Insights into the coiled-coil organization of the Hendra virus phosphoprotein from combined biochemical and SAXS studies

Matilde Beltrandia,b,1, David Blocquela,b,1, Jenny Erales a,b, Pascale Barbierc, Andrea Cavalli d,e,*, Sonia Longhi a,b,2,2

* Aix-Marseille University, Architecture et Fonction des Macromolécules Biologiques (AFMB) UMR 7257, 13288 Marseille, France
b CNRS, AFMB UMR 7257, 13288 Marseille, France
c Institute for Research in Biomedicine, Via Vincenzo Vela 6, 6500 Bellinzona, Switzerland
d Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

A R T I C L E   I N F O
Article history:
Received 20 September 2014
Returned to author for revisions 21 October 2014
Accepted 19 December 2014
Available online 29 January 2015

Keywords:
Henipavirus
Hendra virus
Phosphoprotein
PMD
P multimerization domain
Coiled-coil
Homotrimer
Small-angle X-ray scattering
Cross-linking

A B S T R A C T

Nipah and Hendra viruses are recently emerged paramyxoviruses belonging to the Henipavirus genus. The Henipavirus phosphoprotein (P) consists of a large intrinsically disordered domain and a C-terminal domain (PCT) containing alternating disordered and ordered regions. Among these latter is the P multimerization domain (PMD). Using biochemical, analytical ultracentrifugation and small-angle X-ray scattering (SAXS) studies, we show that Hendra virus (HeV) PMD forms an elongated coiled-coil homotrimer in solution, in agreement with our previous findings on Nipah virus (NiV) PMD. However, the orientation of the N-terminal region differs from that observed in solution for NiV PMD, consistent with the ability of this region to adopt different conformations. SAXS studies provided evidence for a trimeric organization also in the case of PCT, thus extending and strengthening our previous findings on PMD. The present results are discussed in light of conflicting reports in the literature pointing to a tetrameric organization of paramyxoviral P proteins.

Abbreviations: HeV, Hendra virus; NiV, Nipah virus; N, nucleoprotein; L, large protein; P, phosphoprotein; IDP, intrinsically disordered protein; IDR, intrinsically disordered region; PNT, P N-terminal domain; PCT, P C-terminal domain; PMD, P multimerization domain; XD, X domain of P; MeV, measles virus; SeV, Sendai virus; MuV, mumps virus; RDV, Rinderpest virus; RSV, respiratory syncytial virus; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; OD, optical density; IMAC, immobilized metal affinity chromatography; GF, gel filtration; AUC, analytical ultracentrifugation; SEC, size-exclusion chromatography; SDS-PAGE, sodium dodecyl sulphate polyacrylamide electrophoresis; Rs, Stokes radius; Rg, radius of gyration; MALDI-TOF, matrix-assisted laser desorption ionization/time of flight; CD, circular dichroism; TFE, 2,2,2-trifluoroethanol; MRE, mean residue ellipticity values per residue; SAB, suberic acid bis (N-hydroxy-succinimide ester); SAXS, small angle X-ray scattering; Tm, melting temperature

* Corresponding to: Institute for Research in Biomedicine, Via Vincenzo Vela 6, 6500 Bellinzona, Switzerland. Tel.: + 41 0 91 820 0303.
** Corresponding to: AFMB, UMR 7257 CNRS and Aix-Marseille University, 163, avenue de Luminy, Case 932, 13288 Marseille Cedex 09, France. Tel.: + 33 4 91 82 55 80; fax: + 33 4 91 26 67 20.
E-mail addresses: andrea.cavalli@irb.usi.ch (A. Cavalli), Sonia.Longhi@afmb.univ-mrs.fr (S. Longhi).
1 These authors have equally contributed to the work.

Introduction

The Hendra (HeV) and Nipah (NiV) viruses are recently emerged BSL4 pathogens gathered within the Henipavirus genus within the Paramyxoviridae family (Eaton et al., 2007; Wang et al., 2000). The newly identified Cedar virus has also been classified within this genus (Marsh et al., 2012). In henipaviruses, the genome is encapsidated by the nucleoprotein (N) within a helical nucleocapsid that is the substrate used by the polymerase for both transcription and replication. The viral polymerase is a complex made of the L protein and the P protein. The latter is an essential polymerase cofactor as it allows the recruitment of L onto the nucleocapsid template. Beyond its role in tethering the L protein, P is also supposed to serve as a chaperone for N, i.e. to prevent illegitimate self-assembly of N (see Albertini et al. (2005), Blocquel et al. (2012a), Lamb and Parks (2007) and Roux (2005) for reviews on paramyxovirus transcription and replication). Henipavirus N and P proteins were shown to interact with each other (Habchi et al., 2011), being able to form both homologous and heterologous N-P complexes (Blocquel et al., 2012b; Chan et al., 2004; Onn-Furutani et al., 2010).

Using both computational and experimental approaches, we have previously shown that Henipavirus P proteins possess a modular...
organization consisting of both ordered and intrinsically disordered regions (IDRs) (Habchi et al., 2010; Karlin et al., 2003). Intrinsically disordered proteins (IDPs) or IDRs are functional proteins/regions that lack highly populated uniform secondary and tertiary structure under physiological conditions of pH and salinity in the absence of a partner (for reviews on IDPs, see (Chouard, 2011; Dunker et al., 2008a, 2008b; Habchi et al., 2014; Turoverov et al., 2010; Uversky, 2010, 2013; Uversky and Dunker, 2010)). The Henipavirus P proteins contain in fact a spectacularly long N-terminal disordered region (PNT, aa 1–404 in HeV) (Habchi et al., 2010) and a C-terminal moiety (PCT, aa 405–707 in HeV) (Fig. 1). Contrary to PNT, which is common to both P and the auxiliary viral protein V, PCT is unique to the P protein. PCT has a modular organization being composed of alternating disordered and ordered regions: it embraces a predicted disordered region that partly overlaps the V ORF (aa 405–468 in HeV) referred to as spacer, a structured region (PMD, aa 469–578 in HeV) referred to as PMD for P multimerization domain), a disordered linker (aa 579–656 in HeV) and a globular region (aa 657–707 in HeV) referred to as X domain (XD) (Fig. 1) (Habchi et al., 2010). The crystal structure of HeV XD has been solved recently and was shown to consist of a triple α-helical bundle (Communie et al., 2013b), in agreement with previous spectroscopic and modeling studies (Habchi et al., 2011).

In Paramyxovirinae, sequence analyses predict a coiled-coil region within PMD (Habchi et al., 2010; Karlin et al., 2003) (see Fig. 1). The coiled-coil organization has been experimentally confirmed in the case of SeV (Tarbouriech et al., 2000b), rinderpest virus (RDV) (Rahaman et al., 2004), MeV (Blocquel et al., 2014; Communie et al., 2013a), MuV (Cox et al., 2013) and NIV (Blocquel et al., 2013; Bruhn-Johannsen et al., 2014). PMDs from SeV (Tarbouriech et al., 2000b) and MeV were shown to form a tetrameric coiled-coil (Blocquel et al., 2014; Communie et al., 2013a). The tetrameric coiled-coil organization of PMD has also been experimentally confirmed in the case of RDV (Rahaman et al., 2004), RSV (Llorente et al., 2006, 2008) and MuV (Cox et al., 2013). In this latter case, however, the tetramer was found to consist of two sets of parallel helices in opposite orientation, i.e. to be a dimer of two antiparallel coiled-coil dimers (Cox et al., 2013) in striking contrast with all Paramyxoviridae P proteins characterized so far that were all shown to possess a parallel organization.

Strikingly, studies focused on NIV PMD yielded different results depending on whether the protein was studied in solution or by X-ray crystallography. Indeed, several independent biochemical and biophysical approaches, including size-exclusion chromatography (SEC), SDS-PAGE, cross-linking, analytical ultracentrifugation and small-angle X ray scattering (SAXS), consistently converged to show that NIV PMD adopts a trimeric organization in solution (Blocquel et al., 2013). By contrast, the crystallographic structure of NIV PMD reported by Bruhn et al. revealed a tetrameric organization (Bruhn-Johannsen et al., 2014), thus raising interesting questions as to the origin of these discrepancies.

In order to solve this conundrum and to shed light onto the structural organization of Henipavirus phosphoproteins, we herein have undertaken the characterization of both HeV PMD and PCT regions. Using a combination of biochemical and biophysical approaches, we show that like in the case of NIV, HeV PMD forms a homotrimeric coiled-coil structure with an overall elongated shape. Notably, SAXS studies provided evidence for a trimeric organization also in the case of PCT, thus extending and strengthening conclusions based on PMD.

Results and discussion

Cloning, expression and purification of P constructs

The P gene fragments encoding PMD or PCT with a C-terminal histidine-tag were cloned into pDEST14. In both cases, the recombinant product was purified from the soluble fraction of the bacterial lysate (see insets in Fig. 2). The proteins were purified in two steps: IMAC and gel filtration (see insets in Fig. 2). Both PMD and PCT were found to migrate in SDS-PAGE with an apparent molecular mass higher than that expected from their amino acid sequence: indeed PMD was found to have an apparent molecular mass close to 17 kDa (expected mass of 13,524 Da) and PCT a mass of approximately 40 kDa (expected mass of 34,913 Da) (see insets in Fig. 2). In both cases, however, the identity of the purified protein was confirmed by mass spectrometry analysis of the tryptic fragments obtained after digestion of the purified protein band excised from SDS-polyacrylamide gels. This aberrant electrophoretic migration may reflect either an unusual amino acid composition or a polymorphism in the expression system. Further studies are needed to clarify the origin of the molecular mass discrepancy for these proteins.
acid sequence and/or composition, or the presence of predicted disordered regions.

Intriguingly, and as already observed for the cognate NiV PMD (Blocquel et al., 2013), when HeV PMD was analyzed by SDS-PAGE, an additional band of approximately 34 kDa was observed (Fig. 3A). This band was no longer detected when the protein was subjected to prolonged thermal denaturation (Fig. 3A). Furthermore, this additional band was found to be resistant to the addition of 20 mM DTT (Fig. 3A), arguing for a very stable oligomer that resists under denaturing conditions (i.e. in spite of the presence of SDS) and that is not mediated by disulfide bridges.

**Size-exclusion chromatography and cross-linking studies**

SEC experiments yielded an apparent molecular mass of 43 kDa for HeV PMD, with the elution profile being independent from protein concentration (Fig. 2A). This value is not consistent with a monomeric form of the protein (13.5 kDa) and is rather close to the value expected for a trimer (40.5 kDa). Further support to this conclusion comes from the observation that HeV PMD has the same elution profile as NiV PMD (Fig. 2A), a protein that has the same molecular mass as HeV PMD and has already been shown to be an elongated trimer (Blocquel et al., 2013).

---

**Fig. 2.** Gel filtration elution profile and purification of HeV PMD (A) and PCT (B). The arrows point the peak containing the protein of interest. The corresponding apparent molecular mass (MM) and Stokes radius (RS) are indicated. In panel A, the elution profile of NiV PMD (blue) is overlaid. Insets: Coomassie blue staining of an SDS-PAGE showing the bacterial lysate (total fraction, TF), the clarified supernatant (soluble fraction, SN), the non-retained fraction (?-th), the eluent from Immobilized Metal Affinity Chromatography (ENi) and the eluent from Gel Filtration (EGF). M: molecular mass markers (kDa).

**Fig. 3.** Effect of thermal denaturation and cross-linking experiments. (A) Coomassie stained 15% SDS-PAGE showing the electrophoretic migration of purified HeV PMD as obtained either with or without a prolonged thermal denaturation at 95 °C. The migration of a HeV PMD sample not subjected to thermal denaturation and containing 20 mM DTT is also shown. (B) Coomassie stained 15% SDS-PAGE showing the results of cross-linking with SAB (see Materials and methods). The concentration of SAB (μM) is indicated above the gel. Samples in panel B, were thermally denatured for 5 minutes prior to electrophoretic migration.
The Stokes radius ($R_s$) of HeV PMD, as inferred from its apparent molecular mass (Uversky, 2002), is 28.5 Å, a value close to that expected for a globular trimer (27.9 Å) (Uversky, 2002). The slightly higher $R_s$ value with respect to the value expected for a globular trimer could reflect an elongated form of the protein, as already observed for SeV, RDV and NiV PMDs (Blockuel et al., 2013; Rahaman et al., 2004; Tarbouriech et al., 2000a).

In the case of PCT, SEC analyses showed that the protein was eluted in the void volume (data not shown) unless 0.2 M arginine was added to both the protein and the elution buffer (Fig. 2B). In this latter case, its elution profile was borderline with respect to the exclusion volume allowing only a rough estimation of the molecular mass. The estimated value (207 kDa) is twice the value expected for a trimer and 1.5 times greater than the value expected for a tetramer, which can be indicative either of a higher oligomeric state or of an elongated shape combined to the presence of disordered regions that are known to impart an apparently higher molecular mass during SEC (Uversky, 2002). The corresponding $R_s$ is 51.0 Å, a value much higher than those expected for a trimeric or tetrameric globular form of the protein (39.6 Å and 44.1 Å, respectively), again suggesting that the protein either forms higher-order oligomers and/or has an extended shape.

In order to shed light into the oligomeric state of PMD and PCT, we performed cross-linking experiments using SAB, a bifunctional reagent of fixed size (13.1 Å) that reacts with lysines. Experiments carried out with PCT failed to provide insights, as addition of cross-linker resulted in the formation of aggregates of high molecular mass that could not enter into the gel (data not shown). In the case of PMD, the addition of increasing amounts of SAB triggers the progressive accumulation of two additional bands, the apparent molecular mass of which is close to 34 and 55 kDa (Fig. 3B). Taking into account the fact that the monomeric form aberrantly migrates as a 17-kDa protein, these two bands, which correspond to twice

![Fig. 4. Small-angle X-ray scattering experiments of HeV PMD. (A) Experimental SAXS data recorded for $q$ values up to 3.5 nm$^{-1}$. The curves obtained for three protein concentrations (0.77 g/L, blue; 1.36 g/L, red; and 3.31 g/L, light blue) are represented after correction for concentration. (B) Representation of the Guinier plot for the protein at 3.31 g/L. (C) Pair distance distribution, $P(r)$, function of the data for the 3.31 g/L concentration. (D) Kratky plot. (E) Three views of the ab initio envelope calculated with DAMAVER (Volkov and Svergun, 2003). 20 DAMMIF calculations (Franke and Svergun, 2009) were performed and averaged with DAMAVER to produce the average and filtered shape shown in light blue. The structure of a trimeric coiled-coil model of HeV PMD is shown, with the three chains being displayed in three different colors. Docking of the model in the envelope was done manually using the program Chimera. The structural model fits well in the density of the SAXS-derived model.](http://example.com/Figure4.png)
Far-UV CD studies

In order to confirm that HeV PMD adopts a coiled-coil conformation like in the case of its most related paramyxoviral members, we used far-UV spectroscopy. The spectrum of PMD is indicative of a folded protein with an estimated α-helical content of 61% (see inset in Supplementary Fig. S1A). The ratio of the ellipticities at 222 and 208 nm (θ222/208) is greater than 1.0 (see inset in Supplementary Fig. S1A), a property indicative of the presence of interacting helices and already observed also in the case of PMDs from SeV (Tarbouriech et al., 2000a), RDV (Rahaman et al., 2004), MeV (Blocquel et al., 2014) and NIV (Blocquel et al., 2013). Upon addition of 70% of TFE the θ222/208 ratio drops below 1 (Supplementary Fig. S1A). Taking into account the fact that high TFE concentrations are known to disrupt tertiary structure and quaternary structure and to stabilize secondary structure (Lau et al., 1984), these results indicate that HeV PMD forms oligomers through coiled-coil interactions.

To investigate the thermal stability of PMD oligomers, we recorded the CD spectrum at 100 °C along with that obtained upon a stepwise cooling down to 20 °C (Supplementary Fig. S1A). The good superimposition between the initial spectrum and the spectrum obtained after cooling points out the reversibility of thermal unfolding and argues for refolding of the main structural motif (i.e. coiled-coil) (Supplementary Fig. S1A). We also monitored the mean residue ellipticity at 222 nm (MRE222), indicative of the α-helical content, at increasing temperatures and plotted it as a function of temperature (Supplementary Fig. S1B). The unfolding profile shows a single cooperative transition with an inflection point at approximately 65 °C (Supplementary Fig. S1B). Analysis of the thermal unfolding profile yielded an apparent melting temperature (Tm) of 66 °C, a value lower than that determined for NIV PMD (76 °C) (Blocquel et al., 2013).

These results support the conclusion that HeV PMD forms a coiled-coil oligomer that features a relatively high thermal stability, in line with previous observations on related PMDs (Blocquel et al., 2013, 2014; Communie et al., 2013a; Llorente et al., 2008).

The far-UV CD spectrum of PCT is indicative of a protein containing a higher amount of disorder and a lower α-helical content as compared to PMD (Supplementary Fig. S1C and inset), in agreement with the presence of predicted disordered regions. In the case of PCT, the θ222/208 ratio is lower than 1.0, a value reflecting the absence of interacting helices. In addition, and by contrast with HeV PMD, the addition of TFE does not significantly affect this ratio (see inset of Supplementary Fig. S1C), again suggesting the lack of interacting helices through a quaternary structure organization. The lower than 1.0 value of θ222/208 probably reflects the contribution of disordered regions, which are typified by very low ellipticities values at 222 nm (Woody, 2010). Thermal denaturation experiments similar to those described above, yielded comparable results in that thermal unfolding was found to be reversible, thus ruling out possible heat-induced protein aggregation and arguing for proper refolding.

<table>
<thead>
<tr>
<th>Data-collection parameters</th>
<th>PMD</th>
<th>PCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detector</td>
<td>Pilatus (1 M)</td>
<td>Pilatus (1 M)</td>
</tr>
<tr>
<td>Beam geometry</td>
<td>Bending magnet (BM29)</td>
<td>Bending magnet (BM29)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.992</td>
<td>0.992</td>
</tr>
<tr>
<td>q Range (Å⁻¹)</td>
<td>0.028–4.525</td>
<td>0.028–4.525</td>
</tr>
<tr>
<td>Exposure time (s)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Concentration range (g/L)</td>
<td>0.77–3.31</td>
<td>0.44–2.48</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Structural parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I(0) (cm⁻¹) from P(t)</td>
<td>42.0 ± 0.1</td>
<td>115.0 ± 0.2</td>
</tr>
<tr>
<td>I(0) (cm⁻¹) from Guinier</td>
<td>41.6 ± 0.3</td>
<td>114.5 ± 0.3</td>
</tr>
<tr>
<td>Rg (Å) (from P(0))</td>
<td>39.1 ± 0.2</td>
<td>86.7 ± 0.5</td>
</tr>
<tr>
<td>Rg (Å) (from Guinier)</td>
<td>37.2 ± 0.1</td>
<td>87.0 ± 0.1</td>
</tr>
<tr>
<td>Dmax (Å)</td>
<td>132</td>
<td>302</td>
</tr>
<tr>
<td>Molecular-mass determination (kDa)</td>
<td>41.0</td>
<td>114</td>
</tr>
<tr>
<td>Calculated MM from sequence</td>
<td>40.5</td>
<td>104.7</td>
</tr>
</tbody>
</table>

Software employed

| Primary data reduction     | PRIMUS | PRIMUS |
| Data processing            | GNOM | GNOM |
| Ab initio analysis          | DAMMIF | DAMMIF |
| Validation and averaging    | DAMAVER | DAMAVER |
| Validation of structural models | CRYSOLO/FOX | CRYSOLO/FOX |
of the structural motifs of PCT (i.e. PMD and XD). Plotting the MRF$_{222}$ as a function of temperature yielded a typical two-state profile, indicative of a single cooperative transition with an apparent $T_m$ of 68.6 °C (Supplementary Fig. S1D). These results support the conclusion that HeV PCT encompasses both disordered and structured regions adopting a prevalently $\alpha$-helical conformation. Its behavior in terms of reversibility of thermal unfolding and high $T_m$ are reminiscent of that observed in the case of HeV PMD, suggesting that the folding/unfolding behavior of PCT is dominated by PMD and that this latter maintains its structural organization in the context of PCT.

SAXS studies of PMD

In order to achieve insights into the shape and oligomeric state of HeV PMD in solution, we performed SAXS studies. The shapes of the SAXS curves (Fig. 4A) and the Guinier plots obtained (Fig. 4B) are independent of protein concentration, indicating the absence of significant aggregation. Each curve can be well approximated by a straight line in the Guinier region ($qR_g < 1.0$). The slope gives the value of the radius of gyration, $R_g$, while the intercept of the straight line gives the $I(0)$ which is proportional to the molecular mass of the scatterer. Guinier analysis in the low $q$ region gave an $R_g$ of 37.2 ± 0.1 Å, which is in good agreement with the value of 39.1 ± 0.2 determined from the pair distance distribution function $P(r)$ (Tables 1 and 2). The molecular mass determined from the extrapolated scattering intensity at zero angle $I(0)$ is 41.0 kDa, a value in very good agreement with the molecular mass expected for a trimeric form (40.5 kDa) (Table 1). Note that the reliability of this estimation is supported by the precision with which we could estimate protein concentration. Indeed, accurate determination of molecular mass by SAXS analysis is known to critically depend on accurate estimation of particle concentration (Mylonas and Svergun, 2007). Herein, our estimation of HeV PMD concentration based on the theoretical absorption coefficient was found to be in very good agreement with the actual protein concentration determined through analysis of the amino acid composition (see Materials and methods).

The experimentally observed $R_g$ value is 1.7 times larger than that expected (22.1 Å) for a globular protein with an $R_g$ equal to that experimentally observed for HeV PMD (28.5 Å) (Wilkins et al., 1999). It is even much higher than the value expected (17.3 Å) for a sphere with an $R_g$ of 22.4 Å, as determined from the volume of a sphere with 117 × 3 residues (see Materials and methods). The strong discrepancy between the experimentally observed $R_g$ and the value expected for a globular/spherical form indicates that

<table>
<thead>
<tr>
<th>Protein concentration (g/L)</th>
<th>$R_g$ (Å) (Guinier)</th>
<th>$I(0)$ (Guinier) (cm$^{-1}$)</th>
<th>$D_{max}$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMD 0.77</td>
<td>36.5 ± 0.3</td>
<td>41.1</td>
<td>127.6</td>
</tr>
<tr>
<td>PMD 1.36</td>
<td>37.0 ± 0.1</td>
<td>41.8</td>
<td>130.8</td>
</tr>
<tr>
<td>PMD 3.31</td>
<td>37.2 ± 0.1</td>
<td>41.6</td>
<td>132.3</td>
</tr>
<tr>
<td>PCT 0.44</td>
<td>86.4 ± 0.5</td>
<td>114.4</td>
<td>300.8</td>
</tr>
<tr>
<td>PCT 1.25</td>
<td>86.8 ± 0.2</td>
<td>114.7</td>
<td>301.1</td>
</tr>
<tr>
<td>PCT 2.48</td>
<td>87.0 ± 0.1</td>
<td>114.5</td>
<td>301.8</td>
</tr>
</tbody>
</table>

Fig. 5. Comparison between the experimental (green) and calculated (red) HeV PMD scattering curve obtained either with no forced symmetry (P1) or by imposing a P2, P3 or P4 symmetry. The calculated curve corresponds to the average profile provided by DAMMIF. Experimental data presented here were recorded at 3.31 g/L. Values in the insets show the $\chi$ values calculated for the comparison between the predicted SAXS profiles and the experimental data as obtained at different protein concentrations.
the overall structure of HeV PMD is very elongated. The higher than expected \( R_g \) confirms that the higher apparent molecular mass observed upon SEC does indeed reflect an elongated trimer, as already observed for the cognate NiV PMD (Blocquel et al., 2013).

The distances distribution function inferred from the scattering curve of HeV PMD exhibits a maximum at 20 Å, a shoulder at 100 Å and a long tail up to 132 Å (\( D_{\text{max}} \)) typical of an elongated object (Fig. 4C). The \( R_g \) of a thin rod of length \( L = 132 \) Å is 38.1 Å (see Materials and methods), a value in very good agreement with the values estimated from the Guinier plot (37.2 Å) and from the pair distance distribution function (39.1 Å).

The Kratky plot presents a maximum at \( q \approx 0.83 \) nm\(^{-1} \) and a flat region for \( q > 1.75 \) nm\(^{-1} \) (Fig. 4D). The shape of the plot is indicative of a structured protein with possibly an at least partly disordered appendage, as judged from the bell-shape nature of the curve that displays a clear maximum and from the presence of the flat region.

Next, we employed the program DAMMIF to carry out ab initio shape reconstruction from the SAXS data. To this purpose, we used the scattering data obtained at the highest concentration so as to achieve maximal resolution. Several series of independent runs were carried out either with no forced symmetry or by imposing \( P2, P3 \) or \( P4 \) symmetries. Within each symmetry class, models were very reproducible with an average normalized spatial discrepancy (NSD) well below 1.0, indicating structurally similar solutions (Supplementary Fig. S2). The lowest average NSD values were obtained either with no imposed symmetry (NSD = 0.58 ± 0.02) or by imposing a \( P3 \) symmetry (NSD = 0.62 ± 0.64). All generated bead models appeared as elongated cylinders, compatible with a coiled-coil structure (Supplementary Fig. S3 and data not shown). Supplementary Fig. S4 shows, for each symmetry class, the distribution of \( \chi \) values for the 20 models. For each symmetry class, the models resulting from 20 independent DAMMIF runs were superimposed using the DAMAVER suite. After having rejected three and two outliers (see Materials and methods) within the \( P2 \) and \( P3 \) symmetry class, respectively, the program built the final average filtered models that were all found to have an elongated shape (Supplementary Fig. S5). Best results were obtained with the \( P3 \) symmetry, as judged from the DAMMIF average \( \chi \) parameter (Fig. 5).

Of note, best average \( \chi \) values were systematically obtained by imposing a \( P3 \) symmetry even at the intermediate and low protein concentration (see insets in Fig. 5). Global statistical analysis of the obtained \( \chi \) values using Bayesian-Factors (see Materials and methods) showed that differences in \( \chi \) values are meaningful (Supplementary Table S1), thereby supporting the conclusion that the \( P3 \) symmetry is the one that best fits the experimental data. As shown in Fig. 4E, the average filtered model generated by DAMAVER by imposing the \( P3 \) symmetry has an elongated dumbbell shape.

Interestingly, the shape of the HeV PMD envelope is somehow different from that of NiV PMD (Blocquel et al., 2013). Another

![Fig. 6. Comparison of experimental (red) and predicted (green) SAXS data generated by CRYSOL for a dimeric, trimeric and tetrameric model of HeV PMD with the “head” either in the “down” (A) or “up” conformation (B). Experimental data presented here were recorded at 3.31 g/L. Values in the insets show the \( \chi \) values calculated for the comparison between the predicted SAXS profiles of the models and the experimental data as obtained at different protein concentrations (0.77, 1.36 and 3.31 g/L). In all panels models are shown embedded in the ab initio SAXS envelope. Docking of the models in the envelope was done manually using the program Chimera.](image-url)
unexpected difference between the two PMDs resides in their overall dimensions: indeed, although the two domains differ in length by only one residue, HeV PMD was found to have an $R_g$ and a $D_{max}$ much smaller than those of its NiV counterpart (compare 37.2 Å and 132 Å, respectively, to 50.8 Å and 182 Å) (Blocquel et al., 2013). Taking into account the high sequence identity (91%) and the highly similar size of the two P multimerization domains, these results indicate that the amino acid sequence does not uniquely dictate the structure adopted in solution, thereby leading to unexpected differences in the structural parameters. These structural discrepancies between the envelopes of NiV and HeV PMD may reflect the ability of Henipavirus PMDs to undergo conformational changes resulting in forms of different lengths.

Similarly to what we did in the case of NiV PMD (Blocquel et al., 2013), we generated various structural models of HeV PMD differing in their oligomeric state (see Materials and methods) and assessed how they fit into the SAXS envelope. As shown in Fig. 4E, the trimeric model can easily be accommodated in the SAXS envelope. To quantify the extent to which the different models fit the experimental scattering profiles, we used the program CRYSTOL (Svergun et al., 1995), which calculates a theoretical SAXS profile from each model and then compares it to the experimental SAXS profile (Fig. 6A). As we used models and not crystal structures, the best fit was observed in the lower values of $q$ (0–1.5 nm$^{-1}$), which represent the global shape of HeV PMD, whatever the model used (see Fig. 6A). At the lowest and highest protein concentrations, the calculated scattering profile of the trimeric model was found to fit better the experimental profile as compared to the other models, especially in the low $q$ values, thereby yielding the lowest $\chi^2$ value (see insets in Fig. 6A). At the intermediate protein concentration, a slightly lower $\chi^2$ value was obtained for the dimeric form (see insets in Fig. 6A). Global statistical analysis of the obtained $\chi^2$ values using Bayesian-Factors (see Materials and methods) unambiguously showed that the trimeric model is much more likely than both the dimeric and tetrameric models (Supplementary Table S2). Altogether these results indicate that HeV PMD adopts a trimeric organization in solution, as already observed for its close NiV relative.

Interestingly, the N-terminal “head” region is oriented differently in our previous NiV PMD model (hereafter referred to as “up” conformation), being solvent exposed and not packed back onto the coiled-coil as in the models herein described (Blocquel et al., 2013). We thus reasoned that a different orientation of the head could be responsible for the observed differences in the length of the NiV and HeV PMD envelopes. We therefore generated models of a dimeric, trimeric and tetrameric HeV PMD form in which the head is in the “up” configuration (see Materials and methods). We then assessed how these models fit the experimental SAXS profile using CRYSTOL (Fig. 6B). At the lowest and highest protein concentrations, the lowest $\chi^2$ values were obtained for the trimer, while at the intermediate concentration a lower $\chi^2$ value was obtained for the dimer (Fig. 6B). Global statistical analysis performed on all $\chi^2$ values allowed us to conclude that the trimer is the model that is in best agreement with experimental data (see Supplementary Table S2). Interestingly, comparison of the Chi values obtained with the “head-up” and “head-down” trimer clearly designates the latter as the best fitting model (Fig. 6 and Supplementary Table S2).

Altogether, these data support the ability of Henipavirus PMDs to adopt different conformations differing in the orientation of the head and hence in their length. Whether the “up” and “down” conformations are unique to NiV and HeV, respectively, or whether both domains are able to adopt an equilibrium between the two forms remains to be established.

Sedimentation analysis

In view of gaining additional experimental support for a trimeric organization of HeV PMD in solution, we performed sedimentation equilibrium experiments (Fig. 7). Sedimentation equilibrium was performed with various loading protein concentrations (0.6, 1.6 and 2.1 g/L) at 10,000, 12,000, 17,000 and 40,000 rpm at 4 °C. Analysis of the three concentrations at 280 and 257 nm showed a best residuals distribution for a single species model. Multispeed global analysis for each concentration at the two wavelengths lead to a mean molecular mass of 43,500 ± 600 Da, a value close to that expected for a trimer. Equilibrium sedimentation experiments thus confirmed that HeV PMD adopts a trimeric conformation in solution.

SAXS studies of PCT

We next studied PCT by SAXS. The scattering profiles (Fig. 8A) and the obtained Guinier plots (Fig. 8B) were found to be concentration-independent in the 0.44–2.48 g/L range provided that 5 mM DTT was added. Above this concentration threshold however, the experimental profiles indicated at least partial aggregation of the protein (data not shown). We therefore used only the data obtained at concentrations up to this threshold value. Under these conditions, all the scattering curves are linear in the Guinier region. Guinier analysis in the low $q$ region gave an $R_g$ of 87.0 ± 0.1 Å, a value close to that obtained from the pair distance distribution function (86.7 ± 0.5 Å) (Tables 1 and 2). The molecular mass deduced from $I(0)$ is 114 kDa (Table 1). Although this value slightly exceeds the molecular mass expected for a trimeric form (104.7 kDa), it remains nevertheless indicative of a trimeric organization of the protein. In this case, like in the case of PMD, the experimentally observed $R_g$ value is much larger (3.6 times) than the value expected (24.0 Å) for a sphere with an $R_g$ of 31.0 Å, as determined from the volume of a sphere with 310 × 3 residues (see Materials and methods). In this case, the discrepancy between the observed and the expected $R_g$ is even more pronounced than in the case of PMD. The finding that the actual $R_g$
**Fig. 8.** Small-angle X-ray scattering experiments of HeV PCT. (A) Experimental SAXS data recorded for q values up to 3.5 nm\(^{-1}\). The curves obtained for three protein concentrations (0.44 g/L, light blue; 1.25 g/L, magenta; and 2.48 g/L, yellow) are represented after correction for concentration. (B) Representation of the Guinier plot for the protein at 2.48 g/L. (C) Pair distance distribution, \(P(r)\), function of the data for the 2.48 g/L concentration. (D) Kratky plot. (E) Three views of the \textit{ab initio} envelope calculated with DAMAVER (Volkov and Svergun, 2003). 20 DAMMIF calculations (Franke and Svergun, 2005) were performed and averaged with DAMAVER to produce the average and filtered shape shown in blue.
largely exceeds that expected for a sphere reflects the highly elongated nature of the protein, a property arising from the elongated shape of PMD combined with the presence of disordered regions. The very high $R_g$ value also explains the spectacularly high apparent molecular mass observed in SEC studies.

The distance distribution function of PCT displays a maximum at 90 Å and a $D_{max}$ value of 302 Å typical of an elongated object (Fig. 8C). The $R_g$ of a thin rod of length $L=302$ Å is 87.2 Å (see Materials and methods), a value in very good agreement with the values estimated from the Guinier plot (87.0 Å) and from the pair distance distribution (86.7 Å). The shape of the Kratky plot, with a maximum at $q \approx 0.27$ nm$^{-1}$ and a flat region for $q > 0.83$ nm$^{-1}$, is indicative of a mixed protein consisting of a structured and a disordered moiety (Fig. 8D), in agreement with the far-UV CD data.

For ab initio shape reconstruction, we used the scattering data obtained at the highest concentration so as to achieve maximal resolution, and proceeded as described for PMD. For each symmetry class, models were found to be very reproducible with an average NSD below 1.0, indicating structurally similar solutions (Supplementary Fig. S6 and data not shown). All generated bead models have an elongated shape (data not shown). Like in the case of PMD, best results were obtained with the P3 symmetry (data not shown), thus providing additional support for a trimeric organization of the protein. Assuming a P3 symmetry, the models resulting from 20 independent DAMMIF runs were superimposed using the DAMAVER suite, which yielded an average NSD of 0.66 ± 0.23 after having discarded four outliers. The average filtered model generated by DAMAVER has an elongated shape (Fig. 8E).

Altogether, these results confirm and reinforce conclusions based on studies carried out on HeV PMD. They therefore afford additional reliability to the conclusion that the HeV P protein adopts a trimeric organization in solution, like its NIV counterpart.

Conclusions

The results herein presented, along with our previously reported results on NIV PMD (Blocquel et al., 2013), provide experimental evidence for a trimeric organization of Henipavirus phosphoproteins in solution. On the other hand, the crystallographic structure of NIV PMD reported by Bruhn et al. unambiguously shows a tetrameric organization within the crystal (Bruhn-Johannsen et al., 2014).

It is conceivable that Henipavirus P proteins can form both trimers and tetramers depending on the conditions. That coiled-coils are able to modulate their oligomeric state according to the physico-chemical conditions (pH, temperature) or depending on whether they are located inside or outside the cell, has already been reported (Dutta et al., 2001; Lupas and Gruber, 2005). Of even more interest, the GCN4 leucine-zipper domain was shown to adopt different oligomeric states depending on the crystallization conditions, implying that the amino acid sequence does not specify a unique oligomeric state (Oshaben et al., 2012). It is also worthy to emphasize that conflicting experimental evidence are not unique to NIV PMD: indeed SeV PMD had also been shown to form trimers in solution (Curran, 1998; Curran et al., 1995) and to adopt a tetrameric coiled-coil conformation in the crystal (Tarbouriech et al., 2000b). The experimental evidence pointing to a trimeric form of SeV P have been set aside perhaps too rapidly in light of the crystallographic data pointing to a tetrameric organization. However, the finding that both SeV and NIV PMD can form trimers in solution and tetramers in the crystal may reflect their intrinsic ability to adopt different oligomeric states that could be related to different functional forms of the P protein and to the different complexes (i.e. N–P, N–P, P–L) that this latter can form within infected cells. Likewise, the herein proposed ability of the N-terminal “head” region of Henipavirus PMD to adopt different conformations, might play a functional role by imparting specificity in molecular interactivity.

In the same vein, structural comparison among the different crystallographic structures of MeV PMD solved so far unveiled unexpected structural variations (Blocquel et al., 2014; Communie et al., 2013a). Although all the structures have a tetrameric coiled-coil organization, structural comparison unveiled considerable differences not only in the quaternary structure but also in the extent of disorder within the C-terminal region of the coiled-coil (Blocquel et al., 2014). We have previously proposed that the unexpected plasticity and flexibility of MeV PMD could be the first hint of the existence of different functional forms of the P protein reflecting its multifunctional nature and pivotal role in the replicative cycle.

In conclusion, the ability of SeV and Henipavirus P proteins to adopt different oligomeric states, together with the ability of MeV PMD to dynamically sample different forms differing in the degree of compaction and in the extent of disorder, might be the basis for the ability of the P protein to form different complexes critical for transcription and replication, with conformational changes possibly dictating the ability to form a transcriptase versus a replicase complex.

Funding

This work was carried out with the financial support of the Agence Nationale de la Recherche, specific programs “Physico-Chimie du Vivant” (ANR-08-PCVI-0020-01) and “ASTRID”. ANR-11-ASTR-003-01. D.B. was supported by a joint doctoral fellowship from the Direction Générale de l’Armement (DGA) and the CNRS, M.B. was previously supported by an Erasmus Master2 fellowship from the University of Milan and is presently supported by a PhD fellowship from the Italo-French University. J.E. is supported by a post-doctoral fellowship by the Fondation pour la Recherche Médicale (FRM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Materials and methods

Cloning of the PMD and PCT coding regions

The HeV PMD and PCT constructs, encoding respectively residues 469–578 and 404–707 of the HeV P protein (SwissProt accession number O55778.2), all with an hexahistidine tag fused to their C-terminus, were obtained by polymerase chain reaction (PCR). PCR was performed using a synthetic HeV P gene (GenScript), optimized for the expression in Escherichia coli, as template, and Pfu polymerase (Promega). Primers (Operon) were designed to introduce an AttB1 and AttB2 site at the 5’ and 3’ ends, respectively, and to amplify the desired part of the P ORF with a fragment encoding a C-terminal hexahistidine tag. The rationale for the choice of the tag position was to avoid purification of truncated forms arising from possible abortive translation. After digestion with DpnI (New England Biolabs) to remove the methylated DNA template, the PCR products were purified (PCR Purification Kit, Qiagen) and cloned into the pDEST14 vector (Invitrogen) using the Gateway® recombination system (Invitrogen). This vector drives the expression of recombinant products under the control of the T7 promoter. Selection and amplification of the construct was carried out using CaCl₂-competent
E. coli TAM1 cells (Active Motif). The sequences of the coding regions of the constructs were checked by sequencing (GATC Biotech) and found to conform to expectations.

Expression and purification of PMD and PCT constructs

The E. coli strain Rosetta [DE3] pLysS (Novagen) was used for expression of all the P constructs. The pLysS (Novagen) plasmid carries the lysozyme gene, thus allowing tight regulation of the expression of the recombinant gene, as well as a facilitated lysis. Cultures were grown overnight to saturation in LB medium containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. An aliquot of the overnight culture was diluted 1/25 in LB medium and grown at 37 °C. At an OD600 of 0.7, isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of either 0.5 mM (PMD) or 0.2 mM (PCT), and the cells were subsequently grown at 37 °C for 4 h. The induced cells were harvested, washed and collected by centrifugation (5000g, 10 min). The resulting bacterial pellets were frozen at −20 °C.

Bacterial pellets, irrespective of whether they contained PMD or PCT, were resuspended in 5 volumes (v/w) of buffer A (50 mM Tris/HCl pH 8, 1 M NaCl, 10 mM imidazole) supplemented with 0.1 mg/mL lysozyme, 10 µg/mL DNase I, 20 mM MgSO4 and protease inhibitor cocktail (Sigma) (one tablet for a 2 L-culture). After a 20-min incubation with gentle agitation at 4 °C, the cells were disrupted by sonication (using a 750 W sonicator and 5 cycles of 30 s each at 60% power output). The lysates were clarified by centrifugation at 30,000g for 30 min and then purified by immobilized metal affinity chromatography (IMAC). The clarified supernatant from a 1 L-culture was injected onto a 5-mL HiTrap FF column (GE, Healthcare), previously equilibrated in buffer A supplemented with 1 M NaCl. Elution was carried out using a gradient of imidazole (20–500 mM) in buffer A supplemented with 1 M NaCl. Eluents were analyzed by SDS-PAGE for the presence of the desired product. The fractions containing the desired recombinant protein were combined and loaded onto a Superdex 200 HR 16/60 column (GE, Healthcare). PMD was eluted in 50 mM Hepes pH 8, and 300 mM NaCl, while PCT was eluted in 30 mM Tris/HCl pH 8, 5 mM EDTA, 300 mM NaCl, 1 mM PMSF (phenylmethylsulfonylfluoride), and glycerol 10% supplemented with 200 mM arginine. Proteins were concentrated using Centricon Plus-20 (molecular cutoff of 3000 Da for PMD and of 30,000 Da for PCT)—Superdex 200 HR 16/60 column (GE, Healthcare). PMD was eluted in 50 mM Hepes pH 8, and 300 mM NaCl, while PCT was eluted in 30 mM Tris/HCl pH 8, 5 mM EDTA, 300 mM NaCl, 1 mM PMSF (phenylmethylsulfonylfluoride), and glycerol 10% supplemented with 200 mM arginine. Proteins were concentrated using Centricon Plus-20 (molecular cutoff of 3000 Da for PMD and of 30,000 Da for PCT) (Millipore) and then stored at −20 °C.

All purification steps, except for IMAC and gel filtrations, were carried out at 4 °C. Apparent molecular mass of proteins eluted from the gel filtration column was deduced from a calibration carried out with ferritin (440 kDa), catalase (232 kDa), BSA (132 and 66 kDa), Tobacco Etch virus protease (27 kDa), and lysozyme (14 kDa).

Protein concentrations were either calculated using the theoretical absorption coefficients at 280 nm, as provided by the program ProtParam at the EXPASY server (http://www.exASY.ch/tools), or estimated via quantitative analysis of the amino acid composition. In those analyses, PMD samples were dried and hydrolyzed at 110 °C in constant-boiling HCl containing 1% (v/v) phenol for 24 h under reduced pressure and in the absence of oxygen. Amino acids were analyzed on a model Biochrom 30 amino acid analyzer, with the constant-boiling HCl containing 1% (v/v) phenol for 24 h under reduced pressure and in the absence of oxygen. Amino acids were analyzed on a model Biochrom 30 amino acid analyzer, with the amino acids' concentrations estimated through standard solutions that contained all the amino acids except tryptophan. The amino acids’ concentrations were estimated by reaction with ninhydrine. The internal standard was norleucine. Once the actual protein concentration has been determined in that manner, the corresponding experimental absorption coefficient at 280 nm could be derived and found to be very close (4.209 mM−1 cm−1) to the theoretical one (4.470 mM−1 cm−1).

Mass spectrometry (MALDI–TOF)

The identity of the purified PMD and PCT proteins was confirmed by mass spectral analysis of tryptic fragments. Samples for this analysis were obtained by digesting (0.25 µg trypsin) 1 µg of purified recombinant protein obtained after separation onto SDS-PAGE. Mass analysis of the tryptic peptides was performed using an Autoflex II TOF/TOF (Bruker Daltonics, Bremen, Germany). Peptides were analyzed in the Autoflex matrix-assisted laser desorption ionization/time of flight (MALDI–TOF). Spectra were acquired in the linear mode. Samples (0.7 µL containing 15 pmol) were mixed with an equal volume of sinapinic acid matrix solution, spotted on the target, then dried at room temperature for 10 min. The mass standards were either autolytic tryptic peptides or peptide standards (Bruker Daltonics). Peptide fingerprints were obtained and compared in silico with protein digest (Biotools, Bruker Daltonics, Germany).

Analytical ultracentrifugation (AUC)

Equilibrium sedimentation experiments were performed at 4 °C in a Beckman Optima-XL-A analytical ultracentrifuge in six channel centerpiece. Measurements were done at three successive speeds (10,000, 12,000, and 17,000 rpm) by taking scans at 280 nm and 257 nm, when sedimentation equilibrium was reached. High-sedimentation was conducted afterwards for baseline correction. At 4 °C, the partial specific volume of HeV PMD, solvent density and viscosity calculated with SEDNTERP (Laue et al., 1992) were 0.72793 ml/g, 0.99823 g/cm3 and 0.01567 poise, respectively.

For one concentration, a multispeed global analysis was applied to the data using the SEDPHAT program (Vistica et al., 2004) and the best fit was obtained for a single solute.

Calculation of the hydrodynamic radius and of radius of gyration

The theoretical Stokes radii (Rs, in Å) expected for a natively folded (R0) protein with an expected molecular mass (MMtheo) (in Daltons) were calculated according to (Uversky, 2002) the following equation:

\[ \log (R_0^{NP}) = 0.369 \log (MM) - 0.254 \]

(1)

The theoretical Stokes radii (Rs) of a natively folded trimer (Rtrimer) or tetramer (Rtetramer) was calculated as follows:

\[ \log (R_s^{Trimer}) = 0.369 \log (MM) \times 3 - 0.254 \]

(2)

\[ \log (R_s^{Tetramer}) = 0.369 \log (MM) \times 4 - 0.254 \]

(3)

The theoretical radius of gyration (Rg, in Å) expected for a globular protein with a hydrodynamic radius Rs was calculated according to (Wilkins et al., 1999) the following equation:

\[ R_g = (3/5)^{1/2} R_s \]

(4)

The Rs of a thin rod with a length L can be calculated as follows:

\[ R_g^2 = L^2/12 \]

(5)

PMD and PCT consist of 117 and 310 residues, respectively, including the internal methionine and the hexahistidine tag. Using an average volume of 134 Å3 per residue for proteins, and considering a trimeric organization for both PMD and PCT, the radius of a sphere with volume \( V = 4/3 \pi R_s^3 \) would be 22.4 Å in the case of PMD, and 31.0 Å in the case of PCT. According to Eq. (4), the corresponding Rg would be 17.3 Å for PMD and 24.0 Å for PCT.
Cross-linking experiments

A fixed amount of PMD (5 μg) was incubated for 20 h at room temperature with increasing amounts (0–10 mM) of suberic acid bis (N-hydroxy-succinimide ester) (SAB) in 20 mM Hepes pH 7.0, 150 mM NaCl in a final volume of 30 μL. The reactions were stopped by adding Laemli sample buffer and by heating the sample at 95 °C for 5 min. The samples were analyzed by 15% SDS-PAGE followed by Coomassie Blue staining. SAB was first solubilized in DMSO at a concentration of 1% and then diluted in 50 mM Hepes pH 7.0, 150 mM NaCl to the desired concentration.

Far-UV circular dichroism (CD)

For both PMD and PCT, CD spectra were recorded on a Jasco 810 dichrograph using 1-mm thick quartz cells in 10 mM sodium phosphate pH 7 at 20 °C either in the absence or in the presence of increasing concentrations of 2,2,2 trifluoroethanol (TFE). Both CD spectra were measured between 185 and 260 nm, at 0.2 nm/min and were averaged from three independent acquisitions. Mean molar ellipticity values per residue (MRE) were calculated as MRE = 3300 m ΔA/(lcn), where l (path length)=0.1 cm, c (protein concentration in g/l)=0.1, n (number of residues)=117 for PMD and 310 for PCT, m (molecular mass in Daltons)=13,524 for PMD and 34,923 for PCT. Spectra were deconvoluted using the DICHROWEB website (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) which was supported by grants to the BBSRC Centre for Protein and Membrane Structure and Dynamics (CPMSD) (Whitmore and Wallace, 2004, 2008). The CONTINLL deconvolution method was used to estimate the α-helical content using the reference protein set 7.

In order to monitor protein unfolding, measurements at fixed wavelength (222 nm) were performed in the temperature range of 20–100 °C with data pitch 1 °C and a temperature slope of 1 °C/min. The buffer solution without the protein was used as blank. Fitting of experimental data was done using Sigmaplot.

Small angle X-ray scattering (SAXS) measurements and ab initio 3D shape reconstructions

All small-angle X-ray scattering (SAXS) measurements were carried out at the ESRF on beamline BM29 at a working energy of 12.5 keV. The sample-to-detector distance of the X-rays was 2,847 mm, leading to scattering vectors q ranging from 0.028 to 4.525 nm⁻¹ (see Table 1 for details on data collection). The scattering vector is defined as q = 4πlα sin θ, where 2θ is the scattering angle. The exposure time was optimized to reduce radiation damage.

SAXS data were collected at 10 °C using purified protein samples (30 μL each). Protein concentrations and buffers were as follows: 0.77, 1.36 and 3.31 g/L in 20 mM Tris/HCl pH 8, 5 mM EDTA, 150 mM NaCl for PMD; 0.44, 1.25, 2.48 and 4.0 g/L in 20 mM Tris/HCl pH 8, 5 mM EDTA, 500 mM NaCl, and 5 mM DTT for PCT.

Samples were loaded in a fully automated sample charger. Ten exposures of 10 s each were made for each protein concentration and data were combined to give the average scattering curve for each measurement. Any data points affected by aggregation, possibly induced by radiation damages were excluded. The profiles obtained in the range 0.77–3.31 g/L for PMD, and 0.44–2.48 g/L for PCT had the same shape and were flat at low q values indicating the absence of significant aggregation. Then, we used the higher concentration (3.31 g/L for PMD and 2.48 g/L for PCT) to obtain maximal information at high resolution.

Data reductions were performed using the established procedure available at BM29, and buffer background runs were subtracted from sample runs. The Rg and forward intensity at zero angle I(0) were determined with the program PRIMUS (Konarev et al., 2003) according to the Guinier approximation at low values, in a QRg range up to 1.3 as follows:

\[
\ln \left[ I(Q) \right] = \ln [I(0)] - \frac{Q^2 R_g^2}{3}.
\]

The forward scattering intensities were calibrated using bovine serum albumin as reference. The Rg and pair distance distribution function, P(r) were calculated with the program GNOM (Svergun, 1992). The maximum dimension (Dmax) value was adjusted such that the Rg value obtained from GNOM agreed with that obtained from the Guinier analysis.

In the case of PMD, 3D beads models were built by fitting the scattering data with the program DAMMIF (Franke and Svergun, 2009). This program restores a low-resolution shape of the protein as a volume filled with densely packed spheres that reproduces the experimental scattering curve by a simulated annealing minimization procedure. DAMMIF minimizes the interfacial area between the molecule and the solvent by imposing compactness and connectivity constraints. 20 Independent models were generated with DAMMIF either without imposing a symmetry or by imposing a P2, P3 and P4 symmetry. The models resulting from independent runs were superimposed using the DAMAVER suite (Volkov and Svergun, 2003). This yielded an initial alignment of structures based on their axes of inertia followed by minimization of the normalized spatial discrepancy (NSD), which is zero for identical objects and larger than one for systematically different objects. Models with an NSD value greater than the mean value plus twice the standard deviation were discarded.

Accordingly, two and four models were discarded in the case of PMD and PCT, respectively, when imposing the P3 symmetry. The aligned structures were then averaged, giving an effective occupancy to each voxel in the model, and filtered at half-maximal occupancy to produce models of the appropriate volume that were used for all subsequent analyses.

Structural modeling

Structural models of HeV PMD were generated as follows. The model of the tetramer was generated using as template the structure of NiV PMD (pdb code 4N5B) whose sequence is highly similar (98%) to that of HeV PMD. The additional N-terminal five residues, including the internal methionine and the first four residues of the HeV PMD construct (aa 469–472) whose corresponding NiV residues are not defined in the electron density of the NiV PMD structure, were modeled using the SAM-T08 server (Karplus et al., 2003, 2005; Katzman et al., 2008). For each chain of the tetramer, the two last C-terminal HeV PMD residues (P577 and G578), whose cognate NiV PMD residues are not defined in the electron density of the NiV PMD structure, and the C-terminal hexahistidin tag (6xHis tag) were modeled by prolonging the coiled-coil. The final model thus consists of a long coiled-coil region (residues 506–578 plus the 6xHis tag) and an N-terminal region (residues 469–505, hereafter referred to as “head”) that is not involved in formation of the coiled-coil.

To generate the model of the trimmer, we first generated a model of a trimeric coiled-coil of 79 residues in length using the structure of hemagglutinin (pdb code 2HMG) as starting model. Then, using Chimera (Pettersen et al., 2004), we appended at the N-terminus of each of the three chains the “head” fragment in an orientation as much similar as possible to that found in the tetramer while avoiding obvious steric clashes.

The dimeric model of HeV PMD, with a 79-residue coiled-coil region, was generated using the structure of Saccharomyces cerevisiae Atp61e (pdb code 3A7O) as starting model. The “head” was appended at the N-terminus of each of the two chains as described above.

Models with the head oriented as in our previous NiV PMD model (Blocquel et al., 2013), and herein referred to as “head-up” models, were generated from the aforementioned models by
appending the “head” in opposite orientation, i.e. exposed to the solvent rather than packed back onto the coiled-coil region. Note that in these latter models the “head” conserves the same fold as in the “head-down” models, so as to minimize differences among protomers.

All the models were subjected to 10,000 steps of steepest descent energy minimization using the CHARMM19-EEF1 force field (Lazaridis and Karplus, 1999) implemented in the molecular modeling program ALMOST (Fu et al., 2014). They were then further refined to remove steric clashes and improve the agreement with the experimental SAXS profiles. The refinement was performed using ALMOST (Fu et al., 2014) and consisted of 25 rounds. In each round, 100 structures where generated by performing a simulated annealing cycle of 40,000 (80 ps) steps of molecular dynamics starting from a randomly selected structure among the 25 with the lowest energy from the previous round. As energy function we used the CHARMM19-EEF1 force field with an additional term that models the agreement of the structure with the experimental SAXS profile (Forster et al., 2008). Finally the structure with lowest energy among all 25,000 models was selected.

Statistical analysis

To estimate the statistical significance of the individual models we use Bayesian-Factors (Kass and Raftery, 1995). Bayesian-Factors provide a direct way to estimate the posterior odds of two models (hypothesis) A and B by

\[ K = \frac{\text{Pr}(A|\text{data})}{\text{Pr}(B|\text{data})} \]

where \( \text{Pr}(X|\text{data}) \) is the probability of a model \( X \) given a set of data. Applying Bayes theorem, we obtain

\[ K = \frac{\text{P(data|A)}\text{P(A)}}{\text{P(data|B)}\text{P(B)}} \]

The term \( \text{P(A|data)} \) is the prior odds of the two models. This factor can be taken to be equal to 1.0 because it is natural to assume that in the absence of any experimental data all models are equally likely. From this, it follows that in the case of a single data point the Bayesian-Factor of two models A and B is

\[ K = e^{-(1/2)\chi^2} / e^{-(1/2)\chi^2} \]

where

\[ \chi^2 = \frac{(\text{D}_\text{obs} - \text{D}_\text{calc})^2}{\sigma^2} \]

is the chi-square between the observed and the calculated data and \( \sigma \) is the experimental error. In the presence of \( N \) independent data point the previous equation can be easily generalized to

\[ \chi^2 = \sum_{i=1}^{N} \frac{(D_{\text{obs}}^i - D_{\text{calc}}^i)^2}{\sigma_i^2} \]

The chi-square between an observed and a calculated SAXS profile is given by

\[ \chi^2_{\text{SAXS}} = \frac{1}{M} \sum_{i=1}^{M} \frac{(I(q)_i - I(q)_i^\text{calc})^2}{\sigma_i^2} \]

where \( M \) is the number of sampled points. It would, however, be an error to compute the Bayesian-Factors using \( \chi^2 = M \chi^2_{\text{SAXS}} \) because not all M observations are independent. In fact, the SAXS profile \( I(q) \) is the Fourier transform of a band-limited function, the pair-distance distribution \( P(r) \), which is different from zero only for \( 0 < r < d_{\text{max}} \). In this case, the sampling theorem guarantees the ability to reconstruct the entire SAXS profile from a set of Shannon-points \( q_n = n\pi/d_{\text{max}} \) (Rambo and Tainer, 2013) as follows:

\[ l(q) = \sum_{n=0}^{\infty} I \left( \frac{n\pi}{d_{\text{max}}} \right) \sin \left( q d_{\text{max}} - n\pi \right) \]

Because the scattering intensity at any value of \( q \) is the weighted average of the intensities at Shannon-points, it follows that these points form a set of maximally independent observations. From this, it follows that the number of independent observations in a scattering profile truncated at \( d_{\text{max}} \) is as follows:

\[ n = q_{\text{max}} d_{\text{max}} / \pi \]

For HeV PMD the maximal distance obtained from the calculation of the pair-distance distribution is \( d_{\text{max}} = 130 \text{ Å} \). With a \( q_{\text{max}} = 0.5 \text{ Å}^{-1} \) we obtain approximately \( n = 20 \) independent observations. Putting everything together, we get that the final equation for relative probability of two models is

\[ K = e^{-10\chi^2_{\text{Tot}}} / e^{-10\chi^2_{\text{Tot}}} \]

where

\[ \chi^2_{\text{Tot}} = \chi^2_{\text{SAXS}} + \chi^2_{\text{Mass Spec}} + \chi^2_{\text{Binding}} \]

is the sum of the SAXS chi-square at the three measured PMD concentrations.

Acknowledgments

We wish to thank Petra Pernot and Martha Brennich (ESRF) for their help in SAXS data collection, and the ESRF synchrotron for beamtime allocation. We are also grateful to Marion Sevajol and François Ferron (AFMB lab) for their help in SAXS data collection and analysis, and Gerlind Sulzenbacher (AFMB lab) for efficiently managing the AFMB BAG. We also thank Jean-Pierre Andrieu (Seq3A Platform: sequencing and amino acid compositions analyses, Institut de Biologie Structurale, Grenoble, France) for amino acid composition analyses of HeV PMD, and Christophe Flaudrops from the mass spectrometry platform of the IFR48 of Marseille for mass spectrometry analyses. S.L. would like to thank Roberta Pierattelli (CERM, Universita’ degli Studi di Firenze) for co-supervising with her M.B. Finally, we would like to express our gratitude to Johnny Habchi (previously within the AFMB lab and presently at the Dept. of Chemistry, Cambridge, UK) for useful comments and suggestions on the manuscript.

Appendix A. Supplementary information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2014.12.029.

References


