Crystal structure of the Nipah virus phosphoprotein tetramerization domain

Jessica F. Bruhn-Johannsen1, Katherine Barnett1#, Jaclyn Bibby2, Jens Thomas3, Ronan Keegan4, Daniel Rigden3, Zachary A. Bornholdt1, Erica Ollmann Saphire1,5*

1Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, CA 92037
2Department of Chemistry, Robert Robinson Laboratories, University of Liverpool, Liverpool L69 7ZD, UK
3Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK
4RCaH, STFC Rutherford Appleton Laboratory, Chilton OX110FA, UK
5The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037
#Present address: Program in Virology, Harvard University, Cambridge, MA 02115
*To whom correspondence should be addressed: erica@scripps.edu

Running title: Nipah virus P tetramer structure
Abstract

The Nipah virus phosphoprotein, P, is multimeric and tethers the viral polymerase to the nucleocapsid. We present the crystal structure of the multimerization domain of Nipah P: a long, parallel, tetrameric, coiled coil with a small, α-helical cap structure. Across the paramyxoviruses, these domains share little sequence identity; yet retain similar length and structural organization, suggesting a common requirement for scaffolding or spatial organization of the functions of P in the virus life cycle.

Nipah virus is a newly emergent, bat-borne paramyxovirus found in Southeast Asia that causes encephalitis in humans with 40-90% lethality (29, 41). Currently there are no vaccines or anti-viral therapeutics approved for human use (8). Nipah virus has a single-stranded, negative-sense RNA genome that is encapsidated by the nucleoprotein, N, (29) and transcribed and replicated by the polymerase protein, L (31). The phosphoprotein, P, plays an essential role as a polymerase co-factor, enhancing polymerase processivity and allowing for the encapsidation of the newly synthesized viral genomes and anti-genomes (31). In these roles, P serves as a tether between the polymerase and its template, and also serves as a chaperone for nascent, RNA-free N, termed N₀, preventing it from non-specifically binding host RNA (23). P has an additional role in immunosuppression: blocking interferon signaling through binding host STAT-1 (4, 30).

The N-terminal domain of Nipah virus P (NTD, residues 1-469) is intrinsically disordered, and contains the binding site for N₀ (residues 1-50). The C-terminal region of P contains a well-ordered phosphoprotein multimerization domain (PMD, residues 470-
578), a flexible linker, and the X domain (XD, residues 660-709), which mediates binding to the nucleocapsid (10, 22, 23, 33). Multimerization of the P protein is critical for genome replication (11, 42). Crystal structures of the multimerization domains of P from the Sendai, Measles and Mumps paramyxoviruses have been determined (12, 14, 46). All three structures are composed of long, tetrameric coiled coils. Notably, Mumps is antiparallel (14) while Sendai and Measles are parallel (12, 46). There is minimal sequence identity for this domain among these paramyxoviruses (5-26%) (20, 28), and no high-resolution structure is yet described for this key domain of Nipah virus. Blocquel et al. have recently proposed that this domain from Nipah virus forms a trimeric coiled coil, in contrast to the tetrameric coiled coils of the other paramyxoviruses (6). Their findings are based on biophysical data including analytical ultracentrifugation (AUC), chemical cross-linking and small-angle X-ray scattering (SAXS) (6).

In order to provide a high-resolution experimental structure, the multimerization domain of Nipah virus P (residues 470-578) was cloned into pET46 and expressed in Escherichia coli with an N-terminal 6xHis tag followed by a Tobacco Etch Virus (TEV) protease cleavage site. The protein was purified via nickel affinity, ion exchange (MonoQ) and size exclusion chromatography coupled to multi-angle light scattering/refractive index (SEC/MALS). The protein elutes from SEC/MALS with an apparent molecular weight of 52.5 kDa, suggesting that it is a tetramer in solution (57.2 kDa is expected for a tetramer and 42.9 kDa for a trimer). After TEV removal of the 6xHis tag, the protein was crystallized in 0.1M imidazole pH 7.0, 25% PEG MME 550, and was cryo-protected in mother liquor diluted with 15% glycerol. Data to 2.2 Å were collected at the Advanced Light Source (Berkeley, CA) beamline 5.0.2, and processed.
using HKL3000 and d*TREK (34, 35). The crystal used for structure determination belongs to the spacegroup P1 (Table 1). Other crystals, which diffracted to somewhat lower resolution, belong to the spacegroups I422, P42,2, and C2 (Table 1). All crystal forms showed strong off-origin peaks in self-Patterson maps with a distance of 5.15 Å from the origin (2, 38, 39). These peaks correspond to the length of one full turn of an \( \alpha \)-helix and result from intra-helical vectors of long helices all oriented in the same direction (9) and are strongly suggestive of an inherent coiled coil structure (47).

The multimerization domains of Sendai (46), Measles (12) and Mumps virus P protein (14) failed to generate successful molecular replacement solutions. This Nipah structure was eventually determined using the automated pipeline AMPLE (5). Initially employed with \textit{ab initio} models, AMPLE was used here to generate search models by clustering and truncating a set of 100 comparative models of the Nipah protein produced by ROSETTA (40), these being calculated using the Mumps structure (4EIJ) as a template (26% sequence identity) (20, 28). AMPLE then used Molrep (48) to determine the correct positioning for one of these ensemble search models. In the successful case, the search model consisted of an ensemble of 30 structures, each a helical fragment (18 residues). This is the first successful deployment of AMPLE on comparative models deriving from a distantly homologous template.

The solution was refined with Refmac (32) and used as starting point for a near complete C\( \alpha \) trace of the target structure using SHELXE (43). The model was brought closer to completion through the use of multiple rounds of ARP/wARP (27) and Buccaneer (13) to build side chains and missing residues. Refinement was carried out in Phenix (1) followed by manual building in Coot (17).
The asymmetric unit contains eight monomers, which have assembled into two tetrameric coiled coils. Residues 476-576 are visible in electron density maps for most monomers. The Nipah virus coiled coil is a few residues longer than those of the other paramyxoviruses (68 versus 56-64 residues) (21, 49). Analysis of potential quaternary structures using PISA suggests that these two tetramers are energetically stable, with 15,000-20,000 Å² of buried surface area and dissociation energies of 140-200 kcal/mol (26). The two tetramers interact at the base burying 1,600 Å² into a crystal-contact assembled octamer, but neither PISA nor visual inspection can identify any trimeric interaction in this crystal structure. Based on these findings and SEC/MALS, we posit that Nipah multimerization domain forms a tetramer in solution.

The crystal structure reveals that the Nipah coiled coil is parallel, like that of Sendai and Measles viruses, thereby maintaining the N-terminal N0-binding domain and the C-terminal nucleocapsid binding domains on opposite ends of the oligomeric P protein. The crystal structure of Nipah virus P also reveals that each monomer has a two-helix, N-terminal cap attached at the outer top of each helix in the coiled coil. This cap structure is clearly visible in the recently reported SAXS structure (6) and it is likely that this crystal structure would fit nicely into their calculated envelope. In Sendai virus P, the cap structure is formed by three helices instead of two (46). Measles and Mumps virus P lack cap structures (12, 14).

The Nipah tetramer is primarily stabilized by hydrophobic interactions typical of coiled coils (isoleucines, leucines and valines) (37). At the N-terminal end, however, a pocket containing eight water molecules is formed by residue Gly 519, which is flanked by Ser 515 and Asn 522. These are the only waters found inside the channel. The
Measles and Mumps tetramers are similarly hydrophobic with only a few charged/polar residues and waters in their channels. By contrast, the inter-helical channel in Sendai virus P is lined with many aromatic and polar residues, and is filled with 23 water molecules and one calcium ion (21, 46). The multiple water molecules throughout the Sendai virus channel are accommodated by a larger superhelical radius (up to 9Å wide vs. 7.3 to 7.5Å for the other viruses); and a lower superhelical frequency (twist of -1.7°/residue, vs. -2.8 to -3.3°/residue for the other viruses) (21). Hence, although Nipah and Sendai are the only two to have an N-terminal cap, they are assembled internally via different interactions and have different superhelical parameters. The Nipah virus structure is more typical of coiled coils while the Sendai virus assembly is an outlier (21).

An additional difference between the Nipah and Sendai virus coiled coils is that the Nipah coiled coil contains Pro 544 in the middle of its long α-helix, which induces a kink. This kink is more reminiscent of the Leu 342-kinked Measles and the Gly 246-kinked Mumps structures, but only in Nipah does this kink cause a coil frame shift, breaking from ideal Crick parameters for coiled coils (21).

It is interesting to note that Mumps virus P gave a successful molecular replacement solution for Nipah virus P, as Mumps virus P is uniquely antiparallel, while Nipah, Measles and Sendai virus Ps are all parallel (14). The mumps-based model likely worked because its central kink and superhelical frequency (how twisted the oligomer is) are closer to Nipah virus than the other structures. Further, the search model did not contain a cap structure (Mumps lacks a cap, and Nipah and Sendai have differently structured caps), and the search model was a monomer, which could be placed equally well in a parallel or antiparallel fashion in molecular replacement.
Differential scanning calorimetry (DSC) suggests that the tightly coiled, hydrophobically-assembled Nipah virus P multimerization domain is highly stable, only undergoing transitions at 60.1°C, 86.1°C and 98.6°C (Figure 2). These likely reflect temperatures required for dissociation of the tetramer, unraveling of the helices, and unraveling of the cap structure. These data agree with the findings of Blocquel et al. who report transitions at 52°C and 85°C using circular dichroism (CD) (6). The third transition point (~99°C) may not have been observed because the CD study was performed over a range of 20-100°C, while the DSC experiment was performed in 20-130°C. The central value (85°C by CD, 86°C by DSC) is comparable to the 85°C previously measured by circular dichroism (CD) for unraveling of helical secondary structure in Measles virus P (12). Our DSC and SEC/MALS, and additional observations that the purified protein remains stably tetrameric for a year at 4°C, suggest that the Nipah virus P, and likely other paramyxovirus P proteins, do not easily change oligomeric state.

In Sendai virus P, the binding site for L was found to be within the multimerization domain, and several charged residues were implicated in binding (Figure 3) (7). The corresponding residues in Nipah P are Asp 554, Arg 555 and Lys 559. In the Nipah virus structure, these residues form a basic patch flanked by acidic residues. Indeed, a similar central basic patch is observed in all four paramyxovirus P structures (Figure 3), and may function as the L binding site in each of them (7).

In conclusion, the Nipah virus P multimerization domain is a long, parallel, tetrameric, coiled coil organized with an N-terminal cap and a hydrophobic core. Although the multimerization domains from the different paramyxoviruses have low
sequence identity, and differ in the presence or absence of a cap and in the composition
of their internal cores, all four crystal structures illustrate tetrameric coiled coils of
similar length. The conservation and stability of this structural feature suggests that
oligimerization of this type serves an essential scaffolding or organization function in the
paramyxovirus life cycle. Interestingly, the multimerization domain of the P proteins of
rhabdoviruses have little structural similarity with those of paramyxoviruses or with each
other (Figure 1) (15, 24).

**Protein structure and accession number.** Coordinates and structure factors for
the multimerization domain of the Nipah virus phosphoprotein have been deposited into
the Protein Data Bank under accession number 4N5B.

**Acknowledgements**

We would like to thank instructors of the Advanced Photon Source Data
Collection Workshop and CCP4 School for assistance with data collection (Nukri
Sanishvili, APS), data processing (Dominika Borek, UT Southwestern) and interpretation
of Patterson maps (Andrey Lebedev, CCP4, STFC Rutherford Appleton Laboratory). We
thank Kathryn Hastie and Claudia Blattner (TSRI) for help with SEC/MALS and DSC,
respectively. This is manuscript #25023 from The Scripps Research Institute. This work
was supported by NIH training grant T32 AI00760 to the TSRI Department of
Immunology and Microbial Science and an Achievement Rewards for College Scientists
to JBJ. EOS is supported by an Investigators in the Pathogenesis of Infectious Disease
Award from the Burroughs Welcome Fund.
References:


Figure Legends

Fig. 1. (A) Cartoon representations of the multimerization domain of P from Nipah virus, with dimensions, N- and C termini, and residues involved in kinking indicated. (B) Cartoon representations of the multimerization domains of P from three other paramyxoviruses, Sendai, Measles and Mumps, drawn on the same scale as Nipah virus P. Termini and residues involved in kinking of the central helices are indicated. (C) Cartoon representations of the multimerization domains of P from two rhabdoviruses: VSV and Rabies. Both domains are dimers in solution and crystallized as dimers, but VSV has been shown to form a tetramer in the context of the RNA replication machinery (18, 19).

Fig 2. Thermal denaturation curve of the Nipah virus P multimerization domain collected using differential scanning calorimetry (DSC). Normalized molar heat capacity is plotted over the range of 25-130°C. The peak at 60.10°C is sharp and likely corresponds to the dissociation of the tetramer. The following two peaks at 86.11°C and 98.6°C are broad and likely result from a two-state unfolding event: perhaps unfolding of the two helix cap followed by the long helix.

Fig. 3. (A) Electrostatic surface potential representations, generated in APBS (3), of the four paramyxovirus P multimerization domain structures. Residues in a basic patch of Sendai virus P implicated in binding L are indicated (7), as are residues in a corresponding basic patch of Nipah virus P. It is as yet unknown if L binds this site, or a different site in P, as noted for the rhabdovirus VSV (44) (B) Model of the organization...
of the Nipah virus replication machinery, roughly to scale, based on dimensions visualized by electron microscopy [Nipah virus nucleocapsid (45) and VSV L (36, 45)] and X-ray crystallography [this structure and the X domain of Measles virus P (25)]. For clarity, Nipah virus P is illustrated in either its polymerase co-factor function in which it is bound to L and the nucleocapsid (left), or in its role as a chaperone for nascent N⁰ (right). The three main domains of P are indicated: N-terminal domain (NTD), phosphoprotein multimerization domain (PMD) and the X domain (XD). Note that the nucleocapsid is shown in only one conformation (16) and that only about one-third of the actual width of the nucleocapsid is shown. It should also be noted that the L binding site has not been confirmed for Nipah virus.
Table 1. Data Collection and refinement statistics for the multimerization domain of Nipah virus P

<table>
<thead>
<tr>
<th>Space group</th>
<th>Beamline</th>
<th>Wavelength (Å)</th>
<th>Cell dimensions</th>
<th>Resolution rangea (Å)</th>
<th>Completenessb (%)</th>
<th>Rmerge c (%)</th>
<th>Isig</th>
<th>Total no. of reflectionsb</th>
<th>No. of unique reflectionsb</th>
<th>Refinementb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P1</td>
<td>1.0000</td>
<td>48.06, 76.88, 80.75</td>
<td>46.37 - 2.20 (2.28 - 2.20)</td>
<td>97.7 (97.2)</td>
<td>6.3 (29.8)</td>
<td>9.3 (3.0)</td>
<td>195915</td>
<td>51813</td>
<td>18.69</td>
</tr>
<tr>
<td></td>
<td>I422</td>
<td>1.0000</td>
<td>44.67, 44.67, 275.23</td>
<td>40.16 - 2.20 (2.28 - 2.20)</td>
<td>99.8 (100.0)</td>
<td>6.0 (44.1)</td>
<td>14.5 (3.6)</td>
<td>101741</td>
<td>7660</td>
<td>23.82</td>
</tr>
<tr>
<td></td>
<td>P42,222</td>
<td>0.976548</td>
<td>96.09, 96.09, 127.96</td>
<td>33.97 - 2.50 (2.59 - 2.50)</td>
<td>99.9 (99.2)</td>
<td>9.9 (70.4)</td>
<td>9.4 (2.2)</td>
<td>308837</td>
<td>21369</td>
<td>40.10</td>
</tr>
<tr>
<td></td>
<td>C22</td>
<td>1.0000</td>
<td>142.89, 55.64, 77.84</td>
<td>39.71 - 2.65 (2.74 - 2.65)</td>
<td>99.3 (99.4)</td>
<td>10.4 (56.8)</td>
<td>5.2 (1.5)</td>
<td>57710</td>
<td>16320</td>
<td>52.10</td>
</tr>
</tbody>
</table>

*Mean B value, protein (Å²) 40.10
Mean B value, water (Å²) 52.10
Mean B value, all atoms (Å²) 41.80

Ramachandran plot
Most favored region (%) 100.00
Additional favored region (%) 0.00

RMSD bond
deviations

<table>
<thead>
<tr>
<th>Lengths (Å)</th>
<th>0.0026</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angles (°)</td>
<td>0.635</td>
</tr>
</tbody>
</table>

Values in parentheses are for the highest-resolution shell.

A structure was refined only for the highest resolution crystals, belonging to the space group P1.

\[
R_{\text{work}} = \frac{\Sigma_{hkl} ||F_{\text{obs}}| - k |F_{\text{calc}}||}{\Sigma_{hkl} |F_{\text{obs}}|}
\]

\[R_{\text{free}}\] is the same as [R_{\text{work}}] with 5% of reflections chosen at random and omitted from refinement.

The 6xHis tag was not cleaved from this protein.

This protein contained a point mutation, S472D, and the 6xHis tag was not removed.
Figure 1
Figure 2

- 60.10 ± 0.01 °C
- 86.11 ± 0.06 °C
- 98.6 ± 0.1 °C
Figure 3
A N
N NN
B N
C
NNN
N
N
N N N
N
C
P544  
100 Å
 G246  
L342  
CC C C
 C ... 
Nipah Virus
Rabies Virus
Vesicular Stomatitis Virus
C C
RhabdovirusesNipah Virus Rhabdoviruses
<table>
<thead>
<tr>
<th>Space group</th>
<th>P1</th>
<th>I422</th>
<th>P42,2&lt;sup&gt;e&lt;/sup&gt;</th>
<th>C2&lt;sup&gt;f&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>ALS 5.0.2</td>
<td>ALS 5.0.2</td>
<td>ALS 5.0.3</td>
<td>ALS 5.0.2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.976548</td>
<td>1.0000</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>100.56, 100.85, 107.78</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
<td>90, 116.09, 90</td>
</tr>
<tr>
<td>Resolution range&lt;sup&gt;a&lt;/sup&gt; (Å)</td>
<td>46.37 - 2.20 (2.28 - 2.20)</td>
<td>40.16 - 2.20 (2.28 - 2.20)</td>
<td>33.97 - 2.50 (2.59 - 2.50)</td>
<td>39.71 - 2.65 (2.74 - 2.65)</td>
</tr>
<tr>
<td>Completeness&lt;sup&gt;a&lt;/sup&gt; (%)</td>
<td>97.7 (97.2)</td>
<td>99.8 (100.0)</td>
<td>99.9 (99.2)</td>
<td>99.3 (99.4)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt; (%)</td>
<td>6.3 (29.8)</td>
<td>6.0 (44.1)</td>
<td>9.9 (70.4)</td>
<td>10.4 (56.8)</td>
</tr>
<tr>
<td>I/σ&lt;sub&gt;I&lt;/sub&gt;</td>
<td>9.3 (3.0)</td>
<td>14.5 (3.6)</td>
<td>9.4 (2.2)</td>
<td>5.2 (1.5)</td>
</tr>
<tr>
<td>Total no. of reflections&lt;sup&gt;a&lt;/sup&gt;</td>
<td>195915</td>
<td>101741</td>
<td>308837</td>
<td>57710</td>
</tr>
<tr>
<td>No. of unique reflections&lt;sup&gt;e&lt;/sup&gt;</td>
<td>51813</td>
<td>7660</td>
<td>21369</td>
<td>16320</td>
</tr>
<tr>
<td>Refinement&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;&lt;sup&gt;a,c&lt;/sup&gt; (%)</td>
<td>18.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;&lt;sup&gt;a,d&lt;/sup&gt; (%)</td>
<td>23.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean B value, protein (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>40.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean B value, water (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>52.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean B value, all atoms (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>41.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored region (%)</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional favored region (%)</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSD bond deviations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lengths (Å)</td>
<td>0.0026</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angles (°)</td>
<td>0.635</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>