



OPEN

## Ensemble description of the intrinsically disordered N-terminal domain of the Nipah virus P/V protein from combined NMR and SAXS

Marco Schiavina<sup>1,2,5</sup>, Edoardo Salladini<sup>4,5</sup>, Maria Grazia Murralli<sup>1,2,5</sup>, Giancarlo Tria<sup>2,3</sup>, Isabella C. Felli<sup>1,2</sup>✉, Roberta Pierattelli<sup>1,2</sup>✉ & Sonia Longhi<sup>4</sup>✉

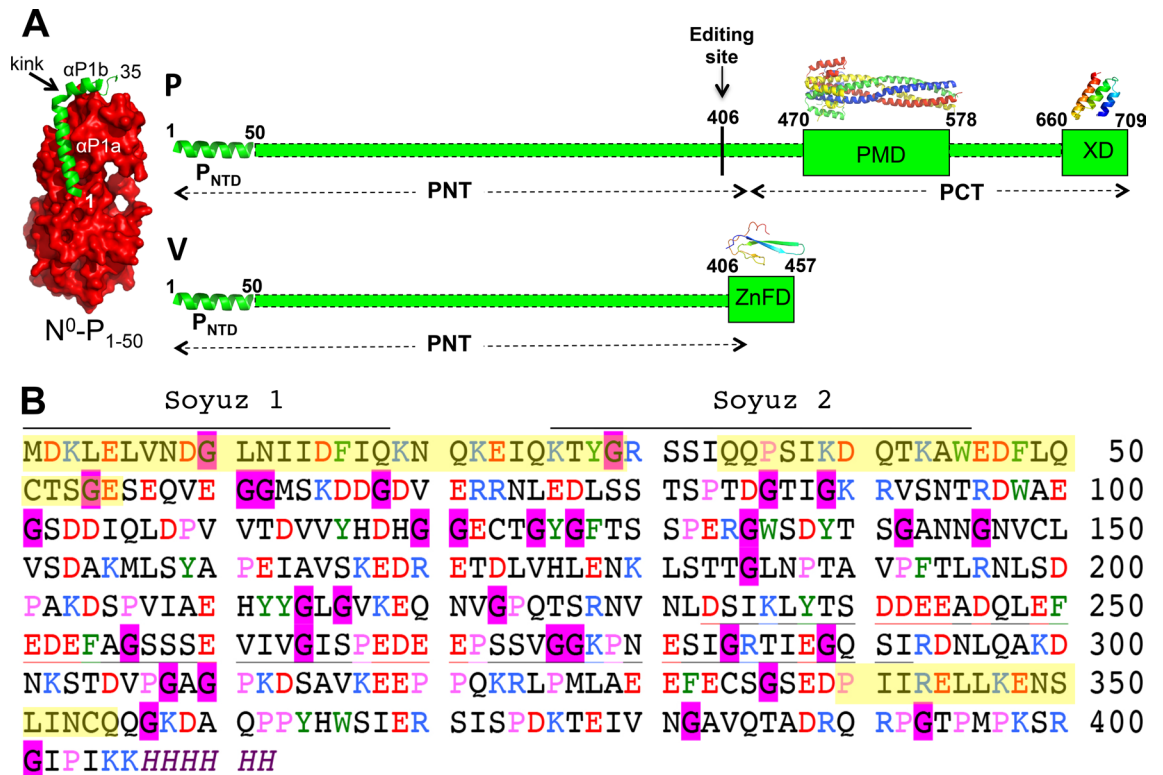
Using SAXS and NMR spectroscopy, we herein provide a high-resolution description of the intrinsically disordered N-terminal domain (PNT, aa 1–406) shared by the Nipah virus (NiV) phosphoprotein (P) and V protein, two key players in viral genome replication and in evasion of the host innate immune response, respectively. The use of multidimensional NMR spectroscopy allowed us to assign as much as 91% of the residues of this intrinsically disordered domain whose size constitutes a technical challenge for NMR studies. Chemical shifts and nuclear relaxation measurements provide the picture of a highly flexible protein. The combination of SAXS and NMR information enabled the description of the conformational ensemble of the protein in solution. The present results, beyond providing an overall description of the conformational behavior of this intrinsically disordered region, also constitute an asset for obtaining atomistic information in future interaction studies with viral and/or cellular partners. The present study can thus be regarded as the starting point towards the design of inhibitors that by targeting crucial protein–protein interactions involving PNT might be instrumental to combat this deadly virus.

The Nipah virus (NiV), together with its close relative Hendra virus (HeV), is a zoonotic paramyxovirus responsible for severe encephalitis in humans. The NiV and HeV have been classified in the *Henipavirus* genus<sup>1</sup> that also comprises the later on discovered Cedar virus (CedV)<sup>2</sup>. Because of their high pathogenic power, broad host range, high interspecies transmission and lack of therapeutics and vaccines, henipaviruses are classified as biosecurity level 4 (BSL-4) pathogens and are considered as potential bio-terrorism agents.

The genome of henipaviruses is made of a non-segmented, single-stranded RNA molecule of negative polarity that is encapsidated by a regular array of nucleoprotein (N) monomers to form a helical nucleocapsid. This N:RNA complex, and not naked RNA, is the substrate used by the viral polymerase for both transcription and replication. The viral polymerase is a complex consisting of the large (L) protein, which bears all the enzymatic activities, and the phosphoprotein (P). Through its interaction with both L and the nucleocapsid, the P protein acts as a tether and recruits L onto the N:RNA template. In addition, P also serves as a chaperon for both L<sup>3</sup> and N in that it is required for proper folding/maturation of L and maintains N in a monomeric, RNA-free form<sup>4</sup>. Therefore P is a pivotal protein endowed with multiple functions critical for both transcription and replication.

The repertoire of P functions is further expanded by the peculiar coding capacity of the P gene. Indeed, beyond the P protein, the P gene also codes for the V and W proteins that are generated through the addition of either one (protein V) or two (protein W) non-templated guanosines at the editing site of the P messenger. The addition of these guanosines triggers a downstream frame-shift. The P, V, and W proteins therefore share a

<sup>1</sup>Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy. <sup>2</sup>Department of Chemistry “Ugo Schiff”, University of Florence, Via della Lastruccia 3-13, 50019 Sesto Fiorentino, Italy. <sup>3</sup>Florence Center for Electron Nanoscopy (FloCEN), University of Florence, Via della Lastruccia 3-13, 50019 Sesto Fiorentino, Italy. <sup>4</sup>Lab. Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Aix-Marseille University and CNRS, 163 Avenue de Luminy, Case 932, Marseille, France. <sup>5</sup>These authors contributed equally: Marco Schiavina, Edoardo Salladini and Maria Grazia Murralli. ✉email: felli@cerm.unifi.it; roberta.pierattelli@unifi.it; sonia.longhi@univ-amu.fr



**Figure 1.** Modular organization of P and V, and amino acid sequence of NiV PNT. (A) Domain organization of P and V showing that P consists of two moieties, PNT and PCT, while V consists of PNT and of a zinc-finger domain (ZnFD). The P editing site is shown. Structured and disordered regions are represented as large and narrow boxes respectively. PNT: N-terminal region of P; PCT: C-terminal region of P. PMD: P multimerization domain; XD: X domain consisting of a triple  $\alpha$ -helical bundle. The  $\alpha$ -MoRE at the N-terminal region of P ( $P_{NTD}$ ), which is partly preconfigured in solution and shown to adopt a stable  $\alpha$ -helical conformation upon binding to the monomeric form of  $N^4$ , is shown as a green helix. The crystal structures of NiV PMD (PDB code 4N5B)<sup>8</sup> and of the NiV  $N^0$ - $P_{NTD}$  complex (PDB code 4CO6)<sup>4</sup> are shown. The homology-derived models of NiV XD<sup>9</sup> and of the ZnFD of V are also shown<sup>10</sup>. All structures were drawn using Pymol 2.0.1 (<https://pymol.org/2/>)<sup>11</sup>. (B) Amino acid sequence of NiV PNT (Uniprot code Q9IK91). Basic and acidic residues are shown in blue and red, respectively. Aromatic residues are shown in green, prolines in pink and glycines in black on a pink background. The soyuz1 and soyuz2 motifs conserved in *Paramyxoviridae* members are shown. The low sequence complexity region (as obtained using SEG<sup>12</sup> with a window size of 25 residues, trigger segment complexity of 3.0 and extension segment complexity of 3.3) is underlined. Fluctuating  $\alpha$ -helices, as observed in NMR studies of NiV P sub-constructs<sup>13</sup> are shown in yellow.

common N-terminal region (referred to as PNT) that constitutes a *bona fide* domain (i.e. a genuine functional unit) as inferred from the genetic organization of the P gene (Fig. 1A). The *Henipavirus* V and W proteins are key players in the evasion of the interferon (IFN)-mediated response via an antagonist activity of IFN signaling<sup>5,6</sup>. V and W bind to STAT1, a key signal transducer in the IFN-mediated antiviral response, through their common PNT region<sup>7</sup>. Binding of STAT1 by V leads to inhibition of STAT1 translocation into the nucleus, whereas binding to W leads to sequestration of STAT1 in the nucleus<sup>7</sup>. The NiV P protein is endowed with anti-IFN function as well, indicating that the PNT domain common to P, V, and W is responsible for the IFN antagonist activity.

We previously showed that PNT from both NiV and HeV is intrinsically disordered<sup>14</sup>. The P C-terminal region (PCT) consists of an alternation of disordered regions and ordered regions (Fig. 1A). In fact, PCT comprises a disordered region (aa 407–469), referred to as « spacer » and overlapping with the reading frame of the C-terminal domain of V and W, a structured region responsible for the oligomerization of P (PMD, aa 470–578)<sup>8,15</sup>, a disordered linker and a structured region (X domain, XD, aa 660–709) with a triple  $\alpha$ -helical bundle fold (Fig. 1A)<sup>9</sup>. We characterized the NiV and HeV V proteins and showed that PNT conserves its disorder also within the V protein, while the zinc-finger domain (ZnFD) has a predominant  $\beta$  conformation<sup>10</sup> in agreement with predictions<sup>16,17</sup>. In that study, we also experimentally showed that the NiV and HeV V proteins interact with DDB1—a cellular protein whose binding to V promotes STAT1 degradation—and concluded that the ZnFD plays a crucial role in strengthening this interaction<sup>10</sup>.

Although our previous study on the V proteins from both NiV and HeV<sup>10</sup> has contributed to illuminating the conformational behavior of these proteins and has provided a conceptual asset to design new antiviral strategies to combat the ability of these viruses to escape the innate immune response, a site-resolved description of these

proteins is still lacking. A major hurdle in this respect arises from the presence of their intrinsically disordered PNT region, which prevents crystallization and whose size constitutes a challenge for NMR studies.

To fill this gap in knowledge, and as a first step towards an atomic description of the NiV V protein, we decided to investigate NiV PNT by combining small-angle X-ray scattering (SAXS) and NMR spectroscopy. When we started this project, high-resolution data were already available for the first 35 residues of NiV PNT. Yabukarski and co-workers indeed solved the X-ray structure of a complex (herein referred to as N<sup>0</sup>-P<sub>1-50</sub>) made of a truncated form of NiV N unable to self-assemble (aa 32–383) and of the first 50 residues of NiV P<sup>4</sup> (Fig. 1A). Only the first 35 residues of P were defined in the electron density. This Molecular Recognition Element (MoRE) forms two  $\alpha$ -helices ( $\alpha$ P1a, residues 1–19;  $\alpha$ P1b; aa 21–28) separated by a kink (Fig. 1A). In that study, the P region encompassing the first 100 residues (P<sub>1-100</sub>) was also investigated by NMR spectroscopy on a small construct. Although the HN correlation NMR spectrum is typical of a disordered protein, the secondary-structure propensities (SSPs) calculated from Ca and C $\beta$  secondary chemical shifts indicate the presence of five fluctuating  $\alpha$ -helices. Upon addition of the N protein, only residues 50–100 remain detectable indicating that they remain flexible in the complex and that the N<sup>0</sup>-binding region of P encompasses residues 1 to 50<sup>4</sup>. This latter region contains two conserved motifs in *Paramyxoviridae* members (i.e. *soyuz1* ad *soyuz2*)<sup>18</sup> (see Fig. 1B).

When we were about to complete our NMR and SAXS characterization of NiV PNT, a study was published reporting the structural description of the NiV P protein<sup>13</sup>. In that study, the authors combined NMR spectroscopy, SAXS, and X-ray crystallography and obtained an ensemble description of this large protein. Notably, they decided to use a “*divide et impera*” approach to disentangle the NMR spectral complexity of the large disordered N-terminal region (residues 1–474). They thus generated several overlapping constructs that altogether cover the entire P protein and then assigned the HSQC spectrum of each construct corresponding to the P disordered regions and to XD. The chemical shift-derived SSPs of the fragments revealed that the long N-terminal intrinsically disordered region and the linker connecting PMD to XD dynamically sample multiple conformations while possessing short regions of residual secondary structure. The comparison between the HSQC spectrum of the full-length NiV P protein with the HSQC spectra of the individual P sub-constructs unveiled an overall good superimposition supporting a scenario where the N-terminal moiety remains highly flexible and retains a conformational behavior in the tetramer similar to that observed in the isolated sub-constructs, with negligible inter-chain contacts<sup>13</sup>.

Herein, we report the results we obtained using cutting-edge, multidimensional NMR spectroscopy approaches applied to the entire PNT domain. In spite of the challenging nature of such an assignment, we succeeded in assigning as much as 91% of the peaks of this long intrinsically disordered domain. The combination of SAXS and NMR information allowed us to obtain an ensemble description of the conformational behavior of the protein in solution.

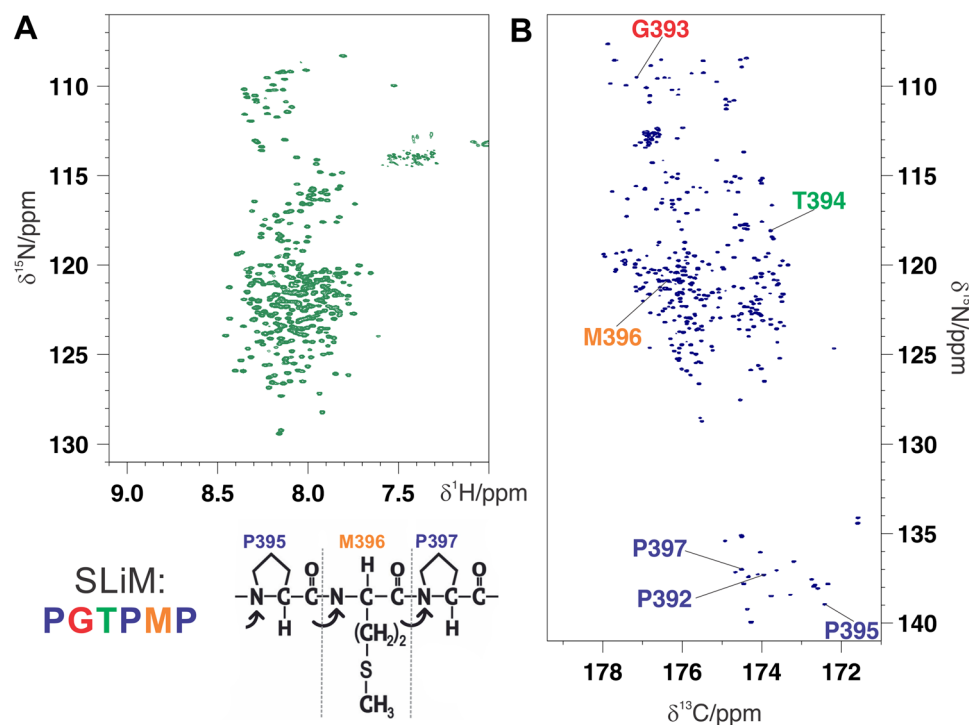
## Results and discussion

**NMR residue-specific structural and dynamic characterization of NiV PNT.** NiV PNT is a very large intrinsically disordered protein (412 residues including the C-terminal His tag and the initial methionine), as can be inferred by inspection of the two 2D spectra, the <sup>1</sup>H–<sup>15</sup>N HSQC (Fig. 2A) and the <sup>13</sup>C′–<sup>15</sup>N CON (Fig. 2B). The latter reveals also signals of proline residues and is characterized by improved chemical shift dispersion, two features that are very useful for the study of intrinsically disordered proteins (IDPs). An assignment strategy that combines <sup>13</sup>C detected with <sup>1</sup>H detected NMR is thus very helpful to provide sufficient information to enable sequence-specific assignment of the resonances of the full-length protein. A series of triple-resonance 3/5D NMR experiments, specifically designed for IDPs based either on <sup>1</sup>H<sup>N</sup> or <sup>13</sup>C detection, were acquired<sup>19,20</sup> (Supplementary Table S1). The projection reconstruction spectroscopy (APSY) approach was exploited in most of the multidimensional experiments to reduce the experimental time while preserving high spectral resolution in the indirect dimensions<sup>21,22</sup>.

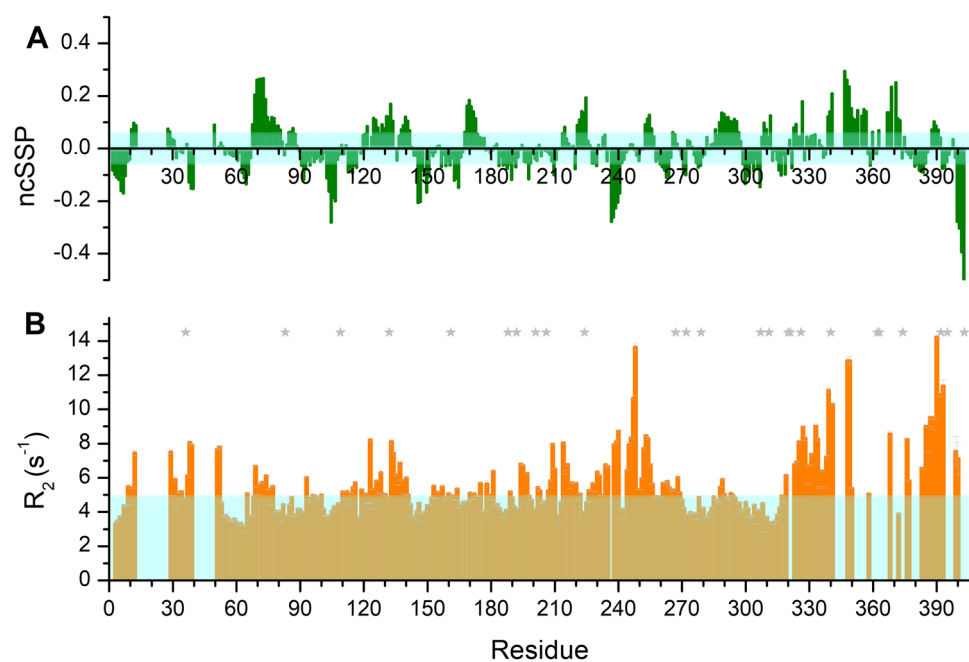
The <sup>13</sup>C′–<sup>15</sup>N CON spectrum of NiV PNT, shown in Fig. 2B, can be considered the reference spectrum for a series of multidimensional CON-based experiments in which correlations to additional nuclear spins in the indirect dimensions of the experiments provide the needed information to achieve the sequence-specific assignment. These multidimensional spectra include the <sup>13</sup>C-detected 3D (H)CBCACON<sup>23</sup>, (H)CBCANCO<sup>24</sup>, (H)COCON<sup>25,26</sup>, and 5D (HCA)CONCACON<sup>27</sup> spectra. Inspection of the spectra constituted the starting point for the sequence-specific assignment<sup>22,27</sup>.

This information was complemented with that available through <sup>1</sup>H<sup>N</sup> detected multidimensional NMR experiments. The 3D BEST-TROSY (BT) triple-resonance experiments<sup>28</sup> HN(CO)CACB, HNCACB, HNCANNH, HNCO, HN(CO)CACB, HNCACB, and HNCOCANNH were acquired to obtain the complementary information needed to complete the assignment, augmented by a 5D BT-HN(COCAN)CONH experiment<sup>29</sup> used to resolve a few ambiguities and to confirm the chemical shift assignment obtained with the 3D spectra.

We could successfully assign the resonances of H<sup>N</sup>, N, C′, C<sup>α</sup>, and C<sup>β</sup> nuclei of as many as 369 residues, including all proline residues (Fig. 2B), thus leading to a percentage of 91% of assigned residues for the entire protein. The chemical shifts of the previously mentioned nuclei are reported in Supplementary Table S2. The assignment is less complete in the N- and C-terminal moieties of the protein and in particular in the stretches encompassing residues 14–27, 41–49 and 342–346 either because of increased transverse relaxation or of residual signal overlap or both. The availability of heteronuclear chemical shifts enables the presence of residual secondary structure to be inferred<sup>30</sup>. The neighbor-corrected structural propensities (ncSSPs) were computed by comparing the experimentally measured chemical shifts of N, C′, C<sup>α</sup>, and C<sup>β</sup> nuclei to the values expected for corresponding random coils<sup>31</sup> using the tool available online at <https://st-protein02.chem.au.dk/ncSPC><sup>32</sup>. The propensities to adopt  $\alpha$ -helical and  $\beta$ -extended conformations, as obtained using the method of Mulder<sup>32</sup>, are shown in Fig. 3A



**Figure 2.** (A) 2D- $^1\text{H}^{15}\text{N}$ -BEST-TROSY and (B)  $^{13}\text{C}$ -detected CON spectra acquired on the full-length NiV PNT. The 2D CON spectrum is much more resolved and enables also the direct detection of the correlations involving the 26 prolines in a clean region of the spectrum. As an example of the importance of those residues, the cross peaks of the residues in a SLiM are reported. The complete assignment, including the  $^{13}\text{C}$  nuclei, is reported in Supplementary Table S2.



**Figure 3.** (A) Neighbor-corrected secondary structure propensity (ncSSP) values obtained through the online ncSSP tool (<https://st-protein02.chem.au.dk/ncSPC/>)<sup>32</sup> using the chemical shift values reported in Supplementary Table S2. Positive and negative values correspond to  $\alpha$ -helical and  $\beta$ -strand propensities, respectively. The light cyan box represents the  $\pm 0.05$  value. Few regions display a moderate  $\alpha$ -helix propensity (SSP > 0.2) while most of the protein is largely disordered. (B)  $^{15}\text{N}$   $R_2$  values reported as a function of residue number. The light cyan box indicates the modal value.



and confirm that the protein is intrinsically disordered, with SSP values in the  $\pm 0.2$  range for the majority of the primary sequence except for the regions encompassing residues 69–73, 237–240, 341–349, and 400–403.

Segments with SSP values between 0.05 and 0.2 indicate very modest deviations from random coil behavior. Moreover, the detection of local structural propensity in a polypeptide crucially depends on the choice of the reference chemical shifts for the random coil state and on possible bias induced by experimental conditions (e.g. buffer, pH, experimental temperature, etc.). Additional information is thus required to characterize the structural and dynamic properties of the protein in addition to nuclear chemical shifts.

To investigate backbone dynamics of NiV PNT, we measured heteronuclear  $^1\text{H}$ – $^{15}\text{N}$  NOE,  $^{15}\text{N}$   $R_2$ , and  $^{15}\text{N}$   $R_1$  values. We could successfully determine relaxation rates for most of the assigned cross-peaks observed in the 2D  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectrum and discarded those whose overlap would have rendered the analysis ambiguous. The relaxation rates show an overall trend consistent with the considerably disordered conformation of PNT.  $^1\text{H}$ – $^{15}\text{N}$  NOE values are all well below 0.5, indicating high flexibility of the backbone (Supplementary Figure S1A). Contrary to the  $^{15}\text{N}$   $R_1$  values that are rather uniform along the polypeptide chain (Supplementary Figure S1B), higher than modal  $^{15}\text{N}$   $R_2$  values were observed in several regions of the primary sequence, i.e. 9–52, 110–141, 207–220, 226–256, 261–269 and 308–406 (Fig. 3B). The segments encompassing residues 9–52 and the C-terminal part of the polypeptide chain (340–406) are regions that are partially unassigned in our experimental conditions. This observation, as well as the absence of correlations in the multidimensional spectra, confirms that these regions are affected by dynamical phenomena leading to high  $R_2$  values, such as the presence of transiently populated secondary structural elements or chemical exchange broadening. The positive SSP values for the residues that can be observed in the regions 339–357 and 367–373 suggest the presence of transiently formed  $\alpha$ -helices in these regions. The following stretch of about 30 amino acids, from residue 376 to the end of the protein primary sequence, is particularly rich in positively charged amino acids (22%), which suggests the possibility that the high nuclear relaxation rates observed in the absence of pronounced local secondary structural elements are due to the occurrence of transient electrostatic interactions (Fig. 1B and Supplementary Figure S2). Interestingly, the region encompassing residues 226–256 and 261–269, which overall shows higher than modal  $^{15}\text{N}$   $R_2$ , is rich in negatively charged residues (33%) and could indeed be engaged in transient interaction with the C-terminal, positively charged stretch (376–406).

It is also interesting to note that this latter region (i.e. 376–406) is rich in positively charged amino acids (with 7 of such residues), and also contains 4 prolines (Fig. 1B and Supplementary Figure S2). This particular amino acid composition of alternating proline residues with amino acids sharing the same charge (positive in this case) could thus have a role in promoting conformations with a local net charge even in the absence of well-defined secondary structural elements. The same pattern of regions rich in amino acids sharing the same charge and proline residues is found in other regions, such as  $^{267}\text{PEDEEP}^{272}$ , although in this case the amino acids are negatively charged. (Fig. 1B and Supplementary Figure S2). The NMR data thus suggest a possible cross-talk between regions enriched in residues with opposite charge. The 238–256 region is negatively charged and can establish interactions with positively charged regions, such as the C-terminal one encompassing residues 376–406 (Fig. 1B and Supplementary Figure S2).

Additional amino acid stretches showing higher than modal  $^{15}\text{N}$   $R_2$  values are 110–141 and 207–220. These do not show evidence of transiently populated secondary structural elements nor highly charged segments. However, looking at the amino acid composition, in particular at amino acid types that are quite rare in IDPs<sup>33</sup>, one can observe a number of aromatic residues in these regions. These account for 22% of the residues in region 110–141 and 23% in region 207–220, indicating that these regions feature a higher amount of aromatic residues with respect to what is observed for the whole protein (6%) (Fig. 1B and Supplementary Figure S2). Aromatic residues are bulky amino acids and might be involved in local transient interactions without inducing specific secondary structural elements explaining the higher  $^{15}\text{N}$  transverse relaxation rates observed, as also observed for other IDPs<sup>34</sup> in which aromatic residues have been shown to promote compact states.

Additional information that can be obtained by NMR derives from the investigation of exchange processes of amide protons with the solvent. Several CLEANEX experiments were acquired with different mixing times to measure the exchange rate constants<sup>35</sup>. Intensities of cross peaks are measured to obtain an estimation of the  $k_{\text{ex}}$  ( $\text{s}^{-1}$ ) (Supplementary Figure S1C). Thus,  $k_{\text{ex}}$  reports on solvent exchange processes of amide protons with the solvent on a per-residue basis. While this observable has been used extensively to characterize globular proteins and identify amide protons that are buried in hydrophobic cores and/or involved in hydrogen bonds through the determination of “protection factors”, in the case of IDPs it is not straightforward to interpret hydrogen exchange data in terms of structural effects as these are not easy to disentangle in a clear way from other effects such as local electrostatic potential, nature of the amino acid, and possible effects of neighboring amino acids in the primary sequence<sup>36,37</sup>. On the other hand, there is no doubt that this observable is rich in information<sup>36–38</sup> and that, as we make progress in understanding the different contributions and in predicting exchange values for random coil conformations<sup>39</sup>, it might become a useful tool to achieve information about the structural and dynamic properties of IDPs.

Inspecting the observed  $k_{\text{ex}}$  for NiV PNT as a function of the primary sequence reveals that some regions present a  $k_{\text{ex}}$  that is higher than the modal value of 1.58 Hz, while some others are well below this value. Higher values are observed for residues 27–30, between two regions with large, positive SSP values at the edges of two transiently populated helices. Another region with high  $k_{\text{ex}}$  is the one encompassing residues 119–147, rich in aromatic residues (17%; Fig. 1B and Supplementary Figure S2), where slightly higher  $^{15}\text{N}$   $R_2$  values were identified, in agreement with the presence of a solvent exposed region. Another region with significantly high rates is the one spanning residues 182–194, a region that is very rich in serine and threonine residues (38% in this segment *versus* 16% in the whole protein), two amino acids that are often solvent exposed and are characterized by high exchange rates with the solvent in general<sup>39</sup> (Fig. 1B and Supplementary Figure S2). The region presenting lower  $k_{\text{ex}}$  values (slow chemical exchange) are the regions encompassing residues 97–122 and 239–263, which are

rich in negatively charged residues (31% in the first segment, 40% in the second one) (Fig. 1B and Supplementary Figure S2), a feature that is known to cause reduced exchange processes<sup>37,40</sup>.

The NMR data clearly suggest that NiV PNT is disordered and very flexible with few short and transient secondary structure elements. The N-terminus shows moderate positive secondary structure propensity for residues 68–81, while some hints of the occurrence of other transiently populated  $\alpha$ -helices can be obtained from the combination of <sup>15</sup>N R<sub>2</sub>, SSP and k<sub>ex</sub> values in the region encompassing residues 10–30. This region contains a segment (1–29) that adopts a helix-kink-helix conformation in the crystal structure of the N<sup>0</sup>–P complex<sup>4</sup>. Subsequent NMR studies confirmed the presence of two transiently populated  $\alpha$ -helices at the N-terminus of NiV PNT also in solution and in the absence of the binding partner<sup>4,13</sup>. The presence of these helical elements mildly populated in our experimental conditions might be responsible for the conformational exchange processes likely causing the broadening of the signals in the full construct used here. Nevertheless, a few amino acids are detectable in this region and do show SSP values that indicate  $\alpha$ -helical propensity (Fig. 3A). The more efficient transverse relaxation in these regions, as compared to the studies performed on shorter constructs, could be due to subtle differences arising from intramolecular interactions.

The overall behavior of PNT in solution might also be influenced by the experimental conditions. Indeed, the buffer used in<sup>13</sup> includes compounds that mitigate protein self-association and aggregation (150 mM NaCl, 50 mM arginine, 50 mM glutamate)<sup>41</sup>. It should be pointed out however that in other cases the same buffer was reported to induce protein compaction by minimizing the effect of intramolecular long-range electrostatic interactions<sup>42</sup>. The causes of these conflicting effects are not well understood<sup>42</sup>.

Similar considerations also hold for the 340–355 region that was shown by Jensen and co-workers to partly sample an  $\alpha$ -helical conformation<sup>13</sup>. Although we could not map the core region of the transient  $\alpha$ -helix we mapped the flanking residues that, also in this case, confirm the presence of a transient  $\alpha$ -helix in this region. This transiently populated  $\alpha$ -helix may correspond to a binding site for one or multiple partners whose identification will require future studies. A possible partner could be the unassembled form of the N protein, by analogy with measles virus. Indeed, the counterpart of this region in measles virus P corresponds to a transiently populated  $\alpha$ -helix ( $\alpha$ 4, aa 190–198) that binds weakly N<sup>0</sup><sup>43</sup>.

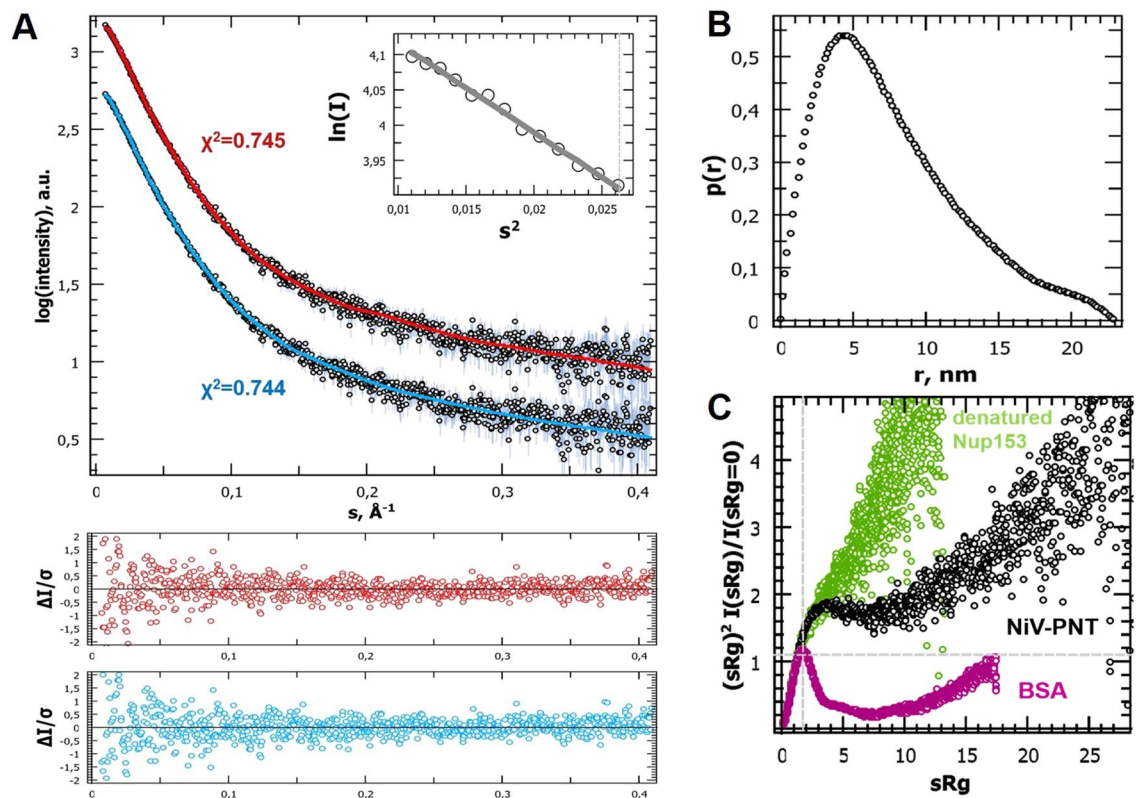
The transiently populated  $\alpha$ -helices are in good agreement with predictions as obtained with both FESS (the fast secondary structure predictor implemented in FIELDS<sup>44</sup> (Supplementary Figure S2) and PSIPRED<sup>45</sup> (Supplementary Figure S3A). In terms of predicted MoREs (see Supplementary Figure S3A), a quite good agreement was found with the experimental data, as well as with already mapped binding sites, including the STAT1 (aa 110–140)<sup>13</sup> and STAT2 (aa 110–140 and 230–237) binding sites<sup>46</sup>. The 110–140 region shown to constitute a weak binding site to STAT1<sup>13</sup> does not exhibit significant SSP values (all the values are below 0.2), a finding that suggests that this region might correspond to an I-MoRE (Irregular-MoRE), i.e. a region that may remain at least partially disordered after binding to the partner. This latter hypothesis is further corroborated by the fact that addition of STAT1 only triggers a decrease in the resonance intensities of the 110–140 region with no concomitant peak shift<sup>13</sup>.

An analysis performed on the ELM database (<https://elm.eu.org/>)<sup>47</sup> highlighted the presence of numerous Short Linear Motifs (SLiMs) along the sequence. SLiMs, also known as linear motifs (LMs), are short stretches of adjacent amino acids mediating protein–protein interactions and occurring within IDRs<sup>48</sup>. SLiMs play crucial roles in cell regulation and SLiM mimicry is often used by viruses to hijack their host cellular machinery<sup>49</sup>. A nice correlation was found between the presence of SLiMs and the dynamic behavior of NiV PNT. Indeed, 22 SLiMs were identified (Supplementary Table S3) and most of them (18) are located within the regions displaying high R<sub>2</sub> values (Supplementary Figure S3B). As an example, the previously mentioned C-terminal region (aa 376–406), enriched in proline residues and in positively charged residues, features several (i.e. 7) SLiMs. These SLiMs are mainly involved in interactions with kinases, phosphatases, SH3 and WW regulatory domains, consistent with a role in hijacking key cell regulatory processes. These interactions can now be investigated at atomic resolution through <sup>13</sup>C-detected experiments tailored for proline residues that allow acquiring quick snapshots in a clean spectral region<sup>50,51</sup>.

**SAXS ensemble characterization of NiV PNT.** Synchrotron SAXS in solution measurements were performed to gain insight into overall NiV PNT conformation and motion. Linearity in the Guinier region at low angles ( $sR_g < 1.0$ ) revealed good data quality, with no indication of protein aggregation (Fig. 4A, inset). The molecular mass determined from the forward scattering intensity at zero angle  $I(0)$  was ~47 kDa indicating a monomer in solution and in agreement with the value (45.3 kDa) computed from the amino acid sequence using the expasy server (<https://web.expasy.org/protparam/>). Although not significantly different, the R<sub>g</sub> extracted from  $P(r)$  is slightly larger (~63 Å) than the one extracted from Guinier analysis (~61 Å). The theoretical R<sub>g</sub> value expected for an IDP of the same length (412 amino acids) calculated using Flory's power law  $R_g^{IDP} = R_0 N^{\nu}$ , where R<sub>0</sub> is  $2.54 \pm 0.01$ , N is number of amino acid and  $\nu$  is  $0.522 \pm 0.001$ <sup>52</sup>, is ~59 Å. The experimentally determined R<sub>g</sub> is therefore very close to the value expected for an IDP (for comparison the R<sub>g</sub> expected for a random coil conformation would be ~72<sup>53</sup>).

$P(r)$  yields a maximal dimension, D<sub>max</sub>, of ~230 Å (Fig. 4B) with a long tail in the  $P(r)$  function, suggesting that the protein tends to assume an overall non-compact conformation<sup>54</sup>. The overall SAXS parameters for NiV PNT are listed in Supplementary Table S4.

The flexible nature of NiV PNT was qualitatively assessed by using the normalized Kratky plot, where the absence of a well-defined bell shape indicates a protein with intrinsically disordered regions (Fig. 4C). To achieve further insights into the conformational behavior of NiV PNT, data obtained from NMR and SAXS experiments were combined in search of an ensemble that quantitatively describes the conformational behavior of NiV PNT in solution. In a first step, a pool of 50,000 structures was generated based on the NMR data. To this end, the ncSSP



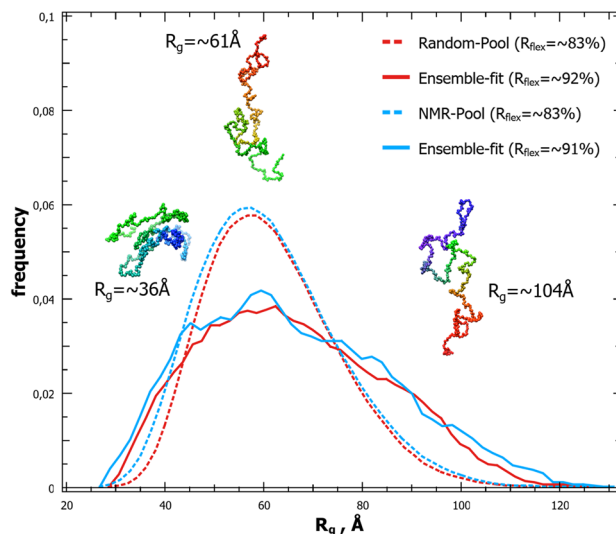
**Figure 4.** SAXS studies of NiV PNT. **(A)** SAXS scattering curve of NiV PNT and ensemble EOM 2.0 fits extracted from two distinct initial pools: (*blue*) generated by Flexible-Meccano considering ncSSP and (*red*) generated by EOM 2.0 including random Ca-only conformers. Inset: Guinier plot of the experimental scattering curve. **(B)** Pair distance distribution function,  $P(r)$ . **(C)** Normalized Kratky plot representation of the scattering data. The normalized scattering plots of a globular (BSA, SASDA32) and denatured (Nup153, SASDEY2) protein are shown for comparative purposes to illustrate that NiV PNT has an overall conformation in between these two extreme examples.

values larger than  $\pm 0.05$  were provided as inputs to Flexible-Meccano<sup>55</sup>. Subsequently, from the NMR-based pool, we used Ensemble Optimization Method (EOM) 2.0 to generate a conformational sub-ensemble that best fits the experimental SAXS data. In order to minimize over-fitting, EOM attempts at minimizing the number of conformers able to fit the experimental data and usually generates ensembles consisting of 5 to 40 conformers. Based on the high flexibility expected, no repetition of conformations in the ensemble was allowed. The scattering curve back-calculated from the selected ensemble (Fig. 4A, *cyan curve*) fits well the experimental SAXS data as judged from both  $\chi^2$  and CorMap<sup>56</sup> ( $\chi^2 = 0.744$ ,  $P = 0.101$ ). The resulting final  $R_g$  distribution, broader than the one generated from the NMR-based pool, indicates that NiV PNT exists in solution as a randomly distributed ensemble of non-compact and highly flexible conformations. Note that successive and independent selections by EOM 2.0 consistently yielded similar  $R_g$  distributions thus attesting the reproducibility of the results (data not shown). The flexibility of the ensemble was quantified as  $\text{Ensemble}_{R_{flex}} = \sim 91\%$  ( $\text{NMR-Pool}_{R_{flex}} = \sim 83\%$ ).

The average  $R_g$  value of the ensemble (66 Å) slightly exceeds the theoretical value expected from Flory's power law (59 Å). This discrepancy may reflect sequence specificities, i.e. specific sequence attributes such as proline content and charge decoration (i.e. net charge per residue, fraction of charged residues and linear distribution of opposite charges) that were shown to be major determinants of IDP conformational properties<sup>57–62</sup>. For an additional discussion of the observed discrepancy see Supplementary text.

A similar investigation was conducted by using a random pool generated by EOM 2.0 without NMR-based restraints ( $\chi^2 = 0.745$ ,  $P = 0.192$ ) (Fig. 4A, *red curve*). EOM 2.0 yielded similar results with randomness quantified as  $\text{Ensemble}_{R_{flex}} = \sim 92\%$  ( $\text{Random-Pool}_{R_{flex}} = \sim 83\%$ ) such that no significant differences was observed when using different pools (Fig. 5).

Although SAXS is poorly sensitive to secondary structure and although the presence of secondary structure elements has been reported to have a limited impact on protein compaction<sup>57,63</sup>, we sought at investigating the impact of the occurrence of the N-terminal transiently populated  $\alpha$ -helix on chain compaction. To this end, using Flexible-Meccano, we generated an initial pool in which an  $\alpha$ -helical conformation was imposed to residues 1–30, with the experimental ncSSPs. Furthermore, we also investigated the final  $R_g$  distribution while imposing a kink centered at residue 20 (as observed in the crystal structure of the  $\text{N}^0\text{-P}_{1-50}$  complex) providing NMR restraints. All the resulting sub-ensembles fit equally well the scattering data as judged from the obtained  $\chi^2$  values and even for these two cases, no differences were observed in terms of the resulting  $R_g$  distributions (Supplementary Figure S4). This probably reflects the fact that SAXS is a low-resolution technique that provides



**Figure 5.** Modeling NiV PNT as a conformational ensemble.  $R_g$  distributions of the initial pools generated by (blue dot) Flexible-Meccano (with NMR secondary structure restraints,  $\overline{R_g} = \sim 62 \text{ \AA}$ ) and (red dot) EOM 2.0 (without restraints,  $\overline{R_g} = \sim 62 \text{ \AA}$ ) respectively, and of the corresponding selected ensembles (solid lines) as obtained using EOM 2.0 (both  $\overline{R_g} = \sim 66 \text{ \AA}$ , see also Fig. 4A). Final best ensembles contain 19 unique conformers in the case of ncSSP NMR-based ensemble and 13 unique conformers in the case of EOM 2.0 random ensemble, respectively. A cartoon representation of three conformers present in the ensembles is also displayed. The structures were drawn using Pymol 2.0.1 (<https://pymol.org/2/>)<sup>11</sup>.

relevant information in terms of chain compactness but fails to capture secondary structure propensities. In addition, previous findings from others and us have highlighted that regular secondary structure does not overly contribute to protein compaction (i.e. more compact forms do not necessarily exhibit an increased content in regular secondary structure)<sup>57,63</sup>. As such, it is not surprising that the presence of transiently populated  $\alpha$ -helices in the initial ensemble does not have a significant impact on the final SAXS-derived sub-ensembles. Nevertheless, comparing SAXS-only ensembles to ensembles generated by the combined use of NMR and SAXS restraints is of interest given the growing interest towards ensemble descriptions of IDPs. In addition, combining NMR and SAXS data enables at least partly overcoming over-fitting as already discussed by Mertens and Svergun<sup>64</sup>.

## Conclusions

Albeit IDPs/IDRs are notoriously prone to undergo proteolytic degradation, we managed in generating a full-length NiV PNT sample stable enough to allow acquiring all the NMR spectra required for the assignment. Although the length of this intrinsically disordered domain is very challenging for NMR studies, the use of multidimensional NMR spectroscopy and of <sup>13</sup>C-detected experiments allowed us to assign as much as 91% of the residues including prolines. The combination of SAXS and NMR data led to ensemble models of NiV PNT showing its conformational heterogeneity.

The presence of such a large disordered region shared by both the V and P proteins is likely related to their promiscuity. Structural disorder is known to serve as a determinant of protein interactivity<sup>65–67</sup>. Indeed V interacts with various cellular partners, such as DDB1<sup>10</sup>, STAT proteins<sup>13,46,68</sup>, CRM1<sup>46,69</sup>, PLK1<sup>70</sup> and, possibly, nuclear factors such as IRF3. All these interactions play a crucial role in the ability of the virus to counteract the antiviral innate immune response of the host. The P protein binds to multiple partners as well, including N and the L protein (for a review see<sup>71</sup>). Therefore, intrinsic disorder represents an ergonomic solution for the virus to encode fewer proteins with more functions while keeping the genome size as small as possible. It is therefore not surprising to find large IDRs in proteins having a broad molecular partnership, such as the *Henipavirus* V and P proteins, and, more generally, in viral proteins<sup>72,73</sup>.

Intrinsic disorder also represents a strategy to alleviate constraints imposed by evolution on overlapping reading frames (such in the case of PNT that overlaps with the C protein) and to buffer the deleterious effect of mutations (i.e. IDRs/IDPs are more tolerant of substitutions compared to globular proteins) (for reviews see<sup>71,73</sup>). It is also well established that intrinsic disorder allows affinity and specificity to be uncoupled (i.e. it enables interactions with low affinity and decent specificity) (see<sup>33,74,75</sup> and references therein cited). However, the occurrence of residual disorder in complexes involving IDPs (in the form of fuzzy appendages) and partial preconfiguration of binding motifs before binding afford a way to attenuate the entropic penalty associated to the disorder-to-order transition thereby, ultimately, modulating the binding affinity. The involvement of IDRs in protein–protein interactions that need to be finely tuned offers an exquisite means to modulate the strength of those interactions: by tuning the extent of preconfiguration of the binding motifs and/or the length of flanking fuzzy appendages, the virus can reach an optimal binding strength. It is therefore conceivable that the presence within NiV PNT of long disordered regions flanking binding motifs, including SLiMs, as well as binding sites



to N<sup>0</sup> or STAT proteins, may exert a role in regulating these critical interactions. The flexibility of these flanking regions would enable regulating both the exposure of binding sites specifically recognized by multiple viral and cellular partners, and the affinity of these interactions. Ultimately, this would result in their ability to orchestrate virus replication, through hijacking of cellular pathways and evasion of the IFN response.

The present results constitute an asset for obtaining atomistic information in future interaction studies involving NiV PNT. Indeed, the availability of the chemical shifts of full-length NiV PNT will be instrumental to map the residues involved in binding to partners such as DDB1 and/or STAT1 in the context of either the isolated PNT domain or the V protein. Once detailed information is available from interaction studies, inhibitors can be rationally conceived. Therefore the present study can be regarded as a starting point towards the design of inhibitors abrogating the ability of this virus to escape the innate immune response. Given the high similarity between NiV and HeV PNT (56%), it is conceivable that the results of these studies could be extrapolated to some extent to the HeV as well.

## Methods

**Protein expression and purification.** The NiV PNT construct, encoding residues 1–406 of the NiV P protein with a C-terminal hexahistidine tag, has been already described<sup>14</sup>. Expression of unlabeled NiV PNT was carried out as previously described<sup>14</sup>.

Isotopically labeled (either <sup>15</sup>N–<sup>13</sup>C or <sup>15</sup>N) NiV PNT samples were prepared by growing at 37 °C transformed *E. coli* T7 cells (New England Biolabs, Ipswich, MA, USA) bearing the pRARE plasmid (Novagen, Madison, WI, USA) in LB medium supplemented with 100 µg mL<sup>-1</sup> ampicillin and 34 µg mL<sup>-1</sup> chloramphenicol. When the OD<sub>600</sub> reached 0.6, the culture was centrifuged at 4000 rpm for 10 min and the pellet was resuspended in ¼ of the initial volume of M9 medium (6 g L<sup>-1</sup> of Na<sub>2</sub>HPO<sub>4</sub>, 3 g L<sup>-1</sup> of KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> of NaCl, 0.246 g L<sup>-1</sup> of MgSO<sub>4</sub>) supplemented with 1 g L<sup>-1</sup> of <sup>15</sup>NH<sub>4</sub>Cl and 2 g L<sup>-1</sup> of either glucose or <sup>13</sup>C-glucose. After one hour at 37 °C, IPTG was added to a final concentration of 0.5 mM, and the cells were subsequently grown at 37 °C for 3.5 h.

NiV PNT was purified as described in<sup>14</sup>, except that 6 M GuHCl was added after the lysis step to both denature bacterial proteases and recover the recombinant protein also from inclusion bodies with the aim of improving protein stability and yield. After 1 h incubation at 4 °C, the sample was clarified and the supernatant was purified through immobilized metal affinity chromatography (IMAC). The fractions containing the recombinant protein were combined, and then loaded onto a Superdex 75 h 16/60 column (GE, Healthcare). The elution buffer was 10 mM sodium phosphate pH 6.5, supplemented with 5 mM EDTA and 5 mM DTT. The fractions containing the protein were collected and conserved at –20 °C.

**NMR spectroscopy experiments.** Immediately before NMR studies, the NiV PNT sample was thawed, centrifuged at 12,000 rpm to remove any possible protein aggregate and then concentrated using 30 kDa Amicon Ultra Centrifugal Filters (Merk Millipore, Darmstadt, Germany). All NMR spectra for assignment were acquired on <sup>13</sup>C–<sup>15</sup>N-isotopically enriched NiV PNT at a concentration of about 200 µM.

The <sup>1</sup>H-detected spectra were acquired at 288 K with a 22.3 T Bruker Avance III 950 NMR spectrometer; the <sup>13</sup>C-detected spectra were acquired at 288 K with a 16.4 T Bruker Avance NEO 700 NMR spectrometer equipped with a cryogenically cooled probehead optimized for <sup>13</sup>C-direct detection. Experimental details are reported in Supplementary Table S1.

All the heteronuclear relaxation experiments (R<sub>1</sub>, R<sub>2</sub> and <sup>1</sup>H–<sup>15</sup>N NOEs) were acquired using <sup>15</sup>N labeled NiV PNT at about 100 µM. The spectra were recorded at 288 K on a 16.4 T Bruker Avance NEO 700 NMR spectrometer equipped with a cryogenically cooled triple resonance probehead. The <sup>15</sup>N R<sub>1</sub> and R<sub>2</sub> experiments were acquired with 8 scans (2048 × 256 points) and a relaxation delay of 3.0 s. To determine the <sup>15</sup>N R<sub>1</sub> the following delays were used: 20 ms, 60 ms, 120 ms, 180 ms, 250 ms, 400 ms, 500 ms, 600 ms, 750 ms, and 900 ms. To determine the <sup>15</sup>N R<sub>2</sub> the following delays were used: 32 ms, 64 ms, 96 ms, 128 ms, 160 ms, 190 ms, 260 ms, 320 ms, 380 ms, 440 ms, and 500 ms. The <sup>1</sup>H–<sup>15</sup>N NOEs experiments were acquired with 64 scans (2048 × 288 points) and a relaxation delay of 6.0 s.

The CLEANEX experiments were acquired on a 100 µM sample, with 16 scans (2048 × 248 points) and a relaxation delay of 3.0 s. The following delays were used: 5 ms, 10 ms, 20 ms, and 30 ms.

**NMR data processing and analysis.** NMR data sets were processed using the Bruker TopSpin 4.0.6 software. CARA<sup>76</sup> and its tool XEASY<sup>77</sup> were used to analyze and annotate the spectra.

The <sup>15</sup>N relaxation rates (R<sub>1</sub> and R<sub>2</sub>) were determined by fitting the cross-peak intensity measured as a function of variable delay, to single-exponential decay using the Bruker Dynamic Center 2.4, available as a stand-alone ancillary software of TopSpin by Bruker. <sup>1</sup>H–<sup>15</sup>N NOE values were obtained as the ratio between peak intensity in spectra recorded with and without <sup>1</sup>H saturation.

The secondary structure propensity (SSP) from heteronuclear chemical shifts was determined by using the neighbor corrected structural propensity calculator (ncSPC) tool<sup>32</sup> available online at <https://st-protein02.chem.au.dk/ncSPC>. The Mulder random coil chemical shift library<sup>31</sup> was chosen for the analysis, and the average window size was left to the standard value of 5.

**SAXS experiments and analysis.** Samples were concentrated using 30 kDa Amicon Ultra Centrifugal Filters (Merk Millipore, Darmstadt, Germany) and loaded onto a Superdex 75 h 16/60 column using 20 mM Tris/HCl pH 8.0, 0.3 M NaCl, 5 mM DTT as elution buffer. Synchrotron X-ray scattering data were collected at ESRF BM29 beamline (Grenoble) using a PILATUS 1 M pixel detector (DÉCTRIS, Baden, Switzerland) at a sample-detector distance of 2.87 m and a wavelength of 0.0992 nm. This setup covers a range of momentum transfer of 0.028 < s < 4.525 nm<sup>-1</sup> (s = 4π sin(θ)/λ, where 2θ is the scattering angle). Samples were loaded using a

robotic sample changer<sup>78</sup> and measured at 20 °C at concentrations ranging from 0.5 to 1.2 mg/mL. 10 independent frames of 1 s each were collected for each sample and data were automatically reduced using an in-house pipeline. Further analysis was done using ATSAS 3.0.1<sup>79</sup>. As no concentration dependence was observed (Supplementary Figure S5), the curve with less noise (at 1.2 mg/mL) was used for the SAXS analysis. The forward scattering  $I(0)$  as well as the  $R_g$  were calculated using the Guinier approximation assuming that, at very small angles ( $s < 1.3/R_g$ ), the intensity is represented as  $I(s) = I(0) \cdot \exp(-(sR_g)^2/3)$ <sup>80,81</sup>. The forward scattering intensities were calibrated using water as reference. Linearity in the Guinier region was used to exclude sample aggregation, and the pair-distance distribution function,  $P(r)$ , from which the  $D_{max}$  and the  $R_g$  were estimated, was computed using GNOM<sup>82</sup>. Qualitative assessment of compactness versus structural disorder was made by transforming the scattering profiles in the so-called normalised Kratky representation [ $(sR_g)^2 * I(s)/I(0)$  vs.  $sR_g$ ]<sup>83</sup>. The MM was derived by placing the scattered intensity on an absolute scale using liquid water as calibrant.

Protein flexibility was quantified using EOM 2.0<sup>84</sup> that assumes coexistence of a range of  $\alpha$ -only conformations in solution for which an average scattering intensity fits the experimental data. A genetic algorithm (GA) is used to select ensembles with varying numbers of conformers. The GA is repeated  $n$  independent times, and the ensemble with the lowest discrepancy considered as the best solution out of  $n$  final ensembles. Furthermore, repetition of GA allows the computation of  $R_g$  distributions so that structural information about the flexibility could be extracted. The width of the distribution is used to derive the flexibility of the particles, whereby a narrow distribution indicates a rather rigid particle and broader distributions are associated with higher flexibility. Using EOM 2.0, systematic quantification of the flexibility was made by using the metric  $R_{flex}$ —which computes the Shannon information entropy of the distributions<sup>84</sup>. Experimental error-independent goodness-of-fit was also confirmed by using the software CorMap that estimates the differences between one-dimensional spectra independently of explicit error estimates, using only data point correlations<sup>86</sup>. All the softwares used for the SAXS data analysis are part of the ATSAS 3.0 package<sup>79</sup>.

**Accession numbers.** The chemical shifts have been deposited in the Biological Magnetic Resonance Data Bank (BMRB) under the code 50370. SAXS data have been deposited in the Small Angle Scattering Biological Data Bank (SASBDB)<sup>85</sup> under code SASDBJ5. The ensemble derived using SAXS constraints only and the one derived from the combined use of SAXS and NMR constraints have been deposited within the Protein Ensemble Database (PED-DB, <https://proteinensemble.org/>)<sup>86</sup> under accession numbers PED00176 and PED00177, respectively. All-atoms pdb files, required for deposition in the PED-DB, were generated from EOM  $\alpha$ -only pdb files using the PD2ca2main server ([http://www.sbg.bio.ic.ac.uk/~phyre2/PD2\\_ca2main/](http://www.sbg.bio.ic.ac.uk/~phyre2/PD2_ca2main/))<sup>87</sup>.

Received: 1 July 2020; Accepted: 27 October 2020

Published online: 11 November 2020

## References

- Wang, L. F. *et al.* The exceptionally large genome of Hendra virus: Support for creation of a new genus within the family Paramyxoviridae. *J. Virol.* **74**, 9972–9979 (2000).
- Marsh, G. A. *et al.* Cedar virus: A novel Henipavirus isolated from Australian bats. *PLoS Pathog.* **8**, e1002836 (2012).
- Bloyet, L. M. *et al.* HSP90 chaperoning in addition to phosphoprotein required for folding but not for supporting enzymatic activities of measles and Nipah virus L polymerases. *J. Virol.* **90**, 6642–6656 (2016).
- Yabukarski, F. *et al.* Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nat. Struct. Mol. Biol.* **21**, 754–759 (2014).
- Park, M. S. *et al.* Newcastle disease virus (NDV)-based assay demonstrates interferon-antagonist activity for the NDV V protein and the Nipah virus V, W, and C proteins. *J. Virol.* **77**, 1501–1511 (2003).
- Fontana, J. M., Bankamp, B. & Rota, P. A. Inhibition of interferon induction and signaling by paramyxoviruses. *Immunol. Rev.* **225**, 46–67 (2008).
- Shaw, M. L., Garcia-Sastre, A., Palese, P. & Basler, C. F. Nipah virus V and W proteins have a common STAT1-binding domain yet inhibit STAT1 activation from the cytoplasmic and nuclear compartments, respectively. *J. Virol.* **78**, 5633–5641 (2004).
- Bruhn-Johannsen, J. F. *et al.* Crystal structure of the Nipah virus phosphoprotein tetramerization domain. *J. Virol.* **88**, 758–762 (2014).
- Habchi, J. *et al.* Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipaviruses. *J. Biol. Chem.* **286**, 13583–13602 (2011).
- Salladini, E., Delazun, V. & Longhi, S. The Henipavirus V protein is a prevalently unfolded protein with a zinc-finger domain involved in binding to DDB1. *Mol. Biosyst.* **13**, 2254–2267 (2017).
- DeLano, W. L. The PyMOL molecular graphics system. *Proteins: Struct. Funct. Bioinf.* **30**, 442–454 (2002).
- Wootton, J. C. Non-globular domains in protein sequences: Automated segmentation using complexity measures. *Comput. Chem.* **18**, 269–285 (1994).
- Jensen, M. R. *et al.* Structural description of the Nipah virus phosphoprotein and its interaction with STAT1. *Biophys. J.* **118**, 2470–2488 (2020).
- Habchi, J., Mamelli, L., Darbon, H. & Longhi, S. Structural disorder within henipavirus nucleoprotein and phosphoprotein: From predictions to experimental assessment. *PLoS ONE* **5**, e11684 (2010).
- Blocquel, D., Beltrandi, M., Eroles, J., Barbier, P. & Longhi, S. Biochemical and structural studies of the oligomerization domain of the Nipah virus phosphoprotein: Evidence for an elongated coiled-coil homotrimer. *Virology* **446**, 162–172 (2013).
- Karlin, D., Ferron, F., Canard, B. & Longhi, S. Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* **84**, 3239–3252 (2003).
- Li, T., Chen, X., Garbutt, K. C., Zhou, P. & Zheng, N. Structure of DDB1 in complex with a paramyxovirus V protein: Viral hijack of a propeller cluster in ubiquitin ligase. *Cell* **124**, 105–117 (2006).
- Karlin, D. & Belshaw, R. Detecting remote sequence homology in disordered proteins: Discovery of conserved motifs in the N-termini of Mononegavirales phosphoproteins. *PLoS ONE* **7**, e31719 (2012).

19. Brutscher, B. *et al.* NMR methods for the study of intrinsically disordered proteins structure, dynamics, and interactions: General overview and practical guidelines. *Adv. Exp. Med. Biol.* **870**, 49–122 (2015).
20. Felli, I. C. & Pierattelli, R. Recent progress in NMR spectroscopy: Toward the study of intrinsically disordered proteins of increasing size and complexity. *IUBMB Life* **64**, 473–481 (2012).
21. Hiller, S., Fiorito, F., Wuthrich, K. & Wider, G. Automated projection spectroscopy (APSY). *Proc. Natl. Acad. Sci. USA* **102**, 10876–10881 (2005).
22. Murralli, M. G. *et al.* <sup>13</sup>C APSY-NMR for sequential assignment of intrinsically disordered proteins. *J. Biomol. NMR* **70**, 167–175 (2018).
23. Bermel, W. *et al.* Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J. Am. Chem. Soc.* **128**, 3918–3919 (2006).
24. Bermel, W. *et al.* H-start for exclusively heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins. *J. Magn. Reson.* **198**, 275–281 (2009).
25. Felli, I. C., Pierattelli, R., Glaser, S. J. & Luy, B. Relaxation-optimised Hartmann–Hahn transfer using a specifically Tailored MOCCA-XY16 mixing sequence for carbonyl–carbonyl correlation spectroscopy in <sup>13</sup>C direct detection NMR experiments. *J. Biomol. NMR* **43**, 187–196 (2009).
26. Mateos, B., Konrat, R., Pierattelli, R. & Felli, I. C. NMR characterization of long-range contacts in intrinsically disordered proteins from paramagnetic relaxation enhancement in <sup>13</sup>C direct-detection experiments. *ChemBioChem* **20**, 335–339 (2019).
27. Bermel, W. *et al.* High-dimensionality <sup>13</sup>C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J. Biomol. NMR* **57**, 353–361 (2013).
28. Solyom, Z. *et al.* BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J. Biomol. NMR* **55**, 311–321 (2013).
29. Piai, A. *et al.* “CON-CON” assignment strategy for highly flexible intrinsically disordered proteins. *J. Biomol. NMR* **60**, 209–218 (2014).
30. Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: Implications for fibrillation. *Protein Sci.* **15**, 2795–2804 (2006).
31. Tamiola, K., Acar, B. & Mulder, F. A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* **132**, 18000–18003 (2010).
32. Tamiola, K. & Mulder, F. A. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem. Soc. Trans.* **40**, 1014–1020 (2012).
33. Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing protein intrinsic disorder. *Chem. Rev.* **114**, 6561–6588 (2014).
34. Mateos, B. *et al.* The ambivalent role of proline residues in an intrinsically disordered protein: From disorder promoters to compaction facilitators. *J. Mol. Biol.* **432**, 3093–3111 (2020).
35. Hwang, T. L., van Zijl, P. C. & Mori, S. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* **11**, 221–226 (1998).
36. Chevelkov, V., Xue, Y., Rao, D., Forman-Kay, J. & Skrynnikov, N. <sup>15</sup>N H/D-SOLEXSY experiment for accurate measurement of amide solvent exchange rates: Application to denatured drkN SH3. *J. Biomol. NMR* **46**, 227–244 (2010).
37. Croke, R., Sallum, C., Watson, E., Watt, E. & Alexandrescu, A. Hydrogen exchange of monomeric alpha-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in *Escherichia coli*. *Protein Sci.* **17**, 1434–1445 (2008).
38. Smith, A., Zhou, L. & Pielak, G. Hydrogen exchange of disordered proteins in *Escherichia coli*. *Protein Sci.* **24**, 706–713 (2015).
39. Bai, Y., Milne, J., Mayne, L. & Englander, S. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).
40. Pontoriero, L., Schiavina, M., Murralli, M., Pierattelli, R. & Felli, I. Monitoring the interaction of alpha-synuclein with calcium ions through exclusively heteronuclear nuclear magnetic resonance experiments. *Angew. Chem. Int. Ed. Engl.* **59**, 18537–18545 (2020).
41. Golovanov, A., Hautbergue, G., Wilson, S. & Lian, L. A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc.* **126**, 8933–8939 (2004).
42. Blobel, J. *et al.* Protein loop compaction and the origin of the effect of arginine and glutamic acid mixtures on solubility, stability and transient oligomerization of proteins. *Eur. Biophys. J.* **40**, 1327–1338 (2011).
43. Milles, S. *et al.* An ultraweak interaction in the intrinsically disordered replication machinery is essential for measles virus function. *Sci. Adv.* **4**, eaat7778 (2018).
44. Piovesan, D., Walsh, I., Minervini, G. & Tosatto, S. C. E. FIELDS: Fast estimator of latent local structure. *Bioinformatics* **33**, 1889–1891 (2017).
45. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
46. Rodriguez, J. J., Cruz, C. D. & Horvath, C. M. Identification of the nuclear export signal and STAT-binding domains of the Nipah virus V protein reveals mechanisms underlying interferon evasion. *J. Virol.* **78**, 5358–5367 (2004).
47. Kumar, M. *et al.* ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res.* **48**, D296–D306 (2020).
48. Van Roey, K. *et al.* Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* **114**, 6733–6778 (2014).
49. Davey, N. E., Trave, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends Biochem. Sci.* **36**, 159–169 (2011).
50. Murralli, M. G., Piai, A., Bermel, W., Felli, I. C. & Pierattelli, R. Proline fingerprint in intrinsically disordered proteins. *ChemBioChem* **19**, 1625–1629 (2018).
51. Schiavina, M. *et al.* Taking simultaneous snapshots of intrinsically disordered proteins in action. *Biophys. J.* **117**, 46–55 (2019).
52. Bernado, P. & Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **8**, 151–167 (2012).
53. Fitzkee, N. & Rose, G. Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. USA* **101**, 12497–12502 (2004).
54. Bernado, P. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur. Biophys. J.* **39**, 769–780 (2010).
55. Ozenne, V. *et al.* Flexible-meccano: A tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **28**, 1463–1470 (2012).
56. Franke, D., Jeffries, C. M. & Svergun, D. Correlation map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods* **12**, 419–422 (2015).
57. Marsh, J. A. & Forman-Kay, J. D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **98**, 2383–2390 (2010).
58. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA* **107**, 8183–8188 (2010).
59. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **110**, 13392–13397 (2013).
60. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32**, 102–112 (2015).

61. Tedeschi, G. *et al.* Conformational response to charge clustering in synthetic intrinsically disordered proteins. *Biochim. Biophys. Acta Gen. Subj.* **1862**, 2204–2214 (2018).
62. Bianchi, G., Longhi, S., Grandori, R. & Brocca, S. Relevance of electrostatic charges in compactness, aggregation, and phase separation of intrinsically disordered proteins. *Int. J. Mol. Sci.* **21**, E6208 (2020).
63. Blocquel, D., Habchi, J., Gruet, A., Blangy, S. & Longhi, S. Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies. *Mol. Biosyst.* **8**, 392–410 (2012).
64. Mertens, H. D. T. & Svergun, D. I. Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Arch. Biochem. Biophys.* **628**, 33–41 (2017).
65. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets. *FEBS J.* **272**, 5129–5148 (2005).
66. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **18**, 343–384 (2005).
67. Haynes, C. *et al.* Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput. Biol.* **2**, e100 (2006).
68. Rodriguez, J. J., Parisien, J. P. & Horvath, C. M. Nipah virus V protein evades alpha and gamma interferons by preventing STAT1 and STAT2 activation and nuclear accumulation. *J. Virol.* **76**, 11476–11483 (2002).
69. Rodriguez, J. J., Wang, L. F. & Horvath, C. M. Hendra virus V protein inhibits interferon signaling by preventing STAT1 and STAT2 nuclear accumulation. *J. Virol.* **77**, 11842–11845 (2003).
70. Ludlow, L. E., Lo, M. K., Rodriguez, J. J., Rota, P. A. & Horvath, C. M. Henipavirus V protein association with Polo-like kinase reveals functional overlap with STAT1 binding and interferon evasion. *J. Virol.* **82**, 6259–6271 (2008).
71. Longhi, S., Bloyet, L. M., Gianni, S. & Gerlier, D. How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication. *Cell. Mol. Life Sci.* **74**, 3091–3118 (2017).
72. Uversky, V. N. & Longhi, S. *Flexible Viruses: Structural Disorder in Viral Proteins* (Wiley, New York, 2012).
73. Xue, B. *et al.* Structural disorder in viral proteins. *Chem. Rev.* **114**, 6880–6911 (2014).
74. Bugge, K. *et al.* Interactions by disorder—A matter of context. *Front. Mol. Biosci.* **7**, 110 (2020).
75. Olsen, J., Teilum, K. & Kragelund, B. Behaviour of intrinsically disordered proteins in protein–protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci.* **74**, 3175–3183 (2017).
76. Keller, R. *The Computer Aided Resonance Assignment Tutorial* (Cantina Verlag, Goldau, 2004).
77. Bartels, C., Xia, T. H., Biller, M., Guntert, P. & Wuthrich, K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **6**, 1–10 (1995).
78. Round, A. R. *et al.* Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. *J. Appl. Crystallogr.* **41**, 913–917 (2008).
79. Franke, D. *et al.* ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
80. Guinier, A. La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Ann. Phys. (Paris)* **12**, 161–237 (1939).
81. Guinier, A. & Fournet, F. *Small Angle Scattering of X-rays* (Wiley Interscience, New York, 1955).
82. Svergun, D. Determination of the regularization parameters in indirect-trasform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495–503 (1992).
83. Durand, D. *et al.* NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45–53 (2010).
84. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* **2**, 202–217 (2015).
85. Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M. & Svergun, D. I. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **43**, D357–363 (2015).
86. Lazar, T. *et al.* PED in 2021: A major update of the protein ensemble database for intrinsically disordered proteins. *Nucl. Acid Res.* (Database issue) in press, (2021).
87. Moore, B. L., Kelley, L. A., Barber, J., Murray, J. W., MacDonald, J. T. High-quality protein backbone reconstruction from alpha carbons using Gaussian mixture models. *J. Comput. Chem.* **34**, 1881–1889 (2013).

## Acknowledgements

This work was carried out with the financial support of the CNRS and of the Fondazione CR Firenze to R.P. E. S. was supported by a joint doctoral fellowship from the Direction Générale de l'Armement (DGA) and Aix-Marseille University. The authors acknowledge the support of the Italian government programme “MIUR—Dipartimenti di Eccellenza 2018-2022” to the Department of Chemistry “Ugo Schiff” of the University of Florence. The authors acknowledge also the support and the use of resources of Instruct-ERIC, a Landmark ESFRI, through the R&D pilot scheme and iNEXT (H2020 Grant # 653706) for providing access to the CERM research infrastructure. The authors are thankful to Julien Perard for his help in SAXS data collection, the ESRF synchrotron for beamtime allocation and to Gerlind Sulzenbacher (AFMB lab) for efficiently managing the AFMB BAG.

## Author contributions

S.L., R.P., and I.C.F. conceived and planned the experiments. E.S. purified and labeled the protein for NMR and SAXS experiments, and performed all the SAXS measurements. E.S. and M.G.M. acquired all the NMR spectra. E.S., M.G.M. and M.S. performed the sequence-specific assignment and analyzed the data under the guidance of R.P. and I.C.F. G.T. processed and analyzed SAXS data. All the authors analyzed the data. S.L. wrote the first draft of the paper, and all the authors contributed to the writing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-76522-3>.

**Correspondence** and requests for materials should be addressed to I.C.F., R.P. or S.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



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



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

© The Author(s) 2020