Nonstructural protein 7 and 8 complexes of SARS-CoV-2

Changhui Zhang

Sichuan University

Li Li

Sichuan University

Jun He

Chinese Academy of Sciences

Cheng Chen

Tianjin University

Dan Su (✉ sudan@scu.edu.cn)

Sichuan University

Research Article

Keywords: SARS-CoV-2, COVID-19, nsp7, nsp8, nsp7-8 complex

Posted Date: October 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-96734/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Protein Science on February 25th, 2021.
See the published version at <https://doi.org/10.1002/pro.4046>.

Abstract

The pandemic outbreak of coronavirus disease 2019 (COVID-19) across the world has led to millions of infection cases and caused a global public health crisis. Current research suggests that SARS-CoV-2 is a highly contagious coronavirus that spreads rapidly through communities. To understand the mechanisms of viral replication, it is imperative to observe coronavirus viral replicase, a huge protein complex comprising up to 16 viral nonstructural and associated host proteins, which is the most promising antiviral target for inhibiting viral genome replication and transcription. Recently, several components of the viral replicase complex in SARS-CoV-2 have been solved to provide a basis for the design of new antiviral therapeutics. Here, we report the crystal structure of the SARS-CoV2 nsp7-8 tetramer, which comprises two copies of each protein representing nsp7's full-length and the C-terminus of nsp8 owing to N-terminus proteolysis during the process of crystallization. We also identified a long helical extension and highly flexible N-terminal domain of nsp8, which is preferred for interacting with single-stranded nucleic acids.

Introduction

In early 2020, the COVID-19 virus swept the world, creating a global pandemic greatest in risk since the influenza pandemic of 1918¹. The virus has been identified as a new coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is an enveloped, positive-strand RNA (+RNA) virus belonging to the family, *Coronaviridae*, order *Nidovirales*, and realm *Riboviria*, with a large genome of approximately 29.8 kb in length²⁻⁵. Sequence analysis suggests that the SARS-CoV-2 genome is highly identical to that of bat coronavirus (BatCoV), which originated from infected bats^{4,6}. The SARS-CoV-2 genome has been predicted to contain 14 functional open-reading frames (ORFs), including ORF1ab, which occupies approximately two-thirds of the 5' end of this genome and encodes two polyproteins (pp1a and pp1ab) using a programmed ribosomal frameshift mechanism caused by a slippery sequence⁷. These two polyproteins contain two proteases, a papain-like proteinase (PLpro, nsp3) and 3C-like proteinase (3CLpro, nsp5), which autoproteolytically processes the polyproteins into 16 nonstructural proteins (nsp1-16)⁸⁻¹¹. These nsps are believed to comprise the major components of the viral replication/transcription complex, a highly dynamic protein-RNA complex that facilitates viral replication and transcription. Most of these nsps are well-known types of enzymes, including a papain-like protease 2 (PL2^{pro}, nsp3)¹², a 3C-like serine protease (M^{pro}, nsp5)¹³, an RNA-dependent RNA polymerase (RdRp, nsp12), an RNA helicase (nsp13), a guanine-N7 methyltransferase (nsp14), an endoribonuclease (nsp15), and 2'O-methyl-transferase (2'O-MTase, nsp16)¹⁴⁻¹⁷. In addition, the replicase gene encodes several small associated proteins, including nsp1 and nsp7-nsp11¹⁵. Proteins nsp7-nsp10 are highly conserved among all coronaviruses (CoVs), *but have no functional homologs outside of coronaviridae*¹⁸. After the emergence of SARS-CoV in 2003, there was an intense effort to characterize the complex structure of such nonstructural proteins¹⁹. Subsequently, the complex structures of nsp7-nsp8, nsp10-nsp14, nsp10-nsp16, and nsp12-nsp7-nsp8 of SARS-CoV have been determined²⁰⁻²³. Very recently, several independent

groups have reported the complex structure of nsp12 RNA-dependent RNA polymerase with cofactors nsp7 and nsp8 from SARS-CoV-2 using cryo-electron microscopy (cryo-EM)²⁴⁻²⁶.

The hexadecameric structure of the nsp7 and nsp8 complex has been solved in SARS-CoV²⁰. The complex resembles a hollow cylinder with a central channel, which can encircle and bind RNA through a strong positive charge in the inner channel as exhibited in biochemical assays. Functional studies have revealed that nsp8 prefers the 5- $\left(\frac{G}{U}\right)\text{C}$ - 3 trinucleotides on the RNA template to initiate the synthesis of complementary short oligonucleotides, and nsp8 is proposed to provide RNA primers required by nsp12 RdRp during replication and transcription²⁷. Deletion of the nsp7 or nsp8 region results in impaired RNA synthesis and a lethal phenotype in MHV²⁸. A report described that nsp8 co-immunoprecipitates with nsp12, nsp9, nsp5, and helicase nsp13, and it also colocalizes with nsp7, nsp9, and nsp10^{19,29}. Therefore, nsp7 and nsp8 seem to play an important role in the viral replication complex.

Herein, we determined the complex structure of nsp7 and nsp8 from SARS-CoV-2 using X-ray crystallography. The 2.5 Å resolution structure is composed of a tetrameric complex of nsp7 and the C-terminus of nsp8. The core architecture of the nsp7-8 tetramer is via composed of two α-helix bundles, which is highly conserved compared with the nsp7-8 hexadecameric structure from SARS-CoV and the cryo-EM structure of the nsp7-8-12 complex from SARS-CoV-2. We also confirmed that the N-terminus of nsp8 is highly flexible with a positive electrostatic surface involved in nucleic acid interactions and complex formation with nsp7 and nsp12.

Results

Overall structure of the SARS-CoV-2 nsp7-8 dimer

The full-length recombinant proteins, GST-tagged nsp7 (residues 1–84) and His₆-tagged nsp8 (residues 1–199) from SARS-CoV-2 were expressed in *Escherichia coli* (E. coli) BL21 (DE3). The nsp7-8 complex was prepared for crystallization with a stoichiometry ratio of 1:1 and further purified by size-exclusion chromatography (SEC). The complex protein was screened for crystallization conditions, and a single crystal was grown for weeks into a proper size for X-ray diffraction data collection (Supplementary Figure S1d-e). The diffraction data were processed to a 2.5 Å resolution with the HKL2000 package³⁰. The final nsp7-8 complex structure was resolved by molecular replacement in the P1 space group packing four nsp7-8 complexes into a unit cell. The N-terminus of nsp8 (residues Ala¹-Ser⁷⁷) could not be traced in the electron density map. The degradation of the nsp8 N-terminus during the process of crystallization was identified by dissolving crystal in water and verified on SDS-PAGE gel (Supplementary Fig. S1a-c).

Each nsp7-8 complex structure exhibits a heterodimer consisting of one copy of nsp7 (residues Met³–Asp⁷⁵) and one copy of the nsp8 C-terminus (residues Glu⁷⁷–Asn¹⁹², nsp8C) (Figure 1a). The nsp7 structure is composed of four α-helices forming an antiparallel helix bundle with helices a1 (residues

Ser⁴-Gln¹⁹), $\alpha 2$ (residues Leu²⁸-Leu⁴¹), $\alpha 3$ (residues Thr⁴⁵-Ser⁶¹), and $\alpha 4$ (residues Ile⁶⁸-Met⁷⁵) (Fig. 1b). The four nsp7 monomers in the unit cell superimpose well (RMSD deviation < 0.5 Å). A previous study reported that nsp7 in feline coronavirus adopts two different conformations in complex with nsp8. As seen in Supplementary Figure S3b and c, the helix bundle conformation of nsp7 is highly conserved in SARS-CoV-2 replicase complexes, including within the structures of the nsp12-7-8 heterodimer (PDB: 7C2K) and nsp12-13-7-8 complex (PDB: 6XEZ). We also compared our structure for SARS-CoV-2 nsp7 within the dimeric complex to the three structures of SARS-CoV nsp7 with coordinates available in the PDB (2AHM, 6NUR, and 5F22). We noticed that the structure of SARS-CoV-2 nsp7 is highly conserved with that of SARS-CoV nsp7 in complex with nsp8 (Supplementary Fig. S3 d-f). The RMS deviation for 62 C α atoms between the two molecules (SARS-CoV-2 nsp7 and SARS-CoV nsp7) is 0.583 Å. These data indicate that the structure of SARS-CoV-2 nsp7 is highly conserved with nsp7s of other CoVs, especially with the structure of SARS-CoV nsp7, likely based on the high sequence identity between these CoVs (Supplementary Fig. S1f).

SARS-CoV nsp8 adopts two different conformations in the hexadecameric complex with nsp7. However, only one conformation of nsp8 was observed in our SARS-CoV-2 nsp7-8 complex. The nsp8 in our structure is incomplete owing to the degradation of the N-terminus protein (residues Ala¹-Ser⁷⁷) during crystallization. Therefore, we named nsp8C for an accurate description. nsp8C is composed of four α -helices ($\alpha 1$, residues Asp⁷⁸-Arg⁹⁶; $\alpha 2$, residues Asp¹⁰¹-Asn¹⁰⁹; $\alpha 3$, residues Tyr¹³⁵-Lys¹³⁹; $\alpha 4$, residues Lys¹⁶⁹-Glu¹⁷¹; $\alpha 5$, residues Ser¹⁷⁷-Asn¹⁷⁹) packed against a four-stranded antiparallel β -sheet ($\beta 1$, residues Leu¹²⁸-Ile¹³²; $\beta 2$, residues Thr¹⁴⁶-Tyr¹⁴⁹; $\beta 3$, residues Ala¹⁵²-Val¹⁶⁰; $\beta 4$, residues Leu¹⁸⁴-Leu¹⁸⁹) is similar to the “golf club” nsp8 structures from SARS-CoV and FIPV (Fig. 1c). Superimposition of the SARS-CoV-2 nsp8C onto the SARS-CoV nsp8 (PDB: 2AHM) and nsp8 in feline coronavirus (FCoV) nsp7-8 complex (PDB:3UB0) indicates that the C-terminal domains of nsp8 are highly conserved in coronaviruses, with RMSD values of 0.671 for 98 C α atoms and 0.668 Å for 103 C α atoms, respectively (Supplementary Fig. S3g and h). Compared with SARS, there are five residues (Y15F, V132I, G136N, N145T, N193S) in SARS-CoV-2 that were mutated, four of which are located at the C-terminus but had no significant effect on structure (Supplementary Figure S1g).

nsp7 interacts tightly with nsp8C to yield a buried surface area of 1467.4 Å² (~26.4% of the whole surface area of nsp7), forming a heterodimer. The interaction interfaces between two proteins consist of a α -helix bundle composed by five α -helices ($\alpha 1$, $\alpha 3$, and $\alpha 4$ from nsp7; $\alpha 1$ and $\alpha 2$ from nsp8C) that interact with each other, which are predominant via hydrophobic interactions, along with nine hydrogen bonds and four salt bridges. The same interface was also observed in the hexadecameric complex of SARS-CoV nsp7-8 (PDB:2AHM), dimeric complex of SARS-CoV nsp7-8C (PDB:6NUR), and FCoV nsp7-8C trimeric complex (PDB:3UB0), which indicates that the core structure of nsp7-8C is highly architecturally conserved among coronaviruses.

Structural basis for SARS-CoV-2 nsp7-8 tetramer

In order to obtain more information about the interfaces and likely biological assemblies of the nsp7-8 complex, we analyzed the crystal structure using the online programs, PISA (Protein, Interfaces, Structures and Assemblies)³¹ and EPPIC (Evolutionary Protein-Protein Interface Classifier)³². The results suggested that nsp7-8 forms a stable symmetric tetramer during crystal packing. The average total surface area buried of the tetramer was 7790 Å². The analysis of the crystal structure showed that a tetramer with a “butterfly” shape was present in the unit cell (Fig. 2a). It is formed by the association of two nsp7-8 dimers, which contact each other with the two helix bundles. According to the PISA calculation, the formation of the tetramer is mainly because of an “α-helix bands” hydrophobic core between nsp7 (α1 and α2) and nsp8 (α1). There are two hydrogen bonds between residues Met94 of nsp8 and Ser4 of nsp7. In addition, Cys8 between the two nsp7 residues forms a disulfide bond to stabilize the tetramer (Figure 2b).

There was also the question of whether the quaternary structure of the SARS-CoV-2 nsp7-8C tetramer exists in solution. First, we performed chemical cross-linking analysis. After treatment with different concentrations (0.005%, 0.01%, and 0.02%) of glutaraldehyde, isolated nsp7, nsp8, and nsp8C displayed cross-linked homopolymers in addition to their monomers (Supplementary Figure S2b-e). When the nsp7-8C complex was cross-linked by glutaraldehyde, we detected the nsp7-8C heterodimer (24 kDa) and heterotetramer (48 kDa) (Supplementary Fig. S2f). We also applied analytical SEC to determine the solution states of the nsp7, nsp8, nsp8C, nsp7-8, and nsp7-8C complexes. As shown in Supplementary Fig. S2a, the nsp7-8 and nsp7-8C complexes eluted with a retention volume of 14.75 ml and 15.12 ml, respectively, representing a tetramer with estimated molecular mass (MM) of 58 kDa and 49 kDa, respectively, which corresponded reasonably well to the MM indicated by cross-linking. nsp7 and nsp8 had retention volumes of 16.95 mL and 15.28 mL, respectively. The estimated MMs calculated from the retention volumes were also in agreement with the MMs of their cross-linked tetramer. Consequently, these data indicate that the tetrameric model of nsp7-8C is a feasible biological assembly.

Nucleic acid-binding domain located in N-terminus of nsp8

Based on our knowledge, the N-terminus of nsp8 is highly dynamic in solution and adopts diverse confirmation depending on the assembly elements in the complex. We calculated several full-length SARS-CoV-2 nsp8 models based on the structures of nsp8 from the SARS-CoV nsp7-8 hexadecamer (PDB: 2AHM). The models present an N-terminal domain containing a long α-helix or several short helices. The surface charge distribution pattern of nsp8 in SARS-CoV-2 reveals a highly positive charge region located on the N-terminus, indicating a potential nucleic acid-binding region (Figure 3a and b). Therefore, the electrophoretic mobility shift assay (EMSA) approach was employed to identify the nucleic acid binding with the wild-type nsp8, specifically two truncated versions nsp8ΔN and nsp8ΔC, respectively. The mobility of the nucleic acid reduced in the lanes of the nsp8 wild-type and nsp7-8 complex indicates that the specificity of the interaction between protein and nucleic acid, nsp8ΔC or nsp7-nsp8ΔC, binds slightly to the nucleic acid, while there is no obvious combination between free nsp7, nsp8ΔN, or nsp7-nsp8ΔN and nucleic acid (Fig. 3e and f). This implies that full-length nsp8 is needed to

Loading [MathJax]/jax/output/CommonHTML/jax.js more, Surface plasmon resonance (SPR) was performed to

investigate whether the nsp8 and nsp7-8 complexes could bind single-stranded nucleic acids. The binding curves of nucleic acids with various binding partners were therefore studied (Fig. 3c and d). The interaction of the nsp8 wild-type to ssRNA showed a K_D of 5.20 μ M while the interaction of the nsp7-8 to ssRNA was 11.5 μ M. Our data support that the N-terminus of nsp8 among coronaviruses is a nucleic acid binding motif. Without the existence of nucleic acid, the N-terminus of nsp8 is highly unstable and prefers to self-aggregation *in vitro*. Very recently, researches on the complex of SARS-CoV-2 nsp12-7-8 complex resolved by Cryo-EM showed that the N-terminus of nsp8 plays pivotal roles in RNA interaction.

Discussion

Compared with structures of SARS-CoV nsp7-8 hexadecamer complex and the heterotrimeric complex of FCoV nsp7-8, the nsp7-8C structure of SARS-CoV-2 forms a heterotetramer consisting with two of nsp7 molecules and two nsp8 C-terminus. Consider N-terminus of nsp8 is highly conserved in coronaviruses and is vital in multimeric assembly of the nsp7-8 complex in SARS-CoV and FCoV, we analyzed the solution state of the SARS-CoV-2 full-length nsp7-8 complex in multiple ways. The SARS-CoV-2 nsp7 and nsp8 were mixed at different ratios before loading to the SEC column, but the nsp7-8 complex always eluted with a retention volume of 14.75 mL with an estimated MM of 58 kDa. Chemical cross-linking also indicated that the 2:2 stoichiometry of nsp7:nsp8 observed in our crystal structure exists in solution, as well. Based on these results, SARS-CoV-2 nsp7 and nsp8 can form a stable tetrameric conformation in the solution state.

The alignment of the structures of all three nsp7-8 complexes (2AHM, 6NUR, and 5F22) in SARS coronaviruses reveals a conserved six-helical bundle contact surface, which is the major interface that forms the nsp7-8C tetramer. The structure of the SARS-CoV nsp7-nsp8 hexadecamer can be regarded as a supercomplex assembled from four nsp7-8 tetramers, which are tightly associated with the long α -helix at the N-terminal of nsp8. The proteolysis product of SARS-CoV nsp8, which loses the N-terminal long α -helix, only forms a tetrameric architecture similar to the structure of SARS-CoV-2 nsp7-8 tetramer (Supplementary Figure S4a and S5). However, it should be noted that the nsp12-nsp7-nsp8 complex in SARS coronaviruses represents the minimal RNA polymerase complex required for viral genomic transcription and replication. There are two cryo-EM structures of RdRp (nsp12) from SARS-CoV-2 bound to nsp7 and nsp8 co-factors, providing a view of the overall architecture of the super complex. However, no hexadecameric or tetrameric forms of nsp7-8 observed in these structures indicated that nsp7-8 may perform diverse biological functions through differential assembly conformations (Supplementary Fig. S4b-c). Very recently, Quan *et al.* also described the SARS-CoV-2 polymerase-RNA complex structure and proposed that the hexadecameric nsp7-8 primase complex is dissociated by half after primer synthesis and delivers the nucleic acid to the polymerase-RNA complex directly. The crystal structure of the nsp7-8 complex from SARS-CoV-2 reported here suggests that the core architecture of the nsp7-8 complex plays a pivotal role in RNA-dependent RNA polymerase assembly.

Methods

Plasmid construction, protein purification, and crystallization

The SARS-CoV-2 nsp7 and nsp8 genes were synthesized and subcloned into the pGEX-6p-1 and pET-DUET1 vectors, respectively. Recombinant proteins nsp7 and nsp8, as well as two truncated nsp8 versions, nsp8 Δ C (residues 1-78) and nsp8 Δ N (residues 77-198), were expressed in *Escherichia coli* (E. coli) BL21 (DE3). The cells were harvested after 18 h induced by 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 16 °C. Next, the cell precipitation was resuspended into buffer with 20 mM Tris pH 8.0 and 150 mM NaCl before the homogenization process. After centrifugation to remove cell debris, the Glutathione S-transferase (GST) tag labeled nsp7 protein was purified with a glutathione affinity column (Qiagen, Hilden, Germany) and the GST tag was released by incubation with PreScission Protease at a final concentration of 0.05 mg/mL in solution. The nsp8 and truncations were purified by Ni-NTA affinity chromatography (Qiagen, Hilden, Germany) following Resource Q (nsp8 and nsp Δ N) or Resource S (nsp8 Δ C) chromatography (Cytiva, USA). The final protein samples were concentrated to 1 mL in Amicon Ultra concentrators (cut-off size of 3 kDa; Millipore, USA) and loaded onto a Superdex™ 200 column (GE Healthcare, USA) for further purification and characterization, with a buffer of 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM dithiothreitol (DTT), and 10% glycerol (v/v). The purity of the nsp7 and nsp8 proteins reached above 95%, as confirmed by SDS-PAGE. Nsp8 or nsp8 Δ N was then mixed with 1 molar excess of nsp7 and passed over the Superdex™ 200 column in the same buffer. Fractions of the nsp7-nsp8 complex were prepared with 1:1 stoichiometry and concentrated to 20 mg/mL for crystallization. The nsp7-nsp8 crystals were initially grown at 20 °C for seven days by using the hanging-drop vapor-diffusion method with a mixture of 1 μ L protein and 1 μ L reservoir solution. One single protein crystal with high X-ray diffraction quality was obtained in 200 mM NaCl, 100 mM Tris-HCl, pH 8.2 and 25% PEG3350. The single crystal was stored in liquid nitrogen for data collection.

Data collection, processing, and structure determination

The diffraction data of nsp7-nsp8 complex crystal were collected at 100K in at the SSRF Beamline BL18U1 (Shanghai, China) at a wavelength of 0.97930 Å. Data were processed and scaled using the HKL2000 package³⁰. The nsp7-nsp8 structure was solved by molecular replacement using Phaser in the CCP4 package with the initial searching coordinates of SARS-CoV nsp7 and nsp8 (PDB: 2AHM)³³. Cycles of refinement and model building were carried out with the REFMAC5, Phenix, and COOT software programs^{34,35}. Model geometry was verified using the SWISS model³⁶. Structural figures were drawn using PyMOL. The data collection and refinement statistics are shown in Table 1.

Cross-linking assay by glutaraldehyde

Cross-linking was carried out by incubating 1 mg/mL of wildtype SARS-CoV-2 nsp7, nsp8, nsp8C, nsp7-8, 100 mM NaCl, and 10 % glycerol (v/v). Different final

concentrations of glutaraldehyde (0.005%, 0.01%, and 0.02% (v/v)) were utilized at 25 °C for 5, 10, and 15 min. The reaction was quenched by adding 1 M Tris pH 8.0 to a final concentration of 0.3 M and analysis was carried out by 14 % SDS-PAGE with detection by Coomassie Brilliant Blue staining.

Analytic size-exclusion chromatography

SEC analysis was conducted with evaluation of the oligomeric state of the nsp7, nsp8, nsp8C, nsp7-nsp8 and nsp7-nsp8C truncated forms using a Superdex™ 200 10/300 GL column (Cytiva, USA). The column was equilibrated with elution buffer in 20 mM Tris (pH 8.0), 300 mM NaCl, 1 mM DTT, and 10 % glycerol (v/v) at a flow rate of 0.3 mL/min followed by protein samples at a concentration of ~1 mg/mL at 4 °C.

Electrophoretic mobility-shift assay

FAM-labeled 12-mer ssRNA oligonucleotide (5'FAM-GCTTTGATTCG-3') was used for the electrophoretic mobility-shift assay. Initially, 2 nmol DNA was incubated with different concentrations of nsp7-nsp8 complex along with free nsp7, nsp8, and truncated nsp8 forms in 10 mM HEPES pH 8.0, 50 mM KCl, 1 mM EDTA, 0.05% Triton-X-100, and 5% Glycerol for 30 min at 4 °C, followed by the addition of 10×loading buffer (250 mM Tris-HCl pH 7.9, 40% glycerol) to the mixture. Samples were then run on 6.5% non-denaturing TBE polyacrylamide gels for 30 min at a voltage of 100 V, and the results were determined with a Bio-Rad ChemiDoc Touch Imaging System (Bio-Rad, USA).

Surface plasmon resonance analysis

The interaction between nsp8 or nsp7-nsp8 complex proteins with 12-mer ssRNA was monitored by SPR using a BIAcore 8K (GE Healthcare) carried out at 16 °C in running buffer composed of 50 mM HEPES pH 7.5, 150 mM NaCl, and 0.05% Tween 20. The serial concentrations of nsp8 or nsp7-nsp8 complex proteins were 0.375, 0.75, 1.5, 3, and 6 mM when testing interactions with 12-mer ssRNA, respectively. The biotin-labeled 12-mer ssRNA was immobilized to a sensor chip SA to a level of ~200 RUs using a BIAcore 8K in the same running buffer described earlier. The resulting data were fitted to a 1:1 binding model, and the binding affinity was calculated using BIAcore 8K Evaluation Software (GE Healthcare).

References

- 1 Kaul, D. An overview of coronaviruses including the SARS-2 coronavirus - Molecular biology, epidemiology and clinical implications. *Curr Med Res Pract* **10**, 54-64, doi:10.1016/j.cmrp.2020.04.001 (2020).
- 2 Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **382**, 727-733, doi:10.1056/NEJMoa2001017 (2020).
- 3 Coronaviridae Study Group of the International Committee on Taxonomy of, V. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature* **579**, 522-524, doi:10.1038/s41586-020-2008-3 (2020).

- 4 Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565-574, doi:10.1016/S0140-6736(20)30251-8 (2020).
- 5 Chan, J. F. *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* **9**, 221-236, doi:10.1080/22221751.2020.1719902 (2020).
- 6 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).
- 7 Graham, R. L., Sparks, J. S., Eckerle, L. D., Sims, A. C. & Denison, M. R. SARS coronavirus replicase proteins in pathogenesis. *Virus Res* **133**, 88-100, doi:10.1016/j.virusres.2007.02.017 (2008).
- 8 Hui, D. S. *et al.* The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* **91**, 264-266, doi:10.1016/j.ijid.2020.01.009 (2020).
- 9 Brian, D. A. & Baric, R. S. Coronavirus genome structure and replication. *Curr Top Microbiol Immunol* **287**, 1-30, doi:10.1007/3-540-26765-4_1 (2005).
- 10 Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921 e910, doi:10.1016/j.cell.2020.04.011 (2020).
- 11 Snijder, E. J. *et al.* Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *Journal of molecular biology* **331**, 991-1004, doi:10.1016/s0022-2836(03)00865-9 (2003).
- 12 Shin, D. *et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*, doi:10.1038/s41586-020-2601-5 (2020).
- 13 Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* **368**, 409-412, doi:10.1126/science.abb3405 (2020).
- 14 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, doi:10.1007/s00705-012-1404-x (2012).
- 15 Thiel, V. *et al.* Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol* **84**, 2305-2315, doi:10.1099/vir.0.19424-0 (2003).
- 16 te Velthuis, A. J., Arnold, J. J., Cameron, C. E., van den Worm, S. H. & Snijder, E. J. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Res* **38**, 203-214, doi:10.1093/nar/gkp904 (2010).

- 17 Subissi, L. *et al.* SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. *Antiviral Res* **101**, 122-130, doi:10.1016/j.antiviral.2013.11.006 (2014).
- 18 Johnson, M. A., Jaudzems, K. & Wuthrich, K. NMR Structure of the SARS-CoV Nonstructural Protein 7 in Solution at pH 6.5. *J Mol Biol* **402**, 619-628, doi:10.1016/j.jmb.2010.07.043 (2010).
- 19 von Brunn, A. *et al.* Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. *PLoS One* **2**, e459, doi:10.1371/journal.pone.0000459 (2007).
- 20 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, doi:10.1038/nsmb999 (2005).
- 21 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).
- 22 Chen, Y. *et al.* Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS Pathog* **7**, e1002294, doi:10.1371/journal.ppat.1002294 (2011).
- 23 Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* **10**, 2342, doi:10.1038/s41467-019-10280-3 (2019).
- 24 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).
- 25 Hillen, H. S. *et al.* Structure of replicating SARS-CoV-2 polymerase. *Nature* **584**, 154-156, doi:10.1038/s41586-020-2368-8 (2020).
- 26 Yin, W. *et al.* Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499-1504, doi:10.1126/science.abc1560 (2020).
- 27 Imbert, I. *et al.* A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J* **25**, 4933-4942, doi:10.1038/sj.emboj.7601368 (2006).
- 28 Deming, D. J., Graham, R. L., Denison, M. R. & Baric, R. S. Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. *J Virol* **81**, 10280-10291, doi:10.1128/JVI.00017-07 (2007).
- 29 Pan, J. *et al.* Genome-wide analysis of protein-protein interactions and involvement of viral proteins in SARS-CoV replication. *PLoS One* **3**, e3299, doi:10.1371/journal.pone.0003299 (2008).
- 30 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).

- 31 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).
- 32 Scharer, M. A., Grutter, M. G. & Capitani, G. CRK: an evolutionary approach for distinguishing biologically relevant interfaces from crystal contacts. *Proteins* **78**, 2707-2713, doi:10.1002/prot.22787 (2010).
- 33 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).
- 34 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
- 35 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
- 36 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).

Declarations

Acknowledgments

We appreciate the staff of the State Key Laboratory of Biotherapy, Sichuan University, during the period of the COVID-19 epidemic. X-ray diffraction experiments were carried out at the Shanghai Synchrotron Radiation Facility (BSRF) BL18U1. We further thank the Beamline staff for approving our urgent proposal submission for this COVID-19 research project. Moreover, we are grateful to Dr. Guangwen Lu for providing us with the associated plasmids. This work was supported by grants from the National Key Research and Development Program of China (2017YFA0505903), the National Natural Science Foundation of China (2018YFC1312300), the Special Research Fund on COVID-19 of Sichuan Province (2020YFS0010), and the urgent project on COVID-19 of West China Hospital, Sichuan University (HX-2019-nCoV-044).

Authors' contributions

D.S. and C.Z. designed the study. C.Z. purified and characterized proteins, cultivated protein crystals, and collected X-ray diffraction data. D.S. solved the crystal structure. D.S. and C.Z. led writing the manuscript. L.L. designed and produced all the figures. C.C. and J.H. contributed to the data analysis and final version of the manuscript. D.S. had final approval of the version to be published.

Conflict of interest: The authors declare that they have no conflicts of interest surrounding the contents of this article.

Tables

Table 1. SARS-CoV-2 nsp7-8 data collection and refinement statistics

	SARS-CoV-2 nsp7-8
<i>Data collection</i>	
Resolution (Å)	2.57
Space group	<i>P1</i>
Unit-cell parameters (Å, °)	43.031 42.99 100.63 90.012 90.035 72.613
Resolution range (Å)	41.06–2.57 (2,662–2.57)
R _{merge} [†] (%)	13.8 (23.6)
R _{pim} [‡] (%)	7.5 (12.7)
Average <i>I</i> /σ(<i>I</i>)	11.6 (4.5)
No. of unique reflections	20426 (1981)
Completeness (%)	93.14 (89.76)
Solvent content (%)	50
Molecules per asymmetric unit	8
<i>Refinement</i>	
R _{work} /R _{free}	0.25/0.35
Ramachandran favored (%)	87.92
Ramachandran outliers (%)	0.67
<i>No. of atoms</i>	
Protein	5809
Water	75
Wilson B value	43.68
<i>Root-mean-square deviations</i>	
Bond length (Å)	0.010
Bond angle (°)	1.239

[†] $R_{\text{merge}} = \frac{\sum_i \sum_h |I_i(h) - \langle I(h) \rangle|}{\sum_h \langle I(h) \rangle}$, where $I_i(h)$ is an individual intensity measurement and $\langle I(h) \rangle$ is the average intensity for all i reflections.

[‡] R_{pim} is approximately estimated by multiplying the R_{merge} value by the factor $\sqrt{N}$, where N is the overall redundancy of the data set.

Figures

Figure 1

Overall structure of the SARS-CoV-2 nsp7-8C complex. (a) The protomeric architecture of SARS-CoV-2 nsp7-8C. The nsp7 and nsp8C proteins are colored in light blue and orange, respectively. (b) The four α helices of nsp7. (c) The structural details of nsp8C. nsp8C contains five α helices and four β strands.

Figure 2

The tetramer structure and interface details of SARS-CoV-2 nsp7-8C. (a) The overall tetramer structure of SARS-CoV-2 nsp7-8C. There is a central symmetry between the two nsp7-8C molecules protomerically. (b) The interface of the SARS-CoV-2 nsp7-8C tetramer. The interactional amino acid residues and distances are represented by abbreviated letters and numbers.

Figure 3

The nsp8 take responsible for nucleic acid-binding. (a-b) The full-length nsp8 SARS-CoV-2 model show as cartoon and surface (blue, positive charge (+60 kBT); red, negative charge (-60 kBT)). (c-d) SPR sensorgram showing the binding kinetics for the SARS-CoV-2 nsp7-8 complex and 12-mer ssRNA. Dates are shown as color lines. (e-f) Single-stranded (ss) RNA-binding abilities of SARS-CoV-2 nsp7-8 as determined by electrophoretic mobility shift assay. The state of the SARS-CoV-2 nsp7-8 complex and the concentration of proteins used in the reaction system are indicated above the gel.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfiles.pdf](#)