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

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

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

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

*To whom correspondence should be addressed.

Alexei Savchenko
Address: HSC B724 3330 Hospital Drive NW, Calgary, AB
Phone: 403-210-7980
Email: alexei.savchenko@ucalgary.ca

1 **<u>Abstract</u>**

2

3 Severe acute respiratory syndrome (SARS) coronavirus-2 (SARS-CoV-2) is a single-stranded,

4 enveloped RNA virus and the etiological agent of the current COVID-19 pandemic. Efficient

5 replication of the virus relies on the activity of nonstructural protein 1 (Nsp1), a major virulence

6 factor shown to facilitate suppression of host gene expression through promotion of host mRNA

7 degradation and interaction with the 40S ribosomal subunit. Here, we report the crystal structure

8 of the globular domain of SARS-CoV-2 Nsp1, encompassing residues 13 to 127, at a resolution

9 of 1.65 Å. Our structure features a six-stranded, capped β-barrel motif similar to Nsp1from

10 SARS-CoV and reveals how variations in amino acid sequence manifest as distinct structural

11 features. Through comparative analysis of structural homologues, we identified a topological

12 signature associated with this protein fold that facilitated modeling of Nsp1 from MERS-CoV.

13 Combining our high-resolution crystal structure with existing data on the C-terminus of Nsp1

14 from SARS-CoV-2, we propose a model of the full-length protein. Our results provide

15 unparalleled insight into the molecular structure of a major pathogenic determinant of SARS-

16 CoV-2.

1   **Introduction**
2
3   In March of 2020, the World Health Organization (WHO) declared Coronavirus Disease

4   2019 (COVID-19) a global pandemic. As of September 2020, there have been more than

5   26,000,00 cases of infection reported globally and approximately 860,000 deaths attributed to

6   COVID-19 [1]. The etiological agent of this pandemic has been identified as Severe Acute

7   Respiratory Syndrome coronavirus-2 (SARS-CoV-2), a member of the *Betacoronavirus* genus

8   and closely related to the SARS-CoV that caused the SARS outbreak of 2002-2004 [2, 3].

9   Coronaviruses infect a diverse array of vertebrates, with infection typically resulting in

10   respiratory disease or gastroenteritis [4]. Their broad host range allows for a substantial

11   reservoir for human infection and mutation facilitates cross-species transmission [5]. There is

12   mounting evidence that SARS-CoV-2 originated via mutation and cross-species transmission of

13   a pangolin coronavirus with which it shares ~97% sequence identity [6]. This highlights the

14   dramatic impact that relatively few amino acid substitutions can have in coronaviruses and

15   underscores the urgent need for characterization of these infectious agents at a molecular level.

16   SARS-CoV-2 is an enveloped, positive-sense RNA virus with a single-stranded genome

17   approximately 30 kb in size. The genome is 5´-capped and 3´poly-adenylated and the first ~2/3

18   of the genome encode two overlapping reading frames that produce polyprotein 1a and 1ab [7].

19   Downstream of this region, the remaining 1/3 of the genome encodes for structural proteins and a

20   number of ORFs that produce accessory proteins largely of unknown function. Polyprotein 1a

21   and 1ab are large polypeptides that are processed post-translationally by viral-encoded proteases

22   to produce non-structural proteins (nsp) 1-16 [8]. The genome also contains 5´ and 3´

23   untranslated regions (UTRs), the former of which plays a critical role in self-recognition that

24   allows for SARS-CoV-2 protein production to occur unabated while host gene expression is

1 suppressed [9]. At the core of this mechanism is Nsp1, a 180 amino acid (AA) protein produced

2 via processing of polypeptide 1a and 1ab by the Papain-like protease domain of Nsp3 (Figure

3 1A) [10].

4   Nsp1 mediates a two-pronged approach to suppression of host gene expression. Firstly, it

5 inhibits translation of host proteins during the initiation stage through interaction with the 40S

6 ribosomal subunit. Secondly, Nsp1 promotes the degradation of host mRNA by endonucleolytic

7 cleavage within the 5´UTR, which in turn leads to accelerated Xrn1-mediated mRNA decay [11,

8 12]. Viral mRNAs are able to avoid the fate of host mRNAs through interaction between Nsp1

9 and the stem-loop 1 (SL1) motif found in the viral 5´ UTR [9]. Nsp1 from SARS-CoV has been

10 shown to inhibit the Type I Interferon response in infected cells, allowing for the virus to

11 circumvent the innate immune response [13]. Expression of SARS-CoV Nsp1 has also been

12 shown to induce the production of chemokines, suggesting this protein may play a role in the

13 "cytokine storm", a maladaptive release of cytokines in response to infection, associated with a

14 number of COVID-19 infections [14, 15]. Thus, Nsp1 has emerged as a major pathogenicity

15 factor that plays a critical role in the coronavirus infection cycle [16]. Deletion or mutation of

16 nsp1 results in viral attenuation in infection models and restores the innate immune response in

17 infected cells [17]. Based on its central role in suppression of the host immune response and

18 essentiality to infection, Nsp1 has been proposed as a therapeutic target for the treatment of

19 COVID-19 [16]. A prerequisite to any investigation into possible interventions targeting Nsp1 is

20 high-resolution structural data that can facilitate robust *in silico* screening. A partial structure

21 corresponding to the N-terminal fragment of Nsp1 from SARS-CoV was resolved by NMR

22 spectroscopy and revealed a unique β-barrel motif. The corresponding domain from SARS-

23 CoV-2 shares 86% sequence identity with its SARS-CoV ortholog, which is a considerable level

4

1 of sequence diversity compared to many other non-structural proteins encoded by the SARS-

2 CoV-2 genome (e.g. Nsp12 shares 96% identity with SARS-CoV ortholog [18]). This highlights

3 the need for a pursuit of structural characterization of SARS-CoV-2 proteins.

4   Here, we report the high-resolution crystal structure of the globular N-terminal domain of

5 Nsp1 from SARS-CoV-2 at 1.65 Å resolution. Our data reveals a high level of structural

6 conservation between Nsp1 of SARS-CoV-2 and SARS-CoV, but also some unique structural

7 features that likely contribute to increased stability of the β-barrel fold in SARS-CoV-2 Nsp1.

8 Comparative analysis reveals additional structural homologues in Nsp1 proteins from

9 *Alphacoronaviruses*, despite low levels of shared sequence identity. These results highlight the

10 critical role this unique protein fold plays in facilitating viral infection and suppression of host

11 gene expression.

12

13 **Results**

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

15   In pursuit of structural characterization of Nsp1 from SARS-CoV-2 we expressed a

16 codon optimised version of the ORF encoding this protein in *E. coli*. Using this expression

17 system, we obtained purified full-length Nsp1 after a two-step purification protocol (see details

18 in Material and Methods). However, we were unable to crystallize the full-length Nsp1 protein,

19 likely due to the presence of flexibly disordered regions at the N- and C-termini.

20   Previous studies of Nsp1 from SARS-CoV reported the presence of a distinct globular

21 domain comprised of residues 13 to 127 [19]. To structurally characterise the N-terminal

22 fragment of Nsp1 from SARS-CoV-2 (Nsp1$_{13\text{-}127}$), we sub-cloned it and expressed it in *E. coli*.

23 The expression level of Nsp1$_{13\text{-}127}$ was comparable to that of the full-length protein and with this
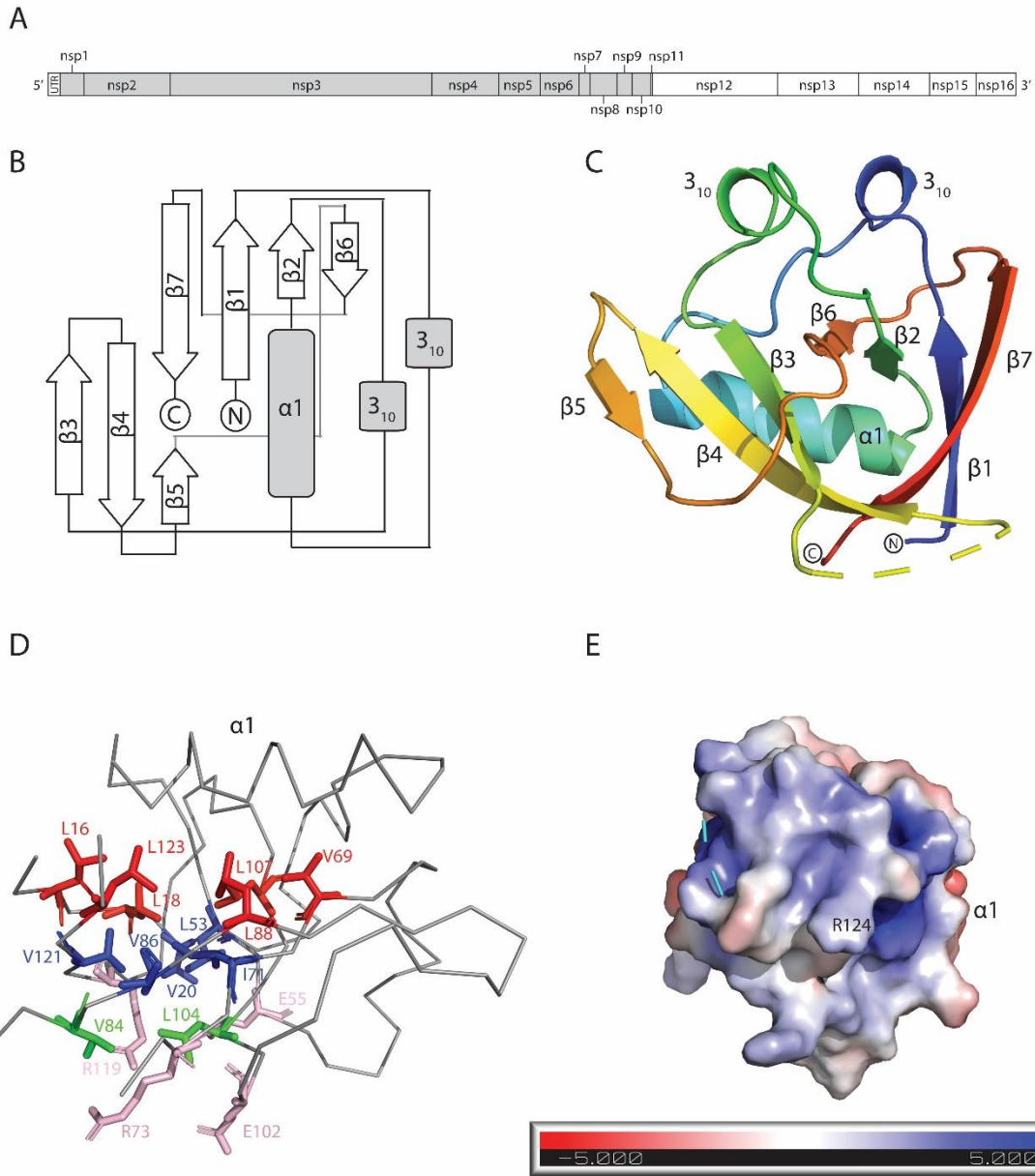
Figure 1. The structure of SARS-CoV-2 Nsp1$_{13-127}$. (A) Schematic of Orf1ab from SARS-CoV-2. The grey shading encompasses the non-structural proteins produced by Orf1a. (B) Topological arrangement of SARS-CoV-2 Nsp1$_{13-127}$. (C) Cartoon representation of the SARS-CoV-2 Nsp1$_{13-127}$ structure coloured 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 sidechains that contribute to the hydrophobic core of the beta barrel shown in stick representation. The three layers of side chains, plus the charged residues at the bottom are coloured and the residues within the layers are labeled. (E) Electrostatic surface potential of the likely RNA-binding interface of SARS-CoV-2 Nsp1$_{13-127}$. R124 which is critical for the interaction with SL1 of the 5´UTR is labeled.

6

1   fragment we obtained well-diffracting crystals in several conditions that enabled structural

2   determination of this portion of Nsp1.

3   The crystal structure of SARS-CoV-2 $Nsp1_{13-127}$ was determined via molecular replacement

4   using the SARS-CoV Nsp1(PDB: 2HSX) solution structure as the search model and refined to a

5   resolution of 1.65 Å (Table 1). The asymmetric unit contained a single molecule of $Nsp1_{13-127}$.

6   Analysis of the $Nsp1_{13-127}$ structure by PDBePISA revealed the protein has a total solvent-

7   exposed area of 6102 $Å^2$ [20].

8   As in the case of SARS-CoV, the structure of SARS-CoV-2 $Nsp1_{13-127}$ features a unique

9   topological arrangement resulting in the formation of a six stranded (n = 6) β-barrel that is

10 primarily antiparallel, with the exception of strands β1 (Q15 – V20) and β2 (C51 – V54) (Figure

11 1B). Additional major structural features include α1 (V35 – D48) helix, which is positioned as a

12 cap along one opening of the β-barrel, two 3_{10} helices that run parallel to each other and the β5

13 strand (I95 – Y97) which is not part of the β-barrel but forms a β-sheet interaction with the β4

14 strand (V84 – L92) (Figure 1C).

15 The core of the β-barrel is highly hydrophobic and is mainly comprised of the side chains of

16 thirteen amino acids organized into three layers. The first layer, which is adjacent to α1 helix, is

17 formed by sidechains of residues L16, L18, V69, L88, L107 and L123 (Figure 1D). The opening

18 of the β-barrel at this layer is obstructed by the α1 helix, which contributes the side chain of L46

19 to the centre of this first layer of core hydrophobic residues. The middle layer of the β-barrel is

20 comprised of the side chains of residues V20, L53, I71, V86 and V121, while the bottom layer

21 features residues V84 and L104. Adjacent to this bottom layer are four solvent-exposed

22 residues; E55, R73, E102 and R119 whose sidechains point inward towards the core of the β-

23

7

1 **Table 1. X-ray crystallographic statistics**

	SARS-Cov-2 Nsp1$_{13-127}$
Data collection	
Space group	$P4_{3}2_{1}2$
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}[1]	0.058 (1.091)
R_{pim}[2]	0.020 (0.387)
I / sI	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	12884
R_{work} / R_{free}[3]	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 (%)	
Favoured	97.1
Allowed	2.9
Disallowed	0

*Values in brackets refer to highest resolution shells.

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

2. $R_{\mathrm{pim}} = \Sigma_{hkl}\sqrt{(1/n-1)}\ \Sigma^{n}_{j=1}|I_{hkl.j} - \langle I_{hkl}\rangle|/\Sigma_{hkl}\Sigma_{j}I_{hk.j.}$.

3. $R = \Sigma|F_{p}^{\mathrm{obs}} - F_{p}^{\mathrm{calc}}|/\Sigma F_{p}^{\mathrm{obs}}$, where F_{p}^{obs} and F_{p}^{calc} are the observed and calculated structure factor amplitudes, respectively.

1   barrel (Figure 1D). Consequently, the hydrophobic core at both ends of the β-barrel is occluded:

2   by the α1 helix on one side and the charged residues (E54, R73, E102, R199) on the other.

3   A distinctive feature of Nsp1 that is evident in the crystal structure of $Nsp1_{13-127}$ is the large

4   number of flexible loops. For one of these loops (A76 – H81), we were unable to resolve the

5   structure due to a lack of interpretable electron density. This region has been previously shown to

6   be highly flexible as evidenced by the multitude of backbone confirmations observed in the

7   NMR structure of SARS-CoV Nsp1 [19]. The positions of other loops in the crystal structure

8   were stabilized via interactions between specific secondary structure elements. Specifically, the

9   two 3_{10} helices are aligned in the crystal structure forming H-bonds between R24 and Q63. This

10   interaction appears to stabilize the position of two of the largest loops in the crystal structure, the

11   β1-α1 (L21 – S34) loop and the β2- β3 (E54 – P67) loop.

12   Electrostatic surface potential analysis of the SARS-CoV-2 $Nsp1_{13-127}$ structure revealed

13   several regions of charged residue colocalization [21]. The most electronegative and

14   electropositive patches appear adjacent on the $Nsp1_{13-127}$ surface, separated by a small

15   electroneutral region corresponding to the α1 helix. The electropositive patch continues around

16   the surface of the protein, where the R124 residue is localized (Figure 1E). This residue has

17   been shown to be essential for the interaction between Nsp1 and SL1 of the 5´ UTR, implicating

18   this charged surface in RNA binding [9].

19

20   ***Comparison with SARS-CoV Nsp1***

21   The overall fold of SARS-CoV-2 $Nsp1_{13-127}$ was highly similar to that of the corresponding

22   SARS-CoV Nsp1 fragment in accordance with the significant primary sequence identity shared

23   between the two proteins (Figure 2A). However, detailed comparative analysis revealed several

1   notable differences between the two structures. The SARS-CoV-2 structure of Nsp1$_{13-127}$

2   features two 3$_{10}$ helices, compared to the single one found in SARS-CoV Nsp1. The primary

3   sequence corresponding to the second 3$_{10}$ helix is completely conserved between SARS-CoV-2

4   and SARS-CoV Nsp1 proteins, suggesting this secondary structure element may be transient in

5   nature or influenced by differences elsewhere in SARS-CoV-2 Nsp1$_{13-127}$. The two 3$_{10}$ helices

6   interact with each other through H-bonds in SARS-CoV-2 Nsp1$_{13-127}$ but this interaction is

7   notably absent in the NMR structure of Nsp1 from SARS-CoV [19].

8   Two of the most striking differences between the two Nsp1 structures are the extension of β4

9   and the presence of an additional β-strand in the case of SARS-CoV-2 Nsp1$_{13-127}$. Amino acid

10   changes at positions 84 (K to V) and 85 (V to M) between SARS-CoV and SARS-CoV-2 Nsp1

11   result in an extension of β4 on the N-terminus by three residues (Figure 2B). This extension

12   results in an increased number of H-bonds formed between β4 and β7 in SARS-CoV-2 Nsp1

13   compared to the pairing between β4 and β6 that occurs in SARS-CoV Nsp1. This extension also

14   facilitates the formation of the β5 strand which is absent in SARS-CoV Nsp1. The β5 and β4

15   strands are part of a small β-sheet that stabilizes the loop region between β5 and β6 (β4 and β5 in

16   SARS-CoV Nsp1) (Figure 2B).

17   Comparison of the electrostatic surface potential of Nsp1$_{13-127}$ from SARS-CoV-2 and SARS-

18   CoV shows additional differences between the two proteins. The SARS-CoV structure features a

19   single patch of electronegative surface potential opposite a separate patch of electropositive

20   surface potential. These two charged surfaces are oriented 180° opposite one another and there

21   is a ring of electroneutral surface potential separating the two charged surfaces (Figure 2C). The

22   SARS-CoV-2 Nsp1$_{13-127}$ structure features an equivalent patch of electronegativity, but the

23   electropositive surface is divided into two by a patch of mild electronegative surface potential.

10

1
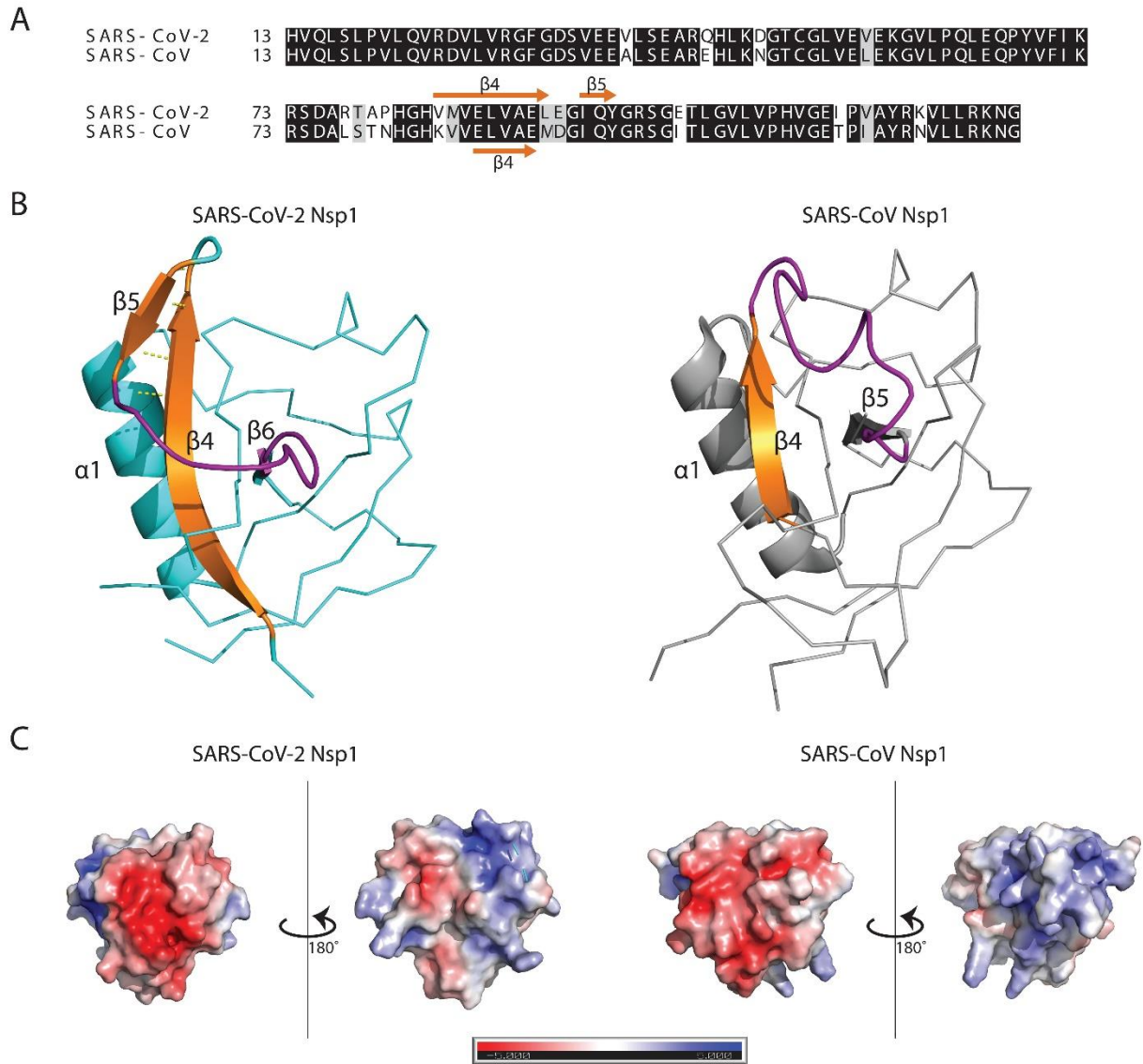
Figure 2. Comparison of SARS-CoV-2 Nsp1$_{13-127}$ to SARS-CoV Nsp1. (A) Alignment of
residues 13-127 of Nsp1 with conserved residues highlighted in black. (B) Ribbon
representation of SARS-CoV-2 and SARS-CoV Nsp1$_{13-127}$. β4 (shown in cartoon representation)
is extended in the SARS-CoV-2 structure allowing for H-bonding with and formation of β5,
which is absent in the SARS-CoV Nsp1structure. The loop region (and equivalent in SARS-CoV
Nsp1) stabilized through interaction between β4 and β5 is coloured in purple. (C) Comparison
of electrostatic surface potential of SARS-CoV-2 and SARS-CoV Nsp1$_{13-127}$.

11

1   Mapping the non-conserved residues onto the SARS-CoV-2 $Nsp1_{13-127}$ structure revealed that

2   most of the differences are in solvent-exposed residues. Distinct patches of variation are

3   observed immediately adjacent to both ends of β4 and in the loop between β3 and β4 (S74 –

4   H83), where 4 of 9 residues differ between SARS-CoV-2 and SARS-CoV (Figure 3A). The

5   remaining sequence variation is broadly distributed across the $Nsp1_{13-127}$ structure. Residues that

6   contribute to the hydrophobic core of the β -barrel are conserved between SARS-CoV-2 and

7   SARS-CoV, suggesting this particular organization is critical to the function of Nsp1.

8

9   *Comparison with other structural homologues*

10   Amino acid sequence-based searches for homologues of Nsp1 reveal a high level of

11   sequence conservation among representatives of Lineage B of *Betacoronaviruses,* which

12   includes several animal viruses closely related to SARS-CoV and SARS-CoV-2. Despite the

13   presence of functionally equivalent Nsp1 proteins, no sequence-based homologues of SARS-

14   CoV-2 Nsp1 were detected in other lineages of *Betacoronaviruses.* This lack of sequence

15   conservation is observed even when comparing SARS-CoV-2 Nsp1 to other coronaviruses

16   known to infect humans, including MERS, HCoV-OC43 and HCoV-HKU1.

17   A search for structural homologues using our high resolution structure of SARS-Cov-2

18    $Nsp1_{13-127}$ identified three closely related structural homologues, all annotated as Nsp1 proteins

19   (Table 2) [22]. Unsurprisingly, the Nsp1 structure from SARS-CoV was identified as the closest

20   homologue, but the search also indicated strong similarity to the Nsp1 proteins from

21   transmissible gastroenteritis virus (TGEV) and porcine epidemic diarrhea virus (PEDV). These

22   *Alphacoronavirus* homologues superimposed with SARS-CoV-2 $Nsp1_{13-127}$ with RMSD values

23   of 2.8 Å and 2.9 Å over 114 and 118 Cα atoms, respectively (Table 2) [23]. Such appreciable
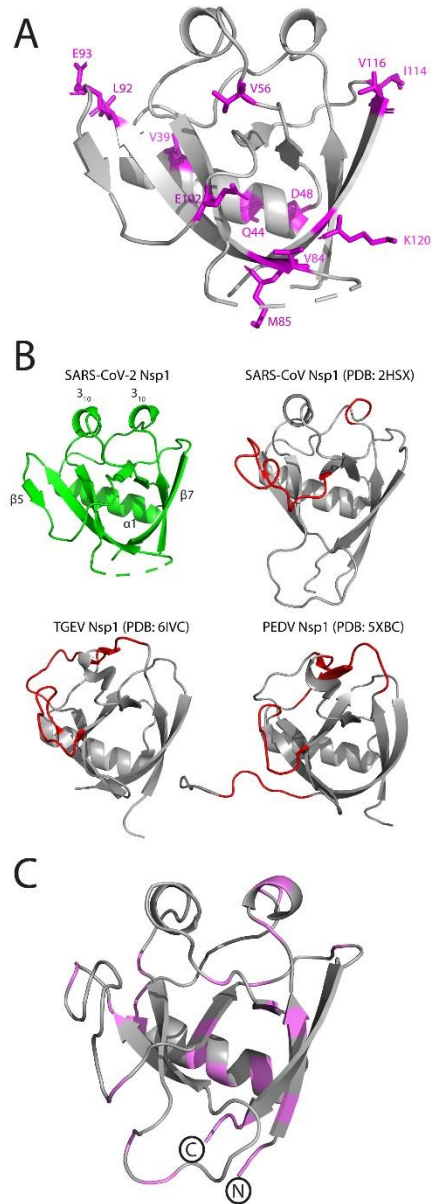
Figure 3. Comparison of the structure of SARS-CoV-2 Nsp1$_{13\text{-}127}$ to structural homologues.
(A) Cartoon representation of the SARS-CoV-2 Nsp1$_{13\text{-}127}$ structure with non-conserved residues between SARS-CoV-2 and SARS-CoV highlighted in magenta. (B) SARS-CoV-2 Nsp1$_{13\text{-}127}$ is shown in green, SARS-CoV Nsp1 (PDB: 2HSX), TGEV Nsp1 (PDB: 6IVC) and PEDV Nsp1 (PDB: 5XBC) are shown in grey with noticeable structural differences (compared to SARS-CoV-2) highlighted in red. Distinctive secondary structure features from the SARS-CoV-2 Nsp1$_{13\text{-}127}$ 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.

13

1   levels of structural conservation, despite the low levels (<15%) of shared sequence identity,

2   between these proteins highlights the plasticity of the unique protein fold found in Coronaviral

3   Nsp1 proteins (Figure 3B).

4   To further examine the plasticity of this fold, we examined the predicted secondary

5   structure of Nsp1 proteins from other coronaviruses known to infect humans (MERS, OC43,

6   HKU1, 229E, NL63). These Nsp1 proteins share minimal primary sequence identity with Nsp1

7   from SARS-CoV-2 and a BLASTP search failed to identify them as homologues (Table S1).

8   Despite the low primary sequence identity, we were able to identify topological fingerprints,

9   regions comprised of 6-7 β-strands and at least one α-helix, in each of the Nsp1 sequences that

10   are indicative of an arrangement capable of forming the capped β-barrel motif observed in the

11   SARS-CoV-2 Nsp1$_{13-127}$ crystal structure (Figure S1). We then modeled the region of MERS-

12   CoV Nsp1 identified through this approach using our crystal structure as the threading template.

13   This resulted in a high-scoring (C-score = 0.48) homology model featuring the capped β-barrel

14   motif characteristic of the structure of SARS-CoV-2 Nsp1$_{13-127}$ (Figure 3C). Mapping the modest

15   primary sequence conservation between MERS-CoV and SARS-CoV-2 onto the homology

16   model revealed it to be distributed throughout the structure (Figure 3C). The β-barrel in the

17   MERS-CoV Nsp1 homology model retains the hydrophobic core observed in SARS-CoV-2,

18   highlighting a level of functional conservation between the two proteins that may not have been

19   predicted through sequence-based comparisons alone. At the C-terminus of the model is the

20   R146/K147 dipeptide (R124/K125 in SARS-CoV-2) that has been shown to be involved in

21   Nsp1-mediated RNA cleavage [24].

22

23

1   **Table 2. Structural homologues of SARS-Cov-2 Nsp1$_{13\text{-}127}$**

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

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

2   Recent cryo-EM data of SARS-CoV-2 Nsp1 in complex with the 40S ribosome were able

3   to resolve the structure of the C-terminal portion of Nsp1 at a resolution of 2.6 Å [25]. This study

4   confirms the presence of a C-terminal α-helix spanning residues S166 to G179 and revealed a

5   second helix spanning residues Y154 to N160 that was not predicted bioinformatically. The

6   resulting helix-turn-helix motif interacts with the ribosome and anchors Nsp1 to the pre-initiation

7   complex.

8   To place the crystal structure of Nsp1$_{13-127}$ in context of full-length protein we analyzed

9   the amino acid sequence using PSIPRED and DISPRRED3 [26, 27]. These tools predicted the N-

10   terminal 12 residues outside of our structure contained a disordered protein-binding region and a

11   largely unstructured C-terminal domain with the exception of a single predicted α-helix. The

12   presence of large disordered regions adjacent to the globular domain of Nsp1 provide strong

13   rationale for why the full-length protein failed to crystallize.

14   Using our crystal structure data in combination with insights into the C-terminal domain

15   of Nsp1, we generated a homology-based model of the full-length Nsp1 protein. This model

16   maintains the capped β-barrel motif observed in our crystal structure and features a largely

17   disordered linker between the end of the globular domain and the C-terminal helix-turn-helix.

18   The C-terminal domain is shown as extending outward from the core of the structure at the

19   bottom of the β-barrel, at the face opposite of where the α1 helix localizes; however the

20   disordered linker connecting this motif from the rest of the structure suggests that the position of

21   this domain is likely to be highly dynamic and capable of multiple configurations (Figure 4).

22   This predicted full-length structure aligns well with the partial density and model proposed by

16

1   the recent cryo-EM structure of the SARS-CoV-2 Nsp1 C-terminus bound to the 40S ribosome

2   [25].

3

4   **<u>Discussion</u>**

5   SARS-CoV-2 Nsp1 arrests translation of host proteins through interaction with the 40S

6   ribosome and in doing so suppresses the innate immune response to viral infection. Through

7   interaction with SL1 in the viral 5´UTR, Nsp1 also distinguishes viral transcripts from host

8   transcripts and allows for viral gene expression to continue unabated, highlighting the critical

9   role this nonstructural protein plays in SARS-CoV-2 pathogenesis. Nsp1 proteins encoded by

10   representatives of the *Coronaviridae* family feature significant amino acid sequence diversity,

11   highlighted by the inability to identify homologues via BLAST of SARS-CoV-2 Nsp1 outside of

12   the B lineage of *Betacoronaviruses*. Leveraging our high-resolution structural data, we were

13   able to identify additional structural homologues of SARS-CoV-2 Nsp1, Nsp1 of TGEV and

14   PEDV, with which it shares low levels of sequence identity. This result is in accordance with the

15   recent data demonstrating that Coronaviral Nsp1 proteins exhibit similar function in suppression

16   of host translation [24]. These observations suggest that the distinctive features of the SARS-

17   CoV-2 Nsp1, highlighted by our structural data, may be critical to its function in self-

18   recognition, via interaction with SL1, of viral RNAs and its role in promoting the degradation of

19   host transcripts. This also provides insight into molecular features that could be expected in

20   functionally similar proteins found in other important human pathogens, such as Nsp1 from

21   MERS coronavirus.

22   Our sequence and structural based queries using SARS-CoV-2 Nsp1were only able to

23   identify homologues within the *Coronaviridae* family. This raises the possibility that the specific

17

1
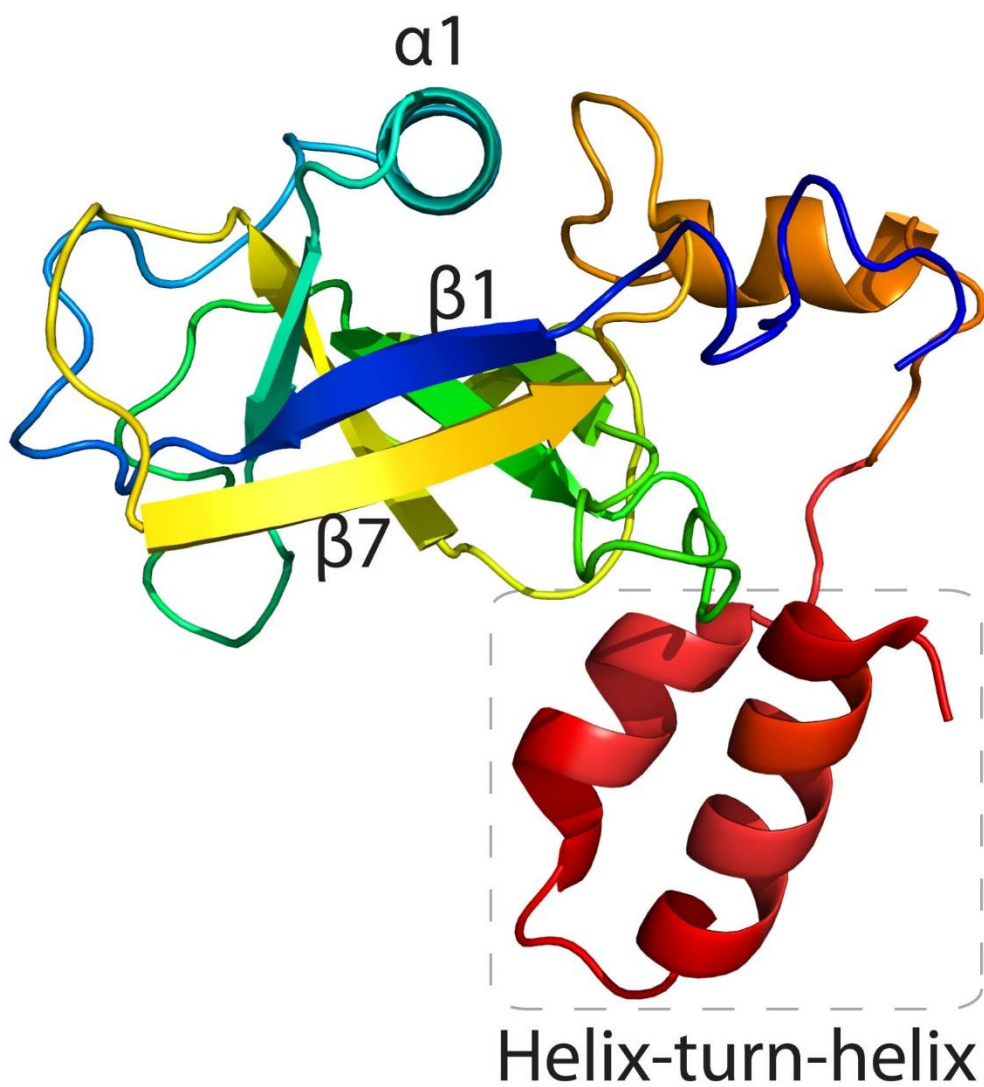
Figure 4. Model of full-length SARS-CoV-2 Nsp1. Cartoon representation of the model of the

full-length SARS-CoV-2 Nsp1 coloured from N-terminus (blue) to C-terminus (red). Distinctive

secondary structure features are labeled

5

6

7

1   fold and functions of Nsp1 are unique to Coronaviruses; however, the failure of sequence-based

2   searches to identify Nsp1 from TGEV and PEDV as homologues of SARS-CoV-2 Nsp1

3   highlights the need for continued structural characterization of proteins involved in viral

4   pathogenesis.

5

6   **Materials and Methods**

7   *Cloning, expression and purification of SARS-CoV2 Nsp1*

8   The portion of Orf1a encoding full-length Nsp1 was codon-optimized for *E. coli* expression,

9   synthesized (Codex DNA) and cloned into the pMCSG53 expression vector at the SspI site via

10   Gibson assembly. The fragment encompassing amino acid residues 13-127 of Nsp1 was PCR

11   amplified and cloned into the same vector via ligation-independent cloning. Plasmids were

12   transformed into the *E. coli* strain BL21(DE3)-Gold for protein expression. The same procedure

13   was used for purification of full-length Nsp1 and $Nsp1_{13-127}$. Cells were grown at 37°C and 200

14   rpm to an OD600 of 0.8, cooled to 20°C then induced with 1 mM IPTG and incubated for 16

15   hours. Cells were harvested via centrifugation at 5000 x g, resuspended in binding buffer (300

16   mM NaCl, 50 mM HEPES pH 7.5, 5 mM imidazole, 5% glycerol) and lysed via sonication.

17   Lysates were centrifuged at 20,000 x g for 45 minutes at 4°C and the supernatant was incubated

18   with Ni-NTA resin and rotated for 1 hour at 4°C. Nsp1 and $Nsp1_{13-127}$ were eluted in elution

19   buffer (300 mM NaCl, 50 mM HEPES pH 7.5, 250 mM imidazole, 5% glycerol) then incubated

20   with Tobacco-etch virus (TEV) protease overnight to cleave the N-terminal polyhistidine tag

21   while dialyzing to remove imidazole. The proteins were then passaged over a second Ni-NTA to

22   remove impurities. $Nsp1_{13-127}$ was immediately dialyzed into precrystallization buffer (300 mM

19

1   NaCl, 10 mM HEPES pH 7.5), while full-length Nsp1was further purified via gel filtration using

2   a Superdex75 column (300 mM NaCl, 10 mM HEPES pH 7.5).

3

4   ***Crystallization***

5   Crystals of $Nsp1_{13-127}$ were grown at 298 K in 0.2 M sodium formate, 20% PEG3350 via the

6   vapour diffusion sitting-drop method. Prior to data collection the crystals were soaked in 0.2 M

7   sodium formate, 20% PEG3350, 30% glycerol and flash frozen in liquid nitrogen.

8

9   ***Data collection, structure determination and refinement***

10   X-ray diffraction data of crystals of $Nsp1_{13-127}$ were collected at Advanced Photon Source

11   Beamline 19 ID (Argonne, Illinoi USA) under cryo-stream at 93.15 K. Diffraction data were

12   processed with HKL3000 suit [28]. Initial phase for $Nsp1_{13-127}$ was obtained by molecular

13   replacement with Molrep using NMR structure of SARS-CoV Nsp1 (PDB: 2HSX) as a search

14   model [29]. Subsequently, the initial electron density map was improved though density

15   modification by parrot and the model was built using buccaneer [29]. Final model was produced

16   by the cycle of manual model building and refinement using Phenix.refine and Coot [30, 31]. All

17   geometry was verified using Phenix validation tools (Ramachandran statistics: Favoured

18   (97.1%), additionally allowed (2.9%), disallowed (0.0%)) and the wwPDB server.

19

20   ***Structure analysis and homology-modeling***

21   Electrostatic potential surfaces were calculated using Pymol using APBS[21]. Structure

22   similarity searches of the Protein Data Bank were performed using the Dali server

23   [22].Secondary structure prediction of Nsp1 homologues was done using PSIPRED [26].

1   Homology modeling of MERS-CoV was performed by I-TASSER using the crystal structure of

2   SARS-CoV-2 $Nsp1_{13-127}$ as the threading template [32, 33]. The model produced by I-TASSER

3   was subjected to further energy minimization using the YASARA energy minimization server to

4   improve stereochemical property [34]. Stereochemical properties of the final model was

5   validated with Ramachandran plot using PROCHECK server and 90% of residues are in the

6   favoured/additionally allowed region [35]. The model of the full-length SARS-CoV-2 Nsp1 was

7   generated using the Robetta structural predication server with the SARS-CoV-2 $Nsp1_{13-127}$

8   structure defined as the template [36].

9

17

18   **<u>Competing Interests</u>**

19   The authors declare that they have no competing interests.

References

1. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-19 in real time.* Lancet Infect Dis, 2020. **20**(5): p. 533-534.

2. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* Nature, 2020. **579**(7798): p. 270-273.

3. 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, 2020. **9**(1): p. 221-236.

4. Weiss, S.R. and S. Navas-Martin, *Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus.* Microbiol Mol Biol Rev, 2005. **69**(4): p. 635-64.

5. Yang, Y., et al., *Two Mutations Were Critical for Bat-to-Human Transmission of Middle East Respiratory Syndrome Coronavirus.* J Virol, 2015. **89**(17): p. 9119-23.

6. Zhang, T., Q. Wu, and Z. Zhang, *Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak.* Curr Biol, 2020. **30**(7): p. 1346-1351 e2.

7. Masters, P.S., *The molecular biology of coronaviruses.* Adv Virus Res, 2006. **66**: p. 193-292.

8. Narayanan, K., et al., *Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression.* Virus Res, 2015. **202**: p. 89-100.

9. Tanaka, T., et al., *Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA.* J Virol, 2012. **86**(20): p. 11128-37.

10. Harcourt, B.H., et al., *Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity.* J Virol, 2004. **78**(24): p. 13600-12.

11. Kamitani, W., et al., *A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein.* Nat Struct Mol Biol, 2009. **16**(11): p. 1134-40.

12. Kamitani, W., et al., *Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation.* Proc Natl Acad Sci U S A, 2006. **103**(34): p. 12885-90.

13. Narayanan, K., et al., *Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells.* J Virol, 2008. **82**(9): p. 4471-9.

14. Law, A.H., et al., *Role for nonstructural protein 1 of severe acute respiratory syndrome coronavirus in chemokine dysregulation.* J Virol, 2007. **81**(1): p. 416-22.

15. Ye, Q., B. Wang, and J. Mao, *The pathogenesis and treatment of the `Cytokine Storm' in COVID-19.* J Infect, 2020. **80**(6): p. 607-613.

16. Zust, R., et al., *Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines.* PLoS Pathog, 2007. **3**(8): p. e109.

17. Wathelet, M.G., et al., *Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain.* J Virol, 2007. **81**(21): p. 11620-33.

1   18.   Zhang, W.F., et al., *Novel Coronavirus Polymerase and Nucleotidyl-Transferase*
2   *Structures: Potential to Target New Outbreaks.* J Phys Chem Lett, 2020. **11**(11): p. 4430-
3   4435.
4   19.   Almeida, M.S., et al., *Novel beta-barrel fold in the nuclear magnetic resonance structure*
5   *of the replicase nonstructural protein 1 from the severe acute respiratory syndrome*
6   *coronavirus.* J Virol, 2007. **81**(7): p. 3151-61.
7   20.   Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline*
8   *state.* J Mol Biol, 2007. **372**(3): p. 774-97.
9   21.   Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the*
10   *ribosome.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.
11 22.   Holm, L. and L.M. Laakso, *Dali server update.* Nucleic Acids Res, 2016. **44**(W1): p.
12   W351-5.
13 23.   Shen, Z., et al., *A conserved region of nonstructural protein 1 from alphacoronaviruses*
14   *inhibits host gene expression and is critical for viral virulence.* J Biol Chem, 2019.
15   **294**(37): p. 13606-13618.
16 24.   Terada, Y., et al., *MERS coronavirus nsp1 participates in an efficient propagation*
17   *through a specific interaction with viral RNA.* Virology, 2017. **511**: p. 95-105.
18 25.   Thoms, M., et al., *Structural basis for translational shutdown and immune evasion by the*
19   *Nsp1 protein of SARS-CoV-2.* Science, 2020.
20 26.   Buchan, D.W.A. and D.T. Jones, *The PSIPRED Protein Analysis Workbench: 20 years*
21   *on.* Nucleic Acids Res, 2019. **47**(W1): p. W402-W407.
22 27.   Jones, D.T. and D. Cozzetto, *DISOPRED3: precise disordered region predictions with*
23   *annotated protein-binding activity.* Bioinformatics, 2015. **31**(6): p. 857-63.
24 28.   Minor, W., et al., *HKL-3000: the integration of data reduction and structure solution--*
25   *from diffraction images to an initial model in minutes.* Acta Crystallogr D Biol
26   Crystallogr, 2006. **62**(Pt 8): p. 859-66.
27 29.   Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta
28   Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.
29 30.   Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular*
30   *structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
31 31.   Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta
32   Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
33 32.   Yang, J. and Y. Zhang, *I-TASSER server: new development for protein structure and*
34   *function predictions.* Nucleic Acids Res, 2015. **43**(W1): p. W174-81.
35 33.   Zhang, C., P.L. Freddolino, and Y. Zhang, *COFACTOR: improved protein function*
36   *prediction by combining structure, sequence and protein-protein interaction information.*
37   Nucleic Acids Res, 2017. **45**(W1): p. W291-W299.
38 34.   Krieger, E., et al., *Improving physical realism, stereochemistry, and side-chain accuracy*
39   *in homology modeling: Four approaches that performed well in CASP8.* Proteins, 2009.
40   **77 Suppl 9**: p. 114-22.
41 35.   Laskowksi, R.A., MacArthur, M.W., Moss, D. S, Thornton, J.M., *PROCHECK - a*
42   *program to check the stereochemical quality of protein structures.* Journal of Applied
43   Crystallography 1993. **26**: p. 283-291.
44 36.   Song, Y., et al., *High-resolution comparative modeling with RosettaCM.* Structure, 2013.
45   **21**(10): p. 1735-42.
46