



HAL
open science

Experimental Evidence of Intrinsic Disorder and Amyloid Formation by the Henipavirus W Proteins

Giulia Pesce, Frank Gondelaud, Denis Ptchelkine, Juliet F Nilsson, Christophe Bignon, Jérémy Cartalas, Patrick Fourquet, Sonia Longhi

► **To cite this version:**

Giulia Pesce, Frank Gondelaud, Denis Ptchelkine, Juliet F Nilsson, Christophe Bignon, et al.. Experimental Evidence of Intrinsic Disorder and Amyloid Formation by the Henipavirus W Proteins. International Journal of Molecular Sciences, 2022, 23 (2), pp.923. 10.3390/ijms23020923. hal-03533043

HAL Id: hal-03533043

<https://hal.science/hal-03533043v1>

Submitted on 18 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Article

Experimental Evidence of Intrinsic Disorder and Amyloid Formation by the *Henipavirus* W Proteins

Giulia Pesce^{1,†}, Frank Gondelaud^{1,†}, Denis Ptchelkine¹, Juliet F. Nilsson¹ , Christophe Bignon¹,
Jérémy Cartalas¹, Patrick Fourquet² and Sonia Longhi^{1,*}

- ¹ 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, France; giulia.pesce@univ-amu.fr (G.P.); frank.gondelaud@univ-amu.fr (F.G.); denis.PTCHELKINE@univ-amu.fr (D.P.); juliet.nilsson@univ-amu.fr (J.F.N.); christophe.bignon@univ-amu.fr (C.B.); jeremycartalas@orange.fr (J.C.)
- ² INSERM, Centre de Recherche en Cancérologie de Marseille (CRCM), Centre National de la Recherche Scientifique (CNRS), Marseille Protéomique, Institut Paoli-Calmettes, Aix Marseille University, 27 Bvd Lei Roure, CS 30059, 13273 Marseille, France; patrick.fourquet@inserm.fr
- * Correspondence: Sonia.longhi@univ-amu.fr
- † Equal contribution to the work.

Abstract: Henipaviruses are severe human pathogens within the *Paramyxoviridae* family. Beyond the P protein, the *Henipavirus* P gene also encodes the V and W proteins which share with P their N-terminal, intrinsically disordered domain (NTD) and possess a unique C-terminal domain. *Henipavirus* W proteins antagonize interferon (IFN) signaling through NTD-mediated binding to STAT1 and STAT4, and prevent type I IFN expression and production of chemokines. Structural and molecular information on *Henipavirus* W proteins is lacking. By combining various bioinformatic approaches, we herein show that the *Henipaviruses* W proteins are predicted to be prevalently disordered and yet to contain short order-prone segments. Using limited proteolysis, differential scanning fluorimetry, analytical size exclusion chromatography, far-UV circular dichroism and small-angle X-ray scattering, we experimentally confirmed their overall disordered nature. In addition, using Congo red and Thioflavin T binding assays and negative-staining transmission electron microscopy, we show that the W proteins phase separate to form amyloid-like fibrils. The present study provides an additional example, among the few reported so far, of a viral protein forming amyloid-like fibrils, therefore significantly contributing to enlarge our currently limited knowledge of viral amyloids. In light of the critical role of the *Henipavirus* W proteins in evading the host innate immune response and of the functional role of phase separation in biology, these studies provide a conceptual asset to further investigate the functional impact of the phase separation abilities of the W proteins.

Keywords: intrinsically disordered proteins/regions; small-angle X-ray scattering; innate immune response evasion; phase separation; fibrillation; biocondensates; Congo red and Thioflavin T binding assays; negative-staining electron microscopy; amyloid-like fibrils; viral proteins



Citation: 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*, 923. <https://doi.org/10.3390/ijms23020923>

Academic Editor: Rashid Giniatullin

Received: 15 December 2021

Accepted: 12 January 2022

Published: 15 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Nipah and Hendra viruses (NiV and HeV) are zoonotic pathogens gathered in the *Henipavirus* genus within the *Paramyxoviridae* family in the *Mononegavirales* order. In human beings they cause severe acute respiratory syndrome, generalized vasculitis and lethal encephalitis. Their natural reservoir are fruit-eating bats of the *Pteropus* genus, the so-called flying foxes [1,2]. HeV came to light in 1994 in the Hendra suburb of Brisbane (Australia) as a new agent responsible for a sudden outbreak of an acute respiratory and neurologic disease within horses [1]. Exposure to body fluids and tissues or excretions of infected horses resulted in transmission to humans, with a high case fatality (~60%). NiV emerged in 1998 in Malaysia as the causative agent of an outbreak of disease in

pigs and humans with 265 human cases of encephalitis and 105 deaths [1]. Since its first emergence, NiV reappeared in 2001 in Bangladesh (where outbreaks of encephalitis, with an average case fatality rate of 80%, have occurred almost annually) and successively in India and in Philippines. Owing to the wide distribution of bats and to the broad spectrum of susceptible animals, additional spillover events into humans are expected to occur in the future. Moreover, and by contrast to HeV, NiV can also be transmitted from human-to-human, with this type of transmission having contributed to more than 50% of the virus spread in Bangladesh [3,4], Philippines [5] and India [6]. The inter-human transmission extends the potential of NiV to cause deadly outbreaks in humans. Although an efficient vaccine against HeV in horses is available, neither vaccines nor therapeutic treatments are available in humans. Because of their high virulence, wide host range and interspecies transmission, HeV and NiV have been classified as biosafety level 4 (BSL4) pathogens and as potential bio-terrorism agents. NiV is even classified in the WHO blueprint list of the eight diseases for which research should be prioritized (<https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts>) (accessed on 14 December 2021).

Henipaviruses have a non-segmented, negative-stranded RNA genome that is encapsidated by the nucleoprotein (N) within a helical nucleocapsid [7]. This nucleocapsid is the template used by the RNA-dependent RNA polymerase (RdRp) for transcription and replication. The RdRp is made of the large (L) protein and the phosphoprotein (P). The P protein is an essential polymerase cofactor as it enables recruitment of L onto the nucleocapsid template, and serves as a chaperon for both L [8–10] and N [11].

The N and P proteins from henipaviruses possess long intrinsically disordered regions (IDRs) [12–14], i.e., regions devoid of stable secondary and tertiary structure [15–19]. The *Henipavirus* P protein consists of a long N-terminal intrinsically disordered domain (referred to as NTD) and a C-terminal region that encompasses both structured and disordered regions (Figure 1) [13,14,20–25].

Like in many paramyxoviruses [26], the P gene from both HeV and NiV also encodes three non-structural proteins: C, V and W. While the C protein is encoded in an alternative reading frame of the P gene, the V and W proteins (~50 kDa) result from the addition of either one (V protein) or two (W protein) non-templated guanosines at the editing site of the P messenger (Figure 1). The latter is located at the end of the region encoding the NTD of P (Figure 1). Consequently, the P, V and W proteins share a common NTD but have distinct C-terminal domains (referred to as P_{CTD}, V_{CTD} and W_{CTD} respectively) (Figure 1).

The V and W proteins are key players in the evasion of the antiviral type I interferon (IFN-I)-mediated response [27–29]. The V protein prevents the detection of viral dsRNA by binding to melanoma differentiation-associated protein 5 (MDA5) and Laboratory of Genetics and Physiology 2 (LGP2) protein through its CTD [30], and to PLK1 (polo-like kinase) through its disordered NTD [31] (Figure 1). The V and W proteins have also an antagonist activity of IFN signaling by targeting Signal Transducers and Activators of Transcription (STAT) proteins [27]. The V and W proteins bind STAT1 through their NTD, with this ability being also conserved in NiV P [20] (Figure 1). The V protein inhibits STAT1 translocation into the nucleus and promotes STAT1 ubiquitination and degradation [32]. The V protein does so by binding to the highly conserved DNA damage-binding protein 1 (DDB1) [33] that is part of the ubiquitin ligase E3 complex. We previously showed that binding of the V protein to DDB1 requires its CTD [33] (Figure 1). The W proteins sequester STAT1 into the nucleus [34] thanks to a nuclear localization signal (NLS), consisting of the KKAR basic sequence present in their CTD (NiV W^{439–442}, HeV W^{437–440}) [35]. This NLS is recognized by importin α 3 [36] and accounts for steady-state location of the W protein in the nucleus [37]. In addition to STAT1, the NiV P, V and W proteins also bind to STAT4, with a common region encompassing residues 114–140 with been shown to be responsible for binding to both STAT proteins [38]. Moreover, the CTD of the NiV V protein binds to STAT5 [38].

Beyond its antagonist activity of IFN signaling [35], NiV W also prevents IFN-I expression [39]. The *Henipavirus* V and W proteins also inhibit the production of chemokines in vitro and modulate the inflammatory response in vivo [40]. NiV W inhibits TLR3 signaling and activation of the TNF α - and IL-1 β -induced NF- κ B canonical pathway [35,39,41]. Finally, *Henipavirus* W proteins bind to 14-3-3 proteins via their CTD, with this interaction having been shown to modulate various cellular processes including apoptosis [42] and inhibition of NF- κ B-induced proinflammatory response [43].

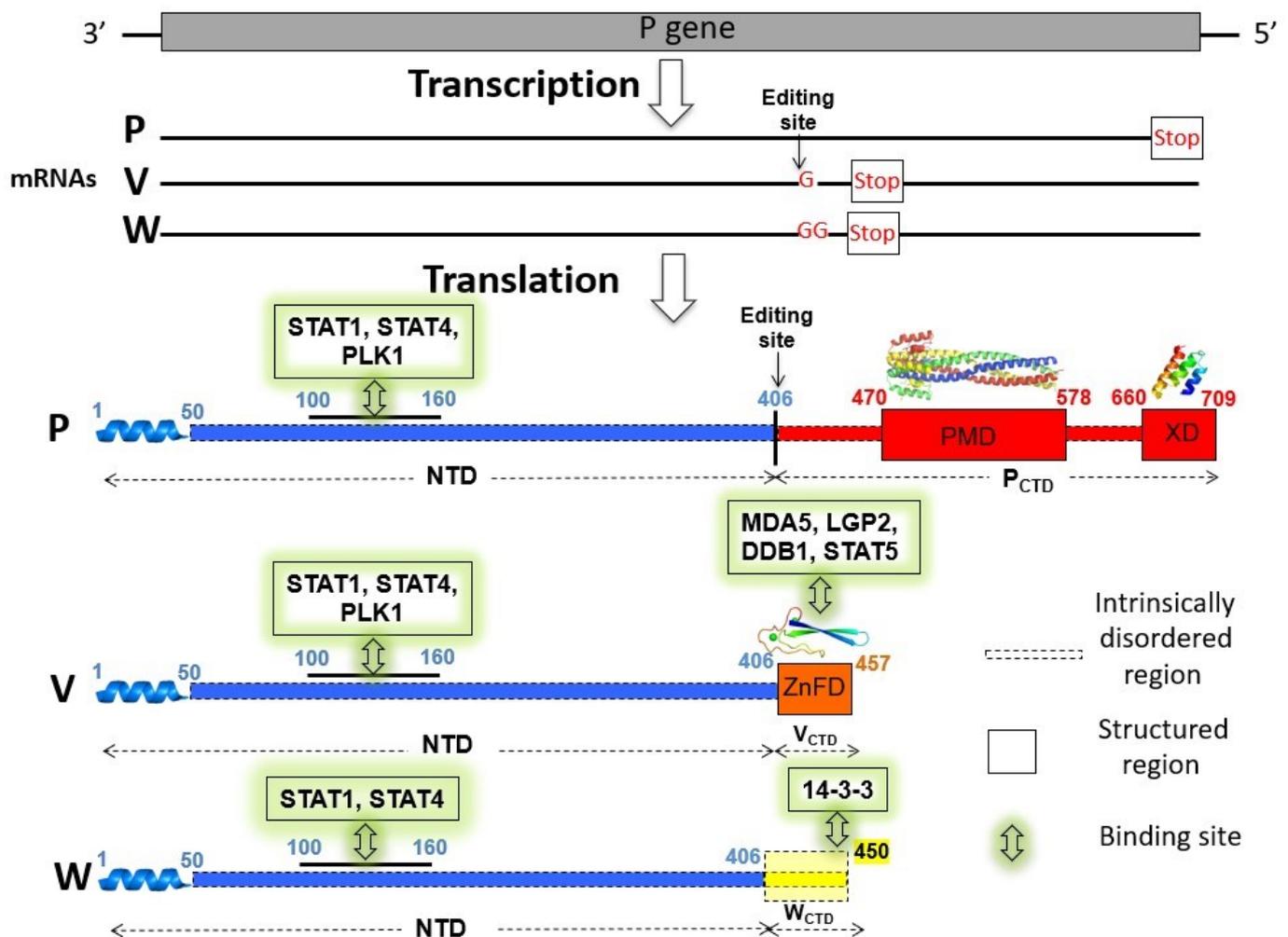


Figure 1. Coding capacity of the P gene and modular organization of the P, V and W proteins. The C protein is not shown for the sake of clarity. Shown is the organization of NiV proteins that is very close to that of their HeV counterpart. NTD: N-terminal region of the P, V and W proteins; P_{CTD}: C-terminal region of the P protein; PMD: P multimerization domain; XD: X domain of the P protein; ZnFD: zinc-finger domain; V_{CTD} and W_{CTD}: C-terminal domain of the V and W proteins. The α -helix at the N-terminus of the P, V and W proteins corresponds to the region adopting a stable α -helical conformation upon binding of P to N (i.e., N^o-P₁₋₅₀ complex) [11] or of V to host cellular transporters [44]. The structures of NiV PMD (PDB code 4N5B) [21], of a model of NiV XD [24] and of a model of the NiV ZnFD (with zinc ions shown as green spheres) [33] are displayed. The W_{CTD} is represented both as structured or disordered to reflect the lack of data on its actual conformation. Well-described interaction sites with human cell partners are shown.

The NTD of the NiV and HeV V proteins was shown to be disordered not only in isolation [13,14,20] but also within the V protein, while their CTD adopts a zinc-finger conformation [33].

Although the *Henipavirus* W proteins have been referred to as intrinsically disordered proteins (IDPs) [29], no study has been published so far supporting their disordered nature. Albeit the disordered nature of the NTD has already been experimentally shown [13,14,20], no data are available on the W protein. The lack of regular secondary structure in the conformation adopted by the W_{CTD} when bound to importin $\alpha 1$ or $\alpha 3$ [36] or to 14-3-3 isoform σ [42] suggests that it may be disordered in isolation. Despite these hints, no experimental data have been gathered so far on the W_{CTD} , nor even any bioinformatic analysis has been performed. In addition, although the NTD is disordered both in isolation and in the context of the V protein, long-range intramolecular contacts between the W_{CTD} and the NTD cannot be ruled out. Importantly, this possible crosstalk between the two moieties of the W protein might cause the NTD to adopt a specific conformation that might impart unique functions to W. For this reason, and in light of the key role that the W protein plays in counteracting the host antiviral response, investigating the conformational properties of this protein is relevant and is a prerequisite toward rational antiviral approaches. To this end, we herein have combined *in silico* analyses and biochemical and biophysical studies of the HeV and NiV W proteins. Our results reveal that their CTD are consistently predicted to be intrinsically disordered and we provide experimental evidence of the disordered nature of the W proteins. In addition, the W proteins were found to phase separate and to form amyloid-like fibrils.

2. Results

2.1. Bioinformatic Analysis of the HeV and NiV W Protein Sequences

The HeV and NiV W native proteins are 448 and 450 residues long, respectively. The two W sequences are close to each other (58% of identity, 71% of similarity) with their CTD displaying a high sequence identity (82%) and the region encompassing residues 270–390 exhibiting the highest divergence (Supplementary Figure S1). The isoelectric point (pI) of both proteins is acidic, while that of their CTDs is basic (Table 1). The hydropathy values of the W proteins are consistent with an overall hydrophilic nature (Table 1).

Table 1. Properties of the amino acid sequences of the W proteins from HeV and NiV.

Proteins	N	pI	f_+	f_-	FCR	NCPR	k	Hydropathy	Disorder Promoting	PDR
HeV W	448	4.93	0.123	0.170	0.292	−0.047	0.223	3.545	0.721	2
HeV W_{NTD}	404	4.64	0.111	0.186	0.297	−0.074	0.200	3.500	0.730	2
HeV W_{CTD}	44	11.89	0.227	0.023	0.250	0.205	0.209	3.952	0.636	2
NiV W	450	4.84	0.117	0.169	0.287	−0.051	0.194	3.674	0.696	2
NiV W_{NTD}	406	4.53	0.106	0.183	0.288	−0.076	0.167	3.677	0.697	2
NiV W_{CTD}	44	11.89	0.227	0.046	0.273	0.182	0.292	3.645	0.682	2

N: residue number; pI: isoelectric point; f_+ : fraction of positively charged residues; f_- : fraction of negatively charged residues; FCR: fraction of charged residues ($f_+ + f_-$); NCPR: net charge per residue, value of the difference between the fraction of positively charged and negatively positively residues, ($f_+ - f_-$); k: linear distribution of opposite charges; Hydropathy: average Kyte-Doolittle hydropathy value rescaled to lie between 0 (least hydrophobic) and 9 (most hydrophobic); Disorder promoting: fraction of residues falling in the disorder promoting category; PDR: phase diagram region. All parameters were obtained using CIDER (<http://pappulab.wustl.edu/CIDER/>) (accessed in 1 April 2021) [45].

The sequence composition of HeV and NiV W proteins was compared to that of proteins within the SWISS-PROT database (Figure 2A). Both domains have a biased sequence composition, being enriched in disorder-promoting residues (T,A,G,R,D,H,Q,K,S,E,P), and in particular in Asp (D), Ser (S) and Pro (P), and depleted in order-promoting residues (W, F, Y, I, M, L, V, N, C) [46]. In line with these observations, both W proteins are predicted to be intrinsically disordered by the mean hydrophobicity/mean net charge ratio [47], as judged from their location in the left-hand side of the RH-plot (Supplementary Figure S2A,B).

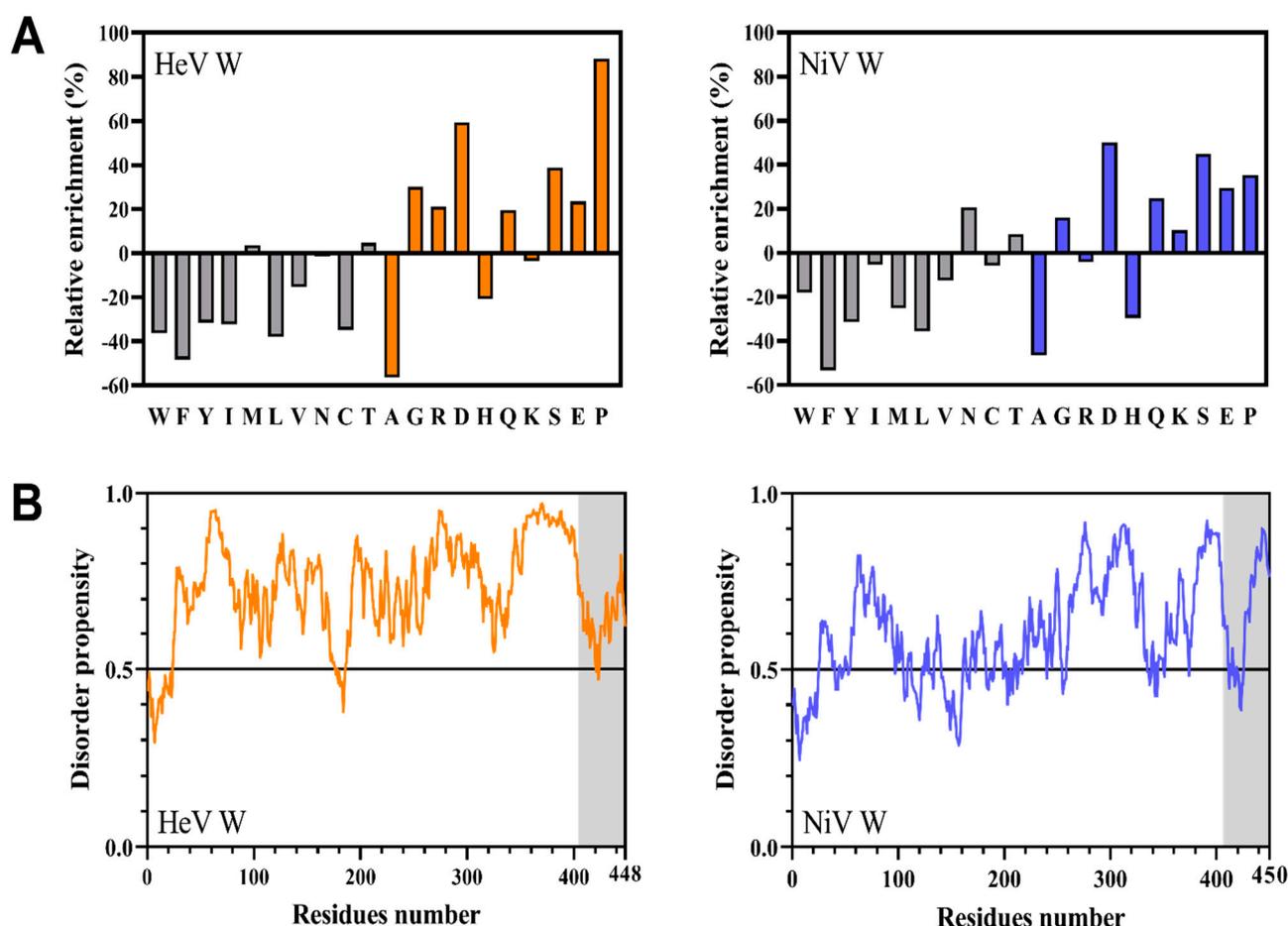


Figure 2. (A) Deviation in amino acid composition from the Swiss-PROT database of HeV and NiV W proteins. The relative enrichment in disorder promoting (orange and blue bars) and depletion in order-promoting (gray bars) residues is shown. Residues have been ordered on the x-axis according to the TOP-IDP flexibility index as described in [46]. (B) Disorder prediction of the W proteins as obtained by IUPred2A [48]. Residues with a disorder score above 0.5 are considered to be disordered. The region corresponding to the CTD is shaded in gray.

Since their respective NTD was already reported to be intrinsically disordered [13,14,20], we were mainly interested in analyzing the sequence properties and contribution of the CTD to the overall disorder content of the proteins. To this end, we analyzed the sequence of the two proteins using IUPred2A, a disorder predictor that provides a *per* residue disorder score (Figure 2B). Both CTD are predicted to be disordered (Figure 2B, gray shaded region), although a sharp drop in the disorder score can be observed over a short stretch of residues (Figure 2B). These short regions of predicted order might correspond to MoRFs (Molecular Recognition Features), also called MoREs (Molecular Recognition Elements), i.e., short regions within IDRs with a propensity to fold upon binding to a partner or ligand [49,50].

In agreement with the IUPred2A prediction, both CTDs are predicted to be disordered by the RH-plot (Supplementary Figure S2C,D) and have a biased sequence composition similar to that of the full-length W proteins (Supplementary Figure S3A).

Since the W proteins possess four (HeV) or six (NiV) cysteines, of which three are conserved (Supplementary Figure S1), we also analyzed the sequences using the redox-dependent prediction implemented in IUPred2A. The redox state modeling option of IUPred2A enables identifying redox-sensitive regions, i.e., regions susceptible to undergo redox-dependent disorder-to-order or order-to-disorder transitions [51]. No such regions were identified (data not shown) consistent with a predicted scenario where intramolecular disulfide bridges (if any) would not trigger a disorder-to-order transition.

Sequence polarity was shown to be a determinant of intrinsically disordered protein (IDP) compaction, with polar IDPs having been found to favor collapsed ensembles in water despite the absence of hydrophobic groups [52]. To assess sequence polarity, the net charge per residue (NCPR, defined as $f_+ - f_-$) [52,53], the total fraction of charged residues (FCR, defined as $f_+ + f_-$), and linear distribution of opposite charges (κ value) [54] were calculated for the full-length W proteins and for their respective NTDs and CTDs (Table 1). For both W proteins, as well as for their respective NTDs, the NCPR is negative, and the FCR values are close (~ 0.29). By contrast, the NCPR of the CTDs is positive and the FCR of NiV W_{CTD} is slightly higher (0.273) than that of HeV W_{CTD} (0.250). According to the predictive diagram of states developed by Pappu and colleagues [55], both the W proteins and their respective domains fall in phase diagram region (PDR) 2, with the CTDs being however very close to the boundary separating PDR1 and PDR2 (Figure 3A,B). PDR 2 corresponds to proteins that adopt conformations likely representing a continuum of possibilities between PDR 1, which accommodates weak polyampholytes or polyelectrolytes adopting a globular conformation, and PDR 3, which embeds strong polyampholytes adopting non-globular, swollen coil-like conformations. The κ values of the full-length proteins and of their domains are low indicating that opposite charges are well mixed and suggesting that these proteins and their domains adopt preferentially extended and swollen, coil-like conformation.

The amino acid sequences of the CTDs were also analyzed using Hydrophobic Cluster Analysis (HCA) [56] (see Appendix A). The HCA plots of the two domains show an enrichment in proline (red stars), glycine (diamonds) and basic residues (blue). Despite this bias toward disorder-promoting residues, the two CTDs are relatively enriched in hydrophobic clusters advocating for the occurrence of at least transiently populated regular secondary structure elements (Figure 3C,D). Notably, the HCA plots of both HeV and NiV W_{CTD} display a higher content in hydrophobic clusters with respect to the HCA plots of the V_{CTD} that are known to adopt a zinc-finger conformation (Supplementary Figure S3B). The latter however are enriched in cysteines that drive folding of these domains [33]. The HCA plot of HeV W_{CTD} features slightly larger hydrophobic clusters suggesting a higher extent of order compared to NiV W_{CTD} (Figure 3C,D). In line with this observation, the disorder promoting score, as provided by CIDER, is slightly larger for NiV W_{CTD} than for the corresponding domain of HeV (Table 1). In further support of a predicted higher order content in HeV W_{CTD} , PSIPRED analysis predicts an additional β -strand in this domain (Figure 3E,F).

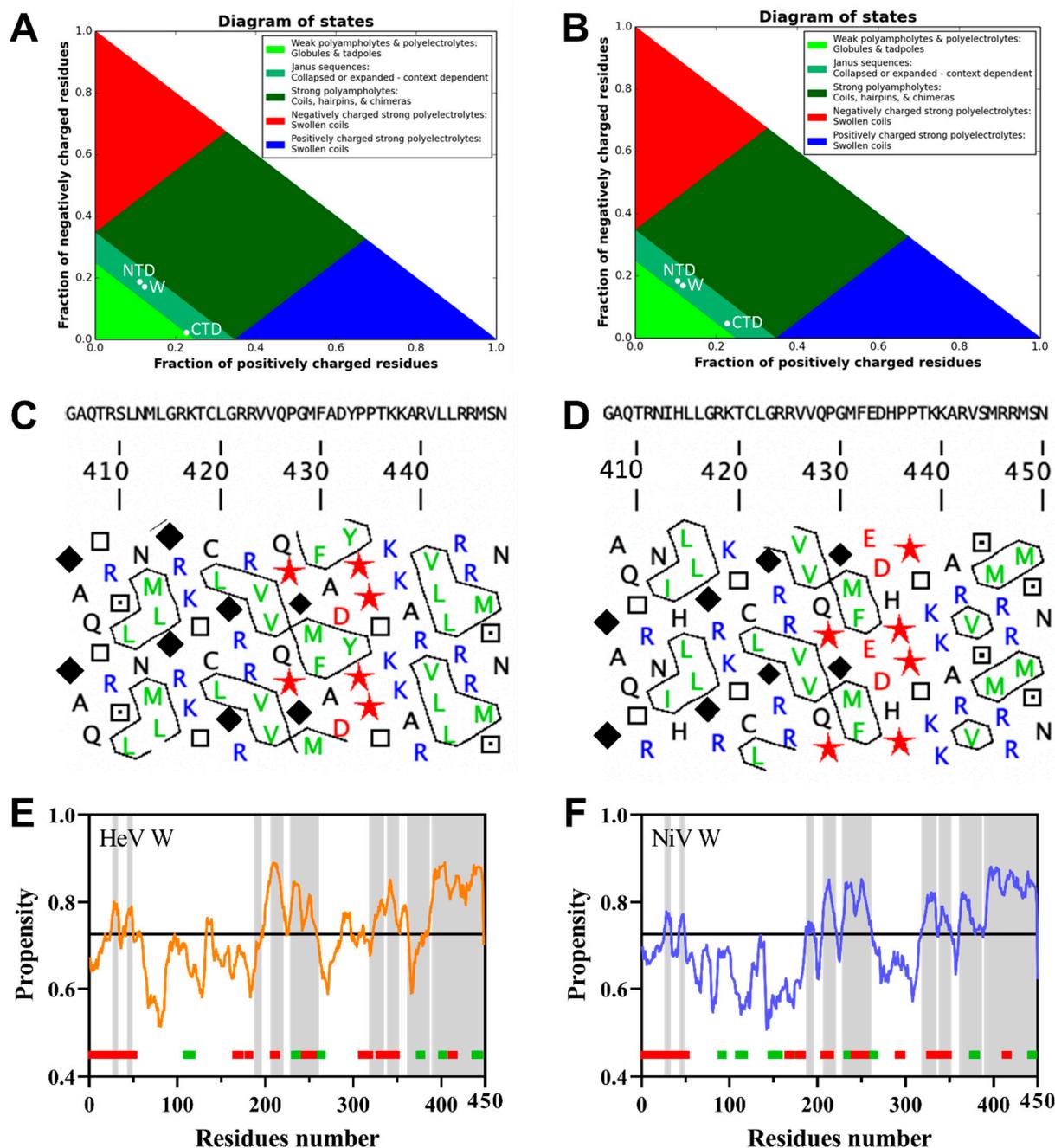


Figure 3. Phase diagram plots of HeV (A) and NiV W (B) proteins as provided by CIDER (<http://pappulab.wustl.edu/CIDER/>) (accessed on 1 April 2021) [45]. HCA plots of the CTD of HeV (C) and NiV (D) W featuring the amino acid sequence above the HCA plot. Red stars: proline residues, black diamonds: glycine residues, squares: threonine residues, dotted squares: serine residues. MoRF and secondary structure prediction within HeV (E) and NiV W (F) proteins as provided by MoRFChibi System [57] and PSIPRED [58] servers, respectively. Predicted MoRFs are shaded in gray. Predicted α -helices and β -strands are shown in red and green respectively.

The pattern of predicted secondary structure elements in HeV W_{NTD} is very similar to the one of the cognate protein, as is the pattern of predicted MoRFs (Figure 3E,F). Predicted MoRFs overlap with predicted secondary structure elements reflecting possible partially pre-configured MoRFs. The largest predicted secondary structure element is an α -helix encompassing the first 50 residues and corresponding to the region adopting a helix-kink-helix conformation in the crystal structure of the N° - P_{1-50} complex of NiV [11]. The other

predicted MoRFs are in good agreement with experimentally observed secondary structure elements as revealed in NiV NTD by NMR spectroscopy [14,20] as well as with already mapped binding sites, including the STAT1 (aa 110–140) [20], STAT2 (aa 110–140 and 230–237) [59] and STAT4 (aa 110–140) [38] binding sites. Interestingly, the CTD of both W proteins is entirely predicted to be a MoRF (Figure 3E,F), reflecting its inherent propensity to fold upon binding and consistent with its expectedly broad molecular partnership as judged from the wide panel of functions specific to the W protein.

Analysis of the amino acid sequence of the CTD of the two W proteins with the Phyre2 protein fold recognition server [60] returned a high confidence (99.6%) model based on the crystal structure of the HeV W_{CTD} in complex with importin $\alpha 1$ or $\alpha 3$ [36]. The model encompasses 59% of the CTD residues and consists of an irregular conformation (data not shown).

Altogether, results support the conclusion that W_{CTD} is an IDR with a high propensity to fold upon binding.

2.2. Expression and Purification of the W Proteins

Because the cognate V proteins devoid of any solubility tag were found to be expressed in inclusion bodies [33], we decided to subclone the W genes into the pETG-20A vector that allows the expression of a cleavable hexahistidine tagged TRX fusion protein. Although the fusion proteins were found to be soluble, attempts at purifying them by IMAC revealed that the majority of the protein did not bind to the resin (data not shown). Moreover, TEV protease failed to cleave off the TRX-His₆ tag (data not shown). We ascribed both events to poor accessibility of the His₆ and TEV cleavage site and resorted to clone the W gene into the pDEST17OI vector that drives the inducible expression of a hexahistidine tagged form of the protein with no solubility tag. As in the case of the V proteins [33], the W proteins expressed from this vector were found to be expressed in inclusion bodies only, irrespective of the IPTG concentration, temperature of induction, and culture medium used (data not shown).

The W proteins were thus purified from inclusion bodies under denaturing conditions in two steps: IMAC and IEC. As shown in Figure 4A,B the final purified W proteins migrate with an apparent molecular mass of ~65 kDa, a value 1.2 times higher than the one expected from the amino acid sequence (~53 kDa). Aberrant electrophoretic migration, which had been already reported for *Henipavirus* NTD [13] and V [33], is a hallmark of IDPs and is mainly due to their typical compositional bias, e.g., high net charge and low hydrophobicity that causes them to bind less SDS than globular proteins [61,62]. In the case of the W proteins, this anomalous electrophoretic behavior could also be at least partly due to their high proline content (8.4% in HeV W and 6.1% in NiV W to be compared to 4.7% for proteins in the SWISS-PROT database) [62,63].

For both proteins, mass spectrometry (MS) analysis of peptides resulting from tryptic digestion confirmed the identity of the purified proteins (Supplementary Figure S4). For both proteins, the profile resulting from tryptic digestion of the protein (Supplementary Figure S4) is consistent with the initial methionine having been cleaved off, an event predicted to occur with an 84% frequency for proteins in which the second residue is a serine [64]. MALDI-TOF MS analysis of the HeV W protein revealed the presence of a species with an average molecular mass of 52,577.55 Da, a value very close (i.e., 3.66 Da of difference) to the expected value calculated from the amino acid sequence without the initial methionine (Figure 4C). In the case of NiV, the experimentally observed average molecular mass (52,656.8 Da) is a bit more divergent (87.54 Da) but still close to the value expected for a form in which the initial methionine has been cleaved off (Figure 4D). Furthermore, for both proteins an additional, much less intense peak with a mass close to that expected for a dimeric form was observed (Figure 4C,D). In further support of the ability of the recombinant proteins to form disulfide-bridged oligomeric species, SDS-PAGE analysis in non-reducing conditions revealed the presence of higher mass species that are no longer detectable upon addition of DTT (Figure 4A,B).

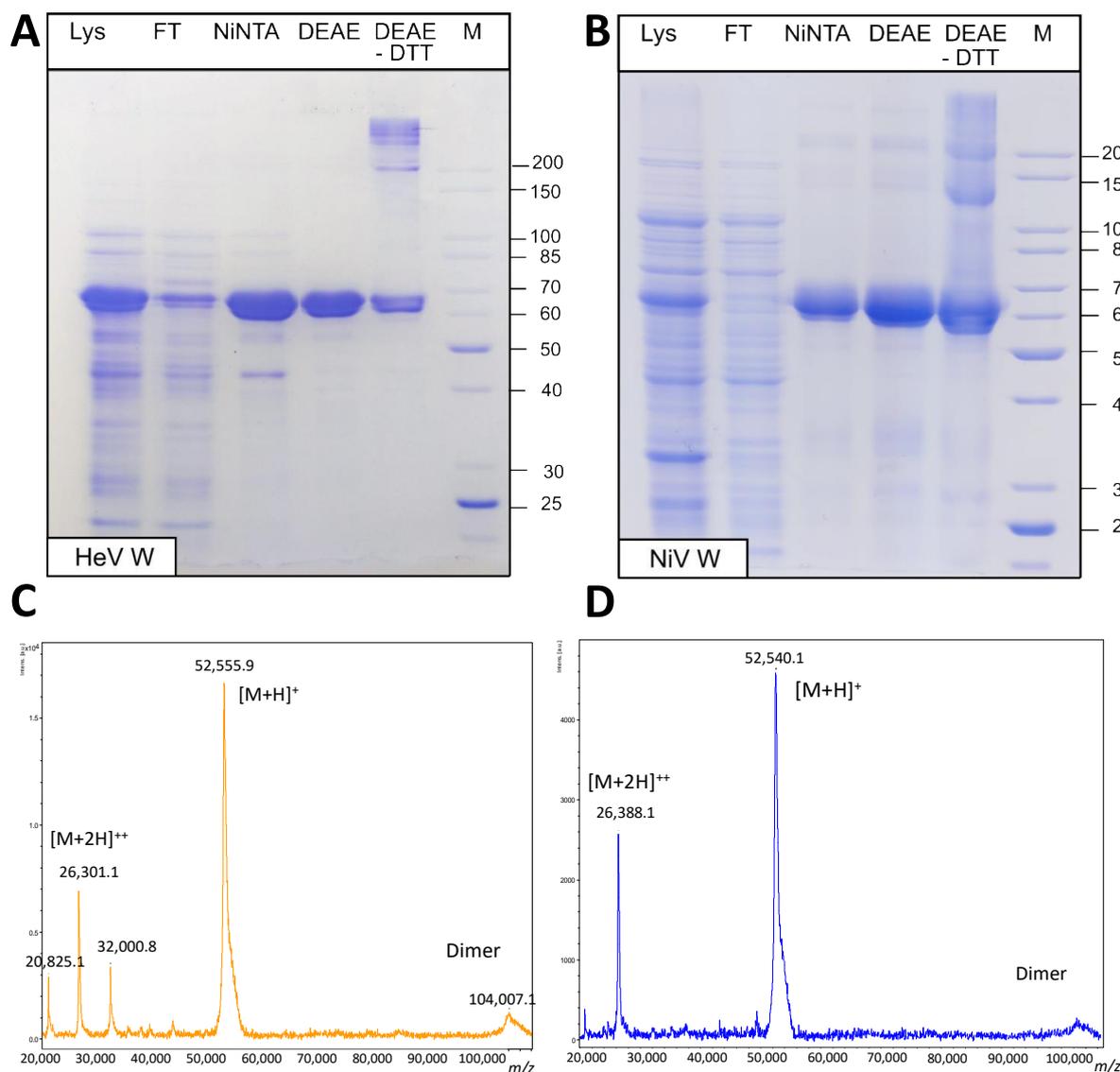


Figure 4. (A,B) Coomassie blue staining of a 15% SDS-PAGE analysis of the purification of the HeV (A) and NiV (B) W proteins. Lys: clarified lysate (soluble *plus* insoluble fraction); FT: flow-through of IMAC; NiNTA: eluent from IMAC; DEAE: eluent from IEC; DEAE–DTT: eluent from IEC under non-reducing conditions; M: molecular mass markers (in kDa). (C,D) MALDI-TOF-TOF mass analysis of HeV (C) and NiV (D) W proteins. The HeV W protein generated an average mass of 52,577.55 Da for a theoretical mass of 52,573.89 Da (without the initial methionine) (C). The NiV W protein generated an average mass of 52,656.8 Da for a theoretical mass of 52,744.15 Da (without the initial methionine) (D).

2.3. Protease Sensitivity of the W Proteins

High sensitiveness to proteolysis is a hallmark of structural disorder (see [62,65] and references therein cited). The use of thermolysin, which has a broad substrate specificity, allows the identification of cleavage sites solely on the basis of the flexibility of the protein substrate. To assess the extent of protease sensitivity, the W proteins were submitted to digestion by thermolysin. As shown in Figure 5, both proteins start to be degraded after 5 min of incubation and are extensively degraded after 45 min, a behavior that is consistent with the lack of a packed core and with an overall solvent accessibility. Conversely, BSA was shown to be resistant to proteolysis even after an incubation of 45 min (Figure 5). Interestingly, the two proteins do not exhibit the same extent of susceptibility to proteolysis. Indeed, while a 45-min digestion of the NiV W protein results in an enrichment in small

(i.e., <35 kDa) molecular mass fragments, in the case of HeV W after 45 min of incubation protein fragments of higher molecular mass (in the 60–25 kDa range), are still discernible (Figure 5).

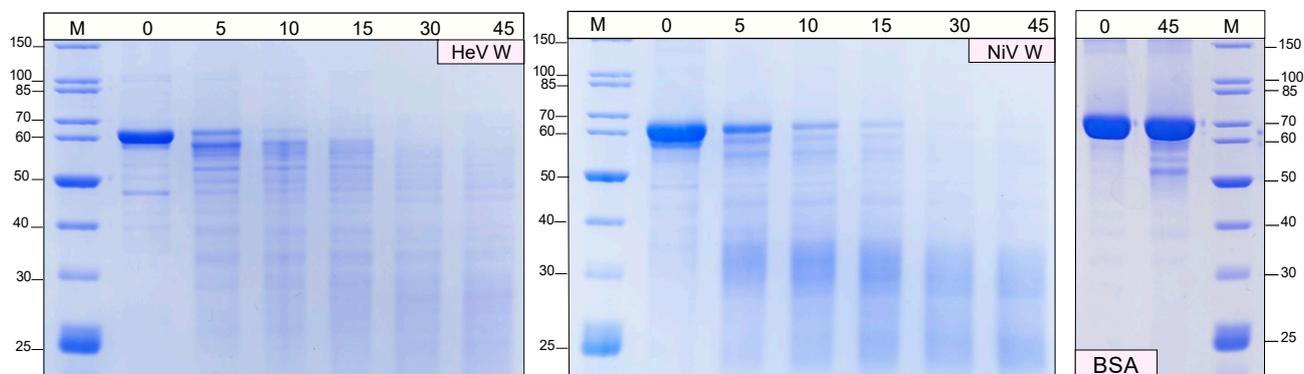


Figure 5. SDS-PAGE analysis of a limited thermolysin digestion of HeV and NiV W proteins and of BSA (control) at various time intervals (minutes). M: molecular mass makers (in kDa).

2.4. Differential Scanning Fluorimetry of the W Proteins

We further analyzed the W proteins by differential scanning fluorimetry (DSF). DSF is a broadly used technique to determine the conformational stability of proteins; it relies on the use of a fluorescent dye whose fluorescence is enhanced when protein hydrophobic cavities become accessible as a result of temperature-induced protein unfolding. IDPs, which do not possess hydrophobic cavities being devoid of a stable 3D structure, are characterized by a fluorescence profile that lacks the typical transition peak observed for structured proteins and is thus rather flat and temperature-independent (see [62,65] and references therein cited). The experimentally observed profile obtained for both the W proteins does not feature any transition peak, consistent with the lack of a stable 3D structure (Figure 6). These results provide additional support for an overall disordered nature of the W proteins, in line with the *in silico* analyses and with the above-described experimental lines of evidence.

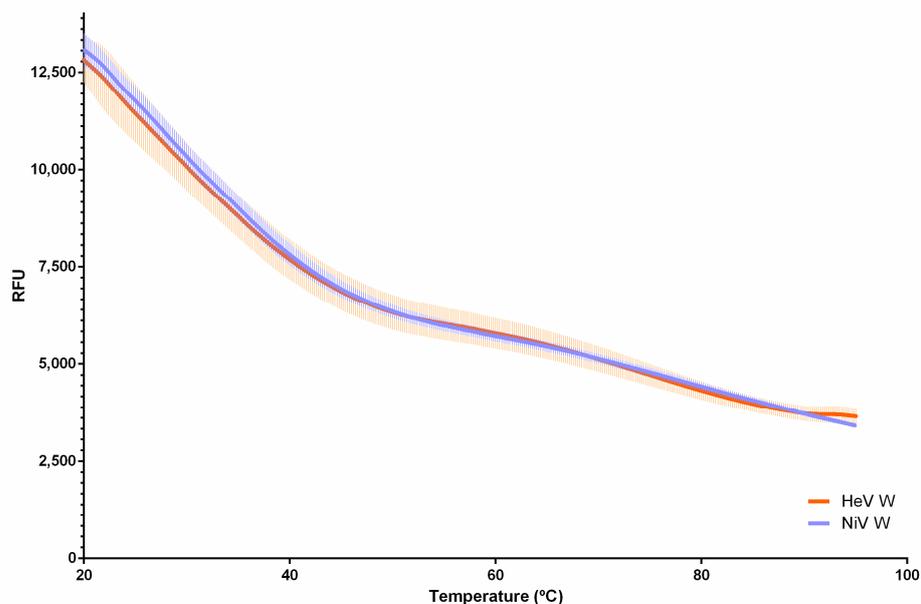


Figure 6. Differential scanning fluorimetry (DSF) of HeV and NiV W proteins in the presence of Sypro Orange in the 20–95 °C temperature range. The results are the mean of 9 replicates. Error bars correspond to the standard deviations.

2.5. Hydrodynamical Properties of the W Proteins from Size Exclusion Chromatography (SEC)

To infer the Stokes radii (R_s) of the HeV and NiV W proteins we used analytical SEC. The elution profile of both proteins in HBS at pH 7 features a major peak at ~16 mL preceded by a shoulder at ~14 mL. However, for both proteins the highest peak is at ~9 mL, indicating that the protein mostly exists in an aggregated form in solution (Figure 7A,C). SDS-PAGE analysis under reducing and non-reducing conditions showed that the species eluted at ~14 mL corresponds to a dimeric form of the protein stabilized by disulfide bridges (data not shown). These results are in line with the MS data that pointed out the presence of a less abundant dimeric species under non-reducing conditions. When we attempted at eliminating the dimeric species by adding DTT, we noticed that both proteins precipitated unless urea, at a concentration of at least 1 M, was added. SEC under reducing conditions was therefore performed by supplementing the buffer with urea.

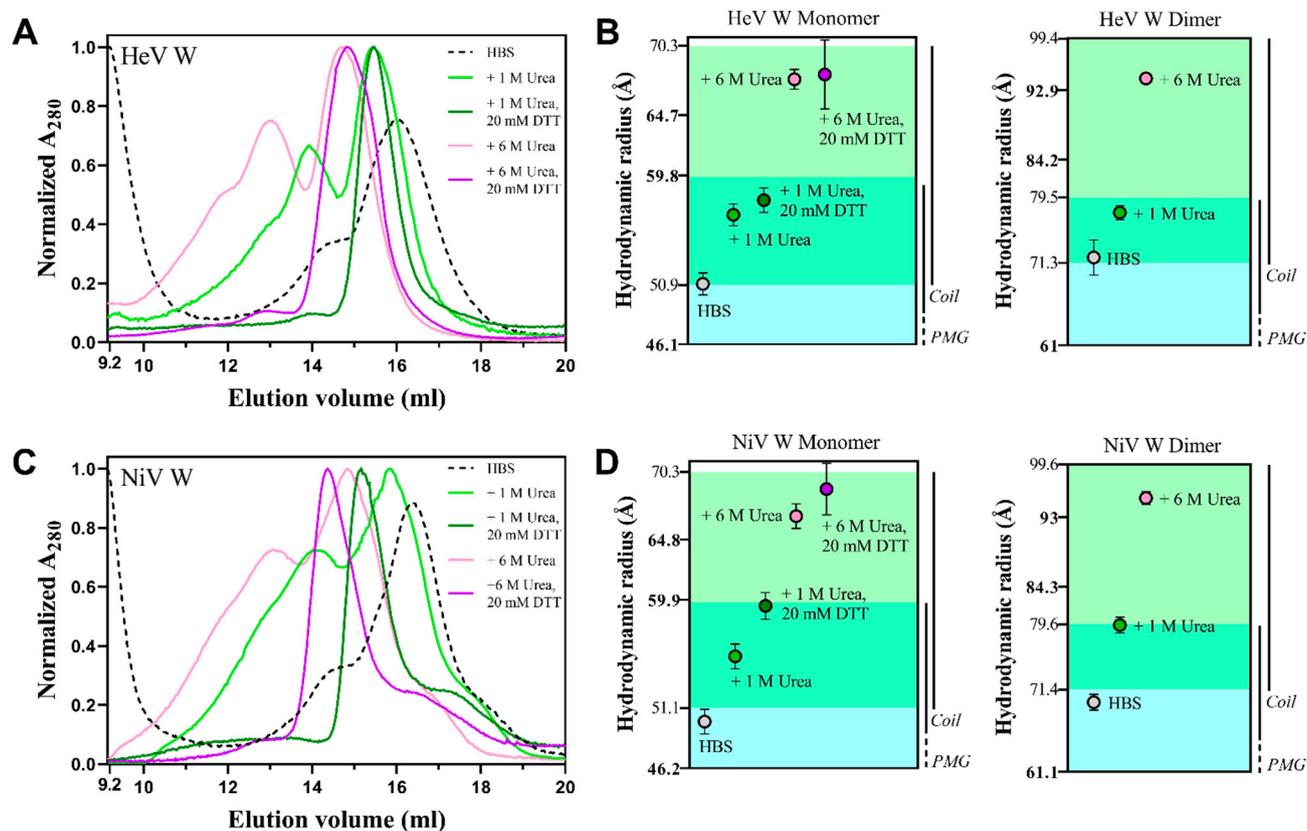


Figure 7. (A,C) Normalized SEC elution profiles of HeV (A) and NiV (C) W, as obtained using various conditions. Shown are the results of one out of three independent experiments. (B,D) Stokes radii (R_s) of the monomeric and dimeric form of HeV (B) and NiV (D) W in the different conditions. The regions corresponding to PMG (light blue) and Coil (green) forms are highlighted. Values shown in bold in panels B and D correspond to the average values expected for PMG (smallest values), Coils (intermediate values) and Denatured (largest values) conformations. The errors bars correspond to the experimentally observed s.d. in three independent experiments.

We first carried out SEC in the presence of 1 M urea to investigate possible effects on the conformational properties of the proteins. The addition of 1 M urea triggers a slight (~0.5 mL), though significant, shift in the elution volume of the peak corresponding to the monomeric form of the proteins (Figure 7A,C), reflecting a ~5.5 Å increase in the derived R_s (Figure 7B,D). Beyond this effect, the addition of 1 M urea also impacts the overall profile. For both proteins, upon addition of 1 M urea, a pronounced peak at ~14 mL is observed reflecting an increased amount of the dimeric form (Figure 7A,C). The latter is paralleled by the disappearance of the major peak at ~9 mL and by the concomitant

accumulation of higher mass species eluted between 10 and 14 mL (Figure 7A,C). As expected, addition of 6 M urea results in an even more pronounced change in the elution volume of the monomeric species that is further reduced by ~ 0.9 mL (Figure 7A,C). This shift in the elution volume entails a change in the R_S that is increased, with respect to the form obtained in native conditions, by ~ 16.6 Å (Figure 7B,D). For both proteins, a very similar trend for the impact of urea, i.e., a progressive reduction in the elution volume and an increase in the R_S , was also observed for the peak of the dimeric form (Figure 7A,B). The sensitivity of the W proteins to urea likely reflects the presence of residual secondary and/or tertiary structure typical of the premolten globule (PMG) state [66].

As expected, for both proteins, the addition of 20 mM DTT triggers the disappearance not only of the peaks corresponding to the dimeric species but also of those corresponding to the higher mass species (Figure 7A,C). Notably, for both proteins, the addition of DTT induces a shift in the elution volume of the peak of the monomeric species (Figure 7A,C), therefore leading to an increase in the corresponding R_S (Figure 7B,D). While in the case of HeV the increase is barely significant, it is more pronounced in the case of NiV (Figure 7B,D). This behavior indicates that the cysteine redox state impacts the compaction of the protein, consistent with a scenario where one (or more) intramolecular disulfide bridges would contribute to compact the chain, and disulfide bridge reduction would lead to expansion of the polypeptide chain. The NiV W protein has 6 cysteines, while the HeV W has only 4. As such, the NiV W has a higher disulfide bridge-mediated compaction potential compared to HeV W, which may explain the more pronounced effect of DTT on NiV W compared to HeV W.

From the elution volume of the major species (16.0 ± 0.08 mL for HeV W and 16.1 ± 0.07 mL for NiV W), as obtained using HBS as elution buffer, the corresponding R_S was estimated to be 51.0 ± 0.9 Å for HeV W and 50.1 ± 0.8 Å for NiV W (Table 2). By comparing the mean measured Stokes radius (R_S^{obs}) for HeV and NiV W with the theoretical Stokes radii expected for various conformational states (R_S^{NF} : natively folded protein; R_S^{PMG} : expected for a PMG; R_S^{U} : fully unfolded form; R_S^{IDP} : expected for an IDP), the two proteins were found to have R_S values ~ 1.7 times larger than the value expected for a natively folded protein, and close to the one expected for a PMG (Table 2). Similarly, the R_S of the dimeric form of the W proteins was estimated to be ~ 72 Å, a value suggesting a disulfide-bridged dimer in a PMG conformation (Figure 7B,D).

These results, while providing additional experimental evidence for the prevalently disordered nature of both W proteins, indicate the presence of some transiently populated secondary and/or tertiary structure typical of the PMG state, also supported by the expansion effect observed in the presence of urea. The compaction index of the W proteins indicates that NiV W is slightly more compact than HeV W (Table 2), in agreement with bioinformatic analyses that overall predict NiV W to be slightly less disordered than HeV W (Figure 2B and Table 1).

2.6. Far-UV Circular Dichroism (CD) Studies of the W Proteins

To evaluate the secondary structure content of the HeV and NiV W proteins, circular dichroism (CD) measurements were performed in the far ultraviolet (UV) region. Under native conditions and neutral pH, both spectra display a large negative peak centered at 200 nm, low intensity in the 220–230 nm region, and low ellipticity at 190 nm (Figure 8A). The spectra are typical of disordered proteins lacking any stable organized secondary structure. For both proteins, spectral deconvolution revealed a high content (about 60%) of unordered structure along with a $\sim 20\%$ content in β -strands (Figure 8B).

Table 2. Stokes radii (R_S^{obs} , Å) of the major SEC peak and expected values for the various conformational states.

Proteins	Mass	R_S^{obs}	R_S^{NF}	R_S^{PMG}	R_S^{U}	R_S^{IDP}	$R_S^{\text{obs}}/R_S^{\text{NF}}$	$R_S^{\text{obs}}/R_S^{\text{PMG}}$	$R_S^{\text{obs}}/R_S^{\text{U}}$	$R_S^{\text{obs}}/R_S^{\text{IDP}}$	CI
HeV W	52,706	51.0 ± 0.9	30.3	46.1	64.7	57.4	1.68	1.11	0.79	0.89	0.40 ± 0.03
NiV W	52,875	50.1 ± 0.8	30.3	46.2	64.8	57.5	1.65	1.08	0.77	0.87	0.43 ± 0.03

R_S^{obs} : experimentally observed Stokes radius (mean value and s.d. from three independent experiments); R_S^{NF} : R_S expected for a natively folded (NF) form; R_S^{PMG} : R_S expected for a premolten globule (PMG); R_S^{U} : R_S expected for a fully unfolded form; R_S^{IDP} : R_S expected for an IDP based on the simple power law model; Mass: molecular mass (Daltons) calculated from the amino acid sequence of the recombinant protein. CI: compaction index.

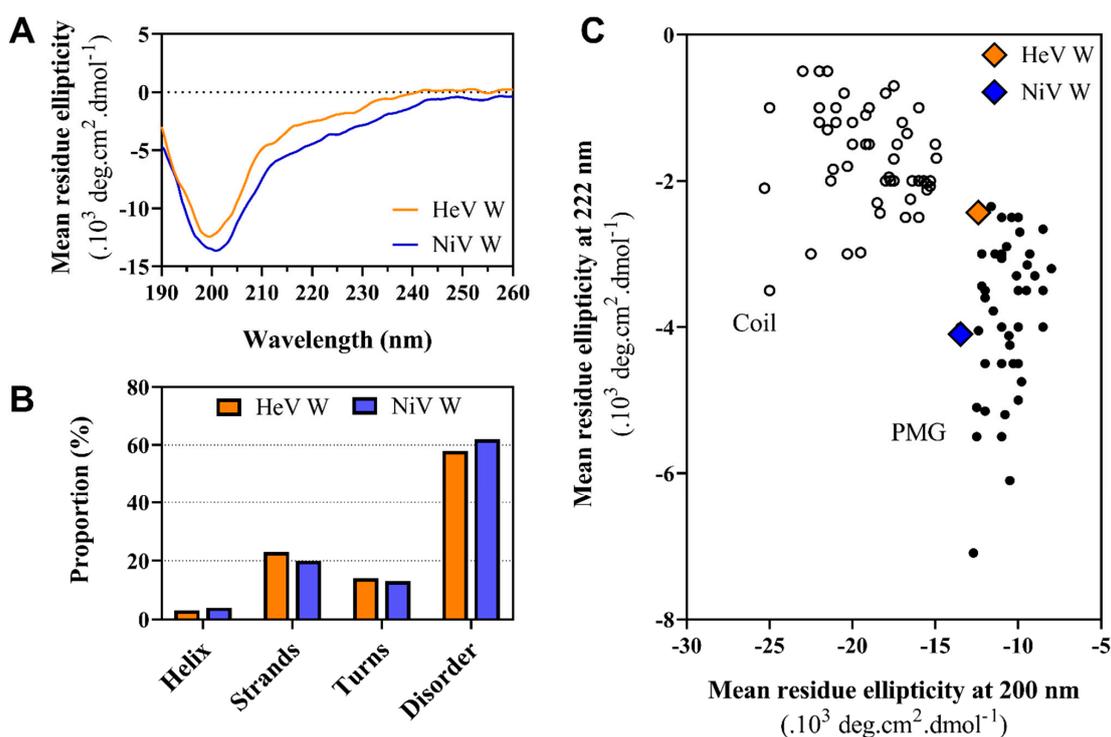


Figure 8. (A) Far-UV CD spectra of HeV and NiV W at 1 μM in 10 mM sodium phosphate pH 7. All spectra were recorded at 20 $^\circ\text{C}$. Data are representative of one out of three independent measurements. (B) Secondary structure content of the two proteins, as derived using Dichroweb (CDSSTR algorithm, set 7). (C) Plot of the molar residue ellipticity (MRE) at 222 nm and at 200 nm of a set of well-characterized unfolded, random coil-like (Coil, empty circles) or PMG-like (PMG, full circles) proteins (from [66]). The position in the plot of HeV W and NiV W is highlighted.

Far-UV CD spectroscopy enables discriminating between IDPs adopting a random coil-like (RC) state and a PMG-like state, based on the ratio of the ellipticity values at 200 and 222 nm [66]. According to their ellipticity values at 200 and 222 nm, both W proteins fall in the PMG-like region of the plot (Figure 8C). These results confirm the hints obtained from SEC experiments and support the classification of the W proteins within the PMG-like subfamily of IDPs.

We next sought at assessing whether disulfide bridges could play a role in stabilizing transiently populated regular secondary structure elements within the W proteins. Since both proteins precipitate in the presence of DTT unless 1 M urea was added, we recorded the CD spectra of the two proteins in the presence of 1 M urea supplemented with 10 mM DTT, and in the presence of solely 1 M urea as a control (Supplementary Figure S5A,B). While the spectrum of NiV W in the presence of 1 M urea is perfectly superimposable onto that obtained in the absence of urea (Supplementary Figure S5B), the spectrum of HeV W slightly deviates from that recorded in sodium phosphate buffer without urea (Supplementary Figure S5A), indicating that at this concentration urea has a poor (HeV) or no (NiV) impact on the secondary structure content. In light of this, the expansion effect triggered by 1 M urea as observed in SEC studies, can be accounted for by uniquely a destabilization of the transiently populated tertiary structure.

Although the spectra of both proteins in the presence of urea and DTT are quite noisy, a significant deviation from the respective control spectra is discernible (Supplementary Figure S5A,B). In particular, a decrease in the ellipticity at 220 and 215 nm is observed, reflecting respectively a reduction in the α and β content (Supplementary Figure S5A,B). These results advocate for a role of disulfide bridges in the stabilization of secondary structure elements and mirror those obtained by analytical SEC pointing to a role of oxidized cysteines in driving protein compaction (Figure 7).

One of the functional advantages of IDPs arises from their ability to contextually change their conformation due to binding or to environmental changes. To assess the folding potential of the W proteins, their far-UV CD spectra were recorded in the presence of TFE. This organic solvent is a secondary structure stabilizer that is used to mimic the hydrophobic environment experienced by proteins during interactions [67,68]. It is thus widely used as a probe of hidden structural propensities to unveil protein regions with a propensity to undergo induced folding [62,65].

For both proteins, addition of TFE induces a gain of α -helicity indicated by the characteristic maximum at 190 nm and double minima at 208 and 222 nm (Figure 9A,D). The estimated α -helical content gradually increases upon increasing the TFE concentration from 0 to 50% (Figure 9B,E). For the two proteins, the spectra display an isodichroic point at 201 nm indicative of a two-state transition (Figure 9A,D). We therefore plotted the percentage of α -helix as a function of the TFE concentration and fitted the data to a sigmoidal curve, corresponding to a two-state transition (Figure 9C,F). The fitting yielded a midpoint of transition at a TFE concentration of $\sim 20\%$ for both proteins. From the fitting procedure, the m value could also be obtained. The latter is a measure of the osmolyte efficacy in folding or unfolding a protein. The m value is typically positive for protecting osmolytes, such as TFE that drive the equilibrium toward the folded state. In agreement, m values of $0.107 \pm 0.013 \text{ kcal mol}^{-1} \text{ M}^{-1}$ and of $0.085 \pm 0.012 \text{ kcal mol}^{-1} \text{ M}^{-1}$ were obtained for HeV W and NiV W, respectively. The two proteins thus have midpoints and m values very close to each other, indicating that they follow a very similar folding pathway. Despite this similarity however, the α -helical propensity of the NiV W protein is more pronounced, as judged from the higher α -helical content of the latter with respect to the cognate protein at each point of the TFE titration.

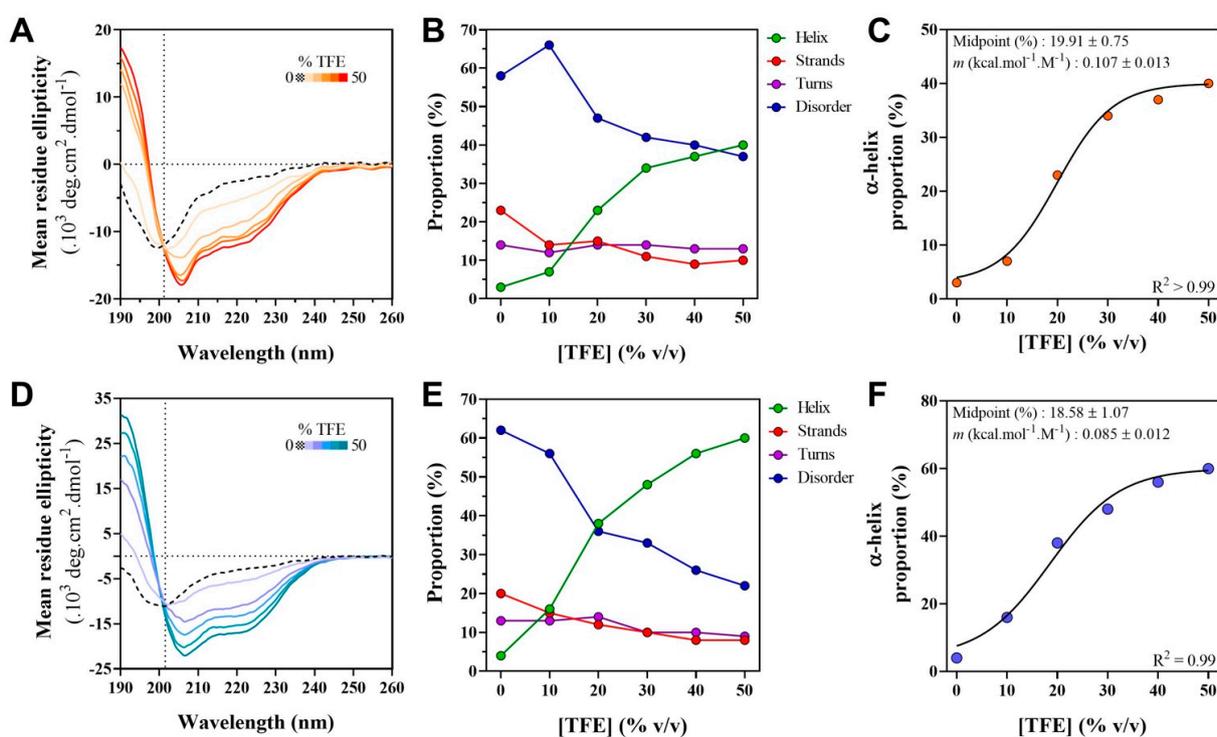


Figure 9. (A,D) Far-UV CD spectra of HeV (A) and NiV (D) W proteins in the presence of increasing concentrations of TFE at 20 °C. Proteins were at 1 μM in 10 mM sodium phosphate pH 7. Data are representative of one out of two independent experiments. The vertical line is centered on the isodichroic point. (B,E) Secondary structure content of HeV (B) and NiV (E) W proteins, as derived using CDSSTR, as a function of TFE concentration. (C,F) Transition diagrams of HeV (C) and NiV (F) W proteins. The midpoint of transition and m values, as derived from the fitting, along with the quality of the fitting (R^2) are shown.

2.7. Small-Angle X-ray Scattering (SAXS) Studies of the W Proteins

To achieve a more quantitative description of the conformational properties of the W proteins in solution, we performed SAXS studies coupled to SEC (SEC-SAXS). To this end, we first recorded SEC-SAXS data under native and non-reducing conditions (i.e., in the absence of urea and DTT) with the hope that the SEC step would enable analysis of the monomeric species on its own. Unfortunately, however, obtained data turned out to suffer from polydispersity and to be unexploitable (data not shown). To eliminate disulfide-bridged oligomeric species, we resorted recording data under reducing conditions. Because DTT causes the proteins to precipitate unless 1 M urea is added, and knowing that the addition of urea dissolves the higher molecular mass aggregates detected by SEC but also visible in solution, we performed SEC-SAXS experiments in the presence of both 1 M urea and 5 mM DTT while being aware that these conditions trigger a slight extension of the chain, as judged from analytical SEC experiments. The rationale for adding 5 mM DTT rather than 20 mM as in analytical SEC experiments, resides in the fact that this DTT concentration was shown to be sufficient to reduce the protein (data not shown) and that we wanted to minimize the concentration of buffer additives so as to achieve maximal contrast between the protein and the solvent.

Despite the presence of minor peaks corresponding to aggregates, for both W proteins the obtained SEC profiles featured a sharp and symmetric peak corresponding to the monomeric form (Supplementary Figure S6A,B). In both cases, Guinier analysis of the deconvoluted scattering curves (Figure 10A,B) revealed linearity in the Guinier region (i.e., for $qR_g < 1.3$), with no indication of protein aggregation (Figure 10A,B insets). The R_g extracted from Guinier analysis is 72.01 ± 0.89 Å for HeV and 70.96 ± 1.27 Å for NiV W. The R_g values derived from Guinier analysis are close to those derived from the $P(r)$ (73.62 ± 0.40 Å for HeV and 73.62 ± 0.62 for NiV W). The theoretical R_g values expected for IDPs of the same length, calculated using Flory's power law (see Equation (9) in Materials and Methods), are 63.5 Å for HeV and 63.6 Å for NiV. The experimentally determined R_g values are therefore larger than those expected for IDPs possessing the same number of residues and close to value expected for chemically denatured proteins (~ 78 Å). This finding is in line with what could be expected given the use of a buffer supplemented with 1 M urea and 5 mM DTT, which jointly were found to trigger a $\sim 13\%$ expansion of the protein chain (see Figure 7). The R_g -based compaction index (CI) (see Equation (10) in Materials and Methods) of NiV W is 0.111, a value slightly, though significantly, larger than that of the cognate HeV W protein (0.089). This finding mirrors differences observed in R_S -based CI as estimated from SEC (see Table 2).

The pairwise distance distribution, $P(r)$, yielded a maximal dimension, D_{max} , of ~ 240 Å for HeV and ~ 245 Å for NiV W (Supplementary Figure S6C) with a long tail in the $P(r)$ function, indicating that the proteins assume an overall non-compact conformation [69]. The ratio between the experimentally observed R_g and R_S is ~ 1.41 for both HeV and NiV W proteins. This value is closer to the one expected for RC-like IDPs (1.5) than to that expected for PMG-like IDPs (1.0) [70], probably due to the expansion effect of urea and DTT. The overall SAXS parameters for the W proteins are listed in Table 3.

The flexible nature of the W proteins was also qualitatively assessed using the Kratky-Debye and the normalized Kratky plots (Figure 10C,D). The presence of a plateau in the Kratky-Debye plot (Figure 10C) and the shape of the normalized Kratky plot (with no clear maximum) (Figure 10D) indicate that the W proteins are intrinsically disordered.

Table 3. SAXS-derived parameters.

Proteins	I(0) cm ⁻¹	R _g (Å) Guinier	R _g P(r) (Å)	R _g (ave) EOM (Å)	D _{max} (Å)	Rflex (%) (pool)	Rflex (%) (ens)	χ ²	p-Value CorMap	R _g ^{IDP} (Å)	CI
HeV W	0.030 ± 2.4 × 10 ⁻⁴	72.01 ± 0.89	73.62 ± 0.40	71.7	240	84.7	82.5	0.52	0.530	63.5	0.089 ± 0.016
NiV W	0.024 ± 3.1 × 10 ⁻⁴	70.96 ± 1.27	73.62 ± 0.62	70.7	245	84.5	82.0	0.44	0.014	63.6	0.111 ± 0.023

I(0): Intensity at zero angle as determined from Guinier approximation; R_g Guinier: R_g values as obtained from Guinier approximation; R_g P(r): R_g as obtained from pairwise distance distribution; R_g (ave) EOM: average value in the final EOM ensemble; D_{max}: maximal intramolecular distance from P(r); Rflex (pool): flexibility index in the initial pool; Rflex (ens): flexibility index in the final ensemble; χ²: quality of the fit between experimental and back-calculated data from the EOM ensemble; p-value: quality of the fit between experimental and back-calculated data from the EOM ensemble, as provided by CorMap. R_g^{IDP}: R_g expected for an IDP based on the simple power law model; CI: compaction index.

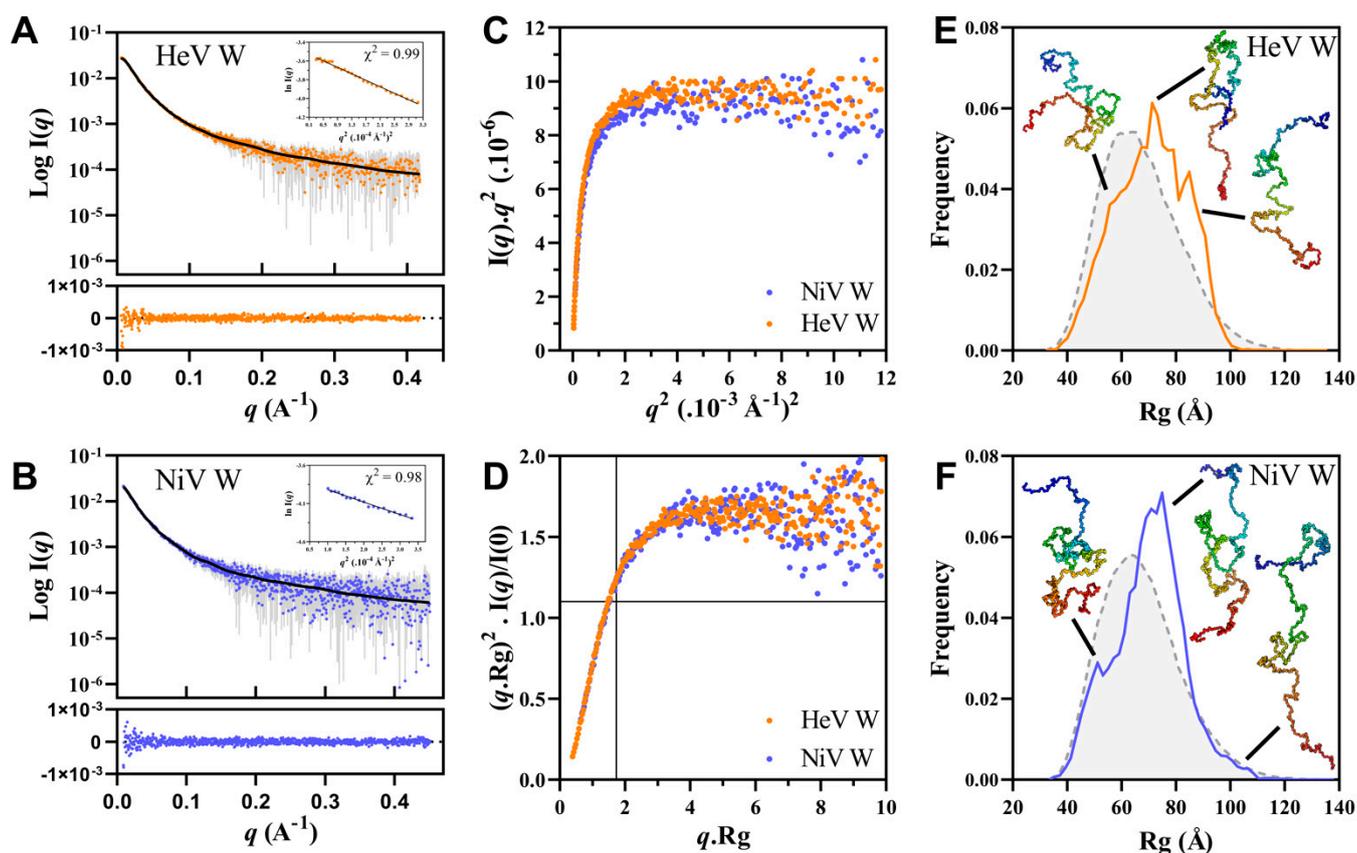


Figure 10. (A,B) Scattering curves of HeV (A) and NiV (B) W proteins and ensemble EOM 2.0 fits (solid line). Below are shown the plots of residuals. Insets: Guinier plots with the quality of the fit (χ^2). Scaled Kratky-Debye (C) and normalized Kratky (D) plots of the scattering data of HeV (orange) and NiV (blue) W. (E,F) R_g distributions of the initial pool generated by EOM 2.0 (gray) and of the corresponding selected ensemble for the HeV (E) and NiV (F) W proteins. A cartoon representation of three conformers present in the ensembles is also displayed. The structures were drawn using Pymol 2.0.1 [71].

To achieve further insights into the conformational behavior of the W proteins, we used SAXS data in search of an ensemble that quantitatively describes the conformational behavior of the proteins in solution. To this end, we used the program EOM 2.0 (see Material & Method section). For both proteins, the scattering curve back-calculated from the selected ensemble (Figure 10A,B solid lines) fits well the experimental SAXS data as judged from both χ^2 and CorMap [72] ($\chi^2 = 0.52$, $p = 0.53$ for HeV W and $\chi^2 = 0.44$, $p = 0.014$ for NiV W). For both proteins, the resulting final R_g (Figure 10E,F) and D_{max} (Supplementary Figure S6D) distributions, are unimodal and close to those of the initial pool, indicating that the W proteins exist in solution as randomly distributed ensembles of non-compact and highly flexible conformations. Please 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 through the R_{flex} that was estimated to be $\sim 82.5\%$ for HeV W and $\sim 82\%$ for NiV W (to be compared to $\sim 84.7\%$ and to 84.5% for the initial HeV and NiV W pool, respectively).

The average R_g value of the two ensembles (71.7 \AA for HeV and 70.7 \AA for NiV) is slightly larger than that of the initial random pool, consistent with the expanding effect of the combined use of urea and DTT.

In conjunction with all the other experimental lines of evidence presented above, SAXS data definitely demonstrate that the W proteins are intrinsically disordered.

2.8. Phase Separation and Fibrillation Abilities of the W Proteins

IDPs and/or IDRs are known for their involvement in a broad range of phase separation behaviors [73–75]. IDRs can drive liquid-liquid phase separation (LLPS) and the resulting biomolecular condensates can undergo “maturation” toward a gel or solid state [74,76–78]. IDRs are also known to form amyloid-like structures that can form either from liquid samples or from hydrogels [79–81]. In line with these properties, we previously reported that HeV V jellifies *in vitro*. The minimal HeV V region responsible for this ability (referred to as PNT3, aa 200–310 of HeV V) was identified within its intrinsically disordered NTD [82,83]. Binding assays to an amyloid-specific dye and negative-staining transmission electron microscopy (TEM) studies, showed that PNT3 forms amyloid-like fibrils [82,83].

Because V and W possess both the minimal region conferring phase separation and fibrillation abilities (i.e., PNT3), we reasoned that W might phase separate and form fibrils as well. To ascertain this ability, we first analyzed the W proteins using Semi-Denaturing Detergent-Agarose Gel Electrophoresis (SDD-AGE). This technique enables distinguishing between amyloid-like protein polymers, which are very stable and withstand treatment with 2% SDS, and aggregates, which are bundles of polymers that dissociate under these conditions [84]. For both W proteins, SDD-AGE showed the presence of SDS-resistant oligomers, whose abundance increases by increasing the temperature of incubation from 4 °C to 37 °C (Figure 11A). By contrast, no such oligomers were observed for a sample of conalbumin preincubated at 65 °C for 30 min (Figure 11A), a condition known to lead to amorphous aggregates [85]. The HeV W protein was found to have a higher propensity to form amyloid-like fibrils as judged from the fact that at 4 °C SDS-resistant aggregates are discernible in the case of HeV but not for NiV (Figure 11A).

Both W proteins were found to phase separate and to form macroscopically observable condensates that sediment at the bottom of Eppendorf tubes, and that bind Congo Red (CR), an amyloid-specific dye (Figure 11B). The CR binding condensate was found to be slightly bigger in the case of the HeV W protein. To quantify this phenomenon, we spectrophotometrically measured the red shift of the absorbance maximum in the CR spectrum of a sample containing either the HeV W or NiV W protein. CR binding to amyloids (i.e., cross β -sheet enriched structures), is known to lead to hyperchromicity and to a red shift. The addition of either HeV W or NiV W, but not of an irrelevant, control IDP with no propensity to fibrillate (i.e., N-Tail), does indeed promote a shift in the CR spectrum from 497 nm to 515 nm (Figure 11C). To assess the stability of the amyloid-like fibrils, the W samples were also heated at 95 °C for five minutes prior to spectrophotometric measurements. Although for both proteins this heating step was found to lead to a significant decrease in the red shift, with this effect being more pronounced in the case of HeV, it does not abolish the ability of the proteins to bind CR, advocating for an at least partial heat stability of the fibrils (Figure 11C).

To further confirm the ability of the W proteins to form amyloid-like fibrils, we carried out binding assays with Thioflavin T (ThT), another well-known amyloid-specific dye [86]. Binding of ThT to the W proteins, but not to the control N-Tail protein, was found to induce an enhancement in the intensity of the ThT fluorescence emission spectrum in a time-dependent manner, although the extent and kinetics of the fluorescence increase is not the same for the two proteins (Figure 11D, upper panel).

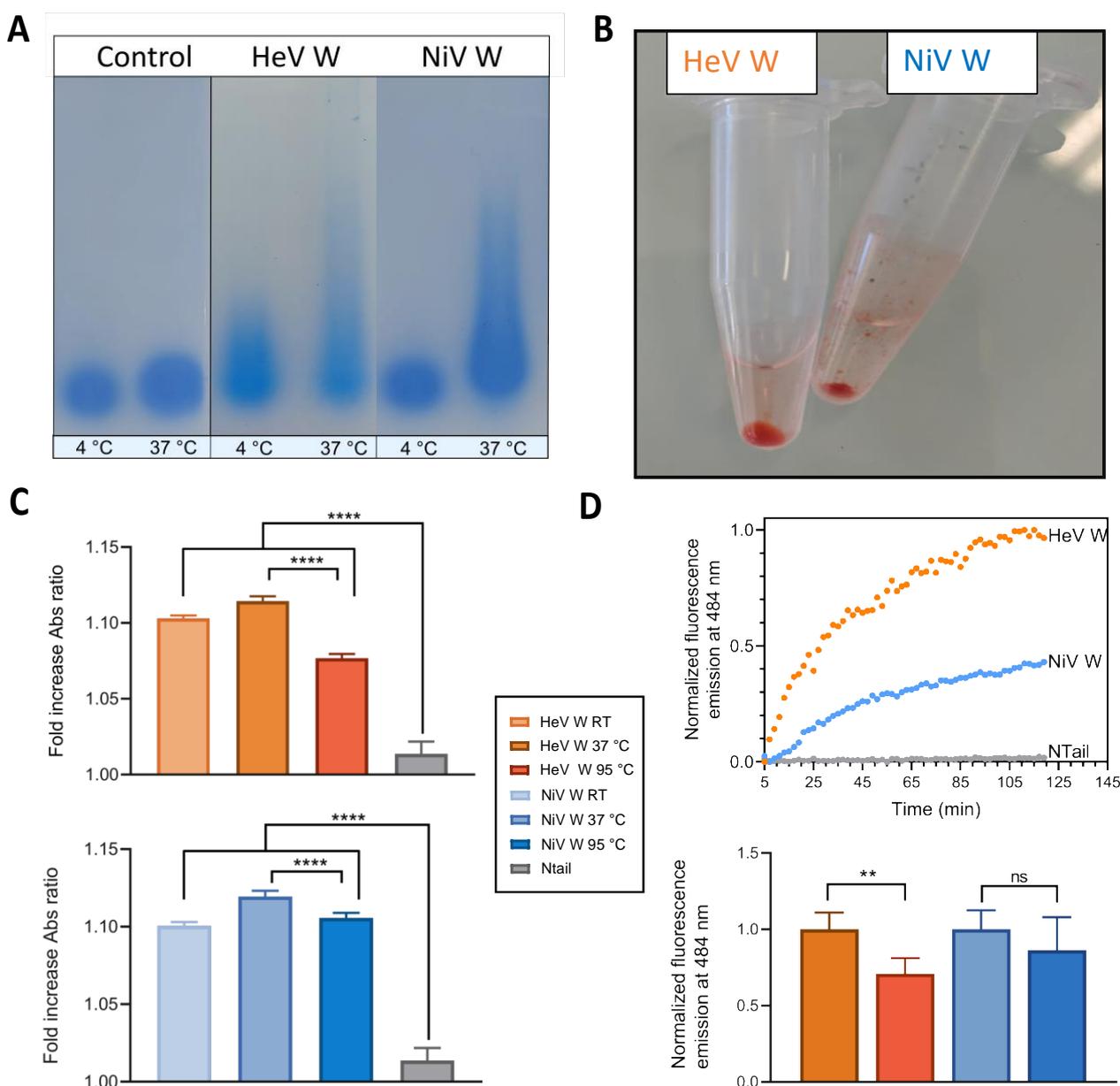


Figure 11. (A) SDD-AGE analysis of HeV and NiV W proteins, as well as of an irrelevant control protein (conalbumin, able to form amorphous aggregates upon heating), following an incubation of 48 h at either 4 °C or 37 °C. (B) CR binding by purified HeV & NiV W proteins. Proteins were incubated at 75 μ M in the presence of 18 μ M CR for three days at room temperature. A dense phase on the bottom of the tube that binds CR can be observed. (C) Fold increase in the ratio between the absorbance at 515 and at 497 nm, with respect to a sample containing CR alone, of a HeV W or NiV W or N-tail sample at 20 μ M after 27 h of incubation at RT or at 37 °C. Following incubation at 37 °C, the W proteins were also heated at 95 °C for five minutes. N-tail was chosen as a control IDP devoid of aggregation and fibrillation propensities. The error bar corresponds to the standard deviation, with $n = 3$. The four asterisks denote a statistically significant difference ($p < 0.0002$) with respect to the control (One-way ANOVA test). (D) Upper panel: Thioflavin T (ThT) binding assays of the W proteins and of a control protein (N-tail) at various time points of incubation at 37 °C. Proteins were incubated at 40 μ M in the presence of 40 μ M ThT. Data are representative of two independent experiments. Lower panel: Normalized fluorescence as obtained after 2 h of incubation at 37 °C followed by an additional incubation step of five minutes at either 37 °C or 95 °C. The error bar corresponds to the standard deviation, with $n = 3$. The two asterisks denote a statistically significant difference ($p < 0.0069$) with respect to the control (One-way ANOVA test). ns: not significant.

In particular, the kinetics of fluorescence enhancement of the NiV W protein is slower and reaches a lower value (~one half of the value reached by the HeV W protein) (Figure 11D, upper panel). Following two hours of incubation at 37 °C, an additional incubation step of the W samples at 95 °C for five minutes was found to trigger a significant decrease in the fluorescence only in the case of HeV (Figure 11D, lower panel). In line with the CR experiments described above, heating does not fully abrogate ThT binding thus supporting ability of the fibrils to partly resist to heat treatment. The lack of significant impact of the heating step on the fluorescence of the NiV W protein confirms the higher heat resistance of the fibrils made of NiV W compared to those made of HeV W.

Finally, we analyzed the purified W proteins using negative-staining TEM, which unambiguously revealed the presence of fibrils for both HeV and NiV W (Figure 12). The formation of amyloid-like fibrils by the W proteins was monitored on samples at different concentrations (10 and 40 µM) and as a function of time. In the case of the HeV W protein, the two samples at the two concentrations show a very similar evolution over time (Figure 12A,B): at t₀, short fibrils are visible in both grids, suggesting an almost instantaneous formation after the removal of the denaturing agent. Statistical analysis of fibril length (end-to-end distance) at t₀ shows a highly similar Gaussian distribution for the two samples, centered on 60 nm for the sample at 10 µM, and at 70 nm for the sample at 40 µM, indicating that at the higher concentration longer fibrils are formed (Figure 12A). At t₂ (10 h of incubation) the protein solution at 40 µM is turbid, and solid floccules are visible to the naked eye (Supplementary Figure S7A). This observation is mirrored in Figure 12B, where for both concentrations three-dimensional fibril clusters are visible. However, the size of the agglomerates observed at t₁ (2 h of incubation) is about 3 times larger than that observed at t₂ (Supplementary Figure S7B). This can be explained by the macroscopic aggregation observed in solution, suggesting that aggregates formed at t₂ from the 40 µM sample are so large to experience fragmentation and breakage during the process of dilution and transfer on the grid. In fact, while in the 10 µM sample agglomerates of variable size and single fibrils are still visible, at 40 µM small agglomerates are present, but single fibrils and intermediate stages of aggregation are absent (Figure 12).

In the case of the NiV W protein, although the final protein concentration in the grids is the same (1.6 µM), the difference between the two concentrations at t₀ is more pronounced, both in terms of fibril length and abundance, probably due to slower formation kinetics (Figure 12C). At t₀ a high percentage of protein is still in the monomeric state, undetectable on the grid. In addition, for the sample at 10 µM, small, elongated structures of a few nm are present, but disappear upon incubation for 10 h (Figure 12D). Those structures are also discernible, although less abundant, in the sample at 40 µM (Figure 12C). As in the case of the less concentrated sample, they disappear upon incubation for either 2 (Supplementary Figure S7C) or 10 h (Figure 12D). This suggests that they might be germinating seeds of fibrils, where elongation takes place. Comparing the two NiV W Gaussians at t₀, one can see an increase in the length of the fibrils with increasing concentration (distribution centered on 40 nm for the 10 µM sample, 60 nm for the 40 µM sample). Upon increasing the incubation time to 10 h (t₂), a moderate elongation of the fibrils can be observed for both concentrations, leading to a homogeneous layer on the grid. In both cases, three-dimensional assembly structures between the fibrils are absent, in contrast to HeV W.

For both proteins, fibril width was also measured (Supplementary Figure S7D), and was found to remain constant over time and to be independent from the initial sample concentration (data not shown). The obtained distributions, averaged over all the incubation times, show that fibrils have a mean width value (i.e., 13.92 ± 3.7 nm for HeV and 17.25 ± 2.5 nm for NiV) that falls within the size range of amyloid-like fibrils (10–20 nm) [87]. Much caution should however be taken when considering these values, as they can be biased by non-constant fibril thickness and negative-staining effects.

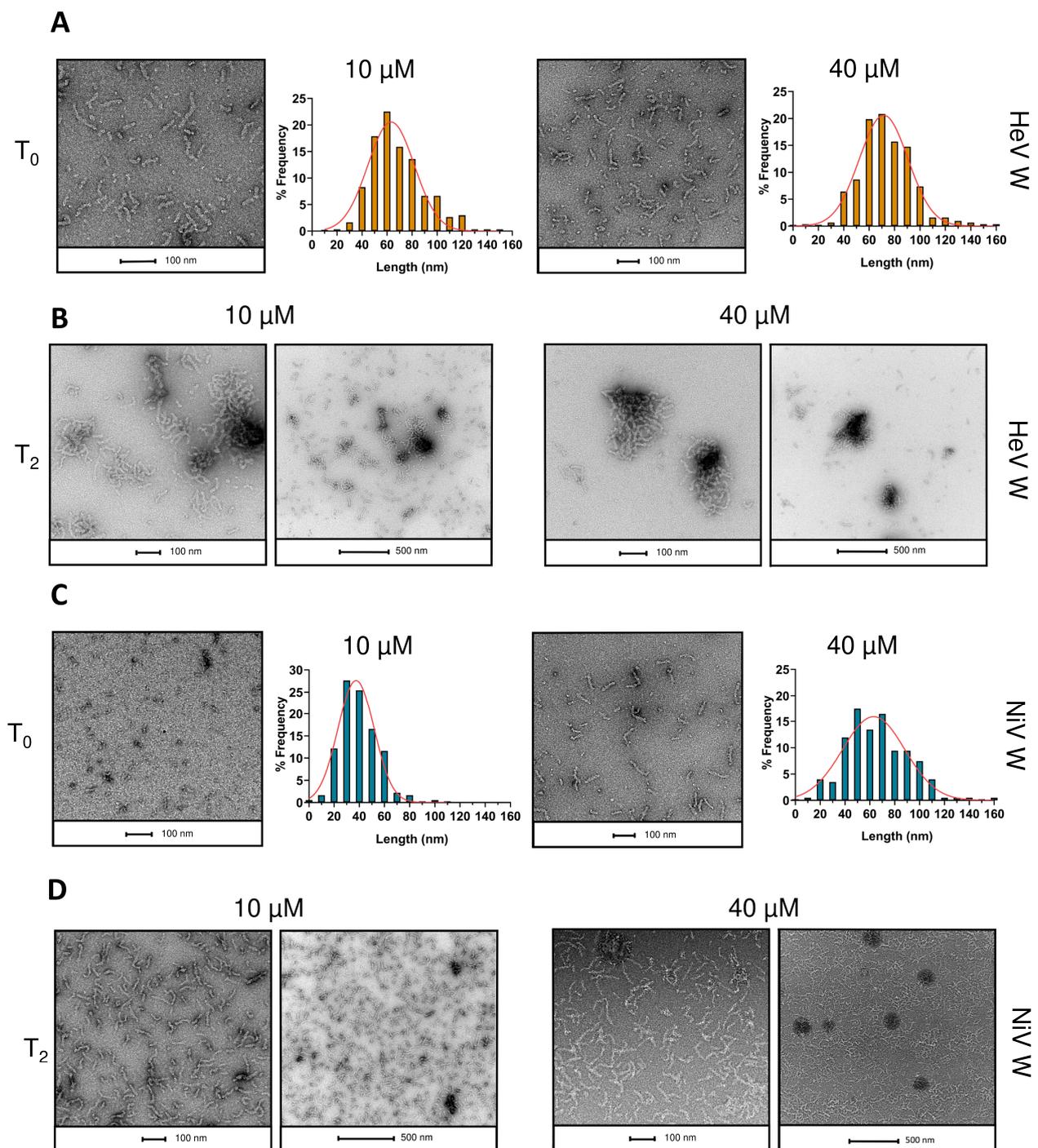


Figure 12. Negative-staining TEM analysis of HeV and NiV W proteins. (A) t_0 (0 h) of incubation at 37 °C and relative fibril length distribution for the HeV W protein at 10 μM (left side) or 40 μM (right side). (B) t_2 (10 h) of incubation at 37 °C for the HeV W protein at 10 μM or 40 μM . (C) t_0 (0 h) of incubation at 37 °C and relative fibril length distribution for the NiV W protein at 10 μM (left side) or 40 μM (right side). (D) t_2 (10 h) of incubation at 37 °C for the NiV W protein at 10 μM or 40 μM . Note that in all cases, samples were diluted to 1.6 μM prior to being deposited on the grid. The distribution of fibril length, as obtained from analysis of the end-to-end distance of 250–300 fibrils for each data set at time zero, is shown. Each bar in the histogram, centered on n , corresponds to fibrils whose length is comprised between $n - 4$ and $n + 5$. The analysis was done using the ImageJ software.

Altogether, these data argue for the ability of the W proteins to form amyloid-like assemblies, consistent with what has been macroscopically observed in aqueous solvents. In agreement with the other experimental lines of evidence, the HeV W protein was found to have a much more pronounced fibrillation propensity, with many long fibrils being already detectable at time zero at both protein concentrations (Figure 12A) contrary to the NiV W proteins for which fibrils can only be detected at 40 μM (Figure 12). Moreover, while the fibrils formed by the HeV W protein tend to form large agglomerates, the fibrils formed by NiV W tend to form a more homogeneous layer.

3. Discussion

By combining *in silico* and experimental studies we herein show that the HeV and NiV W proteins are intrinsically disordered. The W proteins were found to be expressed as inclusion bodies in *E. coli*. This prompted us to purify them under denaturing conditions, a process often used to purify IDPs. Indeed, because IDPs lack stable secondary or tertiary structures, they remain soluble even in extreme condition of pH, temperature or salinity, with this property being even commonly exploited for their purification [88,89].

Despite their overall disordered nature, they possess some transiently populated secondary structure, particularly represented by β -strands, and some fluctuating tertiary structure. Indeed, their experimentally observed Stokes radii are closer to those expected for PMG states than to those expected for IDPs of the same size. The W proteins are more compact than expected for IDPs, a property already observed for the cognate V proteins [33]. Taking into account the high proline content of the W proteins, we first attempted at rationalizing this discrepancy by checking whether the observed hydrodynamic radii were closer to the values expected by the sequence-based model, which takes into account the proline content among other sequence features [90]. The R_S^{IDPseq} radii calculated accordingly (see Equation (5) in Materials and Methods) were however found to be even more divergent from the experimentally observed R_S values, being slightly larger (58.5 \AA for HeV W and 57.8 \AA for NiV W) than those calculated using the simple power law model (57.4 \AA for HeV W and 57.5 \AA for NiV W). Therefore, the sequence determinants that dictate the relative compaction of the W proteins cannot be captured by a model that takes into account only the sequence length, the proline and the charge content. The linear distribution of proline and of charged residues, along with that of other residues whose role still remains to be uncovered, might be responsible for the higher compactness of these proteins. In particular, long-range effects, which are not captured by either the power law or the sequence-based model, may come at play especially taking into account the fact that the W_{CTD} is basic while the W_{NTD} is acidic. Moreover, and of even more relevance, intramolecular disulfide bridges appear to play a role in driving protein compaction, as judged from the fact that the addition of DTT triggers an increase in the Stokes radius.

While the R_S -based CI of HeV W (0.40) is lower than that of HeV V (0.49), the CI of NiV W (0.43) is very close to that of NiV V (0.44). Since the R_S -based CI of HeV NTD is equal to the CI of NiV NTD (0.51) [13,33], differences in compactness between V and W in the two viruses are to be ascribed to differences in their constituent CTDs. In the case of HeV, the higher compactness of V with respect to W can be ascribed to the fact that V_{CTD} adopts a folded Zn-finger conformation while W_{CTD} is predicted to be disordered. By contrast, it is more difficult to rationalize the similarity in the CIs of the NiV V and W proteins. If from one hand this similarity would suggest that the NiV W_{CTD} is more compact than the HeV W_{CTD} , bioinformatics analyses would suggest the opposite scenario. Indeed, the HCA plot, the disorder score, as provided by CIDER (Table 1), and secondary structure predictions predict a slightly higher order content in HeV W_{CTD} . On the other hand, the higher compactness of NiV W_{CTD} could arise from an enrichment in histidine residues in the NiV W_{CTD} compared to HeV W_{CTD} (Supplementary Figure S3), which could promote local and transient folding through π - π or cation- π interactions. Moreover, His residues have an enhanced sensitivity to pH, being protonated under slightly acidic conditions,

which could lead to charge-driven compaction phenomena. Definite answers in this regard await direct experimental investigation of the W_{CTD} from the two proteins.

SEC experiments, as well as MS and SDS-PAGE analyses, highlighted for both proteins the existence of a disulfide-bridged dimer that may be stabilized by one of the three conserved cysteine residues. The fact that the addition of urea triggers a reduction in the elution volume of the dimer from the SEC column, argues for the presence under native conditions of residual secondary and/or tertiary structure within both the monomers building up the dimer. The chain expansion observed for the monomer under reducing conditions, supports the presence of one (or more) intramolecular disulfide bridges that play a role in driving protein compaction. Thus, in contrast with predictions provided by IUPred2A, the cysteine redox state seemingly impacts protein compaction. Specifically, the oxidized state imparts a higher compactness while the reduced state is associated with expansion of the protein chain. In addition, the oxidized state is also characterized by a higher secondary structure content, as judged from far-UV CD studies. In line with these observations, for both proteins, the R_g values experimentally observed under reducing conditions and in the presence of 1 M urea, exceed by ~12% the values expected for IDPs of the same length.

The critical role of the cysteines in dictating the conformational properties of the W proteins is also in line with the experimental observation that the addition of DTT in native conditions (i.e., without urea) leads to rapid, macroscopically visible, aggregation. This behavior strongly supports the involvement of disulfide bridge(s) in preventing the aggregation process of the W proteins. Whether the cysteine redox state may impact the fibrillation process as well, remains to be established and is the focus of current studies that also aim at identifying the cysteine residues involved in the formation of inter- and intramolecular disulfide bridges. Moreover, the biological relevance of a possible role of disulfide bridges in modulating W protein aggregation and fibrillation remains to be assessed in light of the effective ratio between the oxidized and reduced form in the cytoplasm and in the nucleus of infected (or transfected) cells.

CD studies in the presence of TFE unveiled a pronounced α -helical propensity in both W proteins. Please note that although TFE is known to stabilize α -helices more than β -strands, it is worth underscoring that the gain of α -helicity in the presence of TFE is not a general rule and hence truly reflects the inherent structural propensities of the protein under study. Indeed, some IDPs gain little or no α -helical content at high (i.e., 50–90%) TFE concentrations or even fold as β -sheets [91,92]. In further support of a role of TFE as a genuine secondary structure stabilizer, the intrinsically disordered N_{TAIL} domain from three paramyxoviruses undergoes α -helical folding in the presence of 20% TFE exclusively within regions known to undergo partner-induced α -helical folding [93,94]. In light of these considerations, TFE can be regarded as a *bona fide* structure stabilizer that does not promote non-native folding in IDPs. The inherent α -helical propensities of the W proteins can thus be regarded as reflecting the presence of transiently and partly structured α -MoREs, i.e., short regions prone to undergo partner-induced α -helical folding and whose transient α -helical conformation would be stabilized by TFE. In support of this, bioinformatics analyses predict many putative MoREs and α -helices within the W proteins, and their hydrodynamic radii and ellipticities values are consistent with a PMG state. Previous NMR studies focusing on NiV NTD [14,20] provide additional support pointing to the presence of transiently populated α -helices within the W proteins. On the other hand, in light of the ability of the W proteins to form amyloids, it is surprising and puzzling that TFE did not reveal any β -propensities. In this regard, it is noteworthy that although canonical amyloids have a typical cross β -sheet structure, amyloids with a cross- α architecture have been described as well [95,96]. In this latter case, amphipathic α -helices, rather than β -strands, stack perpendicular to the fibril axis and tightly self-associate into sheets to give rise to fibrillar structures with morphological and tinctorial properties similar to those of canonical cross- β amyloids. Whether amyloids formed by *Henipavirus* W proteins have a

cross β -sheet or cross- α architecture remains to be established and awaits structural studies that are currently in progress in our lab.

Although CD studies unveiled a similar disorder content in the two W proteins, the NiV W protein was found to exhibit a markedly higher α -helical propensity. This difference likely arises from the higher α -helical propensity of NiV NTD with respect to HeV NTD [13,33]. Overall, the W proteins exhibit a higher α -helical propensity compared to the V proteins, where the latter were found to gain an α -helical content of $\sim 20\%$ at 30% TFE [33]. This difference likely arises from a propensity of the W_{CTD} to adopt an α -helical folding whereas the V_{CTD} is known to adopt a Zn-finger, β -enriched conformation [33].

The occurrence of partly preconfigured MoREs has been proposed to enable a more efficient start of the folding process induced by a binding partner via a reduction of the entropic cost associated with the disorder-to-order transition [49,50,61,97–99]. In line with this, the W proteins have a broad molecular partnership. Indeed, analysis of the interactome of the two W proteins highlighted 90 interactions in total where 53 partners have been identified for HeV W and 37 for NiV W (Figure 13 and Supplementary Table S1). All of the reported partners are human, except for the murine inhibitor of nuclear factor kappa-B kinase subunit alpha protein, whose sequence is however very close to that of its human counterpart (UniProt Accession number O15111). HeV W and NiV W share 22 common partners including several histones, importins, 14-3-3 proteins, STAT1 and STAT2 proteins and the creatine kinase B-type (Figure 13). Among the identified partners, 31 interact only with HeV W and 15 only with NiV W.

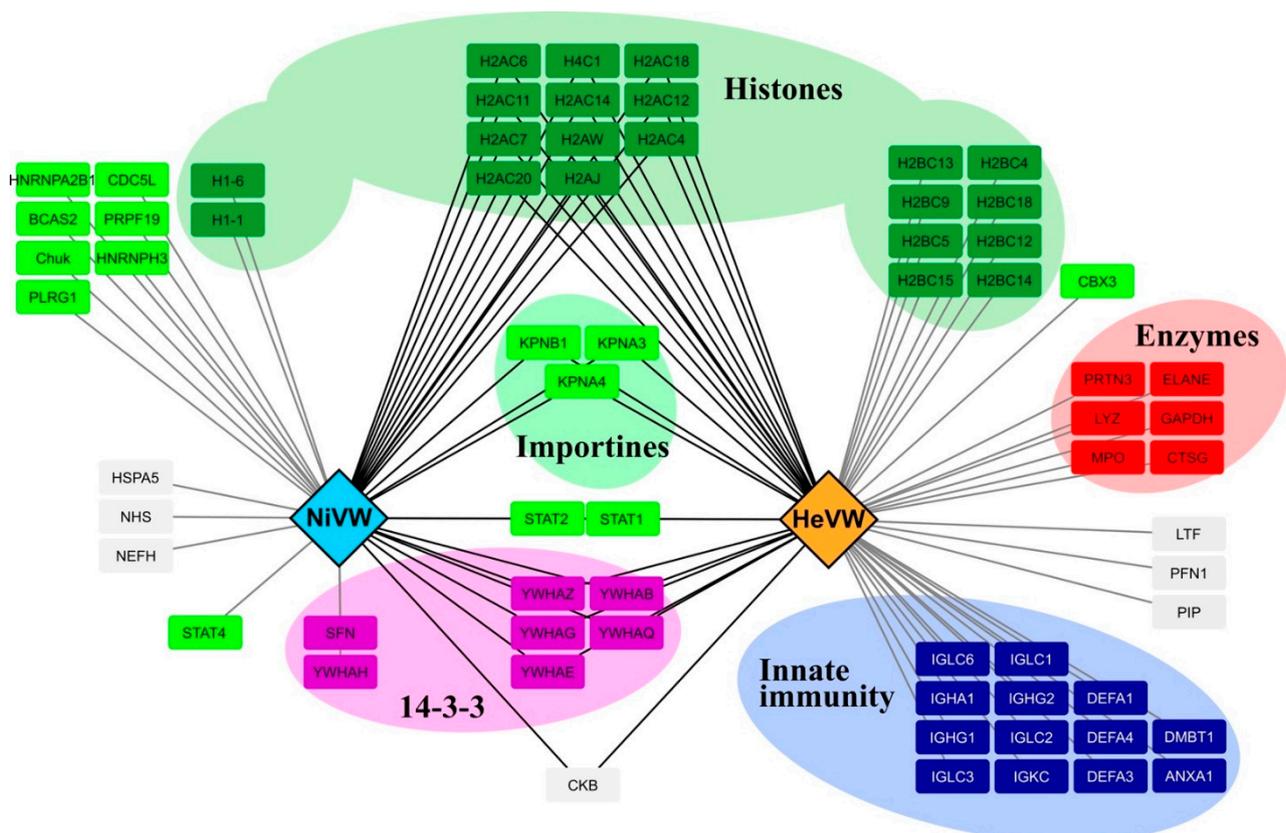


Figure 13. Interactome of the HeV and NiV W proteins as obtained using the INTACT database [100] and literature data mining. HeV W and NiV W are represented as blue and orange diamonds, respectively. The gene names of W protein partners are indicated in the nodes. Black lines connect HeV and NiV W with their common partners while gray lines connect them with their unique partners. Green nodes: nuclear proteins (dark green: histones, bright green: other nuclear proteins). Gray nodes: proteins without particularity.

The majority of the W interacting-proteins are nuclear, mainly represented by histones. This is consistent with the presence of a nuclear localization sequence in the C-terminal part of the W proteins as well as with their reported interaction with importins [35,36]. The W proteins interact with the 14-3-3 protein family regulating apoptosis, metabolic processes and extracellular matrix organization [42], and interfering with the NF- κ B-induced proinflammatory response [43]. A non-neglectable amount of protein partners (13) are implicated in innate immunity, i.e., several immunoglobulins, defensins and annexin A1. Interestingly, six enzymes possessing antimicrobial activities have been found to interact with HeV W. Finally, NiV W interacts with heterogeneous nuclear ribonucleoproteins proteins A2/B1 and H3, where A2/B1 has been found to phase separate [101].

Although the NiV W protein has a higher α -helical propensity, a higher CI (as estimated from both SEC and SEC-SAXS), and a higher predicted order content with respect to its HeV counterpart, it was unexpectedly found to exhibit a slightly higher protease sensitivity. This apparent conflicting behavior might arise from the much higher aggregation propensity of the HeV W protein that may hinder protease cleavage. This observation mirrors our previous studies where the HeV V protein was found to jellify under conditions where NiV V did not [82,83].

In line with the ability of the HeV V PNT3 region, common to both the V and W proteins, to undergo a liquid-to-gel transition and to form amyloid-like fibers [82,83], we herein show that both the HeV and NiV W proteins form phase separated condensates that bind the amyloid-specific CR and ThT dyes.

Phase separation phenomena have attracted much attention in the last decade in light of their role in driving the formation of membrane-less organelles (MLOs) (e.g., Cajal bodies, P-bodies, nucleolus, stress granules, centrosomes, aggresomes, etc.) [79]. LLPS by viral proteins has been implicated in different steps and regulatory processes in viral replication cycles (for reviews see [102–104]). A functionally important manifestation of LLPS driven by viral proteins is the formation of so-called “viral factories”, i.e., cytoplasmic inclusion bodies where viral replication and assembly take place. Indeed, in various *Mononegavirales* members viral factories were shown to be liquid-like MLOs resulting from LLPS of their N and P proteins [102,105–112], with IDRs playing a crucial role in this process [103]. A functionally distinct and less well-known category of LLPS is “LLPS-mediated interference with host cell functions” [103], a process that is not associated with viral replication and assembly: it rather leads to interference of host cell functions through either interaction of the viral condensates with specific host genes [113] or with cellular proteins. Interestingly, the NiV W protein was found to interact with hnRNPA2B1, a protein known to be involved in proteinopathy and amyotrophic lateral sclerosis (ALS), to exhibit an intrinsic tendency to assemble into self-seeding fibrils and to be incorporated in stress granules [114]. As such, NiV W could either interfere with the integrated stress response or recruit host stress granule components to facilitate virus replication, in line with the well-known ability of viruses to regulate stress granules [115]. Interaction of viral condensates with cellular proteins can also lead to sequestration, and hence possible inactivation, of cell proteins involved in the cell innate immune response (see [103,104] and refs therein cited). The formation of fibrillar aggregates by viral proteins, reported in a very limited number of studies, could be revisited in light of LLPS, where fibrils might correspond to solid-like inclusions formed upon maturation of liquid-like condensates. In this regard, it is worthy to note that the SARS-CoV2 nucleocapsid protein was shown to form amyloids in phase separated droplets [116].

In some out of the very few cases reporting the ability of viral proteins to form fibrillar aggregates, fibril formation was shown to be associated with various functional effects, namely (i) membrane disruption and cell toxicity (PB1-F2 from influenza A virus) [117], (ii) blockade of key cellular processes, such as necroptosis or apoptosis (RHIM-containing proteins from *Herpesviridae* members) [118–120], (iii) suppression of host cell RNA synthesis through host transcription factor sequestration in fibrillar aggregates (NSs from Rift valley

fever virus) [121] and (iv) suppression of IFN responses (e.g., silencing of IFN- β expression and degradation of PKR) [122].

Whether the W proteins fibrillate also in the cellular context remains to be established and is the focus of ongoing efforts in our lab. Indeed, one limitation of the present study is that heterologous expression of the W proteins in *E. coli* prevents post-translational modifications (PTMs). *Henipavirus* V proteins are phosphorylated in cellula [123], and phosphorylation of IDRs is known to both trigger conformational transitions [124] and modulate phase separation and fibrillation abilities [79]. Phosphorylation is however the sole PTM that have been reported so far regarding the W proteins.

Under the assumption that fibrillation also occurs *in cellula*, it is tempting to hypothesize that the phase separation and fibrillation abilities of the *Henipavirus* V and W proteins might be related to interference with the host innate immune response. In addition, it should be kept in mind that the minimal region of the HeV V protein conferring the ability to phase separate (i.e., PNT3) [82,83] is also part of the P protein that is a constituent of the viral factories of *Mononegavirales* members [102,103,109]. Therefore, it is conceivable that the *Henipavirus* P protein may phase separate and fibrillate as well, with this property being possibly functionally coupled to the formation of viral factories.

4. Materials and Methods

4.1. Bioinformatic Analyses

Sequence alignment of the HeV W and NiV W (Malaysia strain) proteins (Uniprot code P0C1C6 and P0C1C7, respectively) was performed using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/> (accessed on 1 April 2021)). The RH-plots were generated by the PONDR server (<http://www.pondr.com> (accessed on 1 April 2021)). The phase diagram plots were generated using the CIDER server (<http://pappulab.wustl.edu/CIDER/analysis/> (accessed on 1 April 2021)) [45]. Deviations in amino acid composition of W proteins and of their respective CTDs were computed as previously described [13] using the average amino acid frequencies of the SWISS-PROT database (as obtained from <https://web.expasy.org/protscale/> pscale/A.A.Swiss-Prot.html, 2020-05 release) as the reference value. The average amino acid frequencies of the SWISS-PROT database roughly correspond to the mean composition of proteins in nature. If the average composition of an amino acid X in SWISS-PROT proteins is CSPX, and CPX is the composition of X within a protein P, deviation from the composition of X of SWISS-PROT proteins was defined for P as (CPX-CSPX)/CSPX. The HCA plots were obtained using the MeDor server [125]. Disorder predictions were obtained using the IUPred2A server [48]. MoRFs were predicted using MoRFchibi SYSTEM (<https://morf.msl.ubc.ca/index.xhtml> (accessed on 1 February 2021)) [57]. Secondary structure elements were predicted using the PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/>) (accessed on 1 February 2021) [58].

Structure prediction was carried out using the Phyre2 protein fold recognition server (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) (accessed on 1 April 2021) [60].

The interactome of the HeV W and NiV W proteins was built through data mining and through interrogation of the IntAct database (<https://www.ebi.ac.uk/intact> (accessed on 1 March 2021)) [100]) including protein partners generated by spoke expansion. IntAct is an open-source database for molecular interaction data in which all interactions are derived from literature curation or direct user submissions. The interaction network was then drawn using Cytoscape 3.8.2 [126].

4.2. Generation of the W Expression Constructs

The HeV and NiV (Malaysia strain) W coding sequences were built by overlapping PCR using NTD and CTD coding sequences [13]. To this end, OneTaq[®] DNA polymerase (New England Biolabs, Ipswich, MA, USA) was used with a pDEST14 derivative containing a P gene optimized for the expression in *E. coli* as template [13]. Primers (Operon) were designed to introduce the gene fragment encoding the CTD of the W protein downstream

the gene fragment encoding the NTD. Forward primer was 62 nucleotides in length and was designed to bear at its 5' end an AttB1 site followed by a fragment encoding a Tobacco Etch virus (TEV) protease cleavage site (ENLYFQG) and by a 22-nucleotide long fragment annealing on the 5' end of the P gene. Reverse primer was 183 nucleotide long and was designed to anneal on the last 20 nucleotides of the NTD encoding region, to bear the 132-long fragment encoding the CTD of W followed by two stop codons followed in their turn by an AttB2 site. After digestion with DpnI (New England Biolabs, Ipswich, MA, USA) to remove the methylated DNA template, the amplicons were subsequently transferred into the pDONR vector (Invitrogen, Carlsbad, CA, USA) through a Gateway BP clonase-mediated recombination reaction (Invitrogen, Carlsbad, CA, USA). From the resulting entry vectors each W gene was transferred into either the pETG-20A [127] or the pDEST17OI expression vector using Gateway LR clonase (Invitrogen, Carlsbad, CA, USA). The pETG-20A vector drives the expression under the T7 promoter of a thioredoxin (TRX) tagged fusion protein in which the protein of interest is preceded by a hexahistidine (His₆) tag. The resulting constructs (referred to as pETG-20A/HeVW and pETG-20A/NiVW) encode fusion W proteins in which the TRX tag, the hexahistidine tag and the AttB1-encoded amino acid stretch can be cleaved off by TEV protease digestion thus leading to a protein bearing only a non-native N-terminal glycine residue. The pDEST17OI expression vector is a modified pDEST17 vector (Invitrogen, Carlsbad, CA, USA) in which LacO and LacI encoding sequences were inserted upstream the T7 promoter by swapping pDEST17 and pET-DEST42 (Invitrogen, Carlsbad, CA, USA) PshAI/XbaI fragment in order to allow a better control of protein expression [128]. The resulting constructs (referred to as pDEST17OI/HeVW and pDEST17OI/NiVW) allow the expression under the T7 promoter of N-terminally hexahistidine tagged W proteins in which the native W sequence is preceded by a vector-encoded MSYHHHHHHLESTSLYKKAGS amino acid stretch and by a TEV cleavage site.

4.3. Expression and Purification of the W Proteins

The *E. coli* strain T7pRos was used for the expression of all the recombinant proteins. Cultures were grown overnight to saturation in LB medium containing 100 µg mL⁻¹ ampicillin and 34 µg mL⁻¹ chloramphenicol. An aliquot of the overnight culture was diluted 1/20 into 1 L of TB medium and grown at 37 °C. When the optical density at 600 nm (OD₆₀₀) 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 over night at either 25 °C (pETG20A-derived constructs) or 37 °C (pDEST17OI-derived constructs). The induced cells were harvested, washed and collected by centrifugation (6000 × g, 20 min).

The W proteins fused to TRX were purified as described for the TRX-fused V proteins [33]. The pellets of bacteria transformed with pDEST17OI/W constructs were resuspended in 50 mL per liter of culture of buffer A1 (50 mM Tris/HCl, pH 8, 300 mM NaCl, 8 M urea) and frozen at –20 °C. Upon thawing, the cells were disrupted by sonication as described above and clarified by centrifugation. The supernatant was purified by immobilized metal affinity chromatography (IMAC) using a fast protein liquid chromatography (FPLC) Äkta system (Cytiva, Marlborough, MA, USA). The column was washed with buffer A1 supplemented with 20 mM imidazole and elution was in buffer A1 containing 250 mM imidazole. The eluents from IMAC were analyzed by SDS-PAGE and the fractions containing the protein of interest were pooled and the sample was further purified by ion exchange chromatography (IEC). To this end the sample was desalted using a HiPrep 26/10 desalting column (Cytiva, Marlborough, MA, USA) and buffer B1 (20 mM Hepes pH 7, 50 mM NaCl, 8 M urea). The fractions obtained were then loaded onto a HiPrep DEAE FF 16/10 column (Cytiva, Marlborough, MA, USA) and the protein was eluted using a linear gradient of buffer B2 (20 mM Hepes pH 7, 500 mM NaCl, 8 M urea). The fractions containing the purified recombinant proteins were stored at –20 °C in buffer B2. Prior to all ensuing analyses, urea was removed from the protein samples and the buffer exchanged using Sephadex G-25 medium columns (Cytiva, Marlborough, MA, USA). Both IMAC and IEC were performed at room temperature.

Protein concentrations were calculated using the theoretical absorption coefficients at 280 nm as obtained using the program ProtParam at the EXPASY server (<http://web.expasy.org/protparam/>) (accessed on 1 February 2021).

4.4. Mass Spectrometry

4.4.1. Intact Protein Mass Analysis

Protein masses were determined on purified solution samples. To this end, the proteins in 8 M urea containing buffer were loaded onto a G-25 column (Cytiva, Marlborough, MA, USA) and eluted with 10 mM ammonium acetate. After a concentration step using Zip Tip C4 (Merck Millipore Darmstadt, Germany), 1 μ L of protein at \sim 1.5 mg/mL was mixed with 1 μ L of sinapinic acid matrix solution in 0.3% TFA/CH₃CN (50:50 *v/v*). One μ L of the mix was spotted on the target and analyzed by MALDI-TOF on a Ultraflex III spectrometer (Bruker Daltonics, Wissembourg, France) controlled by the Flexcontrol 3.0 package (Build 51) and operated in the linear mode, using a maximum accelerating potential of 25 kV and a 20,000–100,000 *m/z* range (LP_66kDa method). The laser frequency was fixed to 100 Hz and \sim 1000 shots per sample were cumulated. Four external standards (Protein Calibration Standard II, Bruker Daltonics) were used to calibrate each spectrum to a mass accuracy within 100 ppm. Peak picking was performed using the FlexAnalysis 3.0 software with an adapted analysis method. Parameters used were the centroid peak detection algorithm, S/N threshold fixed to 5 and a quality factor threshold of 30.

4.4.2. Peptide Mass Fingerprintings

Protein bands were excised from SDS-PAGE and digested overnight at 37 °C with high-sequencing-grade trypsin (Promega, Madison, WI, USA) after cysteine reduction in the presence of 10 mM DTT and alkylation with 50 mM iodoacetamide. Peptide extracts were pooled and dried in a centrifugal vacuum system. Samples were reconstituted with 0.1% trifluoroacetic acid in 4% acetonitrile and analyzed by LC-MS/MS using an Orbitrap QExactive Plus Mass Spectrometer (Thermo Electron, Bremen, Germany) online with an Ultimate 3000RSLCnano chromatography system (Thermo Fisher Scientific, Sunnyvale, CA, USA). Peptides were separated on a Dionex Acclaim PepMap RSLC C18 column. First, peptides were concentrated and purified on a pre-column in solvent A (0.1% formic acid (FA) in 2% acetonitrile). In the second step, peptides were separated on a reverse-phase LC EASY-Spray C18 column from Dionex (PepMap RSLC C18, 50 cm \times 75 μ m I.D, 100 Å of pore size, 2 μ m of particle size) at 300 nL/min flow rate. After column equilibration using 4% of solvent B (20% water–80% acetonitrile–0.1% FA), peptides were eluted by a two step linear gradient (4–20% acetonitrile/H₂O; 0.1% FA for 90 min and 20–45% acetonitrile/H₂O; 0.1% FA for 30 min). For peptide ionization in the EASY-Spray nanosource, spray voltage was set at 2.2 kV and the capillary temperature at 275 °C. The Orbitrap QExactive was used in data-dependent mode to switch consistently between MS and MS/MS. MS spectra were acquired with the Orbitrap in the range of *m/z* 400–1600 at a FWHM resolution of 120,000 measured at 400 *m/z*. AGC target was set at 4.0×10^5 with a 50 ms of maximum injection time. The more abundant precursor ions were selected and CID fragmentation was performed in the ion trap to have maximum sensitivity. The number of precursor ions was automatically defined with a maximum injection time of 300 ms. The signal threshold for an MS/MS event was set to 5000 counts. Charge state screening was enabled to exclude precursors with 0 and 1 charge states. Dynamic exclusion was enabled with a repeat count of 1 and a duration of 60 s. The acquired raw data were processed using Proteome Discoverer (version 1.4.1.14, Thermo Fisher Scientific). Spectra were searched using SEQUEST (Thermo Fisher Scientific) against a homemade database comprising 20,150 human and 4306 *E. coli* sequences and the HeV W and NiV W sequences both with and without the initial methionine. Search parameters were: (i) trypsin; two miscleavages allowed; (ii) mass tolerance of 6 ppm for monoisotopic precursor ions and 0.8 ppm for fragment ions from MS/MS; (iii) Cys carbamidomethylation (+57.02146 Da) as a fixed modification; Met oxidation (+15.99491 Da) and N-terminal acetylation (+42.0106 Da) as

variable modifications; (iv) minimum peptide length of four residues. Only high-score peptides were selected. Proteins were identified with a false discovery rate (FDR) of 1%.

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 Superose 6 Increase 10/300 column (Cytiva, Marlborough, MA, USA). Hepes Buffered Saline (HBS, 20 mM Hepes pH 7, 150 mM NaCl) was used as elution buffer unless differently indicated. Typically, 250 μL of purified protein at 2.5 mg mL^{-1} 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 (Thyroglobulin: 85 Å, Ferritin: 61 Å, Aldolase: 48.1 Å, Conalbumin: 36.4 Å, Carbonic anhydrase: 23 Å, RNase A: 16.4 Å and 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 [47]:

$$\log(R_S^{\text{NF}}) = 0.357 \times (\log \text{MM}) - 0.204 \quad (1)$$

$$\log(R_S^{\text{U}}) = 0.521 \times (\log \text{MM}) - 0.649 \quad (2)$$

$$\log(R_S^{\text{PMG}}) = 0.403 \times (\log \text{MM}) - 0.239 \quad (3)$$

The R_S^{NF} , R_S^{U} and R_S^{PMG} expected for a natively folded, and fully unfolded and PMG dimeric form of a protein with molecular mass MM were calculated according to Equations (1)–(3) by replacing $(\log \text{MM})$ with $(\log 2 \times \text{MM})$.

The R_S of an IDP with N residues was also calculated according to [90] using either the simple power law model:

$$R_S^{\text{IDP}} = R_0 N^\nu \quad (4)$$

where $R_0 = 2.49$ and $\nu = 0.509$, or the sequence-based model:

$$R_S^{\text{IDPseq}} = (A P_{\text{Pro}} + B) \times (C |Q| + D) \times S_{\text{his}} \times R_0 N^\nu \quad (5)$$

where $A = 1.24$, P_{Pro} = fractional proline content, $B = 0.904$, $C = 0.00759$, $|Q|$ = absolute value of the difference between the total number of negatively charged and positively charged residues, $D = 0.963$, $S_{\text{his}} = 0.901$ (correction factor for histidine tag).

The compaction index (CI) is expressed as according to [129]:

$$\text{CI} = (R_S^{\text{U}} - R_S^{\text{obs}}) / (R_S^{\text{U}} - R_S^{\text{NF}}) \quad (6)$$

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. Protease Sensitivity Assay

Protein samples in 50 mM Tris/HCl, pH 8, 150 mM NaCl buffer at a final concentration of 50 μM were incubated at room temperature with 2.5 $\mu\text{g mL}^{-1}$ thermolysin (Sigma Aldrich, Saint Louis, MO, USA) in a total reaction volume of 55 μL . The extent of proteolysis was analyzed by 15% SDS-PAGE on 5- μg samples withdrawn from the reaction mixture at specified time intervals after the reaction start. The reaction was stopped by mixing each sample with Laemmli electrophoresis buffer and immediately boiling for 5 min. Bovine serum albumin (BSA) (New England Biolabs, Ipswich, MA, USA) was digested in the same manner and used as a control.

4.7. Differential Scanning Fluorimetry (DSF)

The DSF analysis of the W proteins was carried out in the presence of a fluorescent dye as described in [130,131]. The experiment was conducted using a PCR instrument (Biorad)

and 96-well plates containing 25 μL of mixture per well. Each well contained 21.5 μL of HeV W or NiV W at 1 mg mL^{-1} in HBS and 3.5 μL of SYPRO Orange (ThermoFisher Scientific 7 \times solution in water (prepared from a 5000 \times stock solution in DMSO)). Fluorescent signals were acquired with excitation and emission wavelengths at 485 nm and 625 nm, respectively. Temperature scans were performed from 20 $^{\circ}\text{C}$ to 95 $^{\circ}\text{C}$.

4.8. Circular Dichroism

Far-UV CD spectra were measured using a Jasco J-810 dichrograph (Hachioji, Japan), flushed with N_2 and equipped with a Peltier thermoregulation system set at 20 $^{\circ}\text{C}$. One-mm thick quartz cuvettes were used. Proteins concentrations were 1 μM . Spectra were measured between 260 and 190 nm with a scanning speed of 50 nm/min and a data pitch of 0.2 nm. Response time was set to 4 s and the bandwidth to 2 nm. CD spectra were recorded in 10 mM sodium phosphate pH 7 either in the absence or in the presence of 1 M urea supplemented or not with 10 mM DTT. Each spectrum is the average of ten acquisitions. The spectrum of buffer was subtracted from the protein spectrum. Spectra were smoothed using PRISM.

Far-UV CD spectra were also recorded in the presence of increasing concentrations of TFE (from 10 to 50% v/v). The midpoints of transition and m values were derived upon fitting the data to a sigmoid curve using PRISM.

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

$$[\theta] = 3300m\Delta A/lcn \quad (7)$$

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^{-1} . Numbers of amino acid residues are 476 for HeV W and 478 for NiV W. Molecular masses are 52,706 Da for HeV W and 52,875 Da for NiV W.

The DICHROWEB website (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) (accessed on 1 April 2021), which was supported by grants to the BBSRC Centre for Protein and Membrane Structure and Dynamics (CPMSD) [132], was used to analyze the experimental data in the 190–260 nm range using unsmoothed, subtracted spectra. The content in the various types of secondary structure was estimated using the CDSSTR deconvolution algorithm with the reference protein set 7.

4.9. Small-Angle X-ray Scattering (SAXS)

Samples at 5 mg mL^{-1} in HBS containing 8 M urea were loaded onto a G-25 column (Cytiva, Marlborough, MA, USA) and eluted with freshly prepared HBS supplemented with 1 M urea and 5 mM DTT. The final protein concentration was 2.35 mg mL^{-1} for HeV W and 2.47 mg mL^{-1} for NiV W. Synchrotron X-ray scattering data were collected at SOLEIL (Gif-sur-Yvette, France) on the SWING beamline at a working energy of 12 KeV. Data were collected using a Dectris EIGER 4M detector at a sample-to-detector distance of 2.0 m and a wavelength (λ) of 1.03324 \AA . This setup leads to scattering vectors (q) between 4.56×10^{-4} and 0.540\AA^{-1} (with $q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle).

The calibration was made with water. For both proteins 70 μL were injected onto a BioSec 3–300 SEC column (Agilent, Santa Clara, CA, USA). The elution buffer was HBS, 1 M urea, 5 mM DTT. The flow rate was 0.2 mL min^{-1} and the temperature was 20 $^{\circ}\text{C}$. The exposure was in the continuous mode, with 1 frame sec^{-1} (990 ms exposure and 10 ms dead time). Data reduction and frames subtraction were done manually with the beamline software FOXTROT. Buffer blank frames were taken in the dead volume of the column.

Buffer-subtracted frames were submitted to Gaussian decompositions using the UltraScan solution modeler (US-SOMO) HPLC-SAXS module [133]. Deconvoluted data corresponding to the W proteins were compared with the CorMap test and averaged. The final scattering intensities were analyzed using the ATSAS program package [134]. Linearity in the Guinier region was used to exclude sample aggregation. The few data points differing from the Guinier fitting at low angles were deleted and the useful data

range was determined with SHANUM [134]. The radius of gyration (R_g) was estimated at low angles ($q < 1.3/R_g$) according to the Guinier approximation [135,136]:

$$\ln I(q) = \ln I_0 - \frac{(q^2 R_g^2)}{3} \quad (8)$$

The pair-distance distribution function, $P(r)$, from which the D_{\max} and the R_g were estimated, was computed using GNOM and manually adjusted until a good CorMap p -value ($\alpha > 0.01$) was obtained [137].

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

$$R_g = R_0 N^\nu \quad (9)$$

where N is the number of amino acid residues, R_0 a constant and ν a scaling factor. For IDPs, R_0 is 2.54 ± 0.01 and ν is 0.522 ± 0.01 [138], for chemically denatured (U) proteins R_0 is 1.927 ± 0.27 and ν is 0.598 ± 0.028 [138], and for natively folded (NF) proteins $R_0 = \sqrt{(3/5)} \times 4.75$ and $\nu = 0.29$ [139].

As in the case of the R_g , 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 [129]:

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

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 the R_S -based CI, this index increases with increasing compaction.

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).

Modeling of the W proteins as conformational ensembles was done with the program suite Ensemble Optimization Method (EOM) 2.0 [140] using the default parameters and random coil conformers. EOM first generates an initial pool of 10,000 random, $C\alpha$ -only conformers from the amino acid sequence. Subsequently, from this pool EOM 2.0 generates a conformational sub-ensemble that best fits the experimental SAXS data. To minimize overfitting, EOM attempts at minimizing the number of conformers able to fit the experimental data and usually generates ensembles consisting of 5 to 40 conformers. For both W proteins, the amino acid sequence provided as input to EOM was that of the recombinant W protein without the initial methionine, which was found to be cleaved off. Gaussian distributions of R_g and D_{\max} of the generated conformers were automatically calculated as well as their respective theoretical scattering profile. EOM uses a genetic algorithm (GA) to select an ensemble that best fits the experimental scattering curve. From the width of the R_g distribution the flexibility of the particles can be extracted, whereby a narrow distribution indicates a rather rigid particle and broader distributions indicate higher flexibility. Using EOM 2.0, systematic quantification of the flexibility was made using the metric Rflex—which computes the Shannon information entropy of the distributions [140]. Experimental error-independent goodness-of-fit was also confirmed using the software CorMap that estimates the differences between one-dimensional spectra independently of explicit error estimates, using only data point correlations [72].

SEC-SAXS data have been deposited in the Small-Angle Scattering Biological Data Bank (SASBDB) [141] under codes SASDLK9 and SASDLL9 for the set of data of HeV W and of NiV W, respectively. The HeV and NiV W ensembles derived using SEC-SAXS data have been deposited within the Protein Ensemble Database (PED-DB, <https://proteinensemble.org/>) (accessed on 14 December 2021) [142] under accession numbers PED00204 and PED00205, respectively.

4.10. SDD-AGE Analysis, and Congo Red and Thioflavin T Binding Assays

Semi-Denaturing Detergent-Agarose Gel Electrophoresis (SDD-AGE) analysis was performed on a 1.8% agarose (7 cm × 14 cm) gel in 0.1% SDS, Tris, Acetate, EDTA (TAE) buffer. Both HeV and NiV W (70 μM in HBS, pH 7) were incubated at either 4 °C or 37 °C for 48 h. Samples were then incubated for 10 min at 37 °C with SDD-AGE loading buffer (TAE 2×, 20% glycerol, 8% SDS, 1% Bromophenol blue) and loaded onto gels (60 μL each). A sample of conalbumin (75 kDa, (Cytiva, Marlborough, MA, USA) pre-heated at 65 °C for 30 min, was processed in the same way and used as a control. The electrophoretic run was performed at low voltage (3 V/cm gel length) in ice using 0.1% SDS, TAE as running buffer. The gel was stained with EZBlue™ Gel Staining Reagent (Sigma Aldrich, Saint Louis, MO, USA).

Congo Red (CR, Sigma Aldrich, Saint Louis, MO, USA) was used for both qualitative and quantitative binding assays. In both cases a stock solution of CR at 50 μM in Phosphate Buffer Saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4), 10% ethanol, 0.01% sodium azide was prepared. Qualitative binding assays were performed using the dye at a final concentration of 18 μM in the presence of either HeV W or NiV W at 75 μM in 50 mM sodium phosphate buffer at pH 7.2 in a final volume of 50 μL. The protein samples were previously incubated 3 days at room temperature to enable aggregates to form. At the end of the incubation, CR was added and incubated for 10 min. The samples were then centrifuged, and the sedimented condensate was washed three times with 50 μL of sodium phosphate buffer.

Quantitative, spectrophotometric measurement of CR binding was carried out using the dye at a final concentration of 5 μM in the presence of either HeV W or NiV W at 20 μM in 50 mM sodium phosphate buffer at pH 7.2 in a final volume of 100 μL. The adsorption spectrum was recorded after 27 h of incubation at either RT or 37 °C using a NanoDrop ND-1000 (Thermo Scientific, Waltham, MA, USA) spectrophotometer in the 400–700 nm range. Five μM CR in sodium phosphate buffer without the W protein was used as a control. Following incubation for 27 h at 37 °C, the W samples were also heated at 95 °C for five minutes. In all experiments, the dye was added after the incubation and prior to spectrophotometric measurements. The measles virus N-Tail protein, i.e., an IDP with no propensities to fibrillate [143], was used as a negative control.

The fluorescent dye Thioflavin T (ThT, Sigma Aldrich, Saint Louis, MO, USA) was used to monitor the aggregation kinetics of NiV and HeV W. A stock solution at 2 mM in HBS pH 7 was used. The proteins (40 μM in HBS pH 7) were incubated in the presence of 50 μM ThT for five minutes (equilibration time) before following the aggregation process by fluorimetry for up to two hours. Following incubation for two hours, the samples were also heated at 95 °C for five minutes and fluorescence emission measured again. Fluorescence measurements were carried out on a Cary Eclipse Fluorescence Spectrophotometer (Agilent, Santa Clara, CA, USA) equipped with a Peltier temperature controller set on 37 °C. ThT was excited at 440 nm (slitwidth: 5 nm) and fluorescence emission was recorded between 470 and 520 nm (slitwidth: 10 nm) every two minutes. Spectra were recorded with a scan rate of 300 nm/min with a data interval of 0.5 nm and an averaging time of 0.1 s. The measles virus N-Tail protein was used as a negative control in the same conditions. As an additional control, the fluorescence of a sample containing solely ThT was monitored and found not to increase over time.

4.11. Negative-Staining Transmission Electron Microscopy (TEM)

Two samples at different concentrations (10 and 40 μM) of NiV and HeV W were prepared and analyzed at different times to monitor their evolution. Incubation was carried out at 37 °C in HBS pH 7. Prior to each measurement, the samples were diluted to reach a final concentration of 1.6 μM. Drops of 3 μL of the diluted solution were deposited onto a glow discharge carbon coated grid (Carbon 300 mesh 3 mm Cu, TAAB). Prior to protein deposition, grids were exposed to plasma glow discharge for 20 s using a GloQube (Quorum, Lewes, UK) (Current 25 mA) in order to increase protein adhesion. The grids

were washed three times with 50 μ L of buffer, then washed in 35 μ L 2% (*w/v*) Uranyl Acetate solution (LauryLab, Brindas, France) before incubating them for 1 min in the latter solution. Excess of uranyl was blotted and grids were dried for 1 h at RT. Images were collected using a TECNAI T12 Spirit microscope (Thermo Fisher Scientific, Waltham, MA, USA) operated at 120 kV, and a Veleta 2 K \times 2 K CCD camera (Olympus, Shinjuku, Tokyo, Japan).

5. Conclusions

This study reports the characterization of the conformational properties of the *Henipavirus* W proteins, two key proteins in the host-pathogen race for which molecular data were hitherto conspicuously lacking. Compelling experimental evidence of their disordered nature and ability to fibrillate was obtained. The disordered nature of the W proteins likely endows them with several functional advantages such as (i) ability to interact with a large panel of (host) proteins, (ii) increased exposure of PTM sites, (iii) modulation of interaction strength and accessibility of binding sites through PTMs, (iv) alleviation of structural constraints on overlapping coding regions and mutational robustness (see also [14]), and (v) ability to phase-separate and to form amyloid-like fibrils in the condensed phase.

Notably, W fibrils are formed in near physiological conditions (salt, pH) and at a low protein concentration, thus advocating for a physiologically relevant phenomenon. The present study provides an additional example, among the few reported so far, of a viral protein forming amyloid-like fibrils, therefore significantly contributing to enlarge our currently limited knowledge of viral amyloids.

The present results constitute an asset for further investigating the ability of the *Henipavirus* W proteins to also form fibrils in the cellular context and for unraveling their functional impact. In light of the molecular partnership of the W proteins, our working hypothesis is that the W fibrillar condensates might interfere with the host innate immune and inflammatory response. In the long term, these studies, which are currently ongoing in our lab, are expected to contribute to shed light onto the molecular mechanisms of *Henipavirus* pathogenesis and pave the way to the design of new antiviral strategies aimed at abrogating the ability of these viruses to escape the innate immune response.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms23020923/s1>.

Author Contributions: S.L. conceived, designed and supervised the study and acquired funding; C.B. generated all the bacterial expression constructs; J.C. performed expression and purification of proteins driven by the pETG-20A constructs; P.F. carried out mass spectrometry analyses; J.F.N. performed DSF; D.P. and G.P. performed the TEM analyses; G.P. and F.G. carried out all the other experiments. F.G., G.P., D.P., P.F. and S.L. analyzed and interpreted the data; S.L. generated Figure 1 and the first draft of the manuscript. All the authors wrote and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out with the financial support of the CNRS. It was also partly supported by the French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INSB-0005) and by the Agence Nationale de la Recherche (ANR), specific project Heniphase (ANR-21-CE11-0012-01). 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.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data present in the current study are available from the corresponding author on reasonable request.

Acknowledgments: We thank Aurélien Thureau (SOLEIL) for his help in recording SEC-SAXS data and the SOLEIL synchrotron for beamtime allocation. We are also grateful to Gerlind Sulzenbacher

(AFMB lab) for efficiently managing the AFMB BAG. We heartily thank Cyrille Mathieu and Denis Gerlier (CIRI, Lyon, France), as well as Pierre-Yves Lozach (IVPC, Lyon, France), for the very useful discussions and for providing a very constructive criticism.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

The hydrophobic cluster analysis (HCA) provides a graphical representation of the sequence that enables identifying disordered regions [56]. Although HCA was not originally intended to predict disorder, it is very useful for unveiling disordered regions. HCA outputs can be obtained from <https://mobylye.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::HCA>. HCA (accessed on 1 April 2021) provides a two-dimensional helical representation of protein sequences in which hydrophobic clusters are plotted along the sequence. Glycine residues are represented as diamonds, prolines as stars, threonines as squares and serines as squares containing a dot. Acidic and basic residues are shown in red and in blue, respectively. Hydrophobic residues are shown in green. Disordered regions are recognizable as they are depleted in (or devoid of) hydrophobic clusters. HCA also enables identifying short regions with propensity to fold that appear as regions locally enriched in small hydrophobic clusters within regions otherwise devoid of such clusters.

References

- 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. [[CrossRef](#)] [[PubMed](#)]
- Wang, L.F.; Yu, M.; Hansson, E.; Pritchard, L.I.; Shiell, B.; Michalski, W.P.; Eaton, B.T. The exceptionally large genome of Hendra virus: Support for creation of a new genus within the family Paramyxoviridae. *J. Virol.* **2000**, *74*, 9972–9979. [[CrossRef](#)] [[PubMed](#)]
- Gurley, E.S.; Montgomery, J.M.; Hossain, M.J.; Bell, M.; Azad, A.K.; Islam, M.R.; Molla, M.A.; Carroll, D.S.; Ksiazek, T.G.; Rota, P.A.; et al. Person-to-person transmission of Nipah virus in a Bangladeshi community. *Emerg. Infect. Dis.* **2007**, *13*, 1031–1037. [[CrossRef](#)] [[PubMed](#)]
- Homaira, N.; Rahman, M.; Hossain, M.J.; Epstein, J.H.; Sultana, R.; Khan, M.S.; Podder, G.; Nahar, K.; Ahmed, B.; Gurley, E.S.; et al. Nipah virus outbreak with person-to-person transmission in a district of Bangladesh, 2007. *Epidemiol. Infect.* **2010**, *138*, 1630–1636. [[CrossRef](#)] [[PubMed](#)]
- Ching, P.K.; de los Reyes, V.C.; Sucaldito, M.N.; Tayag, E.; Columna-Vingno, A.B.; Malbas, F.F., Jr.; Bolo, G.C., Jr.; Sejvar, J.J.; Eagles, D.; Playford, G.; et al. Outbreak of henipavirus infection, Philippines, 2014. *Emerg. Infect. Dis.* **2015**, *21*, 328–331. [[CrossRef](#)] [[PubMed](#)]
- Plowright, R.K.; Becker, D.J.; Crowley, D.E.; Washburne, A.D.; Huang, T.; Nameer, P.O.; Gurley, E.S.; Han, B.A. Prioritizing surveillance of Nipah virus in India. *PLoS Negl. Trop. Dis.* **2019**, *13*, e0007393. [[CrossRef](#)]
- Ker, D.; Jenkins, H.; Greive, S.; Antson, A. CryoEM structure of the Nipah virus nucleocapsid assembly. *PLoS Pathog.* **2021**, *17*, e1009740. [[CrossRef](#)] [[PubMed](#)]
- 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. [[CrossRef](#)] [[PubMed](#)]
- 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. USA* **2020**, *117*, 4931–4941. [[CrossRef](#)]
- 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. [[CrossRef](#)]
- Yabukarski, F.; Lawrence, P.; Tarbouriech, N.; Bourhis, J.M.; Delaforge, E.; Jensen, M.R.; Ruigrok, R.W.; Blackledge, M.; Volchkov, V.; Jamin, M. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nat. Struct. Mol. Biol.* **2014**, *21*, 754–759. [[CrossRef](#)] [[PubMed](#)]
- 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. [[CrossRef](#)]
- 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. [[CrossRef](#)] [[PubMed](#)]
- 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. [[CrossRef](#)]

15. Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331. [[CrossRef](#)] [[PubMed](#)]
16. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415–427. [[CrossRef](#)]
17. Tompa, P. Intrinsically disordered proteins: A 10-year recap. *Trends Biochem. Sci.* **2012**, *37*, 509–516. [[CrossRef](#)] [[PubMed](#)]
18. Dunker, A.K.; Babu, M.M.; Barbar, E.; Blackledge, M.; Bondos, S.E.; Dosztányi, Z.; Dyson, H.J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; et al. What’s in a name? Why these proteins are intrinsically disordered: Why these proteins are intrinsically disordered. *Intrinsically Disord. Proteins* **2013**, *1*, e24157. [[CrossRef](#)]
19. Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V.N. Introducing Protein Intrinsic Disorder. *Chem. Rev.* **2014**, *114*, 6561–6588. [[CrossRef](#)]
20. 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. [[CrossRef](#)] [[PubMed](#)]
21. 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. [[CrossRef](#)] [[PubMed](#)]
22. 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* **2013**, *446*, 162–172. [[CrossRef](#)]
23. Beltrandi, M.; Blocquel, D.; Eroles, 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. [[CrossRef](#)] [[PubMed](#)]
24. 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. [[CrossRef](#)] [[PubMed](#)]
25. 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. [[CrossRef](#)]
26. Douglas, J.; Drummond, A.J.; Kingston, R.L. Evolutionary history of cotranscriptional editing in the paramyxoviral phosphoprotein gene. *Virus Evol.* **2021**, *7*, veab028. [[CrossRef](#)]
27. Fontana, J.M.; Bankamp, B.; Rota, P.A. Inhibition of interferon induction and signaling by paramyxoviruses. *Immunol. Rev.* **2008**, *225*, 46–67. [[CrossRef](#)]
28. Audsley, M.D.; Moseley, G.W. Paramyxovirus evasion of innate immunity: Diverse strategies for common targets. *World J. Virol.* **2013**, *2*, 57–70. [[CrossRef](#)]
29. 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*, 1913. [[CrossRef](#)]
30. Childs, K.; Randall, R.; Goodbourn, S. Paramyxovirus V proteins interact with the RNA Helicase LGP2 to inhibit RIG-I-dependent interferon induction. *J. Virol.* **2012**, *86*, 3411–3421. [[CrossRef](#)] [[PubMed](#)]
31. 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.* **2008**, *82*, 6259–6271. [[CrossRef](#)]
32. Ulane, C.M.; Horvath, C.M. Paramyxoviruses SV5 and HPIV2 assemble STAT protein ubiquitin ligase complexes from cellular components. *Virology* **2002**, *304*, 160–166. [[CrossRef](#)] [[PubMed](#)]
33. 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. [[CrossRef](#)]
34. 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.* **2004**, *78*, 5633–5641. [[CrossRef](#)] [[PubMed](#)]
35. Shaw, M.L.; Cardenas, W.B.; Zamarin, D.; Palese, P.; Basler, C.F. Nuclear localization of the Nipah virus W protein allows for inhibition of both virus- and toll-like receptor 3-triggered signaling pathways. *J. Virol.* **2005**, *79*, 6078–6088. [[CrossRef](#)] [[PubMed](#)]
36. Smith, K.M.; Tsimbalyuk, S.; Edwards, M.R.; Cross, E.M.; Batra, J.; Soares da Costa, T.P.; Aragao, D.; Basler, C.F.; Forwood, J.K. Structural basis for importin alpha 3 specificity of W proteins in Hendra and Nipah viruses. *Nat. Commun.* **2018**, *9*, 3703. [[CrossRef](#)]
37. Audsley, M.D.; Jans, D.A.; Moseley, G.W. Nucleocytoplasmic trafficking of Nipah virus W protein involves multiple discrete interactions with the nuclear import and export machinery. *Biochem. Biophys. Res. Commun.* **2016**, *479*, 429–433. [[CrossRef](#)]
38. Keiffer, T.R.; Ciancanelli, M.J.; Edwards, M.R.; Basler, C.F. Interactions of the Nipah Virus P, V, and W Proteins across the STAT Family of Transcription Factors. *mSphere* **2020**, *5*, e00449-20. [[CrossRef](#)]
39. Lo, M.K.; Miller, D.; Aljofan, M.; Mungall, B.A.; Rollin, P.E.; Bellini, W.J.; Rota, P.A. Characterization of the antiviral and inflammatory responses against Nipah virus in endothelial cells and neurons. *Virology* **2010**, *404*, 78–88. [[CrossRef](#)]
40. Satterfield, B.A.; Cross, R.W.; Fenton, K.A.; Agans, K.N.; Basler, C.F.; Geisbert, T.W.; Mire, C.E. The immunomodulating V and W proteins of Nipah virus determine disease course. *Nat. Commun.* **2015**, *6*, 7483. [[CrossRef](#)] [[PubMed](#)]
41. Enchery, F. Etude de la Modulation de la Voie Canonique D’activation de NF-κB par les Protéines non Structurales du virus Nipah. Thèse de Doctorat, Université de Lyon, Lyon, France, 2017.

42. Edwards, M.R.; Hoad, M.; Tsimbalyuk, S.; Menicucci, A.R.; Messaoudi, I.; Forwood, J.K.; Basler, C.F. Henipavirus W Proteins Interact with 14-3-3 To Modulate Host Gene Expression. *J. Virol.* **2020**, *94*, e00373-20. [[CrossRef](#)]
43. Enchery, F.; Dumont, C.; Iampietro, M.; Pelissier, R.; Aurine, N.; Bloyet, L.M.; Carbonelle, C.; Mathieu, C.; Journo, C.; Gerlier, D.; et al. Nipah virus W protein harnesses nuclear 14-3-3 to inhibit NF- κ B-induced proinflammatory response. *Commun. Biol.* **2021**, in press. [[CrossRef](#)]
44. Atkinson, S.C.; Audsley, M.D.; Lieu, K.G.; Marsh, G.A.; Thomas, D.R.; Heaton, S.M.; Paxman, J.J.; Wagstaff, K.M.; Buckle, A.M.; Moseley, G.W.; et al. Recognition by host nuclear transport proteins drives disorder-to-order transition in Hendra virus V. *Sci. Rep.* **2018**, *8*, 358. [[CrossRef](#)] [[PubMed](#)]
45. Holehouse, A.S.; Das, R.K.; Ahad, J.N.; Richardson, M.O.; Pappu, R.V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* **2017**, *112*, 16–21. [[CrossRef](#)]
46. Campen, A.; Williams, R.M.; Brown, C.J.; Meng, J.; Uversky, V.N.; Dunker, A.K. TOP-IDP-scale: A new amino acid scale measuring propensity for intrinsic disorder. *Protein. Pept. Lett.* **2008**, *15*, 956–963. [[CrossRef](#)] [[PubMed](#)]
47. Uversky, V.N. What does it mean to be natively unfolded? *Eur. J. Biochem.* **2002**, *269*, 2–12. [[CrossRef](#)] [[PubMed](#)]
48. Meszaros, B.; Erdos, G.; Dosztanyi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329–W337. [[CrossRef](#)] [[PubMed](#)]
49. Mohan, A.; Oldfield, C.J.; Radivojac, P.; Vacic, V.; Cortese, M.S.; Dunker, A.K.; Uversky, V.N. Analysis of Molecular Recognition Features (MoRFs). *J. Mol. Biol.* **2006**, *362*, 1043–1059. [[CrossRef](#)]
50. Vacic, V.; Oldfield, C.J.; Mohan, A.; Radivojac, P.; Cortese, M.S.; Uversky, V.N.; Dunker, A.K. Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.* **2007**, *6*, 2351–2366. [[CrossRef](#)] [[PubMed](#)]
51. Reichmann, D.; Jakob, U. The roles of conditional disorder in redox proteins. *Curr. Opin. Struct. Biol.* **2013**, *23*, 436–442. [[CrossRef](#)]
52. 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* **2010**, *107*, 8183–8188. [[CrossRef](#)] [[PubMed](#)]
53. Muller-Spath, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Ruegger, S.; Reymond, L.; Nettels, D.; Schuler, B. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14609–14614. [[CrossRef](#)] [[PubMed](#)]
54. 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* **2013**, *110*, 13392–13397. [[CrossRef](#)]
55. 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.* **2015**, *32*, 102–112. [[CrossRef](#)] [[PubMed](#)]
56. Callebaut, I.; Labesse, G.; Durand, P.; Poupon, A.; Canard, L.; Chomilier, J.; Henrissat, B.; Mornon, J.P. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): Current status and perspectives. *Cell. Mol. Life Sci.* **1997**, *53*, 621–645. [[CrossRef](#)] [[PubMed](#)]
57. Malhis, N.; Jacobson, M.; Gsponer, J. MoRFchibi SYSTEM: Software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res.* **2016**, *44*, W488–W493. [[CrossRef](#)]
58. McGuffin, L.J.; Bryson, K.; Jones, D.T. The PSIPRED protein structure prediction server. *Bioinformatics* **2000**, *16*, 404–405. [[CrossRef](#)] [[PubMed](#)]
59. 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.* **2004**, *78*, 5358–5367. [[CrossRef](#)] [[PubMed](#)]
60. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)]
61. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533. [[CrossRef](#)]
62. 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. [[CrossRef](#)]
63. Hames, B. *Gel Electrophoresis of Proteins: A Practical Approach*, 3rd ed.; Oxford University Press: Oxford, NY, USA, 1998.
64. Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R.C.; Giglione, C.; Meinel, T. The proteomics of N-terminal methionine cleavage. *Mol. Cell. Proteom.* **2006**, *5*, 2336–2349. [[CrossRef](#)] [[PubMed](#)]
65. Receveur-Bréchet, V.; Bourhis, J.M.; Uversky, V.N.; Canard, B.; Longhi, S. Assessing protein disorder and induced folding. *Proteins Struct. Funct. Bioinform.* **2006**, *62*, 24–45. [[CrossRef](#)]
66. Uversky, V.N. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* **2002**, *11*, 739–756. [[CrossRef](#)]
67. Hua, Q.X.; Jia, W.H.; Bullock, B.P.; Habener, J.F.; Weiss, M.A. Transcriptional activator-coactivator recognition: Nascent folding of a kinase-inducible transactivation domain predicts its structure on coactivator binding. *Biochemistry* **1998**, *37*, 5858–5866. [[CrossRef](#)] [[PubMed](#)]
68. Vincenzi, M.; Mercurio, F.A.; Leone, M. About TFE: Old and New Findings. *Curr. Protein Pept. Sci.* **2019**, *20*, 425–451. [[CrossRef](#)] [[PubMed](#)]
69. Bernado, P. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur. Biophys. J.* **2010**, *39*, 769–780. [[CrossRef](#)] [[PubMed](#)]

70. Gast, K.; Damaschun, H.; Misselwitz, R.; Muller-Frohne, M.; Zirwer, D.; Damaschun, G. Compactness of protein molten globules: Temperature-induced structural changes of the apomyoglobin folding intermediate. *Eur. Biophys. J.* **1994**, *23*, 297–305. [[CrossRef](#)] [[PubMed](#)]
71. DeLano, W.L. The PyMOL molecular graphics system. *Proteins Struct. Funct. Bioinform.* **2002**, *30*, 442–454.
72. Franke, D.; Jeffries, C.M.; Svergun, D. Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods* **2015**, *12*, 419–422. [[CrossRef](#)]
73. Holehouse, A.S.; Pappu, R.V. Functional Implications of Intracellular Phase Transitions. *Biochemistry* **2018**, *57*, 2415–2423. [[CrossRef](#)]
74. Alberti, S.; Hyman, A.A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 196–213. [[CrossRef](#)] [[PubMed](#)]
75. Fare, C.M.; Villani, A.; Drake, L.E.; Shorter, J. Higher-order organization of biomolecular condensates. *Open Biol.* **2021**, *11*, 210137. [[CrossRef](#)] [[PubMed](#)]
76. Shin, Y.; Brangwynne, C.P. Liquid phase condensation in cell physiology and disease. *Science* **2017**, *357*, eaaf4382. [[CrossRef](#)] [[PubMed](#)]
77. Tsang, B.; Pritišanac, I.; Scherer, S.W.; Moses, A.M.; Forman-Kay, J.D. Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell* **2020**, *183*, 1742–1756. [[CrossRef](#)] [[PubMed](#)]
78. Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176*, 419–434. [[CrossRef](#)] [[PubMed](#)]
79. Boeynaems, S.; Alberti, S.; Fawzi, N.L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L.; et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **2018**, *28*, 420–435. [[CrossRef](#)]
80. Wu, H.; Fuxreiter, M. The Structure and Dynamics of Higher-Order Assemblies: Amyloids, Signalosomes, and Granules. *Cell* **2016**, *165*, 1055–1066. [[CrossRef](#)]
81. Fuxreiter, M.; Vendruscolo, M. Generic nature of the condensed states of proteins. *Nat. Cell Biol.* **2021**, *23*, 587–594. [[CrossRef](#)]
82. Salladini, E.; Debarnot, C.; Delauzun, V.; Murrall, M.G.; Sutto-Ortiz, P.; Spinelli, S.; Pierattelli, R.; Bignon, C.; Longhi, S. Phase transition and amyloid formation by a viral protein as an additional molecular mechanism of virus-induced cell toxicity. *BioRxiv* **2018**, 497024. [[CrossRef](#)]
83. Salladini, E.; Gondelaud, F.; Nilsson, J.; Pesce, G.; Bignon, C.; Murrall, 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 3 phase transition and amyloid formation. *Biomolecules* **2021**, in press. 1324. [[CrossRef](#)]
84. Halfmann, R.; Lindquist, S. Screening for amyloid aggregation by Semi-Denaturing Detergent-Agarose Gel Electrophoresis. *J. Vis. Exp.* **2008**, 838. [[CrossRef](#)]
85. Khan, M.V.; Ishtikhar, M.; Rabbani, G.; Zaman, M.; Abdelhameed, A.S.; Khan, R.H. Polyols (Glycerol and Ethylene glycol) mediated amorphous aggregate inhibition and secondary structure restoration of metalloproteinase-conalbumin (ovotransferrin). *Int. J. Biol. Macromol.* **2017**, *94*, 290–300. [[CrossRef](#)]
86. LeVine, H., 3rd. Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: Detection of amyloid aggregation in solution. *Protein Sci.* **1993**, *2*, 404–410. [[CrossRef](#)] [[PubMed](#)]
87. Boyer, D.R.; Mynhier, N.A.; Saway, M.R. Why amyloid fibrils have a limited width. *BioRxiv* **2021**. [[CrossRef](#)]
88. Livernois, A.M.; Hnatchuk, D.J.; Findlater, E.E.; Graether, S.P. Obtaining highly purified intrinsically disordered protein by boiling lysis and single step ion exchange. *Anal. Biochem.* **2009**, *392*, 70–76. [[CrossRef](#)] [[PubMed](#)]
89. KrishnaKumar, V.G.; Gupta, S. Simplified method to obtain enhanced expression of tau protein from E. coli and one-step purification by direct boiling. *Prep. Biochem. Biotechnol.* **2017**, *47*, 530–538. [[CrossRef](#)]
90. Marsh, J.A.; Forman-Kay, J.D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **2010**, *98*, 2383–2390. [[CrossRef](#)]
91. Van Hoy, M.; Leuther, K.K.; Kodadek, T.; Johnston, S.A. The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. *Cell* **1993**, *72*, 587–594. [[CrossRef](#)]
92. Mouillon, J.M.; Gustafsson, P.; Harryson, P. Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. *Plant Physiol.* **2006**, *141*, 638–650. [[CrossRef](#)]
93. Belle, V.; Rouger, S.; Costanzo, S.; Liquiere, E.; Strancar, J.; Guigliarelli, B.; Fournel, A.; Longhi, S. Mapping alpha-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. *Proteins Struct. Funct. Bioinform.* **2008**, *73*, 973–988. [[CrossRef](#)] [[PubMed](#)]
94. Martinho, M.; Habchi, J.; El Habre, Z.; Nesme, L.; Guigliarelli, B.; Belle, V.; Longhi, S. Assessing induced folding within the intrinsically disordered C-terminal domain of the Henipavirus nucleoproteins by site directed spin labeling EPR spectroscopy. *J. Biomol. Struct. Dyn.* **2013**, *31*, 453–471. [[CrossRef](#)] [[PubMed](#)]
95. Tayeb-Fligelman, E.; Tabachnikov, O.; Moshe, A.; Goldshmidt-Tran, O.; Sawaya, M.R.; Coquelle, N.; Colletier, J.P.; Landau, M. The cytotoxic Staphylococcus aureus PSM α 3 reveals a cross- α amyloid-like fibril. *Science* **2017**, *355*, 831–833. [[CrossRef](#)] [[PubMed](#)]
96. Engelberg, Y.; Landau, M. The Human LL-37(17-29) antimicrobial peptide reveals a functional supramolecular structure. *Nat. Commun.* **2020**, *11*, 3894. [[CrossRef](#)] [[PubMed](#)]
97. Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. Prefolded structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* **2004**, *338*, 1015–1026. [[CrossRef](#)]

98. Lacy, E.R.; Filippov, I.; Lewis, W.S.; Otieno, S.; Xiao, L.; Weiss, S.; Hengst, L.; Kriwacki, R.W. p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat. Struct. Mol. Biol.* **2004**, *11*, 358–364. [[CrossRef](#)]
99. Sivakolundu, S.G.; Bashford, D.; Kriwacki, R.W. Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J. Mol. Biol.* **2005**, *353*, 1118–1128. [[CrossRef](#)] [[PubMed](#)]
100. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; del-Toro, N.; et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **2014**, *42*, D358–D363. [[CrossRef](#)]
101. Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A.P.; Kim, H.J.; Mittag, T.; Taylor, J.P. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **2015**, *163*, 123–133. [[CrossRef](#)]
102. Nevers, Q.; Albertini, A.A.; Lagaudrière-Gesbert, C.; Gaudin, Y. Negri bodies and other virus membrane-less replication compartments. *Biochim. Biophys. Acta Mol. Cell Res.* **2020**, *1867*, 118831. [[CrossRef](#)] [[PubMed](#)]
103. Brocca, S.; Grandori, R.; Longhi, S.; Uversky, V. Liquid-Liquid Phase Separation by Intrinsically Disordered Protein Regions of Viruses: Roles in Viral Life Cycle and Control of Virus-Host Interactions. *Int. J. Mol. Sci.* **2020**, *21*, 9045. [[CrossRef](#)]
104. Pesce, G.; Brocca, S.; Grandori, R.; Longhi, S.; Uversky, A.V. Droplets of life: Role of phase separation in virus replication and compartmentalization. In *Droplets of Life*; Uversky, A.V., Ed.; Elsevier: Amsterdam, The Netherlands, in press.
105. Nikolic, J.; Le Bars, R.; Lama, Z.; Scrima, N.; Lagaudrière-Gesbert, C.; Gaudin, Y.; Blondel, D. Negri bodies are viral factories with properties of liquid organelles. *Nat. Commun.* **2017**, *8*, 58. [[CrossRef](#)] [[PubMed](#)]
106. Heinrich, B.S.; Maliga, Z.; Stein, D.A.; Hyman, A.A.; Whelan, S.P.J. Phase Transitions Drive the Formation of Vesicular Stomatitis Virus Replication Compartments. *MBio* **2018**, *9*, e02290-17. [[CrossRef](#)]
107. Zhou, Y.; Su, J.M.; Samuel, C.E.; Ma, D. Measles Virus Forms Inclusion Bodies with Properties of Liquid Organelles. *J. Virol.* **2019**, *93*, e00948-19. [[CrossRef](#)] [[PubMed](#)]
108. Guseva, S.; Milles, S.; Ringkjøbing Jensen, M.; Salvi, N.; Kleman, J.; Maurin, D.; Ruigrok, R.W.; Blackledge, M. Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. *Sci. Adv.* **2020**, *6*, eaaz7095. [[CrossRef](#)]
109. Guseva, S.; Milles, S.; Jensen, M.R.; Schoehn, G.; Ruigrok, R.W.; Blackledge, M. Structure, dynamics and phase separation of measles virus RNA replication machinery. *Curr. Opin. Virol.* **2020**, *41*, 59–67. [[CrossRef](#)]
110. Etibor, T.A.; Yamauchi, Y.; Amorim, M.J. Liquid Biomolecular Condensates and Viral Lifecycles: Review and Perspectives. *Viruses* **2021**, *13*, 366. [[CrossRef](#)]
111. Dolnik, O.; Gerresheim, G.; Biedenkopf, N. New Perspectives on the Biogenesis of Viral Inclusion Bodies in Negative-Sense RNA Virus Infections. *Cells* **2021**, *10*, 1460. [[CrossRef](#)]
112. Galloux, M.; Risso-Ballester, J.; Richard, C.A.; Fix, J.; Rameix-Welti, M.A.; Eléouët, J.F. Minimal Elements Required for the Formation of Respiratory Syncytial Virus Cytoplasmic Inclusion Bodies In Vivo and In Vitro. *mBio* **2020**, *11*, e01202-20. [[CrossRef](#)]
113. Peng, Q.; Wang, L.; Qin, Z.; Wang, J.; Zheng, X.; Wei, L.; Zhang, X.; Zhang, X.; Liu, C.; Li, Z.; et al. Phase Separation of Epstein-Barr Virus EBNA2 and Its Coactivator EBNALP Controls Gene Expression. *J. Virol.* **2020**, *94*, e01771-19. [[CrossRef](#)] [[PubMed](#)]
114. Kim, H.J.; Kim, N.C.; Wang, Y.D.; Scarborough, E.A.; Moore, J.; Diaz, Z.; MacLea, K.S.; Freibaum, B.; Li, S.; Molliex, A.; et al. Mutations in prion-like domains in hnRNP2B1 and hnRNP1 cause multisystem proteinopathy and ALS. *Nature* **2013**, *495*, 467–473. [[CrossRef](#)] [[PubMed](#)]
115. White, J.P.; Lloyd, R.E. Regulation of stress granules in virus systems. *Trends Microbiol.* **2012**, *20*, 175–183. [[CrossRef](#)]
116. Tayeb-Fligelman, E.; Cheng, X.; Tai, C.; Bowler, J.T.; Griner, S.; Sawaya, M.R.; Seidler, P.M.; Jiang, Y.X.; Lu, J.; Rosenberg, G.M.; et al. Inhibition of amyloid formation of the Nucleoprotein of SARS-CoV-2. *bioRxiv* **2021**. [[CrossRef](#)]
117. Vidic, J.; Richard, C.A.; Péchoux, C.; Da Costa, B.; Bertho, N.; Mazerat, S.; Delmas, B.; Chevalier, C. Amyloid Assemblies of Influenza A Virus PB1-F2 Protein Damage Membrane and Induce Cytotoxicity. *J. Biol. Chem.* **2016**, *291*, 739–751. [[CrossRef](#)]
118. Pham, C.L.; Shanmugam, N.; Strange, M.; O’Carroll, A.; Brown, J.W.; Sierrecki, E.; Gambin, Y.; Steain, M.; Sunde, M. Viral M45 and necroptosis-associated proteins form heteromeric amyloid assemblies. *EMBO Rep.* **2019**, *20*, e46518. [[CrossRef](#)]
119. Steain, M.; Baker, M.; Pham, C.L.L.; Shanmugam, N.; Gambin, Y.; Sierrecki, E.; McSharry, B.P.; Avdic, S.; Slobedman, B.; Sunde, M.; et al. Varicella zoster virus encodes a viral decoy RHIM to inhibit cell death. *PLoS Pathog.* **2020**, *16*, e1008473. [[CrossRef](#)]
120. Shanmugam, N.; Baker, M.; Sanz-Hernandez, M.; Sierrecki, E.; Gambin, Y.; Steain, M.; Pham, C.L.L.; Sunde, M. Herpes simplex virus encoded ICP6 protein forms functional amyloid assemblies with necroptosis-associated host proteins. *Biophys. Chem.* **2021**, *269*, 106524. [[CrossRef](#)]
121. Le May, N.; Dubaele, S.; Proietti De Santis, L.; Billecocq, A.; Bouloy, M.; Egly, J.M. TFIIF transcription factor, a target for the Rift Valley hemorrhagic fever virus. *Cell* **2004**, *116*, 541–550. [[CrossRef](#)]
122. Léger, P.; Nachman, E.; Richter, K.; Tamietti, C.; Koch, J.; Burk, R.; Kummer, S.; Xin, Q.; Stanifer, M.; Bouloy, M.; et al. NSs amyloid formation is associated with the virulence of Rift Valley fever virus in mice. *Nat. Commun.* **2020**, *11*, 3281. [[CrossRef](#)]
123. Shiell, B.J.; Gardner, D.R.; Crameri, G.; Eaton, B.T.; Michalski, W.P. Sites of phosphorylation of P and V proteins from Hendra and Nipah viruses: Newly emerged members of Paramyxoviridae. *Virus Res.* **2003**, *92*, 55–65. [[CrossRef](#)]
124. Darling, A.L.; Uversky, V.N. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front. Genet.* **2018**, *9*, 158. [[CrossRef](#)]
125. Lieutaud, P.; Canard, B.; Longhi, S. MeDor: A metasever for predicting protein disorder. *BMC Genom.* **2008**, *9*, S25. [[CrossRef](#)]

126. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
127. Noguere, C.; Larsson, A.M.; Guyot, J.C.; Bignon, C. Fractional factorial approach combining 4 Escherichia coli strains, 3 culture media, 3 expression temperatures and 5 N-terminal fusion tags for screening the soluble expression of recombinant proteins. *Protein Expr. Purif.* **2012**, *84*, 204–213. [[CrossRef](#)]
128. Vincentelli, R.; Canaan, S.; Campanacci, V.; Valencia, C.; Maurin, D.; Frassinetti, F.; Scappucini-Calvo, L.; Bourne, Y.; Cambillau, C.; Bignon, C. High-throughput automated refolding screening of inclusion bodies. *Protein Sci.* **2004**, *13*, 2782–2792. [[CrossRef](#)] [[PubMed](#)]
129. 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. [[CrossRef](#)]
130. Hamdi, K.; Salladini, E.; O'Brien, D.P.; Brier, S.; Chenal, A.; Yacoubi, I.; Longhi, S. Structural disorder and induced folding within two cereal, ABA stress and ripening (ASR) proteins. *Sci. Rep.* **2017**, *7*, 15544. [[CrossRef](#)] [[PubMed](#)]
131. Yacoubi, I.; Hamdi, K.; Fourquet, P.; Bignon, C.; Longhi, S. Structural and Functional Characterization of the ABA-Water Deficit Stress Domain from Wheat and Barley: An Intrinsically Disordered Domain behind the Versatile Functions of the Plant Abscisic Acid, Stress and Ripening Protein Family. *Int. J. Mol. Sci.* **2021**, *22*, 2314. [[CrossRef](#)] [[PubMed](#)]
132. Whitmore, L.; Wallace, B.A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **2004**, *32*, W668–W673. [[CrossRef](#)]
133. 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. [[CrossRef](#)]
134. 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. [[CrossRef](#)] [[PubMed](#)]
135. Guinier, A. La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Ann. Phys.* **1939**, *12*, 161–237. [[CrossRef](#)]
136. Guinier, A.; Fournet, F. *Small angle Scattering of X-rays*; Wiley Interscience: New York, NY, USA, 1955. [[CrossRef](#)]
137. Svergun, D. Determination of the regularization parameters in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **1992**, *25*, 495–503. [[CrossRef](#)]
138. 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. [[CrossRef](#)]
139. 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. [[CrossRef](#)]
140. Tria, G.; Mertens, H.D.T.; Kachala, M.; Svergun, D. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* **2015**, *2*, 202–217. [[CrossRef](#)] [[PubMed](#)]
141. 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–D363. [[CrossRef](#)] [[PubMed](#)]
142. Lazar, T.; Martínez-Pérez, E.; Quaglia, F.; Hatos, A.; Chemes, L.B.; Iserte, J.A.; Méndez, N.A.; Garrone, N.A.; Saldaño, T.E.; Marchetti, J.; et al. PED in 2021: A major update of the protein ensemble database for intrinsically disordered proteins. *Nucleic Acids Res.* **2021**, *49*, D404–D411. [[CrossRef](#)]
143. Longhi, S.; Receveur-Brechot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S.; Canard, B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J. Biol. Chem.* **2003**, *278*, 18638–18648. [[CrossRef](#)] [[PubMed](#)]