Refolding of lid subdomain of SARS-CoV-2 nsp14 upon nsp10 interaction releases exonuclease activity.

Anna Czarna^{1,*}, Jacek Plewka^{1,*}, Leanid Kresik¹, Alex Matsuda¹, Abdulkarim Karim², Colin Robinson³, Sean O'Byrne², Fraser Cunningham², Irene Georgiou², Magdalena Pachota¹, Grzegorz Popowicz^{5,6}, Paul Graham Wyatt^{2,±}, Grzegorz Dubin^{4,±}, Krzysztof Pyrc^{1,±}

1. Virogenetics Laboratory of Virology, Malopolska Centre of Biotechnology, Jagiellonian University, Gronostajowa 7a, 30-387 Krakow, Poland.
2. Department of Biology, College of Science, Salahaddin University-Erbil, Kirkuk Road, 44002 Erbil, Kurdistan Region, Iraq and Department of Community Health, College of Health Technology, Cihan University-Erbil, 100 street, 44001 Erbil, Kurdistan Region, Iraq.
3. Drug Discovery Unit, Wellcome Centre for Anti-Infectives Research, School of Life Sciences, University of Dundee, Dow Street, Dundee DDI 5EH, U.K.
4. Protein Crystallography Research Group, Malopolska Centre of Biotechnology, Jagiellonian University, Gronostajowa 7a, 30-387 Krakow, Poland.
5. Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany
6. Bavarian NMR Center, Department of Chemistry, Technical University of Munich, Lichtenbergstrasse 4, 85748 Garching, Germany

* These authors contributed equally

± to whom correspondence should be addressed:

K.P.: k.a.pyrc@uj.edu.pl or
 G.D. grzegorz.dubin@uj.edu.pl or
 P.G.W.: P.G.Wyatt@dundee.ac.uk

30 **Abstract**

During the RNA replication, coronaviruses require proofreading to maintain the integrity of their large genomes. Nsp14 associates with viral polymerase complex to excise the mismatched nucleotides. Aside from the exonuclease activity, nsp14 methyltransferase domain mediates cap methylation, facilitating translation initiation and protecting viral RNA from recognition by the innate immune sensors. The nsp14 exonuclease activity is modulated by a protein co-factor nsp10. While the nsp10/nsp14 complex structure is available, the mechanistic basis for nsp10 mediated modulation remains unclear in the absence of nsp14 structure. Here we provide a crystal structure of nsp14 in an apo-form. Comparative analysis of the apo- and nsp10 bound structures explain the modulatory role of the co-factor protein.

40 Further, the structure presented in this study rationalizes the recently proposed idea of nsp14/nsp10/nsp16 ternary complex.

INTRODUCTION

Upon release into the cytoplasm of the host cell, the coronavirus genome is translated into a single non-functional polyprotein 1a/1ab. Its autoproteolytic processing releases a number of nonstructural proteins (nsp) responsible for viral replication. The activity of particular nsp's is relatively well-characterized, but the mechanistic understanding of their interplay is still insufficient¹⁻³.

Nsp14 consists of two domains, each with a distinct catalytic activity. The C-terminal domain is a SAM-dependent methyltransferase, while the N-terminal ExoN domain exhibits 3'-5' exonuclease activity. Recognition of unmodified RNA molecules constitutes a universal antiviral strategy embedded in the host innate immune defences. To avoid detection, coronaviral RNA is capped by viral enzymes, what additionally increases its stability and facilitates translation initiation. Capping involves a series of enzymatic reactions. In the penultimate step, the C-terminal domain of nsp14 catalyzes N7-methylation of a GpppA intermediate. Subsequent methylation by nsp16 results in a functional cap (⁷MeGpppA_{2'}OMe)⁴⁻⁹.

The large coronaviral genome requires replication fidelity to maintain its functionality. While the nsp12-centred polymerase complex is highly processive, it is also error-prone. An unaudited replication would result in excessive accumulation of mutations and the generation of defective progeny virions. Nsp14 associates with the replication complex and removes incorrectly incorporated nucleotides from the 3' end of the newly formed RNA strand. Nsp14 is associated with nsp10, a protein co-factor that modulates nsp14 exonuclease activity in the replication complex^{6, 10}.

The structural basis of nsp14 interaction with nsp10 has been elucidated¹¹. We have recently demonstrated that nsp14, nsp10 and nsp16 form a ternary complex, further modulating nsp14 catalytic activity¹². Despite the detailed structural characterization of the complexes, the mechanistic basis of the modulation of nsp14 exonuclease activity remains elusive. The absence of structural information on nsp14 alone hinders the mechanistic understanding of the consequences of binding with nsp10 and nsp10/nsp16.

Here, we provide the crystal structure of nsp14 in the apo-form (without nsp10), which, together with prior data on nsp10/nsp14 complex, provide insight into the modulatory role of nsp10 on nsp14 exonuclease activity. Furthermore, the presented structure strongly supports the nsp14/nsp10/nsp16 heterotrimer formation model.

RESULTS AND DISCUSSION

Full length SARS-CoV-2 nsp14 exonuclease/methyltransferase was expressed in *E. coli* and purified. The crystal structure of S-adenosylhomocysteine (SAH) bound nsp14 was determined at 2.5 Å resolution (**Table I**). The crystals belonged to P2₁ space group and contained 2 molecules in the asymmetric unit. Nsp14 is characterized by a bimodular structure corresponding to the two distinct catalytic activities (**Figure 1A**). The C-terminal methyltransferase domain of nsp14 (amino acids 287-524) is well defined by the electron density for most of its part save for amino acids 454-465. At the same time, a significant portion of the N-terminal exonuclease domain (amino acids 25-286) is undefined by the electron density, indicating significant flexibility. This, in particular, relates to amino acids 40-

44, 93-103, and 123-154, and is of importance for further discussion. The undefined parts are identical in both chains.

The methyltransferase domain of nsp14 is characterized by a fold atypical to methyltransferases. The structure is built around a central β -sheet composed of 5 instead of 7 canonical β -strands. Further, the methyltransferase domain of nsp14 contains an atypical peripheral zinc finger stabilized by another element that distinguishes nsp14 from the typical methyltransferase fold, a supplementary C-terminal α -helix. The methyltransferase fold within our structure closely resembles that previously seen in the nsp10/nsp14 complex of SARS-CoV-1 (PDB ID 5C8S) with rmsd of 0.57Å for 197 equivalent C α atoms. A reaction product, SAH, is clearly defined by the electron density in our structure (**Figure 1B**), but otherwise, the active site is empty (the methylated reaction product is absent). Interestingly, the active site faces a large, unobstructed water channel. As such, our structure is particularly valuable in terms of inhibitor development (fragment screening and inhibitor soaking).

Figure 1. Crystal structure of apo-nsp14. (A) Arrangement of the molecules in the asymmetric unit. Model A – blue, model B – pink. MTase and ExoN domains within the bimodular structure of nsp14 are distinguished by primary colour shade. Active site residues are indicated in red in one molecule, and the detailed arrangements of the active sites are shown in inserts. (B) Interactions of SAH (orange, stick model) at the active site of nsp14 MTase domain (pink, ribbon). Hydrogen bonds are depicted as dotted lines; waters are indicated as red spheres.

The SAM/SAH-binding region in our structure (331-DxGxPxA-337) diverges from the classical methyltransferase motif I (E/DxGxGxG; ¹³) but is consistent with that described in nsp14 of SARS-CoV-1 ¹⁴. The purine ring of SAH is positioned *via* sulphur/ π interaction with Cys387; hydrophobic interactions contributed by sidechains of Ala353, Val389, Phe367, and direct hydrogen bonds contributed by mainchain of Tyr368. A water-mediated hydrogen bond contributed by Asn388 is additionally observed in model B. Ribose is stabilized by hydrogen bonds contributed by the sidechain of Asp352. Mainchain carbonyl oxygens of Gly333 and Trp385 contribute hydrogen bonds with the cysteine amine, whereas its carboxyl group forms a salt bridge with Arg310. The identified SAH positioning residues are consistent with the conclusions of prior mutagenesis studies ¹⁵. The majority of reported nanomolar MTase inhibitors are SAM competitive ¹⁶. By delineating the detailed molecular interactions in the SAH pocket, our study facilitates the development of SAH mimetic inhibitors of SARS-CoV-2 MTase.

A structure of nsp14 from SARS-CoV-2 has been recently deposited at the protein data bank (PDB ID: 7QGI), which contains an empty active site (no substrates or reaction products).

Comparison with our structure demonstrates that the SAM/SAH binding site in 7QGI is occluded by a loop connecting $\beta 2'$ and $\beta 3'$ strands of a central β -sheet. We initially thought this could represent a mechanism of reaction product removal or a methyltransferase activity regulatory mechanism. However, a closer look at 7QGI crystal structure demonstrates that the place where the loop in question is located in SAH bound structures (nsp10/nsp14 complex from SARS-CoV-1 and this study) is occluded by an adjacent molecule in 7QGI. This suggests that the blocked active site in 7QGI is an artefact of crystal packing rather than representing a physiologically relevant conformation. Supporting the above conclusion, when Ma and colleagues¹⁴ analyzed SARS-CoV-1 nsp10/nsp14 complexes in SAH bound and empty-forms, the loop in question had the same conformation in both SAH bound and empty structure (and comparable to that found in our structure), further suggesting that 7QGI represents a non-physiological conformation within the discussed feature.

The exonuclease domain of nsp14 is built around a five-stranded, twisted β -sheet flanked by α -helices. Such arrangement of the core structure elements follows the overall design of DEDD family exonucleases but contains unique features: the N-terminal region (1-84) and the first zinc finger. The striking features of the exonuclease domain in our structure became evident only upon comparison with the structure of nsp10/nsp14 complex. The N-terminal of nsp14 spanning amino acids 1-70 is folded and oriented differently in compared structures (**Figure 2A, B**). The N-terminal is well defined by electron density in the nsp10/nsp14 complex, but contains little ordered secondary structures. Amino acids 1-56 constitute the first of the two interaction regions with nsp10 (nomenclature of structural features after Ma *et al.*¹⁴). The second interaction region comprises residues 60-67 and 200-202. In apo-nsp14 structure, residues 1-24 are not defined by the electron density. Region 25-70 is completely refolded in apo-nsp14 structure compared to the nsp10/nsp14 complex. For example, residues Thr50-Met58 form an α -helix in apo-nsp14 structure, whereas residues 51-54 form a $\beta 1$ -strand within $\beta 1$, $\beta 5$, $\beta 6$ sheet in the nsp10/nsp14 complex. The remaining residues (25-49, 59-67) contribute completely different intermolecular interactions within apo-nsp14 structure compared to nsp10/nsp14 complex. In the absence of nsp10, the N-terminal region of nsp14 covers the second nsp10 interaction region, burying, among others, residues 200-202. Thereby, the N-terminus of nsp14 acts as a lid, covering the nsp10 binding site in the absence of the protein co-factor. The lid rearranges its structure in the presence of nsp10 to create a significant interaction surface with nsp10.

Nsp10 enhances the 3'-exonucleolytic activity of nsp14 but does not significantly affect the methyltransferase activity. Nsp10 effect on nsp14 was rationalized by nsp10/nsp14 crystal structure, where nsp10 contributes significant contacts with ExoN domain of nps14 but does not interact with the N7-methyltransferase domain¹⁴. The extensive surface area of interaction suggested that nsp10 maintains the structural stability of ExoN, but the nature and extent of instability of apo-nsp14 were not known in the absence of the structure. In the current study, we show that the ExoN domain of nsp14 is partially unfolded in the absence of nsp10. The proximal part of $\beta 2$ - $\beta 3$ sheet adjacent to catalytic Glu92 is undefined by electron density (flexible, unstructured) in apo-nsp14. The same is true for the entire region between $\beta 4$ and $\alpha 3$. Residues constituting $\beta 5$ - $\beta 6$ and $\alpha 2$ in the nsp10/nsp14 complex structure are not defined by the electron density in the apo-nsp14 structure. Significantly, in apo-nsp14, the catalytic metal ion binding site within DEDD motif is partially occluded by the sidechain of

His188. Further, the RNA binding site is partially occluded by the lid fragment (Thr35-Lys47), which buries the nsp10 binding site. Only nsp10 induced refolding of the lid removes the steric occlusion and induces a catalytically capable conformation of nsp14 observed in nsp10/nsp14 complex (Figure 2B).

Figure 2. Nsp10 induced refolding of nsp14 ExoN lid subdomain. (A) Nsp10 (navy blue, ribbon) / Nsp14 (grey, surface) complex overlaid on apo-nsp14 (cyan). The N-terminal lid of nsp14 refolds upon binding of nsp10 as exemplified by Thr50-Met58 region assuming a helical structure in apo-nsp14 and forming a strand of a β-sheet in nsp10/nsp14 complex. In the apo structure, the lid occludes nsp10 binding site at the surface of nsp14. (B) The RNA interaction site of nsp10 (navy blue) /nsp14 (beige) complex (7N0B) is sterically occluded by lid subdomain in apo-nsp14 structure (cyan), providing a rationale for the negligible exonuclease activity of nsp14 in the absence of nsp10. RNA (orange). The apo-nsp14 loop, occluding the RNA binding site, is denoted in dark cyan, dotted line.

We have demonstrated earlier that nsp14, nsp10 and nsp16 form a ternary complex centered around nsp10¹². However, such complex is not supported by the structural data available on nsp10/nsp14 and nsp10/nsp16 binary complexes. When the binary complexes are overlaid on nsp10, the lid of nsp14 clashes with nsp16, demonstrating that ternary complex formation would require a structural rearrangement within the lid. The current study demonstrates that the lid of nsp14 is capable of structural morphing, indirectly supporting the mechanism of triplex formation.

CONCLUSIONS

The current pandemics highlighted the importance of understanding coronavirus physiology. Prior research on other human and animal coronaviruses facilitated an unprecedented pace of vaccine and antiviral development, but the constant genetic drift and the threat of new zoonotic transmission warrant continued effort.

Nsp14 constitutes a valid target in SARS-CoV-2, but the structural information was incomplete. This study provided the structural basis for a mechanistic explanation of nsp10 mediated modulation of nsp14 3'-exonuclease activity, defined the structure of nsp14 N7-methyltransferase in the absence of RNA substrate, and offered crystallization conditions suitable for soaking of small molecules at the active site of nsp14 N7-methyltransferase. Thereby, the study contributes to our understanding of coronavirus physiology and facilitates the effort of anti-coronaviral drug development.

200 MATERIALS AND METHODS

Expression construct

Sequence encoding amino acids 5926 – 6452 of SARS CoV-2 polyprotein 1ab (Uniprot P0DTD1), including D90A and D92A substitutions ⁵, flanked by BamHI and NotI cleavage sites and containing a TEV protease cleavage site was codon-optimized for expression in *Escherichia coli* and obtained by synthesis at GeneArt. The sequence was subcloned into pETDuet-1 expression vector.

Protein expression and purification

210 The plasmid was transformed into *E. coli* BL21 Rosetta2 (DE3) strain. Expression was carried out in Terrific Broth medium (Bioshop) supplemented with 100 µg/ml ampicillin (Sigma-Aldrich). Bacterial cultures were incubated at 37°C until the OD₆₀₀ value reached 1.2 - 1.4. The cultures were cooled to 4°C for 20 min, and the recombinant protein expression was induced with 0.5 mM isopropyl-D-1-thiogalactopyranoside (IPTG; Sigma-Aldrich). The bacteria were incubated at 18°C for additional 18 hours. Bacterial cells were harvested by centrifugation at 7,000 × g for 30 min at 4°C and stored at -20 frozen °C before the purification.

220 Bacterial pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.5, 500 mM NaCl, 5 mM MgCl₂, 10% v/v glycerol, 5 mM β-mercaptoethanol, 10 mM imidazole), supplemented with benzonase (Millipore) and protease inhibitors cocktail (Roche), and lysed by sonication at 80% amplitude for 20 min at 14°C with 3 s on / 2 s off cycles. The fraction of the supernatant containing soluble protein was separated from debris by centrifugation at 25,000 × g for 40 min at 4°C. The clarified supernatant was incubated overnight at 4°C with HisPur Ni-NTA Resin (Thermo Scientific) pre-equilibrated with the lysis buffer. Resin was sequentially washed with lysis buffer, then with wash buffer 1 (50 mM Tris-HCl pH 8.5, 500 mM NaCl, 5 mM MgCl₂, 10% v/v glycerol, 5 mM β-mercaptoethanol, 20 mM imidazole) and wash buffer 2 (50 mM Tris-HCl pH 8.5, 500 mM NaCl, 5 mM MgCl₂, 10% v/v glycerol, 5 mM β-mercaptoethanol, 30 mM imidazole). The immobilized protein was eluted with elution buffer (50 mM Tris-HCl pH 8.5, 250 mM NaCl, 5 mM MgCl₂, 10% v/v glycerol, 5 mM β-mercaptoethanol, 250 mM imidazole). Fractions containing the protein of interest (as determined by SDS-PAGE) were pooled and supplemented with tobacco etch virus (TEV) protease in a 1:10 molar ratio. TEV cleavage was 230 combined with overnight dialysis into a cleavage buffer (50 mM Tris-HCl pH 8.5, 200 mM NaCl, 5 mM MgCl₂, 10% v/v glycerol, 5 mM β-mercaptoethanol, 10 mM imidazole). Reverse chromatography on HisPur Ni-NTA resin pre-equilibrated with the dialysis buffer allowed the cleaved His-tag removal. The protein was further purified by size exclusion chromatography on HiLoad 26/600 Superdex 200 prep grade column (GE Healthcare) equilibrated with the working buffer (50 mM Tris-HCl pH 8.5, 150 mM NaCl, 5 mM MgCl₂, 1 mM TCEP). The purity of the sample was assessed by SDS-PAGE.

Crystallization and data collection

240 The purified protein was concentrated to 11.5 mg/ml. The sample was supplemented with 2mM SAM. Crystallization screening was performed at 4°C using the sitting drop vapour diffusion method. Crystals were obtained from a solution containing 15 % v/v 2-propanol, 0.2 M imidazole pH 7.6 and 500 mM polypropylene glycol 400. Further optimization was performed around initial conditions where the crystals generally appeared after 4 days. The

crystals were cryoprotected in 30% ethylene glycol in the mother liquor and flash-cooled in liquid nitrogen. The diffraction data were collected at Swiss Light Source (SLS, Villigen, Switzerland) beamline X06DA - PXIII. The data were indexed and integrated using XDS ¹⁷. The data were scaled and merged using STARANISO webserver. A molecular replacement solution was found using Phaser ¹⁸ and the structure of SARS-CoV-1 nsp14 (PDB code: 5C8S) as a search model. SAH was identified in the electron density map using CheckMyBlob server (https://checkmyblob.bioreproducibility.org). Restraints were obtained from ligand builder in Coot ¹⁹. The initial model was subjected to several iterations of manual and automated refinement cycles using COOT and REFMAC5, respectively ^{20, 21}. Throughout the refinement, 5% of the reflections were used for cross-validation analysis, and the behaviour of R_{free} was employed to monitor the refinement strategy.

Data analysis

The final model was analyzed and the graphics was prepared using PyMOL (Schrödinger, LLC).

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; to KP).

DECLARATION OF INTERESTS

None.

References

1. Pyrc, K.; Jebbink, M. F.; Berkhout, B.; van der Hoek, L., Genome structure and transcriptional regulation of human coronavirus NL63. *Virology* **2004**, *1*, 7.
- 270 2. Yadav, R.; Chaudhary, J. K.; Jain, N.; Chaudhary, P. K.; Khanra, S.; Dhamija, P.; Sharma, A.; Kumar, A.; Handu, S., Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells* **2021**, *10* (4).
3. Sola, I.; Almazán, F.; Zúñiga, S.; Enjuanes, L., Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annual Review of Virology* **2015**, *2* (1), 265-288.
4. Sevajol, M.; Subissi, L.; Decroly, E.; Canard, B.; Imbert, I., Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. *Virus Res* **2014**, *194*, 90-9.
5. Bouvet, M.; Imbert, I.; Subissi, L.; Gluais, L.; Canard, B.; Decroly, E., 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* **2012**, *109* (24), 9372-7.
- 280 6. Eckerle, L. D.; Becker, M. M.; Halpin, R. A.; Li, K.; Venter, E.; Lu, X.; Scherbakova, S.; Graham, R. L.; Baric, R. S.; Stockwell, T. B.; Spiro, D. J.; Denison, M. R., Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog* **2010**, *6* (5), e1000896.
7. Yan, L.; Ge, J.; Zheng, L.; Zhang, Y.; Gao, Y.; Wang, T.; Huang, Y.; Yang, Y.; Gao, S.; Li, M.; Liu, Z.; Wang, H.; Li, Y.; Chen, Y.; Guddat, L. W.; Wang, Q.; Rao, Z.; Lou, Z., Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell* **2021**, *184* (1), 184-193.e10.
8. Chen, Y.; Cai, H.; Pan, J.; Xiang, N.; Tien, P.; Ahola, T.; Guo, D., Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc Natl Acad Sci U S A* **2009**, *106* (9), 3484-9.
- 290 9. Bouvet, M.; Debarnot, C.; Imbert, I.; Selisko, B.; Snijder, E. J.; Canard, B.; Decroly, E., In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog* **2010**, *6* (4), e1000863.
10. Smith, E. C.; Denison, M. R., Coronaviruses as DNA wannabes: a new model for the regulation of RNA virus replication fidelity. *PLoS Pathog* **2013**, *9* (12), e1003760.
11. Baddock, H. T.; Brolih, S.; Yosaatmadja, Y.; Ratnaweera, M.; Bielinski, M.; Swift, L. P.; Cruz-Migoni, A.; Fan, H.; Keown, J. R.; Walker, A. P.; Morris, G. M.; Grimes, J. M.; Fodor, E.; Schofield, C. J.; Gileadi, O.; McHugh, P. J., Characterization of the SARS-CoV-2 ExoN (nsp14ExoN-nsp10) complex: implications for its role in viral genome stability and inhibitor identification. *Nucleic Acids Res* **2022**.
- 300 12. Matsuda, A.; Plewka, J.; Chykunova, Y.; Jones, A. N.; Pachota, M.; Rawski, M.; Mourão, A.; Karim, A.; Kresik, L.; Lis, K.; Minia, I.; Hartman, K.; Sonani, R.; Schlauderer, F.; Dubin, G.; Sattler, M.; Suder, P.; Popowicz, G.; Pyrc, K.; Czarna, A., Despite the odds: formation of the SARS-CoV-2 methylation complex. *bioRxiv* **2022**, 2022.01.25.477673.
13. Martin, J. L.; McMillan, F. M., SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol* **2002**, *12* (6), 783-93.
14. Ma, Y.; Wu, L.; Shaw, N.; Gao, Y.; Wang, J.; Sun, Y.; Lou, Z.; Yan, L.; Zhang, R.; Rao, Z., Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc Natl Acad Sci U S A* **2015**, *112* (30), 9436-41.
15. Ogando, N. S.; El Kazzi, P.; Zevenhoven-Dobbe, J. C.; Bontes, B. W.; Decombe, A.; Posthuma, C. C.; Thiel, V.; Canard, B.; Ferron, F.; Decroly, E.; Snijder, E. J., Structure-function analysis of the nsp14 N7-guanine methyltransferase reveals an essential role in Betacoronavirus replication. *Proc Natl Acad Sci U S A* **2021**, *118* (49).
- 310 16. Devkota, K.; Schapira, M.; Perveen, S.; Khalili Yazdi, A.; Li, F.; Chau, I.; Ghiabi, P.; Hajian, T.; Loppnau, P.; Bolotokova, A.; Satchell, K. J. F.; Wang, K.; Li, D.; Liu, J.; Smil, D.; Luo, M.; Jin, J.; Fish, P. V.; Brown, P. J.; Vedadi, M., Probing the SAM Binding Site of SARS-CoV-2 Nsp14 In Vitro Using SAM Competitive Inhibitors Guides Developing Selective Bisubstrate Inhibitors. *SLAS Discov* **2021**, *26* (9), 1200-1211.

17. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 2), 125-32.
18. McCoy, A. J., Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **2007**, *63* (Pt 1), 32-41.
19. Emsley, P., Tools for ligand validation in Coot. *Acta Crystallogr D Struct Biol* **2017**, *73* (Pt 3), 203-210.
20. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004**, *60* (Pt 12 Pt 1), 2126-32.
21. Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **2011**, *67* (Pt 4), 355-67.

330

TABLES

Table I

Data collection and refinement statistics. Values provided in parentheses are for the highest resolution shell.

Data collection statistics	
PDB ID	7R2V
Wavelength (Å)	1.00004
Resolution range	86.59 - 2.53
Space group	P 1 21 1
Unit cell	67.18, 100.31, 90.87 (Å)
	90.00, 107.66, 90.00 (°)
Data scaling software	autoPROC (Version 1.0.5), AIMLESS, STARANISO
Total reflections	19,046
Unique reflections	19,046
Multiplicity	4.62 (4.72)
Completeness (%)	100.00 (99.92)
Mean I / sigma (I)	3.80 (0.33)
Wilson B-factor (Å ²)	41.31
R-merge	0.2117 (1.5567)
CC1/2	0.9510 (0.1646)
CC*	1 (0.962)
Structure refinement statistics	
Total reflections	19,046
Unique reflections	19,046
Wilson B-factor	41.31
Reflections in refinement	19,045
Reflections in free set	911
R _{work}	0.201
R _{free}	0.265
FSC average	0.8907
RMSD bonds	0.0056
RMSD angles	1.418
Ramachandran favoured (%)	90.4
Ramachandran allowed (%)	8.8
Ramachandran outliers (%)	0.8
Rotamer outliers	7.6
Clash score	4.9
Overall number of atoms (non-H)	7,061
in macromolecules	6,993
in ligands	20
in solvent	48
Average B-factor	56.4
for macromolecules	56.7
for ligands	44.4
for solvent	29.4

SUPPORTING INFORMATION

Crystal structure of apo-nsp14 in the absence of nsp10.

Materials and Methods

340 Nsp14 construct used in the study

The initial glycine residue is a part of TEV recognition site and remains appended to nsp14 sequence in the purified nsp14.

Nucleotide sequence:

(ggc) atggctgaaaatgtaacaggactctttaagattgtagtaaggtaatcactgggttacatcct
acacaggcacctacacacctcagtgttgacactaaattcaaaactgaagggttatgtgttgacatacc
tggcatacctaaggacatgacctatagaagactcatctctatgatgggttttaaaatgaattatcaag
ttaatgggttacctaacatgtttatcaccgcgaagaagctataagacatgtacgtgcatggattggc
ttcgctgtagctgggtgtcatgctactagagaagctgttggtaccaattttacctttacagctagggtt
350 ttctacaggtgttaacctagttgctgtacctacaggttatgttgatacacctaataatacagattttt
ccagagtttagtgctaaaccaccgcctggagatcaatttaaacacctcataccacttatgtacaaagga
cttccttggaaatgtagtgcgataaagattgtacaaatgttaagtgaacacacttaaaaatctctctga
cagagtcgtattttgtcttatgggcacatggcctttgagttgacatctatgaagtattttgtgaaaatag
gacctgagcgcacctgttgctctatgtgatagacgtgccacatgcttttccactgcttcagacacttat
gcctgttggcatcattctattggatttgattacgtctataatccgtttatgattgatgttcaacaatg
gggttttacaggtaacctacaaagcaaccatgatctgtattgtcaagtccatggtaatgcacatgtag
ctagttgtgatgcaatcatgactaggtgtctagctgtccacgagtgctttgttaagcgtgttgactgg
actattgaatatcctataaattgggtgatgaactgaagattaatgcggcctttagaagggttcaacacat
gggtgttaaagctgcattatttagcagacaaattcccagttcttcacgacattggtaaccctaaagcta
360 ttaagtgtgtacctcaagctgatgtagaatggaagttctatgatgcacagccttgtagtgacaaagct
tataaaatagaagaattattctattcttatgccacacattctgacaaattcacagatgggtgatgcct
attttggaaattgcaatgtcgatagatatcctgctaattccattgtttgtagatttgacactagagtgc
tatctaaccttaacttgccctgggtgtgatgggtggcagtttgtagtataataaacatgcattccacaca
ccagcttttgataaaagtgcctttgttaatttaaaacaattaccatttttctattactctgacagtcc
atgtgagtcctcatggaaaacaagtagtgtcagatatagattatgtaccactaaagtctgctacgtgta
taacacgttgcaatttaggtgggtgctgtctgtagacatcatgctaattgagtacagattgtatctcgat
gcttataacatgatgatctcagctggccttttagcttggtgggtttacaacaatttgatacttataacct
ctggaacactttttacaagacttcag

370 Amino acid sequence

(G) MAENV TGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGFKMNY
QVNGYPNMFITREEAIRHVRAWIGFAVAGCHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTD
FSRVSAKPPPGDQFKHLIPLMYKGLPWNVVRKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVK
IGPERTCCLDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLVCQVHGNAH
VASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADKFPVLHDIGNPK
AIKCV PQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCLFWNCNVD RY PANSIVCRFDTR
VLSN LNLP GCDGGS LYNKHAFHTPAFDKSAFVN LKQLPFFYYSDSPCESHGKQVVS D IDYVPLKSAT
CITRCNLGGAVCRHHANEYRLYLDAYNMMISAGFSLWVYKQFDTYNLWNFTRLQ