Despite the odds: formation of the SARS-CoV-2 methylation complex.

Alex Matsuda[1], Jacek Plewka[1], Yuliya Chykunova[1,2], Alisha N. Jones[3,4], Magdalena Pachota[1,2], Michał Rawski[1], André Mourão[3], Abdulkarim Karim[1], Leanid Kresik[1], Kinga Lis[1,6], Igor Minia[5], Kinga Hartman[7], Ravi Sonani[8], Florian Schlauderer[3], Grzegorz Dubin[8], Michael Sattler[3,4], Piotr Suder[7], Grzegorz Popowicz[3,4,*], Krzysztof Pyrć[1,*], Anna Czarna[1,*]

1. Virogenetics Laboratory of Virology, Malopolska Centre of Biotechnology, Jagiellonian University, Gronostajowa 7a, 30-387 Krakow, Poland.
2. Microbiology Department, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Gronostajowa 7, 30-387 Krakow, Poland.
3. Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany
4. Bavarian NMR Center, Department of Chemistry, Technical University of Munich, Lichtenbergstr. 4, 85748 Garching, Germany.
5. Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical Systems Biology, Berlin 13125, Germany.
6. Faculty of Chemical Engineering and Technology, Cracow University of Technology, Warszawska 24, 31-155, Poland.
7. Department of Analytical Chemistry and Biochemistry, Faculty of Materials Science and Ceramics, AGH University of Science and Technology, Mickiewicza 30, 30-059 Krakow, Poland.
8. Protein Crystallography Research Group, Malopolska Centre of Biotechnology, Jagiellonian University, Gronostajowa 7a, 30-387 Krakow, Poland.

* co-senior and corresponding authors.

Abstract
Abstract
Coronaviruses protect their single-stranded RNA genome with the methylated cap added during the replication. This capping process is carried out by several nonstructural proteins (nsp) encoded in the viral genome. The methylation itself is performed consecutively by two methyltransferases, nsp14 and nsp16, which interact with nsp10 protein acting as a co-factor. The nsp14 protein also carries the exonuclease domain, which also serves as a part of the proofreading system during the replication of the large RNA genome. The available crystal structures suggest that the concomitant interaction between these three proteins is impossible due to the structural clash, and it is generally accepted that the nsp16 and nsp14 bind with the nsp10 separately.

Here, we show that nsp14, nsp10, and nsp16 form a methylation complex despite the odds. Due to spatial proximity, this interaction is beneficial for forming mature capped viral mRNA. Further, it modulates the exonuclease activity of nsp14, protecting the viral RNA at the replication site. Our findings show that nsp14 is more amenable to allosteric regulation and may serve as a molecular target for the therapy.

Introduction

The coronaviral genome is of positive polarity, and it serves as a substrate for the translational machinery of the cell after its release to the cytoplasm. The first and only product of the genomic mRNA translation is the large and non-functional 1a/1ab polyprotein. It maturates by autoproteolytic processing carried out by two viral proteases – M^{pro} and PL^{pro}. This leads to the generation of a set of nonstructural proteins (nsp), which are numbered consecutively and responsible for the whole viral replication process and remodelling of the intracellular environment. Once nsp's reshape the cell to form a viral factory, the genomic RNA is being copied, and a set of subgenomic (sg) mRNAs is produced in a peculiar discontinuous transcription process. These sg mRNAs are monocistronic and serve as templates for the production of structural and accessory proteins required for the formation, assembly and release of progeny viruses[1]. The activity of particular nsp's has been previously described, showing the complex network of interactions and presence of multi-functional nsp's. However, their coordinated action has not been fully understood.

Works by Gao *et al.*[2], Yan *et al.*[3], *Wang et al.*[4] and Kabinger *et al.*[5] shed light on the scaffold of the replicatory complex formed by nsp12 (polymerase) and two co-factors nsp7 and nsp8, that create functional machinery able to replicate the viral RNA. Next, an extended elongation complex was described, where the nsp12/7/8 is accompanied by nsp13 helicase. This complex is suggested to serve as the basic replicatory module[6].

Coronaviruses are known for their large genomes requiring high replication fidelity to maintain their integrity. While the SARS-CoV-2 nsp12 polymerase is highly processive, it is error-prone and does not provide sufficient fidelity. Some time ago, it was shown that a proofreading system is encoded in coronaviral genomes. The nsp14 protein carries an N-terminal exonuclease (ExoN) domain that has been attributed to serve as a part of this system. The ExoN is a member of the DEDDh exonuclease superfamily, which possesses a 3'-5' exonuclease activity and is responsible for removing incorrectly incorporated nucleotides from the 3' terminus of the newly formed RNA. It has also been proposed to play a role during the discontinuous replication of coronaviruses. The nsp14 has been shown to associate with the replicatory complex, with the nsp10 as a co-factor that modulates and enhances the nsp14 exoribonuclease activity[7-9].

The nsp14, however, has yet another function. It takes part in the cap formation after the genome copying is finalized[7]. The capping of the viral mRNAs is essential for their function and integrity and enables translation initiation. Additionally, it protects the mRNA from recognition as foreign by cellular sensors and consequently induction of innate immune

responses[10]. Cap formation is a tightly regulated process consisting of four independent enzymatic reactions. First, the nsp13 triphosphatase removes the γ-phosphate of the 5′-triphosphate end (pppA)[11,12]; next, nsp12 guanylyltransferase (the nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain)[3] transfers the GMP to form the core structure of the cap (GpppA); the GpppA is methylated at the N7 position by the nsp14 N7-methyltransferase domain (^{7Me}GpppA); subsequently, the ribose in the first ribonucleotide is methylated at the 2'-O-position by the nsp16 2'-O-methyltransferase[13-15]. This leads to the formation of a functional cap, and the genome replication is concluded (^{7Me}GpppA$_{2'OMe}$). Interestingly, this process is regulated by the nsp9 protein, which binds the nsp12 near the NiRAN active site[16]. While there is a good basis for understanding the part of the capping process performed by interacting nsp12, nsp13, and nsp14 proteins, the nsp16 remains an orphan.

Here, we studied the interaction between the two methyltransferases - nsp16 and nsp14. While previous *in silico* results suggested that the interaction of nsp14 and nsp16 is impossible due to the steric hindrance on the nsp10 surface, the experimental data proves otherwise. Our data indicate the formation of a functional heterotriplex by nsp10, nsp16 and nsp14. The interaction occurs via the nsp10 protein, where the binding interfaces overlap to some extent. This should render the triplex formation impossible unless structural rearrangement of the nsp14 N-terminal region occurs. Interestingly, in contrast to the remaining part of nsp14, this region is characterized by the absence of secondary structure. Additionally, the nsp10/14(N-terminal) interactions are devoid of "anchors" typical for protein-protein interactions (PPI). There are no deeply buried lipophilic residues, solvent-protected hydrogen bonds and precise shape complementarity present. These anchors appear only in the other part of nsp14. There, the N-terminal α1′ helix of nsp10 forms an interaction with the deep, lipophilic cleft of nsp14 – a very typical PPI interface (**Figure 1**). Therefore, we hypothesized that the triplex formation is possible from the structural point of view, following a rearrangement of the first 50 N-terminal residues of nsp14.

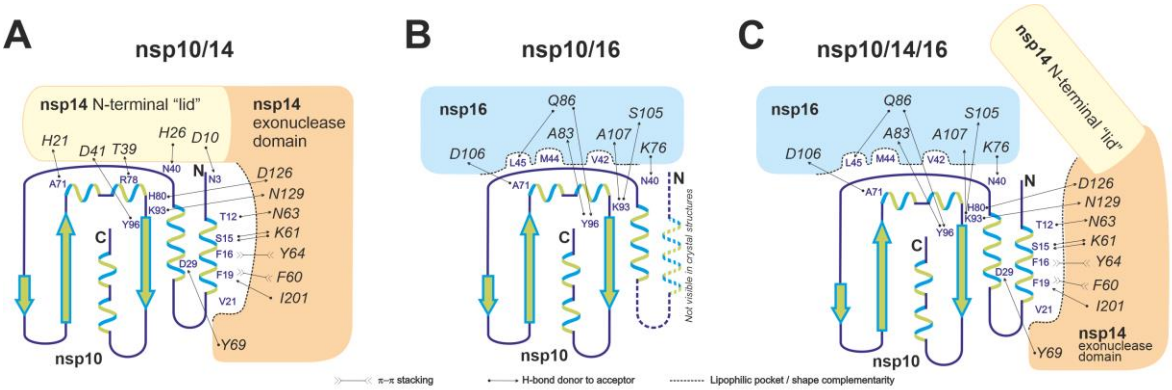

Figure 1. Schematic diagram of the interaction between nsp10, nsp14 and nsp16. (A) Interactions between nsp10 and nsp14 are mainly located within the exonuclease domain, where the α1 helix of nsp10 provides several deeply buried lipophilic and π-stacking interactions. This region is characterized by well-developed shape complementarity. The "lid" region of nsp14, composed of residues 1-50, covers the nsp16 binding site. This region is poorly structured and has poor shape complementarity with the nsp10. Therefore, it is freezing in nsp10-bound form carries a high entropy cost. Only several partially solvent-exposed hydrogen bonds can be identified there. **(B)** In contrast, nsp16 interacts in a similar region as the "lid" of nsp14 but with a well-developed interface involving deep lipophilic pockets and hydrogen bonds shielded from the solvent. The α1 helix of nsp10 is not visible in crystal structures of the nsp10/16 complexes (6W4H, 6YZ1), indicating its flexibility. **(C)** The formation of the ternary nsp10/14/16 complex must be accompanied by a "lid" opening. This can be plausible due to poor interface features and a lack of secondary structure that provides flexibility.

While the need for complex formation is understandable due to the complementary function of the two methyltransferases, we also observed that the nsp16 interaction with the nsp10/14 modulates the exoribonuclease activity. Consequently, we propose that the formation of the complex also regulates the nsp14 activity and limits the non-specific viral RNA degradation observed for the nsp10/14 complex. It is also to be elucidated to what extent such activity may tweak the degradation of cellular mRNA.

Results

Molecular modelling of the nsp10/14/16 complex

High-resolution crystal structures are available for the nsp14 and nsp16 with their common co-factor nsp10. We investigated the data to determine which residues of nsp10 contribute to the binding to each partner. From all the deposited structures of the nsp10/16 complex, we chose one with an almost intact N-terminal α1′ helix of nsp10 (PDB ID: 6WVN) and the nsp10/14 exoribonuclease domain complex (PDB ID: 7DIY). These structures were aligned in PyMol. The nsp10 structure in both complexes is almost identical with RMSD of 0.5, with the α1′ helix flipping as indicated by the arrow in **Figure 2A**.

The overlay of both complexes depicts that nsp16 (cyan) and the first 50 amino acids of nsp14 (green) occupy the same positions. Moreover, both proteins exhibit a high degree of overlap on the nsp10 interfaces, with nsp14 having almost twice as large interface surface area.

The nsp14 interface with nsp10 also involves the N-terminal α1′ helix of nsp10 structurally positioning it at ~130° angle compared to nsp16, which does not interact with it. However, the N-terminal region of nsp14 is flexible and poorly structured.

Interacting residues on nsp14 are mainly present in unstructured loop regions such as the N-terminal coil-coiled region that interacts with nsp10 α1' helix, β-turn 1, and loops between β-sheet 2/3 and β-sheet 7/8. Many of those interactions are present within the region that would overlap with the nsp16 if those complexes are aligned (red region in **Figure 2A** and **2C**). The nsp10/16 interface, on the other hand, is based on a stronger hydrophobic interaction network between, among all, much more structurally rigid central antiparallel β1-sheet of nsp10 as well as helices α2, α3, α4 and a coiled-coil region connecting helix α1 and the sheet β1 and structured α helices 3,4 and 10 as well as β-sheet 4 (**Figure 2B**). Especially, Val42 and Leu45 of nsp10 seem to be essential for the interactions as they are immersed into hydrophobic pockets formed by helices α3, α4, and α10 of nsp16. We, therefore, hypothesize that within the nsp10/nsp14/nsp16 triplex, nsp16 displaces the N-terminal "lid" of nsp14 on the interface of nsp10 as depicted in **Figure 1**.

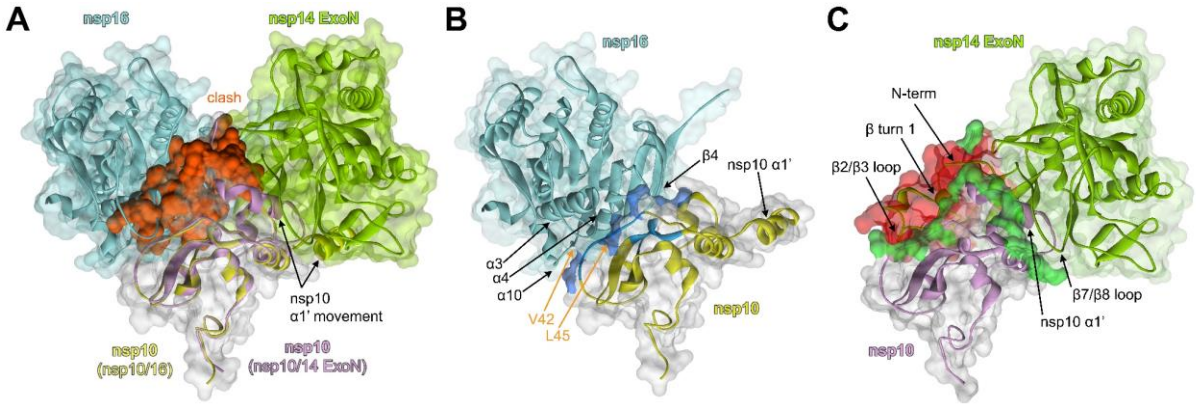

Figure 2. *In silico* modelling of the nsp10/14/16 complex. (1A) The alignment of nsp10/16 (PDB ID: 6WVN) and nsp10/14 ExoN domain (PDB ID: 7DIY). Nsp16 is shown in cyan, and the associated nsp10 is shown in yellow. nsp14 Exon domain is shown in green and the associated nsp10 in purple. The structural clash between the nsp14 ExoN and nsp10/16 surfaces is shown in red. (B) The interface between nsp10 and nsp16. The strong hydrophobic interaction between the nsp10 and nsp16 is shown in blue. The nsp10 residues V42 and L45 are shown as sticks. (C) The interface between nsp10 and nsp14 ExoN. The hydrogen bonds between nsp10 and nsp14 are shown in green. The difference in orientation (~130°) of both nsp10 α1′-helices is evident when interacting with nsp16 or with nsp14. Important structural features involved in the interface are pointed with arrows.

Biochemical confirmation of the triplex formation

Nsp10, nsp14, and nsp16 proteins were co-expressed in *E. coli* and purified. Size Exclusion Chromatography (SEC) was carried out, and the interaction of proteins was evaluated with no prior crosslinking or denaturation. As clearly visible in **Figure 3A**, all three proteins co-

migrate, suggesting strong interaction and heterotriplex formation. To further ensure the content of the complex, it has been analyzed on SDS-PAGE gel, and protein identity was verified using mass spectrometry (**Figure 3B**). Obtained results suggested that the *in silico*-predicted complex is being formed.

The nsp10/14/16 complex formation was also evaluated by Native PAGE in mild conditions at 4°C, to support complex integrity (**Supplementary Figure S1A**). The bands were cut out of the gel, denatured and analyzed by SDS-PAGE (**Supplementary Figure S1B**). Additionally, mass spectrometry analysis of the bands excised from the native gel was carried out to verify the identity of proteins (**Supplementary Table S1**; highlighted in green). Obtained results confirmed the co-migration of all three proteins. Mass spectrometry was also employed to assess the stoichiometry of the complex. By analyzing the signals at 254 and 280 nm, it was established to be 1.2:1:1 (nsp10 to nsp14 to nsp16) (**Supplementary Figure S2** and **Supplementary Table S2**). One may speculate that the used methodology may lead to biased results, as it is impossible to differentiate between the nsp14 and nsp16+nsp10 on a native gel. To verify this hypothesis, nsp14 ExoN domain was co-expressed with nsp110/16. Obtained results show that the co-migration of all three proteins is maintained. The samples were treated with a nuclease to further confirm that the interaction between the proteins is not bridged by co-purified RNA. Again, the co-migration of proteins was not affected (data not shown).

To assess the binding between the proteins in the heterotriplex, MicroScale Thermophoresis (MST) was used. This technique relies on detecting the temperature-related intensity change (TRIC) of the fluorescently labelled target as a function of the concentration of a non-fluorescent ligand. Nsp14 or nsp16 were expressed with a His-tag and labelled with a fluorophore dye specific to the His-tag. As presented in **Figure 3C**, there is an increase in the fraction bound signal upon increasing the concentration of nsp10/16 complex on the nsp14 template that follows Kd fit, what represents the formation of the nsp10/14/16 complex. The binding affinity between nsp14 and nsp10/16 is similar to the one observed between nsp10 and nsp16 (0.28 ± 0.01 and 0.24 ± 0.00 μM, respectively). The binding between nsp10 and nsp14 is approximately an order of magnitude weaker (2.4 ± 0.2 μM). Nsp14 with 50 first N-terminal amino acids ("lid") removed has a similar kD towards nsp10 of 1.5 μM. This shows that “lid” region of nsp14 does not contribute to the binding as predicted from structural analysis. This further confirms the possibility of the triplex formation upon lid rearrangement. nsp16 did not interact with nsp14 alone at tested concentrations, excluding the possibility of nsp14/16 duplex formation.

To further corroborate the interaction between the three proteins, the thermal stability of the complex was investigated using nanoDSF. In this method, the intrinsic fluorescence of aromatic residues is monitored while increasing the temperature to determine the extent of secondary and tertiary structures preserved within the protein at a given temperature. The purity of all analyzed samples was verified with SDS-PAGE prior to the experiment. As depicted in **Figure 3D**, macromolecules are defined by a single peak indicating that the complex partners and domains are connected or have similar melting temperatures. From individually tested proteins, nsp16 is the least stable losing half of its secondary/tertiary structure at 45°C, nsp10 being most stable with a melting temperature of 50°C. Nsp10 and nsp14 form a stable complex with a single peak at 55°C, exhibiting a significant thermal stabilization over nsp14 alone, while nsp10/16 complex retained the melting temperature of 45.5°C. Upon the triplex formation, a single sharp peak at 51.7°C is formed, suggesting that all components of the triplex are structurally related to each other and confirming the nsp10/14/16 complex formation, as there are no other peaks present corresponding to the components of the triplex.

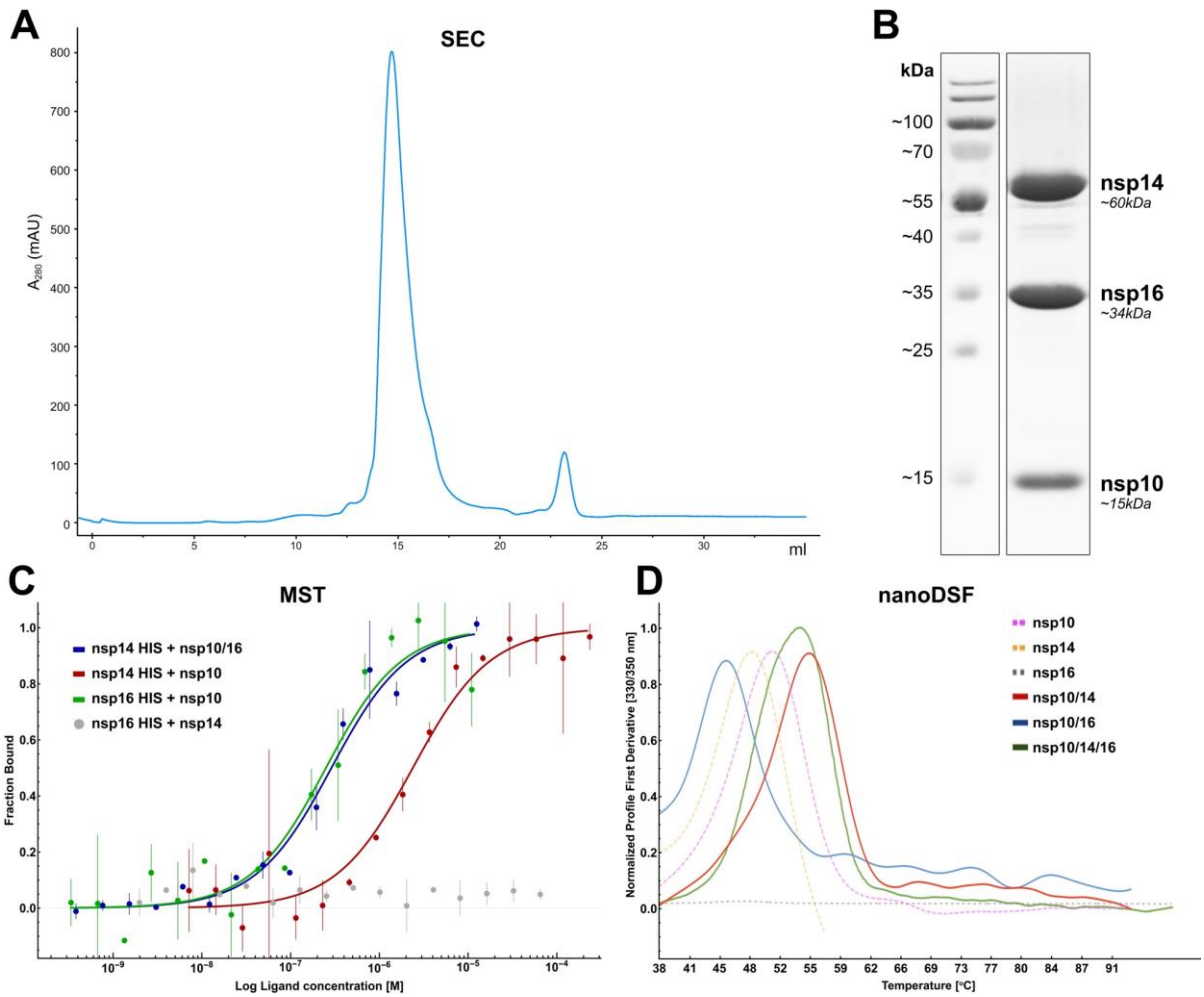

Figure 3. *In vitro* formation of the nsp10/14/16 heterotriplex. (A) SEC chromatogram showing co-migration of nsp14, 16, and 10 proteins in a single peak at ~15 mL. (B) Analysis of contents of the fractions corresponding to the peak with the SDS-PAGE reveals the presence of three proteins in similar quantities that correspond in size to the nsp10, nsp16, and nsp14 (Mw of 15, 33.5 and 60 kDa, respectively). *(C)* MST analysis of the intra-complex affinities. The binding curve for nsp14 His-tag with nsp10/16 complex (blue), nsp14/His-tag with nsp10 (red), nsp16 His-tag with nsp10 (green), and nsp16 His-tag with nsp14 (gray). Experimental data presented in dots with error bars, Kd fits as solid lines. *(D)* The melting profiles for nsp10 (magenta), nsp14 (yellow), nsp16 (grey), nsp10/14 (red), nsp10/16 (blue), nsp10/14/16 (green) determined using nanoDSF.

Functional consequences of nsp10/14/16 complex formation

Eukaryotic viruses modify their RNA to mimic the endogenous nucleic acids to avoid recognition by the innate immune systems. GpppA is first methylated at the N7 position by the nsp14 N7-methyltransferase domain yielding N7MeGpppA[17].

In the assay, N-7 methyltransferase activity is assessed indirectly *via* quantifiable formation of the reaction product SAH using dedicated monoclonal antibodies (**Figure 4A**). The negative controls comprising no substrate (either SAM or RNA) or no protein all yielded maximal signal indicating no production of SAH from SAM. In the assay, nsp14 exhibits no preference for the nascent nucleotide methylating both GpppG and GpppA substrates to the extent similar to

nsp14/10 complex and triplex. Moreover, the activity is hindered upon the addition of pan-methyltransferase inhibitor sinefungin, proving the activity of the enzyme (**Figure 4B**)[18].

Nsp14 also harbours exoribonuclease activity, essential for proofreading during virus replication. However, *in vitro*, the protein acts as a highly processive ribonuclease, unspecifically degrading nucleic acids[19]. As the nuclease activity must be tightly controlled, we assessed whether the complex formation regulates this process. We evaluated the binding and nuclease activity of nsp10/14 compared to the nsp10/16 and the nsp10/14/16 complex. Interestingly, we observed that the nuclease activity of nsp10/14 is reduced for the triplex, suggesting that the nsp14 activity is modulated nsp16, most likely by altering the RNA interaction with the nsp10 co-factor. Obtained data suggest that as the concentration of nsp16 becomes equimolar with the nsp10 and nsp14, the nuclease activity of nsp14 is abrogated (**Figure 4C**).

Both nsp14 and nsp16 rely on nsp10 as an activating co-factor, the degradation activity of nsp14 increasing 35-fold with the presence of nsp10[19-21].

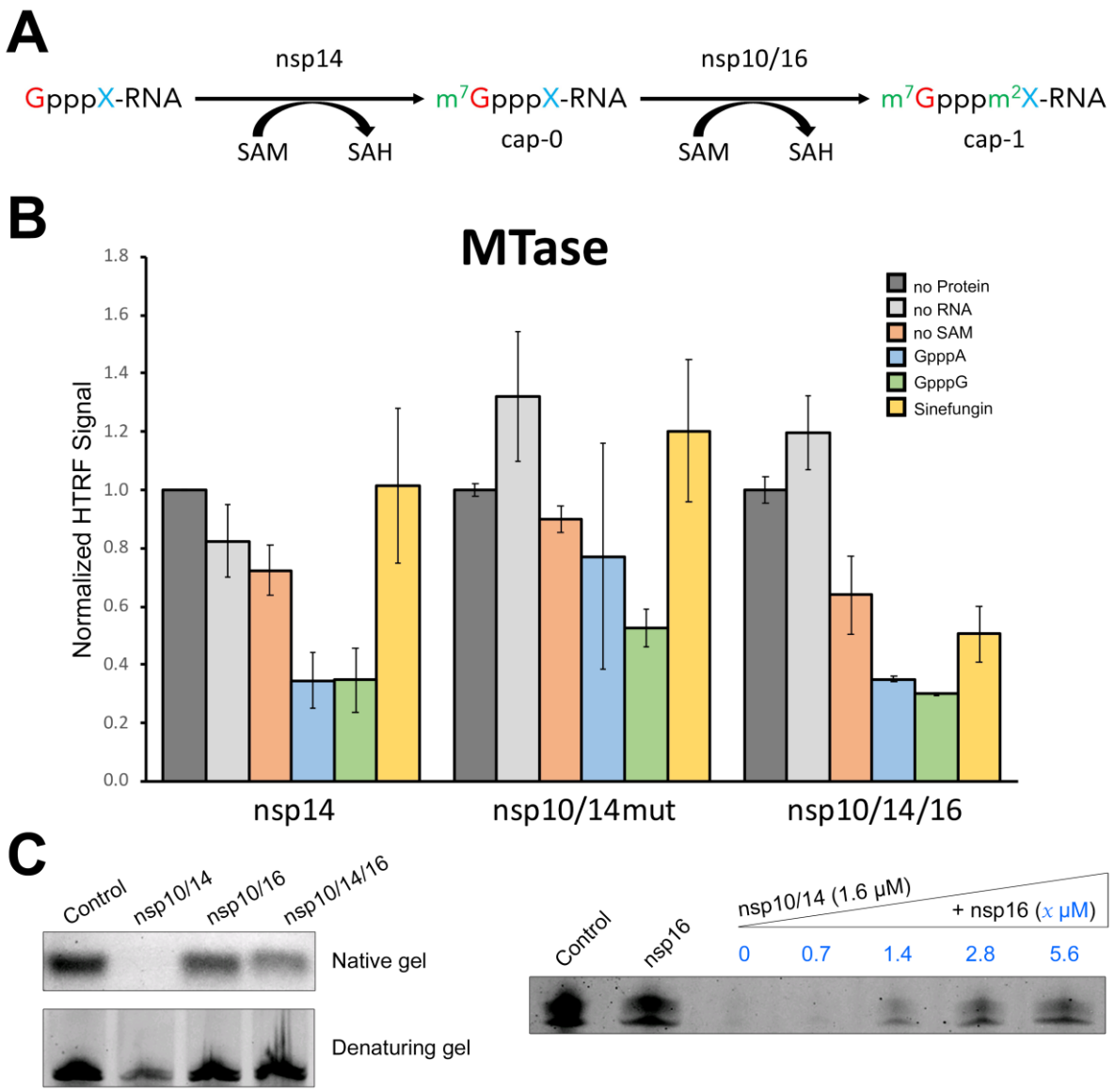

Figure 4. The formation of nsp10/14/16 triplex modulates the ribonuclease but not methyltransferase processivity. (A) The schematic view of mRNA methylation. Blue X represents a nascent nucleotide that could be adenine (A) or guanine (G). First green m^7 represents methylation of the first guanine at position N-7 by nsp14. M^2 is 2′-O methylation of the nascent mRNA nucleotide. **(B)** Results of methylation assay using various substrates: GpppG, GpppA and protein or their respective complexes along with negative controls (no SAM, no protein, no RNA) and positive control: sinefungin. The results were normalized using the SAH calibration curve and denoted µM of methylated product. All experiments were performed in triplicate, and average values with error bars (SD) are shown. **(C, left)** RNA is strongly degraded in vitro by nsp10/14 complex as seen in both native and denaturing RNA gels. This degradation is not seen for nsp16 as it is not a nuclease. Purified triplex is also not degrading the RNA. (right) Titration of nsp10/14 with increased amount of nsp16 causes reduction of exonuclease activity visible on a denaturing RNA gel. The concentration of nsp16 in each sample is indicated in blue. CoV-RNA1-A was used as a substrate.

Structural evaluation

Despite being a low-resolution technique, Small Angle X-ray Scattering (SAXS) is a valuable tool for investigating biomacromolecules under native conditions. Due to the nature of the measurements, SAXS averages over all conformational ensembles for proteins or protein

complexes, accounting for the flexibility of the studied system. Moreover, it allows for a combined Size Exclusion Chromatography-SAXS (SEC-SAXS) measurements to increase understanding of the protein oligomerization or complex formation.

The measurements of nsp14, nsp10/14, nsp10/16 and nsp10/14/16 complexes were performed with a BM29 beamline at the European Synchrotron Radiation Facility (ESRF)[14] in France, in an SEC-SAXS mode[15]. The retention times of tested biomolecules correlate with their molecular weights resulting from their chemical composition and assuming a 1:1(:1) molar ratio[16] from the ATSAS package.

The analysis of the scattering profiles using various tools from the ATSAS package is summarized in **Supplementary Table S3**. According to the Guinier analysis, the radii of gyration of tested complexes (which approximates the size of scatterer) is the biggest for the triplex with 40.3±1.0 Å, which is comparable to both nsp14 alone and in complex with nsp10, 28.0±0.5 and 30.1±1.8 Å, respectively, and nsp10/16 complex being considerably the smallest with 21.0±1.5 Å. This complies with respective molecular weights of nsp14 being a two-domain protein with 60 kDa, nsp10/14 with 75 kDa, nsp10/16 of 48.5 kDa, and a 1:1:1 molecular ratio triplex with the molecular weight of 108.5 kDa. Moreover, this rules out the possibility for a co-migration of nsp10/14 and nsp10/16 complexes. The estimated molecular weights of a scatterer, despite being off, also follow the same pattern where nsp10/16 is the smallest and nsp10/14/16 is the largest. The offset in the molecular weight estimations may arise from the flexibility of tested systems and the use of the same partial specific volume for the molecular mass estimation by used algorithms.

When reconstructed in real space using indirect Fourier transform software GNOM, the profiles present roughly a Gaussian shape, indicating a globular nature of the protein complex with peaks overlapping with radii of gyration generated using Guinier analysis. The profiles of nsp10/14/16 and nsp10/14 also show significant tailing hinting elongation of the complex. The resulting maximal distances within scatterers once more follow the trend of nsp10/16 being the smallest with the longest size of 80.0 Å, nsp14 of 95.8 Å, nsp10/14 of 122.0 Å, and the nsp10/14/16 of 140.0 Å.

The molecular envelopes that account for a low-resolution structure of tested systems in buffer were calculated using ten rounds of DAMMIF software[22] followed by models clustering and averaging and a final round of DAMMIN modelling. The resulting envelopes (**Figure 5**) present the experimental scattering profiles well. A high-resolution structure of nsp10/16, nsp14, nsp10/14, and nsp10/14/16 were fitted in generated envelopes using the SUPCOMB software. Despite an excellent fit to experimental data, the high-resolution models of nsp14

and nsp10/14 (**Figure 5A** and **5B**, respectively) present a poor fit to molecular envelopes suggesting high flexibility of nsp14 alone or in complex with nsp10 in solution or multiple ensembles. The high-resolution structure of nsp10/16 fills the generated envelope tightly, similarly to the generated nsp10/14/16 model. The nsp10/14/16 model was generated by combining the high-resolution structure of nsp10 and nsp16 with nsp14 in PyMol, which was later subjected as a template for the normal mode analysis using the SREFLEX software[23]. SREFLEX uses detected domains within the input, treating them as rigid bodies that could be translated and rotated in space within the constraints of flexible loops of the input model. Domains movements are probed using a random Monte Carlo approach, and the resulting model is evaluated against the experimental data to find a local minimum. The resulting triplex model has a goodness-of-fit to the experimental SAXS data of 1.08, suggesting a likely solution. The high-resolution triplex model and the way it fits in its envelope suggest the assembly of nsp14 with nsp16 via nsp10.

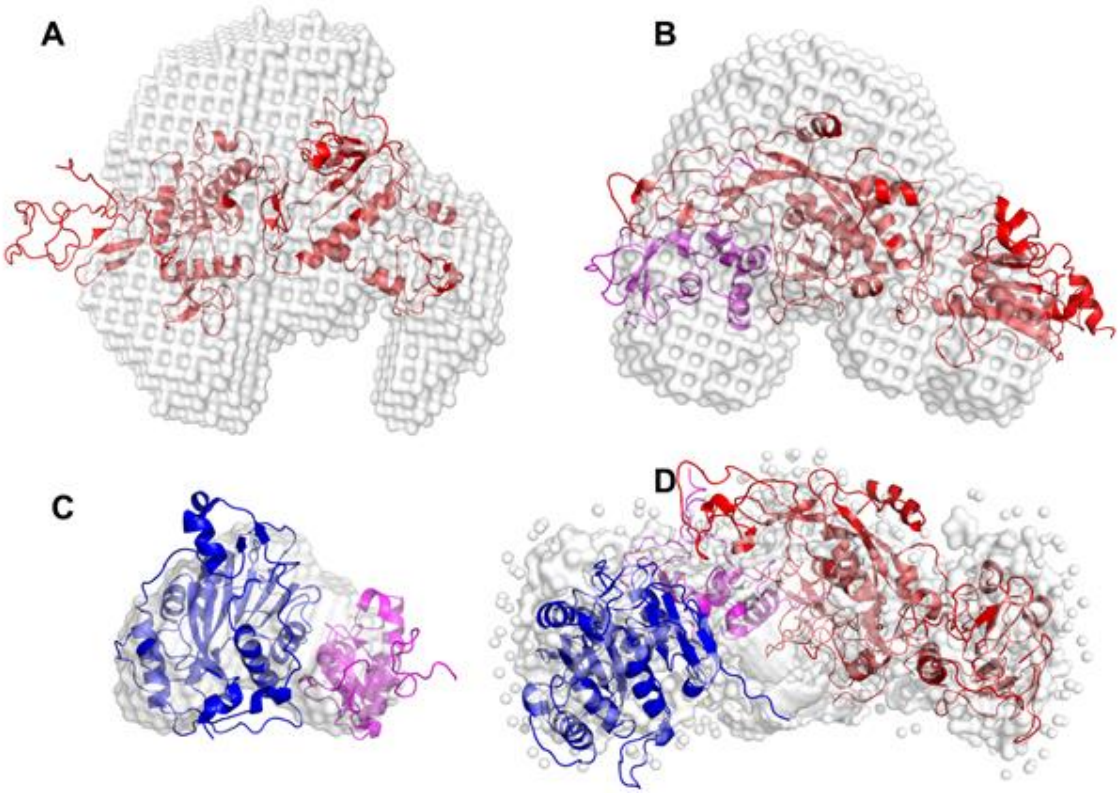

Figure 5. Molecular envelopes of nsp14 (A), nsp10/14 (B), nsp10/16 (C), nsp10/14/16 (D). nsp10 in magenta, nsp14 in red nsp16 in blue, molecular envelopes in grey.

Another classical method to investigate complex formation is negative-stained transmission electron microscopy (NS-TEM). We first monitored the full-length triplex. However, the generated particles yielded non-homogenous classes (Supplement Figure X). Simultaneously,

we tried visualizing the triplex with the truncated methyltransferase domain of nsp14. Here, particles were more homogenous. Obtained 2D classes from negative-stained transmission electron microscopy micrographs show the elongated shape particle of ~10 × 5 nm dimensions from three different viewing directions (**Figure 6A**). The particle possesses a visible inter-band in its centre, indicating the presence of three proteins in the complex. We reconstituted the 3D structure of 20 Å using these class particles. The 3D reconstitution map shows an overall envelope of ~ 10 × 5 × 4.5 nm dimensions with a consensus elongated shape with 2D classes. The rigid-body fitting of the nsp10/14/16 complex structure model into the resulting map results in a reasonable model fitting into the map with correlation coefficient values of 0.59/0.80 (CCmask/CCbox) (**Figure 6B**). The nsp14 and 16 fit well into two globular regions of the map, connected by the electron density corresponding to nsp10. The expected central cavity between nsp14 and nsp16 in the complex is visible in the map indicating an overall matching of the 3D reconstituted envelope and the SAXS-derived model, cross-validating the results.

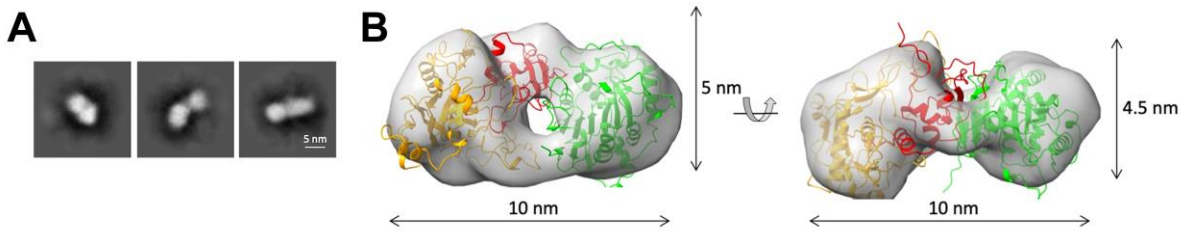

Figure 6. (A) Representative 2D-clasees obtained by template-free 2D classification of particles picked from NS-TEM micrographs of nsp10/14/16 complex. (B) Triplex structure rigid-body fit in to NS-TEM derived 3D reconstitution map from top view (left panel) and side view (right panel). The nsp14, nsp10 and nsp16 are represented as orange, red and green ribbon models, respectively. The 3D reconstitution map is shown as a transparent grey surface.

Discussion

It has been generally acknowledged that nsp16 and nsp14 bind to nsp10 separately. This was a logical and rational conclusion, based on the crystal structure data and the *in silico* modelling. In the available structure of nsp14/nsp10 and nsp16/nsp10 complexes, the overlap between nsp16 and nsp14 structures on the nsp10 interface was obvious. However, the biochemical and biophysical data we present here prove otherwise. It supports the hypothesis depicted in **Figure 1,** presenting nsp10/14/16 ternary complex formation. We re-analyzed the existing data and re-formulated the model. The apparent structural clash between nsp14 and nsp16 on the nsp10 surface can be resolved by rearranging the flexible N-terminal "lid" of the nsp14. This fragment lacks a well-defined secondary structure and has only a few energetically favourable interactions with nsp10. Therefore, the rearrangement of the N-terminal end to

prevent structural hindrances with nsp16 upon the triplex formation is plausible. The remaining parts of the nsp14 have a deep cleft accommodating an nsp10 helix with lipophilic anchoring residues typical for strong protein-protein interactions. Similarly, the nsp16 interface shows significant, well-defined binding elements interacting largely *via* hydrophobic bonds. Interestingly, despite having a similar surface of the interfaces on nsp10, the affinity of nsp14 to nsp10 was ten times lower than observed for nsp16 and nsp10, suggesting that some of the reported interactions in the crystal structure of nsp10/14 ExoN may have been artefacts resulting from the crystal formation. The expendability of the first 50 N-terminal amino acids of nsp14 for the interaction with the nsp10 was further confirmed using the deletion mutants. We show that the affinity of truncated nsp14 without the first 50 N-terminal residues to nsp10 is similar to the full length nsp14/10 complex.

Our hypothesis was further validated using the SEC-SAXS experiment. We probed heterocomplexes of nsp10, nsp14 and nsp16 directly after eluting from the SEC column in native-like conditions. Not only do we see a correct trend in elution volumes from SEC data following the molecular weight of respective complexes, but we can also derive a similar trend from the SAXS scattering profiles calculated *ab initio*, without any model or assumption. These are solely based on the analysis of the features of SAXS scattering profiles, such as radius of gyration or the maximal distance with a scatterer. SAXS-based molecular envelopes of nsp10/16 and nsp10/14/16 are tight and well-defined compared to those of nsp14 or nsp10/14 which present extra volume suggesting their conformational flexibility. Therefore, we postulate that high-resolution X-ray structures capture only one pose of an otherwise flexible ensemble of possible conformations. The molecular envelopes also allow a fit of a model comprising of nsp14/16 without nsp10. However, from the MST data, we know that such configuration does not show any binding even at high concentrations reaching 60 μM. Therefore, only the nsp10/14/16 model could explain this particular shape and volume of the generated envelope of the triplex sample. The extra volume attributed to the nsp10/14 complex seen in SAXS data can also be attributed to the inter-domain motion and flexibility of this two-domain protein. Similar observations regarding the flexibility of the full length nsp14 were made during the TEM or cryoEM efforts. The particles deposited on the grids presented low homogeneity, resulting in fuzzy 2-D classes for TEM and cryoEM (**Supplementary Figures 3** and **4**), hindering further analysis. However, a truncated triplex lacking the nsp14 methyltransferase domain is more stable with particles categorized into well-defined 2D classes on TEM, leading to the reconstruction of a low-resolution 3D model that cross-validates the SAXS-derived model well. Noteworthily, recent CryoEM structures of RdRp complexes where nsp14/nsp10

is present do not show structural hindrance for nsp16 binding besides mentioned N-terminal lid of nsp14.

We speculate that the spatial proximity of both methyltransferases involved in the viral RNA replication and the ExoN domain linked to the proofreading ability of the SARS-CoV-2 genome is well-justified, as the N-7 methylated mRNA product of nsp14 methyltransferase is a substrate for nsp16 O-2 methyltransferase, and the triplex formation does not abrogate the methyltransferase activity of the proteins. Further, the 3'➔5' exoribonuclease activity of nsp10/14 complex is evident and is diminished during the complex formation, most likely due to amended interaction between these two proteins. The modulation of the nsp14 nuclease activity by the nsp16 is beneficial to the virus, as it allows the nsp14 to switch from the proofreading mode to the methylation mode, preventing degradation of the viral nucleic acids.

Materials and Methods

Protein expression and complex purification

Constructs of nsp10 comprising amino acids 4254 – 4392, nsp14 comprising amino acids 5926 – 6452, and nsp16 comprising amino acids 6799 – 7096 of SARS-CoV-2 polyprotein 1ab optimized for expression in *E. coli* were ordered from GeneArt and subcloned into expression vector pETDuet-1. For the nsp14 catalytic mutant, D90 and E92 were replaced with alanines[24]. Plasmids were then co-transformed into *E. coli* strain Bl21.

Transformed BL21 (DE3) *E. coli* cells were grown in Terrific Broth medium (TB; Bioshop) supplemented with 100 µg/ml of ampicillin (Sigma) at 37 °C, overnight and used as a starter culture for the large-scale expression in TB. After the culture reached OD_{600}=1.2-1.4, it was induced with 0.5 mM isopropyl-D-1-thiogalactopyranoside (IPTG; Sigma), and protein expression was carried out at 18 °C for 16 hours. Bacterial pellets were collected by centrifugation at 6,000 g for 10 min at 4 °C, resuspended in lysis buffer (50 mM Tris-HCl pH 8.5, 300 mM NaCl, 5 mM $MgCl_2$, 5% glycerol, 5 mM β-mercaptoethanol) and disrupted by sonication at 80% amplitude for 15 min at 10°C (3s, 3s pulse). The lysed sample was clarified by centrifuging for 1 h at 22,000 rpm at 4°C. The supernatant was collected and incubated with 2 ml of Ni-NTA Agarose (Jena Bioscience) pre-equilibrated with the lysis buffer for 2 hours at 4°C. Purification was carried out in a gravity-flow column. Resin was washed with 50BV (bed volume) of buffer A (50 mM Tris-HCl pH 8.5, 300 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 10 mM imidazole) and 20 BV of buffer B (50 mM Tris-HCl pH 8.5, 300 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 20 mM imidazole). Protein complex was eluted with buffers C, D, E (50 mM Tris-HCl pH 8.5, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 100 mM (C) / 250 mM (D) / 350 mM imidazole (E), 5 × 2BV each). The eluted fractions were concentrated to 5 ml and loaded onto a HiLoad 26/600 Superdex 200 prep grade (GE Healthcare) equilibrated with the SEC buffer (50 mM Tris-HCl pH 8.5, 150 mM NaCl, 5 mM $MgCl_2$, 2 mM β-mercaptoethanol). Confirmed by SDS-PAGE, the protein complex fractions were collected. TEV protease was added in a combination of a final concentration of 10 mM β-mercaptoethanol and incubated with gentle rocking at 4°C for approximately 12h. When the His-tag had been successfully cleaved from the protein complex, the sample was concentrated to 5 ml and loaded over the HiLoad 26/600 Superdex 200 prep grade (GE Healthcare). In order to clear the impurities which have been eluted with the complex, a reverse binding was performed by incubating 15 minutes, rolling at 4°C with 100 µl of Ni-NTA Agarose (GE Healthcare) pre-equilibrated with the SEC buffer. The sample was separated with a gravity-flow column, and the Ni-NTA agarose was washed three times with

the SEC buffer with 150 mM NaCl and with the last wash with SEC buffer supplemented with another 150 mM NaCl (totalling 300 mM NaCl). The flow-through samples contained the clean protein complex, as verified by SDS-PAGE. The protein complex was collected, concentrated, and loaded onto the Superdex 200 Increase 10/300 GL (GE Healthcare) equilibrated with the final SEC buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 2 mM β-mercaptoethanol). The protein complex was collected and concentrated up to 5 mg/ml for further applications such as crystallization, inhibitor, and activity screening. After every stage of purification, the protein content and purity were evaluated with SDS-PAGE and InstantBlue (Sigma) gel staining.

Protein identification from gel bands - LC-MS/MS analysis

Samples were prepared, measured and analyzed as described in Pabis *et al.*[25] with minor changes. Briefly, gel bands were destained by alternating washing with 25% and 50% acetonitrile (ACN) in 25 mM ammonium bicarbonate (ABC). Then, protein reduction was performed with 50 mM DTT in 25mM ABC (45 min of incubation at 37°C) followed by alkylation with 55 mM iodoacetamide (1 h of incubation at room temperature in the dark). In the next steps, gel bands were washed with 50% ACN in 25mM ABC, dehydrated in 100% ACN, dried and rehydrated in 20 µl of trypsin solution (10 ng/ µl in 25 mM ABC). After rehydration, 40 µl of 25 mM ABC was added and samples were left for overnight incubation at 37°C. Protein digestion was stopped by adding trifluoroacetic acid (TFA) to the concentration of about 0.5%. Peptides present in the solution were collected and additionally extracted from the gel by dehydration with 100% ACN. The obtained peptide mixtures were dried and suspended in a loading buffer (2% ACN with 0.05% TFA) for LC-MS/MS analysis carried out with nanoHPLC (UltiMate 3000 RSLCnano System, Thermo Fisher Scientific) coupled to Q Exactive mass spectrometer (Thermo Fisher Scientific). Peptides were loaded onto a trap column (Acclaim PepMap 100 C18, 75 µm × 20 mm, 3 µm particle, 100 Å pore size, Thermo Fisher Scientific) at a flow rate of 5 µl/min and separated on analytical column (Acclaim PepMap RSLC C18, 75 µm × 500 mm, 2 µm particle, 100 Å pore size, Thermo Fisher Scientific) at 50°C with a 60-min gradient of ACN, from 2% to 40%, in the presence of 0.05% formic acid at a flow rate of 250 nl/min. The eluting peptides were ionized in Digital PicoView 550 nanospray source (New Objective) and measured with Q Exactive operated in a data-dependent mode. Top8 method was used with 35 s of dynamic exclusion. MS and MS/MS spectra were acquired with a resolution of 70 000 and 35 000, respectively. The ion accumulation times have been adjusted to ensure parallel filling and detection. The acquired

LC-MS/MS data were processed with the use of Proteome Discoverer platform (v.1.4; Thermo Scientific) and searched with an in-house MASCOT server (v.2.5.1; Matrix Science, London, UK) against the database of common protein contaminants (cRAP database) with manually added sequences of the proteins of interest. The following parameters were applied for the database search: enzyme – trypsin; missed cleavages – up to 1; fixed modifications – carbamidomethyl (C); variable modifications – oxidation (M); peptide mass tolerance: 10 ppm; fragment mass tolerance: 20 mmu. Additionally, the SwissProt database restricted to *E. coli* taxonomy was searched to assess contamination with host proteins.

Stoichiometry determination

For protein quantitation, sample separation was carried out following a simple protocol using the Prominence HPLC system (two LC-20AD pumps, SPD-M20A diode array detector, DGU-20 degasser, all from Shimadzu Corp., Kyoto, Japan). For gradient separation Kinetex 2.6 µm/100A C18 100 mm/2.1 mm ID HPLC column was used (00D-4462AN, Phenomenex, Torrance, CA, USA). Solvents used for separation: A = water + formic acid (99.9:0.1, v/v), B = acetonitrile + formic acid (99.9:0.1, v/v). All solvents were supplied by Merck local distributor (Merck KgaA, Darmstadt, Germany). Gradient was set as follows: t (time) = 0 min, 25% B; t=20 min, 75% B; t=20.5 min 90% B; t=25 min, 90%B; t=25.5 min, 25% B; t=35 min 25% B (end). Flow rate was set to 0.3 ml/min. Diode array detector settings: wavelength range: 200-350 nm (deuterium lamp only), sampling frequency 5 Hz. Data acquisition and data processing were controlled by LCsolution software, ver. 1.25 (Shimadzu Corp., Kyoto, Japan). To confirm protein content under every chromatographic peak taken for protein quantitation, mass spectrometry-based identification was used. Our protocols for protein identification, applied with minor changes, are available elsewhere[26]. Briefly, fractions acquired during protein separation were freeze-dried using CentriVap system (Labconco, Kansas City, MO, USA) and redissolved in 70 µl 50 mM ammonium bicarbonate solution (pH=7.8). Next, reduction and alkylation of cysteine residues were done using DTT and following iodoacetamide (both: 5 mM final concentrations). In both cases 10 min incubation in 80°C with shaking was applied. After cooling down, trypsin (Gold-MS grade, Promega, Madison, WI, USA) was added in a final concentration of 2 pmol per sample. Samples were incubated overnight in 37°C with shaking, then freeze-dried again and redissolved in 30 µl of 4% acetonitrile/water solution acidified by 0.1% formic acid (v/v/v). NanoLC-MS/MS analyses were performed on Ultimate 3000 system (Thermo, Waltham, MA, USA) connected on-line to AmaZon SL equipped with nanoFlowESI ion source (Bruker-Daltonics, Bremen, Germany).

Basic parameters of nanoLC system: column used Acclaim PepMap100, C18, l=10cm/75 µm I.D., precolumn PepMap100, C18, l=1 cm/1 mm I.D., gradient settings: solvent A water with 0.1% formic acid (v/v), solvent B acetonitrile with 0.1% formic acid (purity: MS-grade, both from Merck KGaA, Darmstadt, Germany), t (time) = 0 min 6% B; t = 5 min 6% B; t = 50 min 55% B; t = 50.1 min 80% B; t = 53 min 80% B; t = 54 min 6% B, t = 58 min 6% B (end). Sample injection volume was usually in a range 3-5 µl. Flow rate: 300 nl/min, flow rate for sample introduction on precolumn: 30 µl/min. Mass spectrometer settings were as follows: capillary voltage 4200V, heated capillary temperature 150°C, MS scan range: 375-1600 m/z, MS/MS scan range: 200-2000 m/z, resolution: enhanced, scanning frequency: ca. 1 Hz, ICC (Ion Charge Control) 250 000 ions/trap cycle, fragmentation ions selection range: 450-1600 m/z (with minor exclusions), intensity: at least 1 x 106 units. Both instruments were run under the HyStar ver. 4.1 SR1 (Bruker-Daltonics, Bremen, Germany). Data analysis was performed in Bruker's Compass DataAnalysis 4.4 SR1. Acquired data were converted into Mascot Generic Format (*.mgf) files using built-in scripts and introduced into Mascot search engine (Matrixscience, London, UK), based on SwissProt and in-house created database. Mascot settings: enzyme: trypsin, missing cleavages: 1, taxonomy: all, fixed modifications: carbamidomethylation, variable modifications: oxidation-Met, peptide tolerance: 1.2 Da, #13C: 1, MS/MS tolerance: 0.6 Da, peptide charge: +1,+2,+3, instrument: ion-trap.

MicroScale Thermophoresis

His-tag proteins were labelled with Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation according to manufacturer's guidelines. Labelled proteins were diluted in PBS containing 0.05% Tween-20 up to 80 nM concentration and mixed with tested ligands. Samples were allowed for 30 min incubation at RT prior measurements. Measurements were performed on Monolith NT.115 in duplicates using Excitation Power: 80% and MST Power: high in Monolith NT.115 Capillaries.

NanoDSF

NanoDSF was performed in standard capillaries using Tycho equipment. Protein and their respective complexes were measured at 1 mg/mL using default ramp temperature. The resulting meting temperatures were reported as the first derivative of the fluorescence ratio.

Methyltransferase activity

The methyltransferase activity of wild type triplex complex nsp10/nsp14/nsp16 or triplex complex with mutated nsp14 protein (ExoN mutant and catalytic mutant – please add the info regarding the mutations) was measured using the EPIgeneous Methyltransferase kit from Cisbio as previously described[27]. Individual kit reagents were reconstituted according to the manufacturer's instructions. Briefly, the methyltransferase reaction was incubated for 20 minutes at room temperature in 8 µl reaction volume with 100 nM nsp10/nsp14/nsp16 wild type or mutated triplex complex, 1 µM Ultrapure SAM (Cisbio), 187.5 µM RNA cap analogue (GpppA or GpppG, New England Biolabs) or 18.75 µM Cap 0 RNA oligo (TriLink) in reaction buffer consisting of 20 mM Tris-HCl pH 7.4, 150 mM NaCl, and 0.5 mM DTT. The reaction was quenched by the addition of 2 µl of 5M followed by the addition of 2 µl Detection Buffer 1 (Cisbio) to the reaction mixture. After 10 min, 4 µl of 16× SAH-d2 conjugate solution (Cisbio) was added. After 5 min, 4 µl of 1× α-SAH Tb Cryptate antibody solution was added to the reaction mixture and incubated for 1 hour at room temperature. Homogenous Time-Resolved Fluorescence (HTRF) measurements were performed on a SpectraMax iD5 plate reader (Molecular Devices) according to the manufacturer's guideline (excitation wavelength 340 nm, emission wavelengths 665 and 620 nm, top mode, 100 flashes, optimal gain, z position calculated from the negative control (no enzyme), lag time of 60 µs and the integration time of 500 µs). The resulting data were background-subtracted and normalized as follow. The ratio of 665 to 620 nm wavelength was calculated. The data was background corrected on the averaged signal for the buffer control. Next, the data was normalized for each series individually on the wells not containing the enzyme.

Exonuclease activity

The CoV-RNA1-A RNA was ordered from IDT as a PAGE-purified and desalted oligo. Protein complexes were incubated with 100 ng of RNA in a buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 5% v/v glycerol, 10 mM $MgCl_2$, and 5 mM β-mercaptoethanol, followed by analysis via a 20% urea-denaturing or a native 1% TBE agarose gel.

SAXS

Samples were measured in SEC-SAXS mode at BM29/ESRF, Grenoble France on 12.12.2020. (session ID MX2341). Samples (100 µL) were measured on Agilent AdvanceBio SEC 300

with 50mM Tris-HCl pH 8.5, 150mM NaCl, 5mM $MgCl_2$, and 2 mM β-mercaptoethanol running phase at 0.16 mL/min flowrate. The measurements were performed at 0.99 Å wavelength. The sample to detector distance was set at 2.83 m with Pilatus2M detector for data acquisition.

CryoEM

Out of 100 randomly chosen micrographs, 1000 particles were manually picked without any structural knowledge about the complex to minimize the bias and assigned with 2D classes that were used in the *ab initio* model built. Out of generated 3D classes, one was manually picked and used for the training of TOPAZ neural networks, which in turned picked next interaction of particles that were used to retrain Topaz. With this approach approximately 0.5 M particles were selected to generate 50 2D classes, out of which 19 were manually picked.

Crosslinking and Negative-Staining Transmission Electron Microscopy

Following SEC in HEPES buffer, the purified protein complex was crosslinked with bis(sulfosuccinimidyl)suberate (BS^3, Thermo Scientific), an amine-to-amine crosslinker. The protein sample was incubated with 0.5 mM BS^3 (from 50 mM stock) for 30 min at RT. The crosslinking reaction was quenched with 1M Tris pH 7.5 to a final 50 mM Tris, and incubated 15 min at RT. The excess of BS^3 crosslinker was cleared by sequential dilution and concentration with Tris buffer. Concentrated samples were kept at 4°C or stored at -80°C for further measurements. Negative-stain transmission electron microscopy (NS-TEM) measurements was done in formvar/carbon supported 400 mesh copper grids, suspended in air with a negative lock tweezer. The purified protein-complex (0.03 mg/mL) was applied on glow-discharged Formvar/Carbon supported 400 mesh copper grids, and negatively stained with 1% neutralized uranyl-acetate. Grids were imaged using the JEOL JEM 2100HT electron microscope (Jeol Ltd, Tokyo, Japan) was used at accelerating voltage 200 kV. Images were taken by using 4kx4k camera (TVIPS) equipped with EMMENU software ver. 4.0.9.87.

Image processing and 3-D reconstitution

Collected micrographs were processed using CryoSPARC 3.1.1. Initially, 9'350 particles were picked from micrographs using Blob Picker. Picked particles were subjected to a template-free 2D classification, from which 1'216 particles were selected and subjected for 3D reconstitution using Ab-initio reconstitution job. The nsp10/14/16 complex map derived from SAXS data was used for a rigid-body fit in such 3D-reconstituion map using Dock in map.

Acknowledgements

This work was supported by a subsidy from the Polish Ministry of Science and Higher Education for research on SARS-CoV-2 and a grant from the National Science Center (UMO-2017/27/B/NZ6/02488) and by EU-Horizon2020 ITN OrganoVir grant 812673 (to K.P.), and by the Bayerische Forschungsstiftung grant AZ-1453-20C to G.P and M.S. and a DFG Grant PO 1851/4-1 "Biochemische und strukturelle Charakterisierung des SARS-CoV-2 non-structural protein 16 (Nsp16), eine cap ribose 2'O-methyltransferase." to G.P. and M.S. The mass spectrometry was also partially financed from the subsidy no 16.16.160.557 of the Polish Ministry of Science and Education to P.S. and K.H.

Declaration of interests

None.

Supplementary data

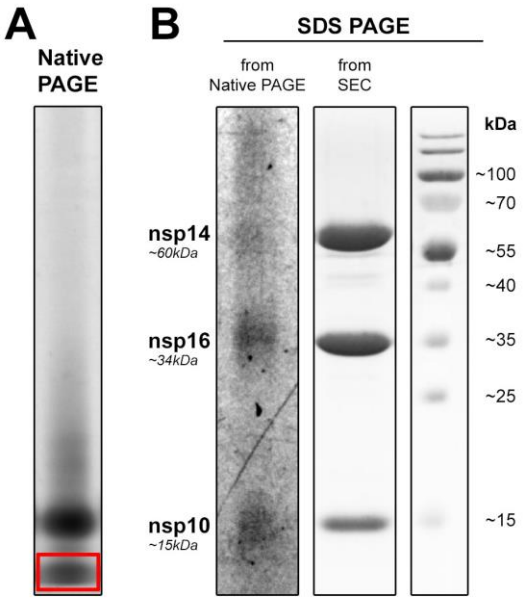

Figure S1. A: Native-PAGE of triplex; B: SDS-PAGE of the band excited from the native-PAGE (red box showing disintegration of the triplex into nsp14, nsp16, nsp10.

Table S1. The qualitative MS analysis of the Native –PAGE gel band indicated in red box.

Description	Score	Coverage
>sp|Nsp14tag_PROTEIN|	10’464.37	85.61
>sp|Nsp14_PROTEIN|	10’156.82	85.28
>sp|Nsp16tag_PROTEIN	9’601.49	86.83
>sp|Nsp10tag_PROTEIN|	7’572.82	100.00
>sp|Nsp10_PROTEIN|	7’408.26	100.00
>sp|K1C10_HUMAN|	3'205.43	56.83
>sp|K2C1_HUMAN|	2'357.77	47.90
>sp|K1C9_HUMAN|	2'156.15	48.96
>sp|TRYP_PIG|	1’814.17	35.50
>sp|K22E_HUMAN|	1’653.50	43.26
>sp|3264P0DTD13569_PROTEIN|	398.47	22.55
>sp|K1C15_SHEEP|	215.66	9.27
>sp|CAS1_BOVIN|	169.83	22.43
>sp|ALBU_HUMAN|	155.74	8.05
>sp|ALBU_BOVIN|	146.29	8.40
>sp|CASB_BOVIN|	59.51	8.48
>sp|MINDY3mod_HUMAN|	21.06	1.34

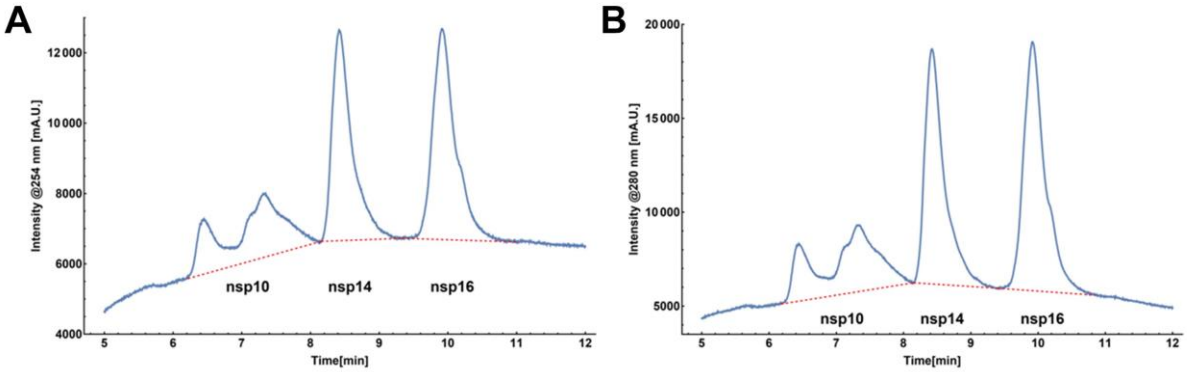

Figure S2. Mass spectroscopy of the triplex at 254 and 280 nm.

Table S2. Summary of the peak volumes from Figure S2 presenting 1.2:1:1 ratio of nsp10/14/16.

	Peak 1 (fr 6.3-8.0)	**Peak 2** (fr. 8.2-9.3)	**Peak 3** (fr. 9.6-10.7)	**Average surface area**	**Total**	**Percent**	**Ratio**
nsp10	64.5% (100)	22% (19.6)	13.5% (10.9)	145’756	205’562	38	**1.2**
nsp14	0% (0.0)	90% (80.4)	10% (8.2)	187’293	167’747	31	**1.0**
nsp16	0% (0.0)	0% (0.0)	100% (80.9)	211’899	171’638	31	**1.0**

Table S3 The summary of the structural parameters derived from scattering profiles.

a. Sample details				
	nsp14	nsp10/14	nsp10/16	nsp10/14/16
Source organism	*Severe acute respiratory syndrome coronavirus 2*			
Source	*E. coli BL21*			
UniProt sequence ID (residues in construct)	P0DTD1 (5926-6452)	P0DTD1 (5926-6452)	-	P0DTD1 (5926-6452)
	-	P0DTC1 (4254-4392)	P0DTC1 (4254-4392)	P0DTC1 (4254-4392)
	-	-	P0DTD1 (6799-7096)	P0DTD1 (6799-7096)
Molecular weight from chemical composition (kDa)	60	75	48.5	108.5
SEC–SAXS column	AdvanceBio Bio SEC 300 column			
Loading concentration (mg/mL)	4	3.5	2.6	2.8
Injection volume (µL)	100			
Flow rate (mg/mL)	0.16			
Running phase composition	50mM Tris, 150mM NaCl, 5mM $MgCl_2$, βME 2mM pH8.5			
b. SAXS data-collection parameters				
Instrument	BM29, ESRF, Grenoble France			
Detector	Pilatus2M			
Wavelength	0.99 Å			
Sample to detector distance	2.83 m			
c. Structural parameters				
	nsp14	nsp10/14	nsp10/16	nsp10/14/16
Guinier analysis				
I(0) (cm^{-1})	8.43±0.13	3.22±0.13	1.73±0.09	5.26±0.11
R_g(Å)	28.0±0.5	30.1±1.8	21.0±1.5	40.3±1.0
qR_g($Å^{-1}$)	0.59-1.51	0.59-1.32	0.52-1.39	0.76 – 1.4
P(r) analysis				
R_g(Å)/ I(0) (cm^{-1})	29.4/8.61	32.5/3.28	21.1/1.64	41.2/5.23
Guinier R_g(Å)/ I(0) (cm^{-1})	29.4/8.61	32.7/3.28	21.2/1.64	41.5/5.23
r_{max}(Å)	95.8	122.0	80.0	140.0
Total quality estimate	0.94	0.78	0.75	0.77
Molecular weight estimate/predicted (kDa)	33.1/60	28.9/75	20.6/48.5	83.2/108.5
Oligomerization state	monomeric			
d. Shape model-fitting results				
DAMMIF (10 runs)	nsp14	nsp10/14	nsp10/16	nsp10/14/16

q_{max} range for fitting(Å^{-1})	0.26	0.26	0.21	0.17
Symmetry, anisotropy assumptions	P1, none			
NSD (standard deviation)	1.42(0.07)	1.10(0.07)	1.29(0.14)	0.80(0.08)
Chi-squared	1.16	1.09	1.07	1.08
Resolution (from SASRES) (Å)	43±3	39.3±3	33.3±3	38.3±3
SASDBD IDs	SASDKT6	SASDKU6	SASDKV6	SASDKW6

Transmission electron microscope analysis of the triplex

To assess the structure of nsp10/14/16 triplex we performed the Transmission electron microscope (TEM) imaging. For this purpose, we crosslinked the triplex using amine reactive suberic acid bis sodium salt as previous attempts with the native complex yielded non-homogenous samples. Crosslinking strengthened the conjugation so that it better withstood the deposition on copper grids. Samples were visualized using JEOL JEM 2100HT electron microscope at 80 kV accelerating voltage. Data was analyzed using cryoSparc software as described in the methodology section. Generated 2D classes clearly show lack of homogeneity of the sample on the grids.

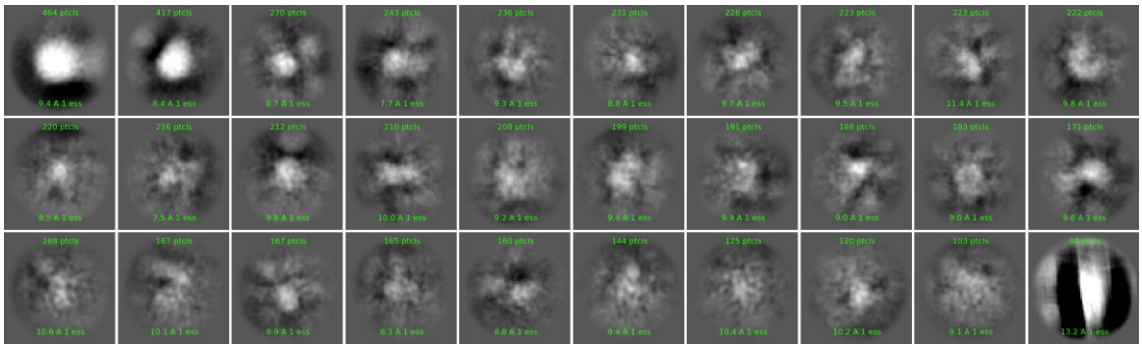

Figure S3. 2D classes picked in cryoSparc for fulllength triplex.

CryoEM

We attempted to solve a high-resolution triplex structure using cryoEM technique. Highly purified triplex solution was vitrified on grids using Vitrobot under various conditions. Resulting grids were measured using Titan Krios G3i at Solaris, Poland. Data analysis was performed using cryoSPARC software. We managed to train the TOPAZ neural networks, to pick approximately 0.5 M particles used to generate 50 2D classes, out of which 20 were manually picked (Figure 6). Generated classes, though noisy, exhibit croissant-like shape with four distinct domains that correspond to nsp16, nsp10 and two domains of nsp14 (highlighted with arrows on the class that shows side-on view). Unfortunately, we were not able to reconstruct a high-resolution 3D structure from collected data. Our best attempt shown below in cyan has ca. 9 Å resolution. The fitted triplex hybrid model based on SAXS data (gold) shows overall good fit with the extra volume that may arise from the flexible nature of the triplex.

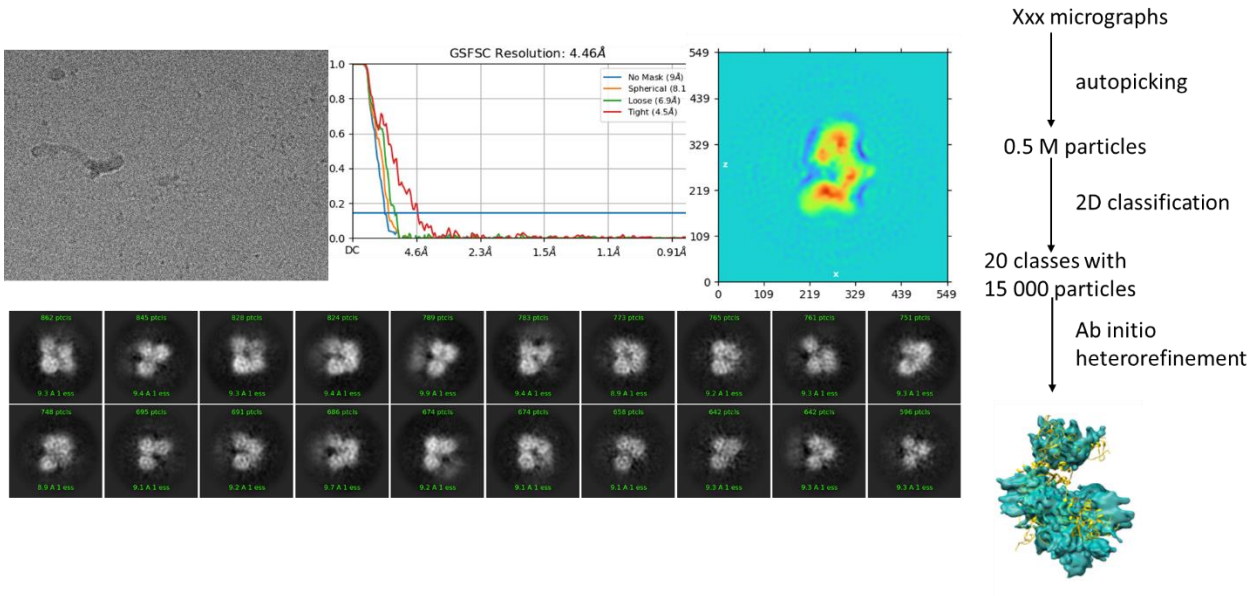

Figure S4.: An overview of cryoEM data analysis in cryoSPARC. The overlay of generated 3D model from cryoEM (cyan) and SAXS model (gold).

REFERENCES

1 Pyrc, K., Jebbink, M. F., Berkhout, B. & van der Hoek, L. Genome structure and transcriptional regulation of human coronavirus NL63. *Virol J* **1**, 7, doi:10.1186/1743-422X-1-7 (2004).

2 Gao, Y. *et al.* Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779-782, doi:10.1126/science.abb7498 (2020).

3 Yan, L. *et al.* Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell* **184**, 184-193 e110, doi:10.1016/j.cell.2020.11.016 (2021).

4 Wang, Q. *et al.* Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell* **182**, 417-428 e413, doi:10.1016/j.cell.2020.05.034 (2020).

5 Kabinger, F. *et al.* Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat Struct Mol Biol* **28**, 740-746, doi:10.1038/s41594-021-00651-0 (2021).

6 Yan, L. *et al.* Coupling of N7-methyltransferase and 3'-5' exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell* **184**, 3474-3485 e3411, doi:10.1016/j.cell.2021.05.033 (2021).

7 Sevajol, M., Subissi, L., Decroly, E., Canard, B. & Imbert, I. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. *Virus Res* **194**, 90-99, doi:10.1016/j.virusres.2014.10.008 (2014).

8 Bouvet, M. *et al.* RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc Natl Acad Sci U S A* **109**, 9372-9377, doi:10.1073/pnas.1201130109 (2012).

9 Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L. & Denison, M. R. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol* **81**, 12135-12144, doi:10.1128/jvi.01296-07 (2007).

10 Daffis, S. *et al.* 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **468**, 452-456, doi:10.1038/nature09489 (2010).

11 Ivanov, K. A. *et al.* Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol* **78**, 5619-5632, doi:10.1128/JVI.78.11.5619-5632.2004 (2004).

12 Ivanov, K. A. & Ziebuhr, J. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. *J Virol* **78**, 7833-7838, doi:10.1128/JVI.78.14.7833-7838.2004 (2004).

13 Decroly, E. *et al.* Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J Virol* **82**, 8071-8084, doi:10.1128/JVI.00407-08 (2008).

14 Chen, Y. *et al.* Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc Natl Acad Sci U S A* **106**, 3484-3489, doi:10.1073/pnas.0808790106 (2009).

15 Bouvet, M. *et al.* In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog* **6**, e1000863, doi:10.1371/journal.ppat.1000863 (2010).

16 Slanina, H. *et al.* Coronavirus replication-transcription complex: Vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2022310118 (2021).

17 Schlee, M. & Hartmann, G. Discriminating self from non-self in nucleic acid sensing. *Nature Reviews Immunology* **16**, 566-580, doi:10.1038/nri.2016.78 (2016).

18 Ahmed-Belkacem, R. *et al.* Synthesis of adenine dinucleosides SAM analogs as specific inhibitors of SARS-CoV nsp14 RNA cap guanine-N7-methyltransferase. *Eur J Med Chem* **201**, 112557, doi:10.1016/j.ejmech.2020.112557 (2020).

19 Riccio, A. A., Sullivan, E. D. & Copeland, W. C. Activation of the SARS-CoV-2 NSP14 3'-5' exoribonuclease by NSP10 and response to antiviral inhibitors. *J Biol Chem* **298**, 101518, doi:10.1016/j.jbc.2021.101518 (2021).

20 Krafcikova, P., Silhan, J., Nencka, R. & Boura, E. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat Commun* **11**, 3717, doi:10.1038/s41467-020-17495-9 (2020).

21 Bouvet, M. *et al.* Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J Biol Chem* **289**, 25783-25796, doi:10.1074/jbc.M114.577353 (2014).

22 Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **42**, 342-346, doi:10.1107/S0021889809000338 (2009).

23 Panjkovich, A. & Svergun, D. I. Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Phys Chem Chem Phys* **18**, 5707-5719, doi:10.1039/c5cp04540a (2016).

24 Ma, Y. *et al.* Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc Natl Acad Sci U S A* **112**, 9436-9441, doi:10.1073/pnas.1508686112 (2015).

25 Pabis, M. *et al.* Molecular basis for the bifunctional Uba4-Urm1 sulfur-relay system in tRNA thiolation and ubiquitin-like conjugation. *EMBO J* **39**, e105087, doi:10.15252/embj.2020105087 (2020).

26 Bodzon-Kulakowska, A. *et al.* Proteomic analysis of striatal neuronal cell cultures after morphine administration. *J Sep Sci* **32**, 1200-1210, doi:10.1002/jssc.200800464 (2009).

27 Basu, S. *et al.* Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of Nsp14 RNA cap methyltransferase. *Biochem J* **478**, 2481-2497, doi:10.1042/bcj20210219 (2021).