JVI Accepts, published online ahead of print on 1 May 2013
J. Virol. doi:10.1128/JVI.00653-13

Structural and Functional Characterization of the Mumps Virus Phosphoprotein

Robert Cox[1], Todd J. Green[1], Sangeetha Purushotham[2], Champion Deivanayagam[2], Gregory J. Bedwell[1], Peter E. Prevelige[1] and Ming Luo[1#]

[1]Department of Microbiology

[2]Department of Vision Sciences

University of Alabama at Birmingham

1720 2nd Ave S

Birmingham, AL 35294

[#] send correspondence to:

Department of Microbiology

University of Alabama at Birmingham

1720 2nd Ave S

Birmingham, AL 35294

Running title: Mumps virus phosphoprotein

18 Telephone: 205-934-4259

19 Fax: 205-934-0480

20 Email: mingluo@uab.edu

21

22 Abstract word count: 248

23 Text word count: 5,926

ABSTRACT

The phosphoprotein (P) is virally encoded by the *Rhabdoviridae* and *Paramyxoviridae* in the order *Mononegavirales*. P is a self-associated oligomer and forms complexes with the large viral polymerase protein (L), the nucleocapsid protein (N), and the assembled nucleocapsid. P from different viruses has shown structural diversities even though their essential functions are the same. We systematically mapped the domains in mumps virus (MuV) P and investigated their interactions with nucleocapsid-like particles (NLPs). Similar to other P proteins, MuV P contains N-terminal, central, and C-terminal domains with flexible linkers between neighboring domains. By pulldown assays, we discovered that in addition to the previously proposed nucleocapsid binding domain (residues 343-391), the N-terminal region of MuV P (residues 1-194) could also bind NLP. Further analysis of binding kinetics was conducted using surface plasmon resonance. This is the first observation that both the N- and C-terminal regions of a negative strand RNA virus P are involved in binding the nucleocapsid. Additionally, we defined the oligomerization domain (P_{OD}) of MuV P as residues 213-277 and determined its crystal structure. The tetrameric MuV P_{OD} is formed by one pair of long parallel α-helices with another pair in opposite orientation. Unlike the parallel orientation of each α-helix in the tetramer of Sendai virus P_{OD}, this represents a novel orientation of a P_{OD} where both the N- and C-terminal domains are at either end of the tetramer. This is consistent with the observation that both the N- and C-terminal domains are involved in binding the nucleocapsid.

Introduction

The phosphoprotein (P) is a multi-functional protein encoded in the genomes of negative strand RNA viruses (NSVs) of the *Rhabdoviridae* and *Paramyxoviridae* in the order *Mononegavirales*. P performs several essential functions in virus replication. It is the cofactor in the viral RNA-dependent-RNA-polymerase (RdRp), a complex that also includes the virally encoded large (L) protein. L harbors the catalytic functions for RNA synthesis and mRNA capping. P also directly binds the nucleocapsid, the active template for viral RNA synthesis. Through this interaction, P delivers the RdRp to the nucleocapsid. In the absence of P, the RdRp cannot recognize the nucleocapsid or gain access to the viral genome. The domain responsible for P to bind the nucleocapsid, P_{NBD}, has been mapped to the C-terminal region for a number of NSVs, including vesicular stomatitis virus (VSV), rabies virus (RABV), measles virus (MeV), Sendai virus (SeV), Mokola virus (MOKV), and mumps virus (MuV) (1-5). Three-dimensional structures of this domain revealed a certain degree of structural homology among different viruses (5-9). The crystal structure of VSV P_{NBD} in complex with a nucleocapsid-like particle (NLP) clearly showed that the P binding site in the nucleocapsid is formed by two neighboring nucleocapsid protein (N) subunits (10). The large extended loop in the C-lobe and a single α–helix (α13) of one N subunit and the same extended loop in an adjacent N subunit make up the U shaped P binding site with α13 at the bottom of the cleft. This high-affinity P binding site could only be created by N subunits that are assembled together in the nucleocapsid. By this mechanism, the cofactor P can dock the RdRp specifically to the nucleocapsid in order to use it as the template for viral RNA synthesis. In addition, P forms a stable complex with N that is

free of RNA, termed N^0-P. This encapsidation-competent complex of N and P exists prior to incorporation of N into the nucleocapsid (11). During virus replication, the N subunit in the N^0-P complex is used to concomitantly encapsidate the newly synthesized viral RNA (12). The domain essential for keeping N free of RNA in the N^0-P complex has been mapped to the N-terminal region ($P_{N°D}$) of P for several NSVs (11, 13-15). A crystal structure of VSV $P_{N°D}$ in complex with a VSV N mutant showed that $P_{N°D}$ sits in the cavity where the viral RNA would be accommodated in the nucleocapsid (16). Occupation of the cavity may be instrumental in allowing P to keep N free of RNA in the N^0-P complex.

Another important feature of P is that its functional form is a self-associated oligomer. Self-association is required for supporting viral RNA synthesis (17), but is not necessary for the association of P with other viral proteins such as L or N. The domain responsible for P oligomerization (P_{OD}) is located in the central region of P. $P_{N°D}$ and P_{NBD} are linked to P_{OD} through flexible loops, which appears to be a common modular structure for all NSV P proteins. The length of P differs greatly among NSVs, and there is no homology for P_{OD}. Furthermore, the structure and the mode of self-association also vary considerably from one P to another. In the case of SeV P (568 amino acids), P_{OD} corresponds to residues 320-433. Each SeV P_{OD} subunit donates a single long helix to form a tetrameric coiled-coil, with all helices being parallel to one another. Short helices at the N-terminal end of SeV P_{OD} contribute to additional tetrameric interactions. For VSV P (265 amino acids), P_{OD} comprises residues 107-177 (18, 19). VSV P_{OD} dimerizes by domain swapping of a β-hairpin from each subunit that participates in collating a β-sheet of four anti-parallel strands on each side. Two parallel α–helices held

93 together by a number of hydrophobic interactions form the core of the VSV P dimer. By
94 comparison, the oligomerization mode of RABV P_{OD} is quite different from that of VSV
95 P_{OD}, despite the two viruses being closely related members of the Rhabdovirus family.
96 RABV P_{OD} spans residues 92 to 131 in P (297 amino acids), and forms a dimer (20).
97 However, each polypeptide contains two anti-parallel helices linked by a loop. The
98 dimer of RABV P_{OD} involves a four-helix bundle between two parallel subunits. This
99 particular orientation places $P_{N°D}$ and P_{NBD} at the same end of RABV P, in contrast to
100 other P proteins in which the two domains are at distant opposite ends of the oligomer
101 (19-21). In VSV, it has been shown that a chimeric P dimer with one P mutant lacking
102 $P_{N°D}$ and another mutant lacking P_{NBD} can fully support VSV RNA synthesis (17). In the
103 case of the RABV P dimer, it is not clear what the functional implications are when all
104 $P_{N°D}$ and P_{NBD} domains are at the same end.

105 For this report, we carried out systematic studies of the domains of MuV P and
106 their possible interactions with the nucleocapsid. Previously, MuV P expressed as a
107 recombinant GST-fusion protein was shown to bind a truncated N (N_{398}) that was
108 described as capable of assembling into NLP (4). In addition, deletion of the last 49
109 residues in MuV P abolished NLP binding, and the C-terminal domain (C343-391) alone
110 was shown to be capable of binding NLP (4). In contrast, nucleocapsid binding by the P
111 protein from other paramyxoviruses reportedly requires interactions with the C-terminal
112 tail region of N (22-25). The previous report appears to suggest that MuV P contains a
113 P_{NBD} that is similar to other paramyxoviruses, but its binding site on the nucleocapsid
114 may be located in a different region of N. The crystal structure of the C-terminal domain
115 of MuV P showed that it contains three packed α–helices that may be induced to fold

when P binds the nucleocapsid (7). We expanded the study of nucleocapsid-binding to include all domains in MuV P. Our structural and functional characterization of MuV P is discussed here with comparison to other NSV Ps.

Materials and Methods

Materials. Restriction enzymes and T4 DNA ligase were purchased from New England BioLabs. pET28b and BL21(DE3) were purchased from Novagen. Primers for PCR were purchased from Invitrogen and Integrated DNA Technologies.

Plasmid construction. The P genes were amplified from a cDNA clone of MuV strain 88-1961 (GenBank: AF467767.2) by PCR. Primers for amplification on the 5' end of the gene fragments contained NheI restriction sites while 3' primers contained XhoI sites. Sequences of the primers can be sent upon request. The P gene or gene fragments and the pET-28b plasmid were each digested with restriction enzymes NheI and XhoI and gel purified. The purified digested P genes were individually ligated in frame with the N-terminal polyhistidine tag, resulting in the vector MuV P-pET28b, or MuV (P clone ID)-pET28b.

Protein expression and purification. Expression vectors were transformed into *E. coli* strain BL21(DE3). Bacteria were cultured in LB broth at 37 °C until OD_{600} reached 0.6. Protein expression was induced with 1 mM IPTG for 18 hours at 18 °C. The cells were harvested by centrifugation and resuspended in binding Buffer A containing 20 mM Tris (pH 7.9), 500 mM NaCl and 5 mM imidazole. The cells were disrupted by sonication, and then centrifuged for 1hr at 18,000rpm. Soluble fractions were collected and passed through a Ni-Affinity column (Chelating Sepharose Fast Flow, GE Healthcare). The

loaded Ni-affinity column was washed with 5 column volume (CV) of binding Buffer A. The loaded column was washed with 5 CV of wash Buffer A containing 20 mM Tris (pH 7.9), 500 mM NaCl, and 50 mM Imidazole. Samples were eluted with elution Buffer A containing 20 mM Tris (pH 7.9), 500 mM Imidazole, and 500 mM NaCl. Proteins were further purified by size exclusion chromatography (Sephacryl S-75 or S-200, GE Healthcare) in size exclusion buffer containing 20 mM Tris pH 7.5 and 500 mM NaCl.

Purification of trypsinized MuV NLP. The MuV N protein was expressed and NLP was purified as previously reported (26). NLP has a ring structure composed of 13 N subunits and a single strand of random RNA. The purified NLP was allowed to incubate overnight at room temperature with trypsin at a 1 mg:100 mg ratio (trypsin:N). Trypsin was then inactivated by the addition of PMSF (1 mM). The digested NLP (NLP_{379}) was concentrated to 8 mg/ml and purified by size exclusion chromatography using a S-400 column (HiPrep Sephacryl S-400, GE Healthcare). The identity of the digested N was confirmed by N-terminal amino acid sequencing and mass spectrometry analysis. The size of the NLP_{379} was consistent with that of a ring structure similar to the full-length NLP.

Identification of protease resistant fragments of the MuV P protein. The full-length P protein was digested using three different proteases: trypsin (TPCK treated, Worthington), α-chymotrypsin (Sigma), and proteinase K (Fisher). The amounts of enzymes used per mg of the P protein were 0.2 U, 0.1 U, 0.2 μg for trypsin, chymotrypsin, and proteinase K, respectively. The P protein was digested with each protease for a maximum of 2 hrs at 37°C. During this time course, aliquots of each digestion mixture were taken at five minute intervals and denatured for analysis in SDS-

PAGE gels. SDS-PAGE gels of digestion reactions were electroblotted onto PVDF membranes and stained with coomassie blue. Bands corresponding to stable P fragments were cut out and sent to the Protein Analysis Core at UTMB (Galveston, TX) for N-terminal amino acid sequencing. Products at optimal time-points from the limited digestions of P were purified using a Gelfree 8100 Fractionation System (Protein Discovery), in accordance with the manufacturer's protocol. Fractions containing pure digestion products were subjected to mass spectrometry analysis at the Protein Analysis Core at UTMB (Galveston, TX).

Bioinformatics analysis

The secondary structure of MuV P protein was predicted using Jpred3 (27). PIV5 V protein (Accession #AAA47882) was used in the Jpred3 homologue search.

Crystallization and data collection of MuV P_{OD}. P161-277 was expressed and purified as described above and concentrated to 5-7 mg/mL. Crystallization conditions at 20°C were screened using Crystal Screen 1 & 2 (Hampton). An initial crystal hit was found in the handing drop set-up with a reservoir solution containing 0.1 M NaAc (pH 4.6) and 2 M NaCl. Typical crystals appeared in hanging drops after 10-14 days at 20°C. Crystals were cryo-protected in a cryosolution containing the reservoir solution supplemented with 25 % glycerol. For preparation of uranyl acetate derivatives, crystals were soaked stepwise in the reservoir solution containing 0.5, 1.0, 1.5 and 2.5 mM uranyl acetate at approximately 10 minute intervals. Uranyl acetate soaked crystals were cryo-protected by the same protocol as native crystals, except for the addition of 2.5 mM uranyl acetate to the cryosolution. Native crystals diffracted X-rays to a

183 resolution of 2.2 Å when exposed to synchrotron X-rays at SER-CAT beamline BL-22
184 BM at APS.

185 **Structure determination and refinement.** The phases of X-ray diffraction data were
186 determined by SIRAS using the program CNS (28). The uranium atom in the uranyl
187 acetate derivative was used as an anomalous scatterer. An initial model was manually
188 built into solvent flattened maps. The model was then subjected to several rounds of the
189 Autobuild routine in PHENIX (29). The resulting model was subjected to several cycles
190 of manual rebuilding using COOT (30) and refinement with PHENIX (29). TLS
191 refinement was performed using the PHENIX program (31, 32)]. The structure has been
192 deposited in the Protein Data Bank under the code 4EIJ. Solvent accessible surfaces
193 were calculated using CNS (28). Structure figures were created using PyMOL (33).
194 Data collection and refinement statistics are shown in Table 1.

195 **Contact area between the helices.** In order to estimate how tight the interactions are
196 between the different helices, calculations of the buried surface areas were performed
197 in CNS (28) (Table 2). The total buried surface area is similar for each of the inter-chain
198 interactions. Among the three possible pairs of contacts, Chain A-B proved to have the
199 highest percentage of buried surface. However, the negligible differences among the
200 three pairs strongly suggest that the oligomeric state of MuV P_{OD} is a tetramer.

201 **Glutaraldehyde crosslinking.** Protein crosslinking studies were carried out with the
202 addition of 0.1% glutaraldehyde stock solution diluted in PBS. Increasing concentrations
203 of glutaraldehyde were added to 50 µl of P161-277 (3 mg/ml) and allowed to incubate

for 4 hours or overnight on ice. The reactions were quenched by addition of 18 μl of SDS-PAGE sample buffer. Samples were denatured and analyzed by SDS-PAGE.

Sedimentation Velocity Analysis. Sedimentation velocity experiments were performed in an XL-A analytical ultracentrifuge (Beckman Coulter) in two channel epon centerpieces. The data were collected at 280 nm**,** and a speed of 40,000 rpm in an An60Ti 4-hole rotor at 20˚C. The concentration of protein loaded was ~1 mg/mL. The data (Table 3) were analyzed with the program SEDFIT using both the c(s) and c(s,ff0) models (http://www.analyticalultracentrifugation.com/default.htm). The c(s,ff0) model was used for molecular weight estimation of the sedimenting species. The buffer density, buffer viscosity, and protein partial specific volume used in these calculations were estimated with the program SEDNTERP (http://bitcwiki.sr.unh.edu/index.php/Main_Page).

Pulldown assay. In order to map N-P interactions, a pulldown assay was performed. A small column containing 50 μl of charged Chelating Sepharose Fast Flow beads was first saturated with the P protein or a P fragment. After allowing the sample to flow through the column, the beads were washed with 10 CV of binding Buffer B, containing 20 mM Tris (pH 7.9), 50 mM NaCl, and 5 mM Imidazole. NLP, or NLP_{379}, in binding Buffer B was added to the loaded column at a 1:1 molar ratio (N:P) and incubated at room temperature for 15 minutes. The NLP or NLP_{379} solution was allowed to flow through. Next, the loaded column was washed with 10 CV of binding Buffer B, followed by wash with 5 CV of wash Buffer B containing 20 mM Tris (pH 7.9), 50 mM NaCl, and 50 mM Imidazole. The protein was eluted with 5 CV of elution Buffer B containing 20 mM Tris (pH 7.9), 50 mM NaCl, and 500 mM Imidazole. The eluate was denatured in

SDS-PAGE sample buffer and electrophoresed in a 12% SDS-PAGE gel. Gels were stained with 1% Coomassie Brilliant blue dye. The full-length MuV P protein was used as a positive control for binding NLP. The purified NLP or NLP_{379} in the absence of any P fragments was used as a negative control.

BIAcore 2000 Analysis. Real time binding analysis of the MuV P protein to the MuV NLP and its truncated form NLP_{379} were analyzed using surface plasmon resonance on a BIAcore 2000 instrument (GE Healthcare). Purified NLP or N_{379} ligands were immobilized on the carboxylated surface of a CM5 sensor chip using an amine-coupling kit. Analytes full-length MuV P and P fragments P1-194, P161-277, and P286-391 were flowed over the immobilized NLP or NLP_{379} CM5 chip at a flow rate of 20 µL/min at 25˚C. Regeneration of the chip surface after each cycle was performed using (0.015 M HEPES in 0.5 M NaCl). The change in the surface plasmon resonance angle is displayed as response units, where 1,000 response units (RU) is equal to 1 ng of analyte bound per nm^2 on the sensor surface. The experiments were carried out in triplicates and kinetic association (K_A) and dissociation (K_D) rate constants were deduced using the 1:1 Langmuir kinetic model with BIA evaluation software.

Results

Digestion Studies of MuV P

In order to study the structure and function of the MuV P protein, a single vector was constructed to express the MuV P protein in *E. coli*. Following expression, his-tagged recombinant P was purified from the soluble fraction of the lysate. The P protein was further purified by size exclusion chromatography and was analyzed by SDS-

PAGE. Purified P migrates to a position consistent with that observed in previous publications, approximately 5 kDa higher than the calculated molecular weight of 41.5 kDa (26). Slow migration has also been observed for VSV P (34), suggesting that NSV P proteins appear to migrate slower in SDS PAGE

The full-length P protein was shown to be susceptible to digestion by residual impurities over a brief time period following purification. In order to define domains within MuV P, limited proteolytic digestion studies were performed. Purified full-length P protein was digested using three different proteases: trypsin, chymotrypsin, and proteinase K. Digestion products were examined by SDS-PAGE, which showed that each protease had a different digestion pattern (Figure 1A-C). In order to identify the N- and C-termini of the stable fragments, individual bands corresponding to digested products were isolated and analyzed by N-terminal amino acid sequencing and mass spectrometry. The results for the limited digestion studies are summarized in Figure 1D. Trypsin digestion resulted in a major stable fragment corresponding to residues 135-391 with an N-terminal sequence of MINRF and a molecular weight of 27.7 kDa. Chymotrypsin digestion resulted in a major stable fragment corresponding to residues 194-391 with an N-terminal sequence of AHPSP and a molecular weight of 21.7 kDa. While bands corresponding to other partially digested products were seen on SDS-PAGE gels for both trypsin and chymotrypsin digestions, the molecular weight of these fragments could not be determined by mass spectrometry. Proteinase K digestion resulted in two stable fragments. Identification of mass by mass spectrometry was not successful. The residue numbers for the two K fragments (K1 and K2), therefore, were designated based on their N-terminal sequence, and the molecular weight

approximated from their position on SDS-PAGE. The N-terminal sequence for K2 was SVISA and likely corresponds to residues 214-391. The N-terminal sequence for K1 was somewhat ambiguous, but the sequence in cycles 3 thru 5 was PSP. This sequence with the size observed on SDS-PAGE of approximately 22 kDa is consistent with a fragment encompassing residues 194-391. The chymotrypsin and K1 fragments are thus the same, and consistent with the cleavage pattern of each enzyme where each cleaves adjacent to the carboxyl group of bulky or aromatic amino acids. The sequence at this cleavage site of P is Y/AHPSP.

Bioinformatics Analysis

Secondary structure prediction by Jpred3 identified nine different regions as probable helical regions (Figure 1E). The prediction also identified four different regions that were likely to have β-strand conformation. In addition, the Jpred3 search identified PIV5 V as a homologue of MuV P. PIV5 V has a 37% identity to MuV P, located within the first 155 residues and 164 residues of MuV P and PIV5 V, respectively. These data were used along with the results of protease digestion to design a clone library of MuV P fragments. These protein fragments were expressed and purified by the same methods used for the full-length P. The yield of purified protein per liter varied for each fragment. A list of cloned P fragments and a summary of protein expression results are shown in Figure 2. P fragments that were consistent with the results of the protease digestion have a better expression level in the soluble fraction.

Structure of the P oligomerization domain

Several fragments were subjected to crystal screens. Crystals were grown with fragment P161-277 and the crystal structure was determined to 2.2 Å resolution. Only residues 213-277 could be fit correctly into the electron density maps. The traced polypeptide represented an appropriate protein volume for the unit cell with a solvent content of 60.04%, assuming a partial specific volume of 0.74 cc/g. Composite omit maps were created to confirm that modeling was carried out correctly. The lack of residues 161-212 is likely due to digestion *in situ* during crystallization as crystal formation only occurred after two weeks. The asymmetric unit contains two chains (A and B) that cover residues 213-273 and 215-277, respectively (Figure 3A). Both chains are composed of a single long α-helix. However, Chain A, unlike Chain B, contains a kink at Gly 246. Chain A also contains a stretch of amino acids with extended conformation at the C-terminal end (residues 272-277) of its long helix.

The crystallized MuV P fragment forms a tightly packed tetramer through non-crystallographic and crystallographic symmetry contacts (Figure 3B). We therefore define this tetramer as MuV P_{OD}. This tetrameric coiled-coil structure is reminiscent of SeV P_{OD} (Figure 3C). However, the tetramer formed by SeV P_{OD} is composed of four parallel helices, whereas the tetramer formed by MuV P_{OD} is composed of two sets of parallel helices that are in opposite orientation (Figure 3B). In the MuV P_{OD} tetramer that is primarily formed with hydrophobic interactions (Table 3), there are two zippers of charged sidechain interactions at each end in addition to hydrophobic surface contacts among the helices. Between the parallel pair of helices, the zipper is formed by Asp229, Glu236 and Asp240 (Chain A) and Arg231 and Lys238 (Chain B) as mentioned above (Figure 4A; inset A). Between the antiparallel pair of helices, the zipper is formed by

317 Lys253 and Lys260 (Chain B) and Asp229, Glu236 and Asp240 (Chain B') (Figure 4B;
318 inset B). The helices wrap around each other and seem to clamp one another through
319 these charge zippers at both ends of the tetrad.

320 **Glutaraldehyde Crosslinking**

321 In order to verify the tetrameric state of the MuV P protein, crosslinking reactions
322 were performed using glutaraldehyde. Initially, full-length P was used in crosslinking
323 experiments. However, because MuV P is highly susceptible to nonspecific digestion
324 (35), the resulting crosslinked products were composed of a mixture of different P
325 fragments, and no clear data could be discerned (data not shown). Since the majority of
326 the known protease sensitive sites are contained within the terminal regions, P161-277
327 was used in place of the full-length P protein and the resulting crosslinked products
328 were much easier to identify. Following analysis in SDS-PAGE gels, bands representing
329 monomers, dimers, trimers, and tetramers could be clearly discerned (Figure 5A).

330

331 **Sedimentation Velocity Analysis**

332 To further confirm the oligomeric state determined by the P_{OD} crystal structure,
333 we carried out ultracentrifugation analyses to measure the sedimentation coefficient of
334 the purified P protein and fragment P161-277. For both P161-277 and full-length P, the
335 data revealed a single dominant species (Figure 5). The molecular weight derived from
336 sedimentation coefficient is 57.8 kDa for P161-277, which corresponds to the theoretical
337 weight of a P161-277 tetramer (57.6 kDa). Furthermore, the molecular weight derived

338 from sedimentation coefficient is 165.5 kDa for full-length P, very close to the theoretical
339 molecular weight for a MuV P tetramer (168 kDa). We therefore concluded that the
340 purified MuV P protein or P161-277 is a tetramer, concurring with the results from the
341 crystal structure.

342 **Nucleocapsid Binding**

343 A number of recombinant P fragments were expressed and purified (Figures 1
344 and 2). In order to determine which parts of the MuV P protein interact with the
345 nucleocapsid, a pulldown assay was developed. Purified P fragments were first added
346 to a small Ni-affinity column in saturating amounts. After allowing the supernatant to
347 flow through, the beads were washed once to remove any unbound protein. NLP of
348 MuV prepared as previously reported (26) was added to the loaded column at a 1:1
349 molar ratio (N:P) and allowed to incubate at room temperature for 15 minutes. The NLP
350 solution was then allowed to flow through. The loaded column was washed to remove
351 any unbound protein. The proteins were then eluted from the column, and the products
352 were denatured and electrophoresed in SDS-PAGE gels. A summary of the results from
353 the pulldown assays is shown in Figure 6. These assays confirmed that the extreme C-
354 terminal domain, P342-391, is a nucleocapsid binding domain as previously reported
355 (4). Other C-terminal fragments including residues 251-391, 286-391 and 342-391 each
356 bound NLP (Figure 6). Unexpectedly, the results showed that the N-terminal region of
357 MuV P (P_{1-194}) was also capable of binding NLP. This has not been reported previously
358 for a NSV P protein and suggests that the functional domains of MuV P may be different
359 from those of other NSVs.

In addition to using NLP formed by the full-length NLP (540 amino acids), binding assays were also conducted using a trypsin digested form of the MuV NLP (NLP_{379}). The digested NLP_{379} was used to determine if P binding could be abolished in NLP lacking the C-terminal region of N. Similar to the full-length NLP, NLP_{379} was also bound by both the N-terminal (1-194, 51-194) and C-terminal (251-391, 286-391, and 342-391) regions of MuV P (Figure 6D,E).

Affinity of P fragments binding NLP

Based upon pulldown experiments, we concluded that MuV P contains two separate nucleocapsid binding domains: the N-terminal domain (P1-194) and the C-terminal domain (P286-391). In order to further characterize the interactions between the two separate nucleocapsid binding domains and NLP, surface plasmon resonance analysis was conducted. The recombinant full-length P, P1-194, P161-277, and P286-391 were investigated for their ability to interact with immobilized NLPs or NLP_{379}. All analytes, except for P161-277 as a negative control, exhibited measurable binding (Figure 7; Table 4). Kinetics experiments were conducted at multiple concentrations (Figure 7). Modeling the association and dissociation of the P analytes with NLP using the 1:1 Langmuir binding model led to estimated K_A values of 2.02×10^5 M^{-1} for N1-194, 2.35×10^6 M^{-1} for P286-391, 2.23×10^6 M^{-1} for the full-length P (Table 4). Using K_D calculations, P286-391 also exhibited a relatively higher affinity for NLP with an estimated dissociation constant of 4.26×10^{-7} M as compared with N1-194 (4.95×10^{-6} M). Next we tested the ability of P to bind immobilized NLP_{379} (Figure 7, E-H). Modeling the association and dissociation of the P analytes with NLP_{379} using the 1:1 Langmuir binding model led to estimated K_A values of 2.17×10^6 M^{-1} for P1-194, 1.01×10^5 M^{-1}

383 for P286-391, 1.61 × 10^6 M^{-1} for the full-length P (Table 4). The estimated K_A and K_D
384 values, P1-194 exhibited a relatively higher affinity for NLP_{379} with an estimated
385 dissociation constant (K_D) of 4.6 × 10^{-7} M as compared with P286-391, which has a K_D
386 value of 9.88 × 10^{-6} M. Control SPR experiments with P161-277, which contains the OD
387 domain, did not show measurable interactions with full length NLP or NLP_{379}. The SPR
388 data indicate a 1.3-fold higher affinity in P binding the full length NLP versus NLP_{379}, but
389 shows a slower dissociation rate (Fig. 7A, E). The curves of Fig. 7F, G show that there
390 are very fast on and off rates for both P1-194 and P286-391 to NLP_{379} as compared to
391 the full length NLP.

392 **Discussion**

393 It has been anticipated that MuV P has a modular structure composed of
394 separated functional domains. Here, we present an experimental approach to more
395 accurately map the functional domains of MuV P. In our proteolytic studies, the N-
396 terminal region of P preceding residue 194 appeared to be less stable than the region
397 that follows since more cleavage sites were found here. This is consistent with
398 observations of other NSV Ps in which the N-terminal region is more disordered
399 compared to the middle and C-terminal regions of P (11, 23, 36, 37). A previous study
400 determined the structure of an extreme C-terminal segment (residues 343-391) of MuV
401 P (7). Our proteolytic digestion study did not identify this smaller region of P as a stable
402 fragment. A possible explanation for not being able to map P343-391 in the current
403 study is that this C-terminal region is part of a larger C-terminal domain. The P343-391
404 region was previously proposed to be MuV P_{NBD} by analogy to other P proteins. P343-

391 may be a molten globule since its crystallization required the addition of stabilizing agents (7). On the other hand, the protein expression data did point to a potential linker between P_{OD} and the C-terminal domain to be from residue 277 to 286. This could suggest that the C-terminal domain of MuV P is larger than previously thought. P proteins of other NSVs were reported to contain flexible regions between $P_{N°D}$, P_{OD}, and P_{NBD} (1, 21, 36, 38-43).

Crystallization studies clearly defined that MuV P_{OD} comprises residues 213-277 when P161-277 was used in crystal growth. The structure of MuV P_{OD} was determined and shows a stable tetramer containing one pair of long parallel α-helices (64 residues) and a second pair of parallel helices in opposite orientation. The presence of P tetramers in solution was confirmed by analytical ultracentrifugation and crosslinking with glutaraldehyde (Figure 5). The unique organization of this tetrameric coiled-coil where parallel helices are joined with antiparallel helices has not been reported for a native protein before. In this distinctive assembly, interactions between the parallel helices are different from those between the antiparallel helices. By contrast, interactions between any neighboring pair of helices are identical in an all parallel or all antiparallel coiled-coil. The helices in the MuV P_{OD} tetramer are held together by a large hydrophobic contact area between each monomer (Table 2) and two zippers of charged sidechain interactions (Figure 4). This interesting orientation of antiparallel helices places the N-terminal and the C-terminal regions at both ends of MuV P_{OD}. This particular structure may well suit the functional requirements of MuV P as discussed below.

427 We have shown that P binds the full length MuV NLP and truncated NLP_{379}.
428 These observations confirm that the P binding site in the MuV nucleocapsid does not
429 fully require the N-tail, same as previously reported (4). In lieu of the apparent
430 dispensability of this N-tail, nucleocapsid binding by MuV P may involve several regions
431 of both N and P. Given this and the novel structure of MuV Pod, P may well be different
432 from other NSV Ps in binding the nucleocapsid. With this open-minded approach, His-
433 tagged fragments covering the full range of MuV P were tested for binding a well
434 prepared NLP composed of 13 N subunits and a piece of random RNA as previously
435 described (26). In the His-tag pulldown assays, full-length MuV P and the previously
436 identified C-terminal domain, P343-391, were shown to bind NLP as expected. What
437 was unexpected is that the N-terminal region of MuV P from residues 1 to 194 could
438 also bind NLP. This indicates that specific nucleocapsid binding may require both the N-
439 terminal and the C-terminal regions of MuV P, a binding mode that is different from all
440 other known NSV Ps. The NLP binding by the N-terminal and C-terminal regions was
441 confirmed by surface plasmon resonance experiments. This analysis showed that the
442 C-terminal region of P has a higher affinity for NLP. When the N-tail was removed by
443 trypsin digestion (NLP_{379}), the C-terminal region of P still binds NLP_{379} with a lower
444 affinity. The affinity of binding NLP_{379} by the C-terminal region of P is similar as that of
445 the N-terminal region of P binding to the full length NLP. These observations suggest
446 that the C-terminal region of P interacts with the N-tail and additional regions of N, and
447 the N-terminal region of P interacts with regions of N not including the N-tail. However,
448 there is a 10.7-fold increase in binding NLP_{379} by the N-terminal region of P versus the
449 full length NLP. This increase could point to the accessibility of the N^0 binding site for P

upon removal of the N-tail. Evidence shows that this site is located in the RNA cavity for other N proteins (11, 36, 37). The N^0 binding motif of P could potentially bind in this site if encapsidated RNA was displaced from NLP_{379}. The overall affinity of the full length P for NLP is lower than that of the individual domains. The affinity values for the full length tetrameric P may be underestimated because the NLP is representative of a single turn of the helical nucleocapsid and thus does not have the multiple binding sites available in the helical nucleocapsid (described below). In addition, the full length P was not very stable in solution.

Comparisons of known structures of P uncovered three modes of P oligomerization (Figure 3B-E). In the SeV P tetramer, oligomerization results in a coiled-coil structure of four long parallel α–helices (21). Similarly, the subunits in the VSV P dimer are also parallel to each other even though the β–sheet on each side is formed by domain swapping (19). The parallel orientation of P subunits in the oligomers places the N-terminal regions on one end and the C-terminal regions on the other. Since specific nucleocapsid binding in these viruses requires only the C-terminal region of P, the N-terminal region may be placed far away from the C-terminal region. In the RABV P dimer, however, the N-terminal and C-terminal regions are each positioned at the same end due to the U-shaped structure that links two antiparallel helices in each P_{OD} subunit (20). There is no indication that the N-terminal region of RABV P is involved in binding the nucleocapsid. It is not clear what functional implications this particular orientation of RABV P has. MuV P represents a third mode of oligomerization in which parallel and antiparallel P subunits are associated together to form a tetramer. In the MuV P tetramer, the dimensions of the helix bundle are similar to those of the coiled-coil

structure of the SeV P tetramer, but the orientations of the helices is different. The antiparallel orientation of the MuV P subunits places two N terminal regions and two C-terminal regions on each end of the tetramer. This organization may be functionally required because both the N-terminal and C-terminal regions of P are involved in binding the nucleocapsid as shown in this report. The two terminal regions should be in close proximity to each other in order to effectively bind the nucleocapsid. The full P binding sites for each terminal regions in the nucleocapsid remain to be identified for MuV. The results shown in Figure 6 and 7 suggest that there are P binding sites within residues 1-379 in addition to the C-terminal tail of N. This mode of binding requires that P bind at least two or more N subunits either in the same turn or in successive turns in the helical nucleocapsid. Since the conserved sequences for initiation of replication (PrE I and II) are located in two successive turns of the nucleocapsid (44) it is possible that P plays a role in recognition of the two sites by the viral RdRp. Our results again demonstrated that NSV P proteins all have a modular structure with separated functional domains and are present as oligomers (Figure 8). While NSV P proteins harbor analogous functions, each may have very different domain organization, and can interact with other viral proteins, especially the nucleocapsid, in a very different manner.

Acknowledgments

We thank the staff of the South East Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source (APS), Argonne National Laboratory, for their assistance in data collection. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. SER-CAT supporting institutions may be found at

496 www.ser-cat.org/members.html.

497 We thank the generosity of the staff at the Stanford Synchrotron Radiation
498 Lightsource (SSRL). "Portions of this research were carried out at the SSRL, a national
499 user facility operated by Stanford University on behalf of the U.S. Department of
500 Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology
501 Program is supported by the Department of Energy, Office of Biological and
502 Environmental Research and by the National Institutes of Health, National Center for
503 Research Resources, Biomedical Technology Program, and the National Institute of
504 General Medical Sciences." The work is supported in part by a NIH grant to ML
505 (AI050066).

506 **Author Contributions:** Conceived and designed the experiments: RC TG ML.
507 Performed the experiments: RC TG SP GB. Analyzed the data: RC TG SP CD GB PP
508 ML. Wrote the paper: RC TG SP CD GB PP ML.

References

1. Takacs, A. M., T. Das, and A. K. Banerjee. 1993. Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. Proc Natl Acad Sci U S A 90:10375-9.

2. Toriumi, H., Y. Honda, K. Morimoto, T. S. Tochikura, and A. Kawai. 2002. Structural relationship between nucleocapsid-binding activity of the rabies virus phosphoprotein (P) and exposure of epitope 402-13 located at the C terminus. J Gen Virol 83:3035-43.

3. Ryan, K. W., and A. Portner. 1990. Separate domains of Sendai virus P protein are required for binding to viral nucleocapsids. Virology 174:515-21.

4. Kingston, R. L., W. A. Baase, and L. S. Gay. 2004. Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins. J Virol 78:8630-40.

5. Assenberg, R., O. Delmas, J. Ren, P. O. Vidalain, A. Verma, F. Larrous, S. C. Graham, F. Tangy, J. M. Grimes, and H. Bourhy. 2010. Structure of the nucleoprotein binding domain of Mokola virus phosphoprotein. J Virol 84:1089-96.

6. Gely, S., D. F. Lowry, C. Bernard, M. R. Jensen, M. Blackledge, S. Costanzo, J. M. Bourhis, H. Darbon, G. Daughdrill, and S. Longhi. 2010. Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. J Mol Recognit 23:435-47.

7. Kingston, R. L., L. S. Gay, W. S. Baase, and B. W. Matthews. 2008. Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. J Mol Biol 379:719-31.

8. Ribeiro, E. A., Jr., A. Favier, F. C. Gerard, C. Leyrat, B. Brutscher, D. Blondel, R. W. Ruigrok, M. Blackledge, and M. Jamin. 2008. Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. J Mol Biol 382:525-38.

9. Blanchard, L., N. Tarbouriech, M. Blackledge, P. Timmins, W. P. Burmeister, R. W. Ruigrok, and D. Marion. 2004. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. Virology 319:201-11.

10. Green, T. J., and M. Luo. 2009. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. Proc Natl Acad Sci U S A 106:11713-8.

11. Chen, M., T. Ogino, and A. K. Banerjee. 2007. Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. J Virol 81:13478-85.

12. Qanungo, K. R., D. Shaji, M. Mathur, and A. K. Banerjee. 2004. Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. Proc Natl Acad Sci U S A 101:5952-7.

13. Mavrakis, M., S. Mehouas, E. Real, F. Iseni, D. Blondel, N. Tordo, and R. W. Ruigrok. 2006. Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. Virology 349:422-9.

14. Majumdar, A., R. Bhattacharya, S. Basak, M. S. Shaila, D. Chattopadhyay, and S. Roy. 2004. P-protein of Chandipura virus is an N-protein-specific chaperone that acts at the nucleation stage. Biochemistry 43:2863-70.

15. Mir, M. A., and A. T. Panganiban. 2006. Characterization of the RNA chaperone activity of hantavirus nucleocapsid protein. J Virol 80:6276-85.

16. Leyrat, C., F. Yabukarski, N. Tarbouriech, E. A. Ribeiro, Jr., M. R. Jensen, M. Blackledge, R. W. Ruigrok, and M. Jamin. 2011. Structure of the vesicular stomatitis virus N-P complex. PLoS Pathog 7:e1002248.

17. Chen, M., T. Ogino, and A. K. Banerjee. 2006. Mapping and functional role of the self-association domain of vesicular stomatitis virus phosphoprotein. J Virol 80:9511-8.

18. Ding, H., T. J. Green, and M. Luo. 2004. Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). Acta Crystallogr D Biol Crystallogr 60:2087-90.

19. Ding, H., T. J. Green, S. Lu, and M. Luo. 2006. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. J Virol 80:2808-14.

20. Ivanov, I., T. Crepin, M. Jamin, and R. W. Ruigrok. 2010. Structure of the dimerization domain of the rabies virus phosphoprotein. J Virol 84:3707-10.

21. Tarbouriech, N., J. Curran, R. W. Ruigrok, and W. P. Burmeister. 2000. Tetrameric coiled coil domain of Sendai virus phosphoprotein. Nat Struct Biol 7:777-81.

22. Curran, J., H. Homann, C. Buchholz, S. Rochat, W. Neubert, and D. Kolakofsky. 1993. The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. J Virol 67:4358-64.

23. Longhi, S., V. Receveur-Brechot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, R. Yeo, S. Finet, and B. Canard. 2003. 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 278:18638-48.

24. Houben, K., D. Marion, N. Tarbouriech, R. W. Ruigrok, and L. Blanchard. 2007. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. J Virol 81:6807-16.

25. Mavrakis, M., A. A. McCarthy, S. Roche, D. Blondel, and R. W. Ruigrok. 2004. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J Mol Biol 343:819-31.

26. Cox, R., T. J. Green, S. Qiu, J. Kang, J. Tsao, P. E. Prevelige, B. He, and M. Luo. 2009. Characterization of a mumps virus nucleocapsidlike particle. J Virol 83:11402-6.

27. Cole, C., J. D. Barber, and G. J. Barton. 2008. The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36:W197-201.

28. Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54:905-21.

29. Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213-21.

30. Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486-501.

31. Painter, J., and E. A. Merritt. 2005. A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. Acta Crystallogr D Biol Crystallogr 61:465-71.

32. Painter, J., and E. A. Merritt. 2006. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 62:439-50.

33. Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 1.3r1.

34. Green, T. J., S. Macpherson, S. Qiu, J. Lebowitz, G. W. Wertz, and M. Luo. 2000. Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. J Virol 74:9515-24.

35. Karlin, D., F. Ferron, B. Canard, and S. Longhi. 2003. Structural disorder and modular organization in Paramyxovirinae N and P. J Gen Virol 84:3239-52.

36. Karlin, D., S. Longhi, V. Receveur, and B. Canard. 2002. The N-terminal domain of the phosphoprotein of Morbilliviruses belongs to the natively unfolded class of proteins. Virology 296:251-62.

37. Leyrat, C., M. R. Jensen, E. A. Ribeiro, Jr., F. C. Gerard, R. W. Ruigrok, M. Blackledge, and M. Jamin. 2011. The N(0)-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient alpha-helices. Protein Sci 20:542-56.

38. Tarbouriech, N., J. Curran, C. Ebel, R. W. Ruigrok, and W. P. Burmeister. 2000. On the domain structure and the polymerization state of the sendai virus P protein. Virology 266:99-109.

39. Choudhary, S. K., A. G. Malur, Y. Huo, B. P. De, and A. K. Banerjee. 2002. Characterization of the oligomerization domain of the phosphoprotein of human parainfluenza virus type 3. Virology 302:373-82.

40. Rahaman, A., N. Srinivasan, N. Shamala, and M. S. Shaila. 2004. Phosphoprotein of the rinderpest virus forms a tetramer through a coiled coil region important for biological function. A structural insight. J Biol Chem 279:23606-14.

41. Llorente, M. T., B. Garcia-Barreno, M. Calero, E. Camafeita, J. A. Lopez, S. Longhi, F. Ferron, P. F. Varela, and J. A. Melero. 2006. Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. J Gen Virol 87:159-69.

650 **42. Gerard, F. C., A. Ribeiro Ede, Jr., A. A. Albertini, I. Gutsche, G. Zaccai, R. W.**
651 **Ruigrok, and M. Jamin. 2007. Unphosphorylated rhabdoviridae**
652 **phosphoproteins form elongated dimers in solution. Biochemistry**
653 **46:10328-38.**

654 **43. Gerard, F. C., A. Ribeiro Ede, Jr., C. Leyrat, I. Ivanov, D. Blondel, S. Longhi,**
655 **R. W. Ruigrok, and M. Jamin. 2009. Modular organization of rabies virus**
656 **phosphoprotein. J Mol Biol 388:978-96.**

657 **44. Kolakofsky, D., L. Roux, D. Garcin, and R. W. Ruigrok. 2005. Paramyxovirus**
658 **mRNA editing, the "rule of six" and error catastrophe: a hypothesis. J Gen**
659 **Virol 86:1869-77.**

660

661 **Figure Legends**

662 **Figure 1.** Limited digestion of MuV P by different proteases. In (A-C), Lanes 1, 2, and 3
663 correspond to molecular weight ladder, undigested P protein, and digested P protein,
664 respectively. (A) Proteinase K digestion resulted in two stable fragments (K1, residues
665 194-391 and K2, residues 214-391) (Lane 3). (B) Trypsin digestion resulted in a major
666 stable fragment (T1, residues 135-391) (Lane 3). Lanes 4-7 correspond to the different
667 fractions separated using the Gelfree 8100 fractionation system. (C) Chymotrypsin
668 digestion resulted in a major stable fragment (C1, residues 194-391) (Lane 4). The
669 identity of the major stable fragments was revealed by N-terminal amino acid
670 sequencing and molecular weight determination, and is summarized in (D). The
671 molecular weights for fragments T1 and C1 were determined through mass
672 spectrometry. The molecular weights of fragments K1 and K2 were approximated from
673 their position on SDS-PAGE. Protease cut sites (*) represent the sequences identified
674 using N-terminal sequencing. (D) Summary of bioinformatics analysis using Jpred3.
675 Predicted helical regions are shown in the schematic to the right of “Helix”. Predicted

676 beta-strand regions are shown in the schematic to the right of “β-strand”. Homology to
677 PIV5 V protein is shown in the schematic to the right of “PIV5 V”.

678 **Figure 2.** Summary of MuV P fragments. A schematic representation of MuV P modules
679 is shown at the top. For expression levels, (+) indicates less than 15 mg/L, (++)
680 indicates 15-25 mg/L, and (+++) indicates greater than 25 mg/L of purified recombinant
681 protein. Fragments were derived from a combination of protease digestion and
682 bioinformatics analysis.

683 **Figure 3.** Structure of MuV P_{OD}. (A) The crystal structure of two parallel extended α-
684 helices is drawn (chain A in green and Chain B in cyan). Chain A contains residues 215-
685 277 and a kink at Gly246 is noted. Chain B contains residues 213-273. (B) The P_{OD}
686 tetramer as found in the crystal. The crystal structure of P_{OD} from SeV (C), VSV (D), and
687 RABV (E) is shown for comparison. The range of residues for each P_{OD} is labeled. In
688 (C-E), each independent polypeptide chain is a different color.

689 **Figure 4.** Tetrameric interactions. (A) Chain A:B interactions. In addition to extensive
690 surface contacts, the main interactions between chains A and B also include a zipper of
691 charged residues: Asp229, Glu236 and Asp240 (Chain A) and Arg231 and Lys238
692 (Chain B). In (A) Chain A is on the left and Chain B is on the right. The other two chains
693 in the tetramer are mono-colored. (B) Chain B:B’ interactions. Residues Lys253 and
694 Lys260 from Chain B (left) also form a zipper of charged residues, Asp229, Glu236 and
695 Asp 240 from the antiparallel Chain B’ (right). This clamping by the zippers occurs on
696 both ends of the tetramer. The tetramer in (B) is rotated left-handed by 90° compared to
697 the tetramer in (A).

698 **Figure 5.** Oligomerization state of MuV P. **(**A) Crosslinking with glutaraldehyde. MuV
699 P161-277 was crosslinked using glutaraldehyde in order to assess the oligomerization
700 state. Crosslinking products were analyzed in SDS-PAGE gels. Lanes 1-5 correspond
701 to 4 hr incubation with glutaraldehyde at the following concentrations: 0%, 0.05%,
702 0.06%, 0.1%, and 0.2%. Lanes 6 and 7 correspond to overnight incubation with 0.2%
703 and 0.3% glutaraldehyde respectively. Bands representing P161-277 monomers (P*1),
704 dimers (P*2), trimers (P*3), and tetramers (P*4) can be seen at ~15 kDa, ~30 kDa, ~45
705 kDa, and ~60 kDa, respectively, and are noted to the right of the gel. (B) Sedimentation
706 Velocity Analysis. Sedimentation coefficients were determined and molecular weights
707 were derived for P161-277 and full-length P (Table 3). The distribution plots showed a
708 single species for each sample. The derived molecular weight for P161-277 (57.8 kDa)
709 corresponded to the theoretical weight of a P161-277 tetramer (57.6 kDa). Furthermore,
710 the molecular weight derived for full-length P (165.5 kDa) was very close to the
711 theoretical molecular weight for a MuV P tetramer (168 kDa).

712 **Figure 6.** NLP pulldown assays. In order to determine which P fragments could interact
713 with NLP, pulldown assays were performed. The header of the gels in (A, B, D, and E)
714 indicate whether the P protein (or P fragments) saturated column was incubated with
715 NLP (+) or NLP_{379}, or not (-). Amino acid numbers corresponding to individual P
716 fragments are labeled on the footer of the gels. Positions for N protein [labeled (N) in
717 panels A and B], and trypsinized N [labeled (NLP379) in panels D and E] are noted (N)
718 (panels A and B), and trypsinized NLP (N_{379}) (panels D and E) are noted with arrows on
719 the right of each gel. (C) Negative Controls. Lanes 1 and 2 are the negative controls for
720 NLP and trypsinized NLP_{379}, respectively, with no P fragments added to the column.

721 **Figure 7**. Evaluation of N-P interactions using surface plasmon resonance. **(**A-D), full-
722 length NLP was immobilized, and P protein analytes (full-length P, P1-194, P286-391,
723 and P161-277 for A-D, respectively) were injected over the CM5 chip surface. All
724 analytes displayed a measurable binding response (1 RU = 1 pg/mm^2). *(*E-H*)* N_{379} was
725 immobilized, and P protein analytes (full-length P, P1-194, P286-391, and P161-277 for
726 E-H, respectively) were injected over the CM5 chip surface. All analytes displayed a
727 measurable binding response. All experiments were carried out in triplicate. K_A and K_D
728 values are listed in Table 4.

729 **Figure 8.** Diagram of the domain organization for four different NSV Ps. P_{OD} is typically
730 located in the central region of P. $P_{N°D}$ and P_{NBD} are linked to P_{OD} through flexible loops.
731 This appears to be a common modular structure for all NSV P proteins.

732

733 **Tables**

734

Table 1. Crystal X-ray data, phasing and refinement statistics.

Data Collection

	Native Crystal	**Uranyl derivative**
X-ray Source	BL-22 BM SER-CAT, APS, Illinois	BL-22 BM SER-CAT, APS, Illinois
Wavelength (Å)	1.0 Å	1.0 Å

Space group	R32	R32
Cell dimensions		
a,b,c (Å)	80.348, 80.348 , 165.17	78.919 , 78.919 , 165.447
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	50-2.2 (2.24-2.2)[1]	50-2.9 (2.96-2.9)
R_{sym}[2]	0.062 (0.38)	0.081 (0.47)
I/σ(I)	26.695 (4.145)	23.11 (3.0)
Completeness	99.9% (100%)	99.6% (97.6%)
Redundancy	4.6 (4.7)	3.5 (3.7)
No. of reflections	49,262	16,324
No. of unique reflections	10,717 (528)	4,601 (233)
Phasing Statistics		
Heavy atom sites		3 x $U(O)_2$
Resolution (Å)		50-2.56 (2.76-2.56)
Figure of merit acentric/centric		0.6566/0.3542 (0.6584/0.3531)
Phasing power (iso/ano)		1.3471/1.3104 (1.2658/1.4415)
R_{cullis} (iso/ano)		0.599/0.7606 (0.6059/0.9150)
Refinement		
Molecules (ASU)	2	

Resolution (Å)	26.6-2.2 (2.31-2.2)
Average B-factor (Å^2) (protein/water)	43.05/46.52
Coordinate Error (max. likelihood based)	0.26
No. of Reflections used	10,308
R_{work}[3]/R_{free}[4]	0.1814/0.2302
Completeness	96.15%
No. atoms	
Protein	942
Water	58
Glycerol	1
rms deviations	
Bond lengths (A)	0.006
Bond angles (degrees)	0.887
Ramachandran favored	98.3%

735 [1]Values for high resolution shell in parentheses

736 [2]Rsym = Σ |(I-<I>)|/ Σ <I> where <I> is the observed intensity

737 [3]R = Σ(||Fobs|-k|Fmodel||)/Σ(|Fobs|)

738 [4]Rfree is obtained for a test set of reflections (9.98%).

739

Table 2. Surface Area Calculations.

Chain ID	Buried Surface (Å^2)	Total Surface Area (Å^2)	Percent Buried (%)
Chain A		6,382	
Chain B		5,824	
A-B	2,771	9,436	29.3
A-A'	2,570	10,193	25.2
B-B'	2,601	9,050	28.7
Tetramer	5,065	13,809	36.7

740

Table 3. Sedimentation Velocity Statistics.

Sample ID	S value	RMSD	f/f0	MW (kDa)
P161-277	2.6	0.009049	1.8525	57.8
P full	3.8	0.01062	2.6522	165.5

741

Table 4. Binding Kinetics

Sample ID	K_A (M^{-1})	K_D (M)
P1-194 + N full	2.02×10^5	4.95×10^{-6}
P286-391 + N full	2.35×10^6	4.26×10^{-7}
P full + N full	2.23×10^6	4.48×10^{-7}
P1-194 + N_{379}	2.17×10^6	4.60×10^{-7}
P286-391 + N_{379}	1.01×10^5	9.88×10^{-6}

P full + N_{379}	1.61×10^{6}	8.65×10^{-7}

742

743

744 **Nomenclature**

745 P - phosphoprotein

746 MuV – mumps virus

747 NSV – negative sense virus

748 RdRp – RNA dependent RNA polymerase

749 POD – phosphoprotein oligomerization domain

750 PNBD – phosphoprotein nucleocapsid binding domain

751 PN°D – phosphoprotein N0 binding domain

752 N – nucleocapsid protein

753 VSV – vesicular stomatitis virus

754 RABV – rabies virus

755 SeV – sendai virus

756 NLP – nucleocapsid like particle

757

758 **Abbreviations**

759 SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

760 CV – column volume

761 mg/L – milligrams per liter

762 ug - microgram

763 ul - microliter

764 Da – Dalton

765 Å – angstrom

766 RU – response unit

767

768