Large-scale recombinant production of the SARS-CoV-2 proteome for high-throughput and structural biology applications

Nadide Altincekic[1,2,#], Sophie Marianne Korn[2,3,#], Nusrat Shahin Qureshi[1,2,#], Marie Dujardin[4,#], Martí Ninot-Pedrosa[4,#], Rupert Abele[5], Marie Jose Abi Saad[6], Caterina Alfano[7], Fabio C. L. Almeida[8,9], Islam Alshamleh[1,2], Gisele Cardoso de Amorim[8,10], Thomas K. Anderson[11], Cristiane Ano Bom[8,12], Chelsea Anorma[13], Jasleen Kaur Bains[1,2], Adriaan Bax[14], Martin Blackledge[15], Julius Blechar[1,2], Anja Böckmann[4,*], Louis Brigandat[4], Anna Bula[16], Matthias Bütikofer[6], Aldo Camacho Zarco[15], Teresa Carlomagno[17,18], Icaro Putinhon Caruso[8,9,19], Betül Ceylan[1,2], Apirat Chaikuad[20,21] Feixia Chu[22], Laura Cole[4], Marquise G. Crosby[23], Karthikeyan Dhamotharan[2,3], Isabella C. Felli[24,25], Jan Ferner[1,2], Yanick Fleischmann[6], Marie-Laure Fogeron[4], Nikolaos K. Fourkiotis[26], Christin Fuks[1], Boris Fürtig[1,2], Angelo Gallo[26], Santosh L. Gande[1,2], Juan Atilio Gerez[6], Dhiman Ghosh[6], Francisco Gomes-Neto[8,27], Oksana Gorbatyuk[28], Serafima Guseva[15], Carolin Hacker[29], Sabine Häfner[30], Bruno Hargittay[1,2], K. Henzler-Wildman[11], Jeffrey C. Hoch[28], Katharina Hohmann[1,2], Marie T. Hutchison[1,2], Kristaps Jaudzems[16], Vanessa de Jesus[1,2], Katarina Jovic[22], Janina Kaderli[6], Gints Kalniņš[31], Iveta Kaņepe[16], Robert N. Kirchdoerfer[11], John Kirkpatrick[17,18], Stefan Knapp[20,21], Robin Krishnathas[1,2], Felicitas Kutz[1,2], Susanne zur Lage[18], Roderick Lambertz[3], Andras Lang[30], Douglas Laurents[32], Lauriane Lecoq[4], Verena Linhard[1,2], Frank Löhr[2,33], Anas Malki[15], Luiza Mamigonian Bessa[15], Rachel W. Martin[13,23], Tobias Matzel[1,2], Damien Maurin[15], Seth W. McNutt[22], Nathane Cunha Mebus Antunes[8,9], Beat H. Meier[6], Nathalie Meiser[1], Miguel Mompeán[32], Elisa Monaca[7], Roland Montserret[4], Laura Moreno Perez[15], Celine Moser[34], Claudia Muhle-Goll[34], Thaís Cristtina Neves Martins[8,9], Xiamonin Ni[20,21], Brenna Norton-Baker[13], Roberta Pierattelli[24,25], Letizia Pontoriero[24,25], Yulia Pustovalova[28], Oliver Ohlenschläger[30], Julien Orts[6], Andrea Thompson Da Poian[9], Dennis J. Pyper[1,2], Christian Richter[1,2], Roland Riek[6], Angus Robertson[14], Anderson de Sá Pinheiro[8,12], Raffaele Sabbatella[7], Nicola Salvi[15], Krishna Saxena[1,2], Linda Schulte[1,2], Marco Schiavina[24,25], Harald Schwalbe[1,2,*], Mara Silber[34], Marcius da Silva Almeida[8,9], Marc A. Sprague-Piercy[23], Georgios A. Spyroulias[26], Sridhar Sreeramulu[1,2], Jan-Niklas Tants[2,3], Kaspars Tārs[31], Felix Torres[6], Sabrina Töws[3], Miguel Á.Treviño[32], Sven Trucks[1], Aikaterini C. Tsika[26], Krisztina Varga[22], Anna Wacker[1,2], Ying Wang[17], Marco E. Weber[6], Julia E. Weigand[35], Christoph Wiedemann[36], Julia Wirmer-Bartoschek[1,2], Maria Alexandra Wirtz Martin[1,2], Johannes Zehnder[6], Martin Hengesbach[1,*], and Andreas Schlundt[2,3,*].

From the [1]Institute for Organic Chemistry and Chemical Biology, Goethe University Frankfurt am Main, Frankfurt 60438, Germany, [2]Center of Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt am Main, Frankfurt 60438, Germany, [3]Institute for Molecular Biosciences, Johann Wolfgang Goethe-University Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt/M., Germany, [4]Molecular Microbiology and Structural Biochemistry (MMSB), UMR 5086, CNRS/Lyon University, France, [5]Institute for Biochemistry, Johann Wolfgang Goethe-University Frankfurt, Max von Laue Str. 7, 60438 Frankfurt/M, Germany, [6]ETH Zurich, Swiss Federal Institute of Technology, Laboratory of Physical Chemistry, Vladimir-Prelog-Weg 2 8093 Zürich, Switzerland, [7]Structural Biology and Biophysics Unit, Fondazione Ri.MED, Palermo, Italy, [8]National Center of Nuclear Magnetic Resonance (CNRMN, CENABIO), Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil, [9]Institute of Medical Biochemistry, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil, [10]Multidisciplinary Center for Research in Biology (NUMPEX), Campus Duque de Caxias Federal University of Rio de Janeiro, Duque de Caxias, RJ, Brazil, [11]Institute for Molecular Virology, University of Wisconsin-Madison, [12]Institute of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil, [13]Department of Chemistry, University of California, Irvine, California 92697-2025, United States, [14]LCP, NIDDK, NIH, [15]Univ. Grenoble Alpes, CNRS, CEA, IBS, Grenoble, F-38000, France, [16]Latvian Institute of Organic Synthesis, Aizkraukles 21, LV-1006 Riga, Latvia, [17]BMWZ and Institute of Organic Chemistry, Leibniz University Hannover, Schneiderberg 38, D-30167 Hannover, [18]Group of NMR-based Structural Chemistry, Helmholtz Centre for Infection Research, Inhoffen-strasse 7, D-38124 Braunschweig, [19]Multiuser Center for Biomolecular Innovation (CMIB), Department of Physics, São Paulo State University (UNESP), São José do Rio Preto, SP,

Brazil, [20]Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Max-von-Laue-Straße 9, Frankfurt 60438, Germany, [21]Structural Genomics Consortium, Buchmann Institute for Molecular Life Sciences (BMLS), Max-von-Laue-Straße 15, Frankfurt 60438, Germany, [22]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, NH, 03824, [23]Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697-3900, United States, [24]Magnetic Resonance Centre (CERM), University of Florence, Sesto Fiorentino, Italy, [25]Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy, [26]Department of Pharmacy, University of Patras, GR-26504, Patras, Greece, [27]Laboratory of Toxicology, Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, RJ, Brazil, [28]Department of Molecular Biology and Biophysics, UConn Health, Farmington CT, USA, [29]Signals GmbH & Co. KG, Graf-von-Stauffenberg-Allee 83, 60438 Frankfurt/M, Germany, [30]Leibniz Institute on Aging – Fritz Lipmann Institute (FLI) Beutenbergstr. 11 D-07745 Jena, Germany, [31]Latvian Biomedical Research and Study Centre, Ratsupites 1, LV-1067 Riga, Latvia, [32]"Rocasolano" Institute for Physical Chemistry (IQFR), Spanish National Research Council (CSIC), c/ Serrano 119, 28006, Spain, [33]Institute of Biophysical Chemistry, Goethe University Frankfurt am Main, Frankfurt 60438, Germany, [34]IBG-4, Karlsruhe Institute of Technology, Postfach 3640, 76021 Karlsruhe, [35]Department of Biology, Technical University of Darmstadt, Schnittspahnstrasse 10, 64287 Darmstadt, Germany, and [36]Martin Luther University Halle-Wittenberg, Institute of Biochemistry and Biotechnology, Charles Tanford Protein Centre, Kurt-Mothes-Str. 3a, D-06120 Halle/Saale, Germany.

[#] These authors contributed equally.

*Corresponding Authors: Anja Böckmann, Martin Hengesbach, Andreas Schlundt, Harald Schwalbe

E-mail: a.bockmann@ibcp.fr, hengesbach@nmr.uni-frankfurt.de, schlundt@bio.uni-frankfurt.de, schwalbe@nmr.uni-frankfurt.de

Running title: Extensive collection of SARS-CoV-2 protein purification protocols

Keywords: COVID-19, SARS-CoV-2, nonstructural proteins, structural proteins, accessory proteins, intrinsically disordered regions and proteins, cell-free synthesis, NMR spectroscopy

Abstract

The highly infectious disease COVID-19 caused by the *Betacoronavirus* SARS-CoV-2 poses a severe threat to humanity, and demands for redirection of scientific efforts and criteria to organized research projects. The international *Covid19-NMR* consortium seeks to provide such new approaches by gathering scientific expertise worldwide. In particular, making available viral proteins and RNAs will pave the way to understanding the SARS-CoV-2 molecular components in detail. The research in *Covid19-NMR* and the resources provided through the consortium are fully disclosed to accelerate access and exploitation. NMR investigations of the viral molecular components are designated to provide the essential basis for further work, including macromolecular interaction studies and high-throughput drug screening.

Here, we present the extensive catalogue of a holistic SARS-CoV-2 protein preparation approach based on the consortium's collective efforts. We provide protocols for the large-scale production of more than 80% of all SARS-CoV-2 proteins or essential parts of them. Several of the proteins were produced in more than one laboratory, demonstrating the high interoperability between NMR groups worldwide. For the majority of proteins, we can produce isotope-labeled samples of HSQC-grade. Together with several NMR-chemical shift assignments made publicly available on *covid19-nmr.com*, we here provide highly valuable resources for the production of SARS-CoV-2 proteins in isotope labeled form.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, SCoV2) is the cause of the early 2020 pandemic coronavirus lung disease 2019 (COVID-19) and belongs to the *Betacoronaviruses*, a genus of the *Coronaviridae* family covering the α–δ genera (1). The large RNA genome of SCoV2 has an intricate, highly condensed arrangement of coding sequences (2). Sequences starting with the main start codon contain an open reading frame 1 (ORF1) which codes for two distinct, large polypeptides (pp), whose relative abundance is governed by the action of an RNA pseudoknot structure element. Upon RNA folding, this element causes a -1 frameshift to allow the continuation of translation, resulting in generation of a 7096 amino acid, 794 kDa polypeptide. If the pseudoknot is not formed, expression of the first ORF generates a 4405 amino acid, 490 kDa polypeptide. Both the short and long polypeptides translated from this ORF (pp1a and pp1ab, respectively) are posttranslationally cleaved by virus-encoded proteases into functional, non-structural proteins (nsps). ORF1a encodes eleven nsps, and ORF1ab additionally the nsps12-16. The downstream ORFs encode structural proteins (S, E, M and N) that are essential components for the synthesis of new virus particles. In between those, additional proteins (accessory/auxiliary factors) are encoded, for which sequences partially overlap (3) and whose identification and classification is a matter of ongoing research (4,5). In total, the number of identified peptides or proteins generated from the viral genome is at least 28 on the evidence level, with an additional set of smaller proteins or peptides being predicted with high likelihood.

High-resolution studies of SCoV and SCoV2 proteins have been conducted using all canonical structural biology approaches, from X-ray crystallography on proteases (6) and methyltransferases (7), over cryo-EM of the RNA polymerase (8,9) to liquid-state (10-17) and solid-state NMR spectroscopy of transmembrane (TM) proteins (18). These studies have significantly improved our understanding on functions of molecular components, and they all rely on the recombinant production of viral proteins in high amount and purity.

Apart from structures, purified SCoV2 proteins are required for experimental and preclinical approaches designed to understand the basic principles of the viral life cycle and processes underlying viral infection and transmission. Approaches range from studies on immune responses (19), antibody identification (20), and interactions with other proteins or components of the host cell (21,22). These examples highlight the importance of broad approaches for the recombinant production of viral proteins.

The 2020-founded research consortium *Covid19-NMR* seeks to support the search for anti-viral drugs using an NMR-based screening approach. This requires the large-scale production of all druggable proteins and RNAs and their NMR resonance assignments. The latter will enable solution structure determination of viral proteins and RNAs for rational drug design as well as the fast mapping of

compound binding sites. We have recently produced and determined secondary structures of SCoV2 RNA *cis*-regulatory elements in near completeness by NMR spectroscopy, validated by DMS-MaPseq (23), to provide a basis for RNA-oriented fragment screens with NMR.

We here compile a compendium of more than 50 protocols (see SI1-SI23) for the production and purification of 23 of the 30 SCoV2 proteins or fragments thereof (summarized in Tables 1 and 2). We defined those 30 proteins as existing or putative ones to our current knowledge (see later discussion). This compendium has been generated in a coordinated and concerted effort between > 30 labs worldwide (Table S1), with the aim of providing pure mg-amounts of SCoV2 proteins. Our protocols include the rational strategy for construct design (if applicable guided by available homologue structures), optimization of expression, solubility, yield, purity and suitability for follow-up work, with a focus on uniform stable-isotope labeling.

We also present protocols for a number of accessory as well as the structural E and M proteins that could only be produced using wheat germ cell-free protein synthesis (WG-CFPS). In SCoV2, accessory proteins represent a class of mostly small and relatively poorly characterized proteins, mainly due to their difficult behavior in classical expression systems. They are often found in inclusion bodies, and difficult to purify in quantities adequate for structural studies. We thus here exploit cell-free synthesis, mainly based on previous reports on production and purification of viral membrane proteins in general (24-26). Besides yields compatible with structural studies, ribosomes in WG extracts further possess an increased folding capacity (27), favorable for those more complicated proteins.

We exemplify in more detail the optimization of protein production, isotope labeling and purification for proteins with different individual challenges: the nucleic acid binding domain of nsp3e, the main protease nsp5, and several auxiliary proteins. For the majority of produced and purified proteins, we achieve > 95% purity, and provide ^{15}N-HSQC spectra as the ultimate quality measure. We also provide additional suggestions for challenging proteins, where our protocols represent a unique resource and starting point exploitable by other labs.

Results

In the following, we provide protocols for the purification of SCoV2 proteins sorted into i) non-structural proteins and ii) structural proteins together with accessory ORFs. Table 1 shows an overview of expression constructs. We use a consequent terminology of those constructs, which is guided by domains, intrinsically disordered regions (IDRs) or other particularly relevant sequence features within them. This study uses the SCoV2 NCBI reference genome entry NC_045512.2, identical to GenBank entry MN908947.3 (2) unless denoted differently in the respective protocols. Any relevant definition of boundaries can also be found in the SI protocols.

As applicable for a major part of our proteins, we further define a standard procedure for the purification of soluble His-tagged proteins that are obtained through the sequence of **I**MAC, TEV/Ulp1 **P**rotease cleavage, **R**everse IMAC and **S**ize exclusion chromatography, eventually with individual alterations, modifications or additional steps. For convenient reading, we will thus use the abbreviation IPRS to avoid redundant protocol description. Details for every protein, including detailed expression conditions, buffers, incubation times, supplements, storage conditions, yields, and stability can be found in the respective SI1-SI23 (see also Tables S1 and S2) and the Tables 1 and 2. Entries in the tables are linked to the respective protocols for convenient download.

Non-structural proteins

We have approached and challenged the recombinant production of a large part of the SCoV2 nsps (Fig. 1), with great success (Table 2). We excluded nsp4 and nsp6 (TM proteins), which are little characterized and do not reveal soluble, folded domains by prediction (28,29). The function of the very short (13 aa) nsp11 is unknown, and it seems to be a mere copy of the nsp12 amino-terminal residues, remaining as protease cleavage product of ORF1a. Further, we left out the RNA-dependent RNA polymerase nsp12 in our initial approach because of its size (> 100 kDa) and known unsuitability for heterologous recombinant production in bacteria. Work on NMR-suitable nsp12 bacterial production is ongoing, while other expert labs have succeeded in purifying nsp12 for cryo-EM applications in different systems (9,30). For the remainder of nsps, we here provide protocols for full-length (fl) proteins or relevant fragments of them.

Nsp1

Nsp1 is the very N-terminus of the polyproteins pp1a and pp1ab and one of the most enigmatic viral proteins, expressed only in α- and β-CoVs (31). Interestingly, nsp1 displays the highest divergence in sequence and size among different CoVs, justifying it as genus-specific marker (32). It functions as host shutoff factor by suppressing innate immune functions as well as host gene expression (33-35). This suppression is achieved by an interaction of the nsp1 C-terminus with the mRNA entry tunnel within the 40S subunit of the ribosome (35,36).

As summarized in Table 1, fl domain boundaries of nsp1 were chosen to contain the first 180 amino acids, in analogy to its closest homologue from SCoV (32). In addition, a shorter construct was designed, encoding only the globular core domain (GD, aa 13-127) suggested by the published SCoV nsp1 NMR structure (10). His-tagged fl nsp1 was purified using the IPRS approach. Protein quality was confirmed by the available HSQC spectrum (Fig. 2). Despite the flexible C-terminus, we were able to accomplish a near-complete backbone assignment (manuscript submitted).

Interestingly, the nsp1 GD was found problematic in our hands despite good expression. We observed insolubility although buffers were used accordingly to the homologue SCoV nsp1 GD (10). Nevertheless, using a protocol comparable to the one for fl nsp1, we were able to record an HSQC spectrum proving a folded protein (Fig. 2).

Nsp2

Nsp2 has been suggested to interact with host factors involved in intracellular signaling (37,38). The precise function, however, is insufficiently understood. Despite its potential dispensability for viral replication in general, it might be a valuable model to gain insights into virulence, due to its possible involvement in regulation of global RNA synthesis (39). We provide here a protocol for the purification of the C-terminal IDR (CtDR) of nsp2 from residues 557 to 601, based on disorder predictions (PrDOS (40)). The His-Trx-tagged peptide was purified by IPRS. Upon dialysis, two IEC steps were performed, first anionic, then cationic, with good final yields (Table 1). Stability and purity were confirmed by an HSQC spectrum (Fig. 2) and a complete backbone assignment (41) (Table 2).

Nsp3

Nsp3, the largest nsp (32), is composed of a plethora of functionally related, yet independent subunits. After cleavage of nsp3 from the full-length ORF1-encoded polypeptide chain, it displays a 1945-residue multi-domain protein, with individual functional entities that are sub-classified from nsp3a to nsp3e followed by the ectodomain embedded in two transmembrane regions and the very C-terminal CoV-Y domain. The soluble nsp3a-3e domains are linked by various types of linkers with crucial roles in the viral life cycle and are located in the so-called viral cytoplasm, which is separated from the host cell after budding off the endoplasmic reticulum and contains the viral RNA (42). Remarkably, the nsp3c sub-structure comprises three sub-domains, making nsp3 the most complex SCoV2 protein. The precise function and eventual RNA-binding specificities of nsp3 domains are not yet understood. We here focus on the nsp3 domains a-e and provide elaborated protocols for additional constructs carrying relevant linkers or combinations of domains (Table 1). Moreover, we additionally present a convenient protocol for the purification of the C-terminal CoV-Y domain.

Nsp3a: The N-terminal portion of nsp3 is comprised of an ubiquitin-like (Ubl) structured domain and a subsequent acidic IDR. Besides its ability to bind ssRNA (43), nsp3a has been reported to interact with the nucleocapsid (44,45), playing a potential role in virus replication. We here provide protocols for the purification of both the Ubl (aa 1-111) as well as fl-nsp3a (aa 1-206), including the acidic IDR (Table 1). Domain boundaries were defined in resemblance to the published NMR structure of SCoV nsp3a (43). His-tagged fl nsp3a and GST-tagged nsp3a Ubl were each purified via the IPRS approach. Nsp3a Ubl yielded mM sample concentrations and displays a well dispersed HSQC spectrum (Fig. 3). Notably, the herein described protocol also enables to purify fl nsp3a including the acidic IDR (Tables 1 and 2). Despite the unstructured IDR overhang, the excellent protein quality and stability allowed for near-complete backbone assignment (Fig. 3, assignment letter in revision).

Nsp3b: Nsp3b is an ADP-ribose phosphatase macrodomain and potentially plays a key role in viral replication. Moreover, the de-ADP ribosylation function of nsp3b protects SCoV2 from antiviral host immune response, making nsp3b a promising drug target (46). As summarized in Table 1, the domain boundaries of the herein investigated nsp3b are residues 207-376 of the nsp3 primary sequence and were identical to published crystal structures with PDB entries 6YWM and 6YWL (unpublished). For purification, we used the IPRS

approach, which yielded pure fl nsp3b (Table 2). Fl nsp3b displays well-dispersed HSQC spectra making this protein an amenable target for NMR-structural studies. In fact, we recently reported near-to-complete backbone assignments for nsp3b in its apo and ADP-ribose bound form (13).

Nsp3c: The SARS unique domain (SUD) of nsp3c has been described as distinguishing feature of SCoVs (32). However, similar domains in more distant CoVs, such as MHV or MERS, have been reported recently (47,48). Nsp3c comprises three distinct globular domains, termed SUD-N, SUD-M and SUD-C, according to their sequential arrangement: N-terminal (N), middle (M) and C-terminal (C). SUD-N and SUD-M develop a macrodomain fold similar to nsp3b and are described to bind G-quadruplexes (49), while SUD-C preferentially binds to purine-containing RNA (50). Domain boundaries for SUD-N and -M, as well as for the tandem domain SUD-NM were defined in analogy to the SCoV homologue crystal structure (49). Those for SUD-C and the tandem SUD-MC were based on NMR solution structures of corresponding SARS CoV homologues (Table 1) (50). SUD-N, SUD-C and SUD-NM were purified using GST affinity chromatography, whereas SUD-M and SUD-MC were purified using His affinity chromatography. Removal of the tag was achieved by thrombin cleavage and final samples of all domains were prepared subsequent to SEC. Except for SUD-M, all constructs were highly stable (Table 2). Overall protein quality allowed for the assignment of backbone chemical shifts for the three single domains (17) as well as good resolved HSQC spectra also for the tandem domains (Fig. 3).

Nsp3d: Nsp3d comprises the papain-like protease (PL^{pro}) domain of nsp3 and, hence, one of the two SCoV2 proteases that are responsible for processing the viral polypeptide chain and generating functional proteins (51). The domain boundaries of PL^{pro} within nsp3 are set by residues 743 and 1060 (Table 1). The protein is particularly challenging, as it is prone to misfolding and rapid precipitation. We prepared His-tagged and His-SUMO-tagged PL^{pro}. The His-tagged version mainly remained in the insoluble fraction. Still, mg quantities could be purified from the soluble fraction, however, greatly misfolded. Fusion to SUMO significantly enhanced protein yield of soluble PL^{pro}. The His-SUMO-tag allowed simple IMAC purification, followed by cleavage with Ulp1 and isolation of cleaved PL^{pro} via a second IMAC. A final purification step using gel filtration led to pure PL^{pro} of both unlabeled and ^{15}N-labeled species (Table 2). The latter has allowed for the acquisition of a promising amide correlation spectrum (Fig. 3).

Nsp3e: Nsp3e is unique to *Betacoronaviruses* and consists of a nucleic acid binding domain (NAB) and the so-called group 2-specific marker (G2M) (52). Structural information is rare; while the G2M is predicted to be intrinsically disordered (53), the only available experimental structure of the nsp3e NAB was solved from SCoV by the Wüthrich lab using solution NMR (11). We here used this structure for a sequence-based alignment to derive reasonable domain boundaries for the SCoV2 nsp3e NAB (Fig. 5a, b). The high sequence similarity suggested to use nsp3 residues 1088-1203 (Table 1). This polypeptide chain was encoded in expression vectors comprising His- and His-GST tags, both cleavable by TEV protease. Both constructs showed excellent expression, suitable for the IPRS protocol (Fig. 5c). Finally, a homogenous NAB species, as supported by the final gel of pooled samples (Fig. 5d), was obtained. The excellent protein quality and stability are supported by the available HSQC (Fig. 3), and a published backbone assignment (14).

Nsp3Y: Nsp3Y is the most C-terminal domain of nsp3 and exists in all coronaviruses (52,54). Together, though, with its preceding regions G2M, TM 1, the ectodomain, TM2, and the Y1-domain it has evaded structural investigations so far. The precise function of the CoV-Y domain remains unclear, but - together with the Y1- domain, might affect binding to nsp4 (55). We were able to produce and purify nsp3Y (CoV-Y) comprising amino acids 1638-1945 (Table 1) yielding 12 mg/L with an optimized protocol that keeps the protein in a final NMR buffer containing HEPES and lithium bromide. Although the protein still shows some tendency to aggregate and degrade (Table 2) and despite its relatively large size, the spectral quality is excellent (Fig. 3). Nsp3 CoV-Y appears suitable for an NMR backbone assignment carried out at lower concentrations in a deuterated background (ongoing).

Nsp5

The functional main protease nsp5 (M^{pro}) is a dimeric cysteine protease (56). Amino acid sequence and 3D structure of SCoV (PDB 1P9U (57)) and SCoV2 (PDB 6Y2E (6)) homologues are highly conserved (Fig. 6a,b). The dimer interface involves the N-termini of both monomers, which puts con-

siderable constraints on the choice of protein sequence for construct design regarding the N-terminus.

We thus designed different constructs differing in the N-terminus: the native N-terminus (wt), a GS mutant with the additional N-terminal residues glycine and serine as His-SUMO fusion, and a GHM mutant with the amino acids glycine, histidine and methionine located at the N-terminus with His-tag and TEV cleavage site (Fig. 6c). Purification of all proteins via the IPRS approach (Fig. 6d-e) yielded homogenous and highly pure protein, analyzed by PAGE (Fig. 6g), mass spectrometry, and 2D [^{1}H,^{15}N]-BEST TROSY spectra (Fig. 6h). Final yields are summarized in Table 2.

Nsp7 & nsp8

Both nsp7 and nsp8 are auxiliary factors of the polymerase complex together with the RNA-dependent RNA polymerase nsp12, and have high sequence homology with SCoV (100% and 99%, respectively) (58). For nsp7 in complex with nsp8 or for nsp8 alone, additional functions in RNA synthesis priming have been proposed (59,60). In a recent study including an RNA-substrate-bound structure (30), both proteins (with two molecules of nsp8 and one molecule of nsp7 for each nsp12 RNA polymerase) were found to be essential for polymerase activity in SCoV2. For both fl proteins, a previously established expression and IPRS purification strategy for the SCoV proteins (61) was successfully transferred, which resulted in decent yields of reasonably stable proteins (Table 2). Driven by its intrinsic oligomeric state, nsp8 showed some tendency towards aggregation, limiting the available sample concentration. The higher apparent molecular weight and limited solubility is also reflected in the success of NMR experiments. While we succeeded in a complete NMR backbone assignment of nsp7 (12), the quality of the spectra obtained for nsp8 is currently limited to the HSQC presented in Fig. 2.

Nsp9

The 12.4-kDa ssRNA-binding nsp9 is highly conserved among *Betacoronaviruses*. It is a crucial part of the viral replication machinery (62), possibly targeting the 3'-end stem-loop II (s2m) of the genome (63). Nsp9 adopts a fold similar to oligonucleotide/oligosaccharide-binding proteins (64), and structural data consistently uncovered nsp9 to be dimeric in solution (62,64-66). Dimer formation seems to be a prerequisite for viral replication (62) and influences RNA-binding (65) despite moderate affinity for RNA *in vitro* (66).

Based on the early available crystal structure of SCoV2 (PDB 6W4B, unpublished), we used the 113 aa full-length sequence of nsp9 for our expression construct (Table 1). Production of either His- or His-GST-tagged fl nsp9 yielded high amounts of soluble protein, both in natural abundance and ^{13}C-, ^{15}N-labeled form. Purification *via* IPRS approach enabled us to separate fl nsp9 in different oligomer states. The earliest eluted fraction represented higher oligomers and was contaminated with nucleic acids and was not possible to concentrate above 2 mg/mL. This was different for the subsequently eluting dimeric fl nsp9 fraction, which had a A260/280 ratio of below 0.7 and could be concentrated to > 5 mg/mL (Table 2). The excellent protein quality and stability is supported by the available HSQC (Fig. 2), and a near-complete backbone assignment (manuscript submitted).

Nsp10

The last functional protein encoded by ORF1a, nsp10, is an auxiliary factor for both, the methyltransferase/exonuclease nsp14 and the 2'-O-methyltransferase (MTase) nsp16. Whereas it is required for the MTase activity of nsp16 (7), it confers exonuclease activity to nsp14 in the RNA polymerase complex in SCoV (67). It contains two unusual zinc finger motifs (68), and was initially proposed to comprise RNA-binding properties. We generated a construct (Table 1) containing an expression and affinity purification tag on the N-terminus as reported for the SCoV variant (68). Importantly, additional Zn^{2+} ions present during expression and purification stabilize the protein significantly (15). The yield during isotope-labeling was high (Table 2), and tests in unlabeled rich medium showed the potential for yields exceeding 100 mg/L. These characteristics facilitated in-depth NMR analysis and a backbone assignment (15).

Nsp13

Nsp13 is a conserved ATP-dependent helicase that has been characterized as part of the RNA synthesis machinery, by binding to nsp12 (69). It represents an interesting drug target, for which the available structure (PDB 6ZSL) serves as an excellent basis (Table 1). The precise molecular function, however, has remained enigmatic, since it is not clear whether the RNA unwinding function is required for making ssRNA accessible for RNA synthesis (70), or whether it is required for proofreading and

backtracking (69). We obtained pure protein using a standard expression vector, generating a His-SUMO-tagged protein. Following Ulp1 cleavage, the protein showed limited protein stability in solution (Table 2).

Nsp14

Nsp14 contains two domains: an N-terminal exonuclease domain, and a C-terminal a MTase domain (67). The exonuclease domain interacts with nsp10 and provides part of the proofreading function that supports the high fidelity of the RNA polymerase complex (71). Several unusual features, such as the unusual zinc finger motifs, set it apart from other DEDD-type exonucleases (72), which are related to both nsp10 binding and catalytic activity. The MTase domain modifies the N7 of the guanosine cap of genomic and subgenomic viral RNAs, which is essential for the translation of viral proteins (36). The location of this enzymatic activity within the RNA synthesis machinery ensures that newly synthesized RNA is rapidly capped and thus stabilized. As a strategy, we used constructs, which allow co-expression of both nsp14 and nsp10 (pRSFDuet and pETDuet, respectively). Production of isolated fl nsp14 was successful, however, with limited yield and stability (Table 2). Expression of the isolated MTase domain resulted in soluble protein with 27.5 kDa mass that was amenable to NMR characterization (Fig. 2), although only under reducing conditions and in the presence of high (0.4 M) salt concentration.

Nsp15

The poly-U-specific endoribonuclease nsp15 was one of the very first SCoV2 structures deposited in the PDB [6VWW, (73)]. Its function has been suggested to be related to the removal of U-rich RNA elements, preventing recognition by the innate immune system (74), even though the precise mechanism remains to be established. The exact role of the three domains (N-terminal, middle, and C-terminal catalytic domain) also remains to be characterized in more detail (73). Here, the sufficient yield of fl nsp15 during expression supported purification of pure protein, which, however, showed limited stability in solution (Table 2).

Nsp16

The MTase reaction catalyzed by nsp16 is dependent on nsp10 as a cofactor (7). In this reaction, the 2'-OH group of nucleotide +1 in genomic and subgenomic viral RNA is methylated, preventing recognition by the innate immune system. Since both nsp14 and nsp16 are in principle susceptible to inhibition by methyltransferase inhibitors, a drug targeting both enzymes would be highly desirable (75). Nsp16 is the last protein being encoded by ORF1ab, and only its N-terminus is formed by cleavage by the M^{pro} nsp5. Employing a similar strategy as for nsp14, nsp16 constructs were designed with the possibility of nsp10 co-expression. Expression of fl nsp16 resulted in good yields, both when expressed isolated and together with nsp10. The protein, however, is in either case unstable in solution, and highly dependent on reducing buffer conditions (Table 2). The purification procedures of nsp16 were adapted with minor modifications from a previous X-ray crystallography study (76).

Structural proteins and accessory ORFs

Besides establishing expression and purification protocols for the nsps, we also developed protocols and obtained pure mg quantities of the SCoV2 structural proteins E, M and N, as well as literally all accessory proteins. With the exception of the relatively well-behaved nucleocapsid (N) protein, SCoV2 E, M and the remaining accessory proteins represent a class of mostly small, and relatively poorly characterized proteins, mainly due to their difficult behavior in classical expression systems.
We used wheat-germ cell-free protein synthesis (WG-CFPS) for the successful production, solubilization, purification, and, in part, initial NMR spectroscopic investigation of ORF3a, ORF6, ORF7b, ORF8, ORF9b and ORF14 accessory proteins, as well as E and M in mg quantities using the highly efficient translation machinery extracted from wheat germs (Fig. 7a-d).

ORF3a

Protein from ORF3a in SCoV2 corresponds to accessory protein 3a in SCoV, with homology of more than 70 % (Table 1). It has 275 amino acids, and its structure has recently been determined (77). The structure of SCoV2 3a displays a dimer, but it can also form higher oligomers. Each monomer has three transmembrane (TM) helices and a cytosolic β-strand rich domain. SCoV2 ORF3a is a cation channel, and its structure has been solved by electron microscopy in nanodiscs. In SCoV, 3a is a structural component and was found in recombinant virus-like particles (78), but is not explicitly needed for their formation. The major challenge for NMR

studies of this largest accessory protein is its size, independent of its employment in solid state or solution NMR spectroscopy.
As most other accessory proteins described in the following, ORF3a has been produced using WG-CFPS, and was expressed in soluble form in presence of Brij-58 (Figure 7c). It co-purified with a small heat-shock protein of the HSP20 family from the wheat-germ extract. The here described protocol is highly similar for the other cell-free synthesized accessory proteins. Where NMR spectra have been reported, the protein has been produced in a $^{2}H,^{13}C,^{15}N$ uniformly labeled form; otherwise, natural abundance amino acids were added to the reaction. The proteins were further affinity purified in one step using Strep-Tactin resin, through the Strep-tag II fused to their N- or C-terminus. For membrane proteins, protein synthesis and also purification were done in the presence of detergent.
About half a milligram of pure protein was generally obtained per mL WG extract, and up to 3 mL wheat germ extract have been used to prepare NMR samples.

ORF3b
The ORF3b protein is a putative protein stemming from a short open reading frame (57 amino acids) with no homology to existing SCoV proteins (79). Indeed, ORF3b gene products of SARS-CoV-2 and SARS-CoV are considerably different, with one of the distinguishing features being the presence of premature stop codons, resulting in the expression of a drastically shortened ORF3b protein (80). However, the SARS-CoV-2 nucleotide sequence after the stop codon shows a high similarity to the SARS-CoV ORF3b. Different C-terminal truncations seem to play a role in the IFN-Antagonistic Activity of ORF3b (80). ORF3b is the only protein which, using WG-CFPS, was not synthetized at all, i.e. it was neither observed in the total cell-free reaction nor in supernatant or pellet. This might be due to the premature stop codon which was not considered. Constructs of ORF3b thus need to be redesigned.

ORF4 (Envelope protein, E)
The SCoV2 envelope (E) protein is a small (75 amino acids), integral membrane protein involved in several aspects of the virus' life cycle, such as assembly, budding, envelope formation, and pathogenecity, as recently reviewed in (81). Structural models for SCoV (82) and the transmembrane helix of SCoV2 (18) E have been established. The structural models show a pentamer with a transmembrane helix. The C-terminal part is polar, with charged residues interleaved, and is positioned on the membrane surface in SCoV. E was produced in a similar manner to ORF3a, using addition of detergent to the cell-free reaction.

ORF5 (Membrane glycoprotein, M)
The M protein is the most abundant protein in the viral envelope, and is believed to be responsible for maintaining the virion in its characteristic shape (83). M is a glycoprotein and aa sequence analyses predict three domains: a C-terminal endodomain, a transmembrane domain with three predicted helices, and a short N-terminal ectodomain. M is essential for viral particle assembly. Intermolecular interactions with the other structural proteins, N and S to a lesser extent, but most importantly E (84), seem to be central for virion envelope formation in coronaviruses, as M alone is not sufficient. Evidence has been presented that M could adopt two conformations, elongated and compact, and that the two forms fulfill different functions (85). The lack of more detailed structural information is in part due to its small size, close association with the viral envelope, and a tendency to form insoluble aggregates when perturbed (85). The M protein is readily produced using cell-free synthesis in the presence of detergent; as ORF3a, it co-purified with a small heat-shock protein of the HSP20 family (Figure 7b). Membrane-reconstitution will likely be necessary to study this protein.

ORF6
The ORF6 protein is incorporated into viral particles, and is also released from cells (83). It is a small protein (61 aa), which has been found to concentrate at the endoplasmic reticulum and Golgi apparatus. In a murine coronavirus model, it was shown that expressing ORF6 increased virulence in mice (86), and results indicate that ORF6 may serve an important role in the pathogenesis during SCoV infection (78). Also, it showed to inhibit expression of certain STAT1-genes critical for the host immune response, and could contribute to the immune evasion. ORF6 expresses very well in WG-CFPS; the protein was fully soluble with detergents and partially soluble without, and was easily purified in presence of detergent, but less efficiently in absence thereof. Solution NMR spectra in the presence of detergent display narrow, but few resonances, which correspond, in addition to the C-terminal

STREP-tag, to the very C-terminal ORF6 protein residues.

ORF7a

SCoV2 protein 7a (121 aa) shows over 85% homology with the SCoV protein 7a. While the SCoV2 7a protein is produced and retained intracellularly, SCoV protein 7a has also been shown to be a structural protein incorporated into mature virions (78). 7a is one of the accessory proteins, of which a (partial) structure has been determined at high resolution for SCoV2 (PDB 6W37). However, the very N-terminal signal peptide, and the C-terminal membrane anchor, both highly hydrophobic, have not been determined experimentally yet.

Expression of ORF7a with a GB1 tag (87) was expected to produce reasonable yields. The IPRS purification resulted in a highly stable protein, as evidenced by the NMR data obtained (Fig. 4).

ORF7b

Protein ORF7b is associated with viral particles in SARS (78). Protein 7b is one of the shortest ORFs with 43 residues. It shows a long hydrophobic stretch, which might correspond to a transmembrane segment. It shows over 93% sequence homology with a bat coronavirus 7b protein (78). There, the cysteine residue in the C-terminal part is not conserved, which might facilitate structural studies. ORF7b has been synthesized successfully both from bacteria and by WG-CFPS in the presence of detergent, and could be purified using a STREP-tag (Table 2). Due to necessity of solubilizing agent and its obvious tendency to oligomerize structure determination, fragment screening, and interaction studies are challenging. However, we were able to record a first promising HSQC as shown in Fig. 4.

ORF8

ORF 8 is believed to be responsible for the evolution of *Betacoronaviruses* and their species jumps (88) as well as to have a role in depressing the host response (89). ORF 8 (121 aa) from SCoV2 does not apparently exist in SCoV on the protein level, despite the existence of a putative ORF. The sequences of the two homologues only show limited identity, with the exception of a small 7 aa segment, where in SCoV the glutamate is replaced by an aspartate. It, however, aligns very well with several coronaviruses endemic to animals, including Paguma and Bat (79). The protein comprises a hydrophobic peptide at its very N-terminus, likely corresponding to a signal peptide; the remaining part does not show any specific sequence features. Its structure has been determined (PDB 7JTL), and shows a similar fold to ORF7a (90). In this study, ORF8 has been used both with (fl) and without signal peptide (ΔORF8). We first tested production of ORF8 in *E. coli*, but yields were low because of insolubility. Both ORF8 versions have then been synthesized in the cell-free system and were soluble in the presence of detergent. Solution-NMR spectra, however, indicate that the protein is either forming oligomers or aggregates.

ORF9a (Nucleocapsid protein, N)

The nucleocapsid protein (N) is important for viral genome packaging (91). The multifunctional RNA-binding protein plays crucial roles in the viral life cycle (92) and its domain architecture is highly conserved among coronaviruses. It comprises the N-terminal intrinsically disordered region (IDR1), the N-terminal RNA-binding globular domain (NTD), a central serine/arginine (SR)-rich intrinsically disordered linker region (IDR2), the C-terminal dimerization domain (CTD), and a C-terminal intrinsically disordered region (IDR3) (93).

N represents a highly promising drug target. We thus focused our efforts not exclusively on the NTD and CTD alone, but in addition also provide protocols for IDR-containing constructs within the N-terminal part.

NTD: The NTD is the RNA-binding domain of the nucleocapsid (93). It is embedded within IDRs, functions of which have not yet been deciphered. Recent experimental and bioinformatic data indicate an involvement in liquid-liquid phase separation (94).

For the NTD, several constructs were designed, also considering the flanking IDRs (Table 1). In analogy to the available NMR (PDB 6YI3, (95)) and crystal (PDB 6M3M, (93)) structures of the SCoV2 NTD, boundaries for the NTD and the NTD-SR domains were designed to span residues 44-180 and 44-212, respectively. In addition, an extended IDR1-NTD-IDR2 (residues 1-248) was designed, including the N-terminal disordered region (IDR1), the NTD domain, and the central disordered linker (IDR2) that comprises the SR region. His-tagged NTD and NTD-SR were purified using IMAC. Final samples were achieved after reverse IMAC of TEV-cleaved protein and yielded approx.

3 mg/L in ^{15}N-labelled minimal medium. High protein quality and stability is supported by the available HSQC spectra (Fig. 4).
The untagged IDR1-NTD-IDR2 was purified by IEC and yielded high amounts of ^{13}C,^{15}N-labeled samples of 12 mg/L for further NMR investigations. The quality of our purification is confirmed by the available HSQC (Fig. 4), and a near-complete backbone assignment of the two intrinsically disordered regions was achieved (manuscript submitted). Notably, despite the structurally and dynamically heterogeneous nature of the N protein, the mentioned N constructs revealed a very good long-term stability as shown in Table 2.

CTD: Multiple studies on the SCoV2 CTD, including recent crystal structures (96,97), confirm the domain as dimeric. Its ability to self-associate seems to be necessary for viral replication and transcription (91). In addition, the CTD was shown to, presumably non-specifically, bind ssRNA (97).
Domain boundaries for the CTD were defined to comprise amino acids 247-364 (Table 1), in analogy to the NMR structure of the CTD from SCoV (PDB 2JW8, (98)). Gene expression of His- or His-GST-tagged CTD yielded high amounts of soluble protein. Purification was achieved *via* IPRS. The CTD eluted as a dimer judged by its retention volume on the size exclusion column and yielded good amounts (Table 2). The excellent protein quality and stability is supported by the available HSQC spectrum (Fig. 4), and a near-complete backbone assignment (16).

ORF9b
Protein 9b (97 aa) shows 73% sequence homology to the SCoV and also to Bat virus (bat-SL-CoV-ZXC21) 9b protein (79). The structure of SCoV2 ORF9b has been determined at high resolution (PDB 6Z4U). Still, a significant portion of the structure was not found to be well ordered. The protein shows a β-sheet-rich structure and a hydrophobic tunnel, in which bound lipid was identified. How this might relate to membrane binding is not fully understood at this point. The differences in sequence between SCoV and SCoV2 are mainly located in the very N-terminus, which was not resolved in the structure (PDB 6Z4U). Another spot of deviating sequence not resolved in the structure is a solvent-exposed loop, which presents a potential interacting segment. ORF9b has been synthesized as a dimer (Fig. 7e) using WG-CFPS, in its soluble form. Spectra show a well-folded protein, and assignments are under way (Fig. 7f).

ORF14 (ORF9c)
ORF14 (73 aa) remains - at this point in time - hypothetical. It shows 89% homology with a bat virus protein (bat-SL-CoVZXC21). It shows a highly hydrophobic part in its C-terminal region, comprising two negatively charged residues, and a charged/polar N-terminus. The C-terminus is likely mediating membrane interaction. While ORF14 has been synthesized in the WG cell-free system in the presence of detergent, and solution NMR spectra have been recorded, they hint at an aggregated protein (Fig. 7e). Membrane-reconstitution of ORF14 revealed an unstable protein, which had been degraded during detergent removal.

ORF10
The ORF10 protein is comprised of 38 aa and is a hypothetical protein with unknown function (99). SCoV2 ORF10 displays 52.4% homology to SCoV ORF9b. The protein sequence is rich in hydrophobic residues, rendering expression and purification challenging. Expression of ORF10 as His-Trx-tagged or His-SUMO tagged fusion protein was possible; however, the ORF10 protein is poorly soluble and shows partial unfolding, even as uncleaved fusion protein. Analytical size exclusion chromatography hints at oligomerization under the current conditions.

Discussion

The ongoing SCoV2 pandemic and its manifestation as the COVID-19 disease call for an urgent provision of therapeutics that will specifically target viral proteins and their interactions with each other and RNAs, which are crucial for viral propagation. Two “classical” viral targets have been addressed in comprehensive approaches soon after the outbreak in December 2019: the viral protease nsp5 and the RNA-dependent RNA polymerase (RdRp) nsp12. While the latter turned out to be a suitable target using the repurposed compound Remdesivir (30), nsp5 is undergoing a broad structure-based screen against a battery of inhibitors in multiple places (6,100), but with, as of yet, limited outcome for effective medication. Hence, a comprehensive, reliable treatment of COVID-19 at any stage after infection has remained unsuccessful.

Further viral protein targets will have to be taken into account in order to provide inhibitors with increased specificity, efficacy, and as preparative starting points for potential following generations of (SARS-)CoVs. Availability of those proteins in recombinant, pure, homogenous and stable form in milligrams is therefore a prerequisite for follow-up applications like vaccination, high-throughput screening campaigns, structure determination and mapping of viral protein interaction networks. We here present, for the first time, a near-complete compendium of SCoV2 protein purification protocols that enable large amounts of pure proteins.

The *Covid19-NMR* consortium was launched by the motivation of providing NMR assignments of all SCoV2 proteins and RNA elements, and an enormous progress has been made since the outbreak of COVID-19 for both components (see Table 2 and (23)). Consequently, we have put our focus on producing proteins in stable-isotope labeled forms for NMR-based applications, e.g. the site-resolved mapping of interactions with compounds (101). Relevant to a broad scientific community, we here report our protocols to perfectly suite any downstream biochemical or biomedical application.

Overall success and protein coverage

As summarized in Table 2, we have successfully purified 80% of the SCoV2 proteins either in fl, or providing relevant fragments of the parent protein. Those include most of the nsps, where all of the known/predicted soluble domains have been addressed (Fig. 1). For a very large part, we were able to obtain protein samples of high purity, homogeneity and fold for NMR-based applications. We would like to point out a number of CoV proteins that, evidenced by their HSQCs, for the first time provide access to structural information, e.g. the PL^{pro} nsp3d and nsp3Y. Particularly for the nsp3 multi-domain protein, we here present soluble samples of almost the complete cytosolic region with more than 120 kDa in form of excellent 2D NMR spectra (Fig. 3), a major part of which fully backbone-assigned. We thus enable the exploitation of the largest and most enigmatic multifunctional SCoV2 protein through individual domains in solution, allowing to study their concerted behavior with single residue resolution. Similarly, for nsp2, we provide a promising starting point for studying the so-far neglected, often uncharacterized and apparently unstructured proteins.

Driven by the fast-spreading COVID-19, we initially left out proteins that either require advanced purification procedures (e.g. nsp12 and S) or where *a priori* information was limited (nsp4 and nsp6). This procedure seems justified with the time-saving approach of our effort in the favor of the less attended proteins. However, we are in the process of collecting protocols for the missing proteins.

Different complexities and challenges

The compilation of protein production protocols, initially guided by information from CoV homologues (Table 1), has confronted us with very different levels of complexity. With some prior expectation towards this, we have shared forces to quickly "work off" the highly conserved soluble and small proteins and soon put focus into the processing of the challenging ones. The difficulties in studying this second class of proteins is due to their limited sequence conservation, no prior information, large molecular weights, insolubility etc.

The nsp3e NAB represents one example where the available NMR structure of the SCoV homologue provided a *bona fide* template for selecting initial domain boundaries (Fig. 5). The transfer of information derived from SCoV was straightforward; the transferability included the available protocol for the production of comparable protein amounts and quality, given the high sequence identity. In such cases, we found ourselves merely to adapt protocols and optimize yields based on slightly different expression vectors and *E. coli* strains.

However, in some cases such transfer was unexpectedly not successful, e.g. for the short nsp1 GD. Despite intuitive domain boundaries with complete local sequence identity seen from the SCoV nsp1 NMR structure, it took considerable efforts to purify an analogous nsp1 construct, which is likely related to the impaired stability and solubility caused by a number of impacting amino acid exchanges within the domain's flexible loops. In line with that, currently available structures of SCoV2 nsp1 have been obtained by crystallography or cryo-EM and include different buffers. As such, our initial design was insufficient in terms of taking into account the parameters mentioned above. However, one needs to consider those particular differences between the nsp1 homologues as one of the most promising target sites for potential drugs as they appear to be hotspots in the CoV evolution and will have essential effects for the molecular networks, both in the virus and with the host (31,36,102,103).

A special focus was put on the production of the SCoV2 main protease nsp5, for which NMR-based screenings are ongoing. The main protease is critical in terms of inhibitor design as it appears under constant selection, and novel mutants remarkably influence structure and biochemistry of the protein (104). In the present study, expression of the different constructs allowed us to characterize the protein both in its monomeric as well as dimeric forms. Comparison of NMR spectra reveals that the constructs with additional amino acids (GS and GHM mutant) display marked structural differences to the wild type protein, while being structurally similar among themselves (Fig. 6h). The addition of two additional residues (GS) interferes with the dimerization interface, although being similar to its native N-terminal amino acids (SGFR). We also introduced an active site mutation that replaces cysteine 145 by alanine (105). Intriguingly, this active site mutation C145A, known to stabilize the dimerization of the main protease (106), supports dimer formation of the GS added construct (GS-nsp5 C145A) shown by its 2D NMR spectrum overlaying with the one of wild type nsp5 (SI4). The NMR results are in line with SEC-MALS analyses (Fig. 6f). Indeed, the additional amino acids at the N-terminus shift the dimerization equilibrium towards the monomer, whereas the mutation shifts it towards the dimer despite the N-terminal aa additions. This example underlines the need for a thorough and precise construct design and the detailed biochemical and NMR-based characterization of the final sample state. The presence of monomers vs. dimers will play an essential role for inhibitor search against SCoV2 proteins as exemplified by the particularly attractive nsp5 main protease target.

Exploiting non-bacterial expression

As a particular effort within this consortium, we included the so-far neglected accessory proteins using a structural genomics procedure supported by wheat-germ cell-free protein synthesis. This approach allowed previously to express a variety of difficult viral proteins in our hands (22-29,83-87). Within the workflow we especially highlight the straightforward solubilization of the membrane proteins through the addition of detergent to the cell-free reaction, which allowed to produce soluble protein in milligram amounts compatible with NMR studies. While home-made extracts were used here, very similar extracts are available commercially (Cell-Free Sciences, Japan) and can thus be implemented by any lab without prior experience. Also, a major benefit of the WG-CFPS system for NMR studies lies in the high efficiency and selectivity of isotopic labeling. In contrast to cell-based expression systems, only the protein of interest is produced (107), which allows to bypass extensive purification steps. In fact, a one-step affinity purification is in most cases sufficient, as shown for the different ORFs in this study. Samples could be produced for virtually all proteins, with the exception of the ORF3b construct used. With new recent insight into the stop codons present in this ORF, constructs will be adapted, which shall overcome the problems of ORF3b production (80).

For two ORFs, 7b and 8, we exploited a paralleled production strategy, i.e. both in bacteria and via cell-free synthesis. For those challenging proteins we were in principle able to obtain pure samples from either expression system. However, for ORF7b we found a strict dependency on detergents for follow-up work from both approaches. ORF8 showed a significantly better solubility when produced in WG extracts compared to bacteria. This shows the necessity of parallel routes to take, in particular for the understudied, biochemically non-trivial ORFs that might represent yet unexplored, but highly specific targets to consider in the treatment of COVID-19.

Downstream structural analysis of ORFs produced with CFPS remains challenging but promising progress is being made in the light of SCoV2. Some solution NMR spectra show the expected number of signals with good resolution (as for example ORF9b). As expected, however, most proteins, cannot be straightforward analyzed by solution NMR in their current form, as they exhibit too large objects after insertion into micelles and/or by inherent oligomerization. Cell-free synthesized proteins can be inserted into membranes through reconstitution (24,26,108-110). Reconstitution will thus be the next step for many accessory proteins, but also for M and E, which were well produced by WG-CFPS. We will also exploit the straight-forward deuteration in WG-CFPS (110-112) that circumvents proton back-exchange, rendering denaturation and refolding steps obsolete (113). Nevertheless, the herein presented protocols for the production of non-nsps by WG-CFPS instantly enable their employment in binding studies and screening campaigns and thus pose a significant contribution to

soon-to-come studies on SCoV2 proteins beyond the classical and convenient drug targets.
Altogether and judged by the ultimate need of exploiting recombinant SCoV2 proteins in vaccination and highly paralleled screening campaigns, we optimized sample amount, homogeneity and long-term stability of samples. Our freely accessible protocols and accompanying NMR spectra now offer a great resource to be exploited for the unambiguous and reproducible production of SCoV2 proteins for the intended applications.

Experimental procedures

Strains, plasmids and cloning

The rationale of construct design for all proteins can be found within the respective protocols in SI1-23. For bacterial production, *E. coli* strains and expression plasmids are given; for WG-CFPS, template vectors are listed. Protein coding sequences of interest have been either obtained as commercial, codon-optimized genes or, for shorter ORFs and additional sequences, annealed from oligonucleotides prior to insertion into the relevant vector. Sub-cloning of inserts, adjustment of boundaries and mutations of genes have been carried out by standard molecular biology techniques. All expression plasmids have been deposited in the AddGene databank containing information about coding sequences, restriction sites, fusion tags and vector backbones.

Protein production and purification

For SCoV2 proteins, we primarily used heterologous production in *E. coli*. Detailed protocols of individual full-length proteins, separate domains, combinations or particular expression constructs as listed in Table 1 can be found in the Supplementary Information (SI1-23).
The ORF3a, ORF6, ORF7b, ORF8, ORF9b and ORF14 accessory proteins, as well as the structural proteins M and E, were produced by wheat-germ cell-free protein synthesis (WG-CFPS) as described in the SI. In brief, transcription and translation steps have been performed separately, and detergent has been added for the synthesis of membrane proteins as described previously (25,114).

NMR spectroscopy

All amide correlation spectra, either HSQC- or TROSY-based, are representative examples. Details on their acquisition parameters and the raw data are freely accessible through https://covid19-nmr.de/ or upon request.

Data availability: Assignments of backbone chemical shifts have been deposited at BMRB for proteins as shown in Table 2, indicated by their respective BMRB IDs. All expression constructs will be deposited as plasmids with Addgene.

Acknowledgments
We thank Leonardo Gonnelli and Katharina Targaczewski for the valuable technical assistance. Part of this work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-05-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG CEA). We acknowledge the Advanced Technologies Network Center of University of Palermo to support infrastructures.

Funding and additional information
This work was supported by Goethe-University (Corona funds), the DFG-funded CRC: "Molecular principles of RNA-based regulation", DFG infrastructure funds (project numbers: 277478796, 277479031, 392682309, 452632086, 70653611) and the state of Hesse (BMRZ), the Fondazione CR Firenze (CERM) and for the IWB-EFRE-program 20007375. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871037. Andreas Schlundt is supported by DFG grant SCHL2062/2-1 and by the JQYA at Goethe through project number 2019/AS01. Work in the lab of Krisztina Varga was supported by a CoRE grant from the University of New Hampshire. The FLI is member of the Leibniz Association (WGL) and financially supported by the Federal Government of Germany and the State of Thuringia. Work in the lab of Rachel W. Martin was supported by NIH (2R01EY021514) and NSF (DMR-2002837). Brenna Norton-Baker was supported by the NSF GRFP. Marquise G. Crosby was supported by NIH (R25 GM055246 MBRS IMSD) and Marc A. Sprague-Piercy was

supported by the HHMI Gilliam Fellowship. Work in the labs of Kristaps Jaudzems and Kaspars Tārs was supported by Latvian Council of Science grant number VPPCOVID2020/1-0014. Work in the UPAT's lab was supported by the INSPIRED (MIS 5002550) project, which is implemented under the Action 'Reinforcement of the Research and Innovation Infrastructure,' funded by the Operational Program 'Competitiveness, Entrepreneurship and Innovation' (NSRF 2014–2020) and co-financed by Greece and the EU (European Regional Development Fund) and the FP7 REGPOT CT-2011-285950 – "SEE-DRUG" project (purchase of UPAT's 700 MHz NMR equipment). Work in the Muhle-Goll lab was supported by the Helmholtz society. Work in the lab of A. Böckmann was supported by the CNRS, the French National Research Agency (ANR, NMR-SCoV2- ORF8), the Fondation de la Recherche Médicale (FRM, NMR-SCoV2-ORF8), and the IR-RMN-THC Fr3050 CNRS. Work in the lab of Beat H. Meier was supported by the Swiss National Science foundation (Grant number 200020_188711), the Günthard Stiftung für Physikalische Chemie and the ETH Zurich. Work in the labs of A. Böckmann and Beat H. Meier was supported by a common grant from SNF (grant 31CA30_196256). This work was supported by the ETH Zurich, the grant ETH 40 18 1 and the grant Krebsliga KFS 4903 08 2019. Work in the lab of the IBS Grenoble was supported by the Agence Nationale de Recherche (France) RA-COVID SARS2NUCLEOPROTEIN and European Research Council Advanced Grant DynamicAssemblies. Work in the Alfano lab was supported by Patto per il Sud della Regione Siciliana - CheMISt grant (CUP G77B17000110001).

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

References

1. Leao, J. C., Gusmao, T. P. L., Zarzar, A. M., Leao Filho, J. C., Barkokebas Santos de Faria, A., Morais Silva, I. H., Gueiros, L. A. M., Robinson, N. A., Porter, S., and Carvalho, A. A. T. (2020) Coronaviridae-Old friends, new enemy! *Oral Dis*
2. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., and Zhang, Y. Z. (2020) A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269
3. Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., and Stern-Ginossar, N. (2020) The coding capacity of SARS-CoV-2. *Nature*
4. Nelson, C. W., Ardern, Z., Goldberg, T. L., Meng, C., Kuo, C. H., Ludwig, C., Kolokotronis, S. O., and Wei, X. (2020) Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic. *Elife* **9**
5. Pavesi, A. (2020) New insights into the evolutionary features of viral overlapping genes by discriminant analysis. *Virology* **546**, 51-66
6. Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., and Hilgenfeld, R. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* **368**, 409-412
7. Krafcikova, P., Silhan, J., Nencka, R., and Boura, E. (2020) Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat Commun* **11**, 3717
8. Yin, W., Mao, C., Luan, X., Shen, D. D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M., Chang, S., Xie, Y. C., Tian, G., Jiang, H. W., Tao, S. C., Shen, J., Jiang, Y., Jiang, H., Xu, Y., Zhang, S., Zhang, Y., and Xu, H. E. (2020) Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499-1504
9. Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L., Ge, J., Zheng, L., Zhang, Y., Wang, H., Zhu, Y., Zhu, C., Hu, T., Hua, T., Zhang, B., Yang, X., Li, J., Yang, H., Liu, Z., Xu, W., Guddat, L. W., Wang, Q., Lou, Z., and Rao, Z. (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779-782
10. 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
11. Serrano, P., Johnson, M. A., Chatterjee, A., Neuman, B. W., Joseph, J. S., Buchmeier, M. J., Kuhn, P., and Wuthrich, K. (2009) Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. *J Virol* **83**, 12998-13008
12. Tonelli, M., Rienstra, C., Anderson, T. K., Kirchdoerfer, R., and Henzler-Wildman, K. (2020) (1)H, (13)C, and (15)N backbone and side chain chemical shift assignments of the SARS-CoV-2 non-structural protein 7. *Biomol NMR Assign*
13. Cantini, F., Banci, L., Altincekic, N., Bains, J. K., Dhamotharan, K., Fuks, C., Furtig, B., Gande, S. L., Hargittay, B., Hengesbach, M., Hutchison, M. T., Korn, S. M., Kubatova, N., Kutz, F., Linhard, V., Lohr, F., Meiser, N., Pyper, D. J., Qureshi, N. S., Richter, C., Saxena, K., Schlundt, A., Schwalbe, H., Sreeramulu, S., Tants, J. N., Wacker, A., Weigand, J. E., Wohnert, J., Tsika, A. C., Fourkiotis, N. K., and Spyroulias, G. A. (2020) (1)H, (13)C, and (15)N backbone chemical shift assignments of the apo and the ADP-ribose bound forms of the macrodomain of SARS-CoV-2 non-structural protein 3b. *Biomol NMR Assign* **14**, 339-346
14. Korn, S. M., Dhamotharan, K., Furtig, B., Hengesbach, M., Lohr, F., Qureshi, N. S., Richter, C., Saxena, K., Schwalbe, H., Tants, J. N., Weigand, J. E., Wohnert, J., and Schlundt, A. (2020) (1)H, (13)C, and (15)N backbone chemical shift assignments of the nucleic acid-binding domain of SARS-CoV-2 non-structural protein 3e. *Biomol NMR Assign* **14**, 329-333

15. Kubatova, N., Qureshi, N. S., Altincekic, N., Abele, R., Bains, J. K., Ceylan, B., Ferner, J., Fuks, C., Hargittay, B., Hutchison, M. T., de Jesus, V., Kutz, F., Wirtz Martin, M. A., Meiser, N., Linhard, V., Pyper, D. J., Trucks, S., Furtig, B., Hengesbach, M., Lohr, F., Richter, C., Saxena, K., Schlundt, A., Schwalbe, H., Sreeramulu, S., Wacker, A., Weigand, J. E., Wirmer-Bartoschek, J., and Wohnert, J. (2020) (1)H, (13)C, and (15)N backbone chemical shift assignments of coronavirus-2 non-structural protein Nsp10. *Biomol NMR Assign*
16. Korn, S. M., Lambertz, R., Furtig, B., Hengesbach, M., Lohr, F., Richter, C., Schwalbe, H., Weigand, J. E., Wohnert, J., and Schlundt, A. (2020) (1)H, (13)C, and (15)N backbone chemical shift assignments of the C-terminal dimerization domain of SARS-CoV-2 nucleocapsid protein. *Biomol NMR Assign*
17. Gallo, A., Tsika, A. C., Fourkiotis, N. K., Cantini, F., Banci, L., Sreeramulu, S., Schwalbe, H., and Spyroulias, G. A. (2020) (1)H,(13)C and (15)N chemical shift assignments of the SUD domains of SARS-CoV-2 non-structural protein 3c: "the N-terminal domain-SUD-N". *Biomol NMR Assign*
18. Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., and Hong, M. (2020) Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat Struct Mol Biol*
19. Esposito, D., Mehalko, J., Drew, M., Snead, K., Wall, V., Taylor, T., Frank, P., Denson, J. P., Hong, M., Gulten, G., Sadtler, K., Messing, S., and Gillette, W. (2020) Optimizing high-yield production of SARS-CoV-2 soluble spike trimers for serology assays. *Protein Expr Purif* **174**, 105686
20. Jiang, H. W., Li, Y., Zhang, H. N., Wang, W., Yang, X., Qi, H., Li, H., Men, D., Zhou, J., and Tao, S. C. (2020) SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. *Nat Commun* **11**, 3581
21. Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Huttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., Haas, P., Polacco, B. J., Braberg, H., Fabius, J. M., Eckhardt, M., Soucheray, M., Bennett, M. J., Cakir, M., McGregor, M. J., Li, Q., Meyer, B., Roesch, F., Vallet, T., Mac Kain, A., Miorin, L., Moreno, E., Naing, Z. Z. C., Zhou, Y., Peng, S., Shi, Y., Zhang, Z., Shen, W., Kirby, I. T., Melnyk, J. E., Chorba, J. S., Lou, K., Dai, S. A., Barrio-Hernandez, I., Memon, D., Hernandez-Armenta, C., Lyu, J., Mathy, C. J. P., Perica, T., Pilla, K. B., Ganesan, S. J., Saltzberg, D. J., Rakesh, R., Liu, X., Rosenthal, S. B., Calviello, L., Venkataramanan, S., Liboy-Lugo, J., Lin, Y., Huang, X. P., Liu, Y., Wankowicz, S. A., Bohn, M., Safari, M., Ugur, F. S., Koh, C., Savar, N. S., Tran, Q. D., Shengjuler, D., Fletcher, S. J., O'Neal, M. C., Cai, Y., Chang, J. C. J., Broadhurst, D. J., Klippsten, S., Sharp, P. P., Wenzell, N. A., Kuzuoglu-Ozturk, D., Wang, H. Y., Trenker, R., Young, J. M., Cavero, D. A., Hiatt, J., Roth, T. L., Rathore, U., Subramanian, A., Noack, J., Hubert, M., Stroud, R. M., Frankel, A. D., Rosenberg, O. S., Verba, K. A., Agard, D. A., Ott, M., Emerman, M., Jura, N., von Zastrow, M., Verdin, E., Ashworth, A., Schwartz, O., d'Enfert, C., Mukherjee, S., Jacobson, M., Malik, H. S., Fujimori, D. G., Ideker, T., Craik, C. S., Floor, S. N., Fraser, J. S., Gross, J. D., Sali, A., Roth, B. L., Ruggero, D., Taunton, J., Kortemme, T., Beltrao, P., Vignuzzi, M., Garcia-Sastre, A., Shokat, K. M., Shoichet, B. K., and Krogan, N. J. (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459-468
22. Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., Cinatl, J., and Munch, C. (2020) Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469-472
23. Wacker, A., Weigand, J. E., Akabayov, S. R., Altincekic, N., Bains, J. K., Banijamali, E., Binas, O., Castillo-Martinez, J., Cetiner, E., Ceylan, B., Chiu, L. Y., Davila-Calderon, J., Dhamotharan, K., Duchardt-Ferner, E., Ferner, J., Frydman, L., Furtig, B., Gallego, J., Grun, J. T., Hacker, C., Haddad, C., Hahnke, M., Hengesbach, M., Hiller, F., Hohmann, K. F., Hymon, D., de Jesus, V., Jonker, H., Keller, H., Knezic, B., Landgraf, T., Lohr, F., Luo, L., Mertinkus, K. R., Muhs, C., Novakovic, M., Oxenfarth, A., Palomino-Schatzlein, M., Petzold, K., Peter, S. A., Pyper, D. J., Qureshi, N. S., Riad, M., Richter, C., Saxena, K., Schamber, T., Scherf, T., Schlagnitweit, J., Schlundt, A., Schnieders, R., Schwalbe, H., Simba-Lahuasi, A., Sreeramulu, S., Stirnal, E., Sudakov, A., Tants, J. N., Tolbert, B. S., Vogele, J., Weiss, L., Wirmer-Bartoschek, J., Wirtz

Martin, M. A., Wohnert, J., and Zetzsche, H. (2020) Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res*

24. Fogeron, M. L., Paul, D., Jirasko, V., Montserret, R., Lacabanne, D., Molle, J., Badillo, A., Boukadida, C., Georgeault, S., Roingeard, P., Martin, A., Bartenschlager, R., Penin, F., and Bockmann, A. (2015) Functional expression, purification, characterization, and membrane reconstitution of non-structural protein 2 from hepatitis C virus. *Protein Expr Purif* **116**, 1-6
25. Fogeron, M. L., Badillo, A., Penin, F., and Bockmann, A. (2017) Wheat Germ Cell-Free Overexpression for the Production of Membrane Proteins. *Methods Mol Biol* **1635**, 91-108
26. Jirasko, V., Lends, A., Lakomek, N. A., Fogeron, M. L., Weber, M., Malar, A., Penzel, S., Bartenschlager, R., Meier, B. H., and Bockmann, A. (2020) Dimer organization of membrane-associated NS5A of hepatitis C virus as determined by highly sensitive 1H-detected solid-state NMR. *Angew Chem Int Ed Engl*
27. Netzer, W. J., and Hartl, F. U. (1997) Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**, 343-349
28. Oostra, M., te Lintelo, E. G., Deijs, M., Verheije, M. H., Rottier, P. J., and de Haan, C. A. (2007) Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. *J Virol* **81**, 12323-12336
29. Oostra, M., Hagemeijer, M. C., van Gent, M., Bekker, C. P., te Lintelo, E. G., Rottier, P. J., and de Haan, C. A. (2008) Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J Virol* **82**, 12392-12405
30. Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature* **584**, 154-156
31. 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
32. Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J., and Gorbalenya, A. E. (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* **331**, 991-1004
33. 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
34. 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
35. 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
36. Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., Kratzat, H., Hayn, M., Mackens-Kiani, T., Cheng, J., Straub, J. H., Sturzel, C. M., Frohlich, T., Berninghausen, O., Becker, T., Kirchhoff, F., Sparrer, K. M. J., and Beckmann, R. (2020) Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **369**, 1249-1255
37. Cornillez-Ty, C. T., Liao, L., Yates, J. R., 3rd, Kuhn, P., and Buchmeier, M. J. (2009) Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *J Virol* **83**, 10314-10318
38. Davies, J. P., Almasy, K. M., McDonald, E. F., and Plate, L. (2020) Comparative multiplexed interactomics of SARS-CoV-2 and homologous coronavirus non-structural proteins identifies unique and shared host-cell dependencies. *bioRxiv*
39. Graham, R. L., Sims, A. C., Brockway, S. M., Baric, R. S., and Denison, M. R. (2005) The nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral replication. *J Virol* **79**, 13399-13411

40. Ishida, T., and Kinoshita, K. (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* **35**, W460-464
41. Mompean, M., Trevino, M. A., and Laurents, D. V. (2020) Towards Targeting the Disordered SARS-CoV-2 Nsp2 C-terminal Region: Partial Structure and Dampened Mobility Revealed by NMR Spectroscopy. *BioRxiv* **preprint**
42. Wolff, G., Limpens, R., Zevenhoven-Dobbe, J. C., Laugks, U., Zheng, S., de Jong, A. W. M., Koning, R. I., Agard, D. A., Grunewald, K., Koster, A. J., Snijder, E. J., and Barcena, M. (2020) A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* **369**, 1395-1398
43. Serrano, P., Johnson, M. A., Almeida, M. S., Horst, R., Herrmann, T., Joseph, J. S., Neuman, B. W., Subramanian, V., Saikatendu, K. S., Buchmeier, M. J., Stevens, R. C., Kuhn, P., and Wuthrich, K. (2007) Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. *J Virol* **81**, 12049-12060
44. Hurst, K. R., Koetzner, C. A., and Masters, P. S. (2013) Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *J Virol* **87**, 9159-9172
45. Khan, M. T., Zeb, M. T., Ahsan, H., Ahmed, A., Ali, A., Akhtar, K., Malik, S. I., Cui, Z., Ali, S., Khan, A. S., Ahmad, M., Wei, D. Q., and Irfan, M. (2020) SARS-CoV-2 nucleocapsid and Nsp3 binding: an in silico study. *Arch Microbiol*
46. Frick, D. N., Virdi, R. S., Vuksanovic, N., Dahal, N., and Silvaggi, N. R. (2020) Molecular Basis for ADP-Ribose Binding to the Mac1 Domain of SARS-CoV-2 nsp3. *Biochemistry* **59**, 2608-2615
47. Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L., and Hilgenfeld, R. (2015) A G-quadruplex-binding macrodomain within the "SARS-unique domain" is essential for the activity of the SARS-coronavirus replication-transcription complex. *Virology* **484**, 313-322
48. Chen, Y., Savinov, S. N., Mielech, A. M., Cao, T., Baker, S. C., and Mesecar, A. D. (2015) X-ray Structural and Functional Studies of the Three Tandemly Linked Domains of Non-structural Protein 3 (nsp3) from Murine Hepatitis Virus Reveal Conserved Functions. *J Biol Chem* **290**, 25293-25306
49. Tan, J., Vonrhein, C., Smart, O. S., Bricogne, G., Bollati, M., Kusov, Y., Hansen, G., Mesters, J. R., Schmidt, C. L., and Hilgenfeld, R. (2009) The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog* **5**, e1000428
50. Johnson, M. A., Chatterjee, A., Neuman, B. W., and Wuthrich, K. (2010) SARS coronavirus unique domain: three-domain molecular architecture in solution and RNA binding. *J Mol Biol* **400**, 724-742
51. Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., Schulz, L., Widera, M., Mehdipour, A. R., Tascher, G., Geurink, P. P., Wilhelm, A., van der Heden van Noort, G. J., Ovaa, H., Muller, S., Knobeloch, K. P., Rajalingam, K., Schulman, B. A., Cinatl, J., Hummer, G., Ciesek, S., and Dikic, I. (2020) Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657-662
52. Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A., Liao, L., Klaus, J. P., Yates, J. R., 3rd, Wuthrich, K., Stevens, R. C., Buchmeier, M. J., and Kuhn, P. (2008) Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J Virol* **82**, 5279-5294
53. Lei, J., Kusov, Y., and Hilgenfeld, R. (2018) Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res* **149**, 58-74
54. Neuman, B. W. (2016) Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles. *Antiviral Res* **135**, 97-107
55. Hagemeijer, M. C., Monastyrska, I., Griffith, J., van der Sluijs, P., Voortman, J., van Bergen en Henegouwen, P. M., Vonk, A. M., Rottier, P. J., Reggiori, F., and de Haan, C. A. (2014) Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology* **458-459**, 125-135
56. Ullrich, S., and Nitsche, C. (2020) The SARS-CoV-2 main protease as drug target. *Bioorg Med Chem Lett* **30**, 127377

57. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., and Hilgenfeld, R. (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763-1767
58. Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Huettenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., Haas, P., Polacco, B. J., Braberg, H., Fabius, J. M., Eckhardt, M., Soucheray, M., Bennett, M. J., Cakir, M., McGregor, M. J., Li, Q., Meyer, B., Roesch, F., Vallet, T., Mac Kain, A., Miorin, L., Moreno, E., Naing, Z. Z. C., Zhou, Y., Peng, S., Shi, Y., Zhang, Z., Shen, W., Kirby, I. T., Melnyk, J. E., Chorba, J. S., Lou, K., Dai, S. A., Barrio-Hernandez, I., Memon, D., Hernandez-Armenta, C., Lyu, J., Mathy, C. J. P., Perica, T., Pilla, K. B., Ganesan, S. J., Saltzberg, D. J., Rakesh, R., Liu, X., Rosenthal, S. B., Calviello, L., Venkataramanan, S., Liboy-Lugo, J., Lin, Y., Huang, X. P., Liu, Y., Wankowicz, S. A., Bohn, M., Safari, M., Ugur, F. S., Koh, C., Savar, N. S., Tran, Q. D., Shengjuler, D., Fletcher, S. J., O'Neal, M. C., Cai, Y., Chang, J. C. J., Broadhurst, D. J., Klippsten, S., Sharp, P. P., Wenzell, N. A., Kuzuoglu, D., Wang, H. Y., Trenker, R., Young, J. M., Cavero, D. A., Hiatt, J., Roth, T. L., Rathore, U., Subramanian, A., Noack, J., Hubert, M., Stroud, R. M., Frankel, A. D., Rosenberg, O. S., Verba, K. A., Agard, D. A., Ott, M., Emerman, M., Jura, N., von Zastrow, M., Verdin, E., Ashworth, A., Schwartz, O., d'Enfert, C., Mukherjee, S., Jacobson, M., Malik, H. S., Fujimori, D. G., Ideker, T., Craik, C. S., Floor, S. N., Fraser, J. S., Gross, J. D., Sali, A., Roth, B. L., Ruggero, D., Taunton, J., Kortemme, T., Beltrao, P., Vignuzzi, M., Garcia-Sastre, A., Shokat, K. M., Shoichet, B. K., and Krogan, N. J. (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*
59. Tvarogova, J., Madhugiri, R., Bylapudi, G., Ferguson, L. J., Karl, N., and Ziebuhr, J. (2019) Identification and Characterization of a Human Coronavirus 229E Nonstructural Protein 8-Associated RNA 3'-Terminal Adenylyltransferase Activity. *J Virol* **93**
60. Konkolova, E., Klima, M., Nencka, R., and Boura, E. (2020) Structural analysis of the putative SARS-CoV-2 primase complex. *J Struct Biol* **211**, 107548
61. Kirchdoerfer, R. N., and Ward, A. B. (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* **10**, 2342
62. Miknis, Z. J., Donaldson, E. F., Umland, T. C., Rimmer, R. A., Baric, R. S., and Schultz, L. W. (2009) Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *J Virol* **83**, 3007-3018
63. Robertson, M. P., Igel, H., Baertsch, R., Haussler, D., Ares, M., Jr., and Scott, W. G. (2005) The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol* **3**, e5
64. Egloff, M. P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E. J., Gorbalenya, A. E., Cambillau, C., and Canard, B. (2004) The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc Natl Acad Sci U S A* **101**, 3792-3796
65. Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., Siddell, S., Poon, L. L., Diprose, J., Alderton, D., Walsh, M., Grimes, J. M., and Stuart, D. I. (2004) The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure* **12**, 341-353
66. 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
67. Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., and Rao, Z. (2015) Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc Natl Acad Sci U S A* **112**, 9436-9441
68. Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Brooun, A., Griffith, M., Moy, K., Yadav, M. K., Velasquez, J., Buchmeier, M. J., Stevens, R. C., and Kuhn, P. (2006) Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *J Virol* **80**, 7894-7901
69. Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P. M. M., Olinares, P. D. B., Maruthi, K., Eng, E. T., Vatandaslar, H., Chait, B. T., Kapoor, T. M., Darst, S. A., and Campbell, E. A. (2020)

Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **182**, 1560-1573 e1513
70. Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., Zheng, L., Ming, Z., Zhang, L., Lou, Z., and Rao, Z. (2019) Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res* **47**, 6538-6550
71. Robson, F., Khan, K. S., Le, T. K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P., and Ng, W. L. (2020) Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Mol Cell* **79**, 710-727
72. Chen, P., Jiang, M., Hu, T., Liu, Q., Chen, X. S., and Guo, D. (2007) Biochemical characterization of exoribonuclease encoded by SARS coronavirus. *J Biochem Mol Biol* **40**, 649-655
73. Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., Michalska, K., and Joachimiak, A. (2020) Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci* **29**, 1596-1605
74. Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., Kao, C. C., and Baker, S. C. (2017) Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc Natl Acad Sci U S A* **114**, E4251-E4260
75. Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E. J., Canard, B., and Decroly, E. (2010) In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog* **6**, e1000863
76. Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N. L., Kiryukhina, O., Wiersum, G., Kim, Y., Jedrzejczak, R., Maltseva, N. I., Endres, M., Jaroszewski, L., Godzik, A., Joachimiak, A., and Satchell, K. J. F. (2020) The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv*
77. Kern, D. M., Sorum, B., Hoel, C. M., Sridharan, S., Remis, J. P., Toso, D. B., and Brohawn, S. G. (2020) Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *bioRxiv*
78. Liu, D. X., Fung, T. S., Chong, K. K., Shukla, A., and Hilgenfeld, R. (2014) Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral Res* **109**, 97-109
79. 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
80. Konno, Y., Kimura, I., Uriu, K., Fukushi, M., Irie, T., Koyanagi, Y., Sauter, D., Gifford, R. J., Consortium, U.-C., Nakagawa, S., and Sato, K. (2020) SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is Increased by a Naturally Occurring Elongation Variant. *Cell Rep* **32**, 108185
81. Schoeman, D., and Fielding, B. C. (2020) Is There a Link Between the Pathogenic Human Coronavirus Envelope Protein and Immunopathology? A Review of the Literature. *Front Microbiol* **11**, 2086
82. Surya, W., Li, Y., and Torres, J. (2018) Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim Biophys Acta Biomembr* **1860**, 1309-1317
83. Huang, Y., Yang, Z. Y., Kong, W. P., and Nabel, G. J. (2004) Generation of synthetic severe acute respiratory syndrome coronavirus pseudoparticles: implications for assembly and vaccine production. *J Virol* **78**, 12557-12565
84. Vennema, H., Godeke, G. J., Rossen, J. W., Voorhout, W. F., Horzinek, M. C., Opstelten, D. J., and Rottier, P. J. (1996) Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J* **15**, 2020-2028
85. Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., Droese, B., Klaus, J. P., Makino, S., Sawicki, S. G., Siddell, S. G., Stamou, D. G., Wilson, I. A., Kuhn, P., and Buchmeier, M. J. (2011) A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol* **174**, 11-22
86. Zhao, J., Falcon, A., Zhou, H., Netland, J., Enjuanes, L., Perez Brena, P., and Perlman, S. (2009) Severe acute respiratory syndrome coronavirus protein 6 is required for optimal replication. *J Virol* **83**, 2368-2373
87. Bogomolovas, J., Simon, B., Sattler, M., and Stier, G. (2009) Screening of fusion partners for high yield expression and purification of bioactive viscotoxins. *Protein Expr Purif* **64**, 16-23

88. Wu, Z., Yang, L., Ren, X., Zhang, J., Yang, F., Zhang, S., and Jin, Q. (2016) ORF8-Related Genetic Evidence for Chinese Horseshoe Bats as the Source of Human Severe Acute Respiratory Syndrome Coronavirus. *J Infect Dis* **213**, 579-583
89. Tan, Y., Schneider, T., Leong, M., Aravind, L., and Zhang, D. (2020) Novel Immunoglobulin Domain Proteins Provide Insights into Evolution and Pathogenesis of SARS-CoV-2-Related Viruses. *mBio* **11**
90. Flower, T. G., Buffalo, C. Z., Hooy, R. M., Allaire, M., Ren, X., and Hurley, J. H. (2020) Structure of SARS-CoV-2 ORF8, a rapidly evolving coronavirus protein implicated in immune evasion. *BioRxiv* **preprint**
91. Luo, H., Chen, J., Chen, K., Shen, X., and Jiang, H. (2006) Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry* **45**, 11827-11835
92. Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., and Huang, T. H. (2014) The SARS coronavirus nucleocapsid protein--forms and functions. *Antiviral Res* **103**, 39-50
93. Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., Zhou, Z., Chen, Q., Yan, Y., Zhang, C., Shan, H., and Chen, S. (2020) Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B* **10**, 1228-1238
94. Chen, H., Cui, Y., Han, X., Hu, W., Sun, M., Zhang, Y., Wang, P. H., Song, G., Chen, W., and Lou, J. (2020) Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res*
95. Dinesh, D. C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., and Boura, E. (2020) Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog* **16**, e1009100
96. Ye, Q., West, A. M. V., Silletti, S., and Corbett, K. D. (2020) Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci*
97. Zhou, R., Zeng, R., Von Brunn, A., and Lei, J. (2020) Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. *Mol. Biomed.* **1**, 1-11
98. Takeda, M., Chang, C. K., Ikeya, T., Guntert, P., Chang, Y. H., Hsu, Y. L., Huang, T. H., and Kainosho, M. (2008) Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J Mol Biol* **380**, 608-622
99. Yoshimoto, F. K. (2020) The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *Protein J* **39**, 198-216
100. Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., and Yang, H. (2020) Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289-293
101. Li, Q., and Kang, C. (2020) A Practical Perspective on the Roles of Solution NMR Spectroscopy in Drug Discovery. *Molecules* **25**
102. 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
103. 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
104. Cross, T. J., Takahashi, G. R., Diessner, E. M., Crosby, M. G., Farahmand, V., Zhuang, S., Butts, C. T., and Martin, R. W. (2020) Sequence Characterization and Molecular Modeling of Clinically Relevant Variants of the SARS-CoV-2 Main Protease. *Biochemistry* **59**, 3741-3756
105. Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P., Shr, H. L., Chang, G. G., Wang, A. H., and Liang, P. H. (2005) Mechanism of the maturation process of SARS-CoV 3CL protease. *J Biol Chem* **280**, 31257-31266

106. Chang, H. P., Chou, C. Y., and Chang, G. G. (2007) Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride. *Biophys J* **92**, 1374-1383
107. Morita, E. H., Sawasaki, T., Tanaka, R., Endo, Y., and Kohno, T. (2003) A wheat germ cell-free system is a novel way to screen protein folding and function. *Protein Sci* **12**, 1216-1221
108. Fogeron, M. L., Badillo, A., Jirasko, V., Gouttenoire, J., Paul, D., Lancien, L., Moradpour, D., Bartenschlager, R., Meier, B. H., Penin, F., and Bockmann, A. (2015) Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins. *Protein Expr Purif* **105**, 39-46
109. Fogeron, M. L., Jirasko, V., Penzel, S., Paul, D., Montserret, R., Danis, C., Lacabanne, D., Badillo, A., Gouttenoire, J., Moradpour, D., Bartenschlager, R., Penin, F., Meier, B. H., and Bockmann, A. (2016) Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus. *J Biomol NMR* **65**, 87-98
110. Jirasko, V., Lakomek, N. A., Penzel, S., Fogeron, M. L., Bartenschlager, R., Meier, B. H., and Bockmann, A. (2020) Proton-Detected Solid-State NMR of the Cell-Free Synthesized alpha-Helical Transmembrane Protein NS4B from Hepatitis C Virus. *Chembiochem* **21**, 1453-1460
111. David, G., Fogeron, M. L., Schledorn, M., Montserret, R., Haselmann, U., Penzel, S., Badillo, A., Lecoq, L., Andre, P., Nassal, M., Bartenschlager, R., Meier, B. H., and Bockmann, A. (2018) Structural Studies of Self-Assembled Subviral Particles: Combining Cell-Free Expression with 110 kHz MAS NMR Spectroscopy. *Angew Chem Int Ed Engl* **57**, 4787-4791
112. Wang, S., Fogeron, M. L., Schledorn, M., Dujardin, M., Penzel, S., Burdette, D., Berke, J. M., Nassal, M., Lecoq, L., Meier, B. H., and Bockmann, A. (2019) Combining Cell-Free Protein Synthesis and NMR Into a Tool to Study Capsid Assembly Modulation. *Front Mol Biosci* **6**, 67
113. Tonelli, M., Singarapu, K. K., Makino, S., Sahu, S. C., Matsubara, Y., Endo, Y., Kainosho, M., and Markley, J. L. (2011) Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J Biomol NMR* **51**, 467-476
114. Takai, K., Sawasaki, T., and Endo, Y. (2010) Practical cell-free protein synthesis system using purified wheat embryos. *Nat Protoc* **5**, 227-238

Abbreviations

The abbreviations used are: aa, amino acid; BEST, Band-selective Excitation Short-Transient; BMRB, Biomagnetic Resonance Databank; CFPS, cell-free protein synthesis; CoV, coronavirus; CTD, c-terminal domain; DEDD, Asp-Glu-Glu-Asp; DMS, dimethylsulfate; E, Envelope protein; ED, ectodomain; fl, full-length; GB1, protein G B1 domain; GD, globular domain; GF, gel-filtration; GST, glutathione-S-transferase; His, His_x-tag; HSP, heat shock protein; HSQC, heteronuclear single quantum coherence; IDP, intrinsically disordered protein; IDR, intrinsically disordered region; IEC, ion exchange chromatography; IMAC, immobilized metal ion affinity chromatography; IPRS, IMAC-protease cleavage-reverse IMAC-SEC; M, Membrane protein; MERS, middle east respiratory syndrome; MHV, murine hepatitis virus; Mpro, main protease; MTase, methyltransferase; N, Nucleocapsid protein; NAB, nucleic acid-binding domain; nsp, nonstructural protein; NTD, N-terminal domain; PLpro, papain-like protease; RdRP, RNA-dependent RNA polymerase; S, Spike protein; SARS, severe acute respiratory syndrome; SEC, size exclusion chromatography; SUD, SARS unique domain; SUMO, small ubiquitin-related modifier; TEV, tobacco etch virus; TM, transmembrane; TROSY, transverse relaxation-optimized spectroscopy; Trx, thioredoxin; Ubl, ubiquitin-like domain; Ulp1, ubiquitin-like specific protease 1; WG, wheat germ

Table 1: SCoV2 protein constructs expressed and purified, given with the genomic position and corresponding PDBs for construct design.

Protein *genome position (nt)* [a]	***Trivial name*** Construct expressed	**Size (aa)**	**Boundaries**	**MW [kDa]**	**Homol. SCoV (%)** [b]	**Template PDB** [c]	**SCoV2 PDB** [d]
nsp1 *266-805*	***Leader***	**180**		**19.8**	**84**		
	Full-length	180	1-180	19.8	83		
	Globular Domain (GD)	116	13-127	12.7	85	2GDT	7K7P
nsp2 *806-2,719*		**638**		**70.5**	**68**		
	C-terminal IDR (CtDR)	45	557-601	4.9	55		
nsp3 *2,720-8,554*		**1,945**		**217.3**	**76**		
a	Ub-like (UBl) domain	111	1-111	12.4	79	2IDY	7KAG
a	Ub-like (UBl) domain + IDR	206	1-206	23.2	58		
b	Macrodomain	170	207-376	18.3	74	6VXS	6VXS
c	SUD-N	140	409-548	15.4	69	2W2G	
c	SUD-NM	267	409-675	29.5	74	2W2G	
c	SUD-M	125	551-675	13.9	82	2W2G	
c	SUD-MC	195	551-745	21.5	79	2KQV	
c	SUD-C	64	680-743	7.3	73	2KAF	
d	Papain-like protease PL^{pro}	318	743-1,060	36	83	6W9C	6W9C
e	NAB	116	1,088-1,203	13.4	87	2K87	
Y	CoV-Y	308	1,638-1,945	34	89		
nsp5 *10,055-10,972*	***Main protease (M^{pro})***	**306**		**33.7**	**96**		
	Full-length [e]	306	1-306	33.7	96	6Y84	6Y84

nsp7 *11,843-12,091*		**83**		**9.2**	**99**		
	Full-length	83	1-83	9.2	99	6WIQ	6WIQ
nsp8 *12,092-12,685*		**198**		**21.9**	**98**		
	Full-length	198	1-198	21.9	97	6WIQ	6WIQ
nsp9 *12,686-13,024*		**113**		**12.4**	**97**		
	Full-length	113	1-113	12.4	97	6W4B	6W4B
nsp10 *13,025-13,441*		**139**		**14.8**	**97**		
	Full-length	139	1-139	14.8	97	6W4H	6W4H
nsp13 *16,237-18,039*	***Helicase***	**601**		**66.9**	**100**		
	Full-length	601	1-601	66.9	100	6ZSL	6ZSL
nsp14 *18,040-19,620*	***Exonuclease/Methyl-transferase***	**527**		**59.8**	**95**		
	Full-length	527	1-527	59.8	95	5NFY	
	MTase domain	240	288-527	27.5	95		
nsp15 *19,621-20,658*	***Endonuclease***	**346**		**38.8**	**89**		
	Full-length	346	1-346	38.8	89	6W01	6W01
nsp16 *20,659-21,552*	***Methyltransferase***	**298**		**33.3**	**93**		
	Full-length	298	1-298	33.3	93	6W4H	6W4H
ORF3a *25,393-26,220*		**275**		**31.3**	**72**		
	Full-length	275	1-275	31.3	72	6XDC	6XDC
ORF4 *26,245-26,472*	***Envelope (E) protein***	**75**		**8.4**	**95**		
	Full-length	75	1-75	8.4	95	5X29	7K3G
ORF5 *26,523-27,387*	***Membrane glycoprotein (M)***	**222**		**25.1**	**91**		
	Full-length	222	1-222	25.1	91		
ORF6 *27,202-27,387*		**61**		**7.3**	**69**		

	Full-length	61	1-61	7.3	69		
ORF7a *27,394-27,759*		**121**		**13.7**	**85**		
	Ectodomain (ED)	66	16-81	7.4	85	1XAK	6W37
ORF7b *27,756-27,887*		**43**		**5.2**	**85**		
	Full-length	43	1-43	5.2	85		
ORF8 *27,894-28,259*		**121**		**13.8**	**32**		
ORF8	Full-length	121	1-121	13.8	32	7JTL	7JTL
ΔORF8	w/o signal	106	16-121	12	41	7JTL	7JTL
ORF9a *28,274-29,533*	***Nucleocapsid (N)***	**419**		**45.6**	**91**		
	IDR1-NTD- IDR2	248	1-248	26.5	90		
	NTD-SR	169	44-212	18.1	92		
	NTD	136	44-180	14.9	93	6YI3	6YI3
	CTD	118	247-364	13.3	96	2JW8	7C22
ORF9b *28,284-28,574*		**97**		**10.8**	**72**		
	Full-length	97	1-97	10.8	72	6Z4U	6Z4U
ORF14 *28,734-28,952*		**73**		**8**	**n.a.**		
	Full-length	73	1-73	8	n.a.		
ORF10 *29,558-29,674*		**38**		**4.4**	**29**		
	Full-length	38	1-38	4.4	29		

[a] Genome position in nt corresponding to SCoV2 NCBI reference genome entry NC_045512.2, identical to GenBank entry MN908947.3 (2).
[b] Sequence identities to SCoV are calculated from an alignment with corresponding protein sequences based on genome sequence of NCBI Reference NC_004718.3.
[c] Representative PDB that was available at the beginning of construct design, either SCoV or SCoV2.
[d] Representative PDB available for SCoV2 (as of December 2020).
[e] additional point mutations in fl-construct have been expressed.
n.a. not applicable

Table 2: Summary of SCoV2 protein production results in *Covid19-NMR*.

Construct expressed	Yields [mg/L][a] or *[mg/mL]* [b]	Results	Comments	BMRB	SI
nsp1					SI1
fl	5	NMR assigned	expression only at > 20°C; after 7 days at 25°C partial proteolysis	50620[c]	
GD	>0.5	HSQC	high expression; mainly insoluble; higher salt increases stability (>250mM)		
nsp2					SI2
CtDR	0.7-1.5	NMR assigned	assignment with His-tag shown in (41)	to come[c]	
nsp3					SI3
UBl	0.7	HSQC	highly stable over weeks; spectrum overlays with UBl+IDR		
UBl+IDR	2-3	NMR assigned	highly stable for > 2 weeks at 25°C	50446[c]	
Macrodomain	9	NMR assigned	highly stable for > 1 week at 25°C and > 2weeks at 4°C;	50387[d] 50388[d]	
SUD-N	14	NMR assigned	highly stable for > 10 days at 25°C;	50448[d]	
SUD-NM	17	HSQC	stable for > 1 week at 25°C		
SUD-M	8.5	NMR assigned	significant precipitation during measure-ment; tendency to dimerize	50516[d]	
SUD-MC	12	HSQC	stable for > 1 week at 25°C		
SUD-C	4.7	NMR assigned	stable for > 10 days at 25°C	50517[d]	
PL^{pro}	12	HSQC	Solubility-tag essential for expression; tendency to aggregate		
NAB	3.5	NMR assigned	highly stable for > 1 week at 25°C; stable for > 5 weeks at 4°C	50334[d]	
CoV-Y	12	HSQC	low temperature (< 25°C) and low concen-trations (> 0.2 M) favor stability; gradual degradation at 25°C; lithium bromide in final buffer supports solubility		
nsp5					SI4
fl	55	HSQC	impaired dimerization induced by artificial N-terminal residues		
nsp7					SI5
fl	17	NMR assigned	stable for several days at 35°C; stable for > 1 month at 4°C	50337[d]	

nsp8					**SI6**
fl	17	HSQC	concentration dependent aggregation; low concentrations favor stability		
nsp9					**SI7**
fl	4.5	NMR assigned	stable dimer for > 4 month at 4°C and > 2 weeks at 25°C;	50621[d] 50622[d]	
nsp10					**SI8**
fl	15	NMR assigned	Zn^{2+} addition during expression and purification increases protein stability; stable for > 1 week at 25°C.	50392	
nsp13					**SI9**
fl	0.5	HSQC	low expression; protein unstable; concentration above 20µM not possible		
nsp14					**SI10**
fl	6	pure protein	not above 50µM; best storage: with 50% (v/v) glycerol; addition of reducing agents		
MTase	10	pure protein	as fl nsp14; high salt (> 0.4M) for increased stability; addition of reducing agents		
nsp15					**SI11**
fl	5	HSQC	tendency to aggregate at 25°C		
nsp16					**SI12**
fl	10	pure protein	addition of reducing agents; 5% (v/v) glycerol favorable; highly unstable		
ORF3a					**SI13**
fl	*0.6*	pure protein	addition of detergent during expression (0.05% Brij-58); stable protein		
E protein					**SI14**
fl	*0.45*	pure protein	addition of detergent during expression (0.05% Brij-58); stable protein		
M protein					**SI15**
fl	*0.33*	pure protein	addition of detergent during expression (0.05% Brij-58); stable protein		
ORF6					**SI16**
fl	*0.27*	HSQC	soluble expression without detergent; stable protein; no expression with STREP-tag at N-terminus		

ORF7a					**SI17**
ED	0.4	HSQC	unpurified protein tends to precipitate during refolding, purified protein stable for 4 days at 25°C		
ORF7b					**SI18**
fl	0.6	HSQC	tendency to oligomerize; solubilizing agents needed		
fl	*0.27*	HSQC	addition of detergent during expression (0.1 % MNG-3); stable protein		
ORF8					**SI19**
fl	*0.62*	HSQC	tendency to oligomerize		
ΔORF8	0.5	pure protein			
N protein					**SI20**
IDR1-NTD-IDR2	12	NMR assigned	high salt (> 0.4M) for increased stability	50618, 50619[c]	
NTD-SR	3	HSQC			
NTD	3	HSQC		34511	
CTD	2	NMR assigned	stable dimer for > 4 month at 4°C and > 3 weeks at 30°C	50518[d]	
ORF9b					**SI21**
fl	*0.64*	HSQC	Expression without detergent, protein is stable		
ORF14					**SI22**
fl	*0.43*	HSQC	addition of detergent during expression (0.05% Brij-58); stable in detergent but unstable on lipid reconstitution		
ORF10					**SI23**
fl	2	HSQC	tendency to oligomerize; unstable upon Tag-cleavage		

[a] Yields from bacterial expression represent the minimal protein amount in mg/L independent of the cultivation medium.
[b] Yields from CFPS represent the minimal protein amount in mg/mL.
[c] *covid19-nmr* BMRB depositions yet to be released.
[d] *covid19-nmr* BMRB depositions.

Figures

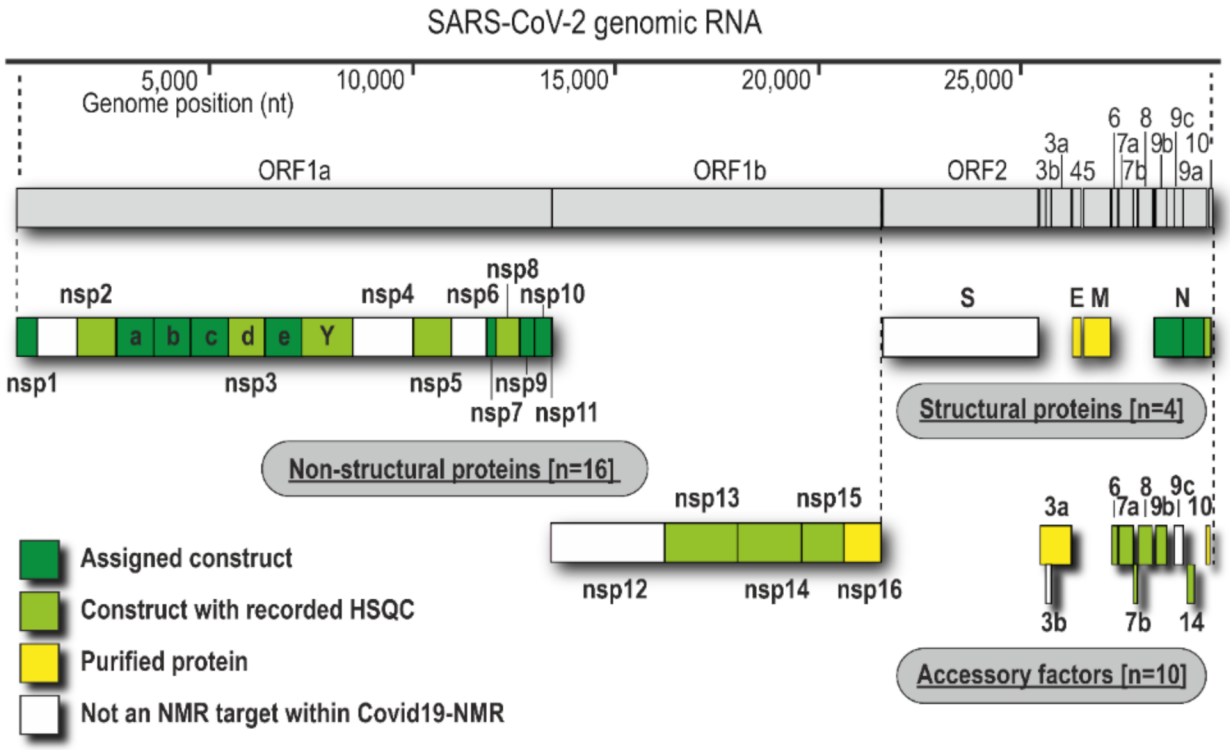

Figure 1: Genomic organization of proteins and current state of analysis or purification. Boxes represent the domain boundaries as outlined in the text and in Table 1. Their position corresponds with the genomic loci. Colors indicate whether the pure proteins were purified (yellow), analyzed by NMR using only HSQC (lime), or characterized in detail including NMR resonance assignments (green).

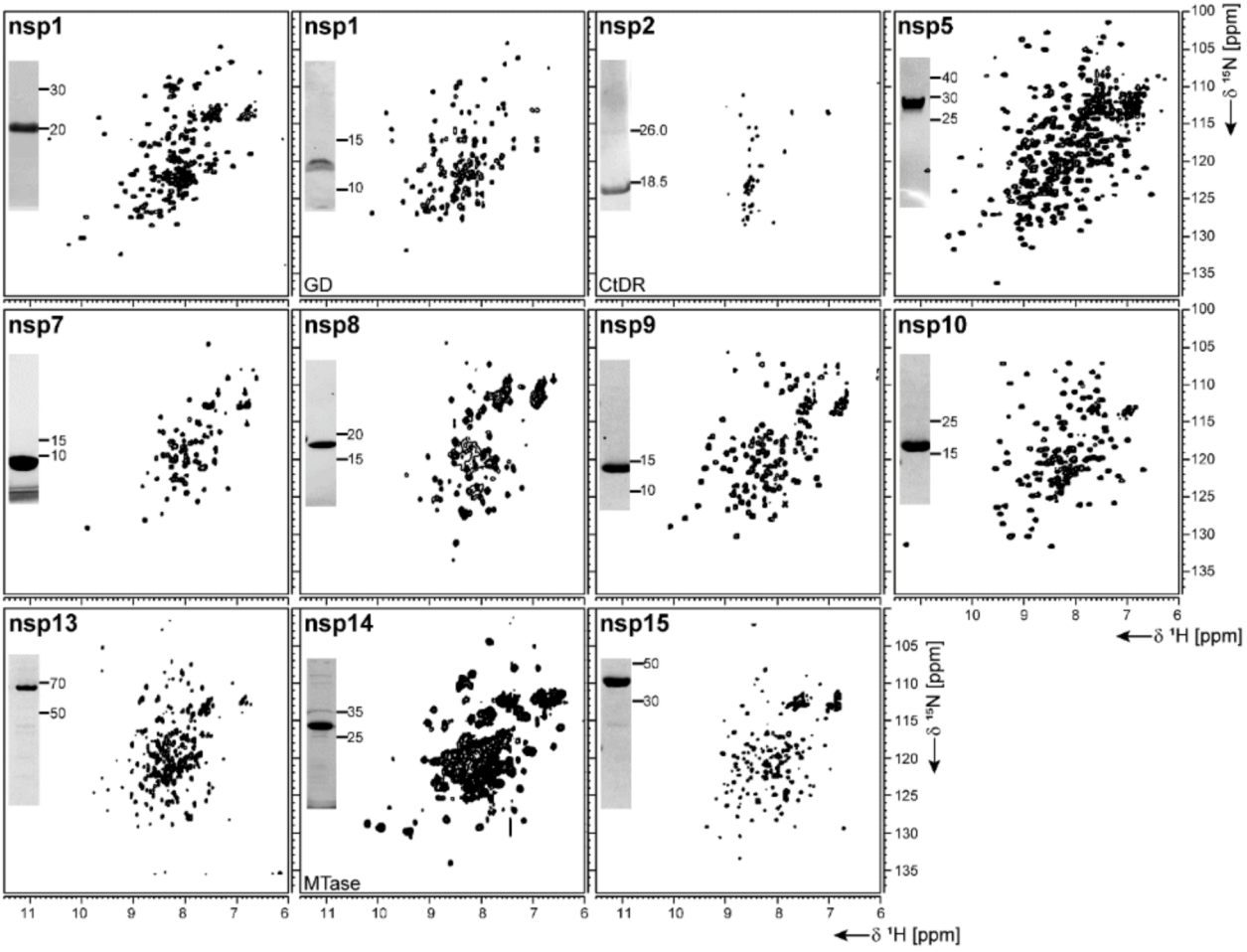

Figure 2. ^{1}H, ^{15}N-correlation spectra of investigated non-structural proteins. Construct names according to Table 1 are indicated unless fl-proteins are shown. A representative SDS-PAGE lane with final samples is included as inset. Spectra for nsp3 constructs are collectively shown in Fig. 3.

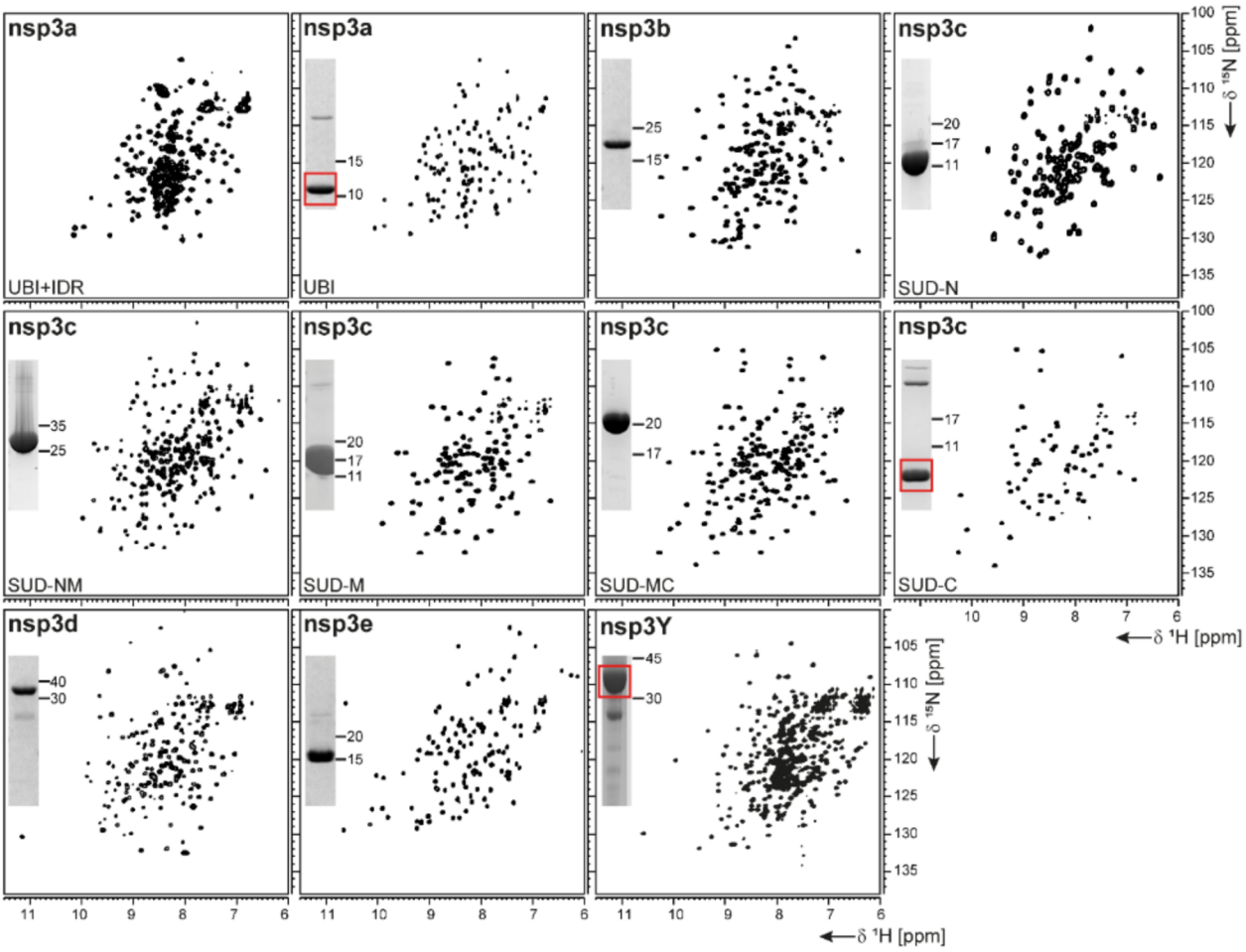

Figure 3. ^{1}H, ^{15}N-correlation spectra of investigated constructs from non-structural protein 3. Construct names of sub-domains according to Table 1 are indicated unless fl-domains are shown. A representative SDS-PAGE lane with final samples is included as inset. Red boxes indicate protein bands of interest.

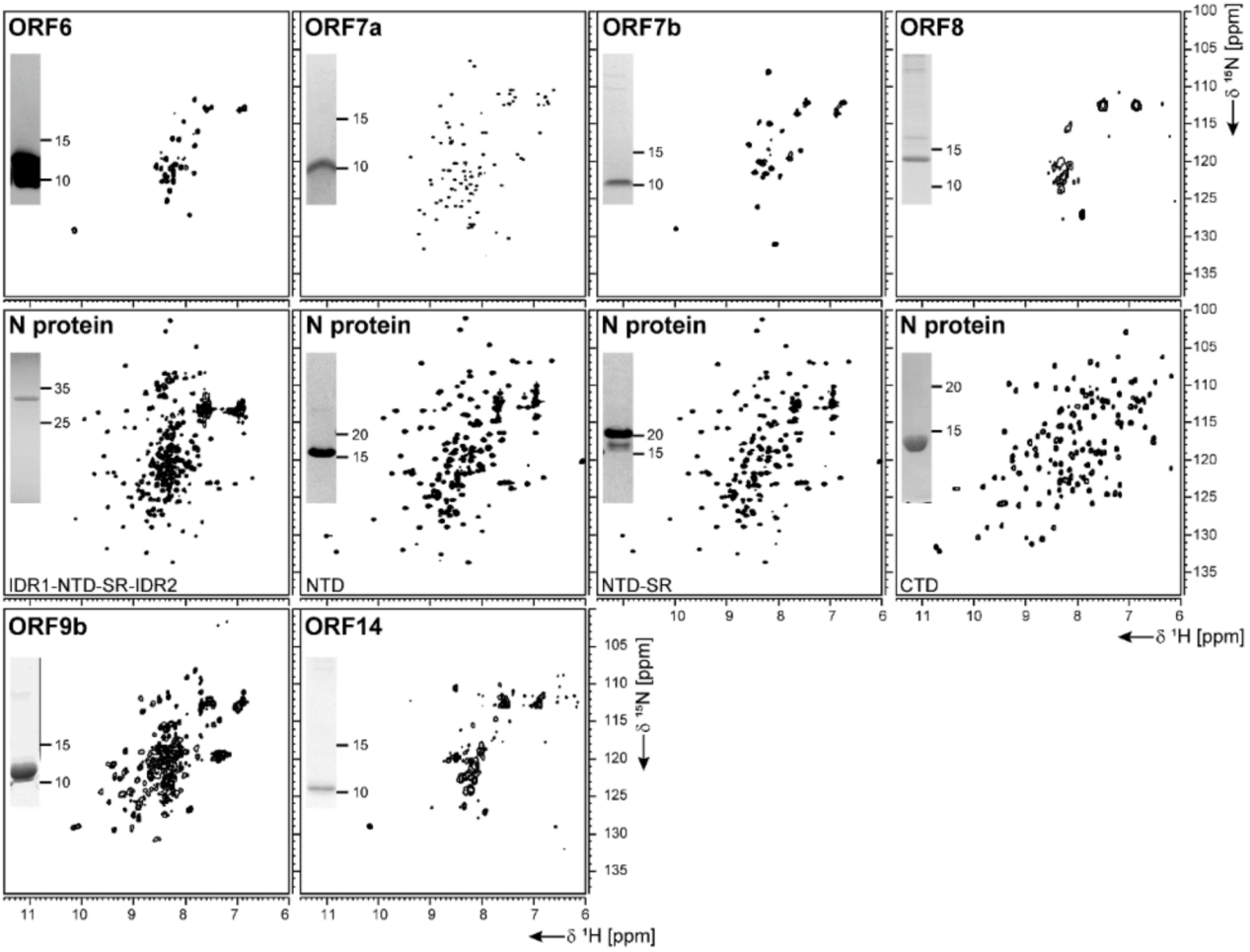

Figure 4. ^{1}H,^{15}N-correlation spectra of investigated structural and accessory proteins. Construct names according to Table 1 are indicated unless fl-proteins are shown. A representative SDS-PAGE lane with final samples is included as inset.

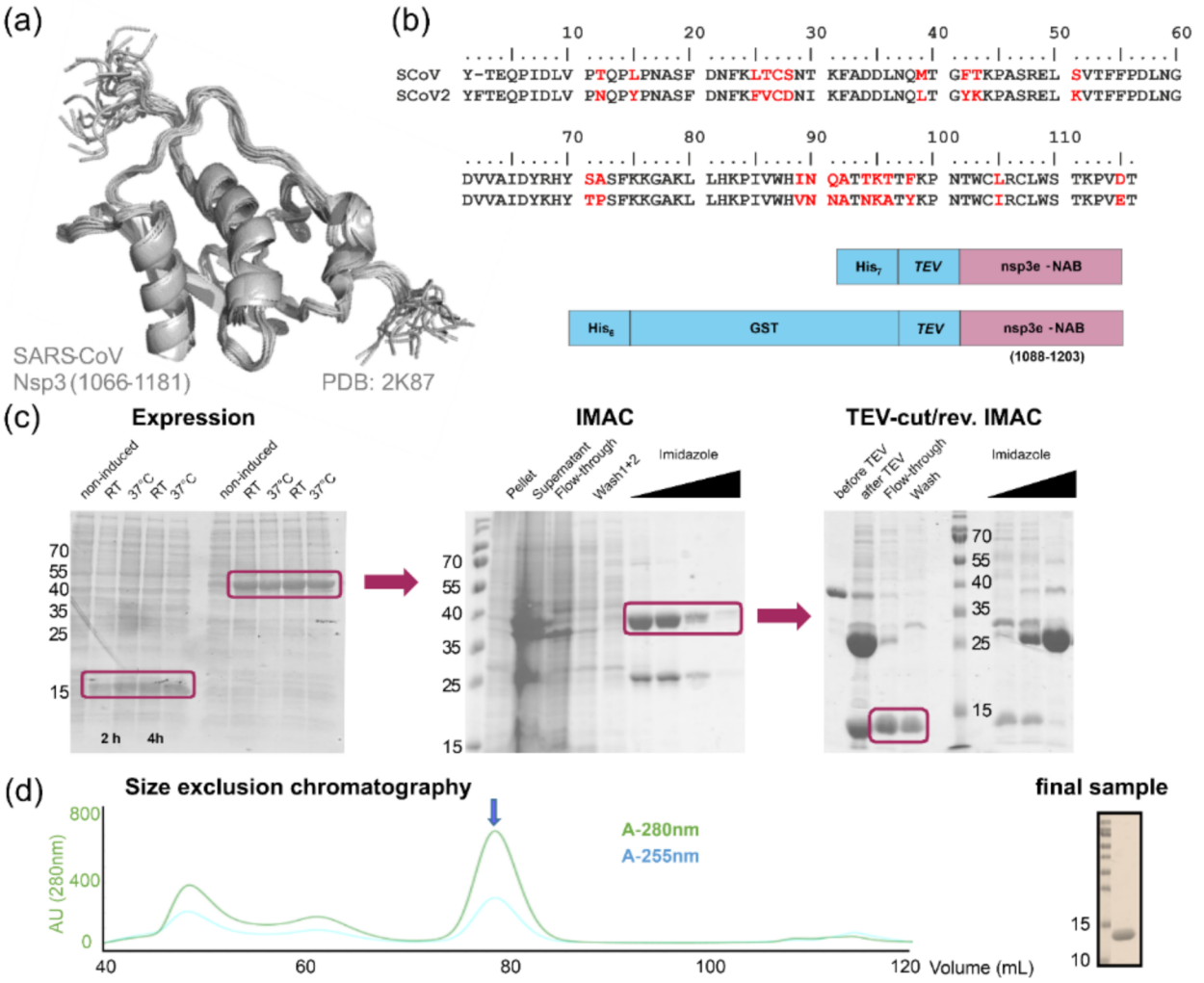

Figure 5. Rationale of construct design, expression, and IPRS purification of the nsp3e nucleic acid binding domain (NAB). (a) NMR structural ensemble of the homologous SCoV nsp3e (11). The domain boundaries as displayed are given. (b) Sequence alignment of SCoV and SCoV2 regions representing the nsp3e locus. Arrows indicate the sequence stretch as used for the structure in panel (a). The analogous region was used for design of the two protein expression constructs shown. (c) Left, SDS-PAGE showing the expression of nsp3e constructs from panel (b) over four hours at two different temperatures. Middle, SDS-PAGE showing the subsequent steps of IMAC. Right, SDS-Page showing steps and fractions obtained before and after TEV/dialysis and reverse IMAC. Boxes highlight the respective sample species of interest for further usage. (d) SEC profile of nsp3e following steps in panel (c) performed with a Superdex 75 16/600 (GE Healthcare) column in the buffer as denoted in SI3. The arrow indicates the protein peak of interest containing monomeric and homogenous nsp3e NAB devoid of significant contaminations of nucleic acids as revealed by the excellent 280/260 ratio. Right, SDS gel shows 0.5 µL of the final NMR sample used for the spectrum in Fig. 3 after concentrating relevant SEC fractions.

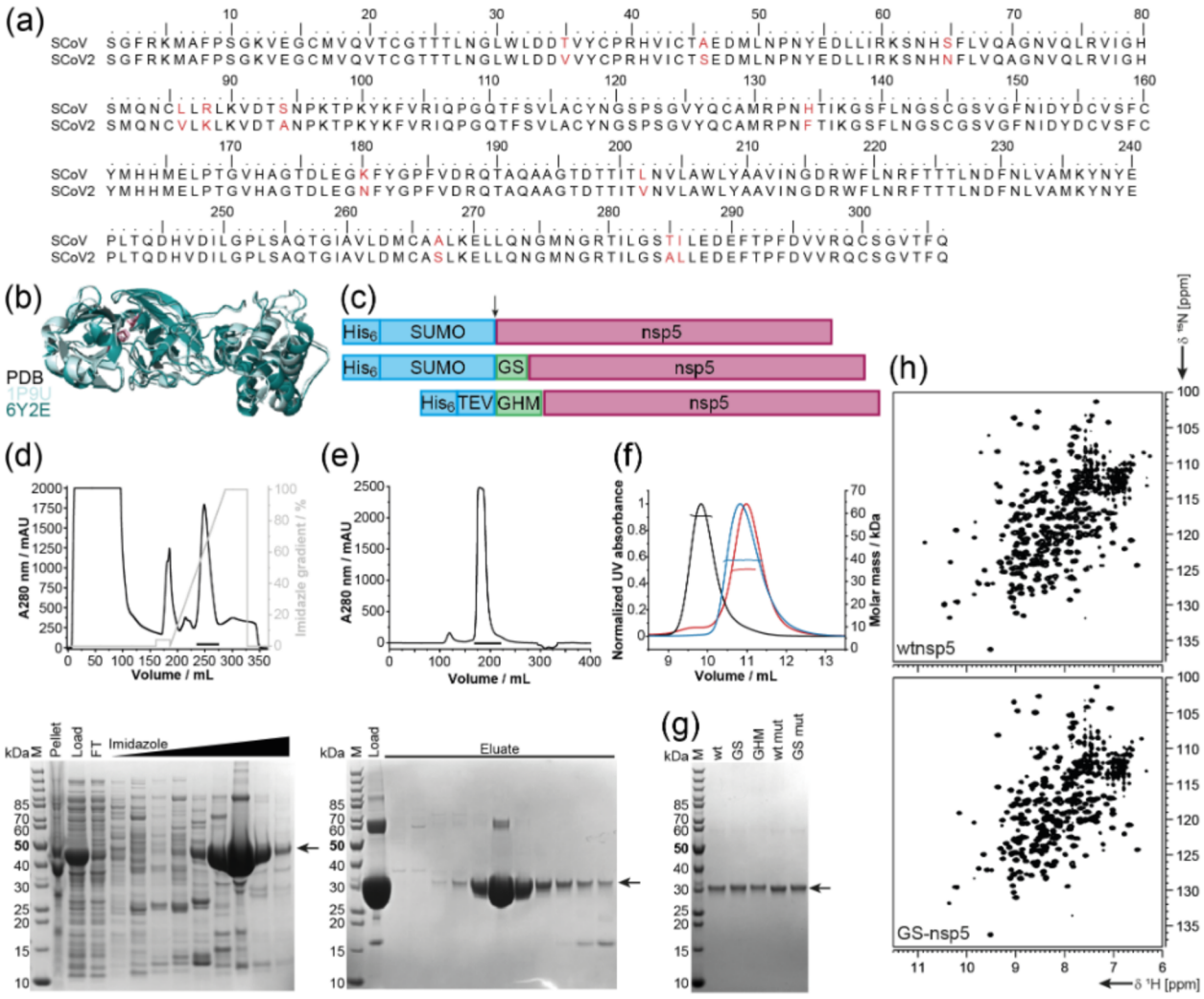

Figure 6. Rationale of construct design, expression, and purification of different nsp5 constructs. (a) Sequence alignment of SCoV and SCoV2 full-length nsp5. (b) X-ray structural overlay of the homologous SCoV (PDB 1P9U, light blue) and SCoV2 nsp5 (PDB 6Y2E, green) in cartoon representation. The catalytic dyad (H41 and C145) is shown in stick representation (magenta). (c) Schematics of nsp5 expression constructs involving purification and solubilization tags (blue), different N-termini and additional aa after cleavage (green), and nsp5 (magenta). Cleavage sites are indicated by an arrow. (d-e) An exemplary purification is shown for wtnsp5. IMAC (d) and SEC (e) chromatograms (upper panels) and the corresponding gels (lower panels). Black bars in the chromatograms indicate pooled fractions. Gel samples are: M: MW standard; Pellet/Load: pellet/supernatant after cell lysis; FT: IMAC flow-through; Imidazole: eluted fractions with linear imidazole gradient; Eluate: eluted SEC fractions from input (Load). (f) A SDS-PAGE showing all purified nsp5 constructs. The arrow indicates nsp5. (g) SEC-MALS analysis with ~0.5 µg of wt nsp5 without additional aa (wtnsp5, black) with GS (GS-nsp5, blue), and with GHM (GHM-nsp5, red)) in NMR buffer on a Superdex 75, 10/300 GL (GE Healthcare) column. Horizontal lines indicate fractions of monodisperse nsp5 used for MW determination. (h) Exemplary [^{15}N, ^{1}H]-BEST-TROSY spectra measured at 298K for the dimeric wtnsp5 (upper spectrum), and monomeric GS-nsp5 (lower spectrum). See SI4 for technical details regarding this figure.

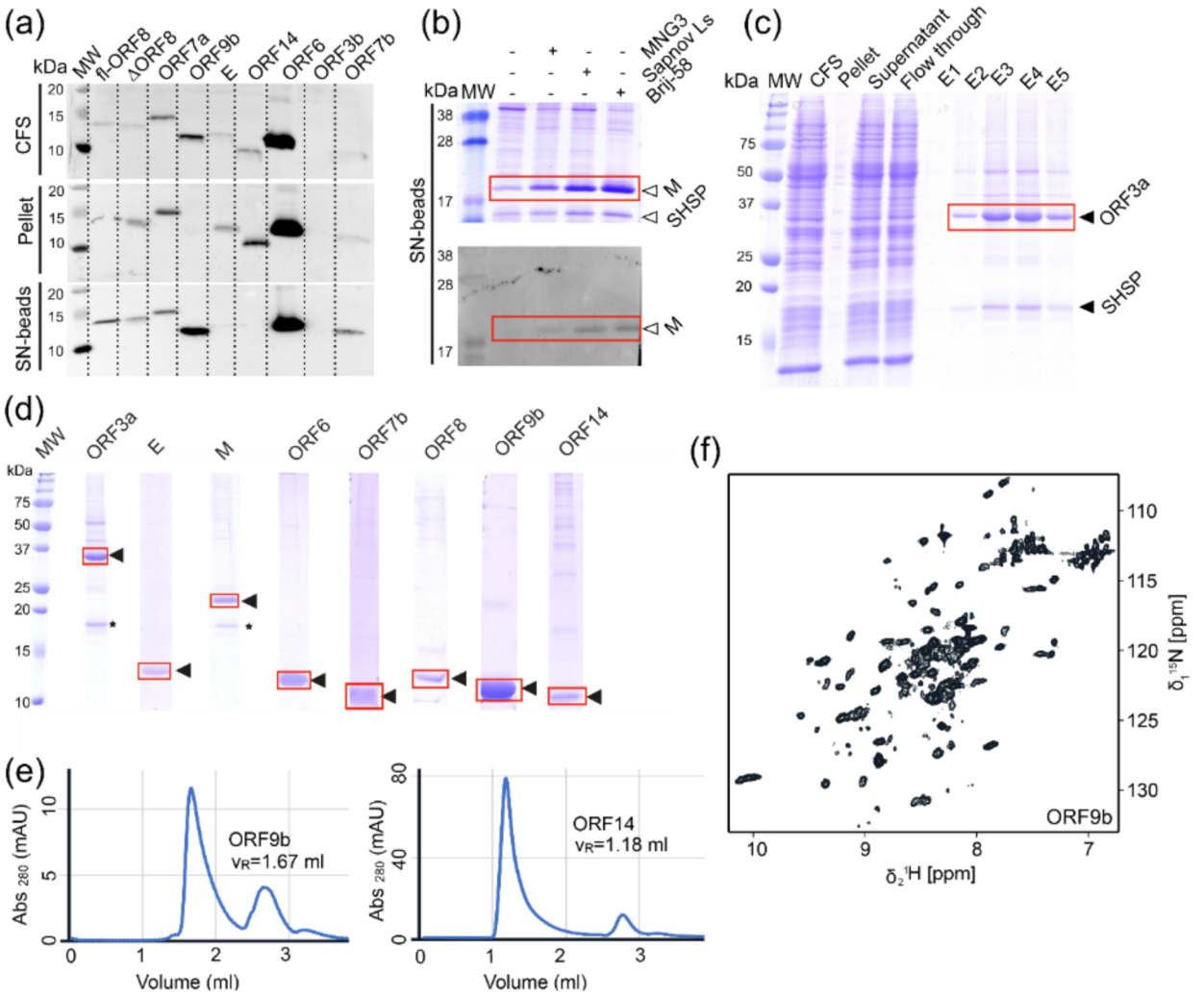

Figure 7. Cell-free protein synthesis of accessory ORFs and structural proteins E and M. (a) Screening for expression and solubility of different ORFs using small-scale reactions. The total cell-free reaction (CFS), the pellet after centrifugation, as well as the supernatant (SN) captured on magnetic beads coated with Strep-Tactin were analyzed. All tested proteins were synthesized, with the exception of ORF3b. MW, MW standard. (b) Detergent solubilization tests using three different detergents, here at the example of the M protein, shown by SDS -PAGE and Western Blot. (c) Proteins are purified in a single step using a Strep-Tactin column. For ORF3a (and also for M), a small heat-shock protein of the HSP20 family is co-purified, as identified by mass spectrometry (see also * in Panel d). (d) SDS-PAGE of the ^{2}H,^{13}C,^{15}N-labeled proteins used as NMR samples. Yields were between 0.2 and 1 mg protein per mL wheat-germ extract used. (e) SEC profiles for two ORFs. Left, ORF9b migrates as expected for a dimer. Right, OFR14 shows large assemblies corresponding to approximately 9 protein units and the DDM detergent micelle. (f) 2D [^{15}N, ^{1}H]-BEST-TROSY spectrum of ORF9b, recorded at 900 MHz in 1h at 298 K, on less than 1 mg of protein. See SI13-19 and SI19-20 for technical and experimental details regarding this figure.