
Accelerated Article Preview

Structure of replicating SARS-CoV-2 polymerase

Received: 27 April 2020

Accepted: 15 May 2020

Accelerated Article Preview Published
online 21 May 2020

Cite this article as: Hillen, H. S. et al.
Structure of replicating SARS-CoV-2
polymerase. *Nature* <https://doi.org/10.1038/s41586-020-2368-8> (2020).

Hauke S. Hillen, Goran Kokic, Lucas Farnung, Christian Dienemann, Dimitry Tegunov & Patrick Cramer

This is a PDF file of a peer-reviewed paper that has been accepted for publication. Although unedited, the content has been subjected to preliminary formatting. Nature is providing this early version of the typeset paper as a service to our authors and readers. The text and figures will undergo copyediting and a proof review before the paper is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

Structure of replicating SARS-CoV-2 polymerase

<https://doi.org/10.1038/s41586-020-2368-8>

Received: 27 April 2020

Accepted: 15 May 2020

Published online: 21 May 2020

Hauke S. Hillen^{1,2}, Goran Kokic^{1,2}, Lucas Farnung^{1,2}, Christian Dienemann^{1,2}, Dmitry Tegunov^{1,2} & Patrick Cramer^{1✉}

The coronavirus SARS-CoV-2 uses an RNA-dependent RNA polymerase (RdRp) for the replication of its genome and the transcription of its genes^{1–3}. Here we present the cryo-electron microscopic structure of the SARS-CoV-2 RdRp in active form, mimicking the replicating enzyme. The structure comprises the viral proteins nsp12, nsp8, and nsp7, and over two turns of RNA template-product duplex. The active site cleft of nsp12 binds the first turn of RNA and mediates RdRp activity with conserved residues. Two copies of nsp8 bind to opposite sides of the cleft and position the second turn of RNA. Long helical extensions in nsp8 protrude along exiting RNA, forming positively charged ‘sliding poles’. These sliding poles can account for the known processivity of the RdRp that is required for replicating the long coronavirus genome³. Our results enable a detailed analysis of the inhibitory mechanisms that underlie the antiviral activity of substances such as remdesivir, a drug for the treatment of coronavirus disease 2019 (COVID-19)⁴.

Coronaviruses are positive-strand RNA viruses that pose a major health risk¹. The novel severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)^{5,6} has caused a pandemic referred to as coronavirus disease 2019 (COVID-19). Coronaviruses use an RNA-dependent RNA polymerase (RdRp) complex for the replication of their genome and for the transcription of their genes^{2,3}. This RdRp complex is the target for nucleoside analogue inhibitors, in particular remdesivir^{7,8}. Remdesivir inhibits coronavirus RdRp^{9,10} and shows antiviral activity in cell culture and animals¹¹. Remdesivir is currently tested in the clinic in many countries¹² and was recently approved for emergency treatment of COVID-19 patients in the United States⁴.

The RdRp of SARS-CoV-2 is composed of a catalytic subunit called non-structural protein (nsp) 12¹³, and two accessory subunits, nsp8 and nsp7^{3,14}. The structure of the RdRp was recently reported¹⁵ and is highly similar to the RdRp of SARS-CoV¹⁶, a zoonotic coronavirus that spread into human population in 2002¹. The nsp12 subunit contains an N-terminal nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain, an interface domain, and a C-terminal RdRp domain^{15,16}. The RdRp domain resembles a right hand and comprises the fingers, palm, and thumb subdomains^{15,16} that are found in all single-subunit polymerases. Subunit nsp7 binds to the thumb, whereas two copies of nsp8 bind to the fingers and thumb subdomains^{15,16}. Structural information is also available for nsp8-nsp7 complexes^{17,18}.

To obtain the structure of the SARS-CoV-2 RdRp in its active form, we prepared recombinant nsp12, nsp8 and nsp7 (Fig. 1a, Methods). When added to a minimal RNA hairpin substrate (Fig. 1b), the purified proteins gave rise to RNA-dependent RNA extension activity, which depended on nsp8 and nsp7 (Fig. 1c). We assembled and purified a stable RdRp-RNA complex with the use of a self-annealing RNA, and collected single-particle cryo-electron microscopy (cryo-EM) data (Extended Data Figure 1, Extended Data Table 1). Particle classification

yielded a 3D reconstruction at a nominal resolution of 2.9 Å (Extended Data Figure 1). This led to a refined structure of the RdRp-RNA complex that showed great details of the RNA in the active center (Extended Data Figure 2).

The structure shows the RdRp enzyme engaged with over two turns of duplex RNA (Fig. 2, Supplementary video 1). The structure resembles that of the free enzyme¹⁵, but additionally reveals a long protruding RNA and extended protein regions in nsp8 (Extended Data Figure 3a). These observations are unique, as RdRp complexes of hepatitis C virus¹⁹, poliovirus²⁰, and norovirus²¹ contain only one turn of RNA and show no features resembling the newly observed nsp8 extensions (Extended Data Figure 3b).

Our structure provides details of the RdRp-RNA interactions (Fig. 3). The nsp12 subunit binds the first turn of RNA between its fingers and thumb subdomains (Fig. 3a, b). The active site is located on the palm subdomain and formed by five conserved nsp12 elements called motifs A-E (Fig. 3b). Motif C binds the RNA 3'-end and contains the residues D760 and D761, which are required for RNA synthesis^{10,14}. The additional nsp12 motifs F and G reside in the fingers subdomain and position the RNA template (Fig. 3b). The observed nsp12 contacts to the RNA product strand may retain short RNA during early steps of RNA synthesis.

As the RNA duplex exits from the RdRp cleft, it forms a second helical turn that protrudes from the nsp12 surface (Fig. 3c). There are no structural elements in the RdRp that restrict extension of the RNA duplex. These observations are consistent with the production of a double-stranded RNA during replication. However, it is unclear whether replication in infected cells results in RNA duplexes or whether RNA strands are separated and if so, how. It is also unknown when and how RNA strands are separated during the transcription of viral genes, which must produce single-stranded product mRNAs that can be translated.

¹Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Göttingen, Germany. ²These authors contributed equally: Hauke S. Hillen, Goran Kokic, Lucas Farnung, Christian Dienemann, Dmitry Tegunov. ✉e-mail: patrick.cramer@mpiibpc.mpg.de

The protruding, exiting RNA duplex is flanked by long α -helical extensions that are formed by the highly conserved¹⁷ N-terminal regions in the two nsp8 subunits (Figs. 2, 3). These prominent nsp8 extensions reach up to 28 base pairs away from the active site and use positively charged residues that are positioned to interact with the RNA backbones (Fig. 3). The two nsp8 extensions differ with respect to their RNA interactions, also arguing for sequence-independent binding. The two nsp8 copies adopt different structures in the RdRp complex, and interact differently with nsp7 and nsp12 subdomains (Extended Data Figure 3c). The nsp8 extensions also adopt different structures in crystals of nsp8-nsp7 complexes^{17,18}, and are mobile in free RdRp^{15,16}. This indicates that the nsp8 extensions are flexible in the RdRp complex and get ordered when an RNA duplex exits from the enzyme.

The interactions of the nsp8 extensions with exiting RNA may explain the processivity of the RdRp, which is required for replicating the very long RNA genome of coronaviruses and other viruses of the nidovirus family³. It is known that nsp8 and nsp7 confer processivity to nsp12¹⁴. It is also known that mutation of nsp8 residue K58 to alanine is lethal for the virus¹⁴. K58 is located in the nsp8 extension, and interacts with exiting RNA around the minor groove (Fig. 3c). The nsp8 extensions may be regarded as 'sliding poles' that slide along exiting RNA to prevent premature dissociation of the RdRp during replication. The sliding poles may serve a function similar to 'sliding clamps' that confer processivity to DNA replication machines²².

To investigate how the RdRp binds the incoming nucleoside triphosphate (NTP) substrate, we superimposed our structure onto the related structure of the norovirus RdRp-nucleic acid complex²¹. This suggested that the NTP-binding site is conserved, including putative contacts between nsp12 and the NTP (Extended Data Fig. 3d, Supplementary video 2). Residues N691, S682 and D623 may recognize the 2'-OH group of the NTP, thereby rendering the RdRp specific for the synthesis of RNA rather than DNA. Our modeling is also consistent with binding of the triphosphorylated form of remdesivir to the NTP site, because there is space to accommodate the additional nitrile group that is present at the 1' position of the ribose ring of remdesivir (Extended Data Fig. 3d).

Recently, the structure of another SARS-CoV-2 RdRp-RNA complex became available²³, and was published²⁴ when our manuscript was under review. Comparison of the two studies shows that the core structures are similar, whereas we additionally observe a second turn of RNA and the nsp8 'sliding poles'. The other study suggested that remdesivir functions as an 'immediate' RNA chain terminator^{23,24}. However, published biochemistry showed that several more nucleotides can be added to RNA following remdesivir incorporation, leading to 'delayed' termination^{10,25}. We note that the latter mechanism can explain how remdesivir escapes removal from the RNA 3'-end by the viral exonuclease nsp14²⁶ that binds the RdRp complex¹⁴. Mechanistic questions on coronavirus replication, transcription, and antiviral targeting can now be studied based on the results presented here.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2368-8>.

- Hilgenfeld, R. & Peiris, M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Res* **100**, 286-295, <https://doi.org/10.1016/j.antiviral.2013.08.015> (2013).
- Snijder, E. J., Decroly, E. & Ziebuhr, J. The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv Virus Res* **96**, 59-126, <https://doi.org/10.1016/bs.avir.2016.08.008> (2016).
- Posthuma, C. C., Te Velthuis, A. J. W. & Snijder, E. J. Nidovirus RNA polymerases: Complex enzymes handling exceptional RNA genomes. *Virus Res* **234**, 58-73, <https://doi.org/10.1016/j.virusres.2017.01.023> (2017).
- Ledford, H. Hopes rise for coronavirus drug remdesivir. *Nature*, <https://doi.org/10.1038/d41586-020-01295-8> (2020).
- Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269, <https://doi.org/10.1038/s41586-020-2008-3> (2020).
- Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, <https://doi.org/10.1038/s41586-020-2012-7> (2020).
- Subissi, L. et al. SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. *Antiviral Res* **101**, 122-130, <https://doi.org/10.1016/j.antiviral.2013.11.006> (2014).
- Jordan, P. C., Stevens, S. K. & Deval, J. Nucleosides for the treatment of respiratory RNA virus infections. *Antivir Chem Chemother* **26**, 2040206618764483, <https://doi.org/10.1177/2040206618764483> (2018).
- Agostini, M. L. et al. Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. *mBio* **9**, <https://doi.org/10.1128/mBio.00221-18> (2018).
- Gordon, C. J., Tcheshnokov, E. P., Feng, J. Y., Porter, D. P. & Gotte, M. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *J Biol Chem*, <https://doi.org/10.1074/jbc.AC120.013056> (2020).
- Sheahan, T. P. et al. Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci Transl Med* **9**, <https://doi.org/10.1126/scitranslmed.aal3653> (2017).
- Kupferschmidt, K. & Cohen, J. Race to find COVID-19 treatments accelerates. *Science* **367**, 1412-1413, <https://doi.org/10.1126/science.367.6485.1412> (2020).
- Ahn, D. G., Choi, J. K., Taylor, D. R. & Oh, J. W. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Arch Virol* **157**, 2095-2104, <https://doi.org/10.1007/s00705-012-1404-x> (2012).
- Subissi, L. et al. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci U S A* **111**, E3900-3909, <https://doi.org/10.1073/pnas.1323705111> (2014).
- Gao, Y. et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, eabb7498, <https://doi.org/10.1126/science.abb7498> (2020).
- Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* **10**, 2342, <https://doi.org/10.1038/s41467-019-10280-3> (2019).
- Zhai, Y. et al. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat Struct Mol Biol* **12**, 980-986, <https://doi.org/10.1038/nsmb999> (2005).
- Xiao, Y. et al. Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. *J Virol* **86**, 4444-4454, <https://doi.org/10.1128/JVI.06635-11> (2012).
- Appleby, T. C. et al. Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* **347**, 771-775, <https://doi.org/10.1126/science.1259210> (2015).
- Gong, P. & Peersen, O. B. Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc Natl Acad Sci U S A* **107**, 22505-22510, <https://doi.org/10.1073/pnas.1007626107> (2010).
- Zamyatkin, D. F., Parra, F., Machin, A., Grochulski, P. & Ng, K. K. Binding of 2'-amino-2'-deoxycytidine-5'-triphosphate to norovirus polymerase induces rearrangement of the active site. *J Mol Biol* **390**, 10-16, <https://doi.org/10.1016/j.jmb.2009.04.069> (2009).
- Moldovan, G. L., Pfander, B. & Jentsch, S. PCNA, the maestro of the replication fork. *Cell* **129**, 665-679, <https://doi.org/10.1016/j.cell.2007.05.003> (2007).
- Yin, W. et al. Structural Basis for the Inhibition of the RNA-Dependent RNA Polymerase from SARS-CoV-2 by Remdesivir. *bioRxiv*, 2020.2004.2008.032763, <https://doi.org/10.1101/2020.04.08.032763> (2020).
- Yin, W. et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*, <https://doi.org/10.1126/science.abc1560> (2020).
- Tcheshnokov, E. P., Feng, J. Y., Porter, D. P. & Gotte, M. Mechanism of Inhibition of Ebola Virus RNA-Dependent RNA Polymerase by Remdesivir. *Viruses* **11**, <https://doi.org/10.3390/v11040326> (2019).
- 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, <https://doi.org/10.1073/pnas.1201130109> (2012).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

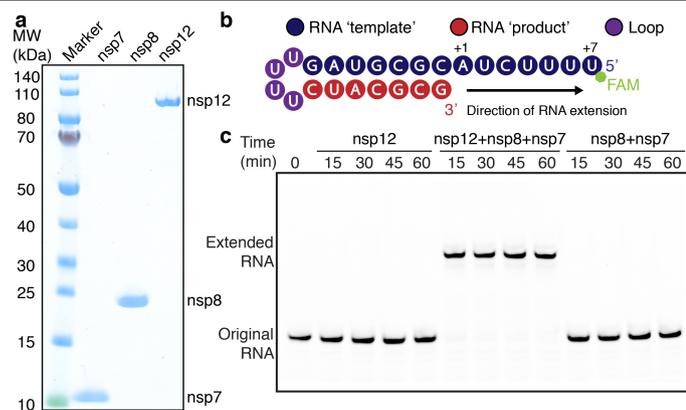


Fig. 1 | Preparation of active SARS-CoV-2 RdRp. **a.** SDS-PAGE analysis of purified SARS-CoV-2 RdRp subunits nsp12, nsp8 and nsp7. The experiment was performed once. **b.** Minimal RNA substrate that folds into a hairpin with 'template' and 'product' regions. The RNA contains a 7-nucleotide fluorescently labeled 5'-overhang. **c.** Incubation of the RdRp subunits (a) with RNA (b) leads to efficient RNA extension. RNAs were separated on a denaturing acrylamide gel and visualized with a Typhoon 95000 FLA Imager. Representative result of three independent experiments (Supplementary Figure 1).

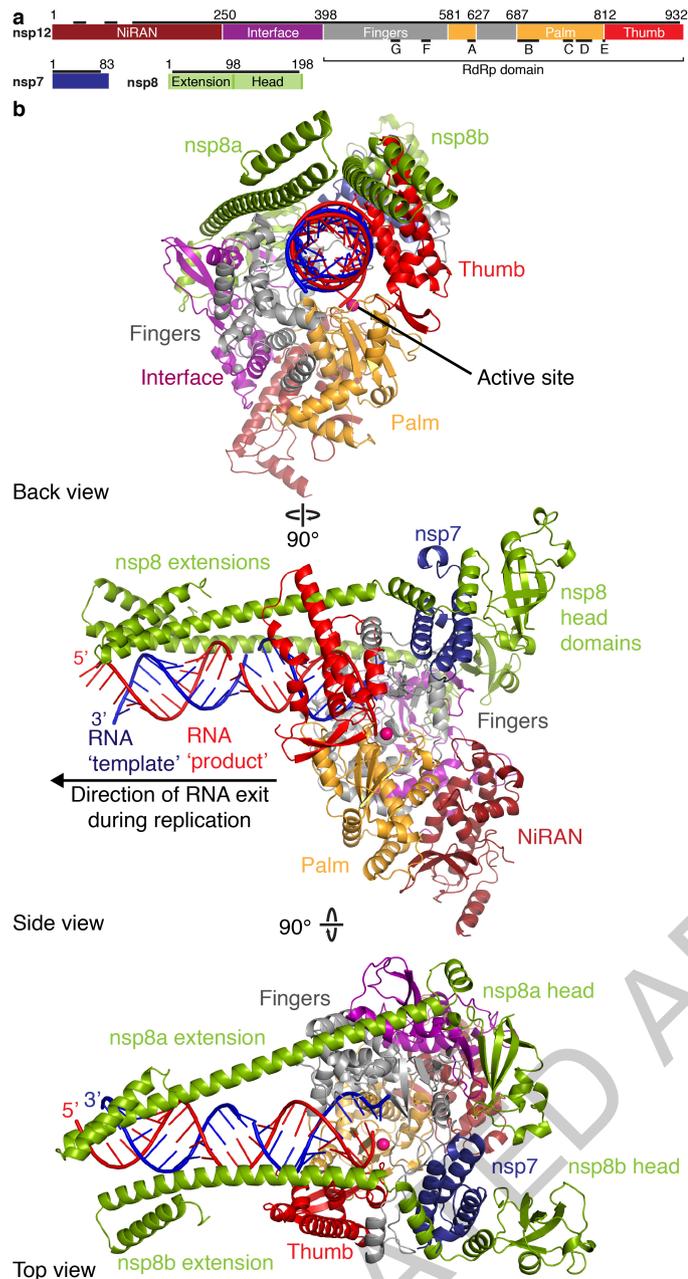


Fig. 2 | Structure of SARS-CoV-2 RdRp-RNA complex. **a.** Domain structure of RdRp subunits nsp12, nsp8, and nsp7. In nsp12, the conserved sequence motifs A-C¹⁶ are depicted. Regions included in the structure are indicated with black bars. **b.** Three views of the structure, related by 90-degree rotations. Color code for nsp12 (NiRAN, interface, fingers, palm, thumb), nsp8, nsp7, RNA template (blue) and RNA product (red) used throughout. The magenta sphere depicts a modeled²¹ metal ion in the active site.

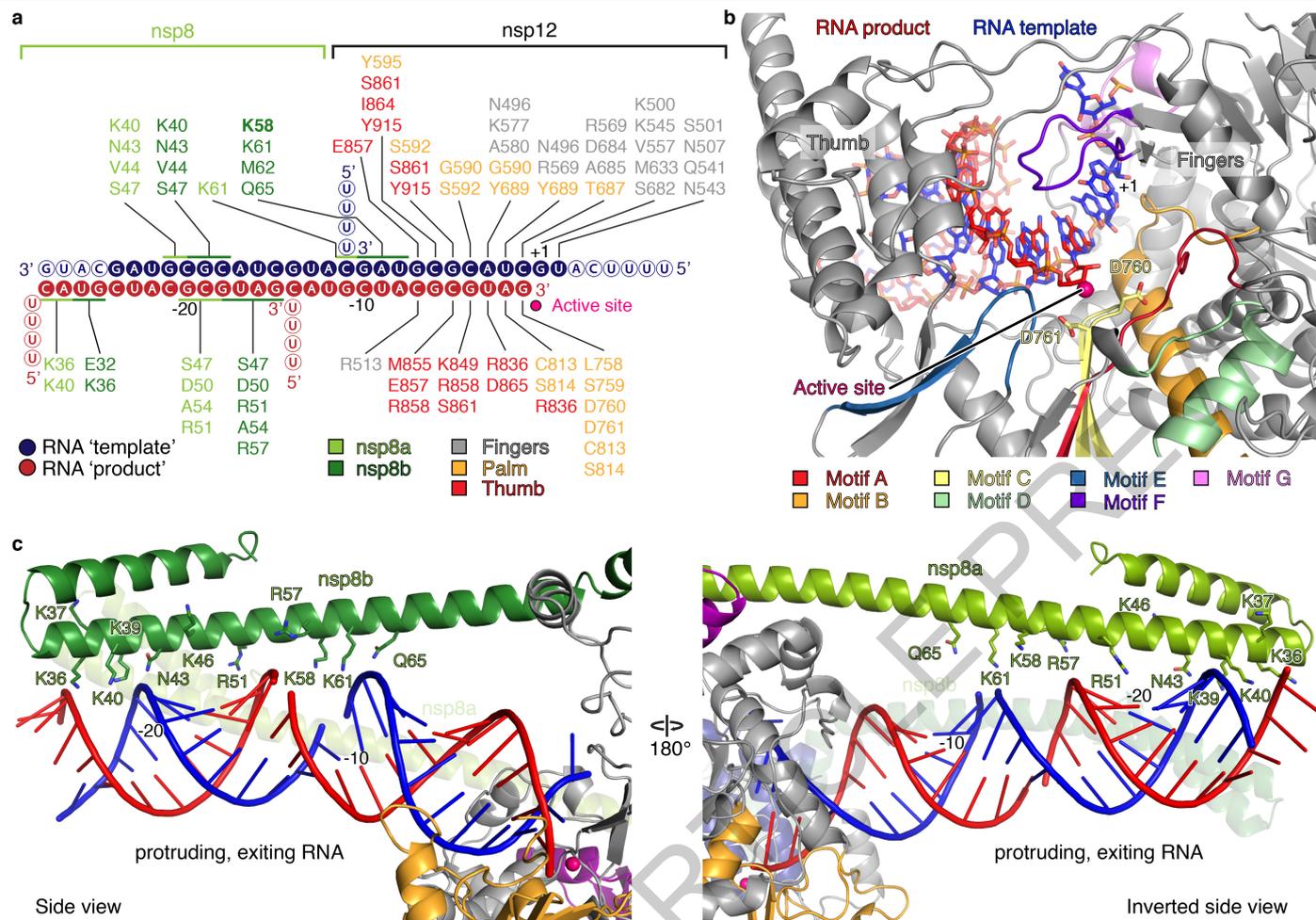


Fig. 3 | RdRp-RNA interactions. **a.** Schematic of protein-RNA interactions. Solid and hollow circles show nucleotides that were included in the structure or invisible, respectively. RNA is assembled from overlapping oligos (Extended Data Figure 1c). RdRp residues in nsp12 within 4 Å of RNA are indicated and contacts depicted with lines. Nsp8 residues and their putative RNA contact

regions (horizontal lines, within 5 Å) are indicated even if side chain density lacks. **b.** Interactions of RdRp active site with the first turn of RNA. Subunit nsp12 is in grey and conserved motifs A-G are colored. Active site residues D760 and D761 shown as sticks. The magenta sphere depicts a modeled²¹ metal ion. **c.** Charged nsp8 residues that may interact with proximal RNA regions.

Article

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Cloning and protein expression

The SARS-CoV-2 nsp12 gene was codon optimized for expression in insect cells. The SARS-CoV-2 nsp8 and nsp7 genes were codon optimized for expression in *Escherichia coli*. Synthesis of genes was performed by GeneArt (ThermoFischer Scientific GENEART GmbH, Regensburg, Deutschland). The gene synthesis products of the respective genes were PCR amplified with ligation-independent cloning (LIC) compatible primer pairs (nsp12: Forward primer: 5'- TACTTC CAA TCC AAT GCA TCT GCT GAC GCT CAG TCC TTC CTG-3'; reverse primer: 5'- TTA TCC ACT TCC AAT GTT ATT ATT GCA GCA CGG TGT GAG GGG-3'; nsp8: Forward primer: 5'- TAC TTC CAA TCC AAT GCA GCA ATT GCA AGC GAA TTT AGC AGC CTG-3', reverse primer: 5'- TTA TCC ACT TCC AAT GTT ATT ACT GCA GTT TAA CTG CGC TAT TTG CAC G-3'; nsp7: Forward primer: 5'- TAC TTC CAA TCC AAT GCA AGC AAA ATG TCC GAT GTT AAA TGC ACC AGC-3', reverse primer: 5'- TTA TCC ACT TCC AAT GTT ATT ACT GCA GGG TTG CAC GAT TAT CCA GC-3'). The PCR products for nsp8 and nsp7 were individually cloned into the pET derived vector 14-B (a gift from S. Gradia, UC Berkeley, Addgene: 48308). The two constructs for nsp8 and nsp7 contain an N-terminal 6xHis tag and a tobacco etch virus protease cleavage site. The PCR product containing codon optimized nsp12 was cloned into the modified pFastBac vector 438-C (a gift from S. Gradia, UC Berkeley, Addgene: 55220) via LIC. The nsp12 construct contained an N-terminal 6xHis tag, followed by an MBP tag, a 10xAsp sequence, and a tobacco etch virus protease cleavage site. All constructs were verified by sequencing.

The SARS-CoV-2 nsp12 plasmid (500 ng) was transformed into DH10EMBacY cells using electroporation to generate a bacmid encoding full-length nsp12. Virus production and expression in insect cells was then performed as described²⁷. Insect cell lines were obtained from Expression Systems (catalogue number: 94-002F and 94-003F) or ThermoFisher (catalogue number: 12659017). Cell lines were not authenticated. No commonly misidentified cell lines were used. After 60 hours of expression in Hi5 cells, cells were collected by centrifugation and resuspended in lysis buffer (300 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 3 mM MgCl₂, 5 mM β-mercaptoethanol, 0.284 μg ml⁻¹ leupeptin, 1.37 μg ml⁻¹ pepstatin, 0.17 mg ml⁻¹ PMSF, and 0.33 mg ml⁻¹ benzamidine). The SARS-CoV-2 nsp8 and nsp7 plasmids were overexpressed in *E. coli* BL21 (DE3) RIL cells grown in LB medium. Cells were grown to an OD₆₀₀ of 0.6 at 37 °C and protein expression was subsequently induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at 18 °C for 16 hours. Cells were collected by centrifugation and resuspended in lysis buffer (300 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 5 mM β-mercaptoethanol, 0.284 μg ml⁻¹ leupeptin, 1.37 μg ml⁻¹ pepstatin, 0.17 mg ml⁻¹ PMSF, and 0.33 mg ml⁻¹ benzamidine).

Protein purification

Protein purifications were performed at 4 °C. After harvest and resuspension, cells of the SARS-CoV-2 nsp12 expression were immediately sonicated for cell lysis. Lysates were subsequently cleared by centrifugation (87,207g, 4 °C, 30 min) and ultracentrifugation (235,000g, 4 °C, 60 min). The supernatant containing nsp12 was filtered using a 5-μm syringe filter, followed by filtration with a 0.8-μm syringe filter (Millipore) and applied onto a HisTrap HP 5 mL (GE Healthcare), preequilibrated in lysis buffer (300 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 3 mM MgCl₂, 5 mM β-mercaptoethanol, 0.284 μg ml⁻¹ leupeptin, 1.37 μg ml⁻¹ pepstatin, 0.17 mg ml⁻¹ PMSF, and 0.33 mg ml⁻¹ benzamidine). After application of the sample, the column was washed with 6 CV high salt buffer (1000 mM NaCl,

50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 3 mM MgCl₂, 5 mM β-mercaptoethanol, 0.284 μg ml⁻¹ leupeptin, 1.37 μg ml⁻¹ pepstatin, 0.17 mg ml⁻¹ PMSF, and 0.33 mg ml⁻¹ benzamidine), and 6 CV lysis buffer. The HisTrap was then attached to an XK column 16/20 (GE Healthcare), prepacked with amylose resin (New England Biolabs), which was pre-equilibrated in lysis buffer. The protein was eluted from the HisTrap column directly onto the amylose column using nickel elution buffer (300 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 500 mM imidazole pH 8.0, 3 mM MgCl₂, 5 mM β-mercaptoethanol). The HisTrap column was then removed and the amylose column was washed with 10 CV of lysis buffer. Protein was eluted from the amylose column using amylose elution buffer (300 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 116.9 mM maltose, 30 mM imidazole pH 8.0, 5 mM β-mercaptoethanol). Peak fractions were assessed via SDS-PAGE and staining with Coomassie. Peak fractions containing nsp12 were pooled and mixed with 8 mg of His-tagged TEV protease (~80% (w/w)). After 12 hours of protease digestion at 4 °C, protein was applied to a HisTrap column equilibrated in lysis buffer to remove uncleaved nsp12, His6-MBP, and TEV. Subsequently, the flow-through containing nsp12 was applied to a HiTrap Heparin 5 column mL (GE Healthcare). The flow-through containing nsp12 was collected and concentrated in a MWCO 50,000 Amicon Ultra Centrifugal Filter unit (Merck). The concentrated sample was applied to a HiLoad S200 16/600 pg equilibrated in size exclusion buffer (300 mM NaCl, 20 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 1 mM MgCl₂, 1 mM TCEP). Peak fractions were assessed by SDS-PAGE and Coomassie staining. Peak fractions were pooled and concentrated in a MWCO 50,000 Amicon Ultra Centrifugal Filter (Merck). The concentrated protein with a final concentration of 102 μM was aliquoted, flash-frozen, and stored at -80 °C until use.

SARS-CoV-nsp8 and nsp7 were purified separately using the same purification procedure, as follows. After cell harvest and resuspension in lysis buffer, the protein of interest was immediately sonicated. Lysates were subsequently cleared by centrifugation (87,200g, 4 °C, 30 min). The supernatant was applied to a HisTrap HP 5 mL column (GE Healthcare), preequilibrated in lysis buffer. The column was washed with 9.5 CV high salt buffer (1000 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 5 mM β-mercaptoethanol, 0.284 μg ml⁻¹ leupeptin, 1.37 μg ml⁻¹ pepstatin, 0.17 mg ml⁻¹ PMSF, and 0.33 mg ml⁻¹ benzamidine), and 9.5 CV low salt buffer (150 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 5 mM β-mercaptoethanol). The sample was then eluted using nickel elution buffer (150 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 500 mM imidazole pH 8.0, 5 mM β-mercaptoethanol). The eluted protein was dialyzed in dialysis buffer (150 mM NaCl, 50 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 5 mM β-mercaptoethanol) in the presence of 2 mg His-tagged TEV protease (nsp7: ~10% (w/w), nsp8: ~6% (w/w)) at 4 °C. After 12 hours, imidazole pH 8.0 was added to a final concentration of 30 mM. The dialyzed sample was subsequently applied to a HisTrap HP 5 mL column (GE Healthcare), preequilibrated in dialysis buffer. The flow-through that contained the protein of interest was then applied to a HiTrap Q 5 mL column (GE Healthcare). The Q column flow-through containing nsp8 or nsp7 was concentrated using a MWCO 3,000 Amicon Ultra Centrifugal Filter (Merck) and applied to a HiLoad S200 16/600 pg equilibrated in size exclusion buffer (150 mM NaCl, 20 mM Na-HEPES pH 7.4, 5% (v/v) glycerol, 1 mM TCEP). Peak fractions were assessed by SDS-PAGE and Coomassie staining. Peak fractions were pooled. Nsp7 with a final concentration of 418 μM was aliquoted, flash-frozen, and stored at -80 °C until use. Nsp8 with a final concentration of 250 μM was aliquoted, flash-frozen, and stored at -80 °C until use. All protein identities were confirmed by mass spectrometry.

RNA extension assays

All RNA oligos were purchased from Integrated DNA Technologies. The RNA sequence used for the transcription assay is /56-FAM/rUrUrUrUrCrArUrGrCrUrArCrGrCrGrUrArGrUrUrUrUrCrUrArCrGrCrG. We

designed a minimal substrate by connecting the template RNA to the RNA primer by a tetraloop, to protect the blunt ends of the RNA duplex and to ensure efficient annealing. RNA was annealed in 50 mM NaCl and 10 mM Na-HEPES pH 7.5 by heating the solution to 75 °C and gradually cooling to 4 °C. RNA extension reactions contained RNA (5 μM), nsp12 (5 μM), nsp8 (15 μM) and nsp7 (15 μM) in 100 mM NaCl, 20 mM Na-HEPES pH 7.5, 5% (v/v) glycerol, 10 mM MgCl₂ and 5 mM β-mercaptoethanol. Reactions were incubated at 37 °C for 5 min and the RNA extension was initiated by addition of NTPs (150 μM UTP, GTP and CTP, and 300 μM ATP). Reactions were stopped by the addition of 2× stop buffer (7M urea, 50 mM EDTA pH 8.0, 1x TBE buffer). Samples were digested with proteinase K (New England Biolabs) and RNA products were separated on 8 cm x 8 cm 20% acrylamide gels in 1X TBE buffer supplemented with 8M urea. 6-FAM labeled RNA products were visualized by Typhoon 95000 FLA Imager (GE Healthcare Life Sciences).

Cryo-EM sample preparation and data collection

An RNA scaffold for RdRP-RNA complex formation was annealed by mixing equimolar amounts of two RNA strands (5'-rUrUrU rUrCrA rUrGrC rUrArC rGrCrG rUrArG-3'; 56-FAM/rCrUrA rCrGrC rG-3') (IDT Technologies) in annealing buffer (10 mM Na-HEPES pH 7.4, 50 mM NaCl) and heating to 75 °C, followed by step-wise cooling to 4 °C. For complex formation, 1.2 nmol of purified nsp12 was mixed with a 1.2-fold molar excess of RNA scaffold and 6-fold molar excess of each nsp8 and nsp7. After incubation at room temperature for 10 min, the EC was subjected to size exclusion chromatography on a Superdex 200 Increase 3.2/300 equilibrated with complex buffer (20 mM Na-HEPES pH 7.4, 100 mM NaCl, 1 mM MgCl₂, 1 mM TCEP). Peak fractions with a volume of 100 μL (absorbance at 280 nm of 0.155 AU, 2 mm path length) corresponding to a nucleic-acid rich high-molecular weight population (as judged by absorbance at 260 nm) were pooled and concentrated in a MWCO 30,000 Vivaspin 500 concentrator (Sartorius) to approx. 20 μL. 3 μL of RdRp-RNA complex were mixed with 0.5 μL of octyl β-D-glucopyranoside (0.003% final concentration) and applied to freshly glow discharged R 2/1 holey carbon grids (Quantifoil). Prior to flash freezing in liquid ethane, the grid was blotted for 6 seconds with a blot force of 5 using a Vitrobot MarkIV (Thermo Fisher Scientific) at 4 °C and 100% humidity.

Cryo-EM data collection was performed with SerialEM²⁸ using a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operated at 300 keV. Images were acquired in EFTM mode with a slit width of 20 eV using a GIF quantum energy filter and a K3 direct electron detector (Gatan) at a nominal magnification of 105,000x corresponding to a calibrated pixel size of 0.834 Å/pixel. Exposures were recorded in counting mode for 2.2 seconds with a dose rate of 19 e⁻/px/s resulting in a total dose of 60 e⁻/Å² that was fractionated into 80 movie frames. Because initial processing showed that the particles adopted only a limited number of orientations in the vitreous ice layer, a total of 8168 movies were collected at 30° stage tilt to yield a broader distribution of orientations. Untilted data was discarded. Motion correction, CTF-estimation, particle picking and extraction were performed on the fly using Warp²⁹.

Cryo-EM data processing and analysis

1.3 million particles were exported from Warp²⁹ to cryoSPARC³⁰, and the particles were subjected to 2D classification. 25% of the particles were selected from classes deemed to represent the polymerase, and refined against a synthetic reference prepared from PDB-6M71. *Ab initio* refinement was performed using particles from bad 2D classes to obtain five 3D classes of 'junk'. These five classes and the first polymerase reconstruction were used as starting references to sort the initial 1.3M particles in supervised 3D classification rather than 2D, as the latter tended to exclude less abundant projection directions. 514k particles (39% from the resulting polymerase class) were subjected to another *ab initio* refinement to obtain five starting references containing four

'junk' classes and the complex of interest. These classes were used as starting references in another supervised 3D classification. 418k particles (82%) from the class representing the complex were exported from cryoSPARC to RELION 3.0³¹. There, all particles were refined in 3D against the reconstruction previously obtained in cryoSPARC, using a mask including only the core part of the polymerase and a short segment of upstream RNA to obtain a 3.1 Å reconstruction. CTF refinement and another round of 3D refinement improved the resolution further to 2.9 Å (Extended Data Figure 2a-c, Map 1). Particles were re-extracted at 1.3 Å/px in a bigger box in Warp to accommodate distant parts of the RNA. Unsupervised 3D classification with local alignment was performed to obtain 2 classes: with nsp8b present, and without. 172k particles with nsp8b present were finally subjected to global (Extended Data Figure 2a-c, Map 2) and focused 3D refinement using a mask including the RNA, nsp8a and nsp8b (Extended Data Figure 2a-c, Map 3).

Model building and refinement

To build the atomic model of the RdRp-RNA complex, we started from the structure of the free SARS-CoV-2 RdRp (PDB-6M71) that was recently adjusted by Tristan Croll (Cambridge University, UK; <https://twitter.com/CrollTristan/status/1247846163061133312>). The structure was rigid-body fit into the cryo-EM reconstruction and adjusted manually in Coot³². Unmodeled density remained for helical segments in the N-terminal regions of both copies of nsp8. These nsp8 extensions were modeled by superimposing the nsp8 model (PDB-2AHM; chain H) from the crystal structure of the nsp7-nsp8 hexadecamer¹⁷, in which the far N-terminal region of nsp8 adopts the same fold. Nsp8a (chain B) showed weaker density than nsp8b (chain D), but the register was faithfully determined by superimposing well-resolved parts (residues 80-97). The most N-terminal helices in nsp8a and nsp8b (residues 6-31) were only visible after low pass filtering of maps to the local resolution of 6 Å and were modeled by superposition of the crystal structure of nsp8 (PDB-2AHM; chain H) with residues 33-55, which positioned these helices within the density in the low-pass filtered map. Sidechains for residues 6-31 were subbed. Careful inspection of the remaining A-form RNA density revealed that in our complex, instead of the originally designed short template-primer duplex (see above), four copies of one of the RNA oligos were annealed to form a pseudo-continuous long RNA duplex. Annealing was mediated by a 10 bp self-complementary region within this RNA oligo (Extended Data Figure 1c). Nucleotides 5-18 of four RNA strands were modeled, whereas the flapped-out nucleotides 1-4 were invisible and excluded. The model was real-space refined using phenix.refine³³ against a composite map of the focused refinement (Maps 1 and 3) and global reconstructions (Map 2) generated in phenix.combine_focused_maps and shows excellent stereochemistry (Extended Data Table 1). Figures were prepared with PyMol and Chimera³⁴.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The electron potential reconstructions and structure coordinates were deposited with the Electron Microscopy Database (EMDB) under accession code EMD-11007 and with the Protein Data Bank (PDB) under accession code 6YYT. We additionally deposited the raw cryo-EM data to the EMPIAR data base under accession code EMPIAR-10409.

- Vos, S. M. et al. Architecture and RNA binding of the human negative elongation factor. *Elife* **5**, <https://doi.org/10.7554/eLife.14981> (2016).
- Mastrorade, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51, <https://doi.org/10.1016/j.jsb.2005.07.007> (2005).
- Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152, <https://doi.org/10.1038/s41592-019-0580-y> (2019).

Article

30. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, <https://doi.org/10.1038/nmeth.4169> (2017).
31. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5-17, <https://doi.org/10.1107/S205225251801463X> (2019).
32. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, <https://doi.org/10.1107/S0907444910007493> (2010).
33. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, <https://doi.org/10.1107/S2059798318006551> (2018).
34. Huang, C. C., Meng, E. C., Morris, J. H., Pettersen, E. F. & Ferrin, T. E. Enhancing UCSF Chimera through web services. *Nucleic Acids Res* **42**, W478-484, <https://doi.org/10.1093/nar/gku377> (2014).

Acknowledgements We thank all members of the Department of Molecular Biology at the Max-Planck-Institute for Biophysical Chemistry for support. We thank Jana Schmitzova for

helpful discussions and advice. H.S.H. was supported by the Deutsche Forschungsgemeinschaft (FOR2848). P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860, SPP2191), Germany's Excellence Strategy (EXC 2067/1- 390729940), the ERC Advanced Investigator Grant TRANSREGULON (grant agreement No 693023) and the Volkswagen Foundation. We thank Henning Urlaub (Max Planck Institute for Biophysical Chemistry) for mass spectrometry.

Author contributions H.S.H., G.K., L.F., C.D. and D.T. designed and carried out experiments and data analysis. P.C. designed and supervised research. All authors interpreted data and wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information

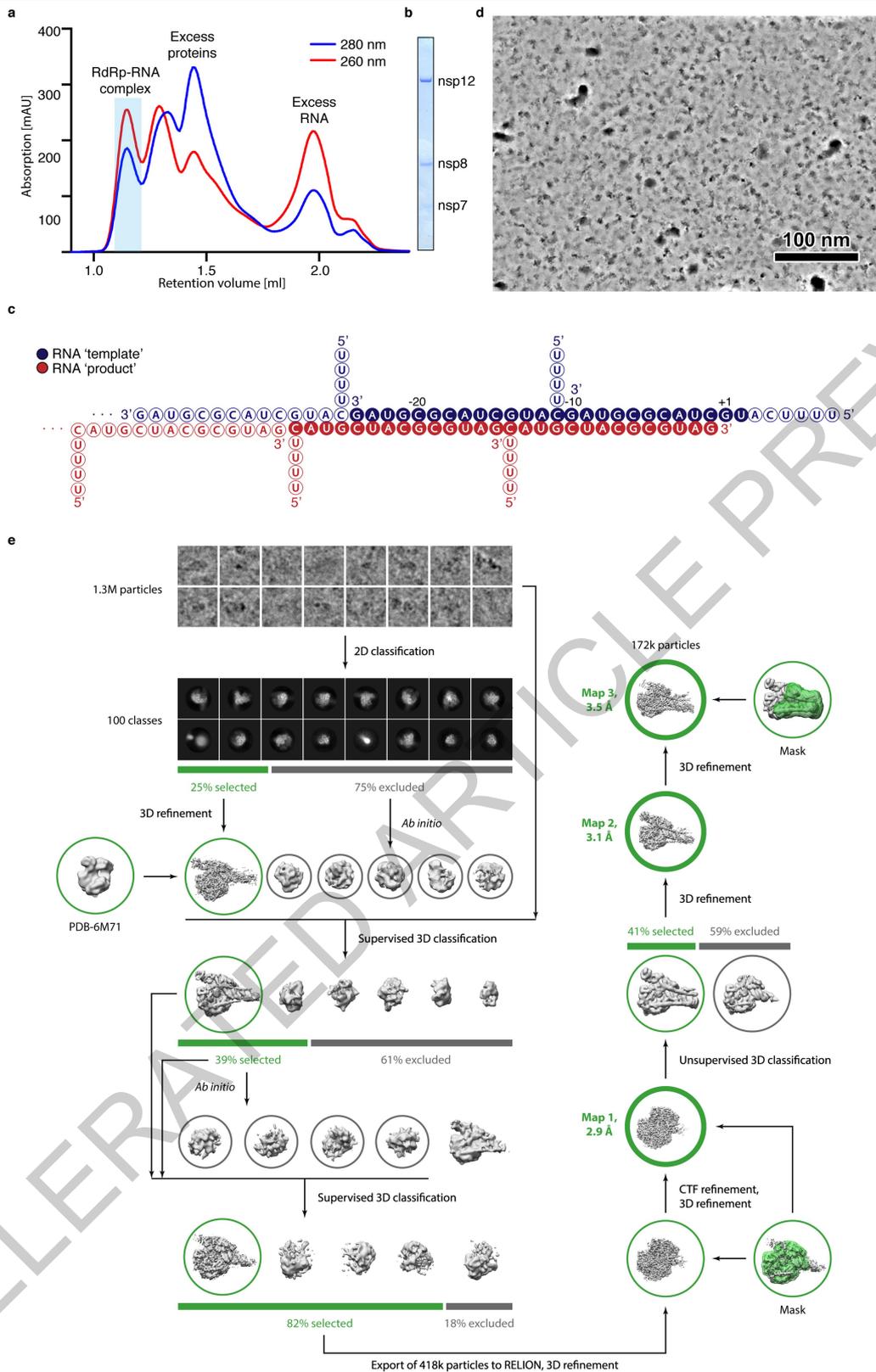
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2368-8>.

Correspondence and requests for materials should be addressed to P.C.

Peer review information *Nature* thanks Bruno Canard and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

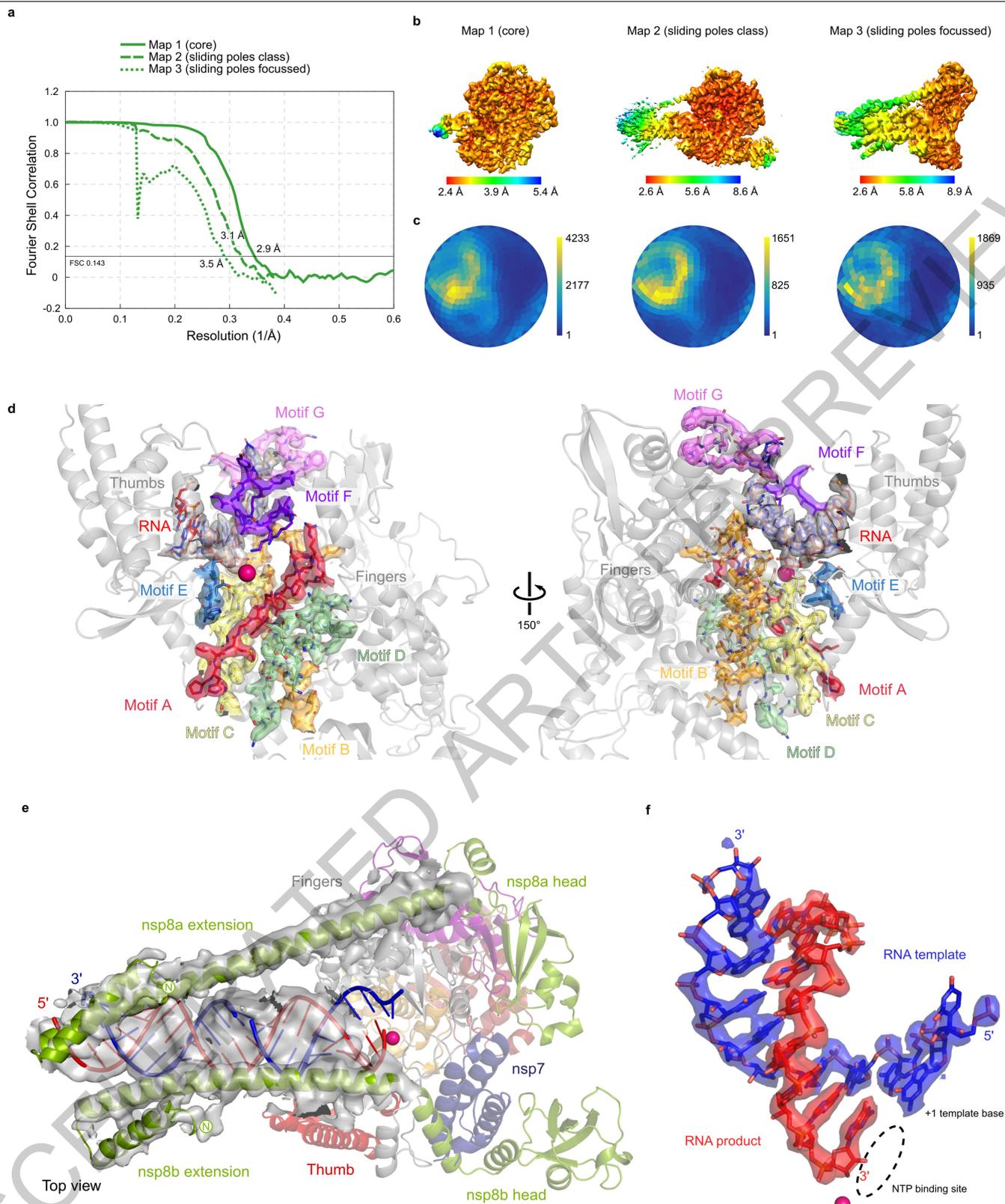
ACCELERATED ARTICLE PREVIEW



Extended Data Fig. 1 | Cryo-EM analysis. Related to Figures 1, 2. a.

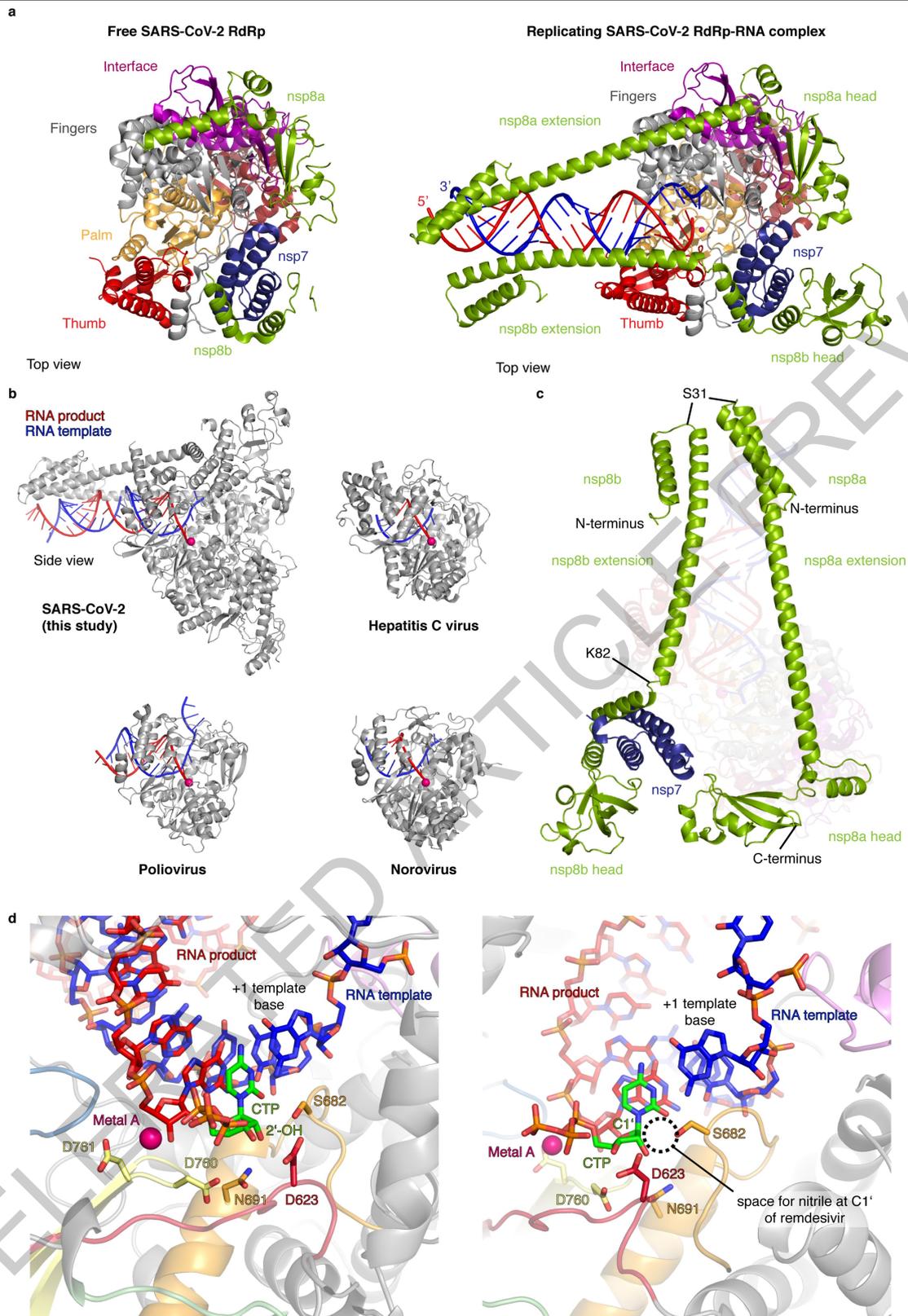
Purification of RdRp-RNA complex by size exclusion chromatography. The peak used for structural analysis is highlighted in blue. The experiment was performed once. **b.** Exemplary SDS-PAGE analysis of purified complex with RdRp subunits labeled. The experiment was performed once. **c.** RNA duplex scaffold formed by oligomerization of a short pseudo-palindromic RNA.

The depicted base pairing gave rise to a pseudo-continuous A-form duplex. Solid and hollow circles show RNA nucleotides that were included in the structure or not visible, respectively. **d.** Example denoised micrograph calculated from two independently measured half sets of 40 frames each. Scale bar, 100 nm. **e.** Cryo-EM processing tree.



Extended Data Fig. 2 | Cryo-EM reconstructions. Related to Figure 2.
a, Fourier shell correlation (FSC) plots for reported reconstructions and resolution estimates. **b**, Local resolution distribution for the reported reconstructions. **c**, Angular distribution plots. Scale shows the number of particles assigned to a particular angular bin. Blue, a low number of particles; yellow, a high number of particles. **d**, Cryo-EM map for the RdRp active center

region including elements with sequence motifs A-G. The active site is indicated by a magenta sphere. **e**, Cryo-EM map for the RNA duplex and the nsp8 extensions. The active site is indicated by a magenta sphere. **f**, Cryo-EM map for the RNA duplex in the active site. The active site is indicated with a magenta sphere, and the NTP site with a dashed circle.



Extended Data Fig. 3 | Structural comparisons. Related to Figures 2, 3.

a. Comparison of the free SARS-CoV-2 RdRp structure¹⁵ (left) and the replicating RdRp-RNA complex (right, this study). Color code as in Fig. 2. **b.** Similar location and orientation of the RNA template-product duplex in RdRp complexes of SARS-CoV-2 virus (top left, this study), hepatitis C virus¹⁹ (top right), poliovirus²⁰ (bottom left) and norovirus²¹ (bottom right). Structures are shown as ribbon models with RNA template and product strands in blue and red, respectively. An active site metal ion is shown as a magenta sphere. Side view as defined in Fig. 2. **c.** The two nsp8 copies adopt different folds within the RdRp-RNA complex. Top view of as in (a) but rotated by 90°. Nsp8 and nsp7 are

opaque and nsp12 transparent. The nsp8b extension has a kink around residue 82, which may be stabilized by nsp7 and nsp12 elements. **d.** Model of substrate nucleoside triphosphate (NTP) in the RdRp active site. A CTP substrate was placed after superposition of the norovirus RdRp-nucleic acid complex structure²¹. Coloring as in Fig. 3b. Active site residues D760, D761, N691, S682 and D623 are shown as sticks, and the modeled active site metal ion is shown as a magenta sphere. When the nucleoside triphosphate form of remdesivir would bind in the NTP site, the nitrile group connected to the ribose C1' position would be accommodated in the space indicated by the dashed circle.

Cryo-EM data collection, refinement and validation statistics

SARS-CoV-2 RdRp-RNA complex (PDB 6YYT)				
	Map 1 core (EMD-11007)	Map 2 sliding poles class (EMD-11007)	Map 3 sliding poles focused (EMD-11007)	Composite map (EMD-11007)
Data collection and processing				
Magnification	105,000 x	105,000 x	105,000 x	
Voltage (kV)	300	300	300	
Electron exposure (e-/Å ²)	60	60	60	
Defocus range (µm)	0.5 – 1.5	0.5 – 1.5	0.5 – 1.5	
Pixel size (Å)	0.834	0.834	0.834	
Symmetry imposed	C1	C1	C1	
Initial particle images (no.)	1,300,000	1,300,000	1,300,000	
Processing pixel size (Å)	0.834	1.3	1.3	
Final particle images (no.)	418,000	171,400	171,400	
Map resolution (Å)	2.9	3.2	3.5	
FSC threshold	0.143	0.143	0.143	
Map resolution range (Å)	2.4 – 5.4	2.6 – 8.6	2.6 – 8.9	
Map sharpening <i>B</i> factor (Å ²)	-93	-60	-95	
Refinement				
Initial model used (PDB code)				6M71
Model resolution (Å)				3.2
FSC threshold				0.5
Model resolution range (Å)				2.4 – 8.9
Model composition				
Non-hydrogen atoms				11,375
Protein residues				1,298
Ligands				54
<i>B</i> factors (Å ²)				
Protein				45.6
Ligand				53.9
R.m.s. deviations				
Bond lengths (Å)				0.004
Bond angles (°)				0.875
Validation				
MolProbity score				1.23
Clashscore				2.82
Poor rotamers (%)				0.36
Ramachandran plot				
Favored (%)				97.12
Allowed (%)				2.88
Disallowed (%)				0.00

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Serial EM 3.8 beta 8

Data analysis

RELION 3.0 beta-2, UCSF Chimera 1.13, UCSF ChimeraX v0.8, Pymol 2.2.2, Coot 0.9, Warp v1.0.7, PHENIX 1.18, cryoSPARC 2.14.2

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The electron density reconstructions and structure coordinates were deposited with the Electron Microscopy Database (EMDB) under accession code EMD-11007 and with the Protein Data Bank (PDB) under accession code 6YYT.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size.
Data exclusions	No data were excluded from the analyses.
Replication	All attempts at replication were successful. Cryo-EM single particle analysis inherently relies on averaging over a large number of independent observations.
Randomization	Samples were not allocated to groups.
Blinding	Investigators were not blinded during data acquisition and analysis because it is not a common procedure for the methods employed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Hi5 cells: Expression Systems, Tni Insect cells in ESF921 media, item 94-002F Sf9 cells: ThermoFisher, Catalogue Number 12659017, Sf9 cells in Sf-9000TM III SFM Sf21 cells: Expression Systems, SF21 insect cells in ESF921 medium, Item 94-003F
Authentication	None of the cell lines were authenticated.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.