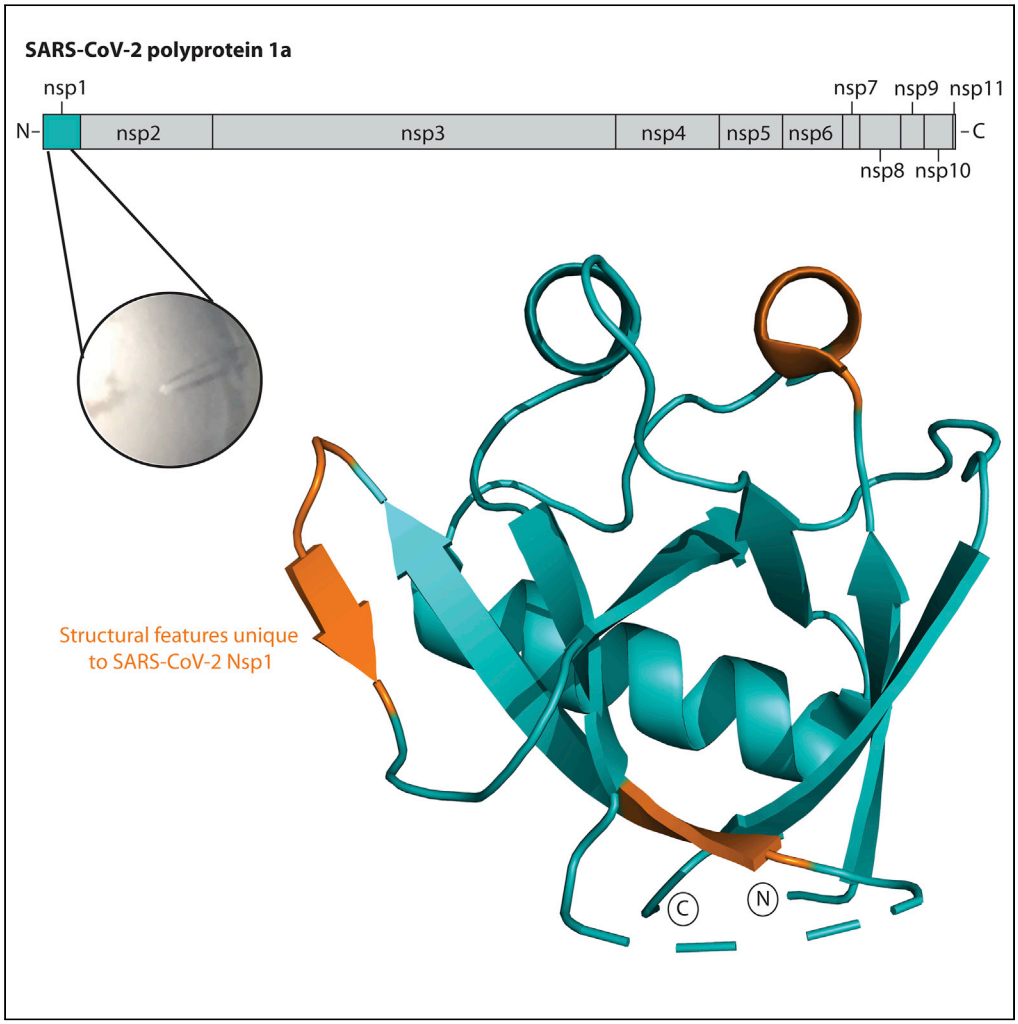
iScience

CellPress
OPEN ACCESS

Article

Structural characterization of nonstructural protein 1 from SARS-CoV-2



Cameron Semper, Nobuhiko Watanabe, Alexei Savchenko

alexei.savchenko@ucalgary.ca

Highlights

SARS-CoV-2 Nsp1 features a capped β-barrel structure, similar to that of SARS-CoV

Distinct structural features distinguish SARS-CoV-2 Nsp1 from its SARS-CoV ortholog

The plasticity of the Nsp1 protein fold is evident through comparison with homologs

Semper et al., iScience 24, 101903
January 22, 2021
https://doi.org/10.1016/j.isci.2020.101903

iScience

CellPress
OPEN ACCESS

Article

Structural characterization of nonstructural protein 1 from SARS-CoV-2

Cameron Semper,[1] Nobuhiko Watanabe,[1,2] and Alexei Savchenko[1,2,3,*]

Summary

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a single-stranded, enveloped RNA virus and the etiological agent of the current coronavirus disease 2019 pandemic. Efficient replication of the virus relies on the activity of nonstructural protein 1 (Nsp1), a major virulence factor shown to facilitate suppression of host gene expression through promotion of host mRNA degradation and interaction with the 40S ribosomal subunit. Here, we report the crystal structure of the globular domain of SARS-CoV-2 Nsp1, encompassing residues 13 to 127, at a resolution of 1.65 Å. Our structure features a six-stranded, capped β-barrel motif similar to Nsp1 from SARS-CoV and reveals how variations in amino acid sequence manifest as distinct structural features. Combining our high-resolution crystal structure with existing data on the C-terminus of Nsp1 from SARS-CoV-2, we propose a model of the full-length protein. Our results provide insight into the molecular structure of a major pathogenic determinant of SARS-CoV-2.

Introduction

In March of 2020, the World Health Organization declared coronavirus disease 2019 (COVID-19) a global pandemic. As of October 2020, there have been more than 35,400,000 cases of infection reported globally and approximately 1,040,000 deaths attributed to COVID-19 (Dong et al., 2020). The etiological agent of this pandemic has been identified as severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), a member of the *Betacoronavirus* genus and closely related to the severe acute respiratory syndrome coronavirus (SARS-CoV) that caused the severe acute respiratory syndrome outbreak of 2002–2004 (Zhou et al., 2020; Chan et al., 2020).

Coronaviruses infect a diverse array of vertebrates, with infection typically resulting in respiratory disease or gastroenteritis (Weiss and Navas-Martin, 2005). Their broad host range allows for a substantial reservoir for human infection and mutation facilitates cross-species transmission (Yang et al., 2015). There is mounting evidence that SARS-CoV-2 originated via mutation and cross-species transmission of a pangolin coronavirus with which it shares ~97% sequence identity (Zhang et al., 2020a). This highlights the dramatic impact that relatively few amino acid substitutions can have in coronaviruses and underscores the urgent need for characterization of these infectious agents at a molecular level.

SARS-CoV-2 is an enveloped, positive-sense RNA virus with a single-stranded genome approximately 30 kb in size. The genome is 5′-capped and 3′poly-adenylated and the first ~2/3 of the genome encode two overlapping reading frames that produce polyprotein 1a and 1ab (Masters, 2006). Downstream of this region, the remaining 1/3 of the genome encodes for structural proteins and a number of open reading frames (ORFs) that produce small accessory proteins. Polyprotein 1a and 1ab are large polypeptides that are processed post-translationally by viral-encoded proteases to produce non-structural proteins (Nsp) 1–16 (Narayanan et al., 2015). The genome also contains 5′ and -3′ untranslated regions (UTRs), the former of which plays a critical role in self-recognition that allows for SARS-CoV-2 protein production to occur unabated while host gene expression is suppressed (Tanaka et al., 2012). At the core of this mechanism is Nsp1, a 180 amino acid (AA) protein produced via processing of polypeptide 1a and 1ab by the Papain-like protease domain of Nsp3 (Figure 1A) (Harcourt et al., 2004).

Nsp1 mediates a two-pronged approach to suppression of host gene expression. Firstly, it inhibits translation of host proteins during the initiation stage through interaction with the 40S ribosomal subunit. Secondly, Nsp1 promotes the degradation of host mRNA by endonucleolytic cleavage within the 5′UTR, which

[1]Department of Microbiology, Immunology and Infectious Disease, University of Calgary, HSC B724 3330 Hospital Drive NW, Calgary, Alberta, T2N 4N1, Canada

[2]Center for Structural Genomics of Infectious Diseases (CSGID)

[3]Lead contact

*Correspondence: alexei.savchenko@ucalgary.ca

https://doi.org/10.1016/j.isci.2020.101903

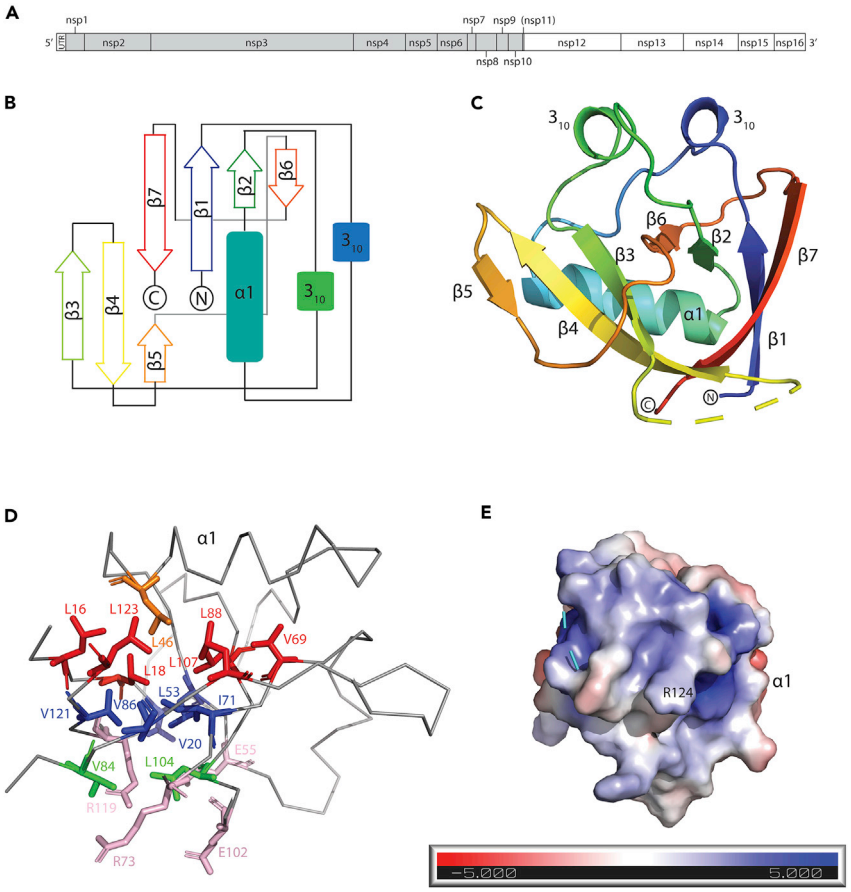


Figure 1. The structure of SARS-CoV-2 Nsp1$_{13\text{-}127}$

(A) Schematic of Orf1ab from SARS-CoV-2. The gray shading encompasses the non-structural proteins produced by Orf1a. Note, nsp11 is not produced by Orf1ab due to ribosomal frameshift (Naqvi et al., 2020). (B) Topological arrangement of SARS-CoV-2 Nsp1$_{13\text{-}127}$.

(C) Cartoon representation of the SARS-CoV-2 Nsp1$_{13\text{-}127}$ structure colored from N-terminus (blue) to C-terminus (red). Secondary structure elements and the N- and C-termini are labeled.

(D) Ribbon depiction of the structure, with side chains that contribute to the hydrophobic core of the beta-barrel shown in stick representation. The three layers of side chains (shown in red, blue, and green), plus the charged residues (shown in pink) at the bottom are colored and the residues within the layers are labeled.

(E) Electrostatic surface potential of the likely RNA-binding interface of SARS-CoV-2 Nsp1$_{13\text{-}127}$. R124 which is critical for the interaction with SL1 of the 5′UTR is labeled.

in turn leads to accelerated Xrn1-mediated mRNA decay (Kamitani et al., 2006, 2009). Viral mRNAs are able to avoid the fate of host mRNAs through interaction between Nsp1 and the stem-loop 1 (SL1) motif found in the viral 5′ UTR (Tanaka et al., 2012). Nsp1 from SARS-CoV has been shown to inhibit the type I interferon response in infected cells, allowing the virus to circumvent the innate immune response (Narayanan et al., 2008). Expression of SARS-CoV Nsp1 has also been shown to induce the production of chemokines, suggesting this protein may play a role in the "cytokine storm", a maladaptive release of cytokines in response to infection, associated with a number of COVID-19 infections (Law et al., 2007; Ye et al., 2020). Thus, Nsp1 has emerged as a major pathogenicity factor that plays a critical role in the coronavirus infection cycle (Zust et al., 2007). Deletion or mutation of Nsp1 results in viral attenuation in infection models and restores the innate immune response in infected cells (Wathelet et al., 2007). Based on its central role in suppression of the host immune response and essentiality to infection, Nsp1 has been proposed as a therapeutic target for the treatment of COVID-19 (Zust et al., 2007). A prerequisite to any investigation into possible interventions targeting Nsp1 is high-resolution structural data that can facilitate robust *in silico* screening. A partial structure corresponding to the N-terminal fragment of Nsp1 from SARS-CoV was resolved by nuclear magentic resonance (NMR) spectroscopy and revealed a unique β-barrel motif. The corresponding domain from

SARS-CoV-2 shares 86% sequence identity with its SARS-CoV ortholog, which is a considerable level of sequence diversity compared to many other non-structural proteins encoded by the SARS-CoV-2 genome (e.g. Nsp12 shares 96% identity with SARS-CoV ortholog (Zhang et al., 2020b)). Recent structural and biochemical studies involving SARS-CoV-2 non-structural proteins have revealed a level of functional divergence that would not have otherwise been predicted based on the high degree of sequence conservation with SARS-CoV orthologs. Biochemical characterization of the Nsp12-Nsp7-Nsp8 replicase complex showed that minor changes in AA sequence compared to SARS-CoV resulted in reduced enzymatic activity (Peng et al., 2020). Structural characterization of Nsp9 identified a novel peptide binding site not predicted bioinformatically or reported for Nsp9 from SARS-CoV (Littler et al., 2020). These findings are illustrative of the need for continued pursuit of structural characterization of SARS-CoV-2 proteins.

Here, we report the high-resolution crystal structure of the globular N-terminal domain of Nsp1 from SARS-CoV-2 at 1.65 Å resolution (PDB: 7K3N). Our data reveal a high level of structural conservation between Nsp1 of SARS-CoV-2 and SARS-CoV but also some unique structural features that likely contribute to increased stability of the β-barrel fold in SARS-CoV-2 Nsp1. Comparative analysis reveals additional structural homologs in Nsp1 proteins from *Alphacoronaviruses*, despite low levels of shared sequence identity. These results highlight the critical role this unique protein fold plays in facilitating viral infection and suppression of host gene expression.

Results

Crystal structure of SARS-CoV-2 Nsp1$_{13\text{-}127}$

In pursuit of structural characterization of Nsp1 from SARS-CoV-2, we expressed a codon-optimized version of the ORF encoding this protein in *E. coli*. Using this expression system, we obtained purified full-length Nsp1 after a two-step purification protocol (see details in Methods). However, we were unable to crystallize the full-length Nsp1 protein, likely due to the presence of flexibly disordered regions at the N- and C-termini.

Previous studies of Nsp1 from SARS-CoV reported the presence of a distinct globular domain comprising residues 13 to 127 (Almeida et al., 2007). To structurally characterize the N-terminal fragment of Nsp1 from SARS-CoV-2 (Nsp1$_{13\text{-}127}$), we sub-cloned it and expressed it in *E. coli*. The expression level of Nsp1$_{13\text{-}127}$ was comparable to that of the full-length protein, and with this fragment, we obtained well-diffracting crystals in several conditions that enabled structural determination of this portion of Nsp1.

The crystal structure of SARS-CoV-2 Nsp1$_{13\text{-}127}$ was determined via molecular replacement using the SARS-CoV Nsp1(PDB: 2HSX) solution structure as the search model and refined to a resolution of 1.65 Å (Table 1). The asymmetric unit contained a single molecule of Nsp1$_{13\text{-}127}$. Analysis of the Nsp1$_{13\text{-}127}$ structure by PDBePISA revealed that the protein has a total solvent-exposed area of 6102 Å^2 (Krissinel and Henrick, 2007).

As in the case of SARS-CoV, the structure of SARS-CoV-2 Nsp1$_{13\text{-}127}$ features a unique topological arrangement resulting in the formation of a six-stranded (n = 6) β-barrel that is primarily antiparallel, with the exception of strands β1 (Q15 – V20) and β2 (C51 – V54) (Figure 1B). Additional major structural features include α1 (V35 – D48) helix, which is positioned as a cap along one opening of the β-barrel, two 3_{10} helices that run parallel to each other, and the β5 strand (I95 – Y97) which is not part of the β-barrel but forms a β-sheet interaction with the β4 strand (V84 – L92) (Figure 1C).

The core of the β-barrel is highly hydrophobic and mainly comprises the side chains of thirteen AAs organized into three layers. The first layer, which is adjacent to α1 helix, is formed by side chains of residues L16, L18, V69, L88, L107, and L123 (Figure 1D). The opening of the β-barrel at this layer is obstructed by the α1 helix, which contributes the side chain of L46 to the center of this first layer of core hydrophobic residues (Figure 1D). The middle layer of the β-barrel comprises the side chains of residues V20, L53, I71, V86, and V121, while the bottom layer features residues V84 and L104. Adjacent to this bottom layer are four solvent-exposed residues: E55, R73, E102, and R119 whose side chains point inward toward the core of the β-barrel (Figure 1D). Consequently, the hydrophobic core at both ends of the β-barrel is occluded: by the α1 helix on one side and the charged residues (E54, R73, E102, R119) on the other.

A distinctive feature of Nsp1 that is evident in the crystal structure of Nsp1$_{13\text{-}127}$ is the large number of flexible loops. For one of these loops (A76 – H81), we were unable to resolve the structure due to a lack of interpretable electron density. This region has been previously shown to be highly flexible as evidenced by the

Table 1. X-ray crystallographic statistics

	SARS-Cov-2 Nsp1$_{13-127}$ (PDB: 7K3N)
Data Collection	
Space group	$P4_32_12$
Cell dimensions	
a, *b*, *c* (Å)	30.0, 37.0, 144.8
α, β, γ (°)	90.0, 90.0, 90.0
Wavelength	0.97913
Resolution (Å)	50.0-1.65
	(1.68-1.65)
R_{meas} [a]	0.058 (1.091)
R_{pim} [b]	0.020 (0.387)
I/s*I*	35.6 (2.0)
Completeness (%)	99.8 (99.5)
Redundancy	8.2 (7.4)
CC1/2	0.997 (0.690)
Refinement	
Resolution (Å)	32.95-1.65
No. reflections	12,884
R_{work}/R_{free} [c]	0.203/0.248
No. atoms	
Protein	828
Water	115
Average B-factors (Å^2)	20.7
Protein	19.2
Water	31.2
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.872
Ramachandran (%)	
Favored	97.1
Allowed	2.9
Disallowed	0

Values in brackets refer to highest resolution shells.

[a] $R_{meas} = \Sigma_{hkl}\sqrt{(n/n-1)}\,\Sigma^{n}_{j=1}|I_{hkl,j} - \langle I_{hkl}\rangle|/\Sigma_{hkl}\Sigma_j I_{hk,j}$.

[b] $R_{pim} = \Sigma_{hkl}\sqrt{(1/n-1)}\,\Sigma^{n}_{j=1}|I_{hkl,j} - \langle I_{hkl}\rangle|/\Sigma_{hkl}\Sigma_j I_{hk,j}$.

[c] $R = \Sigma||F_p^{obs} - F_p^{calc}|/\Sigma F_p^{obs}$, where F_p^{obs} and F_p^{calc} are the observed and calculated structure factor amplitudes, respectively.

multitude of backbone confirmations observed in the NMR structure of SARS-CoV Nsp1 (Almeida et al., 2007). The positions of other loops in the crystal structure were stabilized via interactions between specific secondary structure elements. Specifically, the two 3_{10} helices are aligned in the crystal structure forming H-bonds between R24 and Q63. This interaction appears to stabilize the position of two of the largest loops in the crystal structure, the β1-α1 (L21 – S34) loop and the β2-β3 (E54 – P67) loop.

Electrostatic surface potential analysis of the SARS-CoV-2 Nsp1$_{13-127}$ structure revealed several regions of charged residue colocalization (Baker et al., 2001). The most electronegative and electropositive patches appear adjacent on the Nsp1$_{13-127}$ surface, separated by a small electroneutral region corresponding to the α1 helix. The electropositive patch continues around the surface of the protein, where the R124 residue is localized (Figure 1E). This residue has been shown to be essential for the interaction between Nsp1 and SL1 of the 5′ UTR, implicating this charged surface in RNA binding (Tanaka et al., 2012).

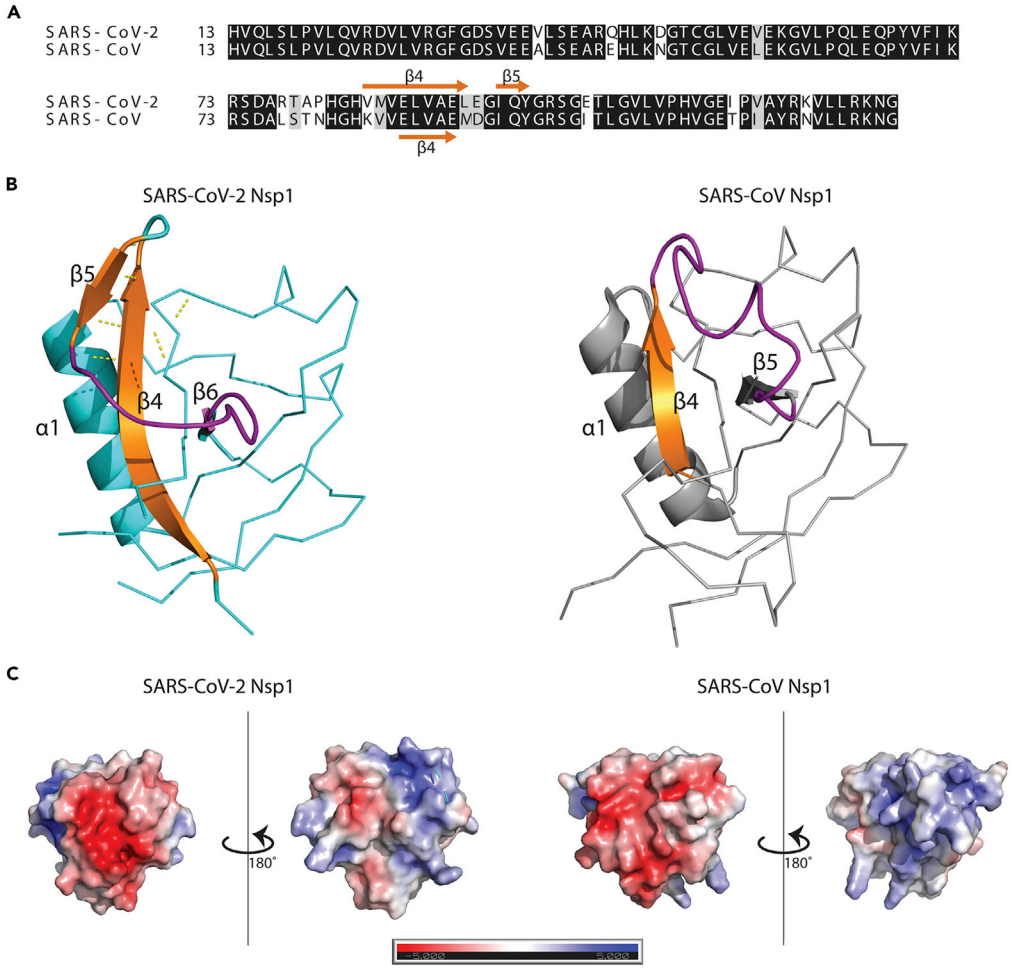


Figure 2. Comparison of SARS-CoV-2 Nsp1$_{13-127}$ to SARS-CoV Nsp1

(A) Alignment of residues 13–127 of Nsp1 with conserved residues highlighted in black.

(B) Ribbon representation of SARS-CoV-2 (PDB: 7K3N) and SARS-CoV (PDB: 2HSX) Nsp1$_{13-127}$. β4 (shown in cartoon representation) is extended in the SARS-CoV-2 structure allowing for H-bonding with and formation of β5, which is absent in the SARS-CoV Nsp1 structure. The loop region (and equivalent in SARS-CoV Nsp1) stabilized through interaction between β4 and β5 is colored in purple.

(C) Comparison of electrostatic surface potential of SARS-CoV-2 and SARS-CoV Nsp1$_{13-127}$.

Comparison with SARS-CoV Nsp1

The overall fold of SARS-CoV-2 Nsp1$_{13-127}$ was highly similar to that of the corresponding SARS-CoV Nsp1 fragment in accordance with the significant primary sequence identity shared between the two proteins (Figure 2A). However, detailed comparative analysis revealed several notable differences between the two structures. The SARS-CoV-2 structure of Nsp1$_{13-127}$ features two 3_{10} helices, compared to the single one found in SARS-CoV Nsp1. The primary sequence corresponding to the second 3_{10} helix is completely conserved between SARS-CoV-2 and SARS-CoV Nsp1 proteins, suggesting that this secondary structure element may be transient in nature or influenced by differences elsewhere in SARS-CoV-2 Nsp1$_{13-127}$. The two 3_{10} helices interact with each other through H-bonds in SARS-CoV-2 Nsp1$_{13-127}$, but this interaction is notably absent in the NMR structure of Nsp1 from SARS-CoV (Almeida et al., 2007).

Two of the most striking differences between the two Nsp1 structures are the extension of β4 and the presence of an additional β-strand in the case of SARS-CoV-2 Nsp1$_{13-127}$. AA changes at positions 84 (K to V) and 85 (V to M) between SARS-CoV and SARS-CoV-2 Nsp1 result in an extension of β4 on the N-terminus by three residues (Figure 2B). This extension results in an increased number of H-bonds formed between β4 and β7 in SARS-CoV-2 Nsp1 compared to the pairing between β4 and β6 that occurs in SARS-CoV

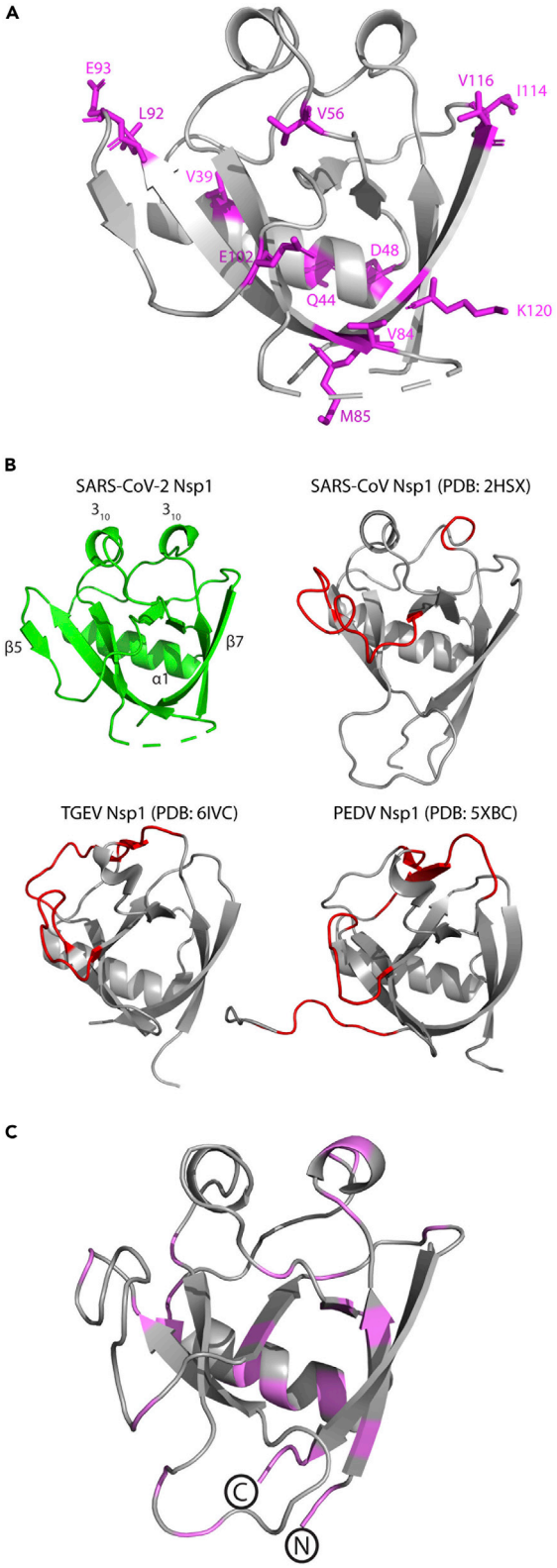

A
E93
L92
V56
V116
I114
V39
E102
D48
Q44
K120
V84
M85
B
SARS-CoV-2 Nsp1
SARS-CoV Nsp1 (PDB: 2HSX)
β5
β7
α1
TGEV Nsp1 (PDB: 6IVC)
PEDV Nsp1 (PDB: 5XBC)
C
C
N

Figure 3. Comparison of the structure of SARS-CoV-2 Nsp1$_{13-127}$ to structural homologs

(A) Cartoon representation of the SARS-CoV-2 Nsp1$_{13-127}$ structure with non-conserved residues between SARS-CoV-2 and SARS-CoV highlighted in magenta.

(B) SARS-CoV-2 Nsp1$_{13-127}$ (PDB: 7K3N) is shown in green; SARS-CoV Nsp1 (PDB: 2HSX), TGEV Nsp1 (PDB: 6IVC), and PEDV Nsp1 (PDB: 5XBC) are shown in gray with noticeable structural differences (compared to SARS-CoV-2) highlighted in red. Distinctive secondary structure features from the SARS-CoV-2 Nsp1$_{13-127}$ structure are labeled.

(C) Homology model of β-barrel domain of Nsp1 from MERS-CoV. Conserved residues between MERS-CoV and SARS-CoV-2 Nsp1 are highlighted in magenta and the N- and C-termini are labeled.

Nsp1. This extension also facilitates the formation of the β5 strand which is absent in SARS-CoV Nsp1. The β5 and β4 strands are part of a small β-sheet that stabilizes the loop region between β5 and β6 (equivalent to β4 and β5 in SARS-CoV Nsp1) (Figure 2B).

Comparison of the electrostatic surface potential of Nsp1$_{13-127}$ from SARS-CoV-2 and SARS-CoV shows additional differences between the two proteins. The SARS-CoV structure features a single patch of electronegative surface potential opposite a separate patch of electropositive surface potential. These two charged surfaces are oriented 180° opposite one another, and there is a ring of electroneutral surface potential separating the two charged surfaces (Figure 2C). The SARS-CoV-2 Nsp1$_{13-127}$ structure features an equivalent patch of electronegativity, but the electropositive surface is divided into two by a patch of mild electronegative surface potential.

Mapping the non-conserved residues onto the SARS-CoV-2 Nsp1$_{13-127}$ structure revealed that most of the differences are in solvent-exposed residues. Distinct patches of variation are observed immediately adjacent to both ends of β4 and in the loop between β3 and β4 (S74 – H83), where 4 of 9 residues differ between SARS-CoV-2 and SARS-CoV (Figure 3A). The remaining sequence variation is broadly distributed across the Nsp1$_{13-127}$ structure. Residues that contribute to the hydrophobic core of the β-barrel are conserved between SARS-CoV-2 and SARS-CoV, suggesting that this particular organization is critical to the function of Nsp1.

Comparison with other structural homologs

AA sequence-based searches for homologs of Nsp1 reveal a high level of sequence conservation among representatives of lineage B of betacoronaviruses, which includes several animal viruses closely related to SARS-CoV and SARS-CoV-2. Despite the presence of functionally equivalent Nsp1 proteins, no sequence-based homologs of SARS-CoV-2 Nsp1 were detected in other lineages of betacoronaviruses. This lack of sequence conservation is observed even when comparing SARS-CoV-2 Nsp1 to other coronaviruses known to infect humans, including Middle East respiratory syndrome coronavirus (MERS-CoV), human coronavirus (HCoV)-OC43, and HCoV-HKU1.

A search for structural homologs using our high-resolution structure of SARS-Cov-2 Nsp1$_{13-127}$ identified three closely related structural homologs, all annotated as Nsp1 proteins (Table 2) (Holm and Laakso, 2016). Unsurprisingly, the Nsp1 structure from SARS-CoV was identified as the closest homolog, but the search also indicated strong similarity to the Nsp1 proteins from transmissible gastroenteritis virus (TGEV) and porcine epidemic diarrhea virus (PEDV). These *Alphacoronavirus* homologs superimposed with SARS-CoV-2 Nsp1$_{13-127}$ with root-mean-square deviation (RMSD) values of 2.8 Å and 2.9 Å over 114 and 118 Cα atoms, respectively (Table 2) (Shen et al., 2019). Such appreciable levels of structural conservation, despite the low levels (<15%) of shared sequence identity, between these proteins highlight the plasticity of the unique protein fold found in coronaviral Nsp1 proteins (Figure 3B).

To further examine the plasticity of this fold, we examined the predicted secondary structure of Nsp1 proteins from other coronaviruses known to infect humans (MERS, OC43, HKU1, 229E, NL63). These Nsp1 proteins share minimal primary sequence identity with Nsp1 from SARS-CoV-2, and a BLASTP search failed to identify them as homologs (Table S1). Despite the low primary sequence identity, we were able to identify topological fingerprints, regions comprising 6–7 β-strands and at least one α-helix, in each of the Nsp1 sequences that are indicative of an arrangement capable of forming the capped β-barrel motif observed in the SARS-CoV-2 Nsp1$_{13-127}$ crystal structure (Figure S1). We then modeled the region of MERS-CoV Nsp1 identified through this approach using our crystal structure as the threading template. This resulted in a high-scoring (C-score = 0.48) homology model featuring the capped β-barrel motif characteristic of the structure of SARS-CoV-2 Nsp1$_{13-127}$ (Figure 3C). Mapping the modest primary sequence conservation

Table 2. Structural homologs of SARS-Cov-2 Nsp1$_{13-127}$

PDB code	DALI *Z* score (RMSD)	% Identity	Protein name	Organism
2GDT	16.4 (1.6)	88	SARS-CoV Nsp1	Severe acute respiratory syndrome virus
6IVC	7.9 (2.8)	14	TGEV Nsp1	Transmissible gastroenteritis virus
5XBC	7.7 (2.9)	15	PEDV Nsp1	Porcine epidemic diarrhea virus

between MERS-CoV and SARS-CoV-2 onto the homology model revealed it to be distributed throughout the structure (Figure 3C). The β-barrel in the MERS-CoV Nsp1 homology model retains the hydrophobic core observed in SARS-CoV-2, highlighting a level of functional conservation between the two proteins that may not have been predicted through sequence-based comparisons alone. At the C-terminus of the model is the R146/K147 dipeptide (R124/K125 in SARS-CoV-2) that has been shown to be involved in Nsp1-mediated RNA cleavage (Terada et al., 2017).

Nsp1$_{13-127}$ structure in context of the full-length SARS-CoV-2 Nsp1

Recent cryogenic electron microscopy (cryo-EM) data of SARS-CoV-2 Nsp1 in complex with the 40S ribosome were able to resolve the structure of the C-terminal portion of Nsp1 at a resolution of 2.6 Å (Thoms et al., 2020) and 2.8 Å (Schubert et al., 2020). These studies confirm the presence of a C-terminal α-helix spanning residues S166 to N178 and revealed a second helix spanning residues P153 to N160 that was not predicted bioinformatically. The resulting helix-turn-helix motif interacts with the ribosome and anchors Nsp1 to the pre-initiation complex.

To place the crystal structure of Nsp1$_{13-127}$ in context of full-length protein, we analyzed the AA sequence using PSIPRED and DISPRRED3 (Buchan and Jones, 2019; Jones and Cozzetto, 2015). These tools predicted the N-terminal 12 residues outside of our structure containing a disordered protein-binding region and a largely unstructured C-terminal domain with the exception of a single predicted α-helix. The presence of large disordered regions adjacent to the globular domain of Nsp1 provides strong rationale for why the full-length protein failed to crystallize.

Using our crystal structure data in combination with insights into the C-terminal domain of Nsp1, we generated a homology-based model of the full-length Nsp1 protein. This model maintains the capped β-barrel motif observed in our crystal structure and features a largely disordered linker between the end of the globular domain and the C-terminal helix-turn-helix. The C-terminal domain is shown as extending outward from the core of the structure at the bottom of the β-barrel, at the face opposite of where the α1 helix localizes; however, the disordered linker connecting this motif from the rest of the structure suggests that the position of this domain is likely to be highly dynamic and capable of multiple configurations (Figure 4). This predicted full-length structure aligns well with the partial density and model proposed by the recent cryo-EM structures of the SARS-CoV-2 Nsp1 C-terminus bound to the 40S ribosome (Thoms et al., 2020).

Discussion

SARS-CoV-2 Nsp1 arrests translation of host proteins through interaction with the 40S ribosome and in doing so suppresses the innate immune response to viral infection. Through interaction with SL1 in the viral 5′UTR, Nsp1 also distinguishes viral transcripts from host transcripts and allows for viral gene expression to continue unabated, highlighting the critical role this nonstructural protein plays in SARS-CoV-2 pathogenesis. Nsp1 proteins encoded by representatives of the *Coronaviridae* family feature significant AA sequence diversity, highlighted by the inability to identify homologs via BLAST of SARS-CoV-2 Nsp1 outside of the B lineage of betacoronaviruses. Leveraging our high-resolution structural data, we were able to identify additional structural homologs of SARS-CoV-2 Nsp1, Nsp1 of TGEV, and PEDV, with which it shares low levels of sequence identity. This result is in accordance with the recent data demonstrating that coronaviral Nsp1 proteins exhibit similar function in suppression of host translation (Terada et al., 2017). These observations suggest that the distinctive features of the SARS-CoV-2 Nsp1, highlighted by our structural data, may be critical to its function in self-recognition, via interaction with SL1, of viral RNAs and its role in promoting the degradation of host transcripts. This also provides insight into molecular features that could be expected in functionally similar proteins found in other important human pathogens, such as Nsp1 from MERS-CoV.

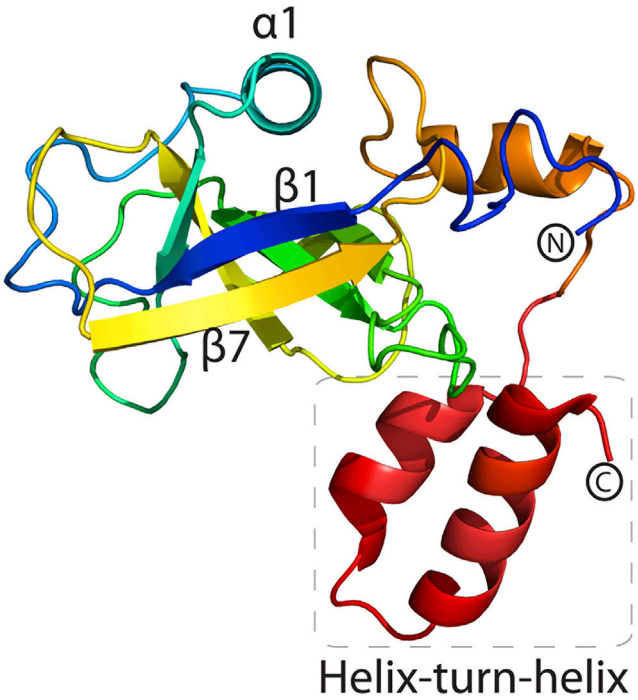


Figure 4. Model of full-length SARS-CoV-2 Nsp1

Cartoon representation of the model of the full-length SARS-CoV-2 Nsp1 colored from N-terminus (blue) to C-terminus (red). Distinctive secondary structure features are labeled.

Our sequence and structural based queries using SARS-CoV-2 Nsp1 were only able to identify homologs within the *Coronaviridae* family. This raises the possibility that the specific fold and functions of Nsp1 are unique to coronaviruses; however, the failure of sequence-based searches to identify Nsp1 from TGEV and PEDV as homologs of SARS-CoV-2 Nsp1 highlights the need for continued structural characterization of proteins involved in viral pathogenesis.

Limitations of the study

The crystal structure presented lacks the C-terminal domain of Nsp1 that has been shown to be critical and sufficient for facilitating interaction with the ribosome and inhibition of translation. Experimentally determining the full-length structure of Nsp1, alone or with an interaction partner, remains an important yet unfulfilled advancement to furthering our understanding of this essential component of SARS-CoV-2 infection.

Resource availability

Lead contact

Further information and requests for resources should be directed to the corresponding author, Alexei Savchenko (alexei.savchenko@ucalgary.ca)

Materials availability

Plasmids used in this study are available upon request.

Data and code availability

The structure of SARS-CoV-2 $Nsp1_{13\text{-}127}$ and all associated data have been deposited to the protein databank (www.rcsb.org) and are available under the accession code PDB: 7K3N.

Methods

All methods can be found in the accompanying Transparent methods supplemental file.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2020.101903.

Acknowledgments

We thank Changsoo Chang and the Structural Biology Center Team at APS for data collection. The structure presented was solved as part of the Center for Structural Genomics of Infectious Diseases (CSGID). This project has been funded in whole or in part with U.S. Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract Nos. HHSN272200700058C, HHSN272201200026C, and HHSN272201700060C.

Author contributions

Conceptualization, C.S. and A.S.; Investigation, C.S.; Formal Analysis, C.S. and N.W.; Writing – Original Draft, C.S. and A.S.; Writing – Review and Editing, C.S., N.W., and A.S.; Funding Acquisition – A.S.

Declaration of interests

The authors declare that they have no competing interests.

Received: October 8, 2020
Revised: October 21, 2020
Accepted: December 3, 2020
Published: January 22, 2021

References

Almeida, M.S., Johnson, M.A., Herrmann, T., Geralt, M., and Wuthrich, K. (2007). Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. J. Virol. *81*, 3151–3161.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and Mccammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U S A *98*, 10037–10041.

Buchan, D.W.A., and Jones, D.T. (2019). The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res. *47*, W402–W407.

Chan, J.F., Kok, K.H., Zhu, Z., Chu, H., To, K.K., Yuan, S., and Yuen, K.Y. (2020). 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.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. *20*, 533–534.

Harcourt, B.H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., Severson, K.M., Smith, C.M., Rota, P.A., and Baker, S.C. (2004). Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. J. Virol. *78*, 13600–13612.

Holm, L., and Laakso, L.M. (2016). Dali server update. Nucleic Acids Res. *44*, W351–W355.

Jones, D.T., and Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics *31*, 857–863.

Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K.G., and Makino, S. (2009). A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. Nat. Struct. Mol. Biol. *16*, 1134–1140.

Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., Kubo, H., and Makino, S. (2006). Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. Proc. Natl. Acad. Sci. U S A *103*, 12885–12890.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797.

Law, A.H., Lee, D.C., Cheung, B.K., Yim, H.C., and Lau, A.S. (2007). Role for nonstructural protein 1 of severe acute respiratory syndrome coronavirus in chemokine dysregulation. J. Virol. *81*, 416–422.

Littler, D.R., Gully, B.S., Colson, R.N., and Rossjohn, J. (2020). Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. iScience *23*, 101258.

Masters, P.S. (2006). The molecular biology of coronaviruses. Adv. Virus Res. *66*, 193–292.

Naqvi, A.A.T., Fatima, K., Mohammad, T., Fatima, U., Singh, I.K., Singh, A., Atif, S.M., Hariprasad, G., Hasan, G.M., and Hassan, M.I. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. Biochim. Biophys. Acta Mol. Basis Dis. *1866*, 165878.

Narayanan, K., Huang, C., Lokugamage, K., Kamitani, W., Ikegami, T., Tseng, C.T., and Makino, S. (2008). Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. J. Virol. *82*, 4471–4479.

Narayanan, K., Ramirez, S.I., Lokugamage, K.G., and Makino, S. (2015). Coronavirus nonstructural protein 1: common and distinct functions in the regulation of host and viral gene expression. Virus Res. *202*, 89–100.

Peng, Q., Peng, R., Yuan, B., Zhao, J., Wang, M., Wang, X., Wang, Q., Sun, Y., Fan, Z., Qi, J., et al. (2020). Structural and biochemical characterization of the nsp12-nsp7-nsp8 core polymerase complex from SARS-CoV-2. Cell Rep. *31*, 107774.

Schubert, K., Karousis, E.D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L.A., Leibundgut, M., Thiel, V., Muhlemann, O., and Ban, N. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. Nat. Struct. Mol. Biol. *27*, 959–966.

Shen, Z., Wang, G., Yang, Y., Shi, J., Fang, L., Li, F., Xiao, S., Fu, Z.F., and Peng, G. (2019). A conserved region of nonstructural protein 1 from alphacoronaviruses inhibits host gene expression and is critical for viral virulence. J. Biol. Chem. *294*, 13606–13618.

Tanaka, T., Kamitani, W., Dediego, M.L., Enjuanes, L., and Matsuura, Y. (2012). Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA. J. Virol. *86*, 11128–11137.

Terada, Y., Kawachi, K., Matsuura, Y., and Kamitani, W. (2017). MERS coronavirus nsp1

participates in an efficient propagation through a specific interaction with viral RNA. Virology *511*, 95–105.

Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., Kratzat, H., Hayn, M., Mackens-Kiani, T., Cheng, J., et al. (2020). Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. Science *369*, 1249–1255.

Wathelet, M.G., Orr, M., Frieman, M.B., and Baric, R.S. (2007). Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. J. Virol. *81*, 11620–11633.

Weiss, S.R., and Navas-Martin, S. (2005). Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiol. Mol. Biol. Rev. *69*, 635–664.

Yang, Y., Liu, C., Du, L., Jiang, S., Shi, Z., Baric, R.S., and Li, F. (2015). Two mutations were critical for bat-to-human transmission of Middle East respiratory syndrome coronavirus. J. Virol. *89*, 9119–9123.

Ye, Q., Wang, B., and Mao, J. (2020). The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. J. Infect. *80*, 607–613.

Zhang, T., Wu, Q., and Zhang, Z. (2020a). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. Curr. Biol. *30*, 1346–1351.e2.

Zhang, W.F., Stephen, P., Theriault, J.F., Wang, R., and Lin, S.X. (2020b). Novel coronavirus polymerase and nucleotidyl-transferase structures: potential to target new outbreaks. J. Phys. Chem. Lett. *11*, 4430–4435.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579*, 270–273.

Zust, R., Cervantes-Barragan, L., Kuri, T., Blakqori, G., Weber, F., Ludewig, B., and Thiel, V. (2007). Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines. PLoS Pathog. *3*, e109.

iScience, Volume *24*

Supplemental Information

Structural characterization of nonstructural protein 1 from SARS-CoV-2

Cameron Semper, Nobuhiko Watanabe, and Alexei Savchenko

Supplemental Information for Semper *et al*

Transparent Methods

Cloning

The portion of Orf1a encoding full-length Nsp1 was codon-optimized for *E. coli* expression, synthesized (Codex DNA) and cloned into the pMCSG53 expression vector at the SspI site via Gibson assembly.

>**cDNA sequence for SARS-CoV-2 Nsp1**
ATGGAAAGTCTGGTACCTGGGTTCAACGAGAAAACCCATGTTCAGCTGAGTTTACCG
GTTCTGCAAGTTCGTGATGTTCTGGTTCGCGGTTTTGGTGATAGCGTTGAAGAAGTTC
TGAGTGAAGCACGTCAACACCTGAAAGATGGTACATGCGGCTTAGTGGAAGTGGAA
AAAGGTGTTTTACCTCAGCTGGAACAGCCGTACGTGTTCATTAAACGCAGCGATGCA
AGAACAGCACCTCATGGTCATGTTATGGTTGAACTGGTGGCAGAACTGGAAGGTAT
CCAGTATGGTAGATCTGGTGAAACACTGGGTGTTTTAGTTCCGCATGTGGGCGAAAT
TCCTGTGGCATACCGTAAAGTGCTGCTGCGTAAAAATGGCAATAAAGGTGCAGGTG
GTCACAGCTATGGTGCCGATCTGAAAAGCTTTGATCTGGGCGATGAATTAGGTACAG
ATCCGTATGAGGACTTCCAGGAAAACTGGAACACCAAGCATAGTAGCGGTGTTACC
CGTGAATTAATGCGCGAACTGAATGGTGGTTAA

The fragment encompassing amino acid residues 13-127 of Nsp1 was PCR amplified using Phusion polymerase (NEB) and cloned into the same vector via ligation-independent cloning.

Primers for amplification of Nsp1$_{13\text{-}127}$

FWD - TACTTCCAATCCAATGCCCATGTTCAGCTGAGTTTACCG
REV - TTATCCACTTCCAATGTTAGCCATTTTTACGCAGCAG

Expression and purification of SARS-CoV2 Nsp1

Plasmids were transformed into the *E. coli* strain BL21(DE3)-Gold for protein expression. The same procedure was used for purification of full-length Nsp1 and Nsp1$_{13\text{-}127}$. Cells were grown at 37°C and 200 rpm to an OD600 of 0.8, cooled to 20°C then induced with 1 mM IPTG and incubated for 16 hours. Cells were harvested via centrifugation at 5000 x *g*, resuspended in

binding buffer (300 mM NaCl, 50 mM HEPES pH 7.5, 5 mM imidazole, 5% glycerol) and lysed via sonication. Lysates were centrifuged at 20,000 x *g* for 45 minutes at 4°C and the supernatant was incubated with Ni-NTA resin and rotated for 1 hour at 4°C. Nsp1 and $Nsp1_{13\text{-}127}$ were eluted in elution buffer (300 mM NaCl, 50 mM HEPES pH 7.5, 250 mM imidazole, 5% glycerol) then incubated with Tobacco-etch virus (TEV) protease overnight to cleave the N-terminal polyhistidine tag while dialyzing to remove imidazole. The proteins were then passaged over a second Ni-NTA to remove impurities. $Nsp1_{13\text{-}127}$ was immediately dialyzed into precrystallization buffer (300 mM NaCl, 10 mM HEPES pH 7.5), while full-length Nsp1was further purified via gel filtration using a Superdex75 column (300 mM NaCl, 10 mM HEPES pH 7.5).

Crystallization

Crystals of $Nsp1_{13\text{-}127}$ were grown at 298 K in 0.2 M sodium formate, 20% PEG3350 via the vapour diffusion sitting-drop method. Prior to data collection the crystals were soaked in 0.2 M sodium formate, 20% PEG3350, 30% glycerol and flash frozen in liquid nitrogen.

Data collection, structure determination and refinement

X-ray diffraction data of crystals of $Nsp1_{13\text{-}127}$ were collected at Advanced Photon Source Beamline 19 ID (Argonne, Illinoi USA) under cryo-stream at 93.15 K. Diffraction data were processed with HKL3000 suit (Minor et al., 2006). Initial phase for $Nsp1_{13\text{-}127}$ was obtained by molecular replacement with Molrep using NMR structure of SARS-CoV Nsp1 (PDB: 2HSX) as a search model (Winn et al., 2011). Subsequently, the initial electron density map was improved though density modification by parrot and the model was built using buccaneer (Winn et al., 2011). Final model was produced by the cycle of manual model building and refinement using Phenix.refine and Coot (Adams et al., 2010, Emsley and Cowtan, 2004). All geometry was

verified using Phenix validation tools (Ramachandran statistics: Favoured (97.1%), additionally allowed (2.9%), disallowed (0.0%)) and the wwPDB server.

Structure analysis and homology-modeling

Electrostatic potential surfaces were calculated using Pymol using APBS(Baker et al., 2001). Structure similarity searches of the Protein Data Bank were performed using the Dali server (Holm and Laakso, 2016).Secondary structure prediction of Nsp1 homologues was done using PSIPRED (Buchan and Jones, 2019). Homology modeling of MERS-CoV was performed by I-TASSER using the crystal structure of SARS-CoV-2 $Nsp1_{13-127}$ as the threading template (Yang and Zhang, 2015, Zhang et al., 2017). The model produced by I-TASSER was subjected to further energy minimization using the YASARA energy minimization server to improve stereochemical property (Krieger et al., 2009). Stereochemical properties of the final model was validated with Ramachandran plot using PROCHECK server and 90% of residues are in the favoured/additionally allowed region (Laskowksi, 1993). The model of the full-length SARS-CoV-2 Nsp1 was generated using the Robetta structural predication server with the SARS-CoV-2 $Nsp1_{13-127}$ structure defined as the template (Song et al., 2013).

Methods References

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-21.

BAKER, N. A., SEPT, D., JOSEPH, S., HOLST, M. J. & MCCAMMON, J. A. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A,* 98**,** 10037-41.

BUCHAN, D. W. A. & JONES, D. T. 2019. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res,* 47**,** W402-W407.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr,* 60**,** 2126-32.

HOLM, L. & LAAKSO, L. M. 2016. Dali server update. *Nucleic Acids Res,* 44**,** W351-5.
KRIEGER, E., JOO, K., LEE, J., LEE, J., RAMAN, S., THOMPSON, J., TYKA, M., BAKER, D. & KARPLUS, K. 2009. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins,* 77 Suppl 9**,** 114-22.
LASKOWKSI, R. A., MACARTHUR, M.W., MOSS, D. S, THORNTON, J.M. 1993. PROCHECK - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26**,** 283-291.
MINOR, W., CYMBOROWSKI, M., OTWINOWSKI, Z. & CHRUSZCZ, M. 2006. HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr,* 62**,** 859-66.
SONG, Y., DIMAIO, F., WANG, R. Y., KIM, D., MILES, C., BRUNETTE, T., THOMPSON, J. & BAKER, D. 2013. High-resolution comparative modeling with RosettaCM. *Structure,* 21**,** 1735-42.
WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISSINEL, E. B., LESLIE, A. G., MCCOY, A., MCNICHOLAS, S. J., MURSHUDOV, G. N., PANNU, N. S., POTTERTON, E. A., POWELL, H. R., READ, R. J., VAGIN, A. & WILSON, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr,* 67**,** 235-42.
YANG, J. & ZHANG, Y. 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res,* 43**,** W174-81.
ZHANG, C., FREDDOLINO, P. L. & ZHANG, Y. 2017. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Res,* 45**,** W291-W299.

Table S1. Amino acid sequence similarity to SARS-CoV-2 Nsp1 of Nsp1 proteins from human pathogenic coronaviruses, related to Figure 3

Protein Name	Organism	% Identity to SARS-CoV-2 Nsp1
MERS-CoV Nsp1	Middle East respiratory syndrome-related coronavirus	19
HCoV-OC43 Nsp1	Human coronavirus OC43	19
HCoV-HKU1 Nsp1	Human coronavirus HKU1	18
HCoV-229E Nsp1	Human coronavirus 229E	9
HCoV-NL63 Nsp1	Human coronavirus NL63	6

Fig S1. Secondary structure analysis of Nsp1 from human-infecting Coronaviruses, related to Figure 3. Within each sequence, the topological fingerprint containing elements that may facilitate formation of the capped β-barrel structure are coloured, while sequence outside of the fingerprint region is depicted in light grey. Key secondary structure elements observed in the SARS-CoV-2 $Nsp1_{13\text{-}127}$ crystal structure are labeled and are coloured accordingly in the ortholog sequences. Unique secondary structure features that lack an equivalent in the SARS-CoV-2 structure but that fall within the predicted β-barrel region are coloured as dark grey.