High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms

Benoît Arragain[1], Juan Reguera[2], Ambroise Desfosses[1], Irina Gutsche[1], Guy Schoehn[1]* and Hélène Malet[1]*

[1]Université Grenoble Alpes, CNRS, CEA, Institute for Structural Biology (IBS), F-38000, Grenoble, France.

[2]Aix-Marseille Université, CNRS, INSERM, AFMB UMR 7257, 13288 Marseille, France.

*Co-corresponding authors:

Hélène Malet: Tel: +33 (0)4 57 42 87 77. Email: helene.malet@ibs.fr

Guy Schoehn: Tel: +33 (0)4 57 42 85 68. Email: guy.schoehn@ibs.fr

Abstract

Negative-strand RNA viruses condense their genome into helical nucleocapsids that constitute essential templates for viral replication and transcription. The intrinsic flexibility of nucleocapsids usually prevents their full-length structural characterization at high resolution. Here we describe purification of full-length recombinant metastable helical nucleocapsid of Hantaan virus (*Hantaviridae* family, *Bunyavirales* order) and determine its structure at 3.3 Å resolution by cryo-electron microscopy. The structure reveals the mechanisms of helical multimerization via sub-domain exchanges between protomers and highlights nucleotide positions in a continuous positively charged groove compatible with viral genome binding. It uncovers key sites for future structure-based design of antivirals that are currently lacking to counteract life-threatening hantavirus infections. The structure also suggests a model of

nucleoprotein-polymerase interaction that would enable replication and transcription solely upon local disruption of the nucleocapsid.

Introduction

The *Bunyavirales* order is one of the largest groups of segmented negative-strand RNA viruses (sNSV) that include many pathogenic strains (Sun et al., 2018). In particular, the *Hantaviridae* family comprises the virus Hantaan (HTNV) that gives rise to haemorrhagic fevers with renal syndrome and the virus Sin Nombre that is linked to severe pulmonary illnesses with fatality rates up to 40%. Neither treatment nor vaccine are currently available to counteract them.

The *Bunyavirales* genome is usually divided into three RNA segments enwrapped by the viral nucleoproteins (NP). The resulting nucleocapsids (NCs) protect the genome and serve as a replication/transcription template for the viral polymerase (Reuter and Krüger, 2018). They coat the genomic and anti-genomic RNA during replication but not the mRNA produced by transcription. As they are specific and essential to the viral cycle, NCs constitute an attractive potential target for antiviral drugs.

Nucleoproteins of segmented NSV (sNSV) present a large diversity of folds (Sun et al., 2018). Most of the available structures have been determined as rings and monomers that present the advantage of being rigid enough for crystallisation. However, the relevant conformations of assembled NPs in the viral context correspond to flexible helices or pearl-necklaces that encapsulate long RNA segments. Helical crystal structures of La Crosse virus NP (*Peribunyaviridae*, LACV) (Reguera et al., 2013) and Crimean Congo Fever Virus NP (*Nairoviridae*, CCFHV) (Wang et al., 2012) have been determined but they correspond to local organisations of pearl-necklace-like native NCs. Influenza double-helical NCs 3D structures have been described at low resolution but display different helical parameters, also

reflecting their malleability (Arranz et al., 2012; Moeller et al., 2012). NC flexibility appears as a hallmark in NSVs as NCs from non-segmented NSV such as Ebola or measles virus require C-terminal NP truncations in order to obtain rigid NCs and high-resolution 3D structures (Gutsche et al., 2015; Sugita et al., 2018; Wan et al., 2017). In this context, HTNV-NCs are intriguing as they were shown to be able to form rather rigid helices of 10 nm diameter within viruses (Battisti et al., 2011; Huiskonen et al., 2010) and during cell infection (Goldsmith et al., 1995). We therefore aimed at obtaining their high-resolution 3D structure, identifying the determinants of NP polymerisation and visualizing RNA organization.

Results

Expression of recombinant full-length HTNV-NP in insect cells led to formation of recombinant NCs that have a diameter consistent with native NCs (Figure 1 – figure supplement 1). Cryo-EM images collected on a Titan Krios enabled structural determination of HTNV-NC at 3.3 Å resolution (Figure 1A, Video 1, Figure 1 – figure supplement 1 and 2, Figure 1 – source data 1). HTNV-NC is a left-handed helix with a pitch of 68.03 Å and 3.6 subunits per turn (Figure 1A and Figure 1 – figure supplement 1). To derive an atomic model of HTNV-NP, the monomeric crystal structure of HTNV-NP comprising residues 113 to 429 (Olal and Daumke, 2016) was fitted into the EM map, and the N-terminal residues and loops not present in the crystal structures were unambiguously built (Figure 1D). The HTNV-NC model was then iteratively rebuilt and the all-atom model refined using stereochemical restraints.

Although we expressed the full-length NP, residues 1 to 79 are missing in the final map, indicating that they do not follow the helical symmetry. Interestingly, a NP_{74-429} construct obtained by trypsin limited proteolysis still forms a rigid helix, which means that the $N\text{-}ter_{1-73}$ is not necessary for helix stabilisation (Figure 2 – figure supplement 1A,B). The $N\text{-}ter_{1-73}$

might correspond to a flexibly-linked coiled-coil as previously visualised in the structure of an N-ter construct (Boudko et al., 2007).

The structure of NP_{80-429} is composed of a core comprising residues 117 to 398 (NP_{core}) that defines two lobes surrounding a positively charged groove (Figure 1B,C, Video 1). The NP_{core} structure is relatively conserved compared to the apo monomeric truncated crystal structure, with a Cα RMSD of 0.808 Å over 205 atoms (Figure 1D). N-terminal and C-terminal arms (Nt_{arm}, Ct_{arm}) are connected to each extremity of NP_{core} by flexible hinges (Figure 1B, Video 1). Superimposition of HTNV-NC with the truncated monomeric structure (Olal and Daumke, 2016) shows that the Ct_{arm} undergoes a 130.4° rotation upon multimerization (Figure 1D). As a result, Nt- and Ct_{arm} of the same subunit embrace each other: residues 97-98 of Nt_{arm} contact residues 423-424 of Ct_{arm}, revealing a unique intra-arms interaction specific to HTNV-NC (Figure 1D, Video 1).

Recombinant HTNV-NC are rigid and remain stable in a large range of salt, pH and temperature conditions (Figure 2 – figure supplement 1C). This can be explained by the multiplicity of interactions between protomers, each NP interacting with six other subunits (Figure 2A, Video 2). Successive subunit interactions rely on exchange of their $Nt_{arm,}$ and Ct_{arm} that make intimate contacts with the core domain of neighboring protomers (Figure 2 A,B,C), resulting in a buried area of 2704 $Å^2$ at each NP-NP interface. The NP_i Nt_{arm} forms an elongated structure that binds to residues 155-160, 177-181, 189-192 and 136-140 of the NP_{i-1} subunit (Figure 2B, Video 2). Since constructs lacking the Nt_{arm} remain monomeric (Olal and Daumke, 2016), the identified contacts appear to be essential for multimerization. The NP_i amphipatic Ct_{arm} binds in a hydrophobic pocket of the protomer NP_{i+1} comprising residues 334-346 and 378-394 (Figure 2C, Video 2). This agrees with the results of double hybrid experiments (Kaukinen et al., 2004; Yoshimatsu et al., 2003) which suggested that interaction of C-terminal helices of neighboring protomers are critical to oligomerization. Another key

actor of multimerization is the β-hairpin$_{152-181}$ that protrudes towards the exterior of the NP-core (Figure 1B). The two β-stands of the hairpin (residues 155-160 and 177-181) interact with the Nt_{arm} residues 101-103 from the following subunit to form a 3-stranded β-sheet (Figure 2A,B, Video 2). As shown by pull-down assays of mutants L102A, V104A, this structure has a main contribution in NC stabilisation (Guo et al., 2015). In addition, the β-hairpin$_{152-181}$ tip (residues 162-175) acts as a clamp and seals the C-terminus-mediated contacts between NP_{i+2} (residues 409 to 419) and NP_{i+3} (residues 385-390), thereby buttressing and rigidifying the metastable helical form of the HTNV-NC (Figure 2A,D, Video 2).

HTNV recombinant NP was expressed in the absence of viral RNA, but after purification the optical density (OD) at 260-280 nm was measured to be around 1.0, which strongly suggests the binding of insect cell RNAs during expression. Consistently, three nucleotides with a partial occupancy can be visualized for each NP (Figure 3A). Conserved residues R197, R314, R368, R146 and K153 interact via salt bridges with RNA phosphates, in line with the versatility of RNA sequence to be incorporated by NP. In addition, F361 stabilizes a stacking interaction network between the nucleotide bases. The nucleotides bind in a continuous positively charged groove oriented towards the interior of the NC. Density corresponding to additional nucleotides is visible in this groove, but present at low occupancy, suggesting the lower affinity of NC-RNA binding in these regions. Interestingly, the observed nucleotides occupancy is in accordance with electromobility shift assays of Sin Nombre NP mutants (Guo et al., 2015): indeed, in this work, residues directly interacting with the three visible RNA nucleotides were identified as the ones displaying the highest affinity with RNA (Guo et al., 2015). Altogether, these results reveal that HTNV-NC structure is compatible with binding of long viral RNA, thereby reinforcing the biological relevance of this structure (Figure 3B,C).

Discussion

Structural determination of a full-length RNA-bound helical NSV NC at 3.3 Å resolution is particularly valuable because the usual intrinsic flexibility of NSV full-length NCs prevents their high-resolution analysis. It fits together the pieces of the jigsaw accumulated over several years of biochemistry and structural analysis on Hantavirus NP. HTNV recombinant NC structure indeed displays 3.6 subunits per turn and is thus consistent with observations of HTNV-NP trimers made by several groups (Alfadhli et al., 2001; Kaukinen et al., 2004). The present structure is also compatible with the proposed model (Kaukinen et al., 2001) which suggested that NPs first trimerize around the viral RNA and then gradually assemble to form longer multimers. The role of Nt_{arm} and Ct_{arm} exchange between successive subunits, identified by double-hybrid and pull-down experiments (Guo et al., 2015; Kaukinen et al., 2004; Yoshimatsu et al., 2003), is in line with HTNV recombinant NC structure. These observations, together with the visualisation of a continuous positively charged groove, strongly suggests that the present structure is biologically pertinent. HTNV recombinant NCs are thus likely to be similar to helical NCs observed within viral particles, although the fact that the latter are less straight implies that at least the β-$hairpin_{152-181}$ might change conformation in the viral context, thereby enabling more flexibility (Battisti et al., 2011; Huiskonen et al., 2010). Other conformations of NCs are in addition likely to exist because the 5' and 3' end of HTNV viral RNA are known to bind to the viral polymerase, implying that the NC must somehow be circularized (Garcin et al., 1995). Accordingly, flexible pearl-necklaces are also visible within virions, in infected cells and in NC extracted from viral particles (Battisti et al., 2011; Goldsmith et al., 1995; Guo et al., 2015; Huiskonen et al., 2010). Combination of helical NCs observed here and flexible pearl-necklaces thus represent relevant genome-encapsidating conformations.

HTNV and LACV NP_{core} share a common fold (Guo et al., 2015; Olal and Daumke, 2016) enabling their structural superimposition (Figure 3-supplement figure 1A) and comparison of

their RNA binding mode. This reveals that the three nucleotides present in HTNV-NC adopt similar conformations as the nucleotides 5, 6 and 7 of LACV-NP (Figure 3-supplement figure 1B) (Reguera et al., 2013). Key residues involved in the binding of HTNV-NC nucleotides, namely R197, R367 and F361 are conserved in LACV-NP (R94, R183 and Y177). Superposition of RNA-bound LACV-NP monomers on HTNV-NC shows that LACV RNA fits reasonably well into the HTNV-NC positively charged groove (Figure 3-supplement figure 1C). However, the proposed HTNV-NC RNA path is slightly shorter, suggesting that each HTNV-NP can contain between 8 and 10 nucleotides (Figure 3 – figure supplement 1C). Modelling of RNA binding in HTNV-NC based on RNA position in LACV-NP (see Material and Methods) indicates that RNA phosphates and riboses interact with residues R146, K150, K153, R197, R314, R339, R367 and R368, while bases are surrounded by residues 113-116, 143-151, 182-188 and 217-222 (Figure 3 – figure supplement 1D).

While the structures of HTNV and LACV NPs share a common fold for their cores, they however differ in their N-terminal organisation, HTNV Nt_{arm} being significantly longer and linked to the unresolved N-terminal$_{1-73}$ region (Figure 3 – figure supplement 1A). The N-terminal$_{1-73}$ region may bring the polymerase in close contact to NP as suggested (Cheng et al., 2014), and thus play a role similar to the intrinsically disordered phosphoproteins and NP C-terminal regions in nsNSV (Ivanov et al., 2011). One may therefore hypothesize that during genome reading by the polymerase, the Nt_{arm} of one HTNV-NP could detach from its adjacent subunit, thereby affecting the conformation of the next NP β-hairpin and removing the seal over the two following NPs. This would provoke a local disruption of the metastable NC and provide RNA access to the polymerase, enabling replication and transcription (Figure 4). This opens new routes for future experiments, such as cryo-EM analysis of *in vivo* reconstituted mini-ribonucleoprotein particles (RNP) or cryo-electron tomography of viral RNPs, that would decipher the exact mode of interaction between NPs, polymerase and RNA during

RNA replication/transcription. The rigidity of HTNV-NC and the absence of disordered phosphoprotein in HTNV should facilitate this study compared to the more complex NP/phosphoprotein/polymerase complex of nsNSV, and thus represents a unique opportunity to unravel significant properties of NSV replication.

In addition to these fundamental aspects, the HTNV-NC structure reveals the position of specific antigenic sites on variable regions of the NC surface (Tischler et al., 2008) thus providing a structure-based rationale for diagnosis (Figure 1 – figure supplement 3). It may also stimulate the design of antivirals, because it defines key regions involved in NP oligomerization and RNA encapsidation.

Material and Methods

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
recombinant DNA reagent	*HTNV-NP*	Geneart		Synthetic gene
peptide, recombinant protein	HTNV-NP	This article	UniProtKB-P05133	HTNV-NP was obtained by expression in insect cell of the synthetic gene *HTNV-NP* mentioned above
software, algorithm	Motioncor2	doi: 10.1038/nmeth.2472		http://msg.ucsf.edu/em/software/motioncor2.html
software, algorithm	Gctf	doi: 10.1016/j.jsb.2015.11.003	RRID:SCR_016500	https://www.mrc-lmb.cam.ac.uk/kzhang/
software, algorithm	Relion2.1 and Relion3	doi: 10.1016/j.jsb.2012.09.006 and doi: 10.1101/421123	RRID:SCR_016274	https://www2.mrc-lmb.cam.ac.uk/relion/index.php?title=Main_Page
software, algorithm	Coot	doi: 10.1107/S0907444910007493	RRID:SCR_014222	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
software, algorithm	RCrane	doi: 10.1107/S0907444912018549		https://pylelab.org/software/rcrane-readme
software, algorithm	PHENIX	doi: 10.1107/S0907444909052925	RRID:SCR_014224	https://www.phenix-online.org/
software, algorithm	Chimera	doi: 10.1002/jcc.20084	RRID:SCR_004097	https://www.cgl.ucsf.edu/chimera/
software, algorithm	Haddock	doi: 10.1038/nprot.2010.32 and 10.1016/j.jmb.2015.09.01		https://haddock.science.uu.nl/

184 **Cloning, Expression and Purification**

185 Sequence-optimized synthetic DNA encoding a N-terminal his-tag, a TEV protease
186 recognition site and the HTNV-NP (NCBI accession code NC_005218) was synthetized
187 (Geneart) and cloned into a pFastBac1 vector between NdeI and NotI restriction sites.

188 The HTNV-NP-expressing baculovirus was generated via the standard Bac-to-Bac method
189 (Invitrogen). For large scale expression, Sf21 cells at $0.5x10^6$ cells/mL concentration were
190 infected by adding 0.1% of virus. Expression was stopped 72h after the day of proliferation
191 arrest.

192 The cells were disrupted by sonication 3 min (10 sec ON, 20 sec OFF, 40% amplitude) on ice
193 in lysis buffer (20 mM Tris-HCl pH 8, 300 mM NaCl, 10 mM $MgCl_2$, 2 mM β-
194 mercaptoethanol and 20 mM Imidazole) with EDTA free protease inhibitor complex (Roche).
195 After lysate centrifugation at 20,000 rpm during 45 min at 4°C, protein from the soluble
196 fraction was loaded on a nickel column, washed with 10 volumes of lysis buffer, 10 volumes
197 lysis buffer supplemented with 50 mM Imidazole and eluted with 5 volumes of lysis buffer
198 supplemented with 500 mM Imidazole. The eluted protein was cleaved with TEV protease in
199 a 20:1 w/w ratio overnight at 4°C in dialysis against lysis buffer resulting in an almost
200 complete cleavage. A second nickel column step was performed to remove unwanted
201 material. The resulting protein was concentrated by ultracentrifugation using Optima XE
202 SW55Ti rotor (Beckman Coulter) and 0.8 mL Ultra Clear tube during 2h at 45,000 rpm, 4°C.
203 Concentrated HTNV-NC present in the 20 µl bottom fraction of each tube was gently
204 resuspended. The 260/280nm absorbance ratio was measured to be around 1 at the end of the
205 purification indicating presence of nucleic acid.

206 Three biologically independent batches of large-scale expression and fifteen biologically
207 independent batches of purifications were performed and gave reproducible results.

208 **Limited proteolysis by trypsin and N-terminal sequencing**

Limited proteolysis of HTNV-NC was performed at 20°C in lysis buffer with a 3:2 w/w ratio of HTNV-NP/trypsin. Proteolysis was stopped by the addition of denaturing SDS-PAGE loading dye and incubation at 95°C for 5 min. Digestion was visualized in 10% acrylamide SDS-Page gels. For N-terminal sequencing, proteins were transferred on PVDF membrane previously incubated in 10 mM CAPS pH 11, 10% methanol buffer. PVDF membrane was stained using 0.1% Coomassie Brilliant Blue R-250, 40% methanol, 1% acetic acid buffer until bands were visible, washed using 50% methanol and dried. Visible bands were cut and subjected to N-terminal sequencing. Amino acid sequence determination based on Edman degradation was performed using an Applied Biosystems gas-phase sequencer model 492 (s/n: 9510287J). Phenylthiohydantoin amino acid derivatives generated at each sequence cycle were identified and quantitated on-line with an Applied Biosystems Model 140C HPLC system using the data analysis system for protein sequencing from Applied Biosystems (software Procise PC v2.1).

Electron microscopy

For negative stain EM grid preparation, 4 µl of sample was applied between mica and carbon layer and stained using sodium silicotungstate (SST) 2%. After removing the mica part, the grid was deposited on the carbon layer and dried at room temperature. Micrographs were collected at 2.5 µm defocus using a FEI Tecnai20 operated at 200 kV on 4k*4k CETA FEI CCD camera.

For cryo-EM grid preparation, Quantifoil grids 400 mesh 2/1 were glow-discharged at 30 mA for 1 min. 3 µl of sample were applied on the resulting glow-discharged grids and excess solution was blotted during 2.5 sec force 7 with a Vitrobot Mark IV (FEI) and the grid frozen in liquid ethane. Two biologically independent datasets, arising from different expression and purification batches, were collected on two high-end cryo-electron microscopes and gave consistent structures. The first cryo-EM dataset was collected on a FEI Polara microscope

operated at 300 kV equipped with a K2 summit GATAN direct electron detector and resulted in a 3.8 Å resolution structure. The second dataset, that gave rise to the 3.3 Å resolution structure presented in the present article, was collected on a FEI Titan Krios operated at 300 kV equipped with a Gatan Bioquantum LS/967 energy filter coupled to a Gatan K2 direct electron detector camera. For this second dataset, automated data collection was performed using EPU FEI software. Zero-loss micrographs were recorded at a 46,860x magnification giving pixel size of 1.067Å with defocus ranging from 0.8 to 3.5 µm. In total, 4328 movies with 28 frames per movie were collected with a total dose of 40 $e^-/Å^2$.

Image processing

Micrographs were initially selected based on visual quality inspection. Movie drift correction was performed in Motioncor2 (Li et al., 2013) excluding the two first frames. CTF parameters were determined in Gctf (Zhang, 2016) (RRID:SCR_016500). All subsequent processing steps were performed in Relion2.1 and Relion3 software (Scheres, 2012; Zivanov et al., 2018) (RRID:SCR_016274). Straight HTNV-NC were manually picked and computationally cut with an inter-box distance of 38 Å along the helical axis into overlapping boxes of 400*400 pixels, resulting in 168,709 extracted segments. 2D classification was used to eliminate bad quality filaments. The best 2D classes were aligned and padded in a square of 1200*1200 pixels. The individual 2D Power Spectra (PS) of each best class were averaged and used for a first estimation of the helical parameters using Fourier Bessel indexing (Cochran et al., 1952). An estimation of the repeat c was first inferred from the layer lines regularly spaced at multiples of ~335.6$Å^{-1}$ (Figure 1 – figure supplement 1E). Several strong meridional layer lines regularly spaced at multiples of 18.64 $Å^{-1}$ indicated the axial rise p between subunits. According to both c estimation and p, pitch P was inferred to be around 67.12$Å^{-1}$ as a strong layer line with a first intensity near the meridian can be seen at l = 5. Therefore, the structure

repeats after u = 18 subunits (18 (u) = 335.6 (c) / 18.64 (p)), in t = 5 turns, resulting in a number of units per turn u/t of ~3.6.

3D helical reconstruction was performed using the 111,248 selected segments from the best 2D classes. A 130 Å diameter featureless cylinder was used as an initial model. Helical symmetry search was performed between -100 ±4 ° for the helical twist (360°/ u/t) and 18.64 ±0.5 Å for the helical rise. Helical twist and rise after refinement were respectively -99.97° and 18.96 Å.

For the final 3D helical reconstruction, poorly aligned segments with an angular tilt > ±20 ° were discarded resulting in 105,665 segments. Local CTF-determination was calculated for each segment using Gctf (Zhang, 2016) and finally CTF-refinement was performed using Relion 3.0[18]. The 10 last frames of each motion corrected micrographs were removed resulting in a total dose of ~22 e-/Å^2. The latest 3D map filtered at 15 Å resolution was used as initial model and helical symmetry search further refined during a last refinement. The resulting final reconstruction displays refined helical parameters of -99.95° for the twist and 18.87Å for the rise. Post-processing, done using a B-factor of -103 Å^2, resulted in a 3.3 Å resolution reconstruction using the FSC 0.143 cutoff criteria (Figure 1 – figure supplement 2A). Local resolution variations were estimated in Relion (Figure 1 – figure supplement 2B).

Observation of nucleotides with partial occupancy suggests that only a fraction of HTNV-NC segments analysed contains RNA. We thus attempted to perform both 2D and 3D classifications in order to separate RNA-bound and apo HTNV-NC segments using several strategies including usage of different masks, subtraction/absence of subtraction of protein density, symmetry release. However, due to the low molecular weight of RNA compared the protein (9 RNA nucleotides bound per HTNV-NP would correspond to 3kDa of RNA versus 50kDa of protein), these classifications did not succeed in separating the apo and RNA-bound states.

Model building and refinement

The monomeric crystal structure of HTNV-NP comprising residues 113 to 429 (PDB code: 5FSG) was initially fitted into the EM density as two separate rigid bodies containing residues 113-398 and 399-429. Loops comprising residues 146-155 and 349-360, β-hairpin 160-172 and the N-terminal arm residues 80 to 112 - previously not present in the crystal structure - were manually built in Coot (Emsley et al., 2010). The three RNA nucleotides visible in the density map were manually built and adjusted using Coot (Emsley et al., 2010) (RRID:SCR_014222) and RCrane (Keating and Pyle, 2012). NP monomeric model was symmetrized according to the helical parameters to form a filament model of 7 subunits. Fiber model was then iteratively rebuilt and all-atom refined using stereochemical and NCS restraints within PHENIX (Adams et al., 2010) (RRID:SCR_014224).

RNA modelling based on LACV NP RNA

In order to model HTNV RNA, two HTNV-NP were extracted from HTNV-NC, together with their respective three nucleotides (called "HTNV-nts"). Five, six or seven LACV-nts were used to link the three HTNV-nts, giving rise to hybrid models containing between eight and ten nucleotides per NP. They were regularized in Coot, minimized in Chimera (Pettersen et al., 2004) (RRID:SCR_004097) and were then subjected to modelling in Haddock (de Vries et al., 2010; van Zundert et al., 2016). During the modelling procedure, the interaction between HTNV-NP and HTNV-nucleotides was maintained in the same conformation as in the EM map using unambiguous distance restraints. In addition, Haddock ambiguous interaction restraints were defined: residues involved in interactions (Guo et al., 2015) were considered as "active", while all solvent accessible surface neighbors of active residues were defined as "passive". The initial starting orientations were not randomized. Rigid body minimization, semi-flexible simulated annealing and flexible explicit solvent refinement was performed. For each starting RNA, the resulting models were very similar and gathered in one

single cluster. The model with the lowest energy and the lowest Haddock score was thus considered as being the modelling result.

Accession numbers

The atomic coordinates and the cryo-EM map have respectively been deposited in the PDB and EMDB under the accession codes 6I2N and EMD-0333.

Figures legends

Figure 1 HTNV-NC structure

(A) HTNV-NC helical organization. The upper part of the cryo-EM map is colored per NP domain, whereas the lower part is shown as transparent and colored per NP protomer. The NP surrounded by dotted lines corresponds to Figure 1B orientation. (B) Domain organization of one monomer extracted from the NC helical assembly. Each domain is colored as in Figure 1A upper part. The newly described β-hairpin (residues 152-181) is labeled and surrounded by dotted lines. (C) Electrostatic surface representation of HTNV-NP monomer. (D) Superimposition of monomeric NP_{core} (Olal and Daumke, 2016) (in gray) and NP from the present structure (colored as in Figure 1 A,B). The Ct_{arm} rotation is highlighted and newly built elements are labeled, shown as non-transparent and surrounded by dotted lines. A close up view of the Ct_{arm}/Nt_{arm} interaction is shown.

Video 1: Overview of HTNV-NC structural organization. The helical arrangement of HTNV-NC is colored as in Figure 1A upper part. HTNV-NP domain organization is colored as in Figure 1B. Ct_{arm}/Nt_{arm} interaction is highlighted.

Figure 2 NP-NP interactions

(A) General view. The filament orientation corresponds to RNA direction from 5' to 3'. Each NP protomer is colored differently. Interacting regions of NP_i with subunits ranging from NP_{i-3} to NP_{i+3} are shown as non-transparent. Positions of Figure 2B, C, D close-up views are indicated with dotted lines. (B) NP_i Nt_{arm} binding site in NP_{i-1}. Residues 101-103 from NP_i Nt_{arm} form a β-sheet with β-strands 155-160 and 177-181 from NP_{i-1}. Hydrogen bonds are shown as black dotted lines. (C) NP_i Ct_{arm} binding site in NP_{i+1}. (D) Binding of NP_i β-hairpin tip (residues 162-175) on NP_{i+2} (residues 409-419) and NP_{i+3} (residues 385-390). NP_i I171 is

plugged into a hydrophobic pocket formed by L409 and L413 of the NP_{i+2} C-terminal helix and M385 and K390 sidechains of the NP_{i+3} C-terminal lobe. Hydrogen bonds between the residue pairs N169-A387 and G170-K390 from NP_i and NP_{i+3} respectively, further stabilize this interaction contributing to the rigidification of HTNV-NC.

Video 2: Interaction of subunit NP_i with six neighboring subunits. Subunit NP_i (in green) interacts: (i) with subunit NP_{i-1} (in yellow) via its Nt_{arm}, (ii) with subunit NP_{i+1} (in orange) via its Ct_{arm}, with subunit NP_{i+2} (in dark blue) and subunit NP_{i+3} (in cyan) via its β-hairpin tip. Due to helical symmetry, NP_{i-3} β-hairpin tip (in light cyan) also interacts with NP_i (same interaction as NP_i-NP_{i+3}). Consistently, NP_{i-2} β-hairpin tip (in red) also interacts with NP_i (same interaction as NP_i-NP_{i+2}).

Figure 3 HTNV-NC RNA binding site

(A) RNA binding mode of 3 nucleotides. EM density of RNA and RNA-binding residues are displayed in transparent grey. Hydrogen bonds are shown as dotted lines. (B) Cut-away view of NC electrostatic potential showing the continuous positively-charged groove. RNA nucleotides are displayed. (C) RNA-binding residues are shown as sticks and colored per subunit. RNA-binding residues defined in (Guo et al., 2015), namely R146*, R197*, R199*, R314*, R339*, R367*, R368* are labeled with stars, while RNA-binding residues identified in the present structure, namely R144, K150, K153 are shown without stars.

Figure 4: Model of HTNV replication and transcription.

(A) Schematics representation of HTNV-NP. Major secondary structures involved in interprotomer interactions are represented as arrows for β-strand and cylinders for α-helices. RNA binding cavity is represented as a clipped part from the NP_{core} region (green oval). (B)

Schematic representation of RNA binding and NP-NP interactions. RNA is shown as a dotted purple line. Main NP-NP interactions between adjacent subunits are indicated. For clarity, NP_i interactions with NP_{i-2}, NP_{i-3}, NP_{i+2} and NP_{i+3} are absent from the schematic representation. (C) Replication working hypothesis model. The polymerase is shown in purple and the newly transcribed/replicated RNA is indicated as a pink dotted line. The model proposes that the polymerase binds to the flexible N-terminal$_{1-73}$ region during replication/transcription in order to move along the NC. This would be reminiscent to P and L binding to flexible sNSV C-terminal region. Binding of the polymerase to the N-terminal$_{1-73}$ region could destabilize the adjacent Nt_{arm} binding. This localized disruption of NP-NP interaction would create a local opening of the NC enabling transient access of the polymerase to few RNA nucleotides for replication/transcription. Such an opening could be possible without disturbing the whole NC as NP to NP contacts are driven not only by the Nt_{arm} but also by the Ct_{arm}. The Ct_{arm} interaction is likely to remain intact even with a local opening of the NC as the C-terminal hinge allows the Ct_{arm} to undergo large rotation.

Figure Supplement legends

Figure 1 – figure supplement 1: Purification, EM analysis and symmetry determination of HTNV-NC (A) SDS-PAGE gel of HTNV-NC purification. T: total fraction, S: soluble fraction, FT: flow through of the 1st Ni-NTA purification, W: wash of the 1st Ni-NTA purification, E1: elution of the 1st Ni-NTA purification, E2: supernatant after TEV cleavage and 2nd Ni-NTA purification, U: concentrated protein after ultra-centrifugation. (B,C) Negative stain (B) and cryo electron micrographs (C) of HTNV-NC. Scale bar is indicated. (D) 2D class average. (E) 2D class average power spectrum, indexing and interpretation enabling symmetry determination.

Figure 1 – figure supplement 2: Resolution of HTNV-NC structure

(A) The gold-standard Fourier shell correlation (FSC) of masked map indicates a resolution of 3.3 Å with the FSC=0.143 criteria. The FSC between the map and the model is 3.4 Å according to FSC=0.5 criterion. (B) Local resolution. (C) Zoom-up view of a representative α-helix that shows clear side chains densities illustrating the quality of the map. (D) The B-factor analysis strongly correlates with the local resolution analysis (B) and shows that the N-terminal hinge and the solvent-exposed region containing residues 242-300 are flexible.

Figure 1 – figure supplement 3: Specific antigenic sites are localized in variable regions of the NC surface.

(A) Surface conservation mapping of HTNV-NC. Regions localized towards NC interior, in particular regions involved in RNA-binding and NP-NP oligomerisation are conserved, whereas HTNV-NC exposed regions are more variable. (B) Epitopes involved in serotype-specific and cross-reactive recognition of HTNV-NC determined in (Tischler et al., 2008) are shown in red. They correspond to surface-exposed regions of HTNV-NC corroborating their necessary accessibility.

Figure 2 – figure supplement 1: Limited proteolysis and stability of recombinant HTNV-NCs

(A) SDS-Page analysis of trypsin digested HTNV-NC. MW represents the molecular weight markers. NC were incubated with trypsin (Tryp) at 20°C for different duration from 0 (NP) to 30 min. After 30 min, a stable construct containing residues 74 to 429 is obtained. Presence of intermediate limited proteolysis fragments containing residues 43-429 and 48-429 strongly suggests that the N-terminal residues 1-73 are not disordered and might correspond to a coiled-coil as observed by (Boudko et al., 2007). (B) Negative stain EM micrograph of truncated construct NP_{74-429} displays NC similar to the ones of full-length NP. (C) Negative

stain micrographs of full-length HTNV-NCs incubated in different pHs and salt concentration as indicated. NC fibers stay stable in harsh conditions. Consecutive freeze-thaw breaks NC in small rigid parts. Scale bars are indicated.

Figure 3 – figure supplement 1: Comparison of LACV and HTNV-NP/NC, RNA-binding model

(A) Superimposition of HTNV and LACV NP_{cores}. HTNV and LACV Nt_{arms} and Ct_{arms} contain the same secondary structures but LACV arms are significantly shorter and differ in orientation. (B) HTNV-NP-bound nucleotides adopt a similar conformation as LACV-NP-bound nucleotides 5 to 7. The nucleotide binding pocket is conserved in HTNV-NP and LACV-NP. (C) Cut-away view of HTNV-NC electrostatic potential. Visible HTNV nucleotides are shown in green. LACV nucleotides position based on superimposition of LACV and HTNV NP_{cores} are shown in salmon. LACV RNA fits reasonably well into HTNV-NC RNA path, however LACV nucleotide 11 from NP_i clashes with LACV nucleotide 1 from NP_{i+1} showing that each HTNV-NP binds less than 11 nucleotides. LACV nucleotides numbering is indicated. (D) Model of HTNV-NC bound to 9 nucleotides. Nucleotides numbering from 1 to 9 is indicated.

Residues interacting with riboses and phosphates are shown as sticks. Regions interacting with bases are colored and specifically labelled.

Acknowledgements

We thank Stephen Cusack for providing the synthetic gene and enabling J.R. and H.M. to initiate this project while being post-doctoral researchers in his lab. We thank Karine Huard, Angélique Fraudeau, Delphine Guilligay, Pascal Fender and Alice Aubert for technical advices on expression and purification; Jean-Pierre Andrieu for N-terminal sequencing; Maria

Bacia and Michael Hons for help with cryo-EM data collection; Leandro Estrozi for advices on helical symmetry determination and discussion on image processing.

This work used the platforms of the Grenoble Instruct-ERIC Center (ISBG: UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The electron microscope facility is supported by the Rhône-Alpes Auvergne Region, the Fondation Recherche Medicale (FRM), the fonds FEDER and the GIS-Infrastrutures en Biologie Sante et Agronomie (IBISA). We acknowledge the European Synchrotron Radiation Facility for provision of microscope time on CM01. We thank all platform staff that enabled us to perform these analyses.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221. https://doi.org/10.1107/S0907444909052925

Alfadhli, A., Love, Z., Arvidson, B., Seeds, J., Willey, J., Barklis, E., 2001. Hantavirus nucleocapsid protein oligomerization. J. Virol. 75, 2019–2023. https://doi.org/10.1128/JVI.75.4.2019-2023.2001

Arranz, R., Coloma, R., Chichón, F.J., Conesa, J.J., Carrascosa, J.L., Valpuesta, J.M., Ortín, J., Martín-Benito, J., 2012. The structure of native influenza virion ribonucleoproteins. Science 338, 1634–1637. https://doi.org/10.1126/science.1228172

Battisti, A.J., Chu, Y.-K., Chipman, P.R., Kaufmann, B., Jonsson, C.B., Rossmann, M.G., 2011. Structural studies of Hantaan virus. J. Virol. 85, 835–841.

https://doi.org/10.1128/JVI.01847-10

Boudko, S.P., Kuhn, R.J., Rossmann, M.G., 2007. The coiled-coil domain structure of the Sin Nombre virus nucleocapsid protein. J. Mol. Biol. 366, 1538–1544. https://doi.org/10.1016/j.jmb.2006.12.046

Cheng, E., Wang, Z., Mir, M.A., 2014. Interaction between hantavirus nucleocapsid protein (N) and RNA-dependent RNA polymerase (RdRp) mutants reveals the requirement of an N-RdRp interaction for viral RNA synthesis. J. Virol. 88, 8706–8712. https://doi.org/10.1128/JVI.00405-14

Cochran, W., Crick, F.H., Vand, V., 1952. The structure of synthetic polypeptides. I. The transform of atoms on a helix. Acta Crystallogr. 5, 581–586. https://doi.org/10.1107/S0365110X52001635

de Vries, S.J., van Dijk, M., Bonvin, A.M.J.J., 2010. The HADDOCK web server for data-driven biomolecular docking. Nat. Protoc. 5, 883–897. https://doi.org/10.1038/nprot.2010.32

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501. https://doi.org/10.1107/S0907444910007493

Garcin, D., Lezzi, M., Dobbs, M., Elliott, R.M., Schmaljohn, C., Kang, C.Y., Kolakofsky, D., 1995. The 5’ ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. J. Virol. 69, 5754–5762.

Goldsmith, C.S., Elliott, L.H., Peters, C.J., Zaki, S.R., 1995. Ultrastructural characteristics of Sin Nombre virus, causative agent of hantavirus pulmonary syndrome. Arch. Virol. 140, 2107–2122.

Guo, Y., Wang, W., Sun, Y., Ma, C., Wang, Xu, Wang, Xin, Liu, P., Shen, S., Li, B., Lin, J., Deng, F., Wang, H., Lou, Z., 2015. Crystal Structure of the Core Region of Hantavirus Nucleocapsid Protein Reveals the Mechanism for Ribonucleoprotein Complex Formation. J.

Virol. 90, 1048–1061. https://doi.org/10.1128/JVI.02523-15

Gutsche, I., Desfosses, A., Effantin, G., Ling, W.L., Haupt, M., Ruigrok, R.W.H., Sachse, C., Schoehn, G., 2015. Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. Science 348, 704–707. https://doi.org/10.1126/science.aaa5137

Huiskonen, J.T., Hepojoki, J., Laurinmäki, P., Vaheri, A., Lankinen, H., Butcher, S.J., Grünewald, K., 2010. Electron cryotomography of Tula hantavirus suggests a unique assembly paradigm for enveloped viruses. J. Virol. 84, 4889–4897. https://doi.org/10.1128/JVI.00057-10

Ivanov, I., Yabukarski, F., Ruigrok, R.W.H., Jamin, M., 2011. Structural insights into the rhabdovirus transcription/replication complex. Virus Res. 162, 126–137. https://doi.org/10.1016/j.virusres.2011.09.025

Kaukinen, P., Koistinen, V., Vapalahti, O., Vaheri, A., Plyusnin, A., 2001. Interaction between molecules of hantavirus nucleocapsid protein. J. Gen. Virol. 82, 1845–1853. https://doi.org/10.1099/0022-1317-82-8-1845

Kaukinen, P., Kumar, V., Tulimäki, K., Engelhardt, P., Vaheri, A., Plyusnin, A., 2004. Oligomerization of Hantavirus N protein: C-terminal alpha-helices interact to form a shared hydrophobic space. J. Virol. 78, 13669–13677. https://doi.org/10.1128/JVI.78.24.13669-13677.2004

Keating, K.S., Pyle, A.M., 2012. RCrane: semi-automated RNA model building. Acta Crystallogr. D Biol. Crystallogr. 68, 985–995. https://doi.org/10.1107/S0907444912018549

Li, X., Mooney, P., Zheng, S., Booth, C., Braunfeld, M.B., Gubbens, S., Agard, D.A., Cheng, Y., 2013. Electron counting and beam-induced motion correction enable near atomic resolution single particle cryoEM. Nat. Methods 10, 584–590. https://doi.org/10.1038/nmeth.2472

Moeller, A., Kirchdoerfer, R.N., Potter, C.S., Carragher, B., Wilson, I.A., 2012. Organization

517 of the influenza virus replication machinery. Science 338, 1631–1634.
518 https://doi.org/10.1126/science.1227270
519 Olal, D., Daumke, O., 2016. Structure of the Hantavirus Nucleoprotein Provides Insights into
520 the Mechanism of RNA Encapsidation. Cell Rep. 14, 2092–2099.
521 https://doi.org/10.1016/j.celrep.2016.02.005
522 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C.,
523 Ferrin, T.E., 2004. UCSF Chimera—A visualization system for exploratory research and
524 analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084
525 Reguera, J., Malet, H., Weber, F., Cusack, S., 2013. Structural basis for encapsidation of
526 genomic RNA by La Crosse Orthobunyavirus nucleoprotein. Proc. Natl. Acad. Sci. U. S. A.
527 110, 7246–7251. https://doi.org/10.1073/pnas.1302298110
528 Reuter, M., Krüger, D.H., 2018. The nucleocapsid protein of hantaviruses: much more than a
529 genome-wrapping protein. Virus Genes 54, 5–16. https://doi.org/10.1007/s11262-017-1522-3
530 Scheres, S.H.W., 2012. RELION: Implementation of a Bayesian approach to cryo-EM
531 structure determination. J. Struct. Biol. 180, 519–530.
532 https://doi.org/10.1016/j.jsb.2012.09.006
533 Sugita, Y., Matsunami, H., Kawaoka, Y., Noda, T., Wolf, M., 2018. Cryo-EM structure of the
534 Ebola virus nucleoprotein–RNA complex at 3.6 Å resolution. Nature 1.
535 https://doi.org/10.1038/s41586-018-0630-0
536 Sun, Y., Li, J., Gao, G.F., Tien, P., Liu, W., 2018. Bunyavirales ribonucleoproteins: the viral
537 replication and transcription machinery. Crit. Rev. Microbiol. 44, 522–540.
538 https://doi.org/10.1080/1040841X.2018.1446901
539 Tischler, N.D., Rosemblatt, M., Valenzuela, P.D.T., 2008. Characterization of cross-reactive
540 and serotype-specific epitopes on the nucleocapsid proteins of hantaviruses. Virus Res. 135,
541 1–9. https://doi.org/10.1016/j.virusres.2008.01.013

van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J., Bonvin, A.M.J.J., 2016. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J. Mol. Biol., Computation Resources for Molecular Biology 428, 720–725. https://doi.org/10.1016/j.jmb.2015.09.014

Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S., Briggs, J.A.G., 2017. Structure and assembly of the Ebola virus nucleocapsid. Nature 551, 394–397. https://doi.org/10.1038/nature24490

Wang, Y., Dutta, S., Karlberg, H., Devignot, S., Weber, F., Hao, Q., Tan, Y.J., Mirazimi, A., Kotaka, M., 2012. Structure of Crimean-Congo hemorrhagic fever virus nucleoprotein: superhelical homo-oligomers and the role of caspase-3 cleavage. J. Virol. 86, 12294–12303. https://doi.org/10.1128/JVI.01627-12

Yoshimatsu, K., Lee, B.-H., Araki, K., Morimatsu, M., Ogino, M., Ebihara, H., Arikawa, J., 2003. The multimerization of hantavirus nucleocapsid protein depends on type-specific epitopes. J. Virol. 77, 943–952.

Zhang, K., 2016. Gctf: Real-time CTF determination and correction. J. Struct. Biol. 193, 1–12. https://doi.org/10.1016/j.jsb.2015.11.003

Zivanov, J., Nakane, T., Forsberg, B., Kimanius, D., Hagen, W.J.H., Lindahl, E., Scheres, S.H.W., 2018. RELION-3: new tools for automated high-resolution cryo-EM structure determination. bioRxiv 421123. https://doi.org/10.1101/421123

A

110 Å

B

N-terminal lobe (117-213)

N-terminal hinge (108-116)

N-terminal arm (80-107)

RNA

C-terminal lobe (214-398)

180°

β-hairpin (152-181)

C-terminal arm (404-429)

C-terminal hinge (399-403)

C

Positively charged groove

Electrostatic potential (kT)

5

0

-5

45°

D

130°

β-hairpin tip 160-172

146-155

349-360

N-terminal arm 80-112

C-terminal arm

Q424

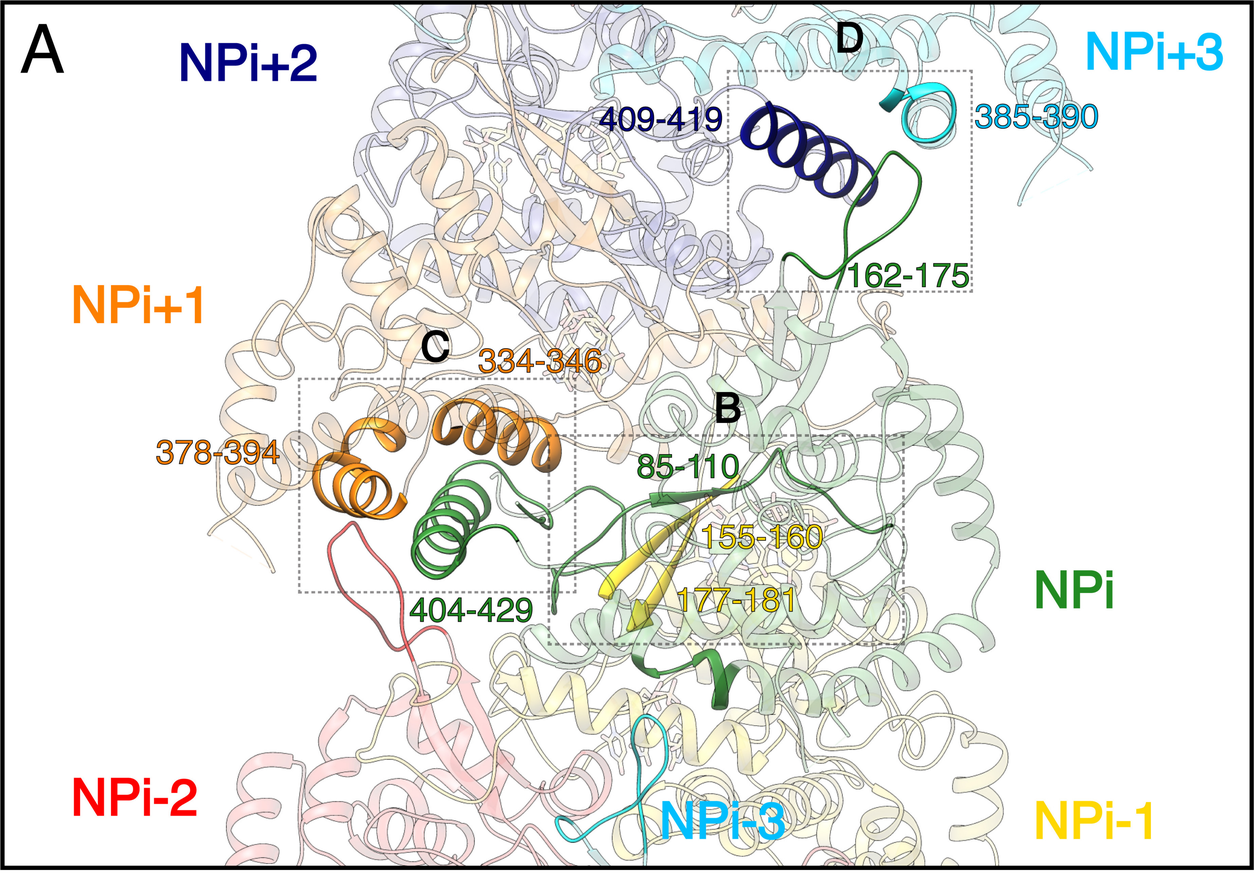
N423

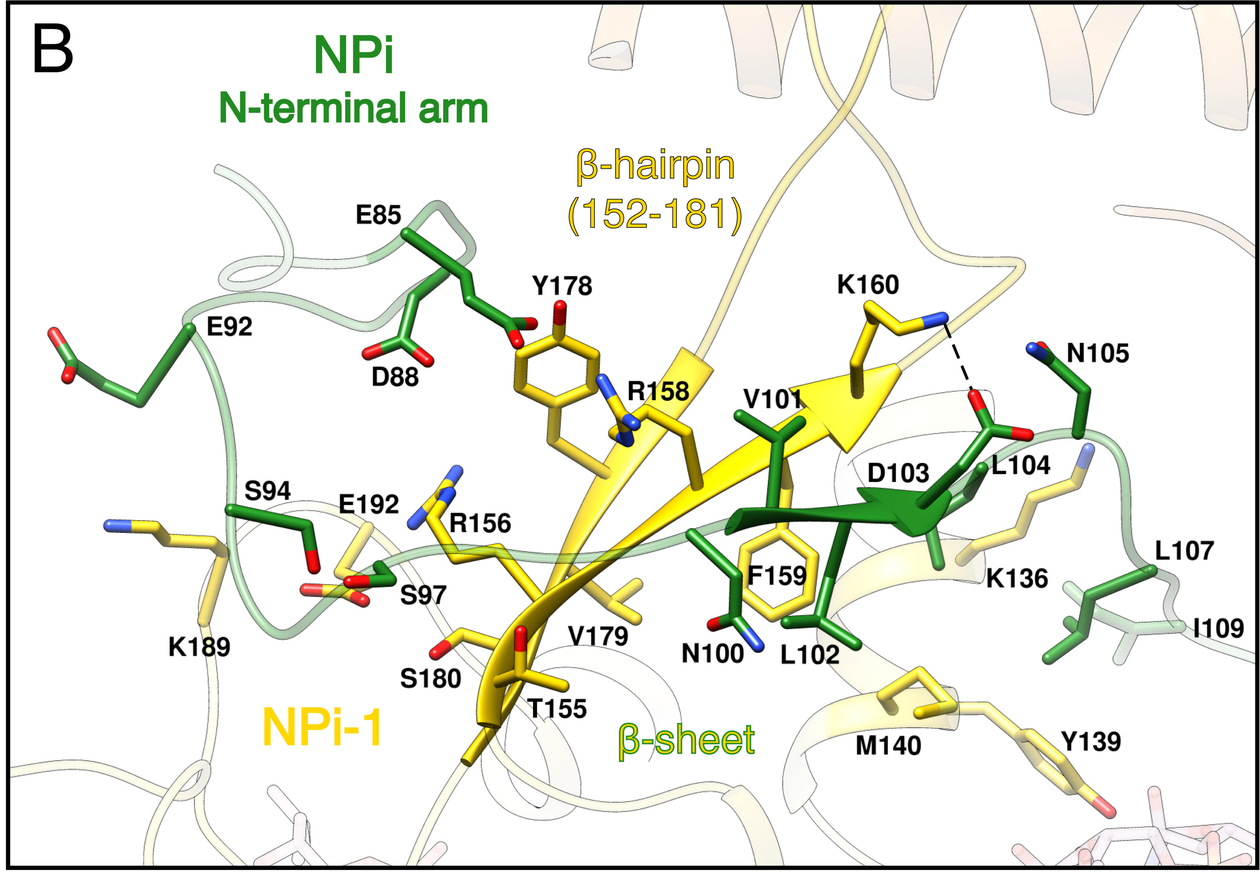
S97

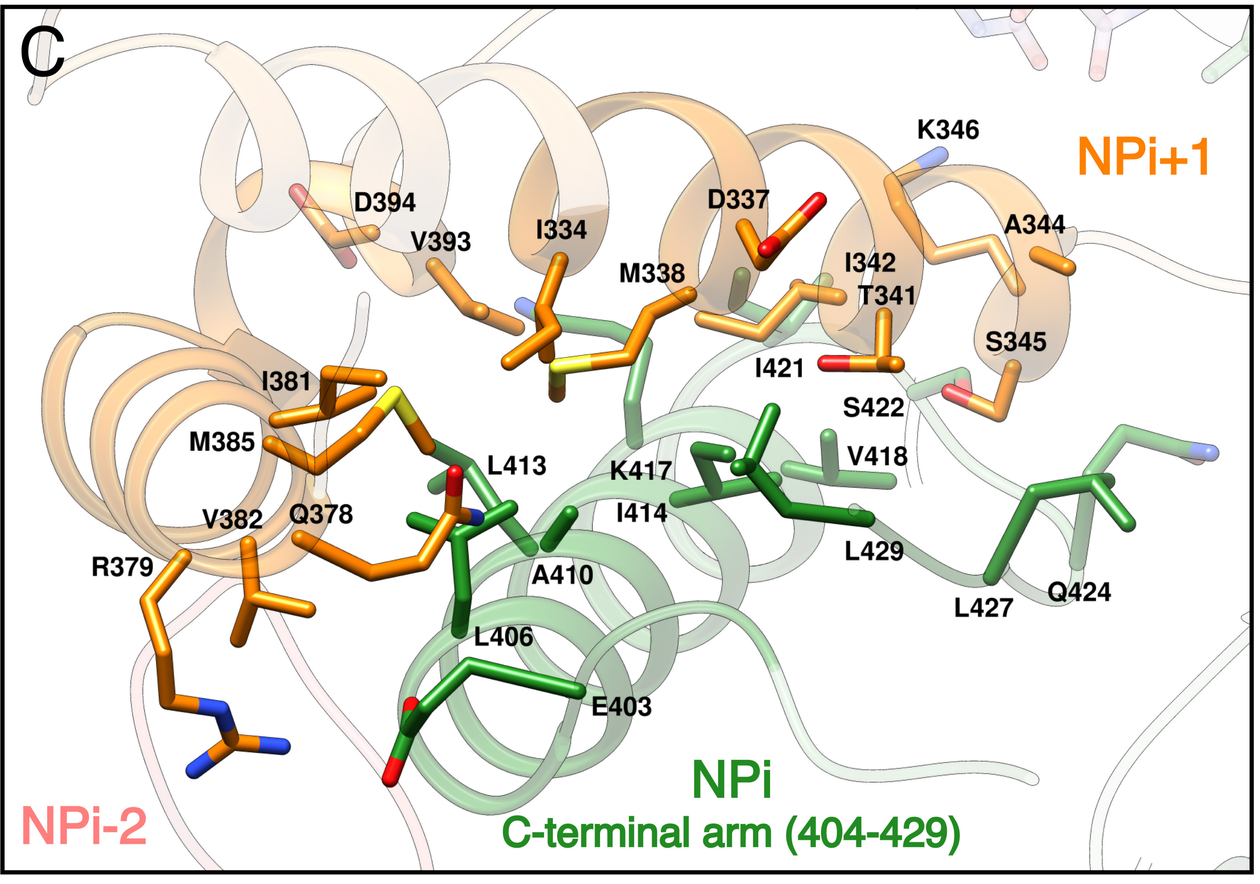
Y98

N-terminal arm

Figure 1


A
NPi+2
NPi+3
NPi+1
D
409-419
385-390
162-175
C
334-346
B
378-394
85-110
155-160
404-429
177-181
NPi
NPi-2
NPi-3
NPi-1


B
NPi
N-terminal arm
β-hairpin
(152-181)
E85
Y178
K160
E92
D88
R158
V101
N105
D103
L104
S94
E192
R156
L107
F159
K136
S97
I109
K189
V179
N100
L102
S180
T155
NPi-1
β-sheet
M140
Y139


C
NPi+1
K346
D394
V393
I334
D337
A344
M338
I342
T341
S345
I421
I381
S422
M385
L413
K417
V418
V382
Q378
I414
R379
L429
A410
L427
Q424
L406
E403
NPi
C-terminal arm (404-429)
NPi-2

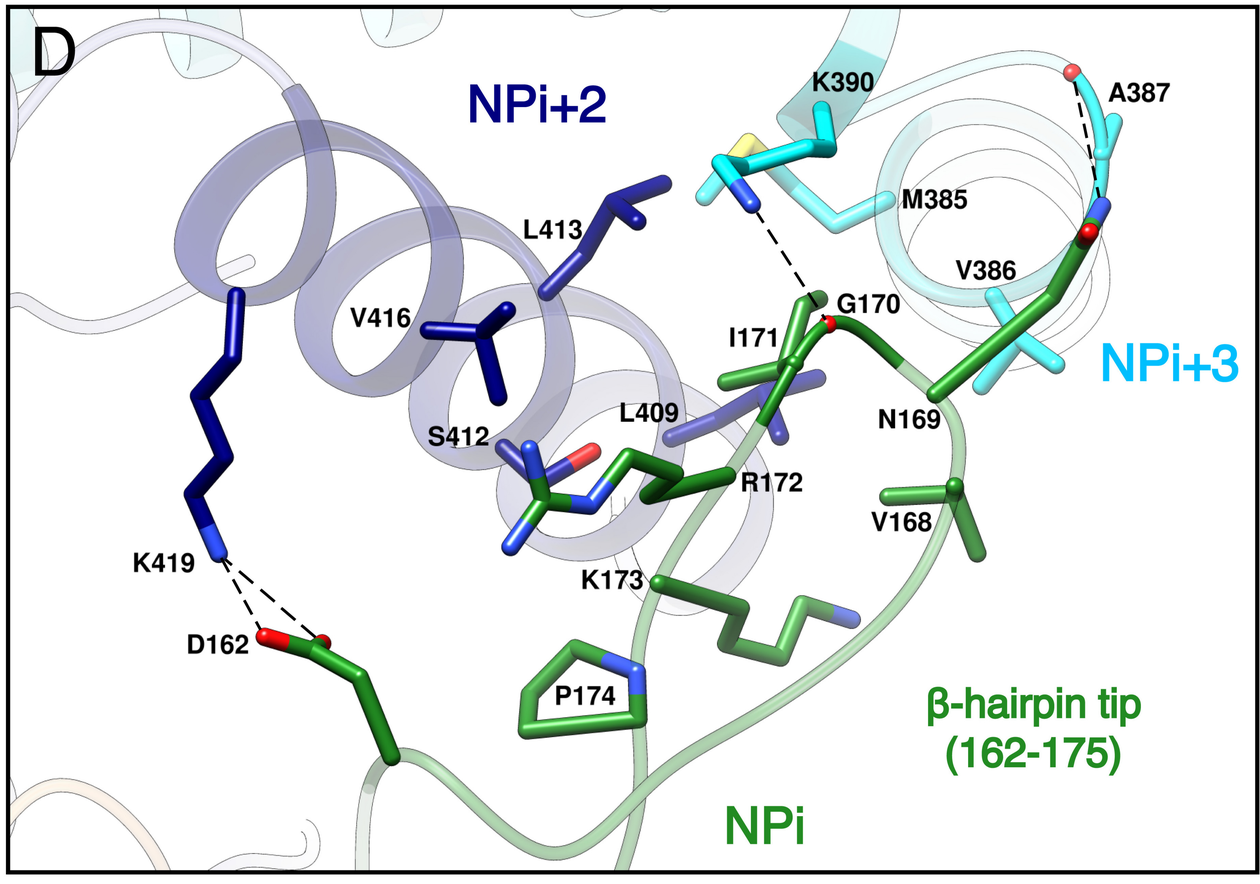
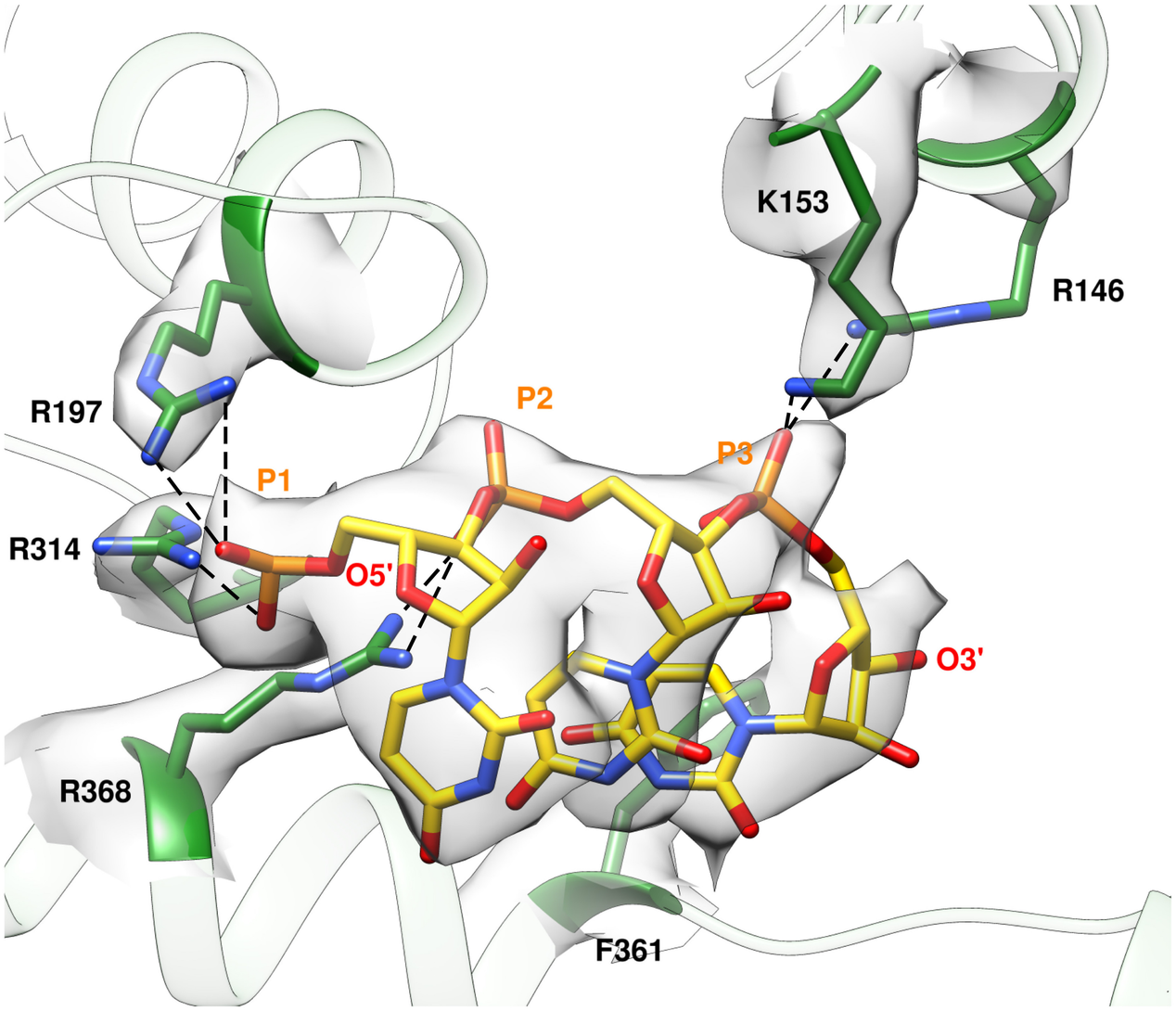
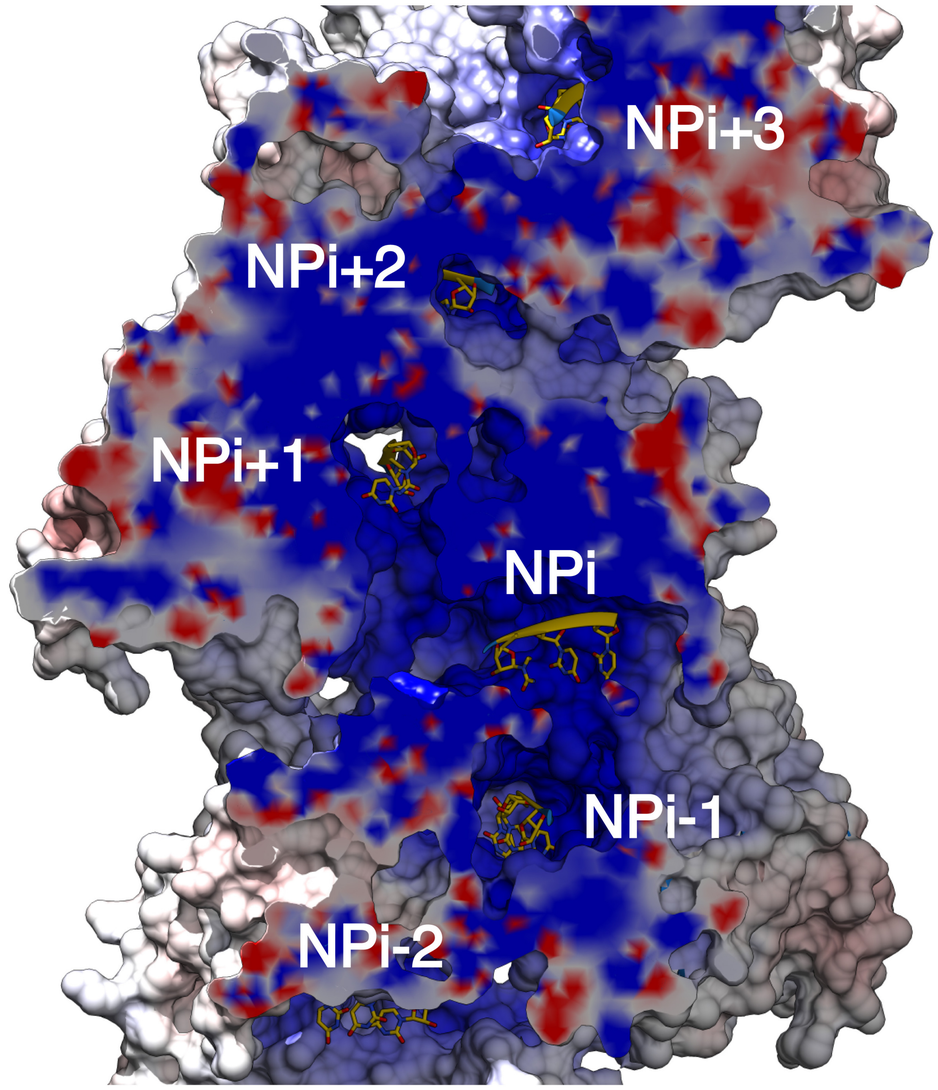

D
NPi+2
K390
A387
M385
L413
V386
G170
V416
I171
NPi+3
L409
N169
S412
R172
V168
K419
K173
D162
P174
β-hairpin tip
(162-175)
NPi

Figure 2

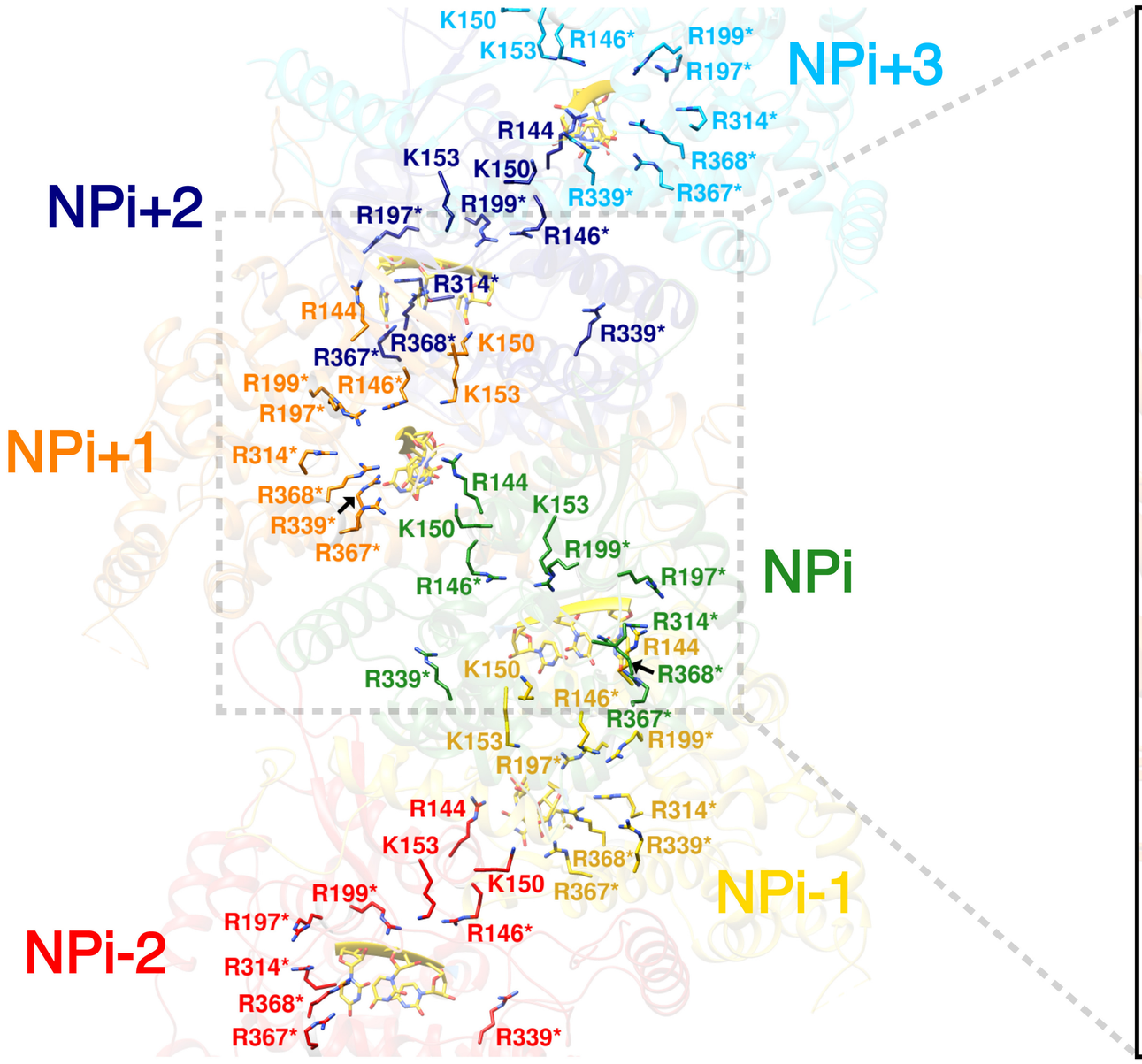
A


K153
R146
R197
P2
P3
P1
R314
O5'
O3'
R368
F361

B


NPi+3
NPi+2
NPi+1
NPi
NPi-1
NPi-2

C


NPi+3
NPi+2
NPi+1
NPi
NPi-1
NPi-2

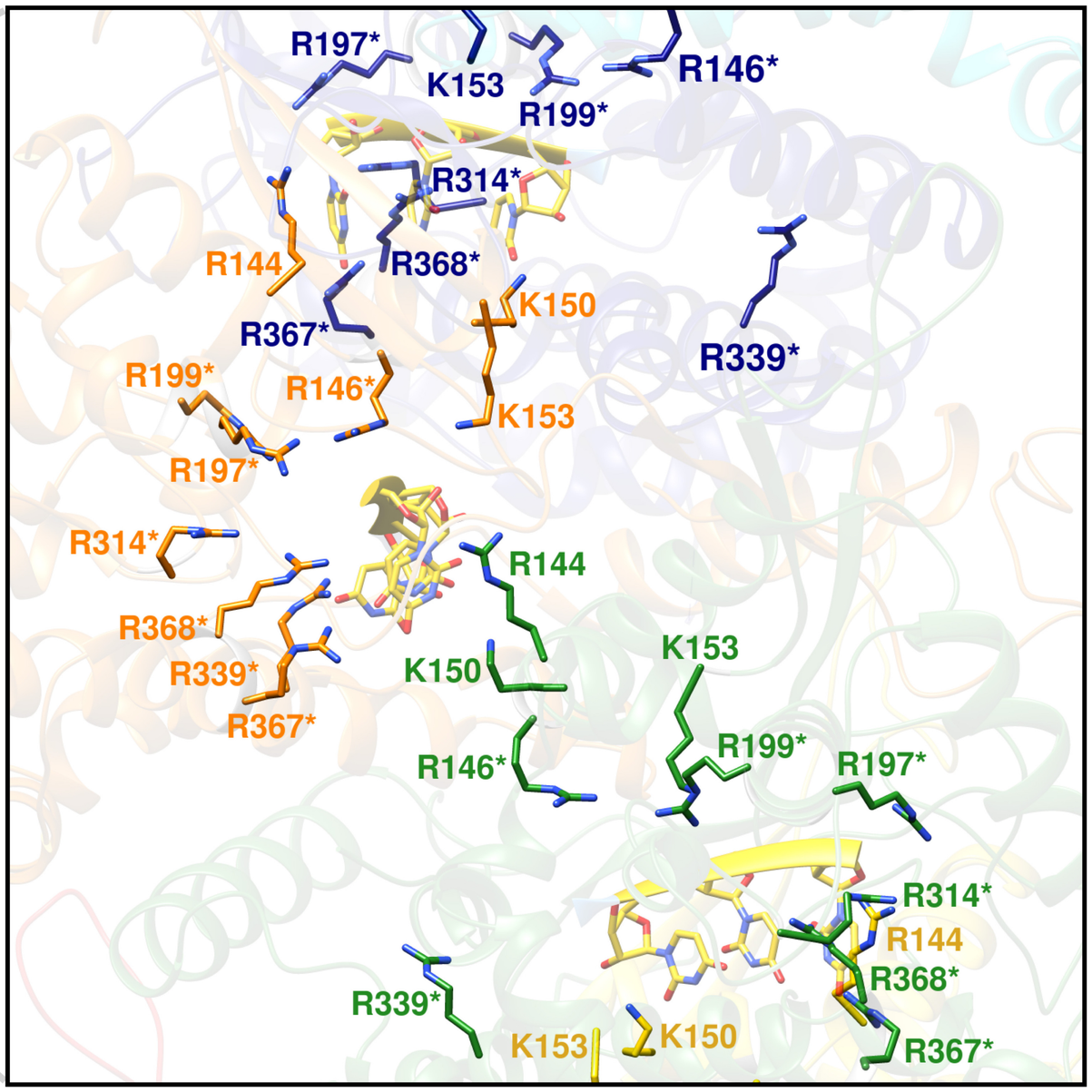

R197*
K153
R146*
R199*
R314*
R144
R368*
R367*
K150
R339*
R199*
R146*
K153
R314*
R144
R368*
R339*
K150
R367*
K153
R199*
R197*
R146*
R314*
R144
R368*
R339*
K150
R367*

Figure 3

A

Helices
334-346
and 378-394

Ct_{hinge}
399-403

Ct_{arm}
404-429

Nt_{arm}
80-107

β-hairpin
152-181

Nt_{hinge}
108-116

Nter
1-73

B

NP_{i-1}

NP_i

NP_{i+1}

NP_{i+2}

Viral RNA 5’

3’

C

Viral RNA
5’

NP_{i-1}

Flexible Ct_{hinge}
399-403

NP_{i+2}

NP_{i+1}

3’

NP_i

Active site

3’

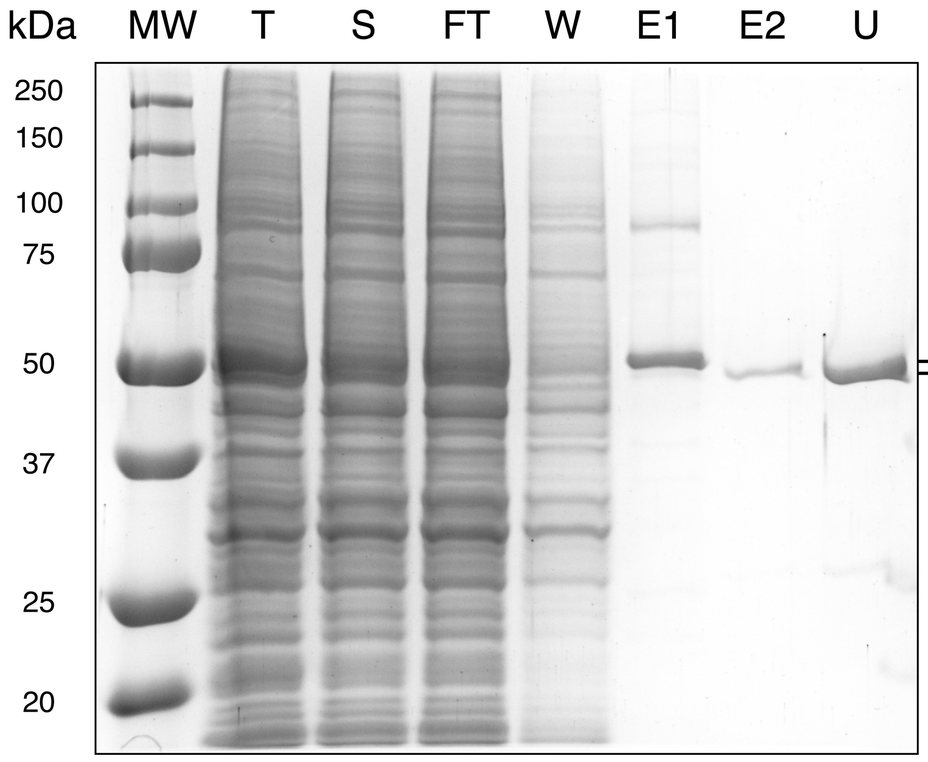
Polymerase

5’

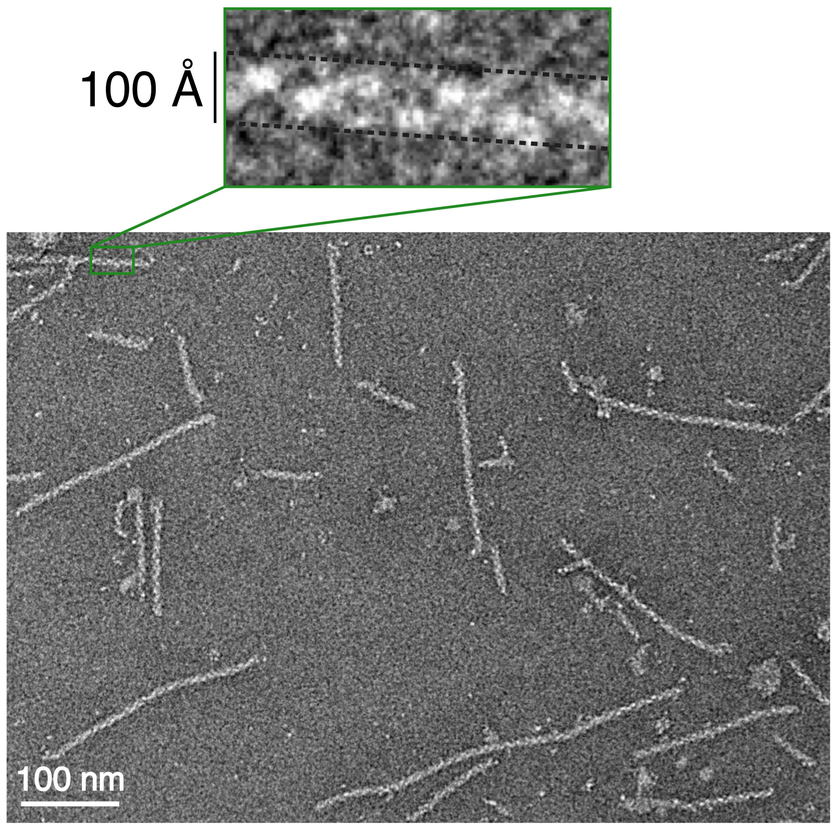
Transcribed/Replicated RNA

Figure 4

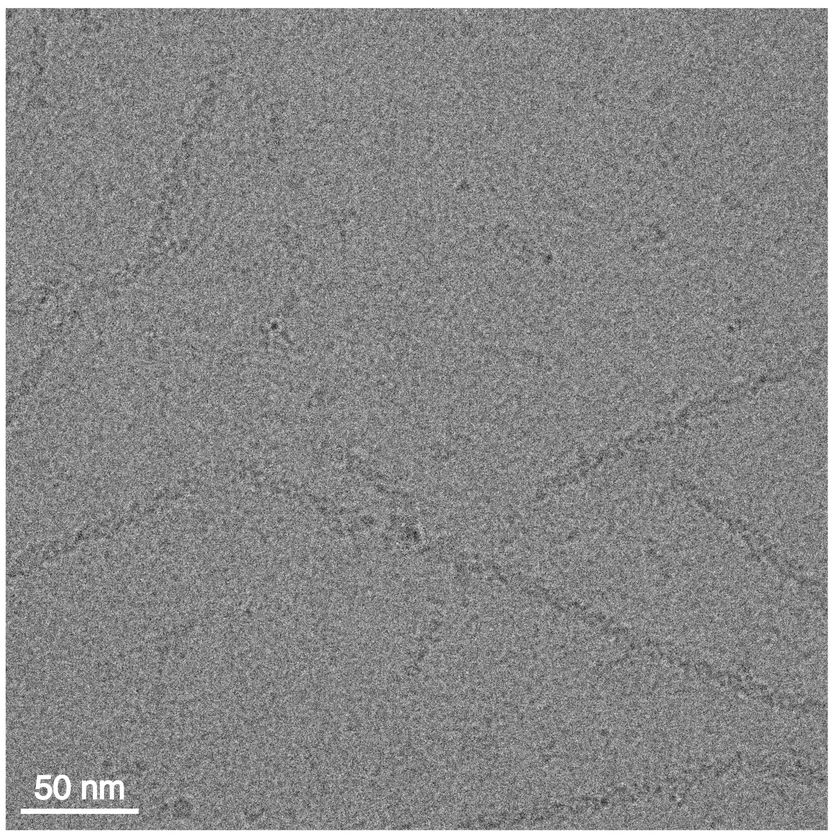
A



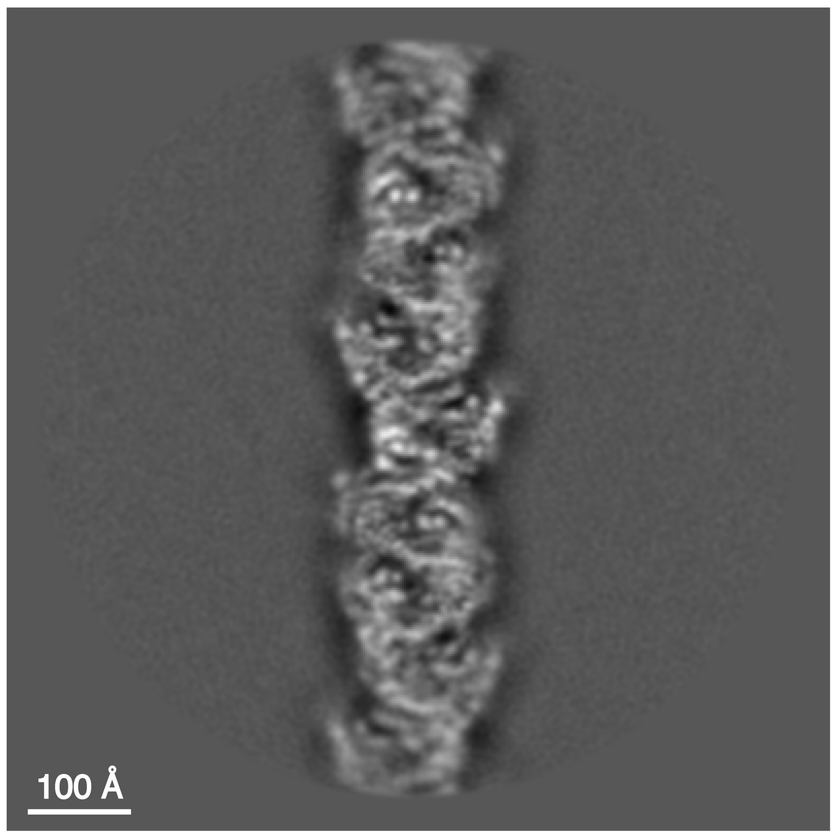
B



C



D



E

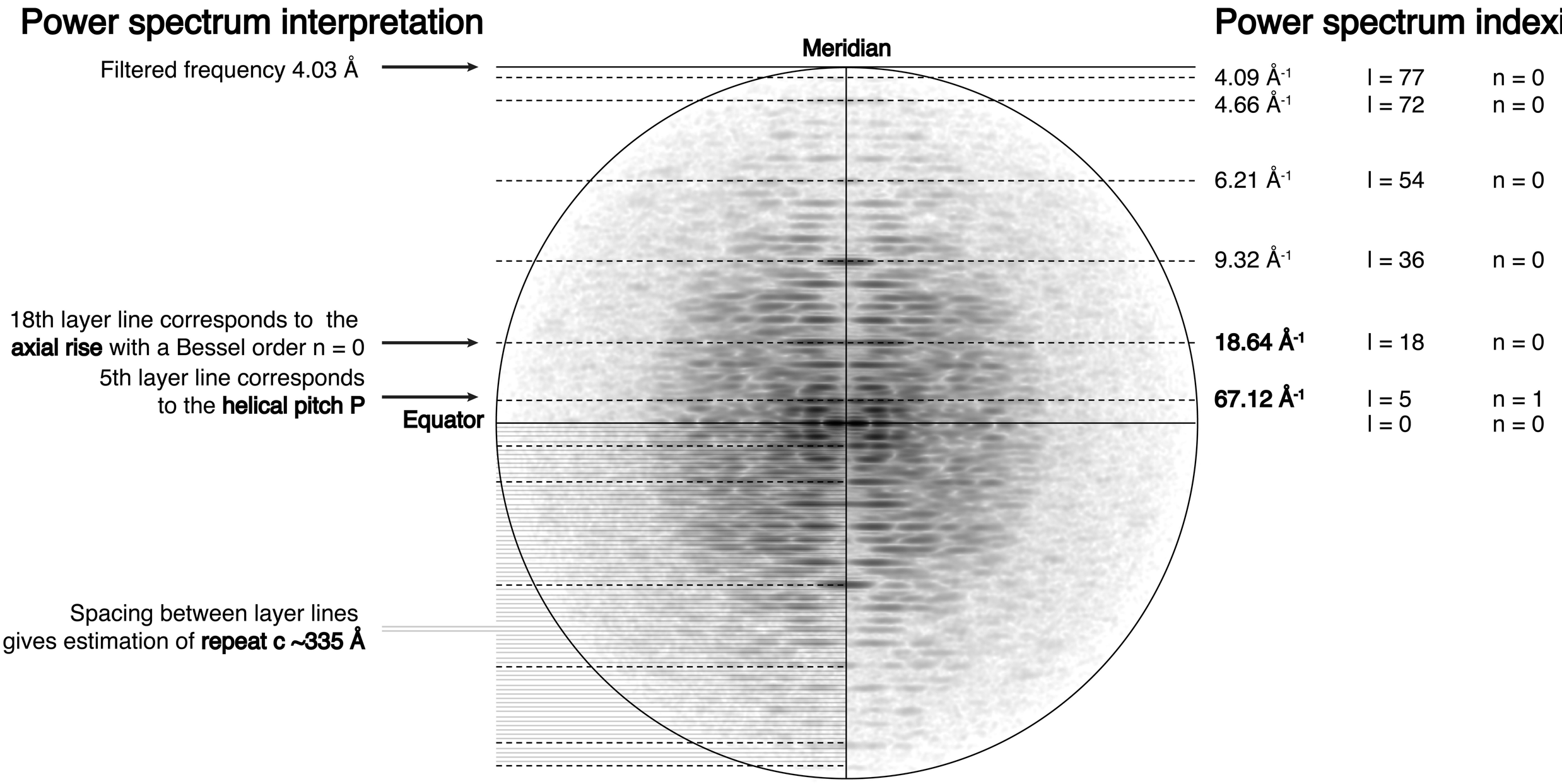
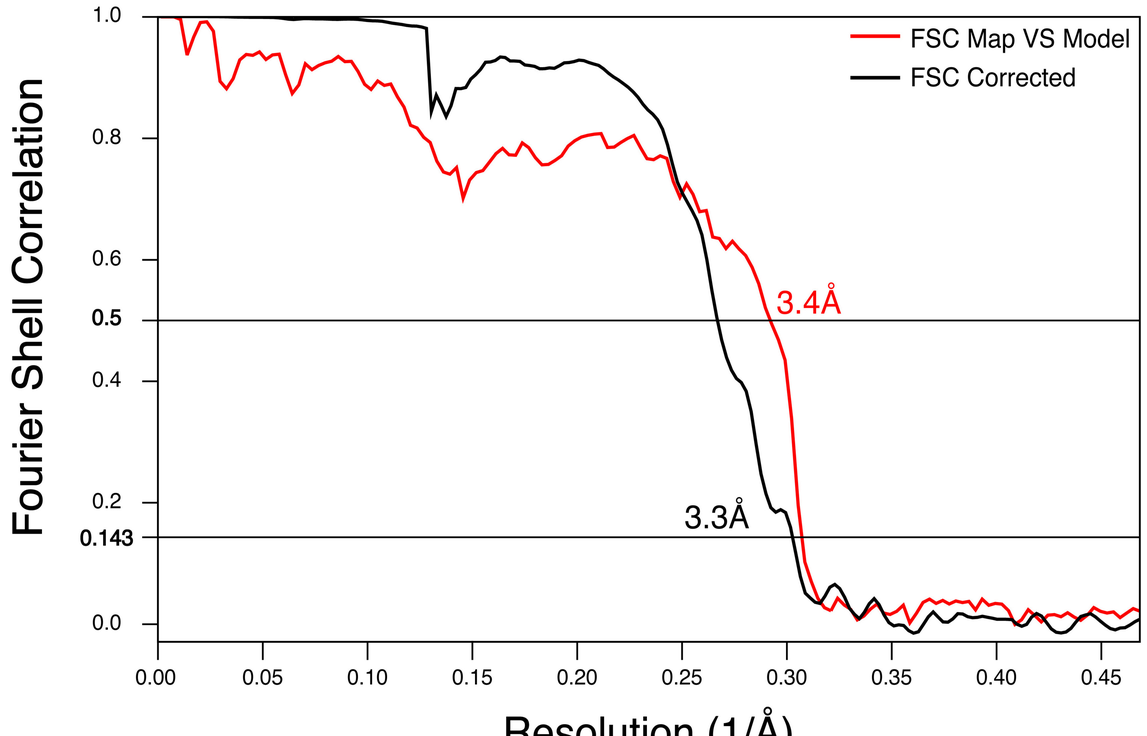
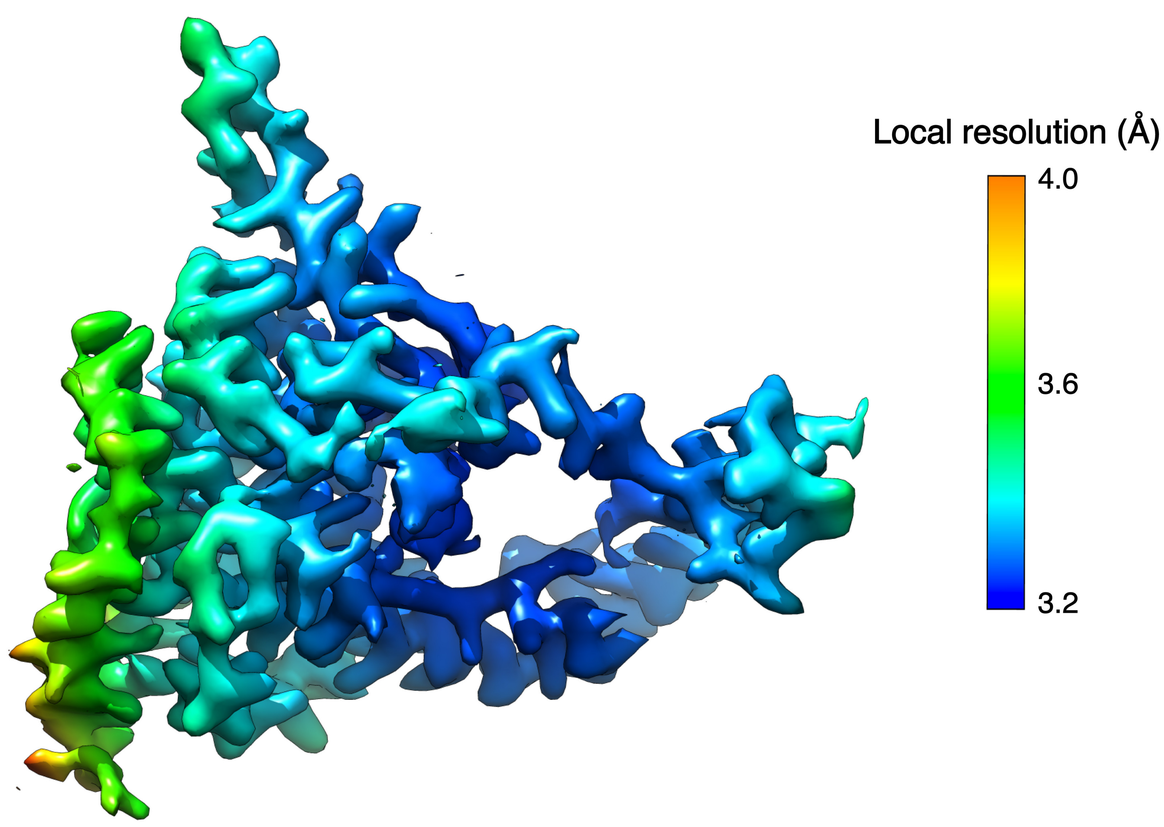


Figure 1 - figure supplement 1

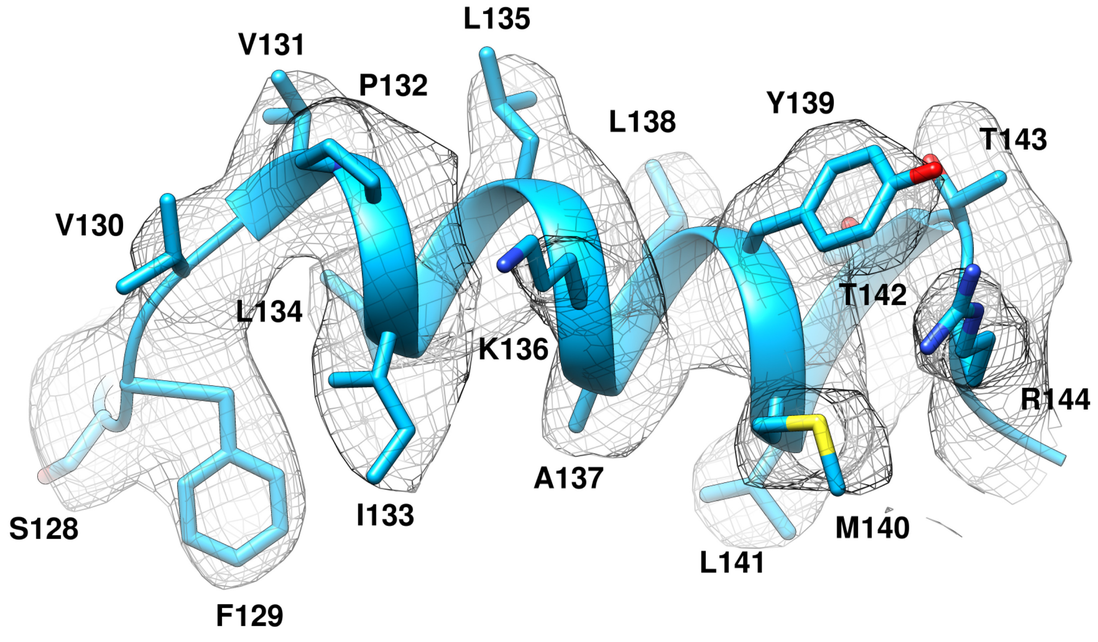
A



B



C



D

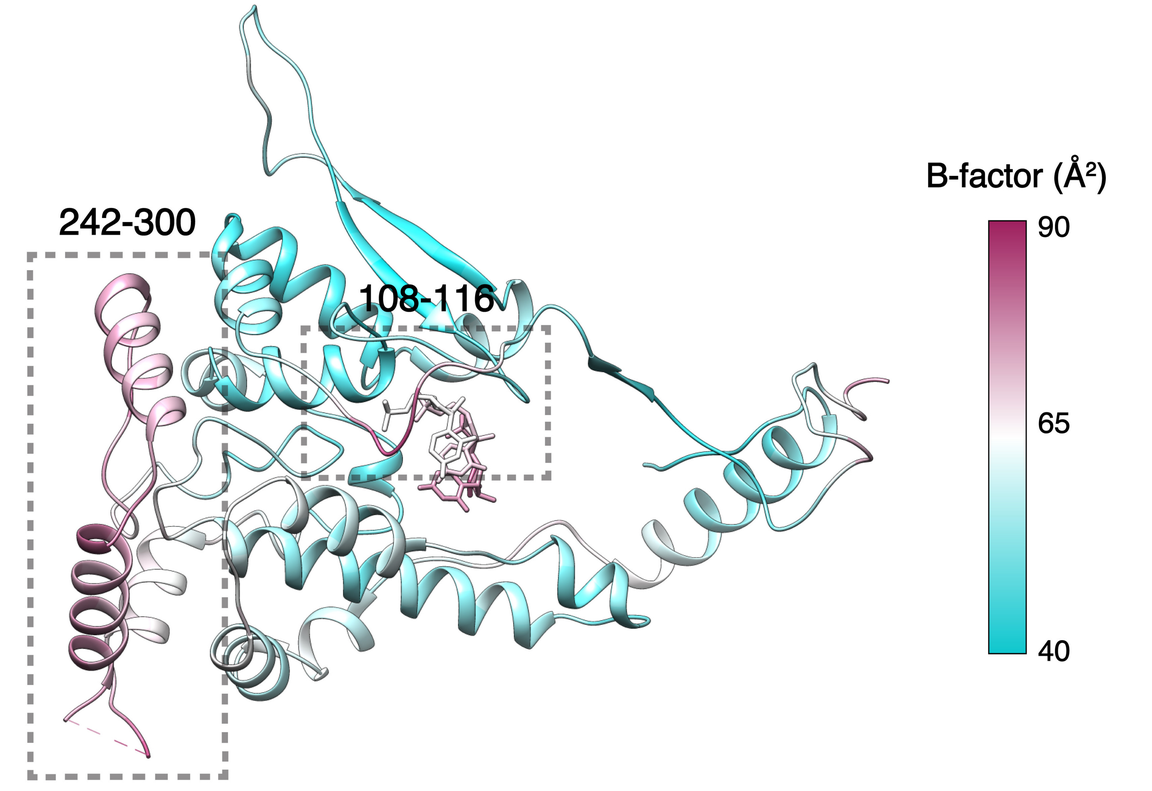
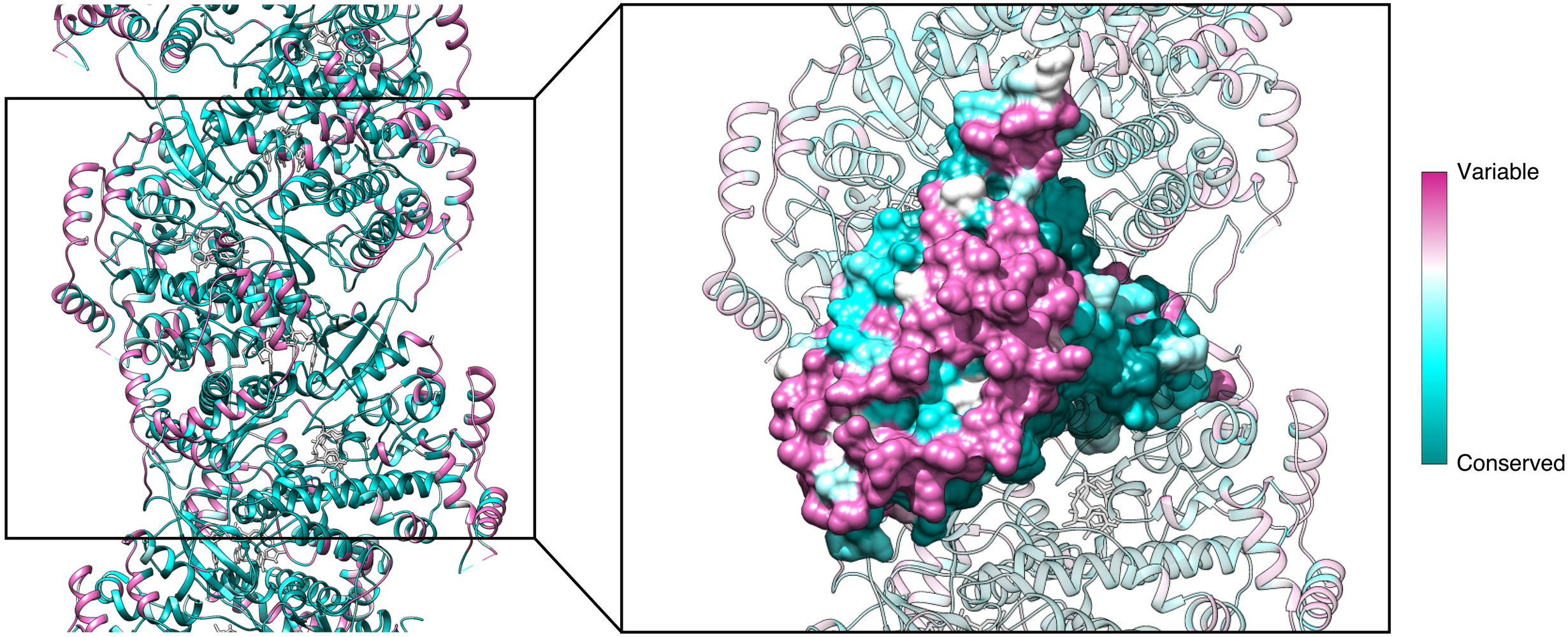


Figure 1 - figure supplement 2

A



B

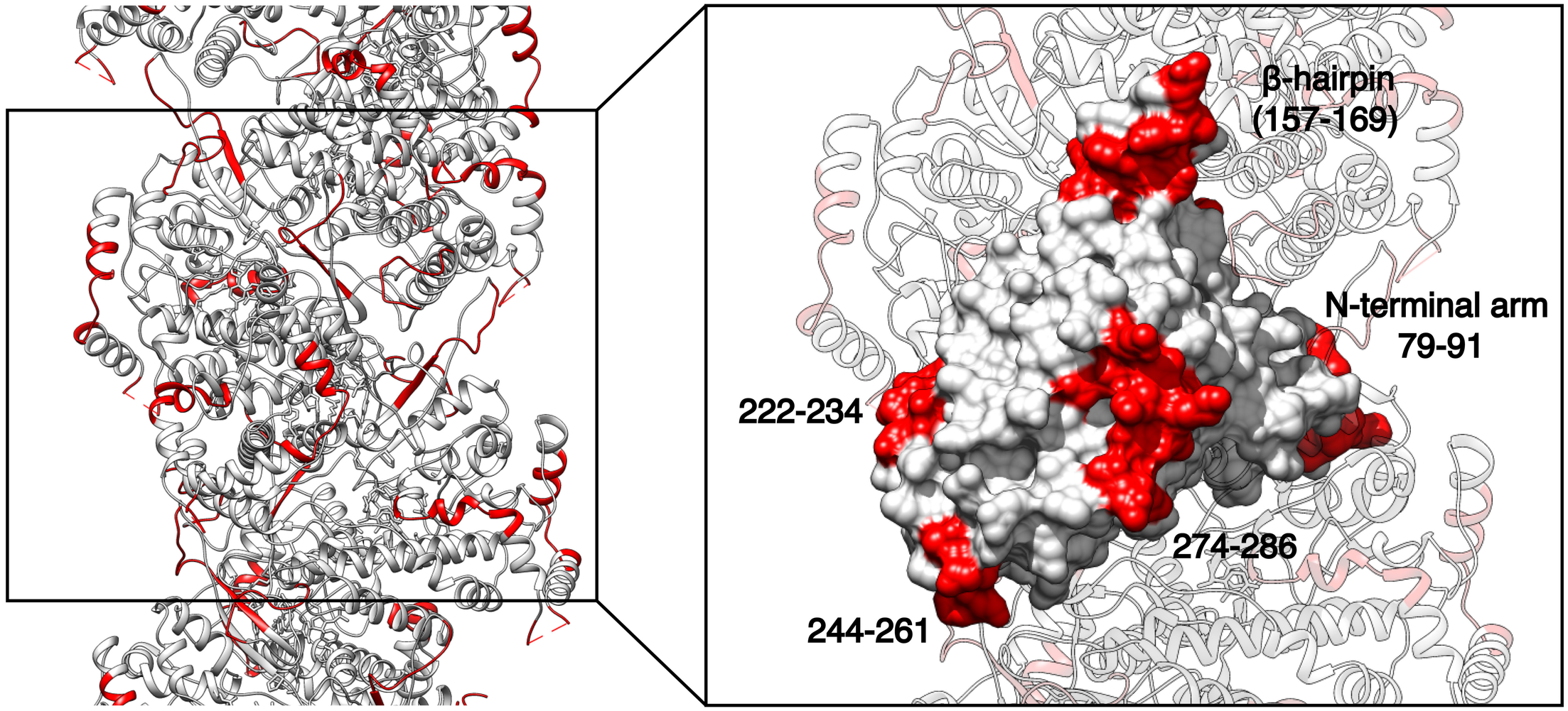
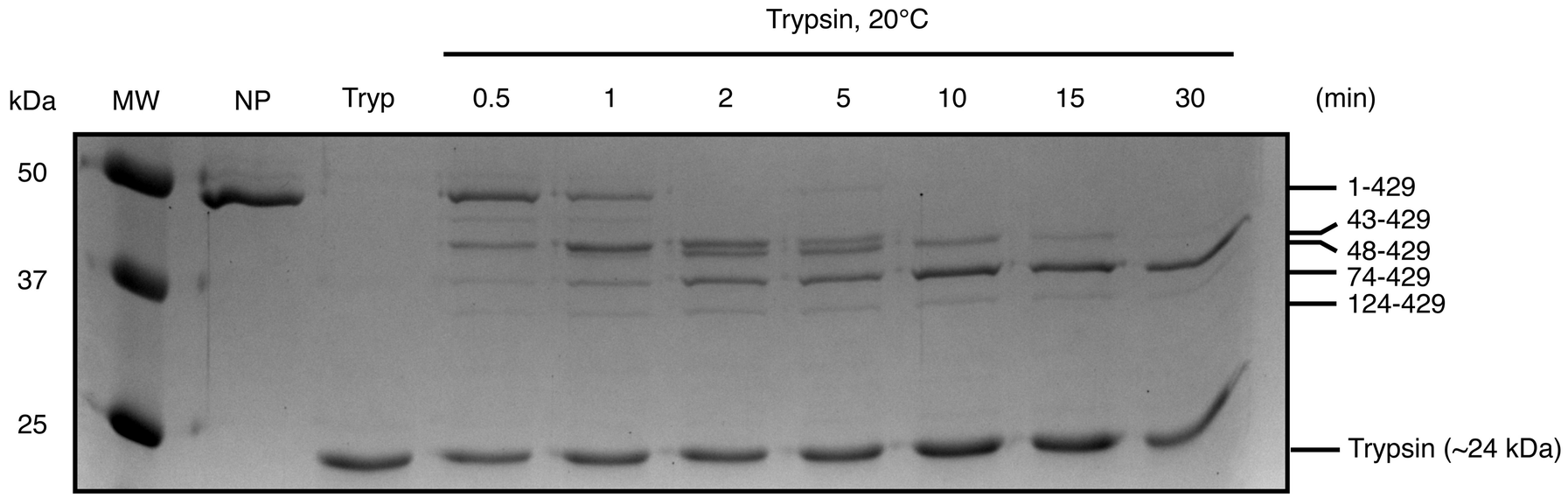
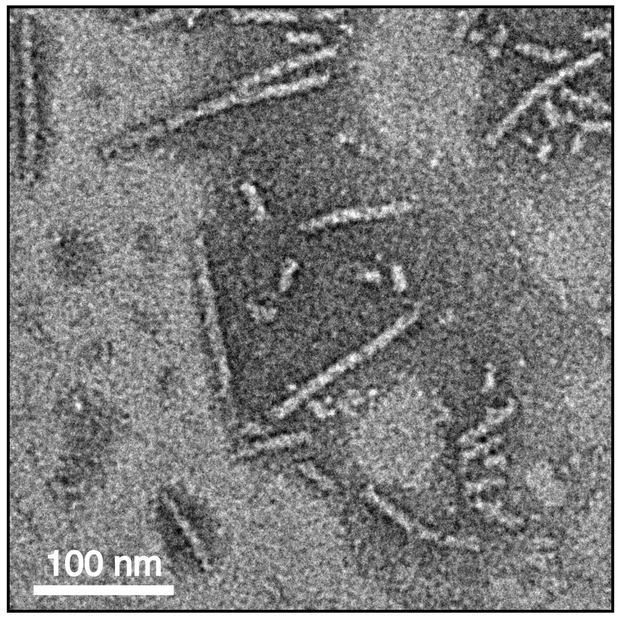


Figure 1 - figure supplement 3

A



B



C

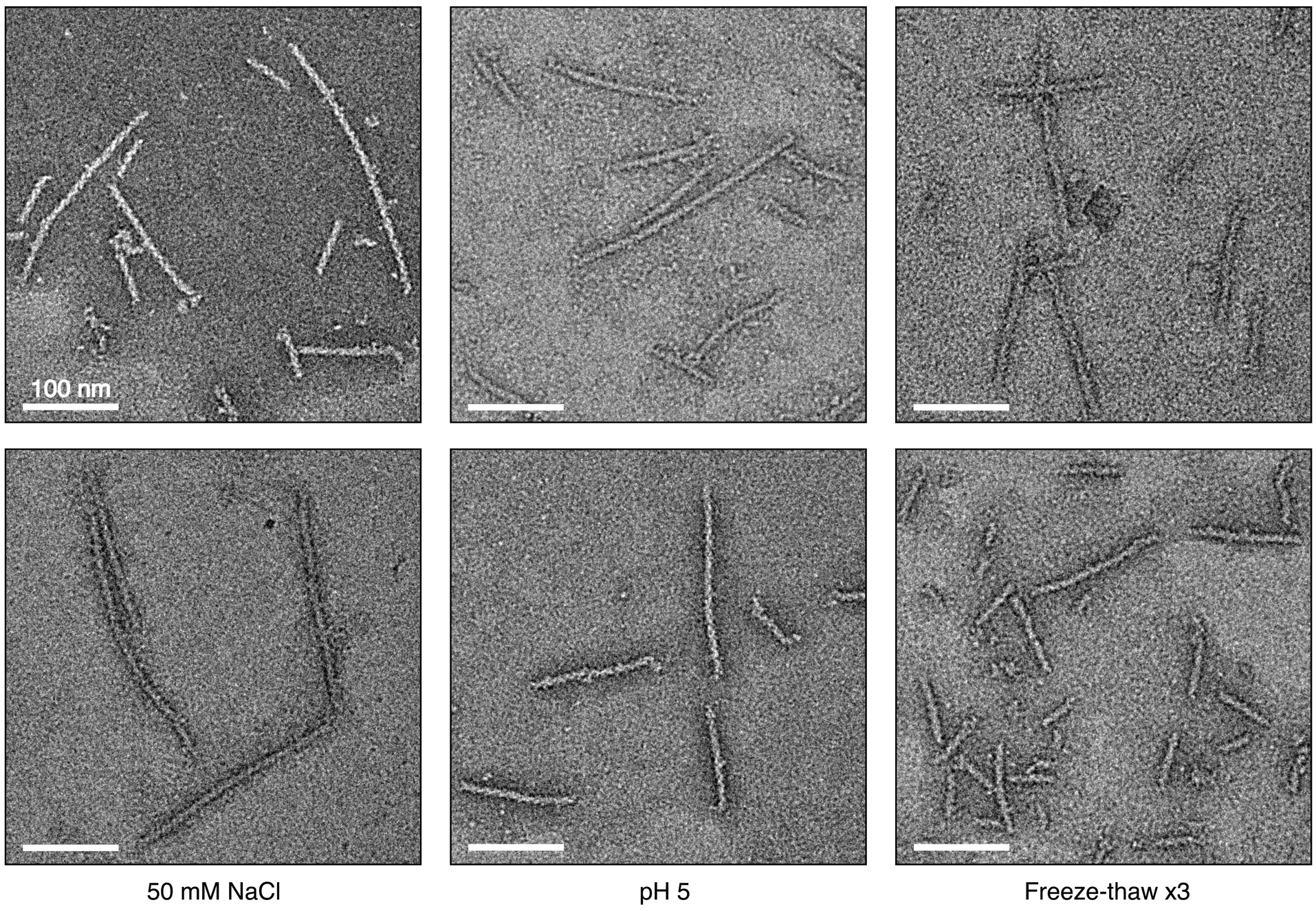
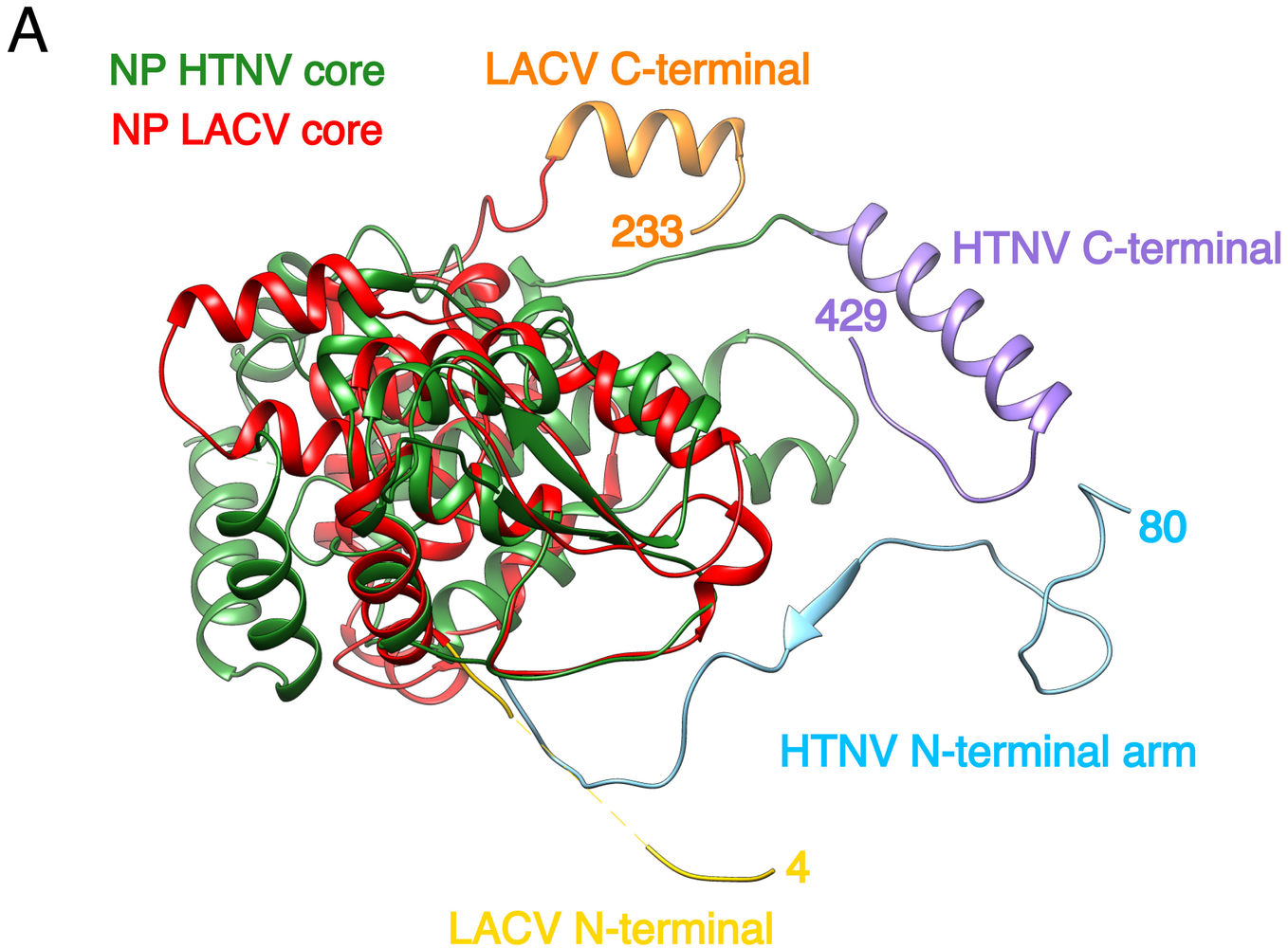
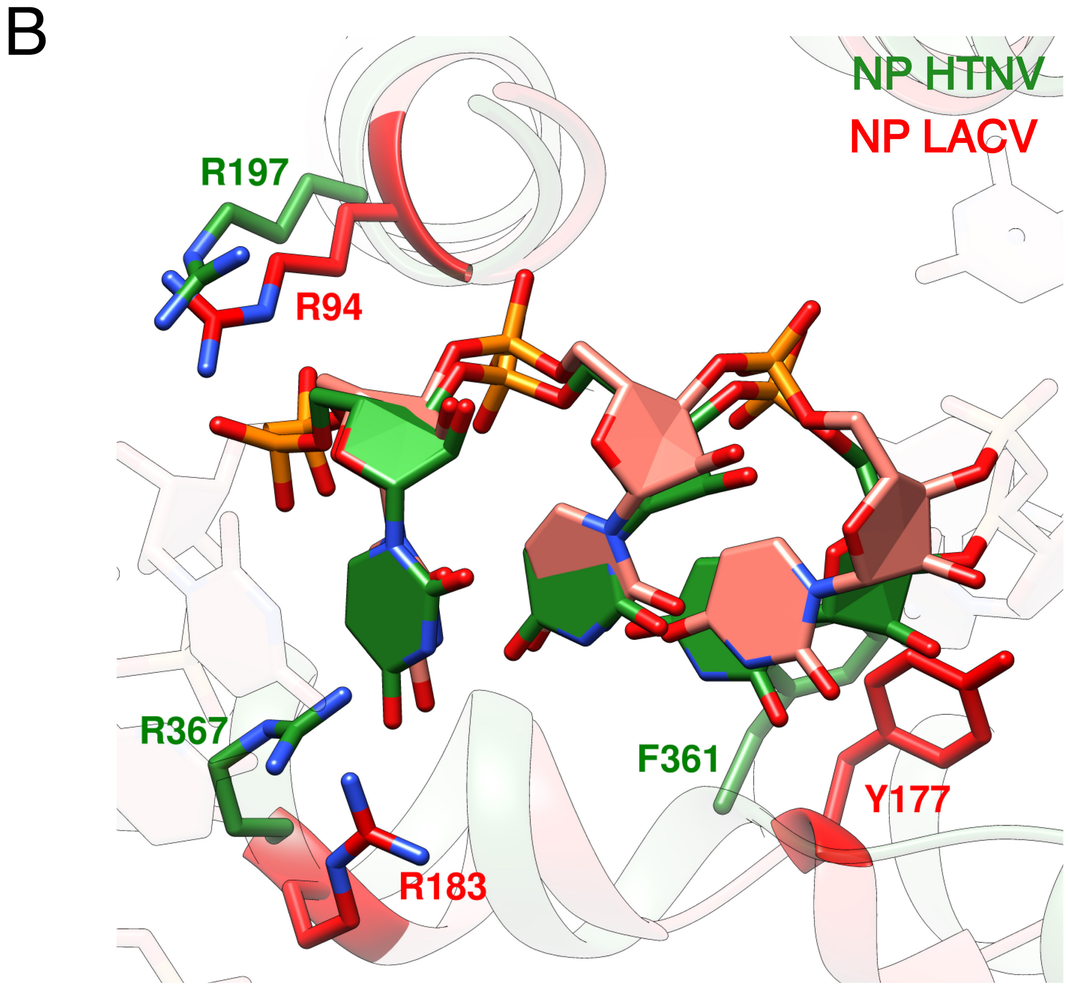


Figure 2 - figure supplement 1





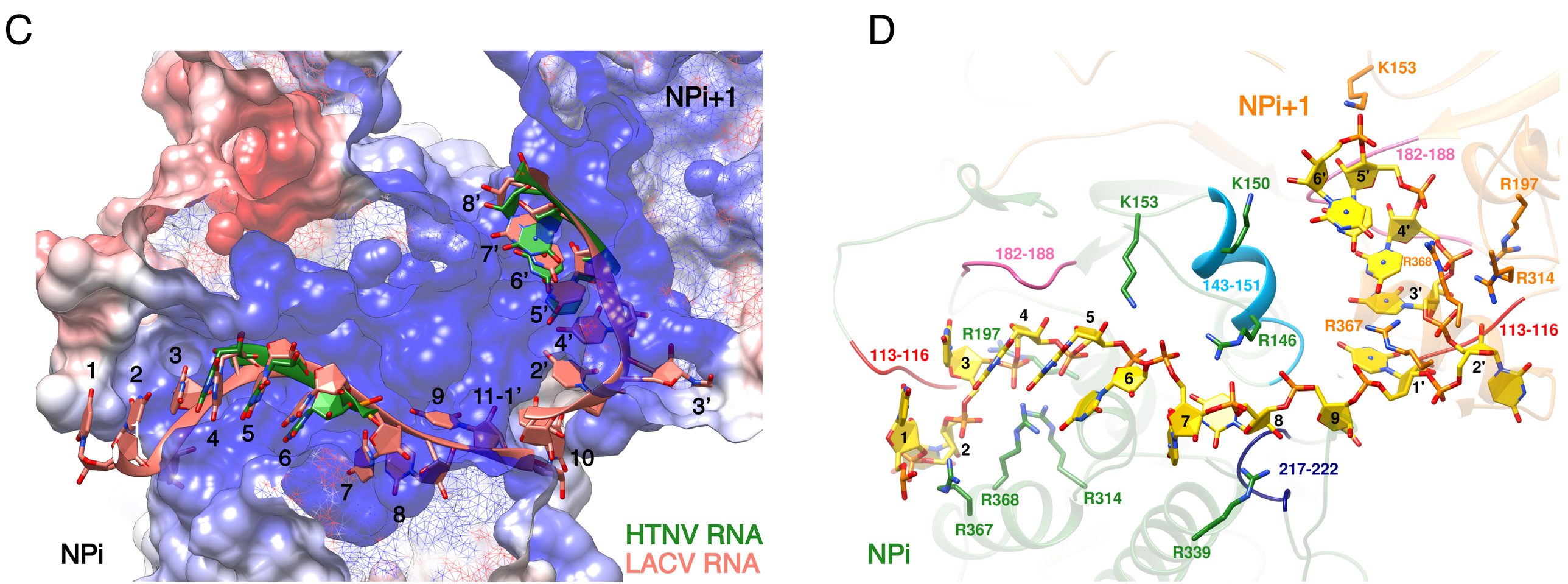


Figure 3 - figure supplement 1