

# UNIVERSITY OF BIRMINGHAM

University of Birmingham  
Research at Birmingham

## Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications

Altincekic, Nadide; Korn, Sophie Marianne; Qureshi, Nusrat Shahin; Dujardin, Marie; Ninot-Pedrosa, Martí; Abele, Rupert; Abi Saad, Marie Jose; Alfano, Caterina; Almeida, Fabio C.L.; Alshamleh, Islam; de Amorim, Gisele Cardoso; Anderson, Thomas K.; Anobom, Cristiane D.; Anorma, Chelsea; Bains, Jasleen Kaur; Bax, Adriaan; Blackledge, Martin; Blechar, Julius; Böckmann, Anja; Brigandat, Louis

DOI:

[10.3389/fmolb.2021.653148](https://doi.org/10.3389/fmolb.2021.653148)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Altincekic, N, Korn, SM, Qureshi, NS, Dujardin, M, Ninot-Pedrosa, M, Abele, R, Abi Saad, MJ, Alfano, C, Almeida, FCL, Alshamleh, I, de Amorim, GC, Anderson, TK, Anobom, CD, Anorma, C, Bains, JK, Bax, A, Blackledge, M, Blechar, J, Böckmann, A, Brigandat, L, Bula, A, Bütikofer, M, Camacho-Zarco, AR, Carlomagno, T, Caruso, IP, Ceylan, B, Chaikuad, A, Chu, F, Cole, L, Crosby, MG, de Jesus, V, Dhamotharan, K, Felli, IC, Ferner, J, Fleischmann, Y, Fogeron, ML, Fourkiotis, NK, Fuks, C, Fürtig, B, Gallo, A, Gande, SL, Gerez, JA, Ghosh, D, Gomes-Neto, F, Gorbatyuk, O, Guseva, S, Hacker, C, Häfner, S, Hao, B, Hargittay, B, Henzler-Wildman, K, Hoch, JC, Hohmann, KF, Hutchison, MT, Jaudzems, K, Jović, K, Kaderli, J, Kalniņš, G, Kaņepe, I, Kirchdoerfer, RN, Kirkpatrick, J, Knapp, S, Krishnathas, R, Kutz, F, zur Lage, S, Lambertz, R, Lang, A, Laurents, D, Lecoq, L, Linhard, V, Löhr, F, Malki, A, Bessa, LM, Martin, RW, Matzel, T, Maurin, D, McNutt, SW, Mebus-Antunes, NC, Meier, BH, Meiser, N, Mompeán, M, Monaca, E, Montserret, R, Mariño Perez, L, Moser, C, Muhle-Goll, C, Neves-Martins, TC, Ni, X, Norton-Baker, B, Pierattelli, R, Pontoriero, L, Pustovalova, Y, Ohlenschläger, O, Orts, J, Da Poian, AT, Pyper, DJ, Richter, C, Riek, R, Rienstra, CM, Robertson, A, Pinheiro, AS, Sabbatella, R, Salvi, N, Saxena, K, Schulte, L, Schiavina, M, Schwalbe, H, Silber, M, Almeida, MDS, Sprague-Piercy, MA, Spyroulias, GA, Sreeramulu, S, Tants, JN, Târs, K, Torres, F, Töws, S, Treviño, M, Trucks, S, Tsika, AC, Varga, K, Wang, Y, Weber, ME, Weigand, JE, Wiedemann, C, Wirmer-Bartoschek, J, Wirtz Martin, MA, Zehnder, J, Hengesbach, M & Schlundt, A 2021, 'Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications', *Frontiers in Molecular Biosciences*, vol. 8, 653148. <https://doi.org/10.3389/fmolb.2021.653148>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

OPEN ACCESS



Edited by:

Qian Han,  
Hainan University, China

Reviewed by:

David Douglas Boehr,  
Pennsylvania State University, United States  
Luis G. Bribea,  
National Polytechnic Institute of Mexico  
(CINVESTAV), Mexico

\*Correspondence:

Anja Böckmann  
a.boeckmann@ibcp.fr  
Harald Schwalbe  
schwalbe@nmr.uni-frankfurt.de  
Martin Hengesbach  
hengesch@nmr.uni-frankfurt.de  
Andreas Schlundt  
schlundt@bio.uni-frankfurt.de

<sup>†</sup>These authors have contributed equally to this work  
and share first authorship

<sup>‡</sup>These authors share last authorship

Specialty section:

This article was submitted to  
Structural Biology,  
a section of the journal  
Frontiers in Molecular Biosciences

Received: 13 January 2021

Accepted: 04 February 2021

Published: 10 May 2021

Citation:

Altıncekcik N, Korn SM, Qureshi NS, Dujardin M, Ninot-Pedrosa M, Abele R, Abi Saad MJ, Alfano C, Almeida FCL, Alshamleh I, de Amorim GC, Anderson TK, Anobom CD, Anorma C, Bains JK, Bax A, Blackledge M, Blechar J, Böckmann A, Brigandat L, Bula A, Bütikofer M, Camacho-Zarco AP, Carlomagno T, Caruso IP, Ceylan B, Chaikuad A, Chu F, Cole L, Crosby MG, de Jesus V, Dhamotharan K, Felli IC, Ferner J, Fleischmann Y, Fogeron M-L, Fourkotis NK, Fuks C, Fürtig B, Gallo A, Gande L, Gerez JA, Ghosh D, Gomes-Neto F, Gorbatyuk O, Guseva S, Hacker C, Häfner S, Hao B, Hargittay B, Henzler-Wildman K, Hoch JC, Hohmann KF, Hutchison MT, Jaudzems K, Jović K, Kaderli J, Kalniņš G, Kaņepe I, Kirchoerfer RN, Kirkpatrick J, Knapp S, Krishnathas R, Kutz F, zur Lage S, Lambert R, Lang A, Laurents D, Lecoq L, Linhard V, Löhner F, Malki A, Bessa LM, Martin RW, Matzel T, Maurin D, McNutt SW, Mebus-Antunes NC, Meier BH, Meiser N, Mompeán M, Monaca E, Montserret R, Mariño Perez L, Moser C, Muhle-Goll C, Neves-Martins TC, Ni X, Norton-Baker B, Pierattelli R, Pontoriero L, Pustovalova Y, Ohlenschläger O, Orts J, Da Poian AT, Pyper DJ, Richter C, Riek R, Riesenra CM, Robertson A, Pinheiro AS, Sabbatella R, Salvi N, Saxena K, Schulte L, Schiavina M, Schwalbe H, Silber M, Almeida MdS, Sprague-Piercy MA, Spyroulas GA, Sreeramulu S, Tants J-N, Tars K, Torres F, Töws S, Treviño MA, Trucks S, Tsika AC, Varga K, Wang Y, Weber ME, Weigand JE, Wiedemann C, Wimer-Bartoschek J, Wirtz Martin MA, Zehnder J, Hengesbach M and Schlundt A (2021) Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. *Front. Mol. Biosci.* 8:653148. doi: 10.3389/fmolb.2021.653148

# Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications

Nadide Altıncekcik<sup>1,2†</sup>, Sophie Marianne Korn<sup>2,3†</sup>, Nusrat Shahin Qureshi<sup>1,2†</sup>, Marie Dujardin<sup>4†</sup>, Marti Ninot-Pedrosa<sup>4†</sup>, Rupert Abele<sup>5</sup>, Marie Jose Abi Saad<sup>6</sup>, Caterina Alfano<sup>7</sup>, Fabio C. L. Almeida<sup>8,9</sup>, Islam Alshamleh<sup>1,2</sup>, Gisele Cardoso de Amorim<sup>8,10</sup>, Thomas K. Anderson<sup>11</sup>, Cristiane D. Anobom<sup>8,12</sup>, Chelsea Anorma<sup>13</sup>, Jasleen Kaur Bains<sup>1,2</sup>, Adriaan Bax<sup>14</sup>, Martin Blackledge<sup>15</sup>, Julius Blechar<sup>1,2</sup>, Anja Böckmann<sup>4,\*†</sup>, Louis Brigandat<sup>4</sup>, Anna Bula<sup>16</sup>, Matthias Bütikofer<sup>6</sup>, Aldo R. Camacho-Zarco<sup>15</sup>, Teresa Carlomagno<sup>17,18</sup>, Icaro Putinhon Caruso<sup>8,9,19</sup>, Betül Ceylan<sup>1,2</sup>, Apirat Chaikuad<sup>20,21</sup>, Feixia Chu<sup>22</sup>, Laura Cole<sup>4</sup>, Marquise G. Crosby<sup>23</sup>, Vanessa de Jesus<sup>1,2</sup>, Karthikeyan Dhamotharan<sup>2,3</sup>, Isabella C. Felli<sup>24,25</sup>, Jan Ferner<sup>1,2</sup>, Yanick Fleischmann<sup>6</sup>, Marie-Laure Fogeron<sup>4</sup>, Nikolaos K. Fourkotis<sup>26</sup>, Christin Fuks<sup>1</sup>, Boris Fürtig<sup>1,2</sup>, Angelo Gallo<sup>26</sup>, Santosh L. Gande<sup>1,2</sup>, Juan Atilio Gerez<sup>6</sup>, Dhiman Ghosh<sup>6</sup>, Francisco Gomes-Neto<sup>8,27</sup>, Oksana Gorbatyuk<sup>28</sup>, Serafima Guseva<sup>15</sup>, Carolin Hacker<sup>29</sup>, Sabine Häfner<sup>30</sup>, Bing Hao<sup>28</sup>, Bruno Hargittay<sup>1,2</sup>, K. Henzler-Wildman<sup>11</sup>, Jeffrey C. Hoch<sup>28</sup>, Katharina F. Hohmann<sup>1,2</sup>, Marie T. Hutchison<sup>1,2</sup>, Kristaps Jaudzems<sup>16</sup>, Katarina Jović<sup>22</sup>, Janina Kaderli<sup>6</sup>, Gints Kalniņš<sup>31</sup>, Iveta Kaņepe<sup>16</sup>, Robert N. Kirchoerfer<sup>11</sup>, John Kirkpatrick<sup>17,18</sup>, Stefan Knapp<sup>20,21</sup>, Robin Krishnathas<sup>1,2</sup>, Felicitas Kutz<sup>1,2</sup>, Susanne zur Lage<sup>18</sup>, Roderick Lambert<sup>3</sup>, Andras Lang<sup>30</sup>, Douglas Laurents<sup>32</sup>, Lauriane Lecoq<sup>4</sup>, Verena Linhard<sup>1,2</sup>, Frank Löhner<sup>2,33</sup>, Anas Malki<sup>15</sup>, Luiza Mamigonian Bessa<sup>15</sup>, Rachel W. Martin<sup>13,23</sup>, Tobias Matzel<sup>1,2</sup>, Damien Maurin<sup>15</sup>, Seth W. McNutt<sup>22</sup>, Nathane Cunha Mebus-Antunes<sup>8,9</sup>, Beat H. Meier<sup>6</sup>, Nathalie Meiser<sup>1</sup>, Miguel Mompeán<sup>32</sup>, Elisa Monaca<sup>7</sup>, Roland Montserret<sup>4</sup>, Laura Mariño Perez<sup>15</sup>, Celine Moser<sup>34</sup>, Claudia Muhle-Goll<sup>34</sup>, Thais Cristina Neves-Martins<sup>8,9</sup>, Xiamonin Ni<sup>20,21</sup>, Brenna Norton-Baker<sup>13</sup>, Roberta Pierattelli<sup>24,25</sup>, Letizia Pontoriero<sup>24,25</sup>, Yulia Pustovalova<sup>28</sup>, Oliver Ohlenschläger<sup>30</sup>, Julien Orts<sup>6</sup>, Andrea T. Da Poian<sup>9</sup>, Dennis J. Pyper<sup>1,2</sup>, Christian Richter<sup>1,2</sup>, Roland Riek<sup>6</sup>, Chad M. Riesenra<sup>35</sup>, Angus Robertson<sup>14</sup>, Anderson S. Pinheiro<sup>8,12</sup>, Raffaele Sabbatella<sup>7</sup>, Nicola Salvi<sup>15</sup>, Krishna Saxena<sup>1,2</sup>, Linda Schulte<sup>1,2</sup>, Marco Schiavina<sup>24,25</sup>, Harald Schwalbe<sup>1,2,\*†</sup>, Mara Silber<sup>34</sup>, Marcus da Silva Almeida<sup>8,9</sup>, Marc A. Sprague-Piercy<sup>23</sup>, Georgios A. Spyroulas<sup>26</sup>, Sridhar Sreeramulu<sup>1,2</sup>, Jan-Niklas Tants<sup>2,3</sup>, Kaspars Tars<sup>31</sup>, Felix Torres<sup>6</sup>, Sabrina Töws<sup>3</sup>, Miguel Á. Treviño<sup>32</sup>, Sven Trucks<sup>1</sup>, Aikaterini C. Tsika<sup>26</sup>, Krisztina Varga<sup>22</sup>, Ying Wang<sup>17</sup>, Marco E. Weber<sup>6</sup>, Julia E. Weigand<sup>36</sup>, Christoph Wiedemann<sup>37</sup>, Julia Wimer-Bartoschek<sup>1,2</sup>, Maria Alexandra Wirtz Martin<sup>1,2</sup>, Johannes Zehnder<sup>6</sup>, Martin Hengesbach<sup>1,\*†</sup> and Andreas Schlundt<sup>2,3,\*†</sup>

<sup>1</sup>Institute for Organic Chemistry and Chemical Biology, Goethe University Frankfurt, Frankfurt am Main, Germany, <sup>2</sup>Center of Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, Frankfurt am Main, Germany, <sup>3</sup>Institute for Molecular Biosciences, Goethe University Frankfurt, Frankfurt am Main, Germany, <sup>4</sup>Molecular Microbiology and Structural Biochemistry, UMR 5086, CNRS/Lyon University, Lyon, France, <sup>5</sup>Institute for Biochemistry, Goethe University Frankfurt, Frankfurt am Main, Germany, <sup>6</sup>Swiss Federal Institute of Technology, Laboratory of Physical Chemistry, ETH Zurich, Zurich, Switzerland, <sup>7</sup>Structural Biology and Biophysics Unit, Fondazione Ri.MED, Palermo, Italy, <sup>8</sup>National Center of Nuclear Magnetic Resonance (CNRMN, CENABIO), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, <sup>9</sup>Institute of Medical Biochemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, <sup>10</sup>Multidisciplinary Center for Research in Biology (NUMPEX), Campus Duque de Caxias Federal University of Rio de Janeiro, Duque de Caxias, Brazil, <sup>11</sup>Institute for Molecular Virology, University of Wisconsin-Madison, Madison, WI, United States, <sup>12</sup>Institute of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, <sup>13</sup>Department of Chemistry, University of California, Irvine, CA, United States, <sup>14</sup>LCP, NIDDK, NIH, Bethesda, MD, United States, <sup>15</sup>Univ. Grenoble Alpes, CNRS, CEA, IBS, Grenoble, France, <sup>16</sup>Latvian Institute of Organic Synthesis, Riga, Latvia, <sup>17</sup>BMWZ and Institute of Organic

Chemistry, Leibniz University Hannover, Hannover, Germany, <sup>18</sup>Group of NMR-Based Structural Chemistry, Helmholtz Centre for Infection Research, Braunschweig, Germany, <sup>19</sup>Multiuser Center for Biomolecular Innovation (CMIB), Department of Physics, São Paulo State University (UNESP), São José do Rio Preto, Brazil, <sup>20</sup>Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt am Main, Germany, <sup>21</sup>Structural Genomics Consortium, Buchmann Institute for Molecular Life Sciences, Frankfurt am Main, Germany, <sup>22</sup>Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH, United States, <sup>23</sup>Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, United States, <sup>24</sup>Magnetic Resonance Centre (CERM), University of Florence, Sesto Fiorentino, Italy, <sup>25</sup>Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy, <sup>26</sup>Department of Pharmacy, University of Patras, Patras, Greece, <sup>27</sup>Laboratory of Toxinology, Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, Brazil, <sup>28</sup>Department of Molecular Biology and Biophysics, UConn Health, Farmington, CT, United States, <sup>29</sup>Signals GmbH & Co. KG, Frankfurt am Main, Germany, <sup>30</sup>Leibniz Institute on Aging—Fritz Lipmann Institute (FLI), Jena, Germany, <sup>31</sup>Latvian Biomedical Research and Study Centre, Riga, Latvia, <sup>32</sup>"Rocasolano" Institute for Physical Chemistry (IQFR), Spanish National Research Council (CSIC), Madrid, Spain, <sup>33</sup>Institute of Biophysical Chemistry, Goethe University Frankfurt, Frankfurt am Main, Germany, <sup>34</sup>IBG-4, Karlsruhe Institute of Technology, Karlsruhe, Germany, <sup>35</sup>Department of Biochemistry and National Magnetic Resonance Facility at Madison, University of Wisconsin-Madison, Madison, WI, United States, <sup>36</sup>Department of Biology, Technical University of Darmstadt, Darmstadt, Germany, <sup>37</sup>Institute of Biochemistry and Biotechnology, Charles Tanford Protein Centre, Martin Luther University Halle-Wittenberg, Halle/Saale, Germany

The highly infectious disease COVID-19 caused by the *Betacoronavirus* SARS-CoV-2 poses a severe threat to humanity and demands the 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 catalog 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](https://covid19-nmr.com), we here provide highly valuable resources for the production of SARS-CoV-2 proteins in isotope-labeled form.

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

## 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 *Betacoronaviruses*, a genus of the Coronaviridae family covering the  $\alpha$ - $\delta$  genera (Leao et al., 2020). The large RNA genome of SCoV2 has an intricate, highly condensed arrangement of coding sequences (Wu et al., 2020). 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 the generation of a 7,096-amino acid 794 kDa polypeptide. If the pseudoknot is not formed, expression of the first ORF generates a 4,405-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, nonstructural proteins (nsps). ORF1a encodes eleven nsps, and ORF1ab additionally encodes the nsps 12–16. 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 (Finkel et al., 2020) and whose identification and classification are a matter of ongoing research (Nelson et al., 2020; Pavesi, 2020). 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, such as X-ray crystallography on proteases (Zhang et al., 2020) and methyltransferases (MTase) (Krafcikova et al., 2020), cryo-EM of the RNA polymerase (Gao et al., 2020; Yin et al., 2020), and liquid-state (Almeida et al., 2007; Serrano et al., 2009; Cantini et al., 2020; Gallo et al., 2020; Korn et al., 2020a; Korn et al., 2020b;

**TABLE 1** | SCoV2 protein constructs expressed and purified, given with the genomic position and corresponding PDBs for construct design.

| Protein genome position (nt) <sup>a</sup> | Trivial name construct expressed          | Size (aa)    | Boundaries  | MW (kDa)     | Homol. SCoV (%) <sup>b</sup> | Template PDB <sup>c</sup> | SCoV2 PDB <sup>d</sup> |
|---|---|--------------|-------------|--------------|------------------------------|---------------------------|------------------------|
| <b>nsp1</b><br>266–805                    | <b>Leader</b>                             | <b>180</b>   |             | <b>19.8</b>  | <b>84</b>                    |                           |                        |
|   | Full-length                               | 180          | 1–180       | 19.8         | 83                           |                           |                        |
|   | Globular domain (GD)                      | 116          | 13–127      | 12.7         | 85                           | 2GDT                      | 7K7P                   |
| <b>nsp2</b><br>806–2,719                  |   | <b>638</b>   |             | <b>70.5</b>  | <b>68</b>                    |                           |                        |
|   | C-terminal IDR (CtDR)                     | 45           | 557–601     | 4.9          | 55                           |                           |                        |
| <b>nsp3</b><br>2,720–8,554                |   | <b>1,945</b> |             | <b>217.3</b> | <b>76</b>                    |                           |                        |
| 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     | 15.5         | 69                           | 2W2G                      |                        |
| c   | SUD-NM                                    | 267          | 409–675     | 29.6         | 74                           | 2W2G                      |                        |
| c   | SUD-M                                     | 125          | 551–675     | 14.2         | 82                           | 2W2G                      |                        |
| c   | SUD-MC                                    | 195          | 551–743     | 21.9         | 79                           | 2KQV                      |                        |
| c   | SUD-C                                     | 64           | 680–743     | 7.4          | 73                           | 2KAF                      |                        |
| d   | Papain-like protease PL <sup>PRO</sup>    | 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                           |                           |                        |
| <b>nsp5</b><br>10,055–10,972              | <b>Main protease (M<sup>PRO</sup>)</b>    | <b>306</b>   |             | <b>33.7</b>  | <b>96</b>                    |                           |                        |
|   | Full-length <sup>e</sup>                  | 306          | 1–306       | 33.7         | 96                           | 6Y84                      | 6Y84                   |
| <b>nsp7</b><br>11,843–12,091              |   | <b>83</b>    |             | <b>9.2</b>   | <b>99</b>                    |                           |                        |
|   | Full-length                               | 83           | 1–83        | 9.2          | 99                           | 6WIQ                      | 6WIQ                   |
| <b>nsp8</b><br>12,092–12,685              |   | <b>198</b>   |             | <b>21.9</b>  | <b>98</b>                    |                           |                        |
|   | Full-length                               | 198          | 1–198       | 21.9         | 97                           | 6WIQ                      | 6WIQ                   |
| <b>nsp9</b><br>12,686–13,024              |   | <b>113</b>   |             | <b>12.4</b>  | <b>97</b>                    |                           |                        |
|   | Full-length                               | 113          | 1–113       | 12.4         | 97                           | 6W4B                      | 6W4B                   |
| <b>nsp10</b><br>13,025–13,441             |   | <b>139</b>   |             | <b>14.8</b>  | <b>97</b>                    |                           |                        |
|   | Full-length                               | 139          | 1–139       | 14.8         | 97                           | 6W4H                      | 6W4H                   |
| <b>nsp13</b><br>16,237–18,039             | <b>Helicase</b>                           | <b>601</b>   |             | <b>66.9</b>  | <b>100</b>                   |                           |                        |
|   | Full-length                               | 601          | 1–601       | 66.9         | 100                          | 6ZSL                      | 6ZSL                   |
| <b>nsp14</b><br>18,040–19,620             | <b>Exonuclease/<br/>methyltransferase</b> | <b>527</b>   |             | <b>59.8</b>  | <b>95</b>                    |                           |                        |
|   | Full-length                               | 527          | 1–527       | 59.8         | 95                           | 5NFY                      |                        |
|   | MTase domain                              | 240          | 288–527     | 27.5         | 95                           |                           |                        |
| <b>nsp15</b><br>19,621–20,658             | <b>Endonuclease</b>                       | <b>346</b>   |             | <b>38.8</b>  | <b>89</b>                    |                           |                        |
|   | Full-length                               | 346          | 1–346       | 38.8         | 89                           | 6W01                      | 6W01                   |
| <b>nsp16</b><br>20,659–21,552             | <b>Methyltransferase</b>                  | <b>298</b>   |             | <b>33.3</b>  | <b>93</b>                    |                           |                        |
|   | Full-length                               | 298          | 1–298       | 33.3         | 93                           | 6W4H                      | 6W4H                   |
| <b>ORF3a</b><br>25,393–26,220             |   | <b>275</b>   |             | <b>31.3</b>  | <b>72</b>                    |                           |                        |
|   | Full-length                               | 275          | 1–275       | 31.3         | 72                           | 6XDC                      | 6XDC                   |
| <b>ORF4</b><br>26,245–26,472              | <b>Envelope (E) protein</b>               | <b>75</b>    |             | <b>8.4</b>   | <b>95</b>                    |                           |                        |
|   | Full-length                               | 75           | 1–75        | 8.4          | 95                           | 5X29                      | 7K3G                   |
| <b>ORF5</b><br>26,523–27,387              | <b>Membrane<br/>glycoprotein (M)</b>      | <b>222</b>   |             | <b>25.1</b>  | <b>91</b>                    |                           |                        |
|   | Full-length                               | 222          | 1–222       | 25.1         | 91                           |                           |                        |
| <b>ORF6</b><br>27,202–27,387              |   | <b>61</b>    |             | <b>7.3</b>   | <b>69</b>                    |                           |                        |
|   | Full-length                               | 61           | 1–61        | 7.3          | 69                           |                           |                        |

(Continued on following page)

**TABLE 1 |** (Continued) SCoV2 protein constructs expressed and purified, given with the genomic position and corresponding PDBs for construct design.

| Protein genome position (nt) <sup>a</sup> | Trivial name construct expressed | Size (aa)  | Boundaries | MW (kDa)    | Homol. SCoV (%) <sup>b</sup> | Template PDB <sup>c</sup> | SCoV2 PDB <sup>d</sup> |
|---|----------------------------------|------------|------------|-------------|------------------------------|---------------------------|------------------------|
| <b>ORF7a</b><br>27,394–27,759             |                                  | <b>121</b> |            | <b>13.7</b> | <b>85</b>                    |                           |                        |
|   | Ectodomain (ED)                  | 66         | 16–81      | 7.4         | 85                           | 1XAK                      | 6W37                   |
| <b>ORF7b</b><br>27,756–27,887             |                                  | <b>43</b>  |            | <b>5.2</b>  | <b>85</b>                    |                           |                        |
|   | Full-length                      | 43         | 1–43       | 5.2         | 85                           |                           |                        |
| <b>ORF8</b><br>27,894–28,259              |                                  | <b>121</b> |            | <b>13.8</b> | <b>32</b>                    |                           |                        |
|   | Full-length                      | 121        | 1–121      | 13.8        | 32                           |                           |                        |
|   | w/o signal peptide               | 106        | 16–121     | 12          | 41                           | 7JTL                      | 7JTL                   |
| <b>ORF9a</b><br>28,274–29,533             | <b>Nucleocapsid (N)</b>          | <b>419</b> |            | <b>45.6</b> | <b>91</b>                    |                           |                        |
|   | 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                   |
| <b>ORF9b</b><br>28,284–28,574             |                                  | <b>97</b>  |            | <b>10.8</b> | <b>72</b>                    |                           |                        |
|   | Full-length                      | 97         | 1–97       | 10.8        | 72                           | 6Z4U                      | 6Z4U                   |
| <b>ORF14</b><br>28,734–28,952             |                                  | <b>73</b>  |            | <b>8</b>    | <b>n.a</b>                   |                           |                        |
|   | Full-length                      | 73         | 1–73       | 8           | n.a                          |                           |                        |
| <b>ORF10</b><br>29,558–29,674             |                                  | <b>38</b>  |            | <b>4.4</b>  | <b>29</b>                    |                           |                        |
|   | Full-length                      | 38         | 1–38       | 4.4         | 29                           |                           |                        |

<sup>a</sup>Genome position in nt corresponding to SCoV2 NCBI reference genome entry NC\_045512.2, identical to GenBank entry MN908947.3.

<sup>b</sup>Sequence identities to SCoV are calculated from an alignment with corresponding protein sequences based on the genome sequence of NCBI Reference NC\_004718.3.

<sup>c</sup>Representative PDB that was available at the beginning of construct design, either SCoV or SCoV2.

<sup>d</sup>Representative PDB available for SCoV2 (as of December 2020).

<sup>e</sup>Additional point mutations in fl-construct have been expressed.

n.a.: not applicable.

Kubatova et al., 2020; Tonelli et al., 2020) and solid-state NMR spectroscopy of transmembrane (TM) proteins (Mandala et al., 2020). These studies have significantly improved our understanding on the 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 (Esposito et al., 2020), antibody identification (Jiang et al., 2020), and interactions with other proteins or components of the host cell (Bojkova et al., 2020; Gordon et al., 2020). These examples highlight the importance of broad approaches for the recombinant production of viral proteins.

The research consortium *COVID19-NMR* founded in 2020 seeks to support the search for antiviral 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 and 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 (Wacker et al., 2020), to provide a basis for RNA-oriented fragment screens with NMR.

We here compile a compendium of more than 50 protocols (see **Supplementary Tables S11–S123**) for the production and purification of 23 of the 30 SCoV2 proteins or fragments thereof (summarized in **Tables 1, 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 (**Supplementary 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 homolog 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 and 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 (Fogeron et al., 2015b; Fogeron et al., 2017; Jirasko



**TABLE 2** | Summary of SCoV2 protein production results in *Covid19-NMR*.

| Construct expressed | Yields (mg/L) <sup>a</sup><br>or (mg/ml) <sup>b</sup> | Results      | Comments   | BMRB  | Supplementary Material |
|---------------------|---|--------------|--|---|------------------------|
| <b>nsp1</b>         |   |              |  |   |                        |
| fl                  | 5   | NMR assigned | Expression only at >20°C; after 7 days at 25°C partial proteolysis   | 50620 <sup>d</sup>                                | SI1                    |
| GD                  | >0.5  | HSQC         | High expression; mainly insoluble; higher salt increases stability (>250 mM)   |   |                        |
| <b>nsp2</b>         |   |              |  |   |                        |
| CtDR                | 0.7–1.5   | NMR assigned | Assignment with His-tag shown in (Mompéan et al., 2020)  | 50687 <sup>c</sup>                                | SI2                    |
| <b>nsp3</b>         |   |              |  |   |                        |
| UBI                 | 0.7   | HSQC         | Highly stable over weeks; spectrum overlays with Ubl + IDR   |   | SI3                    |
| UBI + IDR           | 2–3   | NMR assigned | Highly stable for >2 weeks at 25°C   | 50446 <sup>d</sup>                                |                        |
| Macrodomain         | 9   | NMR assigned | Highly stable for >1 week at 25°C and > 2 weeks at 4°C   | 50387 <sup>d</sup><br>50388 <sup>d</sup>          |                        |
| SUD-N               | 14  | NMR assigned | Highly stable for >10 days at 25°C   | 50448 <sup>d</sup>                                |                        |
| SUD-NM              | 17  | HSQC         | Stable for >1 week at 25°C   |   |                        |
| SUD-M               | 8.5   | NMR assigned | Significant precipitation during measurement; tendency to dimerize   | 50516 <sup>d</sup>                                |                        |
| 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 <sup>d</sup>                                |                        |
| PL <sup>pro</sup>   | 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 <sup>d</sup>                                |                        |
| CoV-Y               | 12  | HSQC         | Low temperature (<25°C) and low concentrations (<0.2 mM) favor stability; gradual degradation at 25°C; lithium bromide in final buffer supports solubility |   |                        |
| <b>nsp5</b>         |   |              |  |   |                        |
| fl                  | 55  | HSQC         | Impaired dimerization induced by artificial N-terminal residues  |   | SI4                    |
| <b>nsp7</b>         |   |              |  |   |                        |
| fl                  | 17  | NMR assigned | Stable for several days at 35°C; stable for >1 month at 4°C  | 50337 <sup>d</sup>                                | SI5                    |
| <b>nsp8</b>         |   |              |  |   |                        |
| fl                  | 17  | HSQC         | Concentration dependent aggregation; low concentrations favor stability  |   | SI6                    |
| <b>nsp9</b>         |   |              |  |   |                        |
| fl                  | 4.5   | NMR assigned | Stable dimer for >4 months at 4°C and >2 weeks at 25°C   | 50621 <sup>d</sup><br>50622 <sup>d</sup><br>50513 | SI7                    |
| <b>nsp10</b>        |   |              |  |   |                        |
| fl                  | 15  | NMR assigned | Zn <sup>2+</sup> addition during expression and purification increases protein stability; stable for >1 week at 25°C                                       | 50392   | SI8                    |
| <b>nsp13</b>        |   |              |  |   |                        |
| fl                  | 0.5   | HSQC         | Low expression; protein unstable; concentration above 20 µM not possible   |   | SI9                    |
| <b>nsp14</b>        |   |              |  |   |                        |
| fl                  | 6   | Pure protein | Not above 50 µM; best storage: with 50% (v/v) glycerol; addition of reducing agents  |   | SI10                   |
| MTase               | 10  | Pure protein | As fl nsp14; high salt (>0.4 M) for increased stability; addition of reducing agents   |   |                        |
| <b>nsp15</b>        |   |              |  |   |                        |
| fl                  | 5   | HSQC         | Tendency to aggregate at 25°C  |   | SI11                   |
| <b>nsp16</b>        |   |              |  |   |                        |
| fl                  | 10  | Pure protein | Addition of reducing agents; 5% (v/v) glycerol favorable; highly unstable  |   | SI12                   |
| <b>ORF3a</b>        |   |              |  |   |                        |
| fl                  | 0.6   | Pure protein | Addition of detergent during expression (0.05% Brij-58); stable protein  |   | SI13                   |
| <b>E protein</b>    |   |              |  |   |                        |
| fl                  | 0.45  | Pure protein | Addition of detergent during expression (0.05% Brij-58); stable protein  |   | SI14                   |

(Continued on following page)

**TABLE 2** | (Continued) Summary of SCoV2 protein production results in *Covid19-NMR*.

| Construct expressed | Yields (mg/L) <sup>a</sup><br>or (mg/ml) <sup>b</sup> | Results      | Comments  | BMRB                                       | Supplementary Material |
|---------------------|---|--------------|---|--|------------------------|
| <b>M Protein</b>    |   |              |   |  | SI15                   |
| fl                  | 0.33  | Pure protein | Addition of detergent during expression (0.05% Brij-58); stable protein   |  |                        |
| <b>ORF6</b>         |   |              |   |  | SI16                   |
| fl                  | 0.27  | HSQC         | Soluble expression without detergent; stable protein; no expression with STREP-tag at N-terminus                  |  |                        |
| <b>ORF7a</b>        |   |              |   |  | SI17                   |
| ED                  | 0.4   | HSQC         | Unpurified protein tends to precipitate during refolding, purified protein stable for 4 days at 25°C              |  |                        |
| <b>ORF7b</b>        |   |              |   |  | 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  |  |                        |
| <b>ORF8</b>         |   |              |   |  | SI19                   |
| fl                  | 0.62  | HSQC         | Tendency to oligomerize   |  |                        |
| ΔORF8               | 0.5   | Pure protein |   |  |                        |
| <b>N protein</b>    |   |              |   |  | SI20                   |
| IDR1-NTD-<br>IDR2   | 12  | NMR assigned | High salt (>0.4 M) for increased stability  | 50618, 50619,<br>50558, 50557 <sup>d</sup> |                        |
| NTD-SR              | 3   | HSQC         |   |  |                        |
| NTD                 | 3   | HSQC         |   | 34511                                      |                        |
| CTD                 | 2   | NMR assigned | Stable dimer for >4 months at 4°C and >3 weeks at 30°C  | 50518 <sup>d</sup>                         |                        |
| <b>ORF9b</b>        |   |              |   |  | SI21                   |
| fl                  | 0.64  | HSQC         | Expression without detergent, protein is stable   |  |                        |
| <b>ORF14</b>        |   |              |   |  | SI22                   |
| fl                  | 0.43  | HSQC         | Addition of detergent during expression (0.05% Brij-58); stable in detergent but unstable on lipid reconstitution |  |                        |
| <b>ORF10</b>        |   |              |   |  | SI23                   |
| fl                  | 2   | HSQC         | Tendency to oligomerize; unstable upon tag cleavage   |  |                        |

<sup>a</sup>Yields from bacterial expression represent the minimal protein amount in mg/L independent of the cultivation medium. *Italic values indicate yields from CFPS.*

<sup>b</sup>Yields from CFPS represent the minimal protein amount in mg/ml of wheat-germ extract.

<sup>c</sup>COVID19-nmr BMRB depositions yet to be released.

<sup>d</sup>COVID19-nmr BMRB depositions.

et al., 2020b). Besides yields compatible with structural studies, ribosomes in WG extracts further possess an increased folding capacity (Netzer and Hartl, 1997), 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 (NAB) 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 <sup>15</sup>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.

## MATERIALS AND METHODS

### Strains, Plasmids, and Cloning

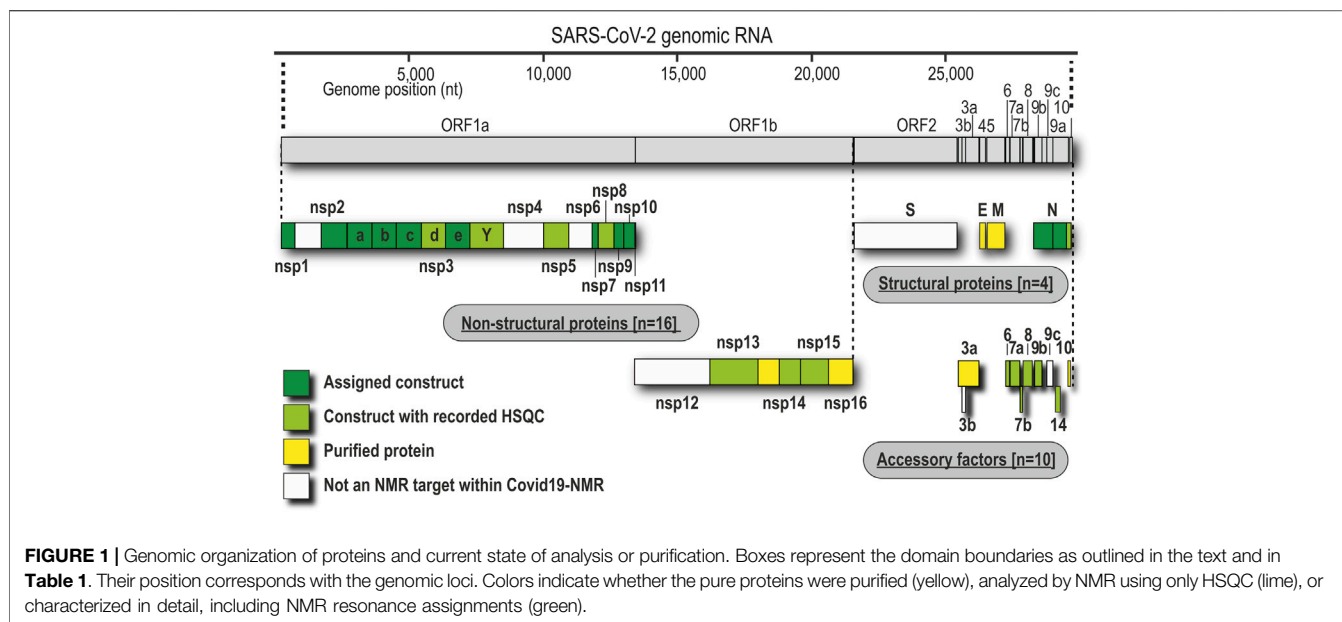
The rationale of construct design for all proteins can be found within the respective protocols in **Supplementary Tables SI1–SI23**. For bacterial production, *E. coli* strains and expression plasmids are given; for WG-CFPS, template

vectors are listed. Protein coding sequences of interest have been obtained as either commercial, codon-optimized genes or, for shorter ORFs and additional sequences, annealed from oligonucleotides prior to insertion into the relevant vector. Subcloning of inserts, adjustment of boundaries, and mutations of genes have been carried out by standard molecular biology techniques. All expression plasmids can be obtained upon request from the *COVID19-NMR* consortium (<https://covid19-nmr.com/>), including 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 (fl) proteins, separate domains, combinations, or particular expression constructs as listed in **Table 1** can be found in the (**Supplementary Tables SI1–SI23**).

The ORF3a, ORF6, ORF7b, ORF8, ORF9b, and ORF14 accessory proteins and the structural proteins M and E were produced by WG-CFPS as described in the **Supplementary Material**. In brief, transcription and translation steps have



been performed separately, and detergent has been added for the synthesis of membrane proteins as described previously (Takai et al., 2010; Fogeron et al., 2017).

## 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.

## RESULTS

In the following, we provide protocols for the purification of SCoV2 proteins sorted into 1) nonstructural proteins and 2) 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 (Wu et al., 2020), 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 IMAC, TEV/Ulp1 Protease cleavage, Reverse IMAC, and Size-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 **Supplementary Tables S11–S123** (see also **Supplementary Tables S1, S2**) and **Tables 1, 2**.

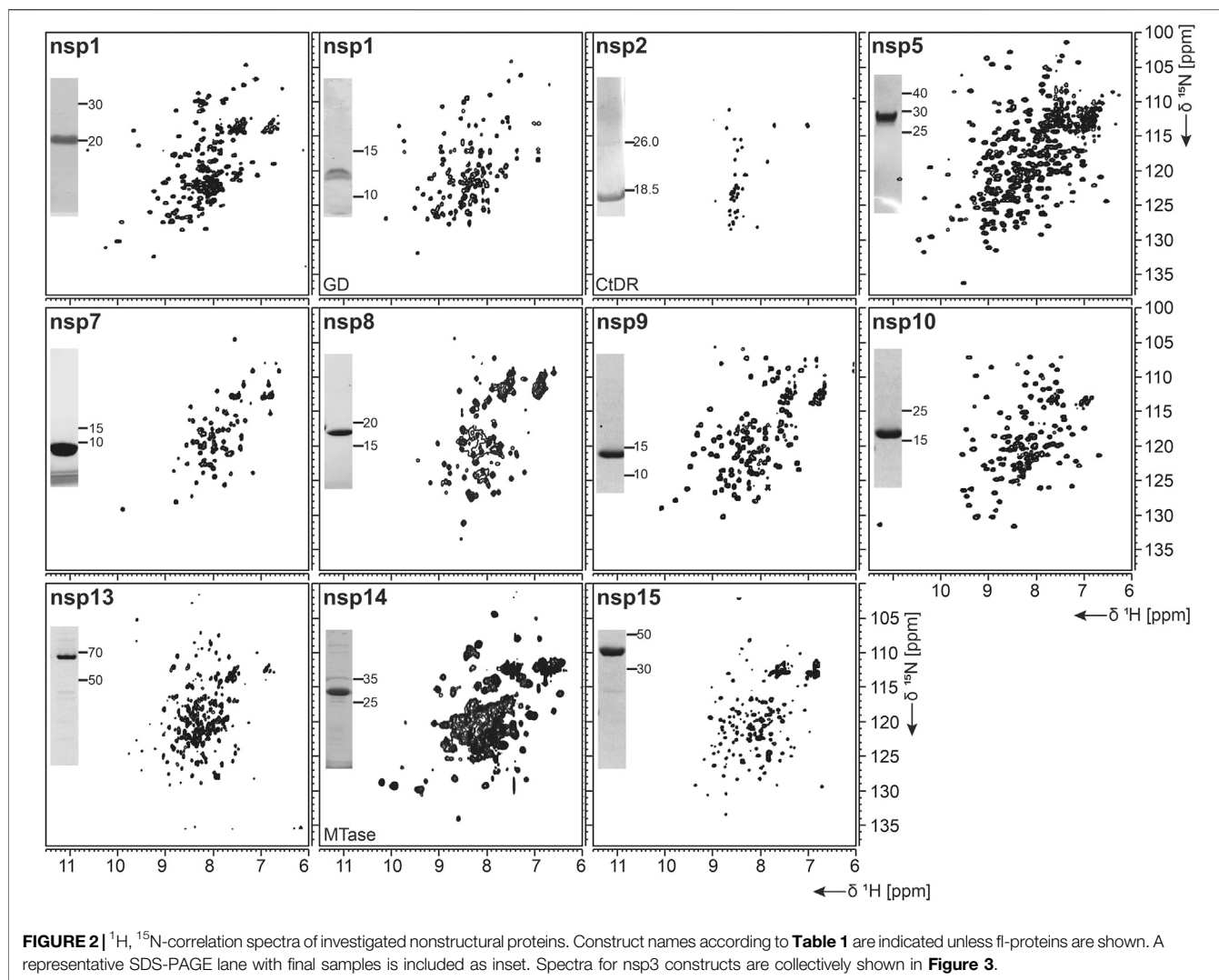
## Nonstructural Proteins

We have approached and challenged the recombinant production of a large part of the SCoV2 nsps (**Figure 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 (Oostra et al., 2007; Oostra et al., 2008). 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 a 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 (Gao et al., 2020; Hillen et al., 2020). For the remainder of nsps, we here provide protocols for 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  $\alpha$ - and  $\beta$ -CoVs (Narayanan et al., 2015). Interestingly, nsp1 displays the highest divergence in sequence and size among different CoVs, justifying it as a genus-specific marker (Snijder et al., 2003). It functions as a host shutoff factor by suppressing innate immune functions and host gene expression (Kamitani et al.,





2006; Narayanan et al., 2008; Schubert et al., 2020). This suppression is achieved by an interaction of the nsp1 C-terminus with the mRNA entry tunnel within the 40 S subunit of the ribosome (Schubert et al., 2020; Thoms et al., 2020).

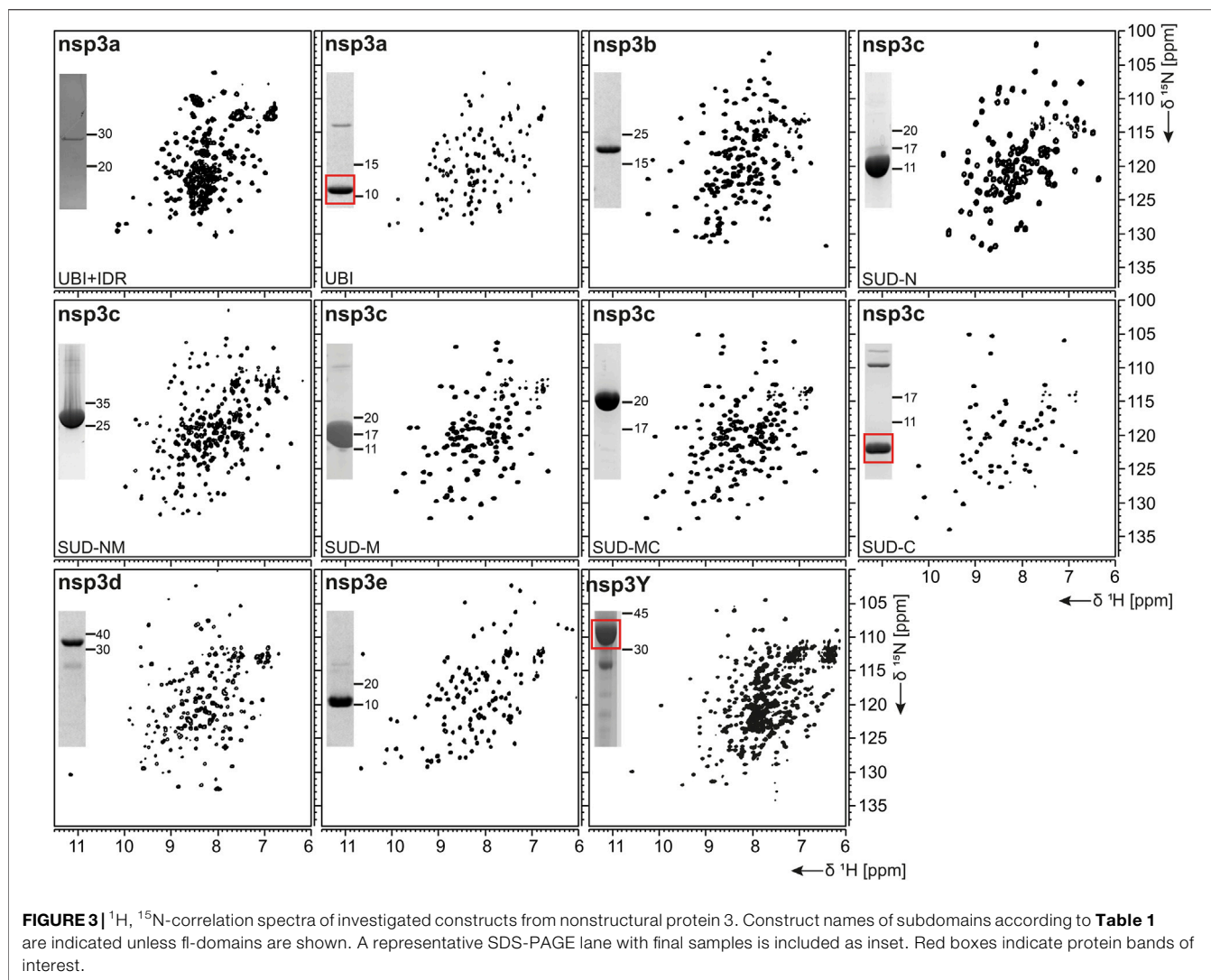
As summarized in **Table 1**, fl-domain boundaries of nsp1 were chosen to contain the first 180 amino acids, in analogy to its closest homolog from SCoV (Snijder et al., 2003). In addition, a shorter construct was designed, encoding only the globular core domain (GD, aa 13–127) suggested by the published SCoV nsp1 NMR structure (Almeida et al., 2007). His-tagged fl nsp1 was purified using the IPRS approach. Protein quality was confirmed by the available HSQC spectrum (**Figure 2**). Despite the flexible C-terminus, we were able to accomplish a near-complete backbone assignment (Wang et al., 2021).

Interestingly, the nsp1 GD was found to be problematic in our hands despite good expression. We observed insolubility, although buffers were used according to the

homolog SCoV nsp1 GD (Almeida et al., 2007). Nevertheless, using a protocol comparable to the one for fl nsp1, we were able to record an HSQC spectrum proving a folded protein (**Figure 2**).

### nsp2

nsp2 has been suggested to interact with host factors involved in intracellular signaling (Cornillez-Ty et al., 2009; Davies et al., 2020). 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 the regulation of global RNA synthesis (Graham et al., 2005). 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 (Ishida and Kinoshita, 2007)]. The His-Trx-tagged peptide was purified by IPRS. Upon dialysis, two IEC steps were performed: first anionic and then cationic, with good final yields (**Table 1**). Stability and



purity were confirmed by an HSQC spectrum (**Figure 2**) and a complete backbone assignment (Mompean et al., 2020; **Table 2**).

### nsp3

nsp3, the largest nsp (Snijder et al., 2003), is composed of a plethora of functionally related, yet independent, subunits. After cleavage of nsp3 from the fl ORF1-encoded polypeptide chain, it displays a 1945-residue multidomain protein, with individual functional entities that are subclassified from nsp3a to nsp3e followed by the ectodomain embedded in two TM 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 (Wolff et al., 2020). Remarkably, the nsp3c substructure comprises three subdomains, 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 a ubiquitin-like (Ubl) structured domain and a subsequent acidic IDR. Besides its ability to bind ssRNA (Serrano et al., 2007), nsp3a has been reported to interact with the nucleocapsid (Hurst et al., 2013; Khan et al., 2020), playing a potential role in virus replication. We here provide protocols for the purification of both the Ubl (aa 1–111) and fl nsp3a (aa 1–206), including the acidic IDR (Ubl + IDR **Table 1**). Domain boundaries were defined similar to the published NMR structure of SCoV nsp3a (Serrano et al., 2007). His-

tagged nsp3a Ubl + IDR and GST-tagged nsp3a Ubl were each purified via the IPRS approach. nsp3a Ubl yielded mM sample concentrations and displayed a well-dispersed HSQC spectrum (**Figure 3**). Notably, the herein described protocol also enables purification of fl nsp3a (Ubl + IDR) (**Tables 1, 2**). Despite the unstructured IDR overhang, the excellent protein quality and stability allowed for near-complete backbone assignment [**Figure 3**, (Salvi et al., 2021)].

### *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 (Frick et al., 2020). 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 available 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 (Cantini et al., 2020).

### *nsp3c*

The SARS unique domain (SUD) of nsp3c has been described as a distinguishing feature of SCoVs (Snijder et al., 2003). However, similar domains in more distant CoVs, such as MHV or MERS, have been reported recently (Chen et al., 2015; Kusov et al., 2015). 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 (Tan et al., 2009), while SUD-C preferentially binds to purine-containing RNA (Johnson et al., 2010). Domain boundaries for SUD-N and SUD-M and for the tandem-domain SUD-NM were defined in analogy to the SCoV homolog crystal structure (Tan et al., 2009). Those for SUD-C and the tandem SUD-MC were based on NMR solution structures of corresponding SCoV homologs (**Table 1**) (Johnson et al., 2010). 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 size-exclusion chromatography (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 (Gallo et al., 2020) and good resolved HSQC spectra also for the tandem domains (**Figure 3**).

### *nsp3d*

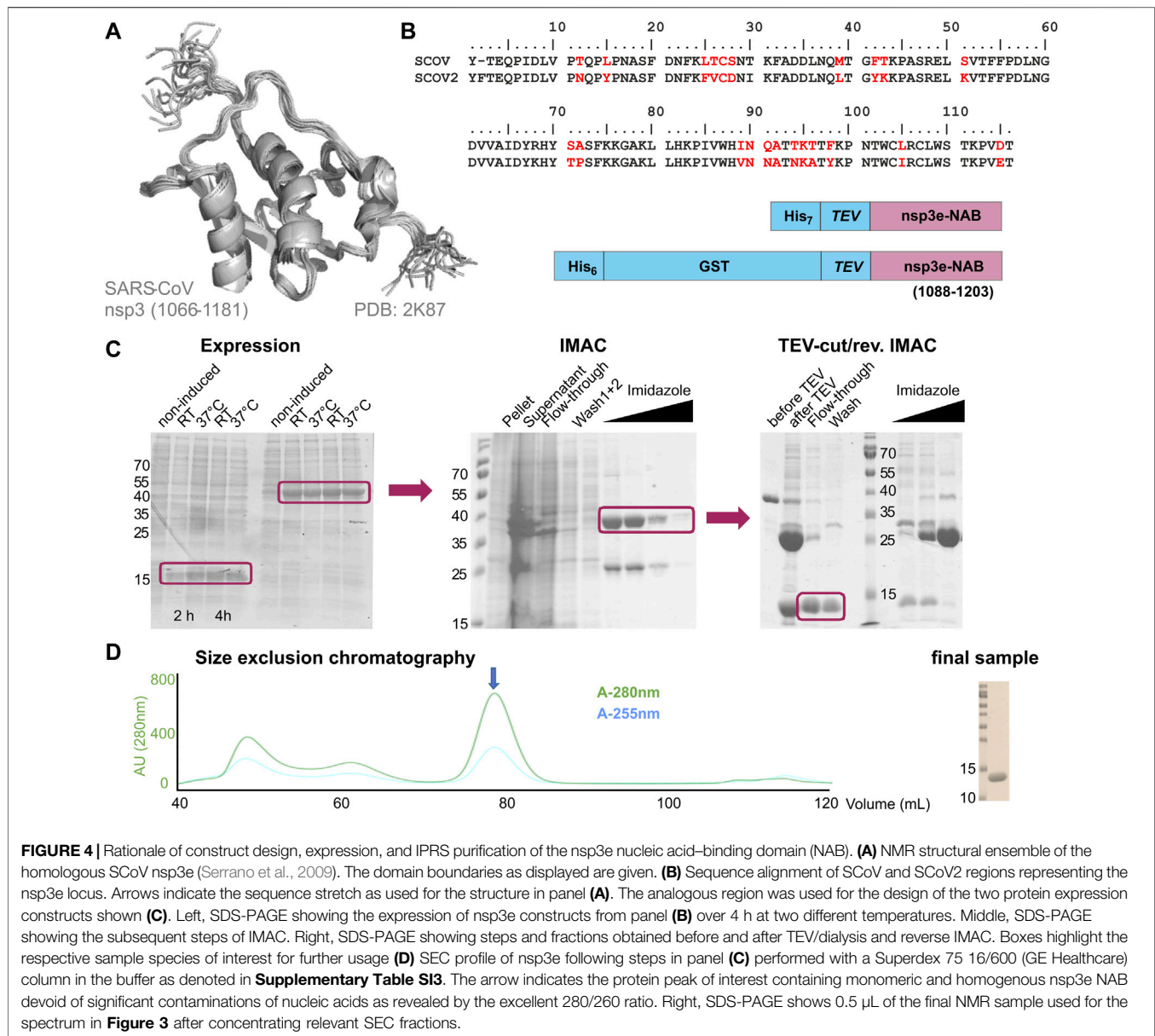
nsp3d comprises the papain-like protease (PL<sup>Pro</sup>) domain of nsp3 and, hence, is one of the two SCoV2 proteases that are responsible for processing the viral polypeptide chain and generating functional proteins (Shin et al., 2020). The domain boundaries of PL<sup>Pro</sup> within nsp3 are set by residues 743 and 1,060 (**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<sup>Pro</sup>. 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<sup>Pro</sup>. The His-SUMO-tag allowed simple IMAC purification, followed by cleavage with Ulp1 and isolation of cleaved PL<sup>Pro</sup> via a second IMAC. A final purification step using gel filtration led to pure PL<sup>Pro</sup> of both unlabeled and 15N-labeled species (**Table 2**). The latter has allowed for the acquisition of a promising amide correlation spectrum (**Figure 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) (Neuman et al., 2008). Structural information is rare; while the G2M is predicted to be intrinsically disordered (Lei et al., 2018); the only available experimental structure of the nsp3e NAB was solved from SCoV by the Wüthrich lab using solution NMR (Serrano et al., 2009). We here used this structure for a sequence-based alignment to derive reasonable domain boundaries for the SCoV2 nsp3e NAB (**Figures 4A,B**). The high sequence similarity suggested using nsp3 residues 1,088–1,203 (**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 (**Figure 4C**). Finally, a homogenous NAB species, as supported by the final gel of pooled samples (**Figure 4D**), was obtained. The excellent protein quality and stability are supported by the available HSQC (**Figure 3**) and a published backbone assignment (Korn et al., 2020a).

### *nsp3Y*

nsp3Y is the most C-terminal domain of nsp3 and exists in all coronaviruses (Neuman et al., 2008; Neuman, 2016). 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, it might affect binding to nsp4 (Hagemeyer et al., 2014). We were able to produce and purify nsp3Y (CoV-Y) comprising amino acids 1,638–1,945 (**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 (**Figure 3**). nsp3 CoV-Y appears suitable for an NMR backbone



**FIGURE 4** | Rationale of construct design, expression, and IPRES purification of the nsp3e nucleic acid-binding domain (NAB). **(A)** NMR structural ensemble of the homologous SCoV nsp3e (Serrano et al., 2009). 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 the design of the two protein expression constructs shown **(C)**. Left, SDS-PAGE showing the expression of nsp3e constructs from panel **(B)** over 4 h 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 **Supplementary Table S13**. 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-PAGE shows 0.5  $\mu$ L of the final NMR sample used for the spectrum in **Figure 3** after concentrating relevant SEC fractions.

assignment carried out at lower concentrations in a deuterated background (ongoing).

### nsp5

The functional main protease nsp5 ( $M^{Pro}$ ) is a dimeric cysteine protease (Ullrich and Nitsche, 2020). Amino acid sequence and 3D structure of SCoV [PDB 1P9U (Anand et al., 2003)] and SCoV2 (PDB 6Y2E [Zhang et al., 2020]) homologs are highly conserved (**Figures 5A,B**). The dimer interface involves the N-termini of both monomers, which puts considerable constraints on the choice of protein sequence for construct design regarding the N-terminus.

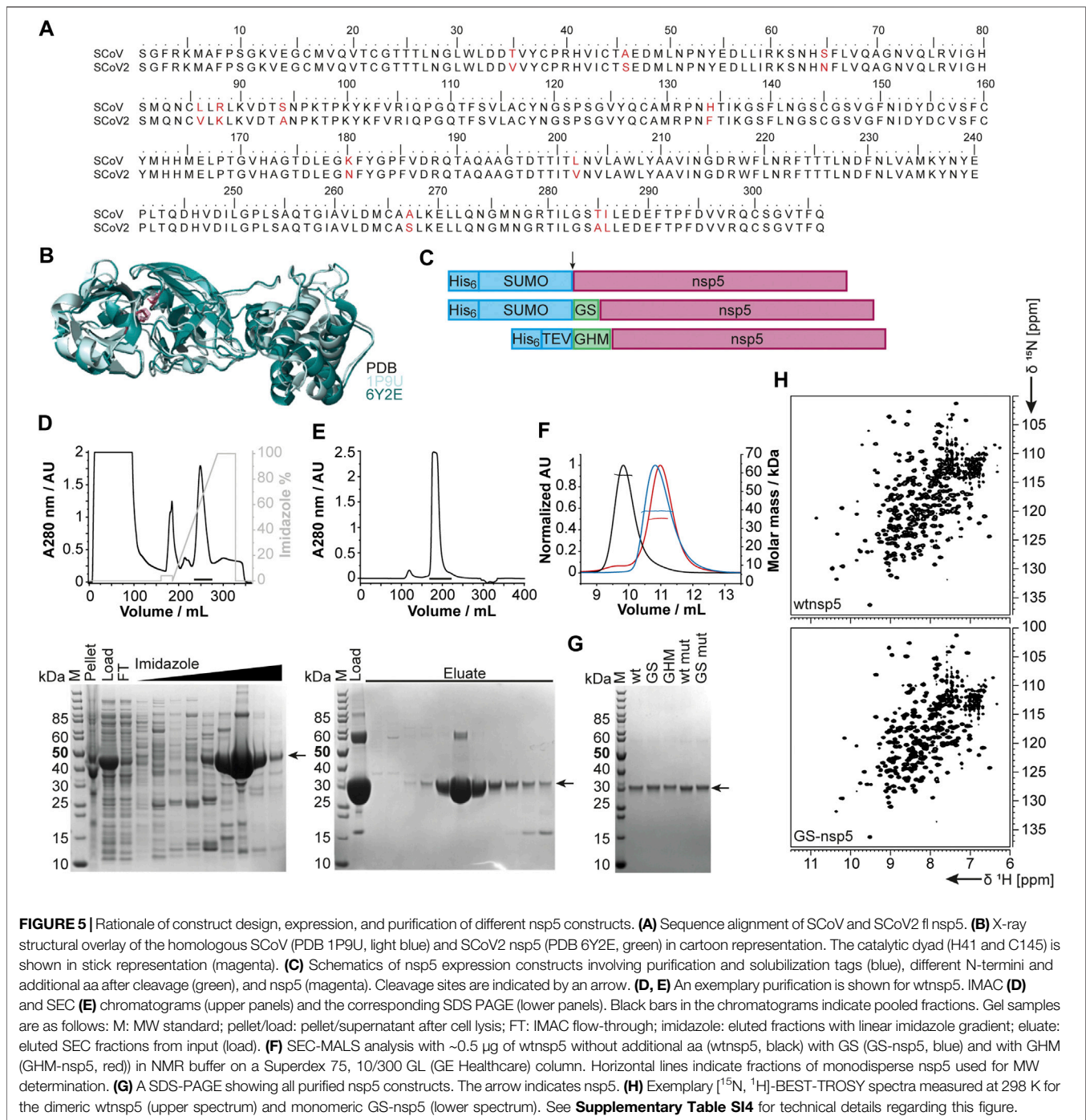
We thus designed different constructs differing in the N-terminus: the native N-terminus (wt), a GS mutant with the additional N-terminal residues glycine and serine as His-SUMO

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 (**Figure 5C**). Purification of all proteins via the IPRES approach (**Figures 5D,E**) yielded homogenous and highly pure protein, analyzed by PAGE (**Figure 5G**), mass spectrometry, and 2D [ $^{15}N$ ,  $^1H$ ]-BEST TROSY spectra (**Figure 5H**). Final yields are summarized in **Table 2**.

### nsp7 and 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) (Gordon et al., 2020). For nsp7 in complex with nsp8 or for nsp8 alone, additional functions in RNA synthesis priming have been proposed (Tvarogova et al., 2019;





Konkolova et al., 2020). In a recent study including an RNA-substrate-bound structure (Hillen et al., 2020), 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 (Kirchdoerfer and Ward, 2019) was successfully transferred, which resulted in decent yields of

reasonably stable proteins (**Table 2**). Driven by its intrinsically oligomeric state, nsp8 showed some tendency toward aggregation, limiting the available sample concentration. The higher apparent molecular weight and limited solubility are also reflected in the success of NMR experiments. While we succeeded in a complete NMR backbone assignment of nsp7 (Tonelli et al., 2020), the quality of the spectra obtained for nsp8 is currently limited to the HSQC presented in **Figure 2**.



### nsp9

The 12.4 kDa ssRNA-binding nsp9 is highly conserved among *Betacoronaviruses*. It is a crucial part of the viral replication machinery (Miknis et al., 2009), possibly targeting the 3'-end stem-loop II (s2m) of the genome (Robertson et al., 2005). nsp9 adopts a fold similar to oligonucleotide/oligosaccharide-binding proteins (Egloff et al., 2004), and structural data consistently uncovered nsp9 to be dimeric in solution (Egloff et al., 2004; Sutton et al., 2004; Miknis et al., 2009; Littler et al., 2020). Dimer formation seems to be a prerequisite for viral replication (Miknis et al., 2009) and influences RNA-binding (Sutton et al., 2004), despite a moderate affinity for RNA *in vitro* (Littler et al., 2020).

Based on the early available crystal structure of SCoV2 nsp9 (PDB 6W4B, unpublished), we used the 113 aa fl 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 in both natural abundance and <sup>13</sup>C- and <sup>15</sup>N-labeled form. Purification *via* the IPRS approach enabled us to separate fl nsp9 in different oligomer states. The earliest eluted fraction represented higher oligomers, 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 are supported by the available HSQC (Figure 2), and a near-complete backbone assignment (Dudas et al., 2021).

### 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. However, it is required for the MTase activity of nsp16 (Krafcikova et al., 2020), it confers exonuclease activity to nsp14 in the RNA polymerase complex in SCoV (Ma et al., 2015). It contains two unusual zinc finger motifs (Joseph et al., 2006) 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 (Joseph et al., 2006). Importantly, additional Zn<sup>2+</sup> ions present during expression and purification stabilize the protein significantly (Kubatova et al., 2020). 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 (Kubatova et al., 2020).

### nsp13

nsp13 is a conserved ATP-dependent helicase that has been characterized as part of the RNA synthesis machinery by binding to nsp12 (Chen et al., 2020b). 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 (Jia et al., 2019) or whether it is required for proofreading and backtracking (Chen

et al., 2020b). 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 the solution (Table 2).

### nsp14

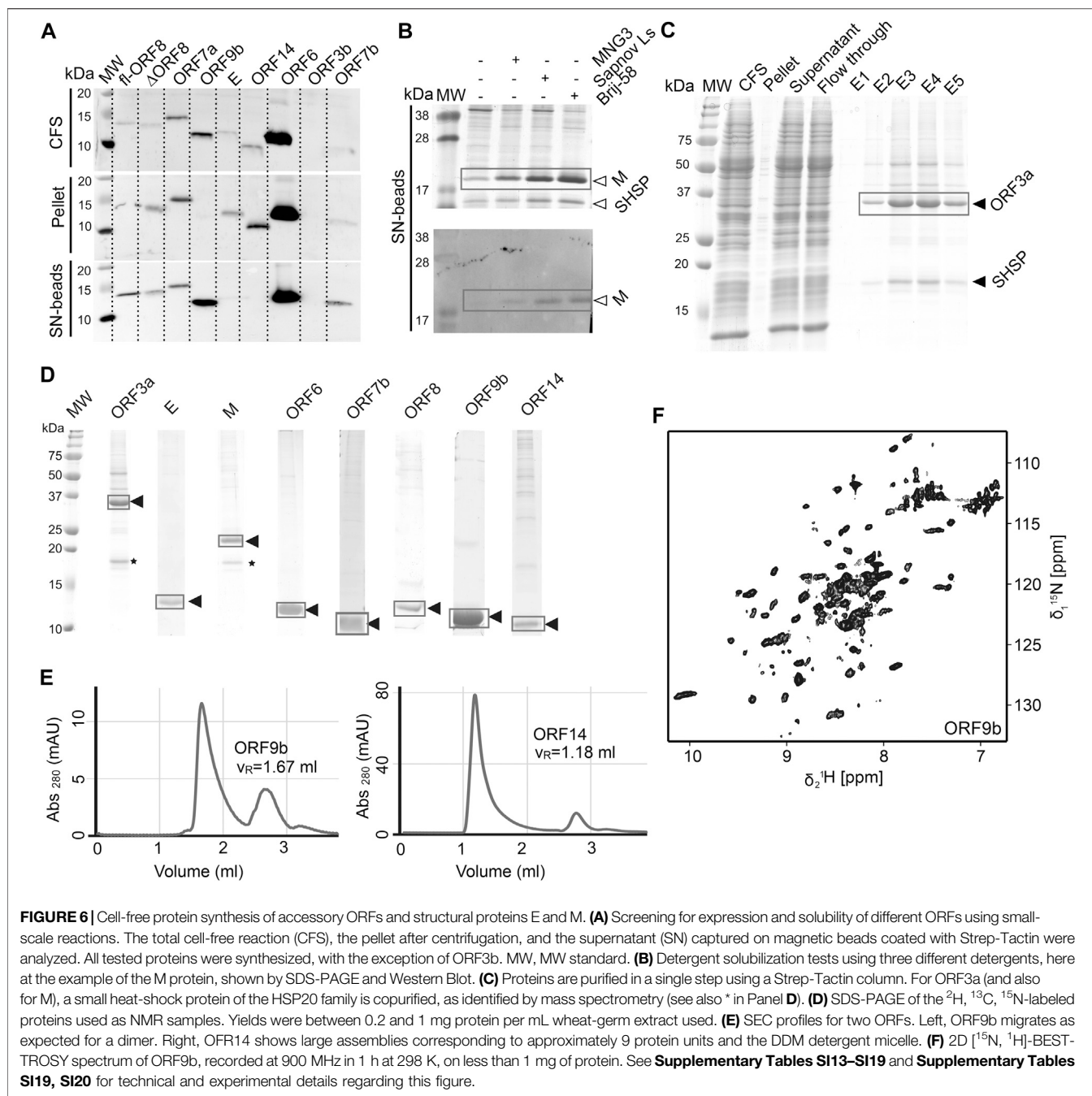
nsp14 contains two domains: an N-terminal exonuclease domain and a C-terminal MTase domain (Ma et al., 2015). The exonuclease domain interacts with nsp10 and provides part of the proofreading function that supports the high fidelity of the RNA polymerase complex (Robson et al., 2020). Several unusual features, such as the unusual zinc finger motifs, set it apart from other DEDD-type exonucleases (Chen et al., 2007), 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 (Thoms et al., 2020). 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 coexpression 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 (Figure 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, (Kim et al., 2020)]. Its function has been suggested to be related to the removal of U-rich RNA elements, preventing recognition by the innate immune system (Deng et al., 2017), 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 (Kim et al., 2020). 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 (Krafcikova et al., 2020). 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 MTase inhibitors, a drug targeting both enzymes would be highly desirable (Bouvet et al., 2010). nsp16 is the last protein being encoded by ORF1ab, and only its N-terminus is formed by cleavage by the M<sup>PRO</sup> nsp5. Employing a similar strategy to that for nsp14, nsp16 constructs were designed with the possibility of nsp10 coexpression. Expression of fl nsp16 resulted in good yields, when expressed both 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 (Rosas-Lemus et al., 2020).

### 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 (Figures 6A–D).

### ORF3a

The protein from ORF3a in SCoV2 corresponds to the 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 (Kern et al., 2020). The structure of SCoV2 3a displays a dimer, but it can also form higher oligomers. Each monomer has three TM helices and a cytosolic  $\beta$ -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 (Liu et al., 2014), 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 the presence of Brij-58 (Figure 6C). It is copurified with a small heat-shock protein of the HSP20 family from the wheat-germ extract. The protocol described here is highly similar to that of the other cell-free synthesized accessory proteins. Where NMR spectra have been reported, the protein has been produced in a  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  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 of 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 ORF (57 aa) with no homology to existing SCoV proteins (Chan et al., 2020). Indeed, ORF3b gene products of SCoV2 and SCoV 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 (Konno et al., 2020). However, the SCoV2 nucleotide sequence after the stop codon shows a high similarity to the SCoV ORF3b. Different C-terminal truncations seem to play a role in the interferon-antagonistic activity of ORF3b (Konno et al., 2020). ORF3b is the only protein that, using WG-CFPS, was not synthesized 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 pathogenicity, as recently reviewed in (Schoeman and Fielding, 2020). Structural models for SCoV (Surya et al.,

2018) and the TM helix of SCoV2 (Mandala et al., 2020) E have been established. The structural models show a pentamer with a TM 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 the addition of detergent to the cell-free reaction.

### ORF5 (Membrane Glycoprotein, M)

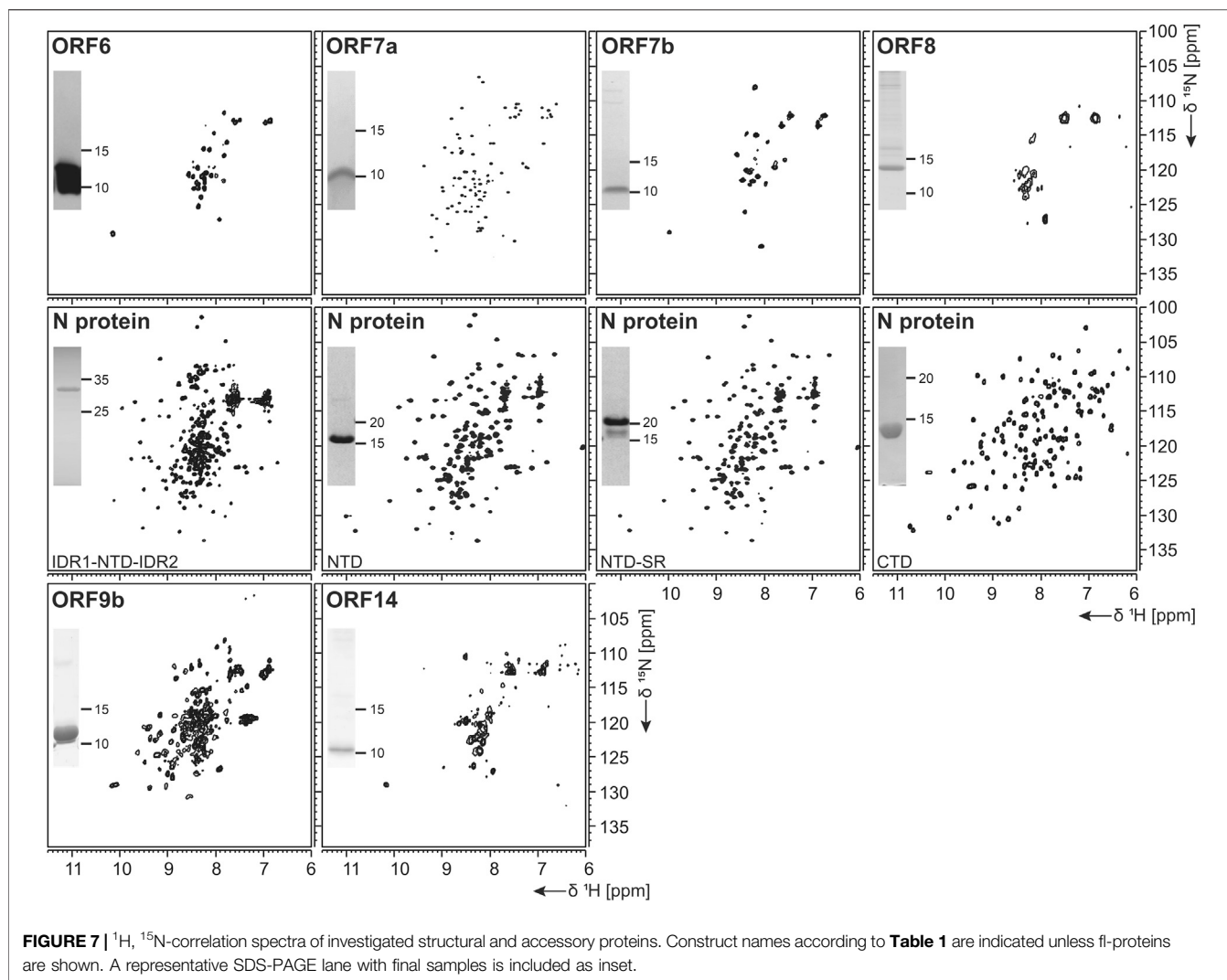
The M protein is the most abundant protein in the viral envelope and is believed to be responsible for maintaining the virion in its characteristic shape (Huang et al., 2004). M is a glycoprotein and sequence analyses predict three domains: A C-terminal endodomain, a TM domain with three predicted helices, 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 (Vennema et al., 1996), 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 (Neuman et al., 2011). 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 (Neuman et al., 2011). The M protein is readily produced using cell-free synthesis in the presence of detergent; as ORF3a, it is copurified with a small heat-shock protein of the HSP20 family (Figure 6B). 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 (Huang et al., 2004). 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 (Zhao et al., 2009), and results indicate that ORF6 may serve an important role in the pathogenesis during SCoV infection (Liu et al., 2014). Also, it showed to inhibit the expression of certain STAT1-genes critical for the host immune response and could contribute to the immune evasion. ORF6 is expressed very well in WG-CFPS; the protein was fully soluble with detergents and partially soluble without them and was easily purified in the presence of detergent, but less efficiently in the 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 (Liu et al., 2014). 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 the ORF7a ectodomain (ED) with a GB1 tag (Bogomolovas et al., 2009) was expected to produce reasonable yields. The IPRS purification resulted in a highly stable protein, as evidenced by the NMR data obtained (Figure 7).

### ORF7b

Protein ORF7b is associated with viral particles in a SARS context (Liu et al., 2014). Protein 7b is one of the shortest ORFs with 43 residues. It shows a long hydrophobic stretch, which might correspond to a TM segment. It shows over 93% sequence homology with a bat coronavirus 7b protein (Liu et al., 2014). 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 the 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 the first promising HSQC, as shown in Figure 7.

### ORF8

ORF 8 is believed to be responsible for the evolution of *Betacoronaviruses* and their species jumps (Wu et al., 2016) and to have a role in repressing the host response (Tan et al., 2020). 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 homologs only show limited identity, with the exception of a small 7 aa segment, where, in SCoV, the glutamate is replaced with an aspartate. It, however, aligns very well with several coronaviruses endemic to animals, including Paguma and Bat (Chan et al., 2020). 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 (Flower et al., 2020). In this study, ORF8 has been used both with (fl) and without signal peptide ( $\Delta$ ORF8). We first tested the 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 forming either oligomers or aggregates.

### ORF9a (Nucleocapsid Protein, N)

The nucleocapsid protein (N) is important for viral genome packaging (Luo et al., 2006). The multifunctional RNA-binding protein plays a crucial role in the viral life cycle (Chang et al., 2014) 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) (Kang et al., 2020).

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.

### N-Terminal Domain

The NTD is the RNA-binding domain of the nucleocapsid (Kang et al., 2020). It is embedded within IDRs, functions of which have not yet been deciphered. Recent experimental and bioinformatic data indicate involvement in liquid-liquid phase separation (Chen et al., 2020a).

For the NTD, several constructs were designed, also considering the flanking IDRs (Table 1). In analogy to the available NMR [PDB 6YI3, (Dinesh et al., 2020)] and crystal [PDB 6M3M, (Kang et al., 2020)] 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) construct 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 IPRS and yielded approx. 3 mg/L in  $^{15}\text{N}$ -labeled minimal medium. High protein quality and stability are supported by the available HSQC spectra (Figure 7).

The untagged IDR1-NTD-IDR2 was purified by IEC and yielded high amounts of  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled samples of 12 mg/L for further NMR investigations. The quality of our purification is confirmed by the available HSQC (Figure 7), and a near-complete backbone assignment of the two IDRs was achieved (Guseva et al., 2021; Schiavina et al., 2021). 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.

### C-Terminal Domain

Multiple studies on the SCoV2 CTD, including recent crystal structures (Ye et al., 2020; Zhou et al., 2020), confirm the domain as dimeric. Its ability to self-associate seems to be necessary for viral replication and transcription (Luo et al., 2006). In addition, the CTD was shown to, presumably nonspecifically, bind ssRNA (Zhou et al., 2020).

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, [Takeda et al., 2008]). 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 are supported by the available HSQC spectrum (Figure 7) and a near-complete backbone assignment (Korn et al., 2020b).

### ORF9b

Protein 9b (97 aa) shows 73% sequence homology to the SCoV and also to bat virus (bat-SL-CoVZXC21) 9b protein (Chan et al., 2020). 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  $\beta$ -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 (Figure 6E) using WG-CFPS in its soluble form. Spectra show a well-folded protein, and assignments are underway (Figure 6F).

### 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 wheat-germ cell-free system in the presence of detergent and solution NMR spectra have been recorded, they hint at an aggregated protein (Figure 6E). 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 (Yoshimoto, 2020). 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 an uncleaved fusion protein. Analytical SEC 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 (Hillen et al., 2020), nsp5 is undergoing a broad structure-based screen against a battery of inhibitors in multiple places (Jin et al., 2020; Zhang et al., 2020), but with, as of yet, the limited outcome for effective medication. Hence, a comprehensive, reliable treatment of COVID-19 at any stage after the infection has remained unsuccessful.

Further viral protein targets will have to be taken into account in order to provide inhibitors with increased specificity and efficacy and preparative starting points for following potential generations of (SARS-)CoVs. Availability of those proteins in a 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 the production of large amounts of pure proteins.

The COVID19-NMR consortium was launched with the motivation of providing NMR assignments of all SCoV2 proteins and RNA elements, and enormous progress has been made since the outbreak of COVID-19 for both components [see Table 2 and (Wacker et al., 2020)]. 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 (Li and Kang, 2020). Relevant to a broad scientific community, we here report our protocols to suite perfectly 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 full 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 (Figure 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<sup>Pro</sup> nsp3d and nsp3Y. Particularly for the nsp3 multidomain protein, we here present soluble samples of

almost the complete cytosolic region with more than 120 kDa in the form of excellent 2D NMR spectra (Figure 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 us 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 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 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 homologs (Table 1), has confronted us with very different levels of complexity. With some prior expectation toward 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 are due to their limited sequence conservation, no prior information, large molecular weights, insolubility, and so forth.

The nsp3e NAB represents one example where the available NMR structure of the SCoV homolog provided a *bona fide* template for selecting initial domain boundaries (Figure 4). 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 homologs 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 (Zust et al., 2007; Narayanan et al., 2015; Shen et al., 2019; Thoms et al., 2020).

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 the structure and biochemistry of the protein (Cross et al., 2020). In the present study, the expression of the different constructs allowed us to characterize the protein in both its monomeric and 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 (**Figure 5H**). The addition of two residues (GS) interferes with the dimerization interface, despite being similar to its native N-terminal amino acids (SGFR). We also introduced an active site mutation that replaces cysteine 145 with alanine (Hsu et al., 2005). Intriguingly, this active site mutation C145A, known to stabilize the dimerization of the main protease (Chang et al., 2007), 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 (**Supplementary Table S14**). The NMR results are in line with SEC-MALS analyses (**Figure 5F**). Indeed, the additional amino acids at the N-terminus shift the dimerization equilibrium toward the monomer, whereas the mutation shifts it toward 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 in the inhibitor search against SCoV2 proteins, as exemplified by the particularly attractive nsp5 main protease target.

## Exploiting Nonbacterial 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 us previously to express a variety of difficult viral proteins in our hands (Fogeron et al., 2015a; Fogeron et al., 2015b; Fogeron et al., 2016; Fogeron et al., 2017; Wang et al., 2019; Jirasko et al., 2020a). 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 the production of 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 (Morita et al., 2003), which allows bypassing extensive purification steps. In fact, 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 (Konno et al., 2020).

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 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 nontrivial 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 (e.g., ORF9b). As expected, however, most proteins cannot be straightforwardly 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 (Fogeron et al., 2015a; Fogeron et al., 2015b; Fogeron et al., 2016; Jirasko et al., 2020a; Jirasko et al., 2020b). 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 straightforward deuteration in WG-CFPS (David et al., 2018; Wang et al., 2019; Jirasko et al., 2020a) that circumvents proton back-exchange, rendering denaturation and refolding steps obsolete (Tonelli et al., 2011). Nevertheless, the herein presented protocols for the production of non-nsp5 by WG-CFPS instantly enable their employment in binding studies and screening campaigns and thus provide 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.

## DATA AVAILABILITY STATEMENT

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 are available as plasmids from <https://covid19-nmr.de/>.

## AUTHOR CONTRIBUTIONS

NA, SK, NQ, MD, MN, ABö, HS, MH, and AS designed the study, compiled the protocols and NMR data, and wrote the manuscript. All authors contributed coordinative or practical work to the study. All authors contributed to the creation and collection of protein protocols and NMR spectra.

## FUNDING

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), the state of Hesse (BMRZ), the Fondazione CR Firenze (CERM), and 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. AS is supported by DFG Grant SCHL 2062/2-1 and by the JQYA at Goethe through project number 2019/AS01. Work in the lab of KV was supported by a CoRE grant from the University of New Hampshire. The FLI is a 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 RM was supported by NIH (2R01EY021514) and NSF (DMR-2002837). BN-B was supported by the NSF GRFP. MC was supported by NIH (R25 GM055246 MBRS IMSD), and MS-P was supported by the HHMI Gilliam Fellowship. Work in the labs of KJ and KT was supported by Latvian Council of Science Grant No. VPP-COVID 2020/1-0014. 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 cofinanced 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 CM-G lab was supported by the Helmholtz society. Work in the lab of ABö 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 BM 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 ABö and BM 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)

## REFERENCES

- Almeida, M. S., Johnson, M. A., Herrmann, T., Geralt, M., and Wüthrich, 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 (7), 3151–3161. doi:10.1128/JVI.01939-06
- 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 (5626), 1763–1767. doi:10.1126/science.1085658
- 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 (1), 16–23. doi:10.1016/j.pep.2008.10.003

RA-COVID SARS2NUCLEOPROTEIN and European Research Council Advanced Grant DynamicAssemblies. Work in the CA lab was supported by Patto per il Sud della Regione Siciliana–CheMIST grant (CUP G77B17000110001). 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). Work at the UW-Madison was supported by grant numbers NSF MCB2031269 and NIH/NIAID AI123498. MM is a Ramón y Cajal Fellow of the Spanish AEI-Ministry of Science and Innovation (RYC2019-026574-I), and a “La Caixa” Foundation (ID 100010434) Junior Leader Fellow (LCR/BQ/PR19/11700003). Funded by project COV20/00764 from the Carlos III Institute of Health and the Spanish Ministry of Science and Innovation to MM and DVL. VDJ was supported by the Boehringer Ingelheim Fonds. Part of this work used the resources of the Italian Center of Instruct-ERIC at the CERM/CIRMMP infrastructure, supported by the Italian Ministry for University and Research (FOE funding). CF was supported by the Stiftung Polytechnische Gesellschaft. Work in the lab of JH was supported by NSF (RAPID 2030601) and NIH (R01GM123249).

## ACKNOWLEDGMENTS

The authors thank Leonardo Gonnelli and Katharina Targaczewski for the valuable technical assistance. IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG CEA). They acknowledge the Advanced Technologies Network Center of the University of Palermo to support infrastructures.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.653148/full#supplementary-material>

- Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., et al. (2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* 583 (7816), 469–472. doi:10.1038/s41586-020-2332-7
- Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E. J., Canard, B., et al. (2010). *In vitro* reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog.* 6 (4), e1000863. doi:10.1371/journal.ppat.1000863
- Cantini, F., Banci, L., Altincekic, N., Bains, J. K., Dhamotharan, K., Fuks, C., et al. (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 (2), 339–346. doi:10.1007/s12104-020-09973-4
- Chan, J. F., Kok, K. H., Zhu, Z., Chu, H., To, K. K., Yuan, S., et al. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* 9 (1), 221–236. doi:10.1080/22221751.2020.1719902

- 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. *Antivir. Res.* 103, 39–50. doi:10.1016/j.antiviral.2013.12.009
- 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 (4), 1374–1383. doi:10.1529/biophysj.106.091736
- Chen, H., Cui, Y., Han, X., Hu, W., Sun, M., Zhang, Y., et al. (2020a). Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res.* 30, 1143. doi:10.1038/s41422-020-00408-2
- Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P. M. M., Olinares, P. D. B., et al. (2020b). Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* 182 (6), 1560–1573. doi:10.1016/j.cell.2020.07.033
- 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 (5), 649–655. doi:10.5483/bmbrep.2007.40.5.649
- 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 (42), 25293–25306. doi:10.1074/jbc.M115.662130
- 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 (19), 10314–10318. doi:10.1128/JVI.00842-09
- Cross, T. J., Takahashi, G. R., Diessner, E. M., Crosby, M. G., Farahmand, V., Zhuang, S., et al. (2020). Sequence characterization and molecular modeling of clinically relevant variants of the SARS-CoV-2 main protease. *Biochemistry* 59 (39), 3741–3756. doi:10.1021/acs.biochem.0c00462
- David, G., Fogeron, M. L., Schledorn, M., Montserret, R., Haselmann, U., Penzel, S., et al. (2018). Structural studies of self-assembled subviral particles: combining cell-free expression with 110 kHz MAS NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* 57 (17), 4787–4791. doi:10.1002/anie.201712091
- 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* [Epub ahead of print]. doi:10.1101/2020.07.13.201517
- Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., et al. (2017). Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc. Natl. Acad. Sci. U.S.A.* 114 (21), E4251–E4260. doi:10.1073/pnas.1618310114
- Dinesh, D. C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., et al. (2020). Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.* 16 (12), e1009100. doi:10.1371/journal.ppat.1009100
- Dudas, F. D., Puglisi, R., Korn, S. M., Alfano, C., Kelly, G., Monaca, E., et al. (2021). Backbone chemical shift spectral assignments of coronavirus-2 non-structural protein nsp9. *Biomol. NMR Assign.* 2021, 1–10. doi:10.1007/s12104-020-09992-1
- Egloff, M. P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., et al. (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 (11), 3792–3796. doi:10.1073/pnas.0307877101
- Esposito, D., Mehalko, J., Drew, M., Snead, K., Wall, V., Taylor, T., et al. (2020). Optimizing high-yield production of SARS-CoV-2 soluble spike trimers for serology assays. *Protein Expr. Purif.* 174, 105686. doi:10.1016/j.pep.2020.105686
- Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., et al. (2020). The coding capacity of SARS-CoV-2. *Nature* 589, 125. doi:10.1038/s41586-020-2739-1
- 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* [Epub ahead of print]. doi:10.1101/2020.08.27.270637
- Fogeron, M. L., Badillo, A., Jirasko, V., Gouttenoire, J., Paul, D., Lancien, L., et al. (2015a). 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. doi:10.1016/j.pep.2014.10.003
- Fogeron, M. L., Badillo, A., Penin, F., and Böckmann, A. (2017). Wheat germ cell-free overexpression for the production of membrane proteins. *Methods Mol. Biol.* 1635, 91–108. doi:10.1007/978-1-4939-7151-0\_5
- Fogeron, M. L., Jirasko, V., Penzel, S., Paul, D., Montserret, R., Danis, C., et al. (2016). Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus. *J. Biomol. NMR* 65 (2), 87–98. doi:10.1007/s10858-016-0040-2
- Fogeron, M. L., Paul, D., Jirasko, V., Montserret, R., Lacabanne, D., Molle, J., et al. (2015b). Functional expression, purification, characterization, and membrane reconstitution of non-structural protein 2 from hepatitis C virus. *Protein Expr. Purif.* 116, 1–6. doi:10.1016/j.pep.2015.08.027
- Frick, D. N., Viridi, 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 (28), 2608–2615. doi:10.1021/acs.biochem.0c00309
- Gallo, A., Tsika, A. C., Fourkiotis, N. K., Cantini, F., Banci, L., Sreeramulu, S., et al. (2020). <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>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.* 2020, 1–5. doi:10.1007/s12104-020-09987-y
- Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., et al. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 368 (6492), 779–782. doi:10.1126/science.abb7498
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583 (7816), 459–468. doi:10.1038/s41586-020-2286-9
- 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 (21), 13399–13411. doi:10.1128/JVI.79.21.13399-13411.2005
- Guseva, S., Perez, L. M., Camacho-Zarco, A., Bessa, L. M., Salvi, N., Malki, A., et al. (2021). <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N Backbone chemical shift assignments of the n-terminal and central intrinsically disordered domains of SARS-CoV-2 nucleoprotein. *Biomol NMR Assign.* doi:10.1007/s12104-021-10014-x
- Hagemeyer, M. C., Monastyrska, I., Griffith, J., van der Sluijs, P., Voortman, J., van Bergen en Henegouwen, P. M., et al. (2014). Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology* 458–459, 125–135. doi:10.1016/j.virol.2014.04.027
- Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. *Nature* 584 (7819), 154–156. doi:10.1038/s41586-020-2368-8
- Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P., et al. (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* 280 (35), 31257–31266. doi:10.1074/jbc.M502577200
- 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 (22), 12557–12565. doi:10.1128/JVI.78.22.12557-12565.2004
- 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 (16), 9159–9172. doi:10.1128/JVI.01275-13
- Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464. doi:10.1093/nar/gkm363
- Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., et al. (2019). Delicate structural coordination of the severe acute respiratory syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res.* 47 (12), 6538–6550. doi:10.1093/nar/gkz409
- Jiang, H. W., Li, Y., Zhang, H. N., Wang, W., Yang, X., Qi, H., et al. (2020). SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. *Nat. Commun.* 11 (1), 3581. doi:10.1038/s41467-020-17488-8
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582 (7811), 289–293. doi:10.1038/s41586-020-2223-y
- Jirasko, V., Lakomek, N. A., Penzel, S., Fogeron, M. L., Bartenschlager, R., Meier, B. H., et al. (2020a). Proton-detected solid-state NMR of the cell-free synthesized



- $\alpha$ -helical transmembrane protein NS4B from hepatitis C virus. *Chembiochem* 21 (10), 1453–1460. doi:10.1002/cbic.201900765
- Jirasko, V., Lends, A., Lakomek, N. A., Fogeron, M. L., Weber, M. E., Malär, A. A., et al. (2020b). Dimer organization of membrane-associated NS5A of hepatitis C virus as determined by highly sensitive 1 H-detected solid-state NMR. *Angew. Chem. Int. Ed.* 60 (10), 5339–5347. doi:10.1002/anie.202013296
- Johnson, M. A., Chatterjee, A., Neuman, B. W., and Wüthrich, K. (2010). SARS coronavirus unique domain: three-domain molecular architecture in solution and RNA binding. *J. Mol. Biol.* 400 (4), 724–742. doi:10.1016/j.jmb.2010.05.027
- Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Brooun, A., Griffith, M., et al. (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 (16), 7894–7901. doi:10.1128/JVI.00467-06
- Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., et al. (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 (34), 12885–12890. doi:10.1073/pnas.0603144103
- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., et al. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* 10 (7), 1228–1238. doi:10.1016/j.apsb.2020.04.009
- Kern, D. M., Sorum, B., Mali, S. S., Hoel, C. M., Sridharan, S., Remis, J. P., et al. (2020). Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *bioRxiv* 17, 156554. doi:10.1101/2020.06.17.156554
- Khan, M. T., Zeb, M. T., Ahsan, H., Ahmed, A., Ali, A., Akhtar, K., et al. (2020). SARS-CoV-2 nucleocapsid and Nsp3 binding: an in silico study. *Arch. Microbiol.* 203, 59. doi:10.1007/s00203-020-01998-6
- Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., et al. (2020). Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci.* 29 (7), 1596–1605. doi:10.1002/pro.3873
- 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 (1), 2342. doi:10.1038/s41467-019-10280-3
- Konkolova, E., Klima, M., Nencka, R., and Boura, E. (2020). Structural analysis of the putative SARS-CoV-2 primase complex. *J. Struct. Biol.* 211 (2), 107548. doi:10.1016/j.jsb.2020.107548
- Konno, Y., Kimura, I., Uriu, K., Fukushi, M., Irie, T., Koyanagi, Y., et al. (2020). SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is increased by a naturally occurring elongation variant. *Cell Rep.* 32 (12), 108185. doi:10.1016/j.celrep.2020.108185
- Korn, S. M., Dhamotharan, K., Fürtig, B., Hengesbach, M., Löhr, F., Qureshi, N. S., et al. (2020a). 1H, 13C, and 15N backbone chemical shift assignments of the nucleic acid-binding domain of SARS-CoV-2 non-structural protein 3e. *Biomol. NMR Assign.* 14 (2), 329–333. doi:10.1007/s12104-020-09971-6
- Korn, S. M., Lambert, R., Fürtig, B., Hengesbach, M., Löhr, F., Richter, C., et al. (2020b). 1H, 13C, and 15N backbone chemical shift assignments of the C-terminal dimerization domain of SARS-CoV-2 nucleocapsid protein. *Biomol. NMR Assign.* 2020, 1–7. doi:10.1007/s12104-020-09995-y
- Krafčikova, 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 (1), 3717. doi:10.1038/s41467-020-17495-9
- Kubatova, N., Qureshi, N. S., Altincekic, N., Abele, R., Bains, J. K., Ceylan, B., et al. (2020). 1H, 13C, and 15N backbone chemical shift assignments of coronavirus-2 non-structural protein Nsp10. *Biomol. NMR Assign.* 2020, 1–7. doi:10.1007/s12104-020-09984-1
- 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. doi:10.1016/j.virol.2015.06.016
- Leao, J. C., Gusmao, T. P. L., Zazar, A. M., Leao Filho, J. C., Barkokebas Santos de Faria, A., Morais Silva, I. H., et al. (2020). Coronaviridae-old friends, new enemy! *Oral Dis.* 2020, 13447. doi:10.1111/odi.13447
- Lei, J., Kusov, Y., and Hilgenfeld, R. (2018). Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antivir. Res.* 149, 58–74. doi:10.1016/j.antiviral.2017.11.001
- Li, Q., and Kang, C. (2020). A practical perspective on the roles of solution NMR spectroscopy in drug discovery. *Molecules* 25 (13), 2974. doi:10.3390/molecules25132974
- 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 (7), 101258. doi:10.1016/j.isci.2020.101258
- Liu, D. X., Fung, T. S., Chong, K. K., Shukla, A., and Hilgenfeld, R. (2014). Accessory proteins of SARS-CoV and other coronaviruses. *Antivir. Res.* 109, 97–109. doi:10.1016/j.antiviral.2014.06.013
- 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 (39), 11827–11835. doi:10.1021/bi0609319
- Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., et al. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc. Natl. Acad. Sci. U.S.A.* 112 (30), 9436–9441. doi:10.1073/pnas.1508686112
- 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.* 27, 1202. doi:10.1038/s41594-020-00536-8
- 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 (7), 3007–3018. doi:10.1128/JVI.01505-08
- 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* [Epub ahead of print]. doi:10.1101/2020.11.09.374173
- 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 (6), 1216–1221. doi:10.1110/ps.0241203
- Narayanan, K., Huang, C., Lokugamage, K., Kamitani, W., Ikegami, T., Tseng, C. T., et al. (2008). Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. *J. Virol.* 82 (9), 4471–4479. doi:10.1128/JVI.02472-07
- 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. doi:10.1016/j.virusres.2014.11.019
- Nelson, C. W., Ardern, Z., Goldberg, T. L., Meng, C., Kuo, C. H., Ludwig, C., et al. (2020). Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic. *Elife* 9, 59633. doi:10.7554/eLife.59633
- Netzer, W. J., and Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388 (6640), 343–349. doi:10.1038/41024
- Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A., et al. (2008). Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J. Virol.* 82 (11), 5279–5294. doi:10.1128/JVI.02631-07
- Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., et al. (2011). A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* 174 (1), 11–22. doi:10.1016/j.jsb.2010.11.021
- Neuman, B. W. (2016). Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles. *Antivir. Res.* 135, 97–107. doi:10.1016/j.antiviral.2016.10.005
- Oostra, M., Hagemeijer, M. C., van Gent, M., Bekker, C. P., te Lintelo, E. G., Rottier, P. J., et al. (2008). Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J. Virol.* 82 (24), 12392–12405. doi:10.1128/JVI.01219-08
- 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 (22), 12323–12336. doi:10.1128/JVI.01506-07
- Pavesi, A. (2020). New insights into the evolutionary features of viral overlapping genes by discriminant analysis. *Virology* 546, 51–66. doi:10.1016/j.virol.2020.03.007



- 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 (1), e5. doi:10.1371/journal.pbio.0030005
- Robson, F., Khan, K. S., Le, T. K., Paris, C., Demirbag, S., Barfuss, P., et al. (2020). Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol. Cell* 79 (5), 710–727. doi:10.1016/j.molcel.2020.07.027
- Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N. L., Kiryukhina, O., Wiersum, G., et al. (2020). The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.04.17.047498
- Salvi, N., Bessa, L. M., Guseva, S., Camacho-Zarco, A., Maurin, D., Perez, L. M., et al. (2021). <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N backbone chemical shift assignments of SARS-CoV-2 nsp3a. *Biomol. NMR Assign.* 2021, 1–4. doi:10.1007/s12104-020-10001-8
- 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. doi:10.3389/fmicb.2020.02086
- Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L. A., et al. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat. Struct. Mol. Biol.* 27 (10), 959–966. doi:10.1038/s41594-020-0511-8
- Serrano, P., Johnson, M. A., Almeida, M. S., Horst, R., Herrmann, T., Joseph, J. S., et al. (2007). Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. *J. Virol.* 81 (21), 12049–12060. doi:10.1128/JVI.00969-07
- Serrano, P., Johnson, M. A., Chatterjee, A., Neuman, B. W., Joseph, J. S., Buchmeier, M. J., et al. (2009). Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. *J. Virol.* 83 (24), 12998–13008. doi:10.1128/JVI.01253-09
- Schiavina, M., Pontoriero, L., Uversky, V. N., Felli, I. C., and Pierattelli, R. (2021). The highly flexible disordered regions of the SARS-CoV-2 nucleocapsid N protein within the 1-248 residue construct: Sequence-specific resonance assignments through NMR. *Biomol NMR Assign.* [in press].
- Shen, Z., Wang, G., Yang, Y., Shi, J., Fang, L., Li, F., et al. (2019). A conserved region of nonstructural protein 1 from alphacoronaviruses inhibits host gene expression and is critical for viral virulence. *J. Biol. Chem.* 294 (37), 13606–13618. doi:10.1074/jbc.RA119.009713
- Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., et al. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* 587 (7835), 657–662. doi:10.1038/s41586-020-2601-5
- Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., et al. (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 (5), 991–1004. doi:10.1016/s0022-2836(03)00865-9
- Surya, W., Li, Y., and Torres, J. (2018). Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim. Biophys. Acta Biomembr.* 1860 (6), 1309–1317. doi:10.1016/j.bbame.2018.02.017
- Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., et al. (2004). The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure* 12 (2), 341–353. doi:10.1016/j.str.2004.01.016
- Takai, K., Sawasaki, T., and Endo, Y. (2010). Practical cell-free protein synthesis system using purified wheat embryos. *Nat. Protoc.* 5 (2), 227–238. doi:10.1038/nprot.2009.207
- Takeda, M., Chang, C. K., Ikeya, T., Güntert, P., Chang, Y. H., Hsu, Y. L., et al. (2008). Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J. Mol. Biol.* 380 (4), 608–622. doi:10.1016/j.jmb.2007.11.093
- Tan, J., Vonnrhein, C., Smart, O. S., Bricogne, G., Bollati, M., Kusov, Y., et al. (2009). The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *Plos Pathog.* 5 (5), e1000428. doi:10.1371/journal.ppat.1000428
- 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 (3). doi:10.1128/mBio.00760-20
- Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., et al. (2020). Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* 369 (6508), 1249–1255. doi:10.1126/science.abc8665
- Tonelli, M., Singarapu, K. K., Makino, S., Sahu, S. C., Matsubara, Y., Endo, Y., et al. (2011). Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J. Biomol. NMR* 51 (4), 467–476. doi:10.1007/s10858-011-9575-4
- Tonelli, M., Rienstra, C., Anderson, T. K., Kirchdoerfer, R., and Henzler-Wildman, K. (2020). <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N backbone and side chain chemical shift assignments of the SARS-CoV-2 non-structural protein 7. *Biomol. NMR Assign.* 2020, 1–5. doi:10.1007/s12104-020-09985-0
- Tvarogová, 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 adenyllyltransferase activity. *J. Virol.* 93 (12), e00291–e00319. doi:10.1128/JVI.00291-19
- Ullrich, S., and Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* 30 (17), 127377. doi:10.1016/j.bmcl.2020.127377
- Vennema, H., Godeke, G. J., Rossen, J. W., Voorhout, W. F., Horzinek, M. C., Opstelten, D. J., et al. (1996). Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* 15 (8), 2020–2028. doi:10.1002/j.1460-2075.1996.tb00553.x
- Wacker, A., Weigand, J. E., Akabayov, S. R., Altincekic, N., Bains, J. K., Banijamali, E., et al. (2020). Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* 48, 12415. doi:10.1093/nar/gkaa1013
- Wang, Y., Kirkpatrick, J., Zur Lage, S., Korn, S. M., Neissner, K., Schwalbe, H., et al. (2021). (<sup>1</sup>H), (<sup>13</sup>C), and (<sup>15</sup>N) backbone chemical-shift assignments of SARS-CoV-2 non-structural protein 1 (leader protein). *Biomol NMR Assign.* doi:10.1007/s12104-021-10019-6
- Wang, S., Fogeron, M. L., Schledorn, M., Dujardin, M., Penzel, S., Burdette, D., et al. (2019). Combining cell-free protein synthesis and NMR into a tool to study capsid assembly modulation. *Front. Mol. Biosci.* 6, 67. doi:10.3389/fmolb.2019.00067
- Wolff, G., Limpens, R. W. A. L., Zevenhoven-Dobbe, J. C., Laugks, U., Zheng, S., de Jong, A. W. M., et al. (2020). A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* 369 (6509), 1395–1398. doi:10.1126/science.abd3629
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579 (7798), 265–269. doi:10.1038/s41586-020-2008-3
- Wu, Z., Yang, L., Ren, X., Zhang, J., Yang, F., Zhang, S., et al. (2016). ORF8-related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J. Infect. Dis.* 213 (4), 579–583. doi:10.1093/infdis/jiv476
- 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.* 29, 1890. doi:10.1002/pro.3909
- Yin, W., Mao, C., Luan, X., Shen, D. D., Shen, Q., Su, H., et al. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 368 (6498), 1499–1504. doi:10.1126/science.abc1560
- 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 (3), 198–216. doi:10.1007/s10930-020-09901-4
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* 368 (6489), 409–412. doi:10.1126/science.abb3405
- Zhao, J., Falcón, A., Zhou, H., Netland, J., Enjuanes, L., Pérez Breña, P., et al. (2009). Severe acute respiratory syndrome coronavirus protein 6 is required for optimal replication. *J. Virol.* 83 (5), 2368–2373. doi:10.1128/JVI.02371-08
- 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 (2), 1–11. doi:10.1186/s43556-020-00001-4
- Züst, R., Cervantes-Barragán, L., Kuri, T., Blakqori, G., Weber, F., Ludewig, B., et al. (2007). Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines. *PLoS Pathog.* 3 (8), e109. doi:10.1371/journal.ppat.0030109

**Conflict of Interest:** CH was employed by Signals GmbH & Co. KG.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Altincekic, Korn, Qureshi, Dujardin, Ninot-Pedrosa, Abele, Abi Saad, Alfano, Almeida, Alshamleh, de Amorim, Anderson, Anobom, Anorma, Bains, Bax, Blackledge, Blechar, Böckmann, Brigandat, Bula, Bütikofer, Camacho-Zarco, Carlomagno, Caruso, Ceylan, Chaikuad, Chu, Cole, Crosby, de Jesus, Dharmotharan, Felli, Ferner, Fleischmann, Fogeron, Fourkiotis, Fuks, Fürtig, Gallo, Gande, Gerez, Ghosh, Gomes-Neto, Gorbatyuk, Guseva, Hacker, Häfner, Hao, Hargittay, Henzler-Wildman, Hoch, Hohmann, Hutchison, Jaudzems, Jović, Kaderli, Kalniņš, Kaņepe, Kirchoerfer, Kirkpatrick, Knapp, Krishnathas, Kutz, zur

Lage, Lambertz, Lang, Laurents, Lecoq, Linhard, Löhr, Malki, Bessa, Martin, Matzel, Maurin, McNutt, Mebus-Antunes, Meier, Meiser, Mompeán, Monaca, Montserret, Mariño Perez, Moser, Muhle-Goll, Neves-Martins, Ni, Norton-Baker, Pierattelli, Pontoriero, Pustovalova, Ohlenschläger, Orts, Da Poian, Pyper, Richter, Riek, Rienstra, Robertson, Pinheiro, Sabbatella, Salvi, Saxena, Schulte, Schiavina, Schwalbe, Silber, Almeida, Sprague-Piercy, Spyroulias, Sreeramulu, Tants, Tars, Torres, Töws, Treviño, Trucks, Tsika, Varga, Wang, Weber, Weigand, Wiedemann, Wirmer-Bartoschek, Wirtz Martin, Zehnder, Hengesbach and Schlundt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## GLOSSARY

- 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** Hisx-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
- M<sup>pro</sup>** Main protease
- MTase** Methyltransferase
- N** Nucleocapsid protein
- NAB** Nucleic acid-binding domain
- nsp** Nonstructural protein
- NTD** N-terminal domain
- PL<sup>pro</sup>** 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.