

## Molecular determinants of fibrillation in a viral amyloidogenic domain from combined biochemical and biophysical studies

Juliet F. Nilsson<sup>1</sup>, Hakima Baroudi<sup>1</sup>, Frank Gondelaud<sup>1</sup>, Giulia Pesce<sup>1</sup>, Christophe Bignon<sup>1</sup>, Denis Pchelkine<sup>1</sup>, Joseph Chamieh<sup>2</sup>, Hervé Cottet<sup>2</sup>, Andrey V. Kajava<sup>3</sup>, and Sonia Longhi<sup>1\*</sup>

<sup>1</sup> Laboratoire Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Aix-Marseille University and Centre National de la Recherche Scientifique (CNRS), 163 Avenue de Luminy, Case 932, 13288 Marseille CEDEX 09, France; [juliet.nilsson@univ-amu.fr](mailto:juliet.nilsson@univ-amu.fr) (J.F.N.); [frank.gondelaud@univ-amu.fr](mailto:frank.gondelaud@univ-amu.fr) (F.G.); [hakima.baroudi@univ-amu.fr](mailto:hakima.baroudi@univ-amu.fr) (H.B.); [giulia.pesce@univ-amu.fr](mailto:giulia.pesce@univ-amu.fr) (G.P.); [christophe.bignon@univ-amu.fr](mailto:christophe.bignon@univ-amu.fr) (C.B.); [denis.PTCHELKINE@univ-amu.fr](mailto:denis.PTCHELKINE@univ-amu.fr) (D.P.); [sonia.longhi@univ-amu.fr](mailto:sonia.longhi@univ-amu.fr) (S.L.)

<sup>2</sup> IBMM, Univ Montpellier, CNRS, ENSCM, 34095 Montpellier, France; [joseph.chamieh@umontpellier.fr](mailto:joseph.chamieh@umontpellier.fr) (J.C.); [herve.cottet@umontpellier.fr](mailto:herve.cottet@umontpellier.fr) (H.C.)

<sup>3</sup> Centre de Recherche en Biologie Cellulaire de Montpellier, UMR 5237, CNRS, Université Montpellier, 34293 Montpellier, France; [andrey.kajava@crbm.cnrs.fr](mailto:andrey.kajava@crbm.cnrs.fr) (A.K.)

\*to whom correspondence should be sent

### Abstract

The Nipah and Hendra viruses (NiV and HeV) are biosafety level 4 human pathogens classified within the *Henipavirus* genus of the *Paramyxoviridae* family. In both NiV and HeV, the gene encoding the Phosphoprotein (P protein), an essential polymerase cofactor, also encodes the V and W proteins. These three proteins, which share an intrinsically disordered N-terminal domain (NTD) and have unique C-terminal domains (CTD), are all known to counteract the host innate immune response, with V and W acting by either counteracting or inhibiting Interferon (IFN) signaling. Recently, using a combination of biophysical and structural approaches, the ability of a short region within the shared NTD (PNT3 region) to form amyloid-like structures was reported. Here, we evaluated the relevance of each of three contiguous tyrosine residues located in a previously identified amyloidogenic motif (EYYY) within HeV PNT3 to the fibrillation process. Our results indicate that removal of a single tyrosine in this motif significantly decreases the ability to form fibrils independently of position, mainly affecting the elongation phase. In addition, we show that the C-terminal half of PNT3 has an inhibitory effect on fibril formation that may act as a molecular shield and could thus be a key domain in the regulation of PNT3 fibrillation. Finally, the kinetics of fibril formation for the two PNT3 variants with highest and lowest fibrillation propensity were studied by Taylor Dispersion Analysis (TDA). The results herein presented shed light onto the molecular mechanisms involved in fibril formation. In addition, the PNT3 variants we generated represent valuable tools to further explore the functional impact of V/W fibrillation in transfected and infected cells.

**Keywords:** Paramyxoviruses, Hendra virus, intrinsically disordered proteins, amyloid-like fibrils, Taylor Dispersion Analysis (TDA), Negative Staining Transmission Electron Microscopy (ns-TEM), PEG precipitation assays, Congo Red, Small-angle X-ray scattering (SAXS).

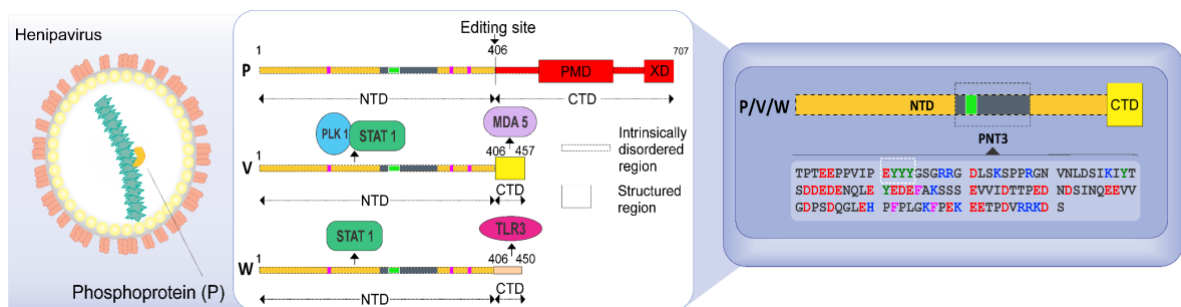
## 1. Introduction

Hendra virus (HeV), together with the closely related Nipah virus (NiV), is a Biosafety Level 4 (BSL-4) pathogen belonging to the *Henipavirus* genus within the *Paramyxoviridae* family [1]. Henipaviruses are zoonotic viruses responsible in humans for severe encephalitis [1]. They are enveloped viruses with a non-segmented, single-stranded RNA genome of negative polarity [2]. Their genome is wrapped by the nucleoprotein (N) within a helical nucleocapsid that is the template used by the viral polymerase for transcription and replication. The polymerase consists of the L protein, which bears all the enzymatic activities, and of the phosphoprotein (P). P serves as an indispensable polymerase co-factor as not only it tethers the L protein onto the nucleocapsid, but also keeps L in a soluble and competent form for transcription and replication [3-5].

The N and P proteins from henipaviruses encompass long intrinsically disordered regions (IDRs) [6-8], *i.e.* regions devoid of stable secondary and tertiary structure [9-12]. The *Henipavirus* P protein consists of a long N-terminal intrinsically disordered domain (NTD) and a C-terminal region that possesses both structured and disordered regions (**Fig. 1**) [7,8,13-18].

As in many paramyxoviruses [19], the P gene from HeV and NiV also encodes the C, V and W non-structural proteins. While the C protein is encoded from an alternative reading frame within the P gene, the V and W proteins (~50 kDa) result from a mechanism of co-transcriptional editing of the P messenger: the addition of either one or two non-templated guanosines at the editing site of the P messenger yields the V and W proteins, respectively (**Fig. 1**). The editing site is located at the end of the NTD-encoding region (**Fig. 1**). The P, V and W proteins therefore share a common NTD but have distinct C-terminal domains (CTDs) (**Fig. 1**). While the CTD of V adopts a zinc-finger conformation [20], the CTD of W is disordered [21].

The V and W proteins are key players in the evasion of the antiviral type I interferon (IFN- $\lambda$ )-mediated response [22-24]. This property relies on their ability to bind to a number of key cellular proteins involved in the antiviral response (**Fig. 1**).



**Figure 1.** Schematic illustration of the HeV particle, of the organization of the P, V and W proteins and sequence of the HeV PNT3 region shared by the P, V, and W proteins. The left panel displays a scheme of the HeV virion, with the genome encapsidated by the nucleoprotein (in green) and the P (yellow) and the Large (beige) proteins attached onto the nucleocapsid. The central panel displays a scheme of the P, V and W protein organization, showing that they share a common N-terminal domain (NTD) and have distinct C-terminal domains (CTD). PMD: P multimerization domain; XD: X domain of P. Interaction sites with proteins associated with the host innate immune response are shown. STAT: Transducers and Activators of Transcription; PLK1: Polo-Like Kinase 1; MDA5: Melanoma Differentiation-Associated protein 5; TLR3: Toll-like receptor 3. Pink bars correspond to cysteine residues within NTD. The localization and the sequence of the PNT3 region within the NTD is shown in grey and its sequence is displayed in the blue, right panel. The EYYY motif is shown as a green square and is underlined within the PNT3 sequence.

We previously reported the ability of the HeV V protein to undergo a liquid-to-gel transition, with a region within the NTD (referred to as PNT3, aa 200-310) (**Fig. 1**) being identified as responsible for this behavior [25]. In those previous studies, we characterized PNT3 using a combination of biophysical and structural approaches. Congo red binding assays, together with negative-staining transmission electron microscopy (ns-TEM) studies, showed that PNT3 forms amyloid-like structures [25]. Noteworthy, Congo red staining experiments provided hints that these amyloid-like fibrils form not only *in vitro* but also *in cellula* after transfection or infection

suggesting a probable functional role. In light of the critical role of the *Henipavirus* V and W proteins in evading the host innate immune response, we previously proposed that in infected cells PNT3-mediated fibrillar aggregates could sequester key cellular proteins involved in the antiviral response such as STAT and 14-3-3 proteins, thereby preventing IFN signaling and abrogating NF- $\kappa$ B-induced proinflammatory response [26]. Consistent with the presence of PNT3, the *Henipavirus* W proteins were shown to be able to form amyloid-like fibrils as well [21].

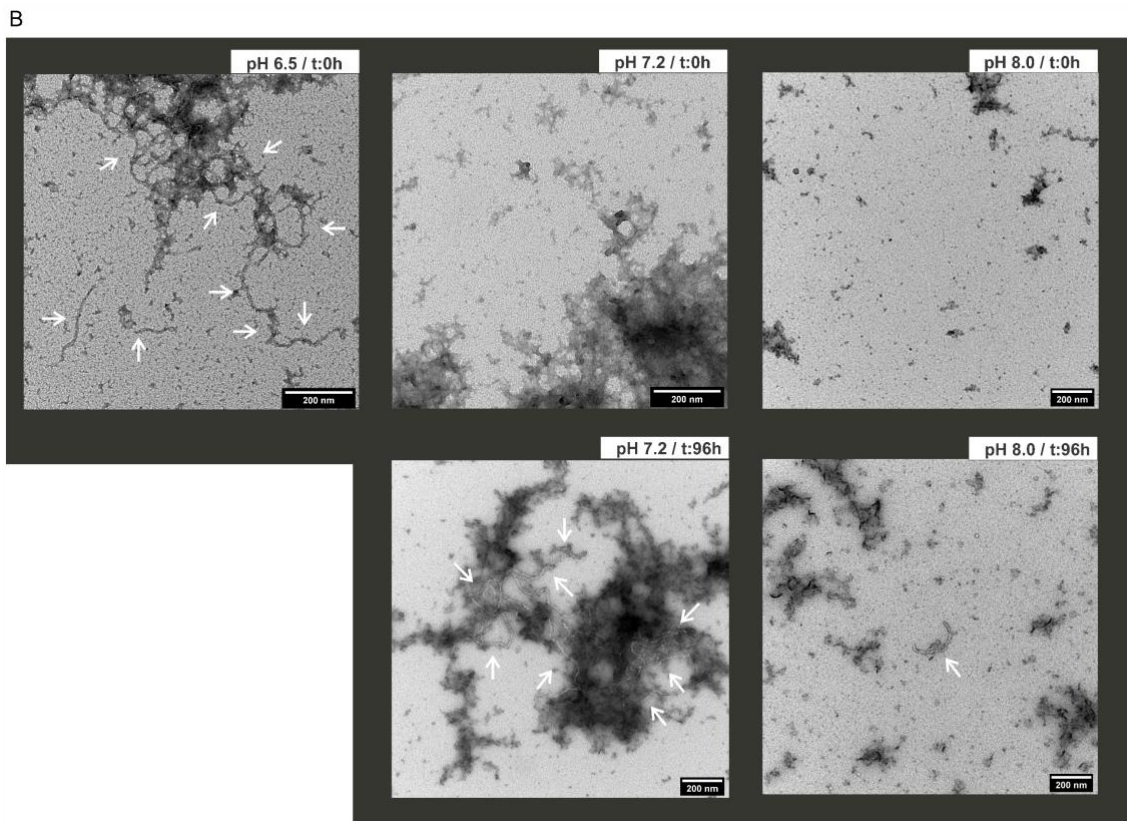
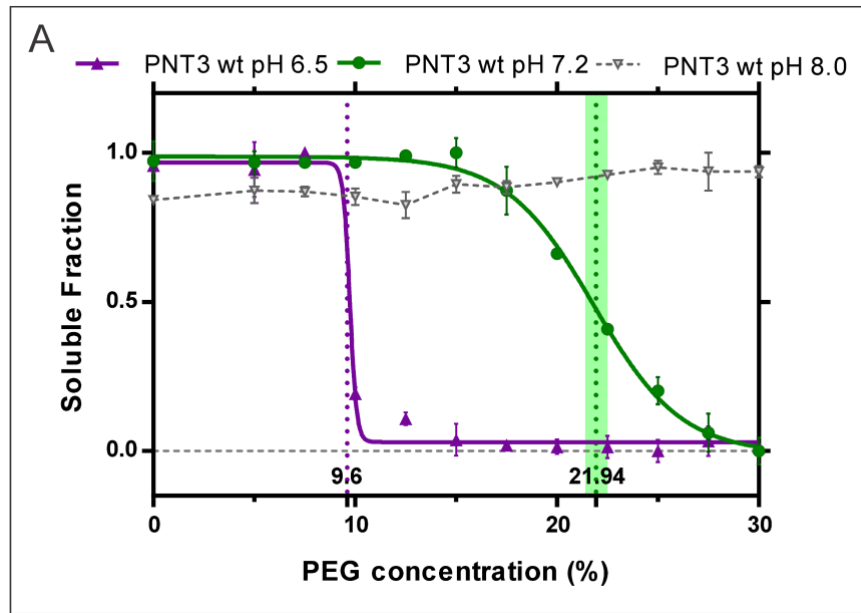
Within PNT3, a motif encompassing three contiguous tyrosines (EYYY) was predicted as an amyloidogenic region [25]. The ArchCandy predictor [27,28] also predicted a fibril architecture in which the three contiguous tyrosines of the motif are part of the first  $\beta$ -strand of a  $\beta$ -strand-loop- $\beta$ -strand motif (see Fig. 4C in [25]). ArchCandy identifies amyloidogenic regions based on their ability to form  $\beta$ -arcades. Indeed, the core structural element of a majority of naturally-occurring and disease-related amyloid fibrils is a  $\beta$ -arcade representing a parallel and in register stacks of  $\beta$ -strand-loop- $\beta$ -strand motifs called  $\beta$ -arches [27]. Substitution of the three contiguous tyrosine residues with three alanine residues yielded a variant (referred to as PNT3<sup>3A</sup>) that was shown to possess a dramatically reduced fibrillation ability, thus providing direct experimental evidence for the predicted involvement of the EYYY motif in building up the core of the fibrils [25].

Here, with the aim of achieving a better understanding of the molecular determinants of HeV PNT3 fibrillation, we designed and characterized a set of additional PNT3 variants that were conceived to either further investigate the EYYY amyloidogenic motif or probe the contribution of the C-terminal half of the protein to the fibrillation process. Results, as obtained by combining various biophysical and structural approaches, show that removal of one out of the three tyrosines of the motif, irrespective of position, is sufficient to lead to a significantly reduced fibrillation ability. In addition, our results revealed that the C-terminal half of PNT3 acts as a natural dampener of the fibrillation process.

## 2. Results and Discussion

### 2.1. Influence of pH on the formation of HeV PNT3 amyloid-like fibrils

We previously documented the ability of the PNT3 region of the HeV V protein to form amyloid-like structures [25]. A number of studies reported an impact of pH on both fibril structure and kinetics [29-31]. As a first step towards an in-depth characterization of PNT3 fibrils, and with the aim of selecting appropriate conditions to investigate the kinetics of fibril formation, we sought at assessing the possible impact of pH on the fibrillation process. To this end, we used a previously described method based on the titration of polyethylene glycol (PEG), a crowding agent, to quantitatively assess the relative solubility of proteins [32]. Hence, after optimization of this method (see Materials and Methods), we performed PEG precipitation assays to evaluate the relative solubility of HeV PNT3 at three different pH values, namely 6.5, 7.2 and 8.0 (**Fig. 2**). From this assay, the PEG<sub>1/2</sub> value, which corresponds to the PEG concentration at which 50% of the protein is still soluble, can be obtained and allows comparing protein aggregation propensities under different conditions [32]. Results display that HeV PNT3 at pH 6.5 shows less relative solubility compared to pH 7.2, indicating a higher aggregation propensity at the lower pH (**Fig. 2A**). This behavior might be at least partly rationalized based on the isoelectric point of the protein (pI=4.6), as proteins are well known to display minimal solubility at pH values close to their pI. The impact of pH on fibril formation, and the correlation with the PEG<sub>1/2</sub> value, was confirmed by ns-TEM (**Fig. 2B**). The obtained micrographs show that at pH 6.5 fibrillar aggregates are present even at time 0, while at pH 7.2 equivalent fibrillar aggregates are only observed after an incubation of 96 hours at 37 °C (*i.e.* no fibrillar aggregates can be detected at time 0, **Fig. 2B**). This trend is further confirmed at pH 8, a condition where PNT3 displays the lowest propensity to form fibrillar aggregates (**Fig. 2**). Notably, in addition to the presence of fibrils, the micrographs obtained at the three pH values also show the presence of amorphous aggregates. These results, beyond advocating for a role of electrostatics in the aggregation process, prompted us to define pH 7.2 as the standard pH value for further studies: the rationale for this choice was that we wanted to be able to monitor the appearance and growth of fibrils over time, while at pH 6.5 fibrils were detected as early as at time zero.



**Figure 2.** Fibrillation propensity of HeV PNT3 *wild-type* (*wt*) at different pHs. **A.** PEG assay and relative solubility of HeV PNT3 *wt* at pH 7.2 (green), 6.5 (violet) and 8.0 (light gray). The vertical lines correspond to PEG<sub>1/2</sub> values (with their 95 % confidence intervals) as obtained after a normalization and fitting step to a sigmoid function. Note that the curve obtained at pH 6.5 exhibits poor fitting thus preventing calculation of the confidence interval. Data points at pH 8.0 could obviously not be fitted. **B.** Ns-TEM of HeV PNT3 *wt* fibrils at three different pH values: 6.5 (at time 0, t:0h), 7.2 and 8.0 (at time 0 and after 96 hours of incubation at 37 °C, t:96h). White arrows indicate fibrils.



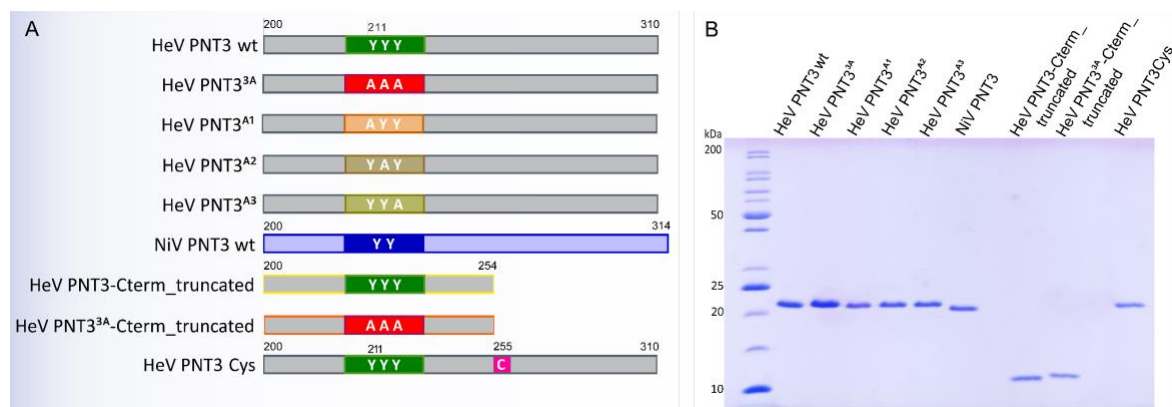
## 2.2 Rational design and generation of PNT3 variants

### 2.2.1. Design of PNT3 variants targeting the EYYY motif

Within the HeV PNT3 region, we previously identified an amyloidogenic motif encompassing three contiguous tyrosine residues (EYYY) [25]. The relevance of this motif for fibril formation was experimentally confirmed through the generation of the PNT3<sup>3A</sup> variant, in which the three tyrosine residues were replaced with three alanine residues, which showed a reduced ability to form amyloid-like fibrils [25]. With the goal of further investigating this amyloidogenic motif and of unveiling whether all the three tyrosine residues were critical for fibril formation or whether only a subset of them was so, we rationally designed three single-site PNT3 variants where each one of the three contiguous tyrosine residues was replaced with one alanine (PNT3<sup>A1</sup>, PNT3<sup>A2</sup> and PNT3<sup>A3</sup>, **Fig. 3A**). In this context, a construct encoding the corresponding PNT3 region from the NiV V protein (NiV PNT3) was also generated taking advantage of the fact that the NiV PNT3 motif encompasses only 2 contiguous tyrosine residues (EHYY) (**Fig. 3A** and **Supplementary Fig. S1**).

### 2.2.2. Design of HeV PNT3 truncated variants devoid of the C-terminal region

As mentioned above, the EYYY motif plays a significant role in fibril formation. However, the fact that the substitution of the triple tyrosine motif only reduces but does not fully abrogate the ability of PNT3 to form amyloid-like fibrils [25] indicates that other motifs and/or sequence attributes, that remained to be identified, contribute to the fibrillation process. Hence, to deepen the characterization and to assess the possible contribution of the PNT3 C-terminal region, we designed a C-terminally truncated HeV PNT3 variant (PNT3 C-term truncated) that lacks the second half of the protein (**Fig. 3A**). In addition, we also designed a C-terminally truncated HeV PNT3 variant where the three contiguous tyrosines of the EYYY motif were replaced with three alanines (PNT3<sup>3A</sup> C-term truncated) (**Fig. 3A**).



**Figure 3. A.** Schematic diagram of HeV PNT3 variants designed and constructed in this work. The EYYY motif of the HeV PNT3 is shown in green. **B.** SDS-PAGE analysis of purified proteins

### 2.2.3. Design of a HeV PNT3 variant bearing a unique cysteine

The N-terminal domain, shared by the HeV P, V, and W proteins, has 3 cysteines distributed along the sequence (**Fig. 1**). Recently, Pesce & Gondelaud et. al suggested that disulfide bridges could be involved in preventing aggregation of the W protein [21]. Thus, in this context, we reasoned that a HeV PNT3 variant bearing a cysteine residue could be useful to investigate the possible impact of disulfide bridge-mediated protein dimerization on the fibrillation abilities of PNT3. We targeted for cysteine substitution the unique alanine residue of PNT3 (Ala255) to yield PNT3 variant A255C, PNT3\_Cys (**Fig. 3A**). The rationale for choosing an alanine, rather than a serine residue which would have enabled a more isosteric substitution, was to introduce an as much conservative as possible substitution, while preserving the content in OH groups, which might play a role in the establishment of stabilizing inter-chain interactions in the core of the fibrils.

#### 2.2.4. Expression and purification of the PNT3 variants

All the proteins were expressed in *E. coli* as hexahistidine tagged forms with no solubility tag. The proteins were purified from the total fraction of the bacterial lysate under denaturing conditions. The proteins were purified by Immobilized Metal Affinity Chromatography (IMAC) and size exclusion chromatography (SEC). The purity of the final purified products was assessed by SDS-PAGE (Fig. 3B). The identity of all the variants was confirmed by mass spectrometry (MS) analysis of peptides resulting from the tryptic digestion of each protein (data not shown).

#### 2.2.5. Conformational characterization of the PNT3 variants

First, we evaluated the hydrodynamic properties of all the variants through analytical SEC. Table 1 shows the Stokes radius ( $R_s$ ) value inferred for each protein, along with the Compaction Index (CI) associated to each variant. By comparing the mean measured Stokes radius ( $R_s^{\text{OBS}}$ ) with the theoretical Stokes radii expected for various conformational states (*i.e.*,  $R_s^{\text{NF}}$ : natively folded protein;  $R_s^{\text{PMG}}$ : expected for a PMG;  $R_s^{\text{U}}$ : fully unfolded form;  $R_s^{\text{IDP}}$ : expected for an IDP), all the proteins were found to have  $R_s$  values consistent with a PMG state [33]. Their compaction indexes are relatively close to each other's, with the notable exception of the HeV PNT3\_C-term\_truncated variant that is much more compact. Strikingly, the introduction of the triple alanine motif in the context of the truncated variant leads to a more extended conformation, a phenomenon already observed, although with a borderline significance, in the context of the full-length PNT3 protein (*cf.* HeV PNT3 *wt* and HeV PNT3<sup>3A</sup> in Table 1). These results suggest that the second half of the protein is a determinant of chain expansion, with this effect being counteracted by the presence of the triple alanine motif.

**Table 1.** Stokes radii ( $R_s^{\text{obs}}$ , Å) of the PNT3 variants as inferred from the elution volume of the major SEC peak. Shown are also the expected values for the various conformational states, along with the ratios between the  $R_s^{\text{obs}}$  and each  $R_s$  state, and compaction index (CI) values.

Proteins	Mass	$R_s^{\text{obs}}$	$R_s^{\text{NF}}$	$R_s^{\text{PMG}}$	$R_s^{\text{U}}$	$R_s^{\text{IDP}}$	$\frac{R_s^{\text{obs}}}{R_s^{\text{NF}}}$	$\frac{R_s^{\text{obs}}}{R_s^{\text{PMG}}}$	$\frac{R_s^{\text{obs}}}{R_s^{\text{U}}}$	$\frac{R_s^{\text{obs}}}{R_s^{\text{IDP}}}$	CI
HeV PNT3 <i>wt</i>	15198	27.0 ± 0.3	19.4	27.9	33.9	30.0	1.39	0.97	0.80	0.90	0.47 ± 0.02
HeV PNT3 <sup>3A</sup>	14922	27.5 ± 0.4	19.3	27.7	33.5	30.0	1.42	0.99	0.82	0.92	0.42 ± 0.03
HeV PNT3 <sup>A1</sup>	15106	26.3 ± 0.8	19.4	27.9	33.8	30.0	1.36	0.94	0.78	0.88	0.52 ± 0.05
HeV PNT3 <sup>A2</sup>	15106	25.5 ± 0.3	19.4	27.9	33.8	30.0	1.31	0.91	0.76	0.85	0.58 ± 0.02
HeV PNT3 <sup>A3</sup>	15106	27.5 ± 0.2	19.4	27.9	33.8	30.0	1.42	0.99	0.81	0.92	0.44 ± 0.01
NiV PNT3 <i>wt</i>	14928	27.3 ± 0.7	19.3	27.7	33.5	30.5	1.41	0.99	0.82	0.90	0.44 ± 0.05
HeV PNT3_C-term_truncated	9020	19.0 ± 0.1	16.1	22.6	25.8	22.7	1.17	0.84	0.73	0.83	0.71 ± 0.01
HeV PNT3 <sup>3A</sup> _C-term_truncated	8743	20.1 ± 0.6	16.0	22.4	25.4	22.7	1.26	0.90	0.79	0.89	0.56 ± 0.06
HeV PNT3Cys	15230	28.0 ± 0.4	19.5	28.0	33.9	30.0	1.44	1.00	0.83	0.93	0.41 ± 0.03

$R_s^{\text{OBS}}$ : experimentally observed Stokes radius (mean value and s.d. from three independent experiments);  $R_s^{\text{NF}}$ :  $R_s$  expected for a natively folded (NF) form;  $R_s^{\text{PMG}}$ :  $R_s$  expected for a pre-molten globule (PMG);  $R_s^{\text{U}}$ :  $R_s$  expected for a fully unfolded form;  $R_s^{\text{IDP}}$ :  $R_s$  expected for an IDP based on the simple power law model; all radii are given in Å. Mass: molecular mass (Daltons) calculated from the amino acid sequence of the recombinant protein. For the  $R_s^{\text{obs}}$  and CI shown are the mean values and s.d., as obtained from three independent experiments.

In order to evaluate the secondary structure content of each variant we performed a Circular Dichroism (CD) analysis in the far ultraviolet (UV) region. All the variants present a spectrum typical of a disordered protein lacking any stable organized secondary structure, as

judged from the large negative peak centered at 200 nm, and from the low ellipticity in the 220–230 nm region and at 190 nm (see [34] and references therein cited) (**Supplementary Fig. S2**). These results indicate that the introduced substitutions and/or the truncation impact only marginally, if at all, the secondary structure content of the protein. They also indicate that NiV PNT3 has a secondary structure content very close to that of its HeV counterpart. The finding that the CD spectra of PNT3 variants bearing the triple alanine motif are virtually superimposable onto those of variants bearing either the naturally occurring triple tyrosine motif or just one Tyr to Ala substitution, rules out the possibility that the expansion effect driven by the triple alanine motif, as observed in SEC studies, may arise from the presence of a transiently populated  $\alpha$ -helix encompassing the motif. Therefore, the mechanism underlying the counteracting effect exerted by the triple alanine motif on chain compaction remains to be elucidated.

To achieve a more quantitative description of the conformational properties of the variants, we carried out Small-Angle X-ray Scattering (SAXS) studies coupled to SEC (SEC-SAXS). We selected a set of representative variants (*i.e.*, HeV PNT3<sup>3A</sup>, NiV PNT3, HeV PNT3\_C-term\_truncated, and HeV PNT3<sup>3A</sup>\_C-term\_truncated) along with HeV PNT3 *wt*. Although we already previously reported SEC-SAXS studies of HeV PNT3 *wt* [25], this sample was herein again investigated under exactly the same conditions used for the other variants, so as to enable meaningful comparisons. For all the five PNT3 proteins, linearity of the Guinier region in the resulting scattering curves (**Supplementary Fig. S3A**) allowed meaningful estimations of the  $R_g$  (**Table 2**). The  $R_g$  values obtained for HeV PNT3<sup>3A</sup>, NiV PNT3, and HeV PNT3 *wt* are very close to each other's clustering in a group with  $R_g$  around 37–39 Å. As expected, the two truncated variants have smaller, and close to each other's,  $R_g$  values (**Table 2**). Notably, all experimental  $R_g$  obtained are close to the theoretical  $R_g^U$ , corresponding to chemically denatured (U) proteins, reflecting a highly extended conformation. Because of this, the  $R_g$ -based CI could not be computed, the numerator ( $R_g^U - R_g^{OBS}$ ) in Eq. 9 being  $\leq 0$  (see Materials & Methods). These results are in contrast with the previous SEC results, where the variants were found to adopt a PMG conformation. A possible explanation for this might be related to the differences in the buffer used in the two techniques. As expected, the five variants were all found to be disordered as judged from the presence of a plateau in the normalized Kratky (**Supplementary Fig. S3B**) and the Kratky-Debye plots (**Supplementary Fig. S3C**). However, the two truncated variants, and particularly the HeV PNT3 C-term truncated one, showed a slight deviation from the pure random-coil regime as observed in the normalized Kratky plot. This deviation is consistent with a slightly more compact conformation, in line with the  $R_s$ -based CI values discussed above.

**Table 2.**  $R_g$  and  $D_{max}$  as obtained from SEC-SAXS studies and expected values for the various conformational states.

Proteins	$I(0) \text{ cm}^{-1}$	$R_g$ (Å) (Guinier)	$D_{max}$ (Å)	$R_g^{IDP}$ (Å)	$R_g^U$ (Å)
HeV PNT3 <i>wt</i>	$0.03 \pm 1.8 \times 10^{-4}$	$36.67 \pm 0.41$	140	32.6	35.9
HeV PNT3 <sup>3A</sup>	$0.022 \pm 8.1 \times 10^{-5}$	$39.48 \pm 0.33$	144	32.6	35.9
NiV PNT3	$0.03 \pm 8.2 \times 10^{-5}$	$37.37 \pm 0.21$	147	33.1	36.5
HeV PNT3 C-term_truncated	$0.018 \pm 4.1 \times 10^{-5}$	$27.53 \pm 0.14$	115	24.5	25.9
HeV PNT3 <sup>3A</sup> C-term_truncated	$0.015 \pm 5.6 \times 10^{-5}$	$27.34 \pm 0.21$	119	24.5	25.9

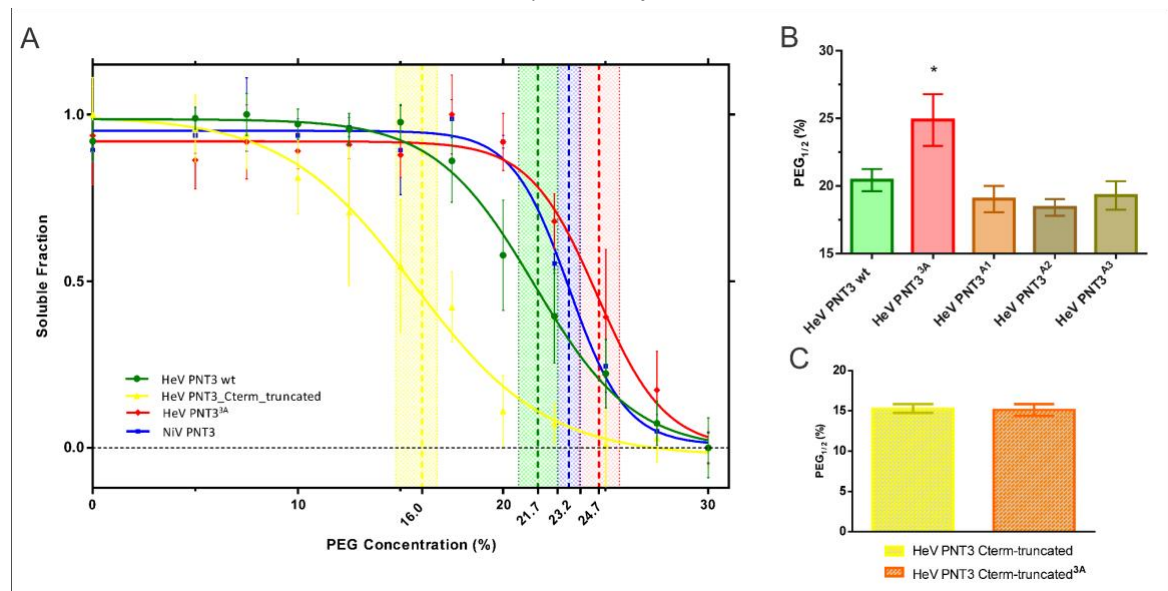
$I(0)$ : Intensity at zero angle as determined from Guinier approximation;  $R_g$  Guinier:  $R_g$  values as obtained from Guinier approximation;  $D_{max}$ : maximal intramolecular distance from  $P(r)$ .  $R_g^{IDP}$ :  $R_g$  expected for an IDP based on the simple power-law model.  $R_g^U$ : theoretical  $R_g$  value expected for chemically denatured (U) proteins.

### 2.3. Relevance of the PNT3 EYYY motif in fibrillation abilities

#### 2.3.1. Aggregation propensity of the EYYY motif PNT3 variants

With the aim of elucidating the contribution of each tyrosine in the EYYY motif to the fibrillation process, we first assessed the aggregation propensity in the set of variants bearing alanine substitutions within the EYYY amyloidogenic motif. To this end, we took advantage of the same PEG solubility assay described above [32]. For each of the five EYYY variants (PNT3<sup>3A</sup>, PNT3<sup>A1, A2, A3</sup> and NiV PNT3), we therefore carried out PEG solubility assays which enabled ranking them based on their estimated PEG<sub>1/2</sub> value.

The PNT3<sup>3A</sup> variant shows a significant increase in the relative solubility compared to PNT3 HeV *wt* (Fig. 4A and 4B), a result in agreement with previous findings that pointed out a much lower propensity to form amyloid-like fibrils for PNT3<sup>3A</sup> [25]. The three variants where only one Tyr was replaced, however, show no significant difference in the PEG<sub>1/2</sub> values compared to HeV PNT3 *wt* (Fig. 4B), suggesting that the removal of one tyrosine does not affect the aggregation propensity of PNT3. By contrast, and interestingly, NiV PNT3 (EHYY) displays an intermediate relative solubility between HeV PNT3 *wt* and HeV PNT3<sup>3A</sup> (Fig. 4A). In light of the results obtained with the HeV PNT3 variants bearing two tyrosines (*i.e.*, PNT3<sup>A1</sup>, PNT3<sup>A2</sup> and PNT3<sup>A3</sup>), the intermediate aggregation propensity of NiV PNT3 more likely arise from differences in the amino acid context rather than from the absence of just one tyrosine.



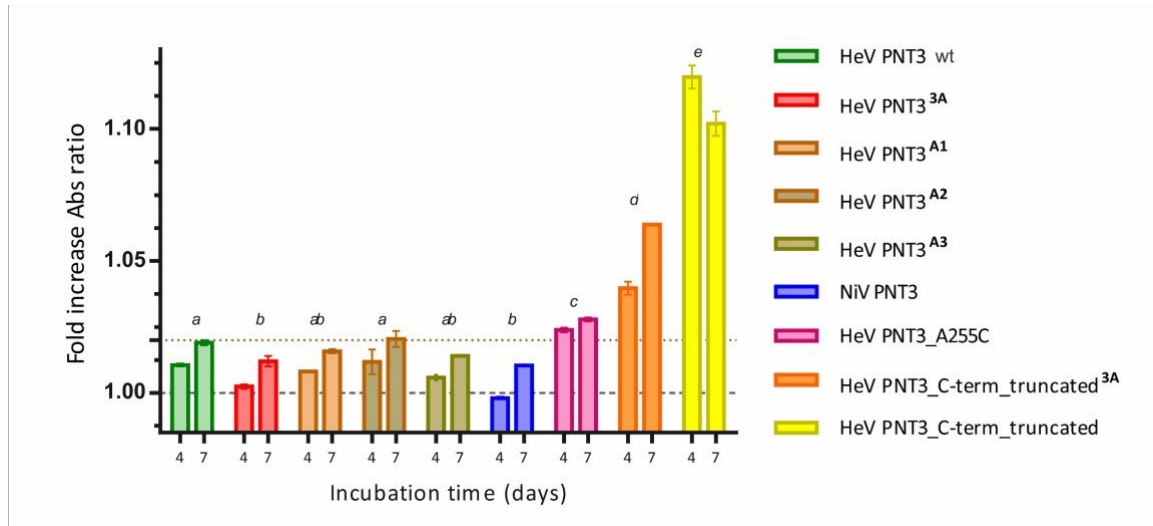
**Figure 4.** PEG solubility assay for the set of PNT3 variants. **A.** Soluble fraction of each variant at different PEG concentrations. HeV PNT3 *wt* (green), HeV PNT3<sup>3A</sup> (red), HeV PNT3<sub>C-term\_truncated</sub> (yellow), NiV PNT3 (blue). Vertical lines represent PEG<sub>1/2</sub> values including their 95 % confidence intervals obtained after a normalization and fitting step to a sigmoid function. **B.** PEG<sub>1/2</sub> values obtained for the *wt*, the triple alanine variant and the single alanine HeV PNT3 variants. **C.** PEG<sub>1/2</sub> values obtained for the C-terminally truncated variants.

#### 2.3.2. Congo Red binding abilities of PNT3 EYYY motif variants

Congo Red (CR) is a widely used dye to document the presence of amyloids: binding of this dye to cross  $\beta$ -sheet structures in fact leads to hyperchromicity and a red shift in the absorbance maximum of the CR spectrum. Hence, to further characterize PNT3 EYYY motif variants, we took advantage of CR binding assays. We compared the binding abilities of the PNT3<sup>A1</sup>, PNT3<sup>A2</sup>, PNT3<sup>A3</sup> and NiV variants to those of both PNT3 *wt* and PNT3<sup>3A</sup>. We spectrophotometrically measured the red shift (from 497 nm to 515 nm) in the absorbance maximum of the CR spectrum of each sample following a four- or seven-days incubation at 37 °C. Results shown in Figure 5 indicate that all the variants promote a shift in the CR spectrum whose amplitude increases with incubation time, suggesting that all the variants are able to progressively form amyloid-like fibrils or at least structures able to bind CR. The previous findings that pointed out that the HeV PNT3<sup>3A</sup> variant has



a reduced ability to bind CR compared to HeV PNT3 *wt* [25] were confirmed here (**Fig. 5**). Two of the three single alanine variants (PNT3<sup>A1</sup>, PNT3<sup>A3</sup>) show an intermediate behavior between HeV PNT3 *wt* and HeV PNT3<sup>3A</sup>, but without significant differences with either the *wt* or the PNT3<sup>3A</sup> variant, while HeV PNT3<sup>A2</sup> displays an ability to bind CR significantly higher than that of HeV PNT3<sup>3A</sup> and similar to that of HeV PNT3 *wt* (**Fig 5**). These results therefore suggest that the central tyrosine in the motif could be less relevant to the fibrillation process. Notably, the NiV PNT3 variant shows a significantly decreased ability to bind CR compared to PNT3 *wt*, similar to that of the triple alanine variant. Thus, as already observed for the aggregation propensity, the reduced CR binding ability of NiV PNT3 likely results from its amino acid context rather than from the fact that it lacks a tyrosine in the motif.



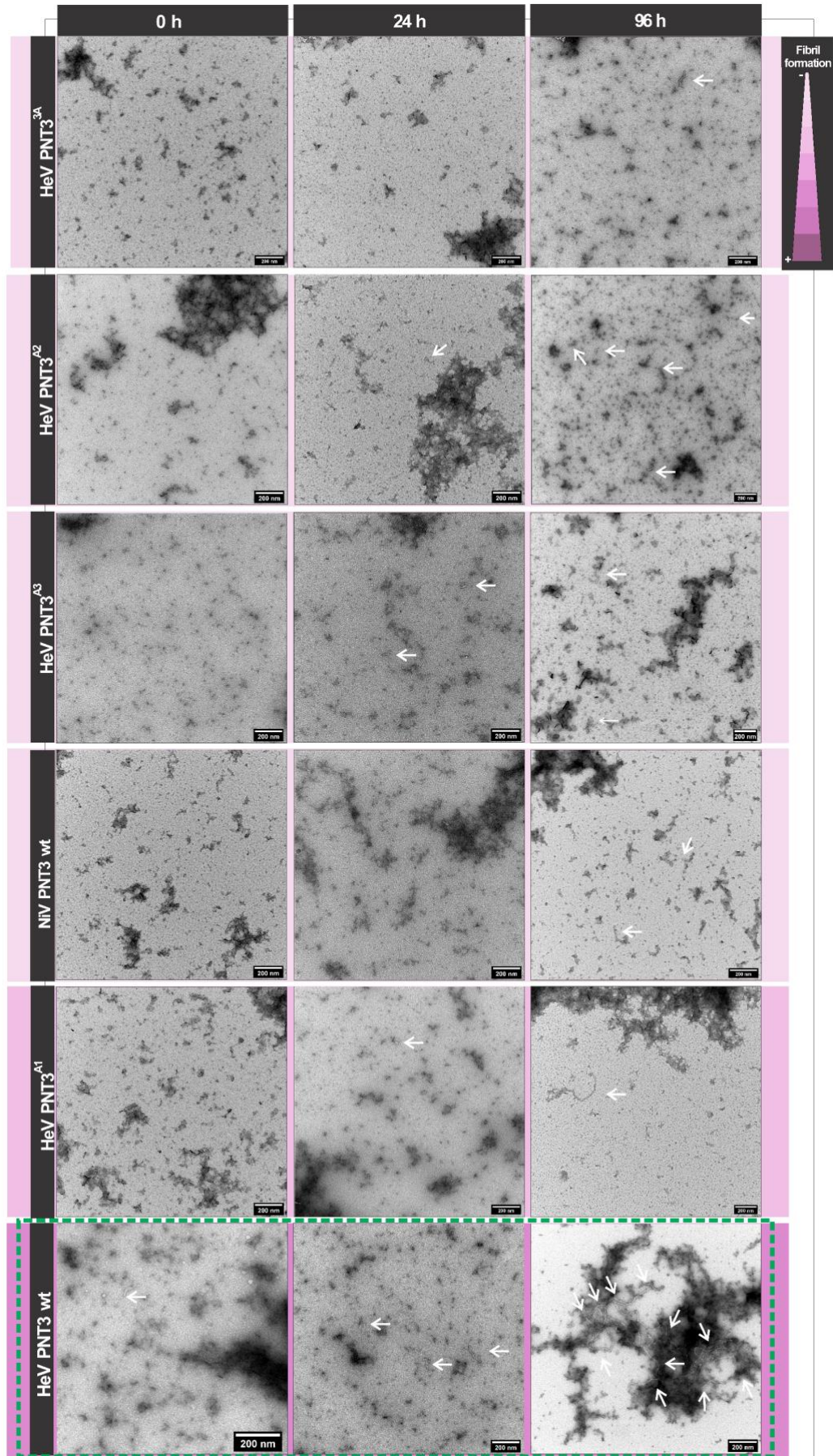
**Figure 5.** Congo Red binding assay of the set of PNT3 variants. The ability to bind CR is represented by the fold increase in the ratio between the absorbance at 515 and at 497 nm, with respect to a sample containing CR alone, of PNT3 samples at 20  $\mu$ M after 4 and 7 days of incubation at 37  $^{\circ}$ C. The error bar corresponds to the standard deviation, with  $n = 3$ . Different letters indicate statistically significant differences ( $p < 0.05$ ) (Two-way ANOVA test; *ab* means lack of statistically significant differences with respect to *a* or *b*).

### 2.3.3. Propensity and time-dependence of fibrillation of the PNT3 EYYY motif variants using negative-staining transmission electron microscopy (ns-TEM)

To directly document fibril formation by the PNT3 EYYY variants as a function of time and to obtain orthogonal experimental evidence corroborating the CR binding assay results, we next carried out ns-TEM studies. These analyses were performed for each of the variants at 0, 24 and 96 h of incubation at 37  $^{\circ}$ C (**Fig. 6**). As shown the **Figure 6**, although HeV PNT3 *wt* in the selected conditions does not form fibrils at time 0, after 24 h of incubation it forms short fibrils, evolving to long fibrils after 96 h. In line with the CR binding results, ns-TEM studies confirmed that all PNT3 EYYY variants, including NiV PNT3, are able to form amyloid-like fibrils but with a significantly decreased ability compared to the *wt*. Specifically, after 24 h short fibrils can be observed in most variants, except for HeV PNT3<sup>3A</sup>. In an attempt at identifying possible significant differences among the EYYY variants in spite of their overall similar behavior, we performed a comprehensive analysis of the number and length of fibrils detected for each of them (**Fig. 7**). This analysis revealed no significant differences in the length of the fibrils obtained after 96 hours of incubation among the 3 variants with a single alanine (PNT3<sup>A1</sup>, PNT3<sup>A2</sup> and PNT3<sup>A3</sup>). However, a slight, though significant, increase is observed in the PNT3<sup>A3</sup> variant in the number of fibrils found *per* picture compared to the other two single alanine variants (**Fig. 7**). This slightly higher number of short fibrils observed for the PNT3<sup>A3</sup> variant may indicate that the last tyrosine in the motif contributes less to the nucleation process compared to the two other tyrosines. Remarkably, the same analysis showed that the NiV PNT3 variant forms significantly longer fibrils compared to the PNT3<sup>A1</sup>, PNT3<sup>A2</sup> and PNT3<sup>A3</sup> variants, therefore confirming that other sequence attributes, beyond the amyloidogenic motif, contribute to the fibrillation process. In other words, fibril formation would

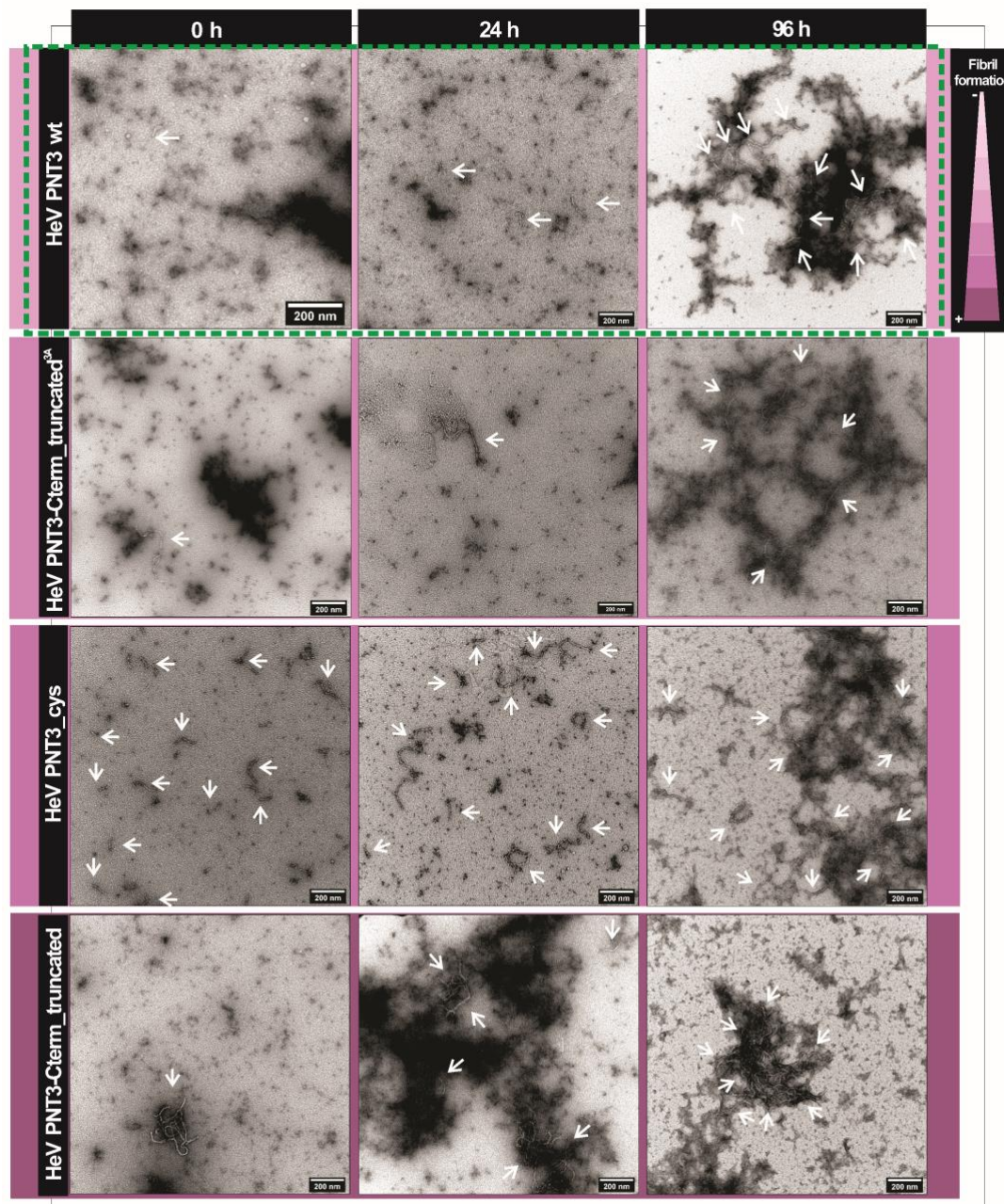
rely not only on stabilizing contacts mediated by the E(H/Y)YY motif but also on additional stabilizing interactions established by other protein regions.

Altogether, CR binding assays and ns-TEM enabled documenting slight differences among the PNT3<sup>A1</sup>, PNT3<sup>A2</sup> and PNT3<sup>A3</sup> variants. Although other regions beyond the amyloidogenic motif were found to contribute to the fibrillation process, results strongly suggest that the absence of a single tyrosine in the EYYY motif leads to a significant decrease in the ability to form fibrils irrespective of the position. In light of the finding that the PNT3<sup>A1</sup>, PNT3<sup>A2</sup> and PNT3<sup>A3</sup> variants are able to form short fibrils, whose length does not increase under the experimental conditions herein used, we can assume that the three tyrosines mainly play a role in the elongation phase. The present results support the involvement of tyrosines in the formation of amyloid-like fibrils, with  $\pi$ - $\pi$  stacking and H-bonding interactions between tyrosines likely allowing to form cross- $\beta$  like architectures, as previously suggested [35].



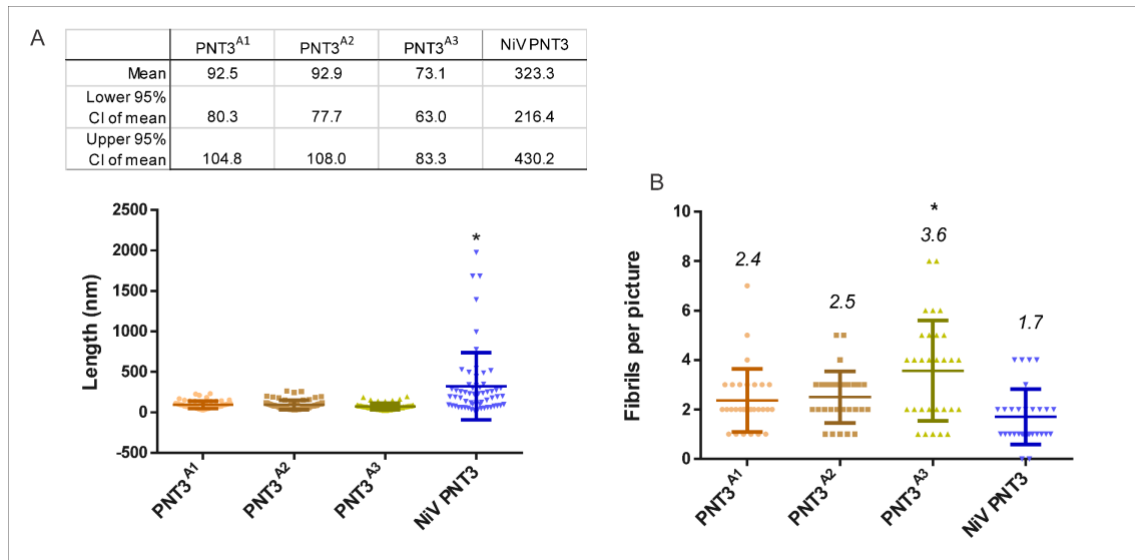
Continue next page





**Figure 6.** Fibril formation as a function of time by ns-TEM. Ns-TEM analysis of PNT3 variants (200  $\mu\text{M}$ ) at time zero and after 24 h or 96 h of incubation at 37  $^{\circ}\text{C}$ . Note that in all cases, samples were diluted to 40  $\mu\text{M}$  prior to being deposited on the grid. White arrows indicate fibrils. The purple gradient represents the propensity to form fibrils.





**Figure 7.** Fibril length of EYYY variants. **A.** For each variant, statistics on fibril length was obtained from analysis of the contour length of 60 fibrils after 96 h of incubation at 37 °C. The upper table shows the main statistic results. **B.** Number of fibrils detected *per* picture for each variant. Upper number indicates the mean value of each variant. The analysis was done using the ImageJ software. (\*) indicates statistically significant differences (One-way Anova test, p-value < .05)

#### 2.4. Impact of the HeV PNT3 C-terminal region in fibrillation abilities

##### 2.4.1. Aggregation propensity of C-terminally truncated PNT3 variants

Taking into account the results presented above that lend support to a scenario where other motifs and/or sequence attributes beyond the amyloidogenic motif could be involved in the fibrillation process, we decided to investigate the impact of the PNT3 C-terminal region. We first studied the aggregation propensity of both PNT3 C-terminally truncated variants. **Figure 4A** shows that the PNT3 C-term truncated variant has the lowest PEG<sub>1/2</sub> value, indicating a strikingly decreased relative solubility compared to all the full-length PNT3 variants. Notably, there are no significant differences in the PEG<sub>1/2</sub> values between the PNT3 C-term truncated and its triple alanine mutant (PNT3<sup>3A</sup> C-term truncated) (**Fig. 4C**). These results indicate that removal of the C-terminal region has a strong impact on the aggregation propensity, with this effect being insensitive to the sequence context of the EYYY motif.

##### 2.4.2. CR binding ability of C-terminally truncated PNT3 variants

Motivated by the results obtained by the PEG solubility assays pointing to a much higher aggregation propensity of both truncated variants, we next carried out CR binding assays. As shown in **Figure 5**, the HeV PNT3 C-terminal truncated variant displays a significantly increased ability to bind CR compared to full-length HeV PNT3 *wt*. In striking contrast with PEG solubility assays that detected no differences in terms of aggregation propensities between the two truncated variants, CR binding assays revealed significant differences between the two variants. In particular, the PNT3<sup>3A</sup> truncated variant has a much-decreased ability to bind CR with respect to the truncated variant bearing a native EYYY motif, hence displaying an intermediate behavior between the full-length and truncated HeV PNT3 *wt* (**Fig. 5**). Therefore, it can be concluded that the C-terminal region, far from being inert, negatively affects the ability of the protein to form CR-binding structures. Albeit the PNT3<sup>3A</sup> truncated variant binds more CR than the *wt* variant, the contribution of the EYYY motif is evident when the two truncated variants are compared. These results, however suggest that the EYYY motif would have only a marginal role in driving the formation of CR-binding structures in the context of C-terminally truncated form.

#### 2.4.3. Propensity and kinetics of fibrillation of C-terminally truncated PNT3 variants using ns-TEM studies

In order to ascertain whether the increased binding to CR and decreased solubility of the truncated variants is actually reflected in a higher fibrillation ability, we analyzed them by NS-TEM studies. In line with expectations, **Figure 6** clearly shows a significantly higher abundance of fibrils, as well as an increased fibril length, in the PNT3 C-terminal truncated variant at short incubation times, thus confirming its increased fibrillation potential. Notably, the PNT3<sup>3A</sup> truncated variant displays a decreased fibrillation ability compared to its *wt* counterpart, similar to the full-length HeV PNT3 *wt* (**Fig. 6**). These findings suggest that the removal of the C-terminal region results in an acceleration of the fibrillation rate, reflecting an interaction between this region and the rest of the sequence that negatively affects the kinetics of fibril formation. Remarkably, similar results were previously documented in the case of the aggregation of  $\alpha$ -synuclein ( $\alpha$ -syn), an extensively characterized protein associated with neurodegeneration and whose transition from a soluble to a fibrillar form is thought to contribute to pathogenesis [36]. Compelling experimental indicates that C-terminal truncation of  $\alpha$ -syn promotes *in vitro* oligomer and fibril formation (see [37] and references therein cited). The middle region of  $\alpha$ -syn, referred to as “non-amyloid component” (NAC) domain, forms the core of  $\alpha$ -syn filaments. The C-terminal region of  $\alpha$ -syn can adopt conformations in which the C-terminus contacts the hydrophobic NAC domain thus shielding it from pathological templating interactions [37]. The negative charge of the C-terminal region has been proposed to contribute to this self-chaperoning activity *via* the establishment of electrostatic interactions [37].

In an attempt at rationalizing the observed self-inhibitory effect of the C-terminal region of PNT3 on fibril formation, we analyzed the charge distribution within the PNT3 sequence (**Supplementary Fig. S4** and **Table S2**). Although full-length PNT3, and its constituent N-terminal and C-terminal regions fall in very close positions in the phase diagram plot, the C-terminal region has a higher fraction of negatively charged residues compared to the N-terminal region (**Supplementary Fig. S4** and **Table S2**). In addition, the full-length form of PNT3 and the truncated variant strongly differ in their net charge at pH 7.0. Taking into account the strong impact of pH on PNT3 fibrillation, where a decrease to pH 6.5 strongly promotes fibrillation and leads to a behavior similar to that of the truncated PNT3 variant at pH 7.2, it is conceivable that the results obtained with the truncated variant could be, at least partly, accounted for by electrostatics, as in the case of  $\alpha$ -syn. A plausible alternative scenario for the self-inhibitory effect of the C-terminal region of PNT3 on fibril formation could be the following: the disordered region downstream the fibril core may hamper fibril formation by slowing the disorder-to-order transition expected to take place in the core of the fibril, through either a purely entropic effect or through a combination of enthalpy and entropy as already documented in the case of fuzzy appendages adjacent to molecular recognition elements (for examples see [38,39]).

Altogether, these findings advocate for a key role of the C-terminal region in regulating the fibrillation properties of PNT3. In particular, the C-terminal region may act either as a molecular shield, as in the case of  $\alpha$ -syn [37], or by slowing down the rate of folding of the core of the fibrils, with this property, irrespective of the underlying mechanisms, being also possibly relevant to biological function *in vivo*. Definite answers on the precise molecular mechanisms and on the possible biological relevance await future studies.

#### 2.5. Impact of a cysteine in the HeV PNT3 sequence on fibrillation abilities

With the aim of elucidating the possible impact of a disulfide bridge-mediated PNT3 dimerization on its fibrillation abilities, we first studied the ability of the HeV PNT3 variant bearing a cysteine residue (PNT3\_Cys) to bind CR. As shown in **Figure 5** this variant shows a significant increased ability to bind CR compared to HeV PNT3 *wt*. Subsequently, the ability of this variant to form fibrils was assessed by ns-TEM in the same conditions used for PNT3 *wt*. **Figure 6** shows that the PNT3\_Cys variant is able to form fibrils even at time 0, indicating a higher fibrillation propensity compared to HeV PNT3 *wt*. Notably, the PNT3\_Cys was the unique variant displaying an enrichment in shortened fibrils (**Fig. 6**). These findings suggest that the presence of one cysteine

in the sequence mainly impacts the nucleation phase. We reasoned that the peculiar fibrillation behavior of this variant could result from disulfide bridge-mediated protein dimerization. However, the addition of DTT was found to exert a negligible impact, as judged from the presence at time 0 of very short fibrils and of long fibrils at 96 hours (**Supplementary Fig. S5**), resulting in an intermediate behavior between PNT3 *wt* and PNT3\_*cys* under non-reducing conditions. Thus, the peculiar behavior of this variant cannot be ascribed to protein dimerization and rather stems from other intrinsic properties that remain to be elucidated.

## 2.6. Characterization of the aggregation process by Taylor Dispersion Analysis

We next sought at shedding light onto the fibrillation kinetics using Taylor Dispersion Analysis (TDA) [40]. TDA is a new technique in the field of protein aggregation that has the notable advantages of being able to (i) capture intermediate species, (ii) quantify early- and late-stage aggregates, and (iii) provide both kinetic and equilibrium constants. Besides, TDA is not dominated by aggregates (as opposite to scattering techniques), and is thus ideally suited to study the molecular mechanisms of protein fibrillation. Recently, this technique successfully allowed obtaining a complete quantitative picture of the aggregation process of both A $\beta$ (1-40) and A $\beta$ (1-42), including the size of the oligomers and protofibrils, the kinetics of monomer consumption, and the quantification of different early- and late-formed aggregated species [40-42].

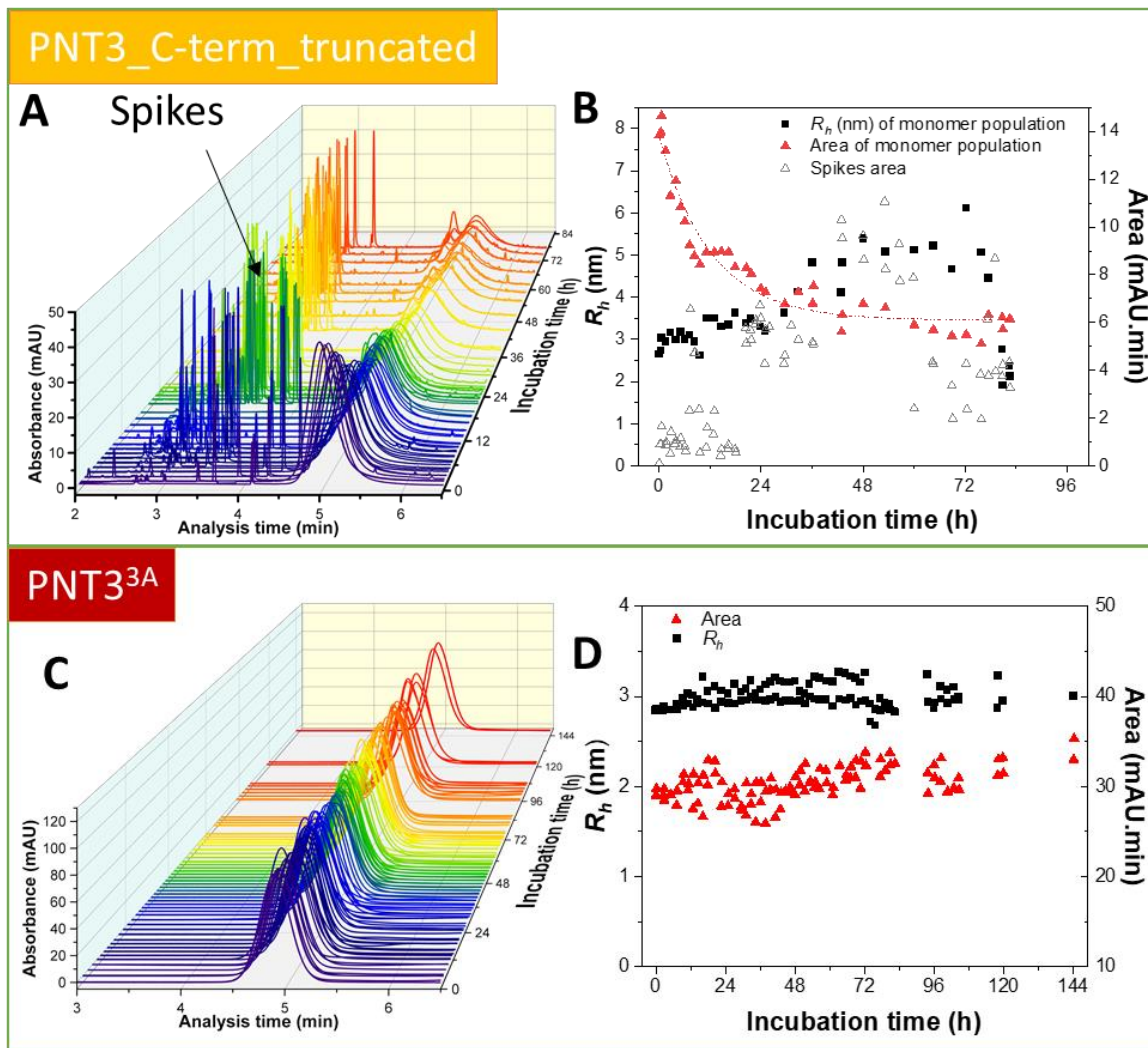
In light of the fibrillation properties of all the PNT3 variants herein investigated, as revealed by the ensemble of studies described above, we decided to focus on the two variants with the most extreme phenotype, namely the variant with the highest fibrillation propensity, *i.e.*, HeV PNT3\_C-term\_truncated, and the least fibrillogenic variant, *i.e.*, PNT3<sup>3A</sup>. **Figure 8A** and **C** show a three-dimensional overview of the obtained taylorgrams during the aggregation process of the two selected variants. In the case of the truncated variant, two major species were detected over time: the monomer (hydrodynamic radius,  $R_h$ , of  $\sim 3$  nm) (**Fig. 8B**) and large aggregates with an  $R_h \geq 400$  nm (see spikes at the beginning of the run in **Fig. 8A**). **Figure 8B** shows that the  $R_h$  value of the soluble fraction increases with time, reaching about the double of its initial value, when this species is present in low proportions. **Figure 8B** also shows that the monomeric population slowly decreases with time. The decrease of the peak area ( $Y$ ) of the monomeric population could be fitted using a 1<sup>st</sup> order exponential decay:

$$Y = A_1 \times \exp\left(-t/t_1\right) + Y_0 \quad (\text{Eq. 1})$$

where  $Y_0 = 6.13 \pm 0.20$ ;  $A_1 = 7.63 \pm 0.34$  and  $t_1 = 12.06 \pm 1.32$  h. This resulted in a good quality fit as judged from the  $R^2$  value of 0.9394. From the fit, the kinetics of aggregation could be deduced, with a characteristic aggregation time of about  $t_1 = 12$  h. From **Figure 8B**, the evolution in the spikes area gives an estimation of the quantity of large aggregated species entering the capillary. The proportion of these aggregated species increases with time until about 48 h, before decreasing because their size becomes too large to enter the capillary or because of precipitation in the sample vial. The presence of spikes at very short times of incubation suggests a fast aggregation process of the truncated form, while the absence of significant intermediate species suggests that the monomers add to the already present aggregates and elongate them.

By contrast, and in agreement with its dramatically reduced fibrillogenic abilities as unveiled by the other approaches herein used, the PNT3<sup>3A</sup> variant does not show any changes during the incubation time, with both  $R_h$  and peak area remaining constant in this period (**Fig. 8C,D**).

In conclusion, the data support a much higher fibrillation ability for the truncated variant as corroborated by the presence of spikes in the elution profiles and by the fast consumption of the monomeric species as compared to PNT3<sup>3A</sup> where no evolution in size nor in area was observed. The aggregation process of the truncated variant follows a first order kinetics, without significant formation of intermediate species between the monomeric and the fibrillar species. In PNT3<sup>3A</sup> the fibrillation process could not be detected, indicating that the fibrils observed by ns-TEM represent a very poorly populated species within the system.



**Figure 8.** Kinetics of fibril formation by Taylor Dispersion Analysis (TDA). Three-dimensional overview of the obtained Taylorgrams during the aggregation process of HeV PNT3-Cterm\_truncated (A) and HeV PNT3<sup>3A</sup> (C) at different incubation times. Analyses were performed using a protein concentration of 200  $\mu$ M in 50 mM phosphate buffer, pH 7.2 at 37  $^{\circ}$ C. Peak area and hydrodynamic radius ( $R_h$ ) evolution of the monomeric species and of the spikes area during the aggregation process of HeV PNT3-Cterm\_truncated (B) and HeV PNT3<sup>3A</sup> (D).

### 3. Conclusions

This study constitutes a comprehensive analysis of the molecular basis of the fibrillation process of a small region (PNT3) within the N-terminal intrinsically disordered domain shared by the HeV P/V/W proteins. Biochemical and biophysical characterization of a set of HeV PNT3 variants bearing alanine substitutions in the amyloidogenic EYYY motif, along with the characterization of the corresponding PNT3 region from the cognate NiV, revealed that each of the three tyrosines in the motif are required for the elongation step of the fibrillation process. Remarkably, the present study also unveiled a role for the C-terminal domain of PNT3 in self-inhibition of fibrillation, possibly reminiscent of the  $\alpha$ -synuclein fibrillation model, and of potential biological significance. Noteworthy, in light of the observation that amyloid-like fibrils form not only *in vitro* but also in the cellular context, it is tempting to hypothesize that the amyloidogenicity of V/W proteins, which both encompass the PNT3 region, could be correlated with the pathogenic (and even encephalitogenic) properties of Henipaviruses. Therefore, the PNT3 variants that we have herein generated constitute valuable tools to further explore the functional impact of V/W fibrillation in transfected and infected cells. The present results therefore set the stage for further



investigations aimed at illuminating the mechanisms underlying the disease as a preliminary step towards the rational design of antivirals.

## 4. Materials and Methods

### 4.1. Generation of the constructs

The pDEST17OI/PNT3 and pDEST17OI/PNT3<sup>3A</sup> expression plasmids, driving the expression of a hexahistidine tagged form of the protein of interest, have already been described [25]. For the construction of expression plasmids encoding the PNT3 variants bearing single alanine substitutions (PNT3<sup>A1</sup>, PNT3<sup>A2</sup>, PNT3<sup>A3</sup>), the pDEST17OI/PNT3 construct was used as template in two separate PCR amplifications using either primers attB1 and specific R\_alaN-PNT3 (PCR1), or primers F\_alaN-PNT3 and attB2 (PCR2), where *N* varies from 1 to 3 (see **Supplementary Table 1**). After DpnI treatment, 1 µl of PCR1 and 1 µl of PCR2 were used as overlapping megaprimers along with primers attB1 and attB2 in a third PCR. After purification, the third PCR product was inserted into the pDEST17OI bacterial expression vector using the Gateway® technology (Invitrogen). This vector allows expression of the recombinant protein under the control of the T7 promoter. The resulting protein is preceded by a stretch of 22 vector-encoded residues (MSYYHHHHHLESTSLYKKAGF) encompassing a hexahistidine tag. The DNA fragment encoding NiV PNT3 (*i.e.*, residues 200-314 of the NiV P/V/W protein) was PCR-amplified using the pDEST17OI/NiV W construct as template [21] and primers NiV PNT3-AttB1 and NiV PNT3-AttB2. After DpnI treatment, the resulting amplicon was cloned into pDEST17OI as described above.

The expression construct encoding the truncated HeV PNT3 variant (PNT3\_C-term\_truncated) was generated by PCR using pDEST17OI/PNT3 [25] as template and attB1 and Trunc\_PNT3\_B2 as primers. After DpnI treatment, the resulting amplicon was cloned in pDEST17OI. The same procedure was used to obtain the HeV PNT3 truncated variant bearing the triple alanine substitution (PNT3<sup>3A</sup>\_C-term\_truncated) except that the pDEST17OI/PNT3<sup>3A</sup> construct [25] was used as template.

The construct encoding the HeV PNT3 variant bearing a cysteine (PNT3\_Cys) was obtained using the pDEST17OI/PNT3 construct [25] as template in two separate PCR amplifications using either primers attB1 and specific PNT3\_C255\_R (PCR1), or primers PNT3\_C255\_F and attB2 (PCR2). After DpnI treatment, 1 µl of PCR1 and 1 µl of PCR2 were used as overlapping megaprimers along with primers attB1 and attB2 in a third PCR. After purification, the third PCR product was inserted into pDEST17. The list and sequence of primers used to generate the above-described constructs is provided in **Supplementary Table S1**. Primers were purchased from Eurofins Genomics. All the constructs were verified by DNA sequencing (Eurofins Genomics) and found to conform to expectations.

### 4.2. Proteins expression and purification

The *E. coli* strain T7pRos was used for the expression of all the recombinant proteins upon transformation of bacterial cells with each of the bacterial expression plasmids described above. Cultures were grown over-night to saturation in LB medium containing 100 µg mL<sup>-1</sup> ampicillin and 34 µg mL<sup>-1</sup> chloramphenicol. An aliquot of the overnight culture was diluted 1/20 into 1 L of TB medium and grown at 37 °C with shaking at 200 rpm. When the optical density at 600 nm (OD<sub>600</sub>) reached 0.5-0.8, isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM, and the cells were grown at 35 °C overnight. The induced cells were harvested, washed and collected by centrifugation (5,000 g, 15 min). The resulting pellets were resuspended in buffer A (50 mM Tris/HCl pH 7.5, 1 M NaCl, 20 mM imidazole) containing 6 M guanidium hydrochloride (GDN). The suspension was sonicated to disrupt the cells (using a 750 W sonicator and 3 cycles of 30 s each at 45 % power output) and then centrifuged at 14,000 g for 30 min at 20 °C. The supernatant was first purified by Immobilized Metal Affinity Chromatography (IMAC) by loading the clarified lysate onto a 5 mL Nickel column (GE Healthcare) pre-equilibrated in buffer A. The affinity resin was washed with 20 column volumes (CV) of buffer A. Proteins were eluted with ~3 CV of buffer A supplemented with 250 mM imidazole. The fractions eluted from the Nickel column were pooled and concentrated in the presence of 6 M GDN up to 1 mM using Centricon concentrators, and the proteins were then frozen at -20 °C. All PNT3 variants were subsequently

subjected to SEC, where the SEC column was equilibrated with buffer B (Sodium Phosphate 50 mM pH 7.2 100 mM NaCl, 5 mM EDTA). The fractions from SEC, were pooled, supplemented with 6M GDN and concentrated (up to ~750  $\mu$ M) and stored at -20 °C. In the case of the PNT3 cysteine variant, the GDN-containing sample was also supplemented with 10 mM DTT. Prior to each subsequent analysis, the samples were loaded onto a Sephadex G-25 medium column (GE Healthcare) to exchange the buffer. The proteins were eluted from G-25 columns using Sodium Phosphate 50 mM buffer at a pH of 7.2 unless differently specified. IMAC and SEC were performed at room temperature (RT).

Protein concentrations were estimated using the theoretical absorption coefficients at 280 nm as obtained using the program ProtParam from the EXPASY server.

The purity of the final purified products was assessed by SDS-PAGE (**Fig. 3B**). The identity of all the purified PNT3 variants generated in this work was confirmed by mass spectrometry analysis of tryptic fragments obtained after digestion of the purified protein bands excised from SDS-polyacrylamide gels (data not shown). The excised bands were analyzed by the mass spectrometry facility of Marseille Proteomics in the same way as previously done for the PNT3 wt variant [25]. Briefly, proteins were reduced by incubation with 100 mM dithiothreitol (DTT) for 45 minutes at 56 °C and free cysteine residues were capped by incubation with 100 mM iodoacetamide for 30 h at 25 °C in the dark. Samples were digested with porcine trypsin (V5111, Promega) at 12,5 ng/ $\mu$ l in 25 mM  $\text{NH}_4\text{HCO}_3$ , for 18 h at 37 °C. Peptides were extracted from the gel with 60% vol/vol acetonitrile in 5% formic acid, dried under vacuum, and reconstituted in 5  $\mu$ l of 50% vol/vol acetonitrile in 0.3% trifluoroacetic acid. Mass analyses of the tryptic fragments were performed on a MALDI-TOF-TOF Bruker Ultraflex III spectrometer (Bruker Daltonics, Wissembourg, France) controlled by the Flexcontrol 3.0 package (Build 51). This instrument was used at a maximum accelerating potential of 25 kV and was operated in reflectron mode and an m/z range from 600 to 4500 (RP\_Wide range\_Method) or 600 to 3400 (RP\_Proteomics\_2015\_Method). The laser frequency was fixed to 200 Hz and approximately 1000-1500 shots by sample were cumulated. Five external standards (Peptide Calibration Standard, Bruker Daltonics) were used to calibrate each spectrum to a mass accuracy within 50 ppm. Peak picking was performed with Flexanalysis 3.0 software (Bruker) with an adapted analysis method. Parameters used were as follows: SNAP peak detection algorithm, S/N threshold fixed to 6 and a quality factor threshold of 30. One  $\mu$ l of sample was mixed with 1  $\mu$ l of a saturated HCCA ( $\alpha$ -cyano-4-hydroxycinnamic acid) solution in acetonitrile/0.3%TFA (1:1) then 1  $\mu$ l was spotted on the target, dried and analyzed with the previously described method. Because of the high sequence similarity between the PNT3<sup>A1,2,3</sup> variants, a MSMS analysis was necessary to unambiguously identify each of them.

#### 4.3. PEG precipitation assay (relative solubility)

The relative solubility of each variant was evaluated at different PEG concentrations using an adaptation of the protocol recently described by Oeller et al. [32]. Briefly, PEG solutions from 0 to 30 % were prepared from 50 % PEG<sub>6000</sub> stock solution. Then, an aliquot of corresponding protein from a stock (at 260  $\mu$ M) was mixed with each PEG solutions to obtain a final concentration of 66  $\mu$ M protein in a 100  $\mu$ L final reaction volume. The assay was performed in 96-well plates sealed with aluminum plate sealers to prevent possible evaporation (Corning). Plates were incubated at 4 °C for 24 hours and centrifugated at maximum velocity (4600  $\times$  g) for 2 hours. Immediately after, 2  $\mu$ L of the supernatant were pipetted to quantify the soluble protein concentration using a ND-1000 Nanodrop Spectrophotometer and theoretical absorption coefficients at 280 nm as obtained using the program ProtParam from the EXPASY server. Each condition was made in triplicate. The soluble fractions obtained were normalized and fitted to a sigmoid function to obtain the PEG<sub>1/2</sub> value that reports on relative solubility (PRISM software). The error on the PEG<sub>1/2</sub> and the quality of the fit were estimated by a 95 % confidence interval analysis (PRISM software).

#### 4.4. Far-UV circular dichroism

CD spectra were measured using a Jasco 810 dichrograph, flushed with N<sub>2</sub> and equipped with a Peltier thermoregulation system. Proteins were loaded into a 1-mm quartz cuvette at 0.06 mg/mL (in 10 mM phosphate buffer at pH 7.2) and spectra were recorded at 37 °C. The scanning speed was 20 nm min<sup>-1</sup>, with data pitch of 0.2 nm. Each spectrum is the average of ten acquisitions. The spectrum

of buffer was subtracted from the protein spectrum. Spectra were smoothed using the “means-movement” smoothing procedure implemented in the Spectra Manager package. As already previously documented, a decrease in the signal spectrum was observed with increasing incubation time and ascribed to fibril formation [25]. Because the different variants have different fibrillation propensities, differences in spectra intensity might reflect differences in the fraction of fibrillar species (which are not detected by CD) rather than *bona fide* spectral differences. We therefore normalized spectra using the maximum negative value of intensity as a normalization factor.

Mean molar ellipticity values per residue (MRE) were calculated as

$$[\theta]=3300 \times m \times \Delta A/(l \times c \times n) \quad (\text{Eq. 2})$$

where  $l$  is the path length in cm,  $n$  is the number of residues,  $m$  is the molecular mass in Daltons and  $c$  is the concentration of the protein in mg mL<sup>-1</sup>.

#### 4.5. Estimation of the hydrodynamic radius by SEC

The hydrodynamic radii (Stokes radii,  $R_s$ ) of the proteins were estimated by analytical SEC using a HiLoad 16/600 Superdex 75 pg column (Cytiva). Buffer B was used as elution buffer. Typically, 250  $\mu$ L of purified protein at 11 mg mL<sup>-1</sup> were injected.

The Stokes radii of proteins eluted from the SEC column were deduced from a calibration curve obtained using globular proteins of known  $R_s$  (Conalbumin: 36.4 Å, Carbonic Anhydrase: 23 Å, RNase A: 16.4 Å Aprotinin:13.5 Å)

The  $R_s$  (in Å) of a natively folded ( $R_s^{NF}$ ), fully unfolded state in urea ( $R_s^U$ ) and natively unfolded premolten globule (PMG) ( $R_s^{PMG}$ ) protein with a molecular mass ( $MM$ ) (in Daltons) were calculated according to [43]:

$$\log (R_s^{NF}) = 0.357 \times (\log MM) - 0.204 \quad (\text{Eq. 3})$$

$$\log (R_s^U) = 0.521 \times (\log MM) - 0.649 \quad (\text{Eq. 4})$$

$$\log (R_s^{PMG}) = 0.403 \times (\log MM) - 0.239 \quad (\text{Eq. 5})$$

The  $R_s$  (in Å) of an IDP with  $N$  residues was also calculated according to [44] using the simple power-law model:

$$R_s^{IDP} = R_0 N^{\nu} \quad (\text{Eq. 6})$$

where  $R_0 = 2.49$  and  $\nu = 0.509$ . The compaction index ( $CI$ ) is expressed according to [45]:

$$CI = (R_s^U - R_s^{obs}) / (R_s^U - R_s^{NF}) \quad (\text{Eq. 7})$$

This parameter, which allows comparison between proteins of different lengths, in principle varies between 0 and 1, with 0 indicating minimal compaction and 1 maximal compaction.

#### 4.6. Small-angle X-ray scattering (SAXS)

In order to ensure maximal monodispersity of the sample, SAXS studies were coupled to SEC. SEC-SAXS data were collected at SOLEIL (Gif-sur-Yvette, France) as described in **Table 3**. In both cases, the calibration was made with water. Sample from each PNT3 variant at 5 mg mL<sup>-1</sup> in buffer B containing 6M GDN was injected onto an AdvanceBio SEC 2.7  $\mu$ m (Agilent) SEC column. Elution was carried out in buffer C (50 mM sodium phosphate buffer at pH 7.2). Data reduction and frames subtraction were done with the beamline software FOXTROT. Gaussian decomposition was performed using the UltraScan solution modeler (US-SOMO) HPLC-SAXS module [46] or Chromixs (manual frames selection) [47] and the final deconvoluted scattering curves were submitted to the SHANUM program [48] to remove noisy, non-informative data at high angles.

**Table 3.** SEC-SAXS data acquisition parameters.

Instrument	SOLEIL Synchrotron (Gif-sur-Yvette, France) Beamline Swing
X-rays wavelength (Å)	1.033
Energy (keV)	12
Detector type	Dectris EIGER 4M

Sample-to-detector distance (m)	2.0
q-range	0.003 – 0.549 Å <sup>-1</sup>
Temperature (°C)	20
<b>Samples</b>	
Concentration (mg mL <sup>-1</sup> )	5
Sample volume (μL)	50
Gel filtration column	AdvanceBio SEC 2.7 μm (Agilent)
Flow rate (mL min <sup>-1</sup> )	0.3
Buffer	50 mM sodium phosphate pH 7.2 (buffer C)

The data were analyzed using the ATSAS program package [48]. The radius of gyration ( $R_g$ ) and  $I(0)$  were estimated at low angles ( $q \cdot R_g < 1.3$ ) according to the Guinier approximation [49,50]:

$$\ln[I(q)] = \ln[I_0] - \frac{q^2 R_g^2}{3} \quad (\text{Eq. 8})$$

The pairwise distance distribution functions  $P(r)$ , from which the  $D_{max}$  and the  $R_g$  were estimated, were calculated with the program GNOM [51] and manually adjusted until a good CorMap p-value ( $\alpha > 0.01$ ) was obtained [51].

The theoretical  $R_g$  value (in Å) expected for various conformational states was calculated using Flory's equation

$$R_g = R_0 N^\nu \quad (\text{Eq. 9})$$

where  $N$  is the number of amino acid residues,  $R_0$  a constant and  $\nu$  a scaling factor. For IDPs,  $R_0$  is  $2.54 \pm 0.01$  and  $\nu$  is  $0.522 \pm 0.01$  [52], for chemically denatured (U) proteins  $R_0$  is  $1.927 \pm 0.27$  and  $\nu$  is  $0.598 \pm 0.028$  [52], and for natively folded (NF) proteins  $R_0 = \sqrt{(3/5)} \times 4.75$  and  $\nu = 0.29$  [53].

As in the case of the  $R_s$ , the  $CI$  allows comparing the degree of compaction of a given IDP, through comparison of the observed  $R_g$  to the reference values expected for a fully unfolded and a folded conformation of identical mass. The  $CI$  referred to the  $R_g$  can be calculated as follows [45]:

$$CI = (R^U - R_g^{obs}) / (R_g^U - R_g^{NF}) \quad (\text{Eq. 10})$$

where  $R_g^{obs}$  is the experimental value for a given protein, and  $R_g^U$  and  $R_g^{NF}$  are the reference values calculated for a fully unfolded (U) and natively folded (NF) form as described above. Akin to the  $R_s$ -based  $CI$ , this index increases with increasing compaction.

The overall conformation and the flexibility of the proteins was assessed with the dimensionless Kratky plot ( $(qR_g)^2 I(q)/I_0$  vs  $qR_g$ ) and the Kratky-Debye plot ( $q^2 I(q)$  vs  $q^2$ ).

SEC-SAXS data have been deposited in the Small Angle Scattering Biological Data Bank (SASBDB) [54] under codes SASDQB7, SASDQC7, SASDQD7, SASDQE7 and SASDQF7 for the set of data of PNT3 *wt*, PNT3<sup>3A</sup>, NiV PNT3, PNT3\_C-terminal truncated and PNT3<sup>3A</sup>\_C-terminal truncated, respectively.

#### 4.7. Congo red binding assays

Quantitative measurement of Congo Red (CR, Sigma-Aldrich) binding (CR shift assay) was carried out by using protein samples containing each PNT3 variant at 20 μM (in Buffer C) and 5 μM of CR in a final volume of 100 μL. The samples were then incubated at 37 °C for 4 or 7 days. The adsorption spectrum of the CR-containing samples was recorded using a PHERAstar FSX Microplate Reader (BMG LABTECH) in the 350-600 nm wavelength range. A solution of 5 μM CR in Buffer C without the protein was used as a control to normalize the analysis. Experiments were carried out in triplicate. Statistical analysis was done using Two-way ANOVA test implemented in the PRISM software.

#### 4.8. Negative-staining transmission electron microscopy (TEM)

All the variants, at a concentration of 200 μM, were prepared and analyzed at different times to monitor their evolution (0, 24, 96 h). Incubation was carried out at 37 °C in Buffer C. Prior to each measurement, the samples were diluted to reach a final concentration of 40 μM. EM grids (carbon



coated copper grids, 300 mesh, Agar Scientific) were exposed to plasma glow discharge for 20 seconds using GloQube (Quorum) (Current 15mA) in order to increase protein adhesion. Drops of 3.5  $\mu$ L of the diluted protein solutions were deposited onto glow-discharged grids. After 1min incubation with the sample, the grids were washed three times with 50  $\mu$ L of buffer C, once in 35  $\mu$ L 1 % (w/v) Uranyl acetate solution (LauryLab-France) and then stained for 1 min in the latter solution. Excess of uranyl was blotted and grids were left to dry for 1 h at RT. Images were collected on Tecnai 120 Spirit TEM microscope (Thermo Fisher Scientific) operated at 120 kV using a Veleta 2Kx2K CCD camera (Olympus).

#### 4.9. Kinetic protein aggregation study by Taylor Dispersion Analysis (TDA)

TDA was performed as already described [40] using an Agilent 7100 (Waldbronn, Germany) capillary electrophoresis system with bare fused silica capillaries (Polymicro Technologies) having 60 cm  $\times$  50  $\mu$ m i.d. dimensions and a detection window at 51.5 cm. New capillaries were conditioned with the following flushes: 1 M NaOH for 30 min and ultrapure water for 30 min. Between each analysis, the capillaries were rinsed with Buffer C (2 min). Samples were injected hydrodynamically on the inlet end of the capillary (30 mbar, 6 s, injected volume is about 6.1 nL corresponding to less than 1% of the capillary volume to the detection point). Experiments were performed using a mobilization pressure of 100 mbar. The temperature of the capillary cartridge was set at 37  $^{\circ}$ C. The vial carousel was thermostated using an external circulating water bath from Instrumat (France). The solutes were monitored by UV absorbance at 198 nm. The mobile phase was Buffer C. (viscosity at 37  $^{\circ}$ C is  $7.54 \times 10^{-4}$  Pa s). Samples obtained after the desalting column were diluted to reach 700  $\mu$ L at 200  $\mu$ M solution, and were immediately transferred to a vial and incubated at 37  $^{\circ}$ C in the capillary electrophoresis instrument's carousel. The aggregation was conducted by injecting the sample every 1 h. The total number of TDA runs for each sample was about 120. The taylorgrams were recorded with Agilent Chemstation software and then exported to Microsoft Excel for subsequent data processing. Data were fitted to a first order exponential decay according to equation 1. This resulted in a good quality fit as judged from the Reduced Chi-Sqr: 0.43723; R-Square (COD)= 0.93943, and Adj. R-Square= 0.93576. When necessary, the elution peaks were fitted with the sum of  $n$  Gaussian functions (in this work:  $n \leq 3$ ) as already described [40,41].

**Acknowledgements.** We thank Aurélien Thureau (SOLEIL) and Anton Popov (ESRF) for their help in recording SEC-SAXS data. We thank both the ESRF and the SOLEIL synchrotrons for beamtime allocation. We are also grateful to Gerlind Sulzenbacher (AFMB lab) for efficiently managing the AFMB BAG. We thank Patrick Fourquet for mass spectrometry analyses done using the mass spectrometry facility of Marseille Proteomics ([marseille-proteomique.univ-amu.fr](http://marseille-proteomique.univ-amu.fr)), supported by IBISA (Infrastructures Biologie Santé et Agronomie), Plateforme Technologique Aix-Marseille, the Cancéropôle PACA, Région Sud-Alpes-Côte d'Azur, the Institut Paoli-Calmettes, the Centre de Recherche en Cancérologie de Marseille (CRCM), Fonds Européen de Développement Régional and Plan Cancer.

**Funding.** This work was carried out with the financial support of the Agence Nationale de la Recherche (ANR), specific project Heniphase (ANR-21-CE11-0012-01). It was also partly supported by the French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INSB-0005) and by the CNRS. F. G. is supported by a post-doctoral fellowship from the FRM (Fondation pour la Recherche Médicale). G. P. is supported by a joint doctoral fellowship from the AID (Agence Innovation Défense) and Aix-Marseille University. J. N. is supported by a postdoctoral fellowship from the Infectiopôle Sud.

## References

1. Eaton, B.T.; Mackenzie, J.S.; Wang, L.F. Henipaviruses. In *Fields Virology*, 5th ed.; Fields, B.N., Knipe, D.M., Howley, P.M., Eds.; Lippincott-Raven: Philadelphia, 2007; pp. 1587-1600.
2. Eaton, B.T.; Broder, C.C.; Middleton, D.; Wang, L.F. Hendra and Nipah viruses: different and dangerous. *Nat. Rev. Microbiol.* **2006**, *4*, 23-35, doi:10.1038/nrmicro1323.
3. Bloyet, L.M.; Welsch, J.; Enchery, F.; Mathieu, C.; de Breyne, S.; Horvat, B.; Grigorov, B.; Gerlier, D. HSP90 Chaperoning in Addition to Phosphoprotein Required for Folding but Not for Supporting Enzymatic Activities of Measles and Nipah Virus L Polymerases. *J. Virol.* **2016**, *90*, 6642-6656, doi:10.1128/JVI.00602-16.
4. Abdella, R.; Aggarwal, M.; Okura, T.; Lamb, R.A.; He, Y. Structure of a paramyxovirus polymerase complex reveals a unique methyltransferase-CTD conformation. *Proc Natl Acad Sci U S A* **2020**, *117*, 4931-4941, doi:10.1073/pnas.1919837117.
5. Bloyet, L.M.; Schramm, A.; Lazert, C.; Raynal, B.; Hologne, M.; Walker, O.; Longhi, S.; Gerlier, D. Regulation of measles virus gene expression by P protein coiled-coil properties. *Sci Adv* **2019**, *5*, eaaw3702, doi:10.1126/sciadv.aaw3702.
6. Karlin, D.; Ferron, F.; Canard, B.; Longhi, S. Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* **2003**, *84*, 3239-3252, doi:10.1099/vir.0.19451-0
7. Habchi, J.; Mamelli, L.; Darbon, H.; Longhi, S. Structural Disorder within Henipavirus Nucleoprotein and Phosphoprotein: From Predictions to Experimental Assessment. *PLoS ONE* **2010**, *5*, e11684, doi:10.1371/journal.pone.0011684.
8. Schiavina, M.; Salladini, E.; Murralli, M.G.; Tria, G.; Felli, I.C.; Pierattelli, R.; Longhi, S. Ensemble description of the intrinsically disordered N-terminal domain of the Nipah virus P/V protein from combined NMR and SAXS. *Sci Rep* **2020**, *10*, 19574, doi:10.1038/s41598-020-76522-3.
9. Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321-331., doi:10.1006/jmbi.1999.3110.
10. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415-427, doi:10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7
11. Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **2012**, *37*, 509-516, doi:10.1016/j.tibs.2012.08.004.
12. Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V.N. Introducing Protein Intrinsic Disorder. *Chem Rev* **2014**, *114*, 6561-6588, doi:10.1021/cr400514h.
13. Jensen, M.R.; Yabukarski, F.; Communie, G.; Condamine, E.; Mas, C.; Volchkova, V.; Tarbouriech, N.; Bourhis, J.M.; Volchkov, V.; Blackledge, M.; et al. Structural Description of the Nipah Virus Phosphoprotein and Its Interaction with STAT1. *Biophys. J.* **2020**, *118*, 2470-2488, doi:10.1016/j.bpj.2020.04.010.
14. Bruhn-Johannsen, J.F.; Barnett, K.; Bibby, J.; Thomas, J.; Keegan, R.; Rigden, D.; Bornholdt, Z.A.; Saphire, E.O. Crystal structure of the Nipah virus phosphoprotein tetramerization domain. *J. Virol.* **2014**, *88*, 758-762, doi:10.1128/JVI.02294-13.
15. Blocquel, D.; Beltrandi, M.; Erales, 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* **2013**, *446*, 162-172, doi:10.1016/j.virol.2013.07.031.
16. Beltrandi, M.; Blocquel, D.; Erales, J.; Barbier, P.; Cavalli, A.; Longhi, S. Insights into the coiled-coil organization of the Hendra virus phosphoprotein from combined biochemical and SAXS studies. *Virology* **2015**, *477*, 42-55, doi:10.1016/j.virol.2014.12.029.
17. Habchi, J.; Blangy, S.; Mamelli, L.; Ringkjøbing-Jensen, M.; Blackledge, M.; Darbon, H.; Oglesbee, M.; Shu, Y.; Longhi, S. Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipaviruses. *J. Biol. Chem.* **2011**, *286*, 13583-13602, doi:10.1074/jbc.M111.219857
18. Communie, G.; Habchi, J.; Yabukarski, F.; Blocquel, D.; Schneider, R.; Tarbouriech, N.; Papageorgiou, N.; Ruigrok, R.W.; Jamin, M.; Ringkjøbing-Jensen, M.; et al. Atomic

- resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. *PLoS Pathog* **2013**, *9*, e1003631, doi:e1003631.
19. Douglas, J.; Drummond, A.J.; Kingston, R.L. Evolutionary history of cotranscriptional editing in the paramyxoviral phosphoprotein gene. *Virus Evolution* **2021**, *7*, doi:10.1093/ve/veab028.
  20. Salladini, E.; Delauzun, V.; Longhi, S. The Henipavirus V protein is a prevalently unfolded protein with a zinc-finger domain involved in binding to DDB1. *Mol Biosyst* **2017**, *13*, 2254-2267, doi:10.1039/c7mb00488e.
  21. Pesce, G.; Gondelaud, F.; Ptchelkine, D.; Nilsson, J.F.; Bignon, C.; Cartalas, J.; Fourquet, P.; Longhi, S. Experimental Evidence of Intrinsic Disorder and Amyloid Formation by the Henipavirus W Proteins. *Int J Mol Sci* **2022**, *23*, doi:10.3390/ijms23020923.
  22. Fontana, J.M.; Bankamp, B.; Rota, P.A. Inhibition of interferon induction and signaling by paramyxoviruses. *Immunol. Rev.* **2008**, *225*, 46-67, doi:10.1111/j.1600-065X.2008.00669.x.
  23. Audsley, M.D.; Moseley, G.W. Paramyxovirus evasion of innate immunity: Diverse strategies for common targets. *World J Virol* **2013**, *2*, 57-70, doi:10.5501/wjv.v2.i2.57.
  24. Tsimbalyuk, S.; Cross, E.M.; Hoad, M.; Donnelly, C.M.; Roby, J.A.; Forwood, J.K. The Intrinsically Disordered W Protein Is Multifunctional during Henipavirus Infection, Disrupting Host Signalling Pathways and Nuclear Import. *Cells* **2020**, *9*, doi:10.3390/cells9081913.
  25. Salladini, E.; Gondelaud, F.; Nilsson, J.; Pesce, G.; Bignon, C.; Murralli, M.G.; Horvat, B.; Fabre, R.; Pierattelli, R.; Kajava, A.V.; et al. Identification of a region in the common amino-terminal domain of Hendra virus P, V and W proteins responsible for phase transition and amyloid formation. *Biomolecules* **2021**, *11*, 1324, doi:10.3390/biom11091324.
  26. Gondelaud, F.; Pesce, G.; Nilsson, J.F.; Bignon, C.; Ptchelkine, D.; Gerlier, D.; Mathieu, C.; Longhi, S. Functional benefit of structural disorder for the replication of measles, Nipah and Hendra viruses. *Essays in Biochemistry* **2022**, *in press*, doi:10.1042/EBC20220045.
  27. Kajava, A.V.; Baxa, U.; Steven, A.C. Beta arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils. *FASEB J.* **2010**, *24*, 1311-1319, doi:10.1096/fj.09-145979.
  28. Ahmed, A.B.; Znassi, N.; Château, M.T.; Kajava, A.V. A structure-based approach to predict predisposition to amyloidosis. *Alzheimers Dement* **2015**, *11*, 681-690, doi:10.1016/j.jalz.2014.06.007.
  29. Swaminathan, R.; Ravi, V.K.; Kumar, S.; Kumar, M.V.; Chandra, N. Lysozyme: a model protein for amyloid research. *Adv Protein Chem Struct Biol* **2011**, *84*, 63-111, doi:10.1016/b978-0-12-386483-3.00003-3.
  30. Tseng, W.H.; Chen, S.H.; Hiramatsu, H. pH-controlled stacking direction of the  $\beta$ -strands in peptide fibrils. *Sci Rep* **2020**, *10*, 22199, doi:10.1038/s41598-020-79001-x.
  31. Pfefferkorn, C.M.; McGlinchey, R.P.; Lee, J.C. Effects of pH on aggregation kinetics of the repeat domain of a functional amyloid, Pmel17. *Proc Natl Acad Sci U S A* **2010**, *107*, 21447-21452, doi:10.1073/pnas.1006424107.
  32. Oeller, M.; Sormanni, P.; Vendruscolo, M. An open-source automated PEG precipitation assay to measure the relative solubility of proteins with low material requirement. *Sci Rep* **2021**, *11*, 21932, doi:10.1038/s41598-021-01126-4.
  33. Uversky, V.N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **2002**, *11*, 739-756., doi:10.1110/ps.4210102.
  34. Schramm, A.; Bignon, C.; Brocca, S.; Grandori, R.; Santambrogio, C.; Longhi, S. An arsenal of methods for the experimental characterization of intrinsically disordered proteins - How to choose and combine them? *Arch Biochem Biophys* **2019**, *676*, 108055, doi:10.1016/j.abb.2019.07.020.
  35. Anand, B.G.; Prajapati, K.P.; Shekhawat, D.S.; Kar, K. Tyrosine-Generated Nanostructures Initiate Amyloid Cross-Seeding in Proteins Leading to a Lethal Aggregation Trap. *Biochemistry* **2018**, *57*, 5202-5209, doi:10.1021/acs.biochem.8b00472.
  36. Hoyer, W.; Cherny, D.; Subramaniam, V.; Jovin, T.M. Impact of the acidic C-terminal region comprising amino acids 109-140 on alpha-synuclein aggregation in vitro. *Biochemistry* **2004**, *43*, 16233-16242, doi:10.1021/bi048453u.

37. Sorrentino, Z.A.; Giasson, B.I. The emerging role of  $\alpha$ -synuclein truncation in aggregation and disease. *J. Biol. Chem.* **2020**, *295*, 10224-10244, doi:10.1074/jbc.REV120.011743.
38. Gruet, A.; Dosnon, M.; Blocquel, D.; Brunel, J.; Gerlier, D.; Das, R.K.; Bonetti, D.; Gianni, S.; Fuxreiter, M.; Longhi, S.; et al. Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. *FEBS J.* **2016**, *283*, 576-594, doi:10.1111/febs.13631.
39. Troilo, F.; Bonetti, D.; Bignon, C.; Longhi, S.; Gianni, S. Understanding Intramolecular Crosstalk in an Intrinsically Disordered Protein. *ACS Chem Biol* **2019**, *14*, 337-341, doi:10.1021/acscchembio.8b01055.
40. Deleanu, M.; Hernandez, J.F.; Cipelletti, L.; Biron, J.P.; Rossi, E.; Taverna, M.; Cottet, H.; Chamieh, J. Unraveling the Speciation of  $\beta$ -Amyloid Peptides during the Aggregation Process by Taylor Dispersion Analysis. *Anal Chem* **2021**, *93*, 6523-6533, doi:10.1021/acs.analchem.1c00527.
41. Deleanu, M.; Deschaume, O.; Cipelletti, L.; Hernandez, J.F.; Bartic, C.; Cottet, H.; Chamieh, J. Taylor Dispersion Analysis and Atomic Force Microscopy Provide a Quantitative Insight into the Aggregation Kinetics of  $A\beta$  (1-40)/ $A\beta$  (1-42) Amyloid Peptide Mixtures. *ACS Chem Neurosci* **2022**, *13*, 786-795, doi:10.1021/acscchemneuro.1c00784.
42. Deleanu, M.; Hernandez, J.F.; Cottet, H.; Chamieh, J. Taylor dispersion analysis discloses the impairment of  $A\beta$  peptide aggregation by the presence of a fluorescent tag. *Electrophoresis* **2022**, *in press*.
43. Uversky, V.N. What does it mean to be natively unfolded? *Eur. J. Biochem.* **2002**, *269*, 2-12., doi: 10.1046/j.0014-2956.2001.02649.x.
44. Marsh, J.A.; Forman-Kay, J.D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **2010**, *98*, 2383-2390, doi:10.1016/j.bpj.2010.02.006.
45. Brocca, S.; Testa, L.; Sobott, F.; Samalikova, M.; Natalello, A.; Papaleo, E.; Lotti, M.; De Gioia, L.; Doglia, S.M.; Alberghina, L.; et al. Compaction properties of an intrinsically disordered protein: sic1 and its kinase-inhibitor domain. *Biophys. J.* **2011**, *100*, 2243-2252, doi:10.1016/j.bpj.2011.02.055.
46. Brookes, E.; Rocco, M. Recent advances in the UltraScan SOLUTION MOdeller (US-SOMO) hydrodynamic and small-angle scattering data analysis and simulation suite. *Eur. Biophys. J.* **2018**, *47*, 855-864, doi:10.1007/s00249-018-1296-0.
47. Panjkovich, A.; Svergun, D.I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **2018**, *34*, 1944-1946, doi:10.1093/bioinformatics/btx846.
48. Manalastas-Cantos, K.; Konarev, P.V.; Hajizadeh, N.R.; Kikhney, A.G.; Petoukhov, M.V.; Molodenskiy, D.S.; Panjkovich, A.; Mertens, H.D.T.; Gruzinov, A.; Borges, C.; et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr* **2021**, *54*, 343-355, doi:10.1107/s1600576720013412.
49. Guinier, A. La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Ann Phys (Paris)* **1939**, *12*, 161-237.
50. Guinier, A.; Fournet, F. *Small angle scattering of X-rays*; Wiley Interscience: New York., 1955.
51. Svergun, D. Determination of the regularization parameters in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **1992**, *25*, 495-503., doi:10.1107/S0021889892001663.
52. Bernado, P.; Blackledge, M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys. J.* **2009**, *97*, 2839-2845, doi:10.1016/j.bpj.2009.08.044.
53. Wilkins, D.K.; Grimshaw, S.B.; Receveur, V.; Dobson, C.M.; Jones, J.A.; Smith, L.J. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **1999**, *38*, 16424-16431., doi: 10.1021/bi991765q
54. 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.* **2015**, *43*, D357-363, doi:10.1093/nar/gku1047.