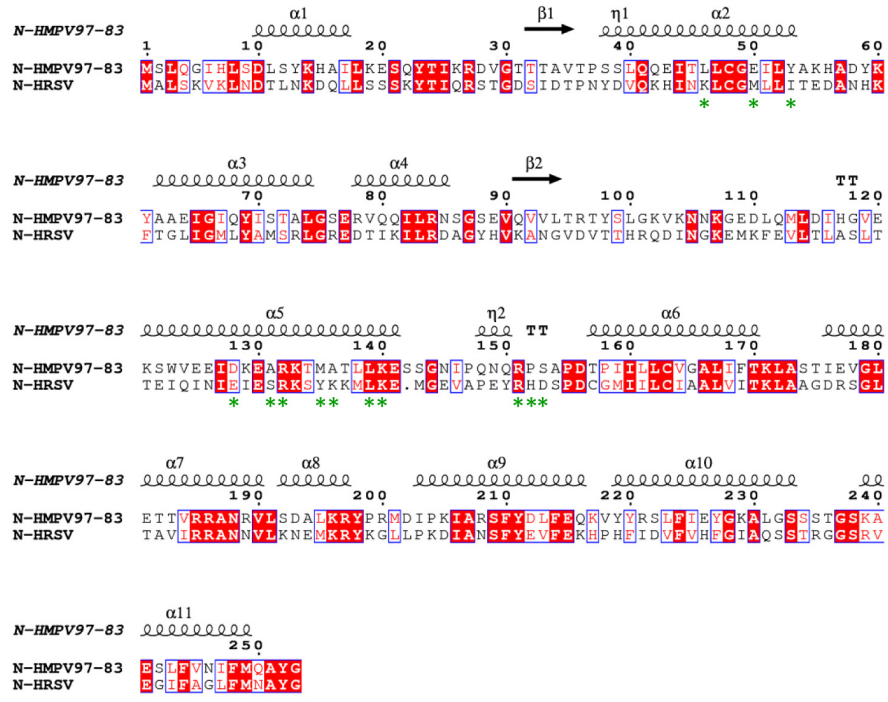
E

HMPV EDDIYQLIM
RSV DNDLSLED F

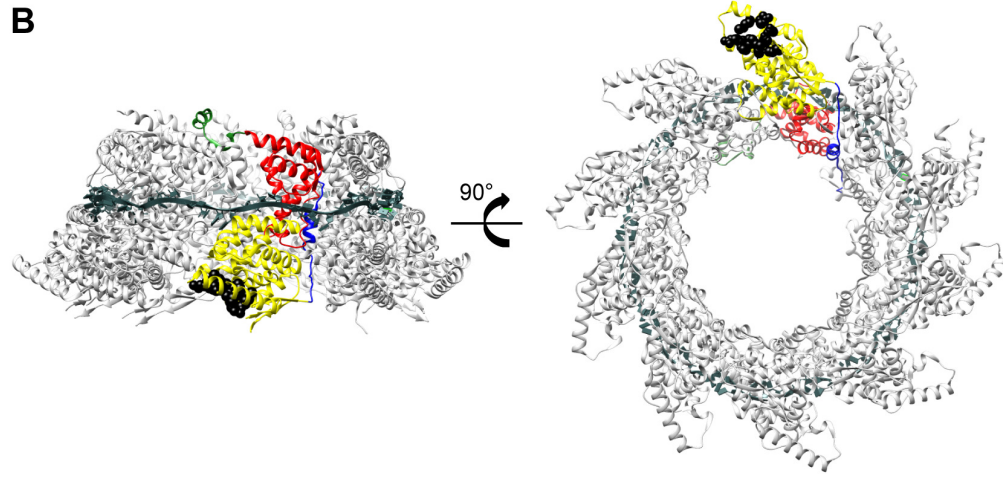
F

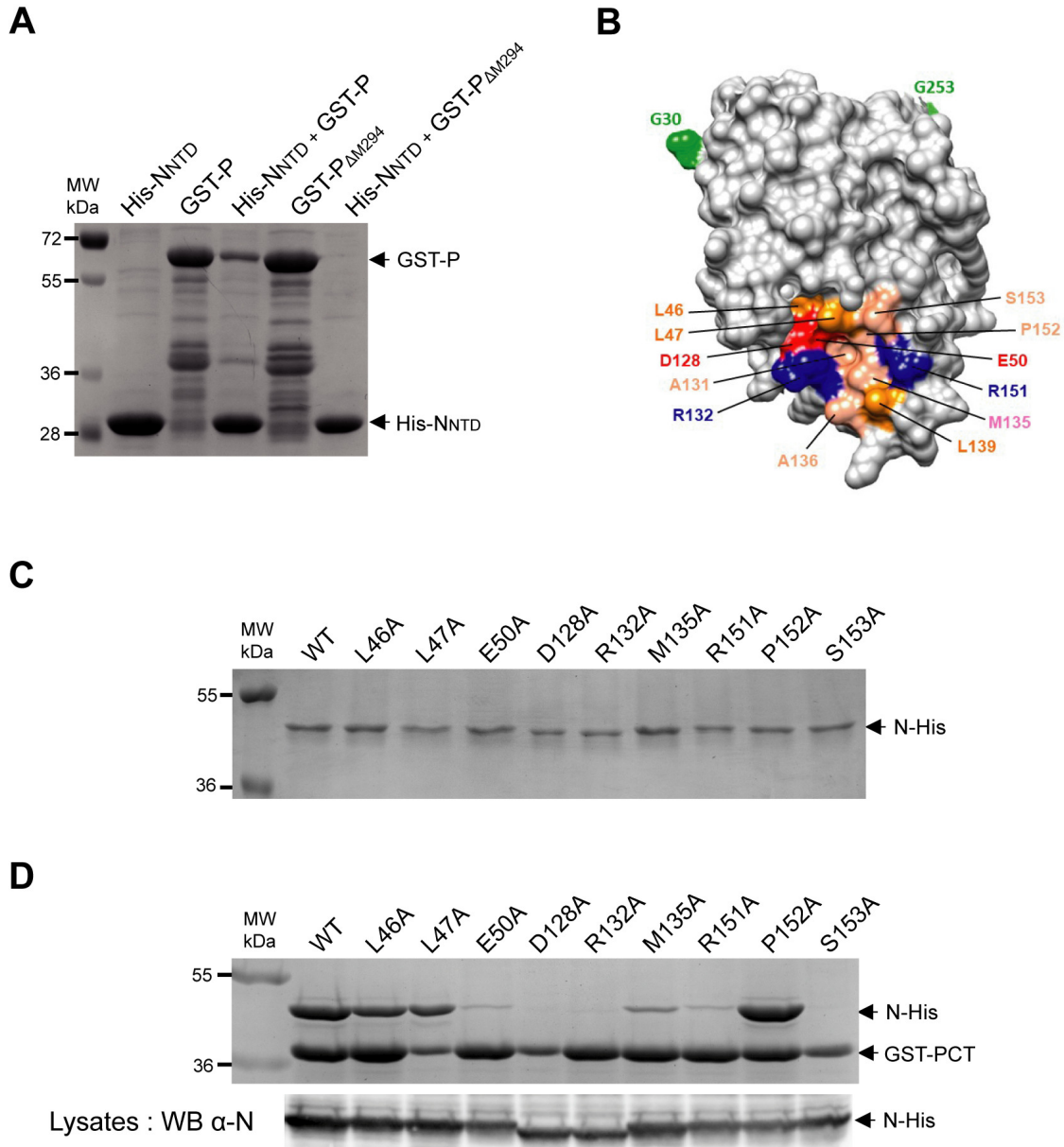


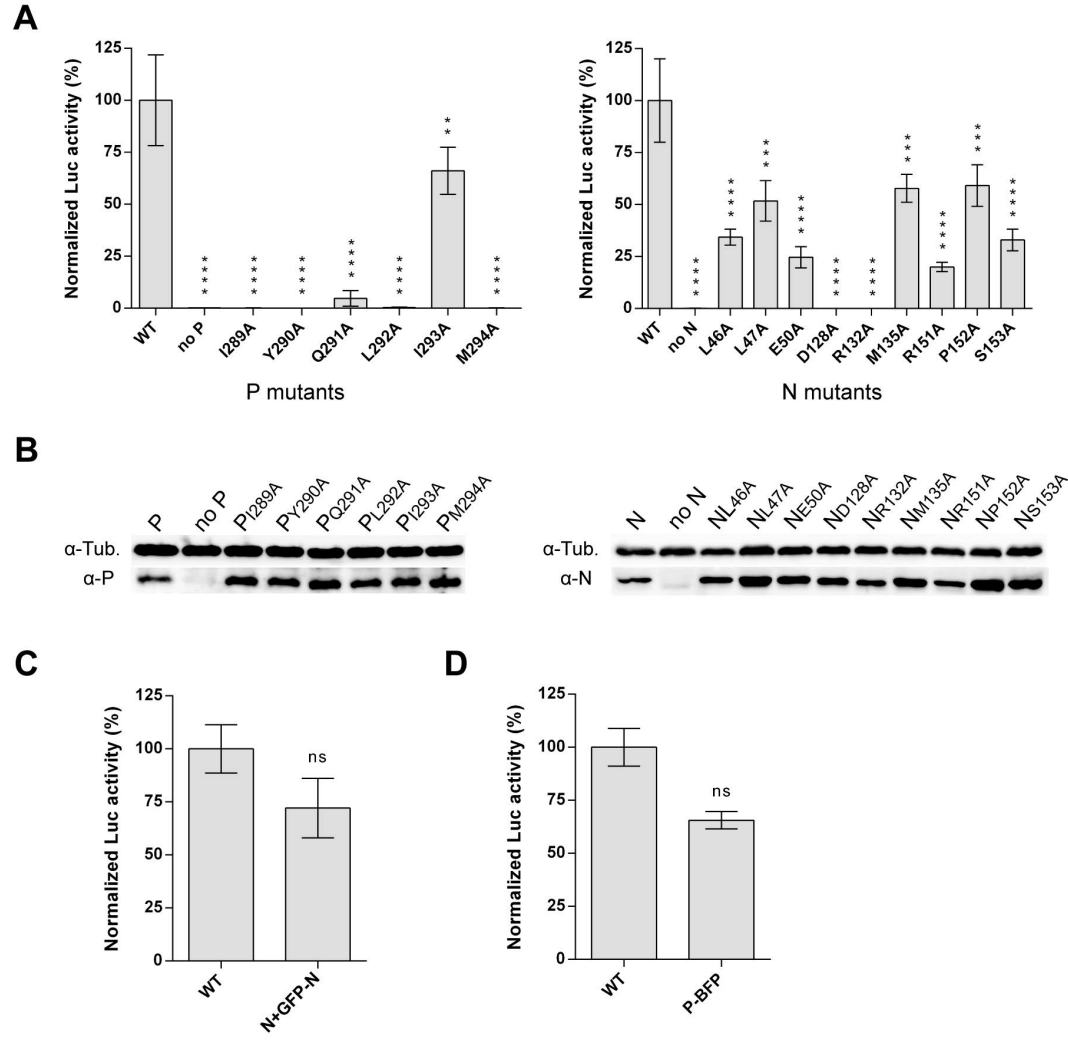
A

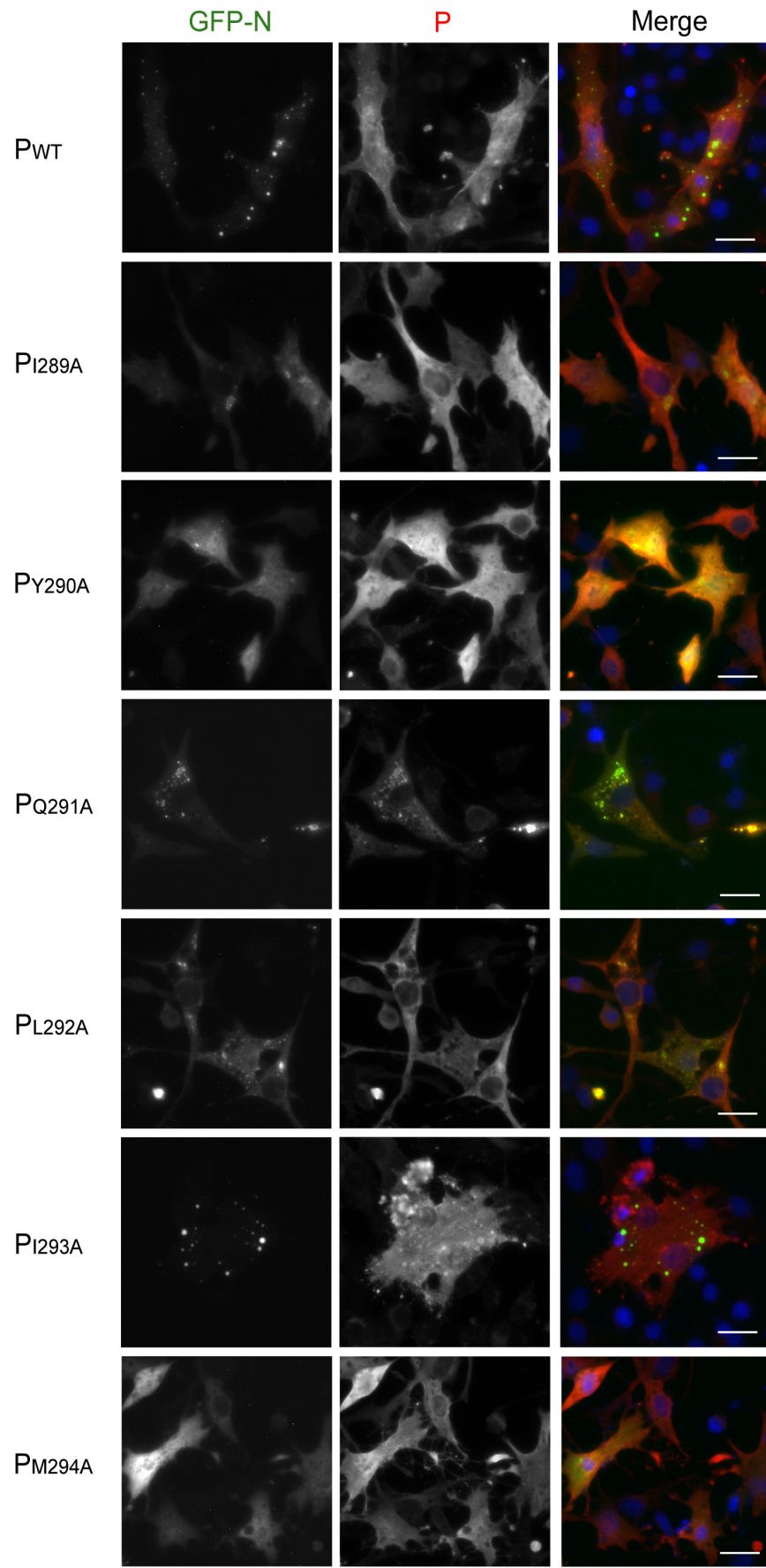


B









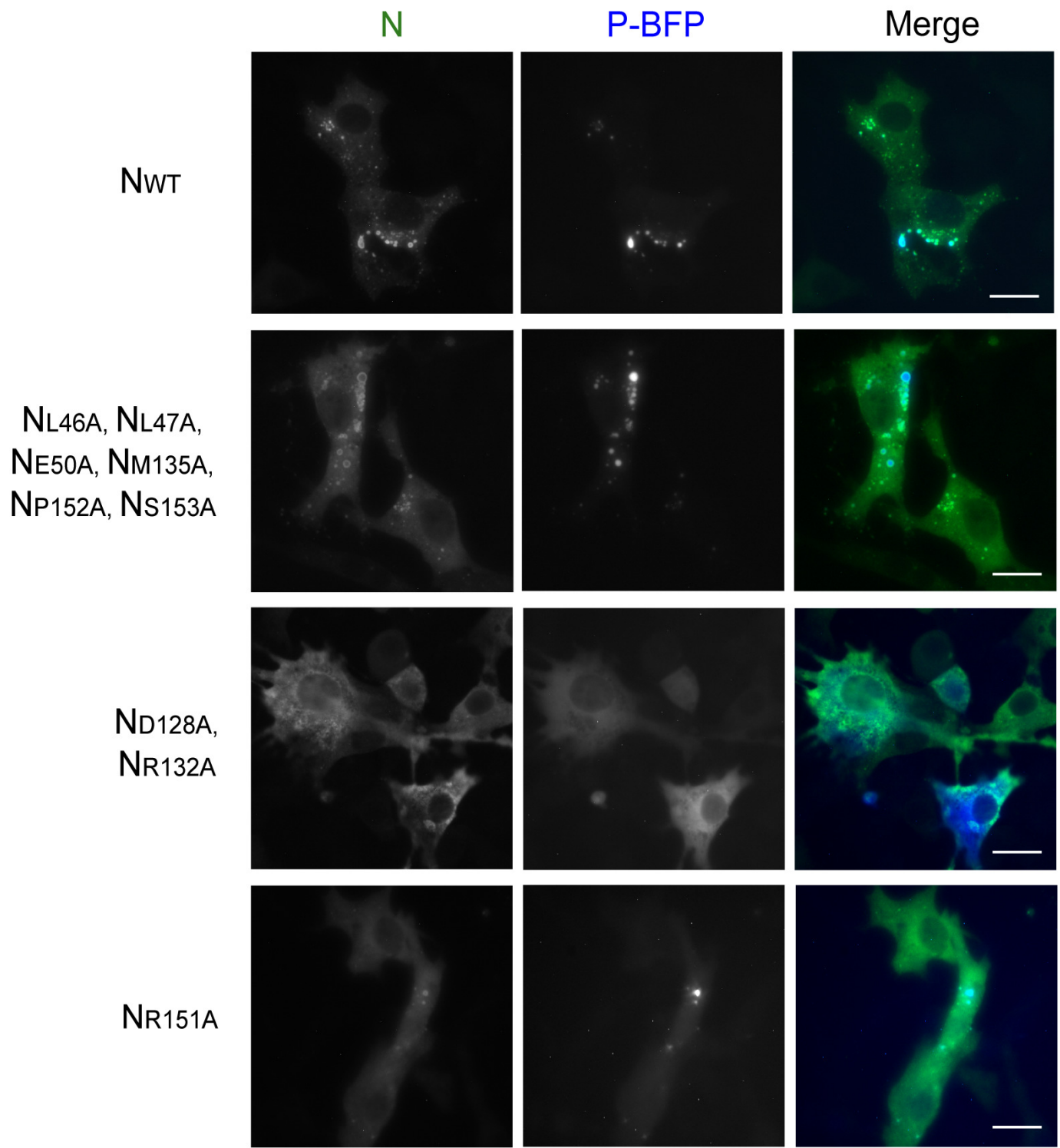
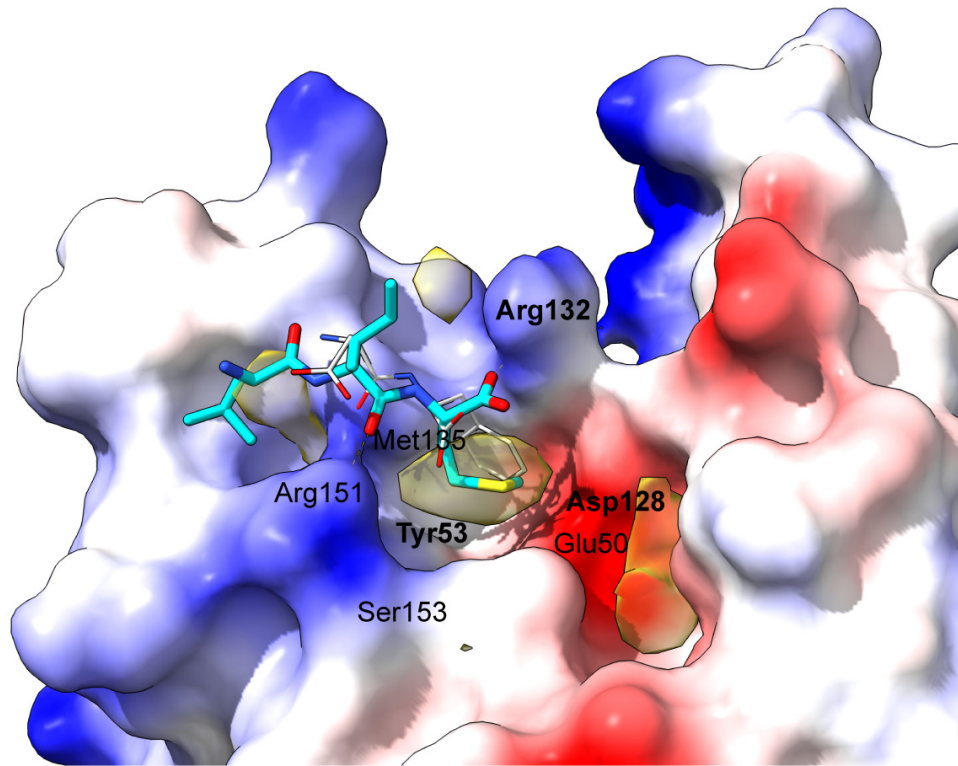


Table 1. Characterization of P and N mutants

	Mutation	Pulldown of N by GST-PCT	Polymerase activity *	Observation of IBs
P mutants	I289A	-	-	-
	Y290A	-	-	-
	Q291A	+	5%	+/-
	L292A	-	-	-
	I293A	+	65%	+
	M294A	-	-	-
N mutants	L46A	+	35%	+
	L47A	+	50%	+
	E50A	+/-	25%	+
	D128	-	-	-
	R132A	-	-	-
	M135A	+/-	60%	+
	R151A	+/-	20%	+/-
	P152A	+	60%	+
	S153A	-	30%	+

* Values correspond to rounded-up averages ; -, no polymerase activity.

A



B

