Crystal Structure of the Measles Virus Nucleoprotein Core in Complex with an N-Terminal Region of Phosphoprotein

Sergey G. Guryanov, Lassi Liljeroos, Prasad Kasaragod, Tommi Kajander, Sarah J. Butcher

Department of Biological Sciences and Institute of Biotechnology, University of Helsinki, Helsinki, Finland

ABSTRACT

The enveloped negative-stranded RNA virus measles virus (MeV) is an important human pathogen. The nucleoprotein (N\textsuperscript{0}) assembles with the viral RNA into helical ribonucleocapsids (NC) which are, in turn, coated by a helical layer of the matrix protein. The viral polymerase complex uses the NC as its template. The N\textsuperscript{0} assembly onto the NC and the activity of the polymerase are regulated by the viral phosphoprotein (P). In this study, we pulled down an N\textsuperscript{0}-1,408 fragment lacking most of its C-terminal tail domain by several affinity-tagged, N-terminal P fragments to map the N\textsuperscript{0}-binding region of P to the first 48 amino acids. We showed biochemically and using P mutants the importance of the hydrophobic interactions for the binding. We fused an N\textsuperscript{0} binding peptide, P\textsubscript{1-48}, to the C terminus of an N\textsuperscript{0} 21-408 fragment lacking both the N-terminal peptide and the C-terminal tail of N protein to reconstitute and crystallize the N\textsuperscript{0}-P complex. We solved the X-ray structure of the resulting N\textsuperscript{0}-P chimeric protein at a resolution of 2.7 Å. The structure reveals the molecular details of the conserved N\textsuperscript{0}-P interface and explains how P chaperones N, preventing both self-assembly of N\textsuperscript{0} and its binding to RNA. Finally, we propose a model for a preinitiation complex for RNA polymerization.

IMPORTANCE

Measles virus is an important, highly contagious human pathogen. The nucleoprotein N binds only to viral genomic RNA and forms the helical ribonucleocapsid that serves as a template for viral replication. We address how N is regulated by another protein, the phosphoprotein (P), to prevent newly synthesized N from binding to cellular RNA. We describe the atomic model of an N-P complex and compare it to helical ribonucleocapsid. We thus provide insight into how P chaperones N and helps to start viral RNA synthesis. Our results provide a new insight into mechanisms of paramyxovirus replication. New data on the mechanisms of phosphoprotein chaperone action allows better understanding of virus genome replication and nucleocapsid assembly. We describe a conserved structural interface for the N-P interaction which could be a target for drug development to treat not only measles but also potentially other paramyxovirus diseases.

Measles virus (MeV) belongs to the Paramyxoviridae family, which includes several other human pathogens, like respiratory syncytial (RSV), mumps, and parainfluenza viruses. It has a helical ribonucleocapsid (NC) containing a nonsegmented single-strand RNA (ssRNA) genome wrapped around the outside the nucleoprotein (N) helix (1). The helical NC is active in both transcription and replication. During virus assembly, the matrix protein forms an additional helix covering the majority of the NC, potentially inhibiting transcription and promoting packaging into progeny virions (2). There are still only limited data on the detailed molecular interactions required to go from replication initiation to packaging of nascent RNA. The availability of N in a chaperoned, assembly-competent state with the phosphoprotein (P) versus the assembled helical state is thought to be critical to these processes. N is composed of an ordered N\textsubscript{CORE} region (amino acids 1 to 391) and an intrinsically disordered N\textsubscript{TAIL} region (amino acids 392 to 525) (Fig. 1A). N\textsubscript{CORE} contains two domains (N\textsubscript{STD} and N\textsubscript{CTD}) flanked by N- and C-terminal arms (N\textsubscript{ARM} and C\textsubscript{ARM}). A recent atomic model of the MeV NC from a cryo-electron microscopic (cryo-EM) reconstruction revealed the molecular details of N oligomerization mediated by exchange of the N\textsubscript{ARM} and C\textsubscript{ARM} between consecutive N monomers and showed the RNA-binding site on the groove between the two N\textsubscript{CORE} domains (1).

P is a modular protein comprised of an ordered tetramerization domain, MD (amino acids 304 to 377), forming a parallel four-helix coiled-coil (3), and an extreme C-terminal domain (CTD), XD, alternating with disordered regions (Fig. 1A). For transcription and replication, the RNA polymerase (L) in complex with P attaches to the NC via an interaction between the XD domain in P and the molecular recognition element (MoRE) (Fig. 1A) in N (4–7). The three-helix bundle in XD binds a helix from N’s MoRE element to facilitate this interaction (7). P has a second role: it binds N\textsubscript{CORE} through its N-terminal soyuz1 motif (8) and performs a chaperone function required to keep newly synthesized N from binding to cellular RNA (9). This RNA-free N\textsuperscript{0} is then transferred from the N\textsuperscript{0}-P complex to the nascent NC by a currently unknown mechanism.

Whereas the XD and MD domains of P have been well characterized, the interaction between the P N terminus and the N in the
RNA binding groove; therefore, the chaperoning mechanism of P appears to be remarkably different from that in VSV. In the present study, we addressed MeV N⁰-P complex formation and structure. We expressed and purified MeV N⁰-P complexes from *Escherichia coli* in a monodisperse form and mapped the location of the N-binding region on P to the first 48 N-terminal amino acids. Then we designed a chimeric N-P protein that was crystallized to reconstitute the N⁰-P complex and solved the structure at a resolution of 2.7 Å. We also characterized the mode of interaction between the P N terminus and N⁰ and showed the importance of hydrophobic interactions. Based on the structural data, we describe conformational changes upon RNA binding and propose a model for the preinitiation complex for RNA replication and transcription.

**MATERIALS AND METHODS**

**Cloning and expression.** All constructs were derived from reverse-transcribed N and P genes of an MeV wild-type isolate (a gift from I. Davidkin, Helsinki, Finland) (2). The N coding sequence was identical to the GenBank sequence for the Halonen strain (accession number U01996). The P coding sequence differed from GenBank sequence AF266288 for the Edmonton strain by three nucleotides: one was synonymous, and the other two resulted in the amino acids G225 and D492. Truncated P constructs were generated by PCR and cloned into Ncol and Xhol sites in pET41(a). The P constructs had an N-terminal glutathione S-transferase followed by a hexahistidine sequence (GST-H₆ tag) for purification. N₁-408 and N₂₁-408 were constructed similarly but were cloned into pET22(b) with a stop codon added to the 3′ end and did not contain any tags. For the N₁₂₁-408–P₁-48 chimera, the N₁₂₁-408 and P₁-48 coding sequences were amplified by PCR to generate megapeptides with overlapping sequences. Then the megapeptides were annealed and extended. The product was amplified with primers coding for a N-terminal H₁₂₅-TEV tag, N₂₁-408–P₁-48 and N₁₂₁-408–P₁-48 were coexpressed in *E. coli* Rosetta (DE3) (Merck Millipore). Expression was induced at an optical density at 600 nm (OD₆₀₀) of 0.5 with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and was allowed to proceed for 16 to 20 h at 22°C. Cells were collected by low-speed centrifugation and frozen at −80°C as pellets until use.

**Protein purification.** N₁-408 and GST-H₆–P₁-48 were coexpressed in *E. coli* Rosetta (DE3). The cell pellet was resuspended in buffer A (20 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, 2 mM CaCl₂ [pH 8.0]) supplemented with 200 μg/ml of lysozyme and one EDTA-free protease inhibitor tablet/25 ml (Thermo Scientific). Cells were lysed with a French press at 22,000 lb/in², cell debris was spun down by low-speed centrifugation (11,000 × g for 15 min at 4°C), and the resulting supernatant was incubated with Ni-loaded IMAC beads (GE Healthcare) for 45 min at room temperature. After a washing with buffer A, the beads were exchanged into buffer B (20 mM Tris-HCl, 150 mM NaCl, 10 mM MgCl₂, 2 mM ATP [pH 8.0]) and incubated at 37°C for 10 min. Next, the beads were exchanged into buffer C (20 mM Tris-HCl, 150 mM NaCl, 2 mM CaCl₂ [pH 8.0]). The N₁₂₁-408–P₁-48 heterocomplex was then released from the beads by an overnight digestion with enterokinase light chain (New England BioLabs). For 2 ml of Ni-IMAC beads with protein from 1 liter of cell culture, 0.16 g of enzyme was used. The released protein was then concentrated with Millipore Amicon Ultra-4 30-kDa-cutoff spin concentrators and polished with size exclusion chromatography (SEC) using a Superdex 200 column (GE Healthcare). Peak fractions were collected and concentrated to the desired concentration with the same concentrator.

N⁰-P complex is less well described. The dual function for P has been established for many viruses of the *Mononegavirales* order, and the crystal structures of the vesicular stomatitis virus (VSV), Ebola virus, and Nipah virus (NiV) N⁰-P complexes have been solved (10–12). In VSV N⁰-P, the P N-terminal amino acids 17 to 31 formed an amphipathic α-helix and occupied a hinge region in N adjacent to the RNA-binding site (13). In NiV N⁰-P, P amino acids 1 to 35 formed two α-helices separated by a kink (11). Interestingly, the NiV P binding site does not overlap the predicted

**FIG 1**  Protein constructs and N⁰-P complex analysis. (A) Domain structure of measles N and P proteins and protein constructs used in this study. Numbers refer to amino acid positions. (B) Coexpressed N and GST-H₆–P constructs after elution from glutathione Sepharose beads. Lanes: 1, marker; 2, P₁–38 and N₁–408; 3, P₁–48 and N₁–408; 4, P₁–58 and N₁–408; 5, P₁–68 and N₁–408; 6, P₁–78 and N₁–408; 7, GST-H₆ and N₁–408. Numbers on the left are molecular size markers. (C) Electron microscopy of negatively stained N₁₂₁-408–GST-H₆–P₁-48 heterocomplex (left) and flowthrough from the glutathione Sepharose beads containing NC-like helical particles (right). Images are at the same scale. Scale bar, 100 nm. (D) Data from the SEC-MALLS experiment with the N₁₂₁-408 and P₁-48 chimera. The solid line is the absorbance profile; blue crosses show the molecular mass distribution. For comparison, the UV₂₅₀ Profile of the N₂₁-408–P₁-48 chimera is shown with a dashed line.
tagged chimera was resuspended in buffer D (20 mM Tris-HCl, 150 mM NaCl, 2 mM MgCl₂, 20 mM imidazole [pH 8.0]) supplemented with 10 μg/ml of lysozyme, 1 μg/ml of DNase I (Sigma-Aldrich), and 0.5 mM Pefabloc (Roche). Cells were lysed, and the lysate was cleared and incubated with Ni-IMAC as described above. After a washing with buffer D, protein was eluted with buffer D supplemented with 0.2 M imidazole. Eluted protein was incubated with TEV protease (purified in-house) overnight at 4°C. Cleaved protein was purified on a Superdex 200 column in buffer E (20 mM Tris-HCl, 150 mM NaCl [pH 8.0]), the monomer peak was collected, and uncleaved protein was removed by passing through an Ni-IMAC column. Purified protein was concentrated as described above.

**N⁰-P heterocomplex interaction experiments.** To find the minimal length of N that stably interacted with N₁-408, 5 different P constructs were coexpressed with N₁-408 as described above. The cells were lysed by sonication in phosphate-buffered saline (PBS) supplemented with 0.17 mg/ml of lysozyme. The cleared lysates were incubated with glutathione beads in PBS for 30 min at room temperature. After 3 washes with PBS, the proteins were eluted with 10 mM reduced glutathione in 50 mM Tris-HCl (pH 8.0).

To test the stability of the interactions, the purified N₁-408-GST-H₆-P₁-48 heterocomplex was bound in 50 mM sodium phosphate (pH 7.4), 10 mM imidazole, and 300 mM NaCl to Ni-IMAC beads and eluted with 20 mM Tris (pH 8.0) supplemented with one of the following: NaCl at 0, 0.5, 1, or 2 M; KCl at 0.5, 1, or 2 M; urea at 2, 4, or 8 M; or Trition X-100 at 0.1% or 1%. To probe N₂₁-408 heterocomplex formation with mutated GST-H₆-P constructs, cell lysates of individually expressed proteins were mixed and incubated overnight at 4°C. Lysate samples were incubated with Ni-IMAC beads and washed with lysis buffer E. The samples were eluted with buffer E supplemented with 0.2 M imidazole. Eluates were analyzed by SDS-PAGE.

**SEC-MALLS experiment.** To analyze the exact stoichiometry of the P₁-38-N₁-408 heterodimeric complex, 47 μl of the complex (1 mg/ml) released by enterokinase digestion was run on a Superdex 200 10/300 GL column in buffer E supplemented with 0.2 M imidazole (pH 8.0). 5% (wt/vol) sodium phosphotungstate (pH 7.0). Grids were imaged with an FEI T200 transmission electron microscope, and images were collected with a Gatan Ultrascan 4000 charge-coupled-device (CCD) camera.

**Structure determination.** Crystals of the N₂₁-408-P₁-48 chimeric protein were grown by sitting-drop vapor diffusion (2°C) by mixing 200 nl of protein (8 mg/ml) with 200 nl of reservoir (0.1 M sodium citrate [pH 5.2], 3% polyethylene glycol 8000). Crystals were cryoprotected in mother liquor containing 20% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected at the Diamond Light Source beamline I03. The data set was processed and scaled using the xia2 program (with DLS and AIMLESS) (15, 16). A summary of the data collection is given in Table 1. The structure was solved by molecular replacement using PHASER (17). The N-terminal domain (NTD) and C-terminal domain (CTD), based on mass spectrometric analysis of the fragments observed, were used as search models separately. The model was rebuilt using several cycles of autobuilding and refinement with PHENIX (18) and manual rebuilding with COOT (19). No density was observed for the N regions from 21 to 30, 119 to 120, and 133 to 139 and the P region from 39 to 48, and therefore, they were left out of the model. The last refinement cycles were done using TLS models (two TLS groups). The final refinement statistics are summarized in Table 1. The R factors (Rwork/Rfree) of the refined structure are 21.1%/26.6% (Table 1). In the Ramachandran plot, 90% of the residues in the structure are in the most favored regions.

**Structure analysis.** All the structure illustrations were prepared using UCSF Chimera software (20). Interface surface was estimated using the PDBePISA server (21). Calculation of the relative angle between the N domains in N⁰-P versus NC structure (PDB code 4UFT) was done using Modeller software (22) as described earlier (23). Structure alignments and root mean square deviation (RMSD) value calculations were made using UCSF Chimera. Dali multiple structural alignment (24) was used to generate the corresponding primary sequence alignment, followed by phylogenetic tree generation by PHYLIP in Unipro UGENE software (25).

**Protein structure accession number.** Final refined coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession code 5E4V.

### RESULTS

#### Mapping the interaction of the MeV NCORE⁰ with the N-terminal region of P and crystallization of the complex.

In order to obtain a well-structured N⁰-P complex, we analyzed the protease sensitivity of N by limited trypsin proteolysis (data not shown). Based on mass spectrometric analysis of the fragments observed, we cloned a C-terminally truncated construct containing the first 408 amino acids, N₁-408, thus excluding the disordered C-terminal Tₐₙ region (7).

To screen for NCORE interaction in the N-terminal region of P, we used coexpression of N₁-408 together with GST-hexahistidine (GST-H₆) fusions with P₁-38, P₁-48, P₁-58, P₁-68, or P₁-78 (Fig. 1A) and analyzed the interactions by GST affinity chromatography. All of the P constructs readily interacted with N₁-408 and could be
Hence, we designed a chimeric construct, H6–TEV–N21-408. After extensive efforts, the heterodimeric complex failed to crystallize. The gel filtration mobility (Fig. 1D) and the additional dimer peak. The solution state of the N21-408 chimera suggests that the P1-48 sequence is bound to the P-binding site with NCORE resides within the first 38 amino acids. We analyzed negatively stained N1-408 and N21-408 eluate and the P N-terminal region in some paramyxoviruses has been described previously (8, 11). In the MeV NCORE-P complex, the binding interface is mostly composed of conserved hydrophobic residues (Fig. 3A and B, in blue). To biochemically probe the binding of the P N-terminal region to NCORE, we screened for dissociation of N1-408 from GST-H6-P1-48 bound on Ni-IMAC beads under different conditions and looked for release of N1-408. The screen was designed to include conditions which would hinder either ionic or hydrophobic interactions between the proteins. NaCl or KCl concentrations ranging between 0 and 2 M did not cause significant release of N1-408, whereas 0.1% and 1% Triton X-100 caused release of N1-408 from the complex (Fig. 3C) to levels similar to those obtained with 4 and 8 M urea, respectively.

To evaluate the role of P’s hydrophobic amino acid residues in N0 binding, we expressed N21-408 lacking the NTarm. We found that N21-408 was insoluble and formed NC-like particles upon expression (data not shown), similar to N1-408, but could bind to P1-48 in vitro. We pulled down N21-408 with GST-H6-tagged wild-type P1-48 and its mutants (Fig. 3D). While wild-type P1-48 can efficiently bind N21-408, replacements of hydrophobic by negatively charged amino acids (L13D, 116D, and L19E) severely affected the interaction. Replacement of L13 by the small amino acid Ala also strongly affected the interaction, possibly due to the increased solvent accessibility of the binding interface. In line with this observation, we still observed residual binding of the shorter P1-48 peptide where most of the interacting hydrophobic amino acids were retained (Fig. 3D, lane 9). Thus, hydrophobic interactions make a major contribution to P N-terminal region and NCORE binding.

**Comparison of the MeV RNA-bound helical form with the chaperoned form.** Direct comparison of the MeV NCORE-P and helical NC structures reveals several factors that could contribute to P’s chaperone activity; these include conformational changes (Fig. 4) as well as the position of P and RNA binding (Fig. 5 and 6). The largest difference between the two structures is that there is a relative domain movement in the MeV N in the NCORE-P complex, compared to the helical RNA-bound NC form (Fig. 4A and B) (1). By aligning either the CTD or NT domain only, we measured an ~40° relative rotation of the two domains, with the hinge occurring between α9 and α10 (Fig. 4F). The RMSD for the individual domains in the two different MeV N conformations were calculated (Table 2). This comparison indicated that the 4.3-Å resolution cryo-EM structure agrees well with our crystal structure, and the changes seen between the structures could be interpreted reliably (Fig. 4A and B). Besides bending of the hinge between helices α9 and α10, helix α6 forming the lower lobe of the RNA binding cleft differs between the two different states. The helix undergoes both a shift and a rotation around its axis by half a turn (Fig. 4D). The helix movement increases the proximity of the two

**Interaction of NCORE with P is mainly hydrophobic.** Conservation of the NCORE binding interface for P and of the P N-terminal region in some paramyxoviruses has been described previously (8, 11). In the MeV NCORE-P complex, the binding interface is mostly composed of conserved hydrophobic residues (Fig. 3A and B, in blue). To biochemically probe the binding of the P N-terminal region to NCORE, we screened for dissociation of N1-408 from GST-H6-P1-48 bound on Ni-IMAC beads under different conditions and looked for release of N1-408. The screen was designed to include conditions which would hinder either ionic or hydrophobic interactions between the proteins. NaCl or KCl concentrations ranging between 0 and 2 M did not cause significant release of N1-408, whereas 0.1% and 1% Triton X-100 caused release of N1-408 from the complex (Fig. 3C) to levels similar to those obtained with 4 and 8 M urea, respectively.

To evaluate the role of P’s hydrophobic amino acid residues in N0 binding, we expressed N21-408 lacking the NTarm. We found that N21-408 was insoluble and formed NC-like particles upon expression (data not shown), similar to N1-408, but could bind to P1-48 in vitro. We pulled down N21-408 with GST-H6-tagged wild-type P1-48 and its mutants (Fig. 3D). While wild-type P1-48 can efficiently bind N21-408, replacements of hydrophobic by negatively charged amino acids (L13D, 116D, and L19E) severely affected the interaction. Replacement of L13 by the small amino acid Ala also strongly affected the interaction, possibly due to the increased solvent accessibility of the binding interface. In line with this observation, we still observed residual binding of the shorter P1-48 peptide where most of the interacting hydrophobic amino acids were retained (Fig. 3D, lane 9). Thus, hydrophobic interactions make a major contribution to P N-terminal region and NCORE binding.

**Comparison of the MeV RNA-bound helical form with the chaperoned form.** Direct comparison of the MeV NCORE-P and helical NC structures reveals several factors that could contribute to P’s chaperone activity; these include conformational changes (Fig. 4) as well as the position of P and RNA binding (Fig. 5 and 6). The largest difference between the two structures is that there is a relative domain movement in the MeV N in the NCORE-P complex, compared to the helical RNA-bound NC form (Fig. 4A and B) (1). By aligning either the CTD or NT domain only, we measured an ~40° relative rotation of the two domains, with the hinge occurring between α9 and α10 (Fig. 4F). The RMSD for the individual domains in the two different MeV N conformations were calculated (Table 2). This comparison indicated that the 4.3-Å resolution cryo-EM structure agrees well with our crystal structure, and the changes seen between the structures could be interpreted reliably (Fig. 4A and B). Besides bending of the hinge between helices α9 and α10, helix α6 forming the lower lobe of the RNA binding cleft differs between the two different states. The helix undergoes both a shift and a rotation around its axis by half a turn (Fig. 4D). The helix movement increases the proximity of the two

**Interaction of NCORE with P is mainly hydrophobic.** Conservation of the NCORE binding interface for P and of the P N-terminal region in some paramyxoviruses has been described previously (8, 11). In the MeV NCORE-P complex, the binding interface is mostly composed of conserved hydrophobic residues (Fig. 3A and B, in blue). To biochemically probe the binding of the P N-terminal region to NCORE, we screened for dissociation of N1-408 from GST-H6-P1-48 bound on Ni-IMAC beads under different conditions and looked for release of N1-408. The screen was designed to include conditions which would hinder either ionic or hydrophobic interactions between the proteins. NaCl or KCl concentrations ranging between 0 and 2 M did not cause significant release of N1-408, whereas 0.1% and 1% Triton X-100 caused release of N1-408 from the complex (Fig. 3C) to levels similar to those obtained with 4 and 8 M urea, respectively.

To evaluate the role of P’s hydrophobic amino acid residues in N0 binding, we expressed N21-408 lacking the NTarm. We found that N21-408 was insoluble and formed NC-like particles upon expression (data not shown), similar to N1-408, but could bind to P1-48 in vitro. We pulled down N21-408 with GST-H6-tagged wild-type P1-48 and its mutants (Fig. 3D). While wild-type P1-48 can efficiently bind N21-408, replacements of hydrophobic by negatively charged amino acids (L13D, 116D, and L19E) severely affected the interaction. Replacement of L13 by the small amino acid Ala also strongly affected the interaction, possibly due to the increased solvent accessibility of the binding interface. In line with this observation, we still observed residual binding of the shorter P1-48 peptide where most of the interacting hydrophobic amino acids were retained (Fig. 3D, lane 9). Thus, hydrophobic interactions make a major contribution to P N-terminal region and NCORE binding.

**Comparison of the MeV RNA-bound helical form with the chaperoned form.** Direct comparison of the MeV NCORE-P and helical NC structures reveals several factors that could contribute to P’s chaperone activity; these include conformational changes (Fig. 4) as well as the position of P and RNA binding (Fig. 5 and 6). The largest difference between the two structures is that there is a relative domain movement in the MeV N in the NCORE-P complex, compared to the helical RNA-bound NC form (Fig. 4A and B) (1). By aligning either the CTD or NT domain only, we measured an ~40° relative rotation of the two domains, with the hinge occurring between α9 and α10 (Fig. 4F). The RMSD for the individual domains in the two different MeV N conformations were calculated (Table 2). This comparison indicated that the 4.3-Å resolution cryo-EM structure agrees well with our crystal structure, and the changes seen between the structures could be interpreted reliably (Fig. 4A and B). Besides bending of the hinge between helices α9 and α10, helix α6 forming the lower lobe of the RNA binding cleft differs between the two different states. The helix undergoes both a shift and a rotation around its axis by half a turn (Fig. 4D). The helix movement increases the proximity of the two
sides of the interdomain cleft in NCORE\textsuperscript{0-P} and therefore collapses the NC RNA-binding site. In addition, the surface electrostatic charge distribution changes quite dramatically depending on the \textit{H9251 N6} conformation. In NCORE\textsuperscript{0-P}, a new negatively charged groove is evident on the NTD surface that could potentially bind RNA (Fig. 5). It has a contribution from the conserved \textit{R194} that interacts with the RNA backbone in NC (1). In NC, \textit{Y199} stacks with \textit{Y260}, a key residue that regulates RNA binding pocket size. In NCORE\textsuperscript{0-P}, \textit{Y199} faces the exterior and \textit{W196} occupies the space instead, thus potentially participating in the local stacking configuration.

Alignment of the NCTD domains of NCORE\textsuperscript{0-P} and NC (1) models (Fig. 4A and C) shows that the P N termini of MeV, NiV, and PIV5. Asterisks indicate residues making contacts with N\textsuperscript{0}. Conserved residues have a colored background. (C) SDS-PAGE of protein released from the N_{1\textsuperscript{408}-\text{GST-H}_{1\textsuperscript{P_{48}}}} heterodimer complex bound to Ni-IMAC beads when subjected to different conditions. Lanes: 1, marker; 2, 0 M NaCl; 3, 0.5 M NaCl; 4, 1 M NaCl; 5, 2 M NaCl; 6, 0.5 M KCl; 7, 1 M KCl; 8, 2 M KCl; 9, 2 M urea; 10, 4 M urea; 11, 8 M urea; 12, 0.1% Triton X-100; 13, 1% Triton X-100. (D) SDS-PAGE of pull-down of N_{1\textsuperscript{408}} by GST-H\textsubscript{1\textsuperscript{P_{48}}} and its mutants. Lanes: 1, markers; 2, control N_{1\textsuperscript{408}} only; 3, control GST-H\textsubscript{1\textsuperscript{P_{48}}}wt only; 4, N_{1\textsuperscript{408}} plus GST-H\textsubscript{1\textsuperscript{P_{1\textsuperscript{48}}}}wt; 5, N_{1\textsuperscript{408}} plus GST-H\textsubscript{1\textsuperscript{P_{1\textsuperscript{48}}}}L13D; 6, N_{1\textsuperscript{408}} plus GST-H\textsubscript{1\textsuperscript{P_{1\textsuperscript{48}}}}L13A; 7, N_{1\textsuperscript{408}} plus GST-H\textsubscript{1\textsuperscript{P_{1\textsuperscript{48}}}}I16D; 8, N_{1\textsuperscript{21-408}} plus GST-H\textsubscript{1\textsuperscript{P_{8-48}}}. Note that some of the P mutations affected the electrophoretic mobility of the tagged constructs due to impaired SDS binding to the protein molecules with changed net charge (compare lanes 5 and 7 with lane 3). The deletion construct P\textsubscript{8-48} migrates faster, reflecting its shorter amino acid sequence. Numbers on the left in panels C and D are molecular size markers.

Comparison to other virus nucleoproteins. Comparison of the N structures from MeV and other viruses of the Mononegavirales order reveals their structural similarity (Fig. 7). Each protein...
that the RNA-binding site rather than binds on the opposite side to PIV5 N. 32% amino acid sequence identity to NiV N and 24% identity to VSV N, in VSV N, the RSMD reported between the two states was less than 1 Å (10), reflecting the fact that both VSV states were crystallized in a ring form that probably constrained the conformation. In contrast to MeV, VSV P blocks the RNA-binding site rather than binds on the opposite side of the molecule. For the other N0-P complexes reported, NiV N shows 32% amino acid sequence identity to NiV N and 24% identity to PIV5 N.

To our knowledge, MeV and VSV are the only Mononegavirales members with both the N0-P and N-RNA complexes available. Compared to the 40° rotation transition between the N0-P and the RNA-bound states in MeV N, in VSV N, the RSMD reported between the two states was less than 1 Å (10), reflecting the fact that both VSV states were crystallized in a ring form that probably constrained the conformation. In contrast to MeV, VSV P blocks the RNA-binding site rather than binds on the opposite side of the molecule. For the other N0-P complexes reported, NiV N is the closest, with a binding site similar to that for MeV N0-P (Fig. 3 and 6), and relative domain positioning (referred to as an “open conformation” in NiV [11]). None of the other RNA-bound states are from a helical NC structure; rather, they are all ring structures, but in PIV5, RSV, and rabies virus, the RNA-bound states also indicate domain positioning similar to that of MeV NC (“closed state”) (11, 27). These comparisons emphasize the importance of the flexibility in the interdomain region in regulating N’s interactions with other viral components.

DISCUSSION
Here, for the first time in paramyxovirus research, our data allow direct comparison of the structures of the nucleoprotein from the same virus in two functional states: a P-bound naive state and an RNA-bound helical assembly. Our X-ray crystallographic atomic model at higher resolution confirms, complements, and improves upon the recently published cryo-EM reconstruction (1). The overall fold of MeV N is most similar to those of NiV and PIV5 N proteins (Fig. 7) (11, 27). In addition, the structure of MeV P1-48 and its binding site are very similar to those of NiV N0-P complex (11). Noting the extensive, conserved hydrophobic interactions of the P protein helix αP1 and N (8, 11), we showed by mutation of the hydrophobic residues in P and by biochemical analyses that we could indeed impair the binding interaction.

Our findings suggest that both N domains mostly preserve their fold upon transition from the N0-P to the NC state. Notably, N in both N0-P and NC has a flexible region between residues 118 and 140 composed of a well-defined αN helix (residues 124 to 130) flanked by unresolved regions. In NiV N, the αN helix is longer and only one unresolved region was left, whereas in PIV5, there are no gaps here. This region is on the outer surface of the NC. Hence, this flexible region could interact with the flexible C terminus of N or with the polymerase complex.

How does P act as a chaperone? The roles of P are at least 2-fold: first, to act as a chaperone to keep N0 in its RNA-free, soluble, monomeric form, and second, to position the polymerase complex for polymerization. In the role as a chaperone, it has been proposed that in the NiV N0-P complex, binding of P to N locks the open conformation by rigidifying the NC CTD structure (11). Our model, however, suggests a significant impact from steric interference between the P N terminus and NC CTD (Fig. 6). Alignment of the NC CTD domains of the MeV N0-P and NC (1) models (Fig. 4A and C) shows that the P N terminus will clash with helix αN of N in the RNA-bound conformation, thus favoring the RNA-free conformation in N0-P. The flexibility of the N molecule is therefore an inherently important part of our model, compared to the published NiV model (11). We have additional evidence that in the NC, N can assume different conformations. The pitch of the protease-treated NC used for high-resolution structure determination is 5.0 nm (1) and imposes a rigidity on the helix that was important for image processing. However, the recombinant full-length protein forms flexible helices with pitches ranging from 5.0 to 6.6 nm (28, 29), and those imaged inside virions have a pitch of 6.4 nm (2). In the latter, the rigidity of the NC helix is enforced by interaction with an outer layer of matrix protein. Where the matrix is lacking, the NC is flexible. From the current work, at least two flexible regions could affect the twist and pitch, the twisting of the two domains (28), induced by the interdomain hinge region described above, and the conformation of the αN helix. Confident assignment of amino acids Trp196 and Tyr199 in this helix in both the cryo-EM and X-ray electron densities showed rotation and elongation of the αN helix (Fig. 4D), reflecting the intrinsic flexibility in this part of the molecule. Noticeably, in both the NiV N0-P and PIV5 N-RNA structures (11, 27), a loop preceding the corresponding helix is unresolved; this loop flexibility further supports the inherent mobility of the αN helix.
What determines whether N binds P or RNA? The RNA binding in the NC state is favored by specific arrangement of the amino acid residues from both the N_{NTD} and N_{CTD} (1). In the N_{0}-P state, the rotation of the N domains forces the overlap of these two binding surfaces; hence, we hypothesize that the RNA binding affinity is reduced. This is supported by two observations. First, regarding the solubility of the chimera in an E. coli cell lysate, we found a predominance of RNA-free monomers and dimers rather than helical assemblies, even in the presence of E. coli RNA, compared to what was found with N expressed alone. Hence, the N_{0}-P interaction hampered NC assembly and binding to RNA in our study. Second, the surface charge distribution of the chimera is altered, changing and shrinking the position of the positively charged surface in the N_{0}-P compared to the NC. This suggests that N’s affinity for P in our constructs was higher than for RNA. There is probably a balance in the cell, during infection, dictated by the local concentrations of the relevant components and the avidity of N for RNA and its neighboring N subunits that together orchestrate the assembly of the NC. The flexibility of N facilitates its exchange between its binding partners, P and RNA. P can further regulate helix assembly through sterically impeding both side-to-side and vertical growth of the helix through occupying the same sites as both the NT_{arm} of the N_{i-1} protomer and CT_{arm} of the N_{i+1} protomer (Fig. 6).

Model for the formation of a preinitiation complex. According to the current paramyxovirus models, both transcription and replication are initiated at the 3’ end of the genomic RNA (30). The linear unidirectional organization of the “herringbone” NC means that the 3’ and 5’ ends of the NC do not present the same molecular surface due to this polarity. In addition, the transition between the bulk of the helix to the tip means that there is an extra potential site for P binding on the last molecule of the NC at the 3’ end (Fig. 8). The specific architecture of the pointed 3’ end of the NC could thus facilitate the recognition of the initiation site and assembly of the preinitiation complex through the interactions of P, L, N, and RNA. We propose a simple model for formation of a preinitiation complex, as shown in Fig. 8. In this model, the first interactions occur between the RNA polymerase complex, L-P, and NC through P’s XD domains (Fig. 1A) in a low-affinity interaction with the flexible extended N_{TAIL} (5, 31). This transient interaction allows one-dimensional diffusion of the polymerase complex along the NC. The accurate positioning on the tip occurs when the P N-terminal region binds to a vacant NT_{arm} binding site at the NC’s 3’ end. Binding of P may initiate NC uncoiling, as has been observed with mumps virus (32), to facilitate the release of the genomic RNA 3’ end from the RNA-binding groove. RNA release from the NC 3’ end by P is indirectly supported by the ability of the P N-terminal region to dissociate the N_{TAIL} assembly where effectively all NT_{arm} sites are vacant (Fig. 3D). Upon RNA 3’-end release, it may transiently bind to the exposed positively charged patch on N by its sugar–phosphate backbone (Fig. 5). The polymerase complex is then positioned for the entry of the first 6 nucleotides of the RNA 3’ end in to the active site of L (33–35). Bipartite promoter recognition by the polymerase complex is required for genomic RNA replication (36). In analogy to mumps virus, this may require further uncoiling of the NC, promoted by P (32). Elongation will also require NC uncoiling to expose the RNA. The processivity of the polymerase will promote this, and additional P may be injected into the helical assembly, resulting in local NC uncoiling and template RNA exposure according to the cartwheel model (37). N could be recycled onto the NC once the polymerase complex has passed due to the transient association with the P CTD. The presence of assembled matrix on the NC during early stages of infection will necessitate additional disassembly which is as yet not understood.

In conclusion, our MeV N0-P structure and its comparison to
the previously reported NC state provide insight into MeV NC assembly and polymerase activity.

ACKNOWLEDGMENTS

We thank the following people for their help: Eevakaisa Vesanen for electron microscopy, Juho Kellosalo for SEC-MALLS experiments, Tuula Nyman for mass spectrometry, and the Finnish Instruct National Associate Center, the National Biocenter Finland Cryo-Electron Microscopy Unit, the Crystallization Facility, and the Proteomics and Metabolomics Unit for the use of their facilities.

FUNDING INFORMATION

Sigrid Juselius Foundation provided funding to Sarah Jane Butcher. Biocenter Finland provided funding to Tommi Kajander and Sarah Jane Butcher. Suomen Akatemia (Academy of Finland) provided funding to Sarah Jane Butcher under grant numbers 139178 and 275199.

L.L. was a fellow of the Viikki Doctoral Programme in Molecular Biosciences. The Diamond Light Source beam line I03 is acknowledged for provision of synchrotron beam time. The research leading to these results received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013) under BioStruct-X (grant agreement no. 283570). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES


