AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

STRUCTURE AND ASSEMBLY

Structure of Nonstructural Protein 1 from SARS-CoV-2

Lauren K. Clark,[a] Todd J. Green,[b] Chad M. Petit[a]

[a]Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham School of Medicine, Birmingham, Alabama, USA
[b]Department of Microbiology, University of Alabama at Birmingham School of Medicine, Birmingham, Alabama, USA

ABSTRACT The periodic emergence of novel coronaviruses (CoVs) represents an ongoing public health concern with significant health and financial burdens worldwide. The most recent occurrence originated in the city of Wuhan, China, where a novel coronavirus (severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]) emerged causing severe respiratory illness and pneumonia. The continual emergence of novel coronaviruses underscores the importance of developing effective vaccines as well as novel therapeutic options that target either viral functions or host factors recruited to support coronavirus replication. The CoV nonstructural protein 1 (nsp1) has been shown to promote cellular mRNA degradation, block host cell translation, and inhibit the innate immune response to virus infection. Interestingly, deletion of the nsp1-coding region in infectious clones prevented the virus from productively infecting cultured cells. Because of nsp1's importance in the CoV life cycle, it has been highlighted as a viable target for both antiviral therapy and vaccine development. However, the fundamental molecular and structural mechanisms that underlie nsp1 function remain poorly understood, despite its critical role in the viral life cycle. Here, we report the high-resolution crystal structure of the amino globular portion of SARS-CoV-2 nsp1 (residues 10 to 127) at 1.77-Å resolution. A comparison of our structure with the SARS-CoV-1 nsp1 structure reveals how mutations alter the conformation of flexible loops, inducing the formation of novel secondary structural elements and new surface features. Paired with the recently published structure of the carboxyl end of nsp1 (residues 148 to 180), our results provide the groundwork for future studies focusing on SARS-CoV-2 nsp1 structure and function during the viral life cycle.

IMPORTANCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of the COVID-19 pandemic. One protein known to play a critical role in the coronavirus life cycle is nonstructural protein 1 (nsp1). As such, it has been highlighted in numerous studies as a target for both the development of antivirals and the design of live-attenuated vaccines. Here, we report the high-resolution crystal structure of nsp1 derived from SARS-CoV-2 at 1.77-Å resolution. This structure will facilitate future studies focusing on understanding the relationship between structure and function for nsp1. In turn, understanding these structure-function relationships will allow nsp1 to be fully exploited as a target for both antiviral development and vaccine design.

KEYWORDS SARS-CoV-2, nonstructural protein 1, coronavirus, severe acute respiratory syndrome, X-ray crystallography, COVID-19

The periodic emergence of novel coronaviruses (CoVs) represents an ongoing public health concern with significant health and financial burdens worldwide. In the past 2 decades, three highly pathogenic human CoVs have emerged from zoonotic events. In 2002, an outbreak of atypical pneumonia termed severe acute respiratory syndrome (SARS) appeared in the Guangdong Province of Southern China. The etiological agent

Citation Clark LK, Green TJ, Petit CM. 2021. Structure of nonstructural protein 1 from SARS-CoV-2. J Virol 95:e02019-20. https://doi.org/10.1128/JVI.02019-20.

Editor Rebecca Ellis Dutch, University of Kentucky College of Medicine

Address correspondence to Todd J. Green, tgreen@uab.edu, or Chad M. Petit, cpetit@uab.edu.

Received 15 October 2020
Accepted 19 November 2020
Accepted manuscript posted online 24 November 2020
Published 28 January 2021

for this disease was found to be a novel coronavirus named SARS coronavirus (SARS-CoV) (1–3). In 2012, a novel coronavirus named Middle East respiratory syndrome coronavirus (MERS-CoV) was isolated from a patient who died in Saudi Arabia after presenting with acute respiratory distress and kidney injury (4). The most recent occurrence originated in the city of Wuhan, China, where a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged causing severe respiratory illness and pneumonia (5, 6). The resulting disease was named coronavirus disease 2019 (COVID-19), and a pandemic was declared by the World Health Organization in March of 2020.

CoVs are members of the *Coronaviridae* family of viruses that contain large positive-sensed, single-stranded RNA genomes of approximately 30 kb in length. Approximately two-thirds of the genome, located in the 5′ genomic region, contains open reading frame 1a (ORF1a) and open reading frame 1b (ORF1b). To express both polyproteins, coronaviruses encode a slippery sequence and an RNA pseudoknot that cause a −1 ribosomal frameshift (7, 8). ORF1a and ORF1ab encode nsps 1 to 11 and 1 to 16, respectively, and are translated into large polyproteins which are subsequently cleaved into individual nonstructural proteins (nsps) by virally encoded proteases (9). Genes that encode the structural proteins (spike, envelope, membrane, and nucleocapsid) are located in the remaining third of the genome with a variable number of strain-dependent ORF accessory proteins present between the structural genes (10).

Nsp1 is the N-terminal cleavage product released from the replicase polyprotein by the virally encoded papain-like proteinase (nsp3d; PLpro) (11). It has been shown to promote cellular mRNA degradation, block host cell translation, and inhibit the innate immune response to virus infection (12–20). Deletion of the nsp1-coding region in infectious clones prevented the virus from productively infecting cultured cells (21). Also, mutations preventing the release of nsp1 from the nascent ORF1a polyprotein substantially limited virus viability (22). While there have been multiple studies describing nsp1 function, there have been limited structure-function studies focusing on nsp1. These types of studies are needed to fully exploit nsp1 as a target for the development of both antivirals and rational vaccine. The availability of high-resolution structures is a critical first step in establishing structure-function relationships of proteins and is the focus of the manuscript.

Here, we report the high-resolution crystal structure of the globular portion of SARS-CoV-2 nsp1 to a resolution of 1.77 Å. A comparison of our structure with the SARS-CoV-1 nsp1 structure reveals how mutations alter the conformation of flexible loops. These changes in conformation induce the folding of novel secondary structural elements that, in turn, generate surfaces that are distinct from those of SARS-CoV-1 nsp1 in both contour and feature. Comparison with α-CoV nsp1 structures shows that the nsp1 globular domain has a shared structural homology while having low sequence similarity. Our results provide the groundwork for future studies focusing on the structure-function mechanisms by which SARS-CoV-2 nsp1 functions during the viral life cycle.

RESULTS

Expression of SARS-CoV-2 nsp1. To obtain the high-resolution structure of nsp1 from SARS-CoV-2, we based our expression construct on the previously solved structure of nsp1 from SARS-CoV-1 (23). In this study, it was determined that nsp1 of SARS-CoV-1 includes a globular domain of residues 13 to 121 that is between disordered regions consisting of resides 1 to 12 and 122 to 179. We generated an initial expression construct with N-terminal truncations beginning at amino acid residue 13 and continuing to residue 127. Using nuclear magnetic resonance spectroscopy, heteronuclear single quantum coherence triple resonance spectral analysis indicated that the amino terminus of our construct beginning at amino acid residue 13 prematurely eliminated a β-strand (data to be reported elsewhere). We generated a second expression construct

TABLE 1 Data collection and refinement statistics

Parameter	Value(s) for SARS-CoV-2 nsp1[a]
Data collection statistics	
Wavelength (Å)	1.54
Resolution range (Å)	35.42–1.77 (1.80–1.77)
Space group	P 43 21 2
Unit cell	36.6, 36.6, 141.2, 90, 90, 90
No. of total reflections	137,056
No. of unique reflections	10,106
Multiplicity	13.5
Completeness (%)	99.8 (96.2)
Mean I/sigma(I)	40.38 (2.24)
Wilson B-factor	18.58
R_{merge}[b]	0.071 (0.634)
R_{pim}[c]	0.018 (0.376)
CC1/2	0.998 (0.759)
Refinement statistics	
No. of reflections used in refinement	10,071 (965)
No. of reflections used for R_{free}	1,008 (97)
R_{work}/R_{free}[d]	0.187/0.2128
Model statistics	
No. of nonhydrogen atoms	1,026
Macromolecules (no.)	912
Solvent (no.)	114
Protein residues (no.)	117
RMS[e] (bonds)	0.007
RMS (angles)	1.02
Ramachandran favored (%)	96.52
Ramachandran allowed (%)	3.48
Ramachandran outliers (%)	0
Average B-factor	23.69
Macromolecules	22.96
Solvent	29.47

[a]Values between parentheses correspond to the highest-resolution shells.
[b]R-merge = $\Sigma_{hkl}\ \Sigma_i|I_i(hkl) - [I(hkl)]|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $\Sigma_i I_i(hkl)$ is the *i*-the measurement of reflection hkl and [I(hkl)] is the weighted mean of all measurements.
[c]$R_{pim} = \Sigma_{hkl}[1/(N-1)]^{1/2}\ \Sigma_i|I_i(hkl) - [I(hkl)]|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $\Sigma_i I_i(hkl)$ is the *i*-th measurement of reflection hkl, [*I*(hkl)] is the weighted mean of all measurements, and *N* is the redundancy for the hkl reflection.
[d]$R_{work}/R_{free} = \Sigma_{hkl}|F_o - F_c|/\Sigma_{hkl}|F_o|$, where F_c is the calculated and F_o is the observed structure factor amplitude of reflection hkl for the working/free set, respectively; 10% of reflections were reserved for the R_{free} set.
[e]RMS, root mean square.

with N-terminal truncations beginning at amino acid residue 10 continuing to residue 127. Both nsp1 constructs were expressed and purified to homogeneity.

Crystal Structure of nsp1$_{10–127}$ from SARS-CoV-2. Crystallization trials were initiated with nsp1$_{10–127}$ and nsp1$_{13–127}$. Crystals were obtained with each construct, though the number of crystallization conditions was greater and crystal quality better with the longer construct, nsp1$_{10–127}$. For this reason, structural studies were carried out with this construct. The structure of the globular domain of nsp1 was determined by molecular replacement with the core of SARS-CoV-1 nsp1 (PDB identifier [ID] 2HSX). Final data collection and refinement statistics are shown in Table 1. The final model contained residues 10 to 126 (Fig. 1). The tertiary fold of nsp1 is composed of regular secondary structural elements arranged sequentially as β1-η1-α1-β2-η3-β3-β4-β5-β6-β7 with η1 and η2 being 3_{10} helices. A seven-stranded β-barrel is formed with a mixture of parallel and antiparallel β-strands with contacts between β2, β3, β4, β5, and β6 strands as well as β1 and β7 strands. The β-strands consist of residues 13 to 20 (β1), 51 to 54 (β2), 68 to 73 (β3), 84 to 92 (β4), 95 to 97 (β5), 103 to 109 (β6), and 117 to 123 (β7). The η1 (residues 23 to 25) and η2 (residues 61 to 63) helices are positioned across one barrel opening, while α1 (residues 34 to 49) is positioned alongside the barrel making contact with the β4 strand.

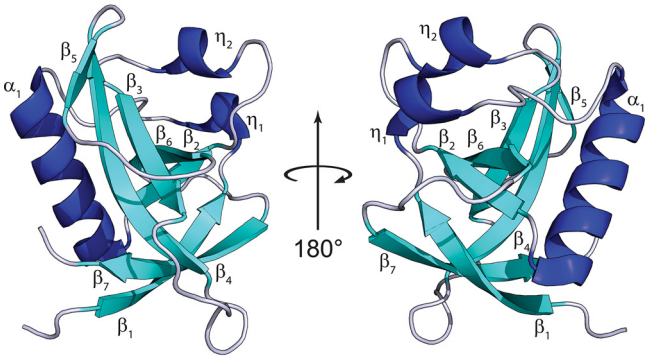

FIG 1 Crystal structure of SARS-CoV-2 nsp1. Ribbon diagram of the SARS-CoV-2 nsp1$_{10–126}$ with the α-helices in blue, β-strands in cyan, and loops in gray. All secondary structural elements are labeled. Coordinates have been deposited under PDB ID 7K7P.

Comparison with solution structure of nsp1$_{13–127}$ from SARS-CoV-1. The nsp1 portion of ORF1ab is well conserved when comparing SARS-CoV-1 and SARS-CoV-2 (84% identity) (Fig. 2A). This relatively high conservation is also applicable to the subset of amino acids (10 to 126) that comprise our structure (86%). This level of conservation

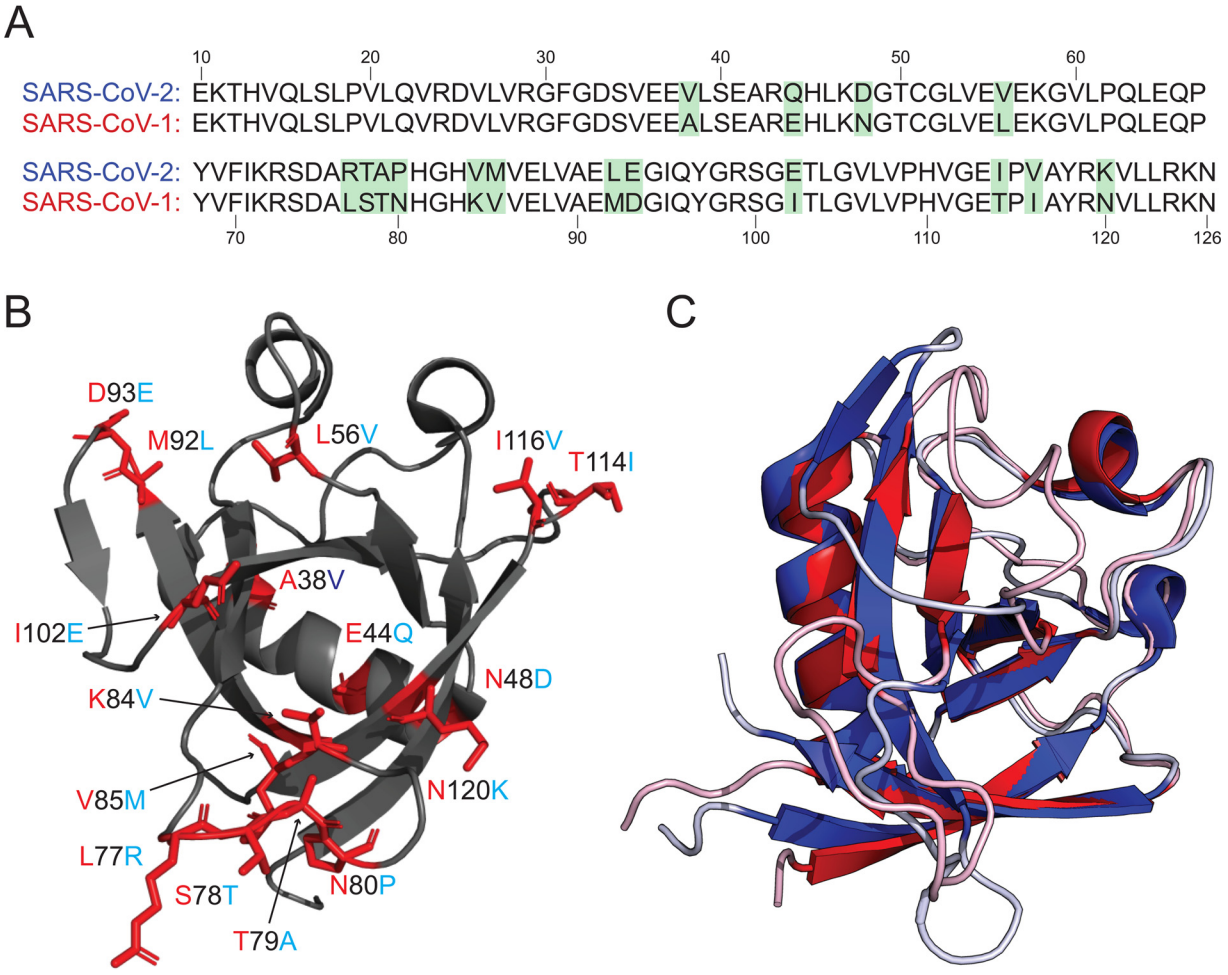

FIG 2 Comparison of the SARS-CoV-2 and SARS-CoV-1 nsp1 structures. (A) Sequence alignment between the nsp1 structures derived from SARS-CoV-2 and SARS-CoV-1 with amino acid differences highlighted in green. (B) Ribbon diagram of the SARS-CoV-2 nsp1$_{10–127}$ with amino acid differences with SARS-CoV-1 nsp1$_{12–127}$. The amino acid position is labeled in black with the sequence identity of SARS-CoV-1 at that position labeled in red and SARS-CoV-2 in light blue. (C) Overlay of the structures of nsp1 derived from SARS-CoV-1 (red) and SARS-CoV-2 (blue). PDB ID 2HSX for SARS-CoV-1 nsp1$_{12–127}$ and 7K7P for SARS-CoV-2 nsp1$_{10–127}$.

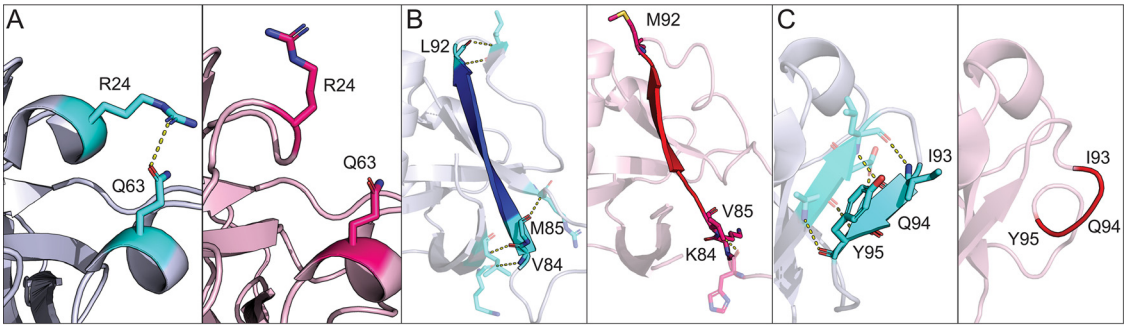

FIG 3 Differences in secondary structural elements between SARS-CoV-2 and SARS-CoV-1 nsp1. (A) Ribbon diagram of the stabilization via polar contacts of the 3_{10} helix present in SARS-CoV-2 nsp1 but not in SARS-CoV-1 nsp1. Amino acids specified in the text are indicated in cyan and pink for SARS-CoV-2 and SARS-CoV-1, respectively. (B) Extension of the β4 strand in SARS-CoV-2 (blue) relative to that in SARS-CoV-1 (red). Mutations between the two viruses are labeled with their respective amino acid identity and position. Each mutation is also indicated in cyan for SARS-CoV-2 and pink for SARS-CoV-1. (C) An additional β-strand is present in the SARS-CoV-2 nsp1. Amino acids that compose this segment of both proteins are labeled and indicated in cyan for SARS-CoV-2 and pink for SARS-CoV-1. For all panels, SARS-CoV-2 nsp1 is in blue gray, SARS-CoV-1 nsp1 is in light pink, and polar contacts are indicated by yellow dashed lines.

between CoVs is remarkable given the limited amino acid sequence homology among nsp1s from different β-CoVs (19). A number of mutations (9 of 16) occur in the large loops in the structure (Fig. 2B). It is widely understood that loops often play critical roles in protein-protein interactions, making these mutations particularly interesting (24, 25).

We note that the structure is composed of approximately the same portions of nsp1 (residues 10 to 127) as the previously solved nsp1 structure from SARS-CoV-1 (residues 13 to 127). The fold of SARS-CoV-2 nsp1 is generally well conserved compared to the previously solved structure of nsp1 from SARS-CoV-1 (Fig. 2C). However, there are several notable differences between the structures upon detailed analysis. One such difference is the presence of an additional 3_{10} helix (η1) in the SARS-CoV-2 structure (residues 23 to 25). We note that the primary sequence composing this additional 3_{10} helix is identical between SARS-CoV-1 and SARS-CoV-2. Additionally, there are no apparent proximal mutations in SARS-CoV-2 that would stabilize this 3_{10} helix. The stabilization of this helix is due to a polar interaction between R24 in α1 and Q63 in α3 (Fig. 3A) present in only SARS-CoV-2 nsp1.

In addition to the formation of a new 3_{10} helix in nsp1 from SARS-CoV-2, there are also mutations that result in extension of β-strands within the interior of the protein when comparing the two proteins. The β4 strand is extended by four amino acids and is now composed by amino acids 84 to 92 in SARS-CoV-2 as opposed to amino acids 87 to 91 in SARS-CoV-1. This is due to the K84V, V85M, and M92L mutations in positions adjacent to the β4 strand (Fig. 3B). This extension facilitates the formation of a new β-strand (β5) not found in the original SARS-CoV-1 structure. This strand is composed of amino acids 95 to 97 and is stabilized by main chain contacts with residues 90 to 92 (Fig. 3C). The β6 strand is also extended by two amino acids, and so it is now composed of amino acids 103 to 109 rather than amino acids 105 to 109. This extension is due to additional main chain interactions. Finally, β1, β3, and β6 (β5 in SARS-CoV-1) strands are also extended by a single amino acid.

One striking difference between the structures is the alternative conformations of two major loops located between β3 and β4 (loop 1) and between β4 and β6 (loop 2, β4 and β5 in SARS-CoV-1) (Fig. 4). The residues that make up these loops in SARS-CoV-2 nsp1 form a much more significant network of polar contacts with respect to those in the analogous loops in SARS-CoV-1 nsp1. These changes in local stability and conformation facilitate the formation of a new β-strand (β5) not found in SARS-CoV-1 nsp1. Analyses of loop 1 (residues 73 to 84) and loop 2 (residues 92 to 104) indicate average C_α displacements of 16.6 Å and 27.6 Å, respectively. As loops play critical roles in

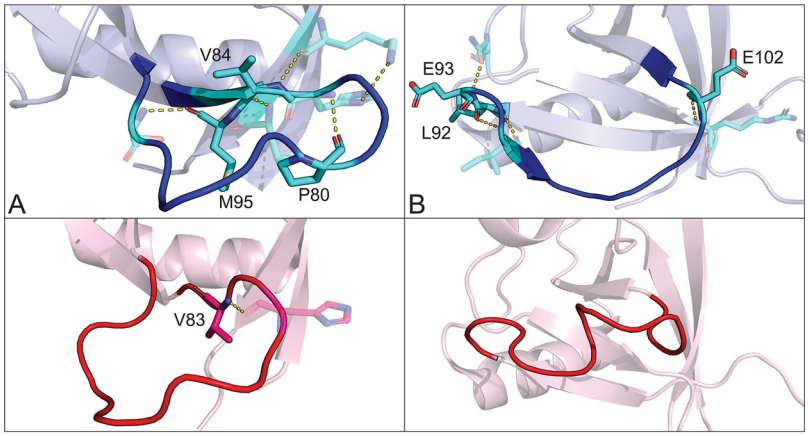

FIG 4 Conformational differences in major loops between SARS-CoV-2 and SARS-CoV-1 nsp1. Ribbon diagram of SARS-CoV-2 and SARS-CoV-1 nsp1 with the loops located between β3 and β4 (A) and β4 and β6 (B) indicated in blue for SARS-CoV-2 and red for SARS-CoV-1. Residues mutated between the two viruses that are also involved in polar contacts are labeled as well as indicated in cyan for SARS-CoV-2 and pink for SARS-CoV-1. Polar contacts are indicated by yellow dashed lines.

protein-protein interactions, relative variations in these loops between SARS-CoV-1 and SARS-CoV-2 are particularly interesting as they may alter host-pathogen interactions involving nsp1 (24, 25).

Structural differences between the two proteins result in differences in the electrostatic surface potential. These differences are the result of conformational differences and mutations presented on the surfaces of the proteins. Figure 5 contrasts four of these regions. Both SARS-CoV-1 and SARS-CoV-2 nsp1 share a large common patch of electronegative potential (Fig. 5A and E); however, in SARS-CoV-2 nsp1, the electronegative surface potential has an increased intensity and clustering has increased. This patch in SARS-CoV-2 nsp1 is generated by D33, E36, E37, E41, Q44, H45, D48, E65, Q66, and E93. These residues comprise three mutations from SARS-CoV-1 nsp1 and include E44Q, N48D, and D93E. In addition, a valley between the upper and lower portions of the patch in SARS-CoV-1 residues, between residues D33 and E65, is closed in SARS-CoV-2 by the reorientation of the main chain flanking Q66. This reorientation allows the sidechain of Q66 to form a new hydrogen bond with the amide nitrogen of E93. A dominant electropositive pocket and extended surface is also observed in both struc-

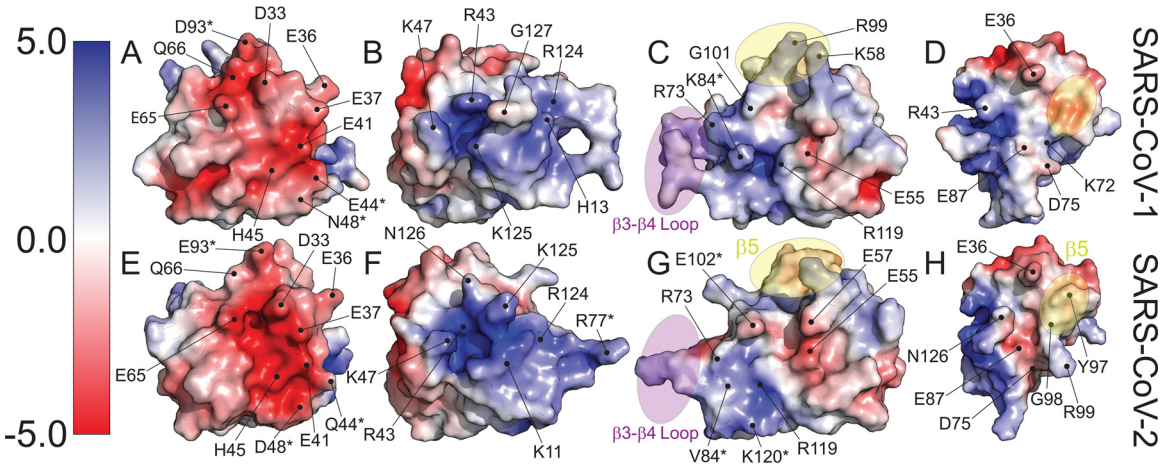

FIG 5 Differences in electrostatic surface potential between SARS-CoV-1 nsp1 and SARS-CoV-2 nsp1. (A to H) Surface models of both proteins with amino acids contributing to the differences in electrostatic character labeled. Secondary structural elements that contribute to differences in surface contours are highlighted in purple for the loop between β3 and β4 and in yellow for β5, found in SARS-CoV-2 but not SARS-CoV-1 nsp1. Residues that are mutated between the two viruses are indicated with an asterisk (*).

tures (Fig. 5B and F). In SARS-CoV-2 nsp1, this region is composed of K11, R43, K47, R77, R124, and K125. The contiguous pit of electronegativity observed in SARS-CoV-2 nsp1 is disrupted in the SARS-CoV-1 structure by rearrangement of the C terminus allowing G127 to protrude between R43 and K125. Two smaller juxtaposed electropositive and electronegative patches are shown in Fig. 5C and G. There are three mutations relative to SARS-CoV-1 nsp1 in this area: K84V, I102E, and N120K. In the SARS-CoV-2 structure, two mutations contribute to the negative surface character of this region. The mutated side chain of V84 occupies the space of K84 in SARS-CoV-1, while K120 buttresses the bottom of the patch. In the adjacent electropositive patch, E102 sits in the positive pocket surrounded by E55, E57, and E102. The conformation of the loop between β3 and β4, as well as the restructured strand (β5) and loop between β4 and β6 (β4 and β5 in SARS-CoV-1), contributes to vastly different surface contours (Fig. 5C and G). Finally, a new electronegative crevasse is generated by the restructuring of strand β5 and the trailing loop leading to strand β6 and subsequent reposition of the concluding residues of β7 through the C terminus (residues 122 to 126) in SARS-CoV-2 (Fig. 5D and H). Two prominent residues in this new crevasse are D75 and E87. These residues are identical in SARS-CoV-1; however, the surrounding residues are positioned so that they are more obscured from the surface. E87 (SARS-CoV-1) is sandwiched between K72 and R124, appearing to salt bridge with K72.

Structural homology between SARS-CoV-2 nsp1 and α-CoV nsp1s. The structures of SARS-CoV-1 nsp1 and SARS-CoV -2 nsp1 represent the known nsp1 structures of β-CoVs. To date, the structures of five additional nsp1 structures originating from viruses in the α-coronavirus genera have been published. The sequence identity between SARS-CoV-2 nsp1 and these α-CoV nsp1s is extremely low at 9% to 13%, with only marginally more sequence similarity at 21% to 27%. There is, however, easily identifiable structural homology between nsp1s derived from α- and β-CoVs. The structure of SARS-CoV-2 was structurally aligned with the structures of nsp1s from porcine transmissible gastroenteritis coronavirus strain Purdue (3ZBD) (26), porcine epidemic diarrhea virus (5XBC) (27), transmissible gastroenteritis virus (6IVC) (28), swine acute diarrhea syndrome coronavirus (6LPA) (29), and feline infectious peritonitis virus (6LP9) (29). Aligned structures are illustrated in Fig. 6, with statistical analysis data compiled in Table 2. All of the structures share a core topological similarity. However, there are substitutions and insertions in the linear arrangement of the secondary structure elements (Fig. 6C). The first 3_{10} helix of the SARS-CoV-2 structure is swapped for strand β2 in the α-CoV structures, though essentially occupying the same space. Strand β4 represents an insertion in the α-CoV structures. The strand forms a short parallel β-strand configuration with β2. As noted above, SARS-CoV-2 has a new strand β5, not observed in SARS-CoV-1. This strand is also a deviation from the topology of the α-CoV nsp1 structures. The rest of the secondary structural elements follow a common fold and alignment, being β1-α1-β2-η2-β3-β4-β6-β7 for SARS-CoV-2 and β1-α1-β3-η1-β5-β6-β7-β8 for α-CoVs (Fig. 6H).

DISCUSSION

Here, we report the high-resolution structure of nsp1 derived from SARS-CoV-2, the etiological agent of COVID-19. The interior of the protein is composed of a seven-stranded β-barrel with an α-helix positioned on the side of the barrel and two 3_{10} helices across one end of the barrel. nsp1 is highly homologous to SARS-CoV-1 nsp1 in both structure and sequence. However, there are a number of differences between the two proteins that may alter their function in the context of their respective viral life cycles. One difference is the presence of secondary structural elements, an additional β-strand (β5) and 3_{10} helix, in SARS-CoV-2 nsp1 that are not in SARS-CoV-1 nsp1. Several of the major loops in SARS-CoV-2 nsp1 are also in alternative conformations due to the increased polar contacts between amino acids that compose them and the globular domain. Finally, there are differences in the electrostatic surface potential

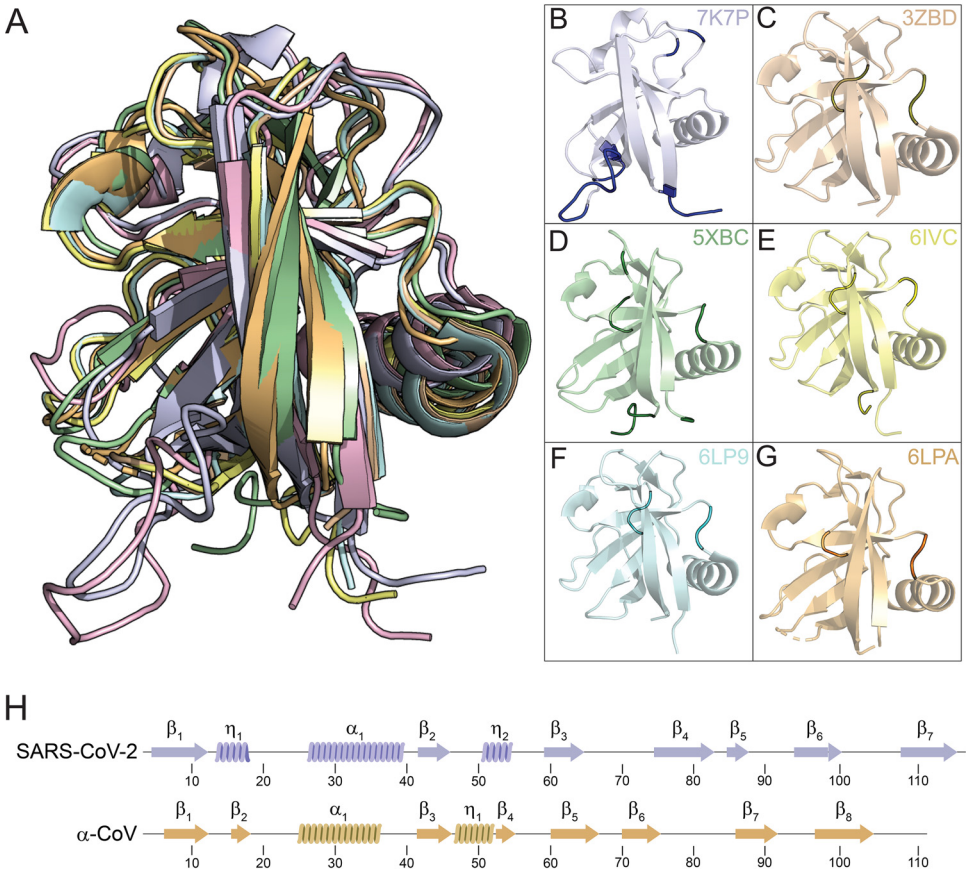

FIG 6 Structural homology of β- and α-coronavirus nsp1s. Structures of nsp1s from SARS-CoV-2 (7K7P), porcine transmissible gastroenteritis coronavirus strain Purdue (3ZBD), porcine epidemic diarrhea virus (5XBC), transmissible gastroenteritis virus (6IVC), swine acute diarrhea syndrome coronavirus (6LPA), and feline infectious peritonitis virus (6LP9) are shown structurally aligned (A) and in common orientations (B to G). (B to G) The common core shared with SARS-CoV-2 is shown in pale coloring, with deviations shown in darker shade coloring. (H) Secondary structural elements of SARS-CoV-2 and α-CoV nsp1 with positions indicated.

resulting from amino acid differences between the two proteins. Given the critical role that nsp1 plays in the CoV life cycle, these differences may underlie fundamental differences between the SARS-CoV-2 and SARS-CoV-2 such as virulence, pathogenicity, and infectivity.

nsp1 is a virulence determinant of CoVs that plays critical roles during the viral life cycle such as suppressing host gene expression (14, 30, 31). This suppression of host gene expression is necessary for viral replication and allows evasion of the cellular immune response. In a recent study, cryo-electron microscopy (cryo-EM) was used to show that the carboxyl terminus (residues 148 to 180) of SARS-CoV-2 nsp1 binds to the

TABLE 2 Structural similarity of coronavirus nsp1s

Category	Data for: SARS-CoV-1	Feline infectious peritonitis virus	Swine acute diarrhea syndrome coronavirus	Porcine transmissible gastroenteritis coronavirus strain Purdue	Transmissible gastroenteritis virus	Porcine epidemic diarrhea virus
Genus		β-CoV	α-CoV	α-CoV	α-CoV	α-CoV
α-CoV						
Residues aligned (*n*)	106	99	97	98	100	97
RMSD[a]	1.71	3.31	3.24	3.2	3.2	3.13
Sequence identity (%)	73	12	12	13	13	9
Sequence similarity (%)	79	26	27	25	26	21
PDB ID	2HSX	6LP9	6LPA	3ZBD	6IVC	5XBC

[a]RMSD, root mean square deviation.

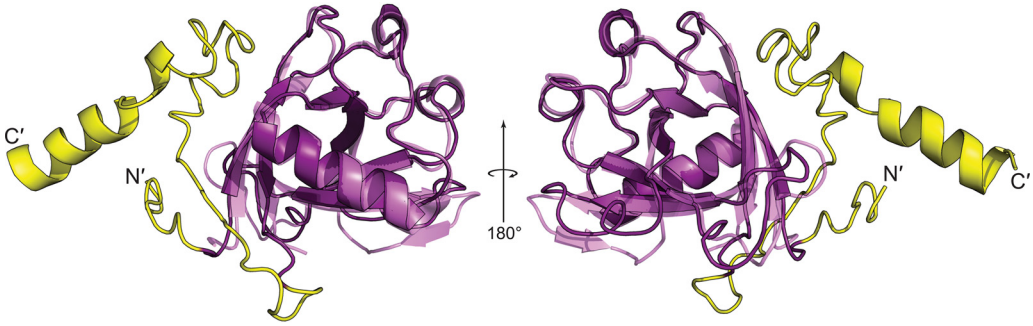

FIG 7 Structural prediction of full-length SARS-CoV-2 nsp1. Ribbon diagram of full-length SARS-CoV-2 nsp1 modeled using the I-TASSER platform.

40S ribosomal subunit and inhibits translation (32, 33). Specifically, the structures revealed that this fragment of nsp1 binds to the mRNA entry channel of the 40S subunit, which effectively blocks translation (32, 33). We modeled the structure of full-length nsp1 using the I-TASSER platform with our structure as a template (34–36). As shown in Fig. 7, residues 1 to 9 and 131 to 162 contain no discernible secondary structure, with the remaining residues adopting two α-helices. This is compatible with the cryo-EM structure that indicates that the carboxyl terminus (residues 148 to 180) binds to the 40S ribosomal subunit (32, 33).

Additionally, there is a high degree of structural and sequence homology between nsp1 derived from SARS-CoV-1 and from SARS-CoV-2. The rare conservation among different β-CoVs may provide insight into SARS-CoV-2 nsp1 function from previous studies that focused on SARS-CoV-1 nsp1 (19). For example, previous studies have shown that nsp1 from SARS-CoV-1 is unique among CoVs and is hypothesized to contribute to the exceptional pathogenesis of SARS-CoV-1 in humans (37, 38). In fact, nsp1 control of interferon expression may be crucial to SARS-CoV-1 infection given that the virus is interferon sensitive (15, 39, 40). Further studies will be needed to determine which functions are common between the two proteins and if there are any functional differences that alter their respective viral lifecycles.

There is an urgent need to develop both vaccines and antivirals against SARS-CoV-2 and any potential novel CoVs that pose a threat to global health. Because of its critical role, nsp1 has been identified as a target for both the development of broad-spectrum antivirals and rational live-attenuated vaccine design (19, 20, 41–47). This fact was underscored by a previous study that focused on the function of nsp1 during the SARS-CoV-1 life cycle which concluded the following. "How SARS-CoV nsp1 functions remains to be answered. Removal of the nsp1 protein function from the virus (48) will enable determination of its effects on replication and/or virulence *in vivo*; however, these experiments will require careful design and analysis because indiscriminate deletion of nsp1 sequence would remove *cis*-acting elements required for replication" (49). This conclusion is even more relevant during the COVID-19 pandemic. Our findings, combined with the structures of the carboxyl end of nsp1, are the first step to a comprehensive understanding of the structure-function relationship of SARS-CoV-2 nsp1.

Note that during the preparation of the manuscript, an independent X-ray crystal structure of SARS-CoV-2 nsp1 was deposited in PDB (ID 7K3N). The two structures, PDB ID 7K7P (this work) and 7K3N, have an RMSD of 0.351 Å based on C_α positioning, indicating that they are highly correlated.

MATERIALS AND METHODS

Protein expression and purification. The nsp1 segment of SARS-CoV-2 was subcloned into a 6× His-SMT3 fusion T7 expression vector (pE-SUMO; Life Sensors) using primers to amplify residues 10 to 127 of ORF1a. Expression plasmids were transformed into DE3 Star BL21 *Escherichia coli* cells and grown in LB medium containing 200 μg/ml of kanamycin. The nsp1 cultures were induced at an optical density

at 600 nm (OD_{600}) of 0.7 with 1 mM isopropyl β-D-1-thiogalactopyranoside for 48 h at 18°C. Cells were lysed via sonication in buffer containing 20 mM HEPES, 500 mM NaCl, 5 mM imidazole, 5 mM dithiothreitol (DTT), and 1.3 mg/ml lysozyme at pH 7.8. Clarified lysates were loaded onto a His tag purification column (Roche), and the protein was eluted in a buffer containing 20 mM HEPES, 500 mM NaCl, 250 mM imidazole, and 2.5 mM DTT at pH 8. The 6× His-SMT3 tag was removed from the N terminus with recombinant SUMO protease. The cleaved sample was then reloaded onto the His tag purification column to separate the 6× His-SMT3 tag from nsp1. Buffer exchange was performed with size exclusion chromatography using a Hi-Prep 26/60 Sephacryl S-100 HR (GE). The final buffer is composed of 20 mM $NaPO_4$, 200 mM NaCl, 1 mM EDTA, 50 mM arginine, 50 mM glutamic acid, and 0.02% sodium azide at pH 6.8.

Crystallization, data collection, structure determination, and refinement. Purified SARS-CoV-2 nsp1 was concentrated to 7 mg/ml and screened for crystallization conditions against commercially available kits using a Mosquito liquid handling robot (SPT Labtech). Diffraction quality crystals were obtained under conditions from multiple kits. An initial X-ray diffraction data set was collected with the Pilatus 200K detector (Dectris) atop a MicroMax-007 HF rotating anode X-ray generator (Rigaku) from a crystal grown in 0.1 M morpholineethanesulfonic acid (MES) monohydrate (pH 6.5) and 12% (wt/vol) polyethylene glycol 20000 and cryoprotected with a supplement of 20% glycerol. Raw intensity data were processed with the HKL3000 package (50). Initial phases were solved by molecular replacement with Phaser (51) using the core structure of PDB ID 2HSX. The structure was initially traced and refined with the Autobuild routine in PHENIX (52), followed by manual editing in Coot (53) and refinement with phenix.refine. This nearly complete model was used as the model for molecular replacement with a final high-resolution data set collected from a crystal grown in 0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate trihydrate (pH 6.5), and 30% (wt/vol) polyethylene glycol 8000 and cryoprotected with a supplement of 20% glycerol. The model was further traced manually and refined with phenix.refine. Data collection and refinement statistics are displayed in Table 1. All ribbon and surface illustrations of protein structures were prepared with PyMOL (Schrödinger) (54). Alignment statistics were determined with structure alignment tools from the RCSB (55).

Data availability. Structure factors and final refined coordinates have been deposited in the PDB with accession code 7K7P.

ACKNOWLEDGMENTS

This work was funded by the National Institutes of Health grants AI134693 (C.M.P.) and AI116738 (T.J.G.) in addition to the UAB School of Medicine and Hugh Kaul Precision Medicine Institute Urgent COVID-19 Research Award (C.M.P.). The O'Neal Comprehensive Cancer's X-ray crystallography core facility is supported through NIH P30CA13148.

REFERENCES

1. Drosten C, Preiser W, Gunther S, Schmitz H, Doerr HW. 2003. Severe acute respiratory syndrome: identification of the etiological agent. Trends Mol Med 9:325–327. https://doi.org/10.1016/s1471-4914(03)00133-3.
2. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ, Group SW, SARS Working Group. 2003. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 348:1953–1966. https://doi.org/10.1056/NEJMoa030781.
3. Peiris JS, Yuen KY, Osterhaus AD, Stohr K. 2003. The severe acute respiratory syndrome. N Engl J Med 349:2431–2441. https://doi.org/10.1056/NEJMra032498.
4. Al-Omari A, Rabaan AA, Salih S, Al-Tawfiq JA, Memish ZA. 2019. MERS coronavirus outbreak: implications for emerging viral infections. Diagn Microbiol Infect Dis 93:265–285. https://doi.org/10.1016/j.diagmicrobio.2018.10.011.
5. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi HW, Lo SK, Chan KH, Poon VK, Chan WM, Ip JD, Cai JP, Cheng VC, Chen H, Hui CK, Yuen KY. 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395:514–523. https://doi.org/10.1016/S0140-6736(20)30154-9.
6. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.
7. Baranov PV, Henderson CM, Anderson CB, Gesteland RF, Atkins JF, Howard MT. 2005. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. Virology 332:498–510. https://doi.org/10.1016/j.virol.2004.11.038.
8. Brierley I, Digard P, Inglis SC. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell 57:537–547. https://doi.org/10.1016/0092-8674(89)90124-4.
9. Ziebuhr J, Snijder EJ, Gorbalenya AE. 2000. Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. J Gen Virol 81:853–879. https://doi.org/10.1099/0022-1317-81-4-853.
10. Sawicki SG, Sawicki DL, Siddell SG. 2007. A contemporary view of coronavirus transcription. J Virol 81:20–29. https://doi.org/10.1128/JVI.01358-06.
11. Ziebuhr J. 2005. The coronavirus replicase. Curr Top Microbiol Immunol 287:57–94. https://doi.org/10.1007/3-540-26765-4_3.
12. Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S. 2011. SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. PLoS Pathog 7:e1002433. https://doi.org/10.1371/journal.ppat.1002433.
13. Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S. 2011. Alphacoronavirus transmissible gastroenteritis virus nsp1 protein suppresses protein translation in mammalian cells and in cell-free HeLa cell extracts but not in rabbit reticulocyte lysate. J Virol 85:638–643. https://doi.org/10.1128/JVI.01806-10.
14. Kamitani W, Huang C, Narayanan K, Lokugamage KG, Makino S. 2009. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. Nat Struct Mol Biol 16:1134–1140. https://doi.org/10.1038/nsmb.1680.
15. Kamitani W, Narayanan K, Huang C, Lokugamage K, Ikegami T, Ito N, Kubo H, 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. https://doi.org/10.1073/pnas.0603144103.
16. Lei L, Ying S, Baojun L, Yi Y, Xiang H, Wenli S, Zounan S, Deyin G, Qingyu Z, Jingmei L, Guohui C. 2013. Attenuation of mouse hepatitis virus by deletion of the LLRKxGxKG region of Nsp1. PLoS One 8:e61166. https://doi.org/10.1371/journal.pone.0061166.
17. Lokugamage KG, Narayanan K, Huang C, Makino S. 2012. Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation. J Virol 86:13598–13608. https://doi.org/10.1128/JVI.01958-12.
18. Narayanan K, Huang C, Lokugamage K, Kamitani W, Ikegami T, Tseng CT, 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. https://doi.org/10.1128/JVI.02472-07.
19. Narayanan K, Ramirez SI, Lokugamage KG, 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. https://doi.org/10.1016/j.virusres.2014.11.019.
20. Zust R, Cervantes-Barragan L, Kuri T, Blakqori G, Weber F, Ludewig B, 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. https://doi.org/10.1371/journal.ppat.0030109.
21. Brockway SM, Denison MR. 2005. Mutagenesis of the murine hepatitis virus nsp1-coding region identifies residues important for protein processing, viral RNA synthesis, and viral replication. Virology 340:209–223. https://doi.org/10.1016/j.virol.2005.06.035.
22. Galan C, Enjuanes L, Almazan F. 2005. A point mutation within the replicase gene differentially affects coronavirus genome versus minigenome replication. J Virol 79:15016–15026. https://doi.org/10.1128/JVI.79.24.15016-15026.2005.
23. Almeida MS, Johnson MA, Herrmann T, Geralt M, 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. https://doi.org/10.1128/JVI.01939-06.
24. Gavenonis J, Sheneman BA, Siegert TR, Eshelman MR, Kritzer JA. 2014. Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. Nat Chem Biol 10:716–722. https://doi.org/10.1038/nchembio.1580.
25. Siegert TR, Bird MJ, Makwana KM, Kritzer JA. 2016. Analysis of loops that mediate protein-protein interactions and translation into submicromolar inhibitors. J Am Chem Soc 138:12876–12884. https://doi.org/10.1021/jacs.6b05656.
26. Jansson AM. 2013. Structure of alphacoronavirus transmissible gastroenteritis virus nsp1 has implications for coronavirus nsp1 function and evolution. J Virol 87:2949–2955. https://doi.org/10.1128/JVI.03163-12.
27. Shen Z, Ye G, Deng F, Wang G, Cui M, Fang L, Xiao S, Fu ZF, Peng G. 2018. Structural basis for the inhibition of host gene expression by porcine epidemic diarrhea virus nsp1. J Virol 92:e01896-17. https://doi.org/10.1128/JVI.01896-17.
28. Shen Z, Wang G, Yang Y, Shi J, Fang L, Li F, Xiao S, Fu ZF, 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. https://doi.org/10.1074/jbc.RA119.009713.
29. Shen Z, Yang Y, Yang S, Zhang G, Xiao S, Fu ZF, Peng G. 2020. Structural and biological basis of alphacoronavirus nsp1 associated with host proliferation and immune evasion. Viruses 12:812. https://doi.org/10.3390/v12080812.
30. Jimenez-Guardeno JM, Regla-Nava JA, Nieto-Torres JL, DeDiego ML, Castano-Rodriguez C, Fernandez-Delgado R, Perlman S, Enjuanes L. 2015. Identification of the mechanisms causing reversion to virulence in an attenuated SARS-CoV for the design of a genetically stable vaccine. PLoS Pathog 11:e1005215. https://doi.org/10.1371/journal.ppat.1005215.
31. Tanaka T, Kamitani W, DeDiego ML, Enjuanes L, 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. https://doi.org/10.1128/JVI.01700-12.
32. Thoms M, Buschauer R, Ameismeier M, Koepke L, Denk T, Hirschenberger M, Kratzat H, Hayn M, Mackens-Kiani T, Cheng J, Straub JH, Sturzel CM, Frohlich T, Berninghausen O, Becker T, Kirchhoff F, Sparrer KMJ, Beckmann R. 2020. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. Science 369:1249–1255. https://doi.org/10.1126/science.abc8665.
33. Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler LA, Leibundgut M, Thiel V, Muhlemann O, Ban N. 2020. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. Nat Struct Mol Biol 27:959–966. https://doi.org/10.1038/s41594-020-0511-8.
34. Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738. https://doi.org/10.1038/nprot.2010.5.
35. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. Nat Methods 12:7–8. https://doi.org/10.1038/nmeth.3213.
36. Yang J, Zhang Y. 2015. Protein structure and function prediction using I-TASSER. Curr Protoc Bioinformatics 52:5.8.1–5.8.15. https://doi.org/10.1002/0471250953.bi0508s52.
37. Stadler K, Masignani V, Eickmann M, Becker S, Abrignani S, Klenk HD, Rappuoli R. 2003. SARS–beginning to understand a new virus. Nat Rev Microbiol 1:209–218. https://doi.org/10.1038/nrmicro775.
38. Skowronski DM, Astell C, Brunham RC, Low DE, Petric M, Roper RL, Talbot PJ, Tam T, Babiuk L. 2005. Severe acute respiratory syndrome (SARS): a year in review. Annu Rev Med 56:357–381. https://doi.org/10.1146/annurev.med.56.091103.134135.
39. Cinatl J, Morgenstern B, Bauer G, Chandra P, Rabenau H, Doerr HW. 2003. Treatment of SARS with human interferons. Lancet 362:293–294. https://doi.org/10.1016/S0140-6736(03)13973-6.
40. Spiegel M, Pichlmair A, Muhlberger E, Haller O, Weber F. 2004. The antiviral effect of interferon-beta against SARS-coronavirus is not mediated by MxA protein. J Clin Virol 30:211–213. https://doi.org/10.1016/j.jcv.2003.11.013.
41. Carbajo-Lozoya J, Ma-Lauer Y, Malesevic M, Theuerkorn M, Kahlert V, Prell E, von Brunn B, Muth D, Baumert TF, Drosten C, Fischer G, von Brunn A. 2014. Human coronavirus NL63 replication is cyclophilin A-dependent and inhibited by non-immunosuppressive cyclosporine A-derivatives including alisporivir. Virus Res 184:44–53. https://doi.org/10.1016/j.virusres.2014.02.010.
42. Dawar FU, Tu J, Khattak MN, Mei J, Lin L. 2017. Cyclophilin A: a key factor in virus replication and potential target for anti-viral therapy. Curr Issues Mol Biol 21:1–20. https://doi.org/10.21775/cimb.021.001.
43. de Wilde AH, Pham U, Posthuma CC, Snijder EJ. 2018. Cyclophilins and cyclophilin inhibitors in nidovirus replication. Virology 522:46–55. https://doi.org/10.1016/j.virol.2018.06.011.
44. Pfefferle S, Schopf J, Kogl M, Friedel CC, Muller MA, Carbajo-Lozoya J, Stellberger T, von Dall'Armi E, Herzog P, Kallies S, Niemeyer D, Ditt V, Kuri T, Zust R, Pumpor K, Hilgenfeld R, Schwarz F, Zimmer R, Steffen I, Weber F, Thiel V, Herrler G, Thiel HJ, Schwegmann-Wessels C, Pohlmann S, Haas J, Drosten C, von Brunn A. 2011. The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. PLoS Pathog 7:e1002331. https://doi.org/10.1371/journal.ppat.1002331.
45. Tanaka Y, Sato Y, Sasaki T. 2013. Suppression of coronavirus replication by cyclophilin inhibitors. Viruses 5:1250–1260. https://doi.org/10.3390/v5051250.
46. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY. 2016. Coronaviruses - drug discovery and therapeutic options. Nat Rev Drug Discov 15:327–347. https://doi.org/10.1038/nrd.2015.37.
47. Wathelet MG, Orr M, Frieman MB, Baric RS. 2007. Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. J Virol 81:11620–11633. https://doi.org/10.1128/JVI.00702-07.
48. Yount B, Curtis KM, Fritz EA, Hensley LE, Jahrling PB, Prentice E, Denison MR, Geisbert TW, Baric RS. 2003. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. Proc Natl Acad Sci U S A 100:12995–13000. https://doi.org/10.1073/pnas.1735582100.
49. Connor RF, Roper RL. 2007. Unique SARS-CoV protein nsp1: bioinformatics, biochemistry and potential effects on virulence. Trends Microbiol 15:51–53. https://doi.org/10.1016/j.tim.2006.12.005.
50. Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326. https://doi.org/10.1016/S0076-6879(97)76066-X.
51. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. J Appl Crystallogr 40:658–674. https://doi.org/10.1107/S0021889807021206.
52. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213–221. https://doi.org/10.1107/S0907444909052925.
53. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and

development of Coot. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493.
54. Schrodinger, LLC. 2015. The PyMOL molecular graphics system, version 2.0. Schrodinger, LLC, New York, NY.
55. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, Christie C, Dalenberg K, Duarte JM, Dutta S, Feng Z, Ghosh S, Goodsell DS, Green RK, Guranovic V, Guzenko D, Hudson BP, Kalro T, Liang Y, Lowe R, Namkoong H, Peisach E, Periskova I, Prlic A, Randle C, Rose A, Rose P, Sala R, Sekharan M, Shao C, Tan L, Tao YP, Valasatava Y, Voigt M, Westbrook J, Woo J, Yang H, Young J, Zhuravleva M, Zardecki C. 2019. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res 47:D464–D474. https://doi.org/10.1093/nar/gky1004.