Modeling SARS-CoV2 proteins in the CASP -commons experiment

Article

Accepted Version

Protein
DOI10.1002/prot.26231

Modeling SARS-CoV2 proteins in the CASP-commons experiment

Running title: CASP-COVID

Andriy Kryshtafovych [1], Genome Center, University of California, Davis, California 95616, USA; akryshtafovych@ucdavis.edu

John Moult [2], Institute for Bioscience and Biotechnology Research, Department of Cell Biology and Molecular genetics, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850, USA; jmoult@umd.edu

Wendy M. Billings [3], Department of Physics & Astronomy, Brigham Young University, N361 ESC, BYU, Provo, UT 84602; wendybillings7@gmail.com

Dennis Della Corte [3], Department of Physics & Astronomy, Brigham Young University, N361 ESC, BYU, Provo, UT 84602; dennis.dellacorte@byu.edu

Krzysztof Fidelis [1], Genome Center, University of California, Davis, California 95616, USA; kfidelis@ucdavis.edu

Sohee Kwon [4], Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea; sohee95@snu.ac.kr

Kliment Olechnovič [5], Institute of Biotechnology, Life Sciences Center, Vilnius University, Saulėtekio 7, Vilnius, LT 10257, Lithuania; kliment.olechnovic@bti.vu.lt

Chaok Seok [4], Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea; chaok@snu.ac.kr

Česlovas Venclovas [5], Institute of Biotechnology, Life Sciences Center, Vilnius University, Saulėtekio 7, Vilnius, LT 10257, Lithuania; ceslovas.venclovas@bti.vu.lt

Jonghun Won [4], Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea; cozki@snu.ac.kr

Received: May 05, 2021;Revised: Aug 23, 2021;Accepted: Aug 26, 2021

collaborative authors:

AlphaFold team; DeepMind, London, UK EC4A 3TW; alphafold@deepmind.com

Badri Adhikari; University of Missouri-St. Louis; adhikarib@umsl.edu

Recep Adiyaman; School of Biological Sciences, University of Reading, Reading, RG6 6EX, UK; r.adiyaman2@reading.ac.uk

Joaquim Aguirre-Plans; Structural Bioinformatics Lab (GRIB-IMIM), Department of Experimental and Health Science, University Pompeu Fabra, 08005 Barcelona, Catalonia; joaquim.aguirre@upf.edu

Ivan Anishchenko; Department of Biochemistry, University of Washington, Seattle, WA 98195, USA; Institute for Protein Design, University of Washington, Seattle, WA 98195, USA; aivan@uw.edu

Minkyung Baek; Department of Biochemistry, University of Washington, Seattle, WA 98195, USA; Institute for Protein Design, University of Washington, Seattle, WA 98195, USA; minkbaek@uw.edu

David Baker; Department of Biochemistry, University of Washington, Seattle, WA 98195, USA; Institute for Protein Design, University of Washington, Seattle, WA 98195, USA; Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA; dabaker@uw.edu

Frederico Baldassarre; KTH Royal Institute of Technology, 100 44 Stockholm, Sweden; baldassarre.fe@gmail.com

Jacob Barger; University of Missouri-St. Louis; jsbp67@mail.umsl.edu

Sutanu Bhattacharya; Department of Computer Science and Software Engineering, Auburn University, Auburn, AL 36849, USA; szb0134@auburn.edu

Debswapna Bhattacharya; Department of Computer Science and Software Engineering, Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA; bhattacharyad@auburn.edu

Mor Bitton, Department of Computer Science, Ben Gurion University of the Negev, morbitt@post.bgu.ac.il

Renzhi Cao, Department of Computer Science, Pacific Lutheran University, caora@plu.edu,

Jianlin Cheng; Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA; chengji@missouri.edu

Charles Christoffer; Computer Science, Purdue University, West Lafayette, IN, USA; christ35@purdue.edu

Cezary Czaplewski; Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland, cezary.czaplewski@ug.edu.pl

Arne Elofsson; Science for Life Laboratory and Dep of Biochemistry and Biophysics, Stockholm University, 106 91 Stocholm Sweden; arne@bioinfo.se

Eshel Faraggi; Research and Information Systems, LLC, and Physics Department, Indiana University Purdue University Indianapolis, Indianapolis, IN 46202, USA; efaraggi@gmail.com

Michael Feig; Department of Biochemistry and Molecular Biology, Michigan State University, MI 48824, USA; mfeiglab@gmail.com

Narcis Fernandez-Fuentes; IBERS, Aberystwyth University, Aberystwyth SY23 3EE, United Kingdom; naf4@aber.ac.uk

Nick Grishin; HHMI, Department of Biophysics and Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA; Nick.Grishin@utsouthwestern.edu

Sergei Grudinin; Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France; sergei.grudinin@inria.fr

Zhiye Guo; Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA; zggc9@mail.missouri.edu

Yuya Hanazono; Institute for Quantum Life Science, National Institutes for Quantum and Radiological Science and Technology, Tokai, Ibaraki, 319-1106, Japan; hanazono.yuya@qst.go.jp

Demis Hassabis; DeepMind, London, UK; UK EC4A 3TW; dhcontact@google.com

Bryce Hedelius; Department of Physics and Astronomy, Brigham Young University, Provo, UT, 84602; bhedelius@gmail.com

Lim Heo; Department of Biochemistry and Molecular Biology, Michigan State University, MI 48824, USA, huhlim@gmail.com

Naozumi Hiranuma; Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA 98195, USA; Institute for Protein Design, University of Washington, Seattle, WA 98195, USA; hiranumn@uw.edu

Cassandra Hunt, Department of Computer Science, Pacific Lutheran University, cass.j.hunt@plu.edu,

Ilia Igashov; Moscow Institute of Physics and Technology, 141701 Dolgoprudniy, Russia and Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France; igashov.is@phystech.edu

Takashi Ishida; Department of Computer Science, School of Computing, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo, 152-8550, Japan; ishida@c.titech.ac.jp

Robert L. Jernigan; Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa 50011, USA; jernigan@iastate.edu

David Jones; Department of Computer Science, University College London, Gower Street, London WC1E 6BT, United Kingdom; d.t.jones@ucl.ac.uk

John Jumper; DeepMind, London, UK EC4A 3TW; jumper@google.com

Maria Kadukova; Moscow Institute of Physics and Technology, 141701 Dolgoprudniy, Russia and Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France; mn.kadukova@gmail.com

Shaun Kandathil; Department of Computer Science, University College London, Gower Street, London WC1E 6BT, United Kingdom; s.kandathil@ucl.ac.uk

Chen Keasar, Department of Computer Science, Ben Gurion University of the Negev, keasar@bgu.ac.il

Daisuke Kihara; [1] – Department of Biological Sciences, Purdue University, West Lafayette, IN, USA; [2] – Computer Science, Purdue University, West Lafayette, IN, USA; dkihara@purdue.edu

Lisa Kinch; HHMI, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA; lkinch@chop.swmed.edu

Yasuomi Kiyota; School of Pharmacy, Kitasato University, Tokyo 108-8641, Japan; kiyotay@pharm.kitasato-u.ac.jp

Andrzej Kloczkowski, Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205, and Department of Pediatrics, The Ohio State University, Columbus, OH 43205, USA; Andrzej.Kloczkowski@nationwidechildrens.org

Pushmeet Kohli; DeepMind, London, UK EC4A 3TW; pushmeet@google.com

Mateusz Kogut; Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland, mateusz.kogut.mk@gmail.com

Elodie Laine; Sorbonne Université, CNRS, IBPS, Laboratoire de Biologie Computationnelle et Quantitative (LCQB), 75005 Paris, France; elodie.laine@sorbonne-universite.fr

Cade Lilley; Department of Computer Science, Pacific Lutheran University, lilleycr@plu.edu,

Jian Liu; Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA; jl4mc@mail.missouri.edu

Adam Liwo; Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland, adam.liwo@ug.edu.pl

Emilia Lubecka; Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, G. Narutowicza 11/12, 80-233 Gdańsk, Poland, emilubec@eti.pg.edu.pl

Arup Mondal; Department of Chemistry, University of Florida, Gainesville, FL, 32603; arup.mondal@chem.ufl.edu

Connor J. Morris; Department of Physics and Astronomy, Brigham Young University, Provo, UT, 84602; con.morris09@gmail.com

Liam McGuffin; School of Biological Sciences, University of Reading, Reading, RG6 6EX, UK; l.j.mcguffin@reading.ac.uk

Alexis Molina; Electronic and Atomic Protein Modelling Group, Life Sciences Department, Barcelona Supercomputing Center, 08034, Barcelona, Catalonia; alexis.molina@bsc.es

Tsukasa Nakamura; Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, 980-8579, Japan; t.nakamura@sb.ecei.tohoku.ac.jp

Baldo Oliva; Structural Bioinformatics Lab (GRIB-IMIM), Department of Experimental and Health Science, University Pompeu Fabra, 08005 Barcelona, Catalonia; baldo.oliva@upf.edu

Alberto Perez; Department of Chemistry, Quantum Theory Project, University of Florida, Gainesville, FL, 32603; perez@chem.ufl.edu

Gabriele Pozzati; Science for Life Laboratory and Dep of Biochemistry and Biophysics, Stockholm University, 106 91 Stocholm Sweden; gabriele.pozzati@scilifelab.se

Daipayan Sarkar; Department of Biological Sciences, Purdue University, West Lafayette, IN, USA; sarkar30@purdue.edu

Rin Sato; [1-] Department of Computer Science, School of Computing, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo, 152-8550, Japan; [2-] Real World Big-Data Computation Open Innovation Laboratory(RWBC-OIL), National Institute of Advanced Industrial Science and Technology (AIST), Aomi, Koto-ku, Tokyo 135-0064, Japan; sato@cb.cs.titech.ac.jp

Torsten Schwede; [1-] Biozentrum, University of Basel, Basel 4056, Switzerland; [2-] SIB Swiss Institute of Bioinformatics, Basel 4056, Switzerland; torsten.schwede@unibas.ch

Bikash Shrestha; University of Missouri-St. Louis; bsmmy@mail.umsl.edu

Tomer Sidi, Department of Computer Science, Ben Gurion University of the Negev, siditom@post.bgu.ac.il

Gabriel Studer; [1-] Biozentrum, University of Basel, Basel 4056, Switzerland
[2-] SIB Swiss Institute of Bioinformatics, Basel 4056, Switzerland; gabriel.studer@unibas.ch

Md Hossain Shuvo; Department of Computer Science and Software Engineering, Auburn University, Auburn, AL 36849, USA; mzs0149@auburn.edu

Mayuko Takeda-Shitaka; School of Pharmacy, Kitasato University, Tokyo 108-8641, Japan; shitakam@pharm.kitasato-u.ac.jp

Yuma Takei; [1-] Department of Computer Science, School of Computing, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo, 152-8550, Japan; [2-] Real World Big-Data Computation Open Innovation Laboratory(RWBC-OIL), National Institute of Advanced Industrial Science and Technology (AIST), Aomi, Koto-ku, Tokyo 135-0064, Japan; takei@cb.cs.titech.ac.jp

Genki Terashi; Department of Biological Sciences, Purdue University, West Lafayette, IN, USA; gterashi@purdue.edu

Kentaro Tomii; 1 - Artificial Intelligence Research Center (AIRC), National Institute of Advanced Industrial Science and Technology (AIST), Koto-ku, Tokyo 135-0064, Japan; 2 - AIST-Tokyo Tech Real World Big-Data Computation Open Innovation Laboratory (RWBC-OIL), National Institute of Advanced Industrial Science and Technology (AIST), Koto-ku, Tokyo 135-0064, Japan; k-tomii@aist.go.jp

Yuko Tsuchiya; Artificial Intelligence Research Center (AIRC), National Institute of Advanced Industrial Science and Technology (AIST), Koto-ku, Tokyo 135-0064, Japan; yuko.tsuchiya@aist.go.jp

Kathryn Tunyasuvunakool; DeepMind, London, UK EC4A 3TW; ktkool@google.com

Björn Wallner; Department of Physics, Chemistry, and Biology, Linköping University, Sweden; bjorn.wallner@liu.se

Tianqi Wu; Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA; tianqiwu@mail.missouri.edu

Jinbo Xu; Toyota Technological Institute at Chicago, Chicago, IL 60637, USA; jinboxu@gmail.com

Yu Yamamori; Artificial Intelligence Research Center (AIRC), National Institute of Advanced Industrial Science and Technology (AIST), Koto-ku, Tokyo, 135-0064, Japan; yu.yamamori@aist.go.jp

Chengxin Zhang; Department of Computational Medicine and Bioinformatics, University of Michigan, 100 Washtenaw Ave, Ann Arbor, MI 48109; zcx@umich.edu

Yang Zhang; Department of Computational Medicine and Bioinformatics, University of Michigan, 100 Washtenaw Ave, Ann Arbor, MI 48109; zhng@umich.edu

Wei Zheng; Department of Computational Medicine and Bioinformatics, University of Michigan, 100 Washtenaw Ave, Ann Arbor, MI 48109; zhengwei@umich.edu

Correspondence to: Andriy Kryshtafovych**;** Genome Center, University of California, Davis, California 95616, USA; E-mail: akryshtafovych@ucdavis.edu

DATA AVAILABILITY: The data that supports the findings of this study are available in the supplementary material of this article.

Keywords: CASP, SARS-CoV-2, COVID, EMA, protein structure prediction, model accuracy

Abbreviations:

CASP: Critical **A**ssessment of **S**tructure **P**rediction;

SARS-CoV-2: Severe **A**cute **R**espiratory **S**yndrome – **Co**rona**V**irus-**2**;

CASP-commons (CASP-COVID): CASP community-wide experiment on modeling SARS-CoV-2 proteins causing the **co**rona**vi**rus **d**isease;

EMA: estimates of model accuracy

TBM: template-based modeling

FM: free modeling

Abstract

CASP (Critical Assessment of Structure Prediction) is an organization aimed at advancing the state of the art in computing protein structure from sequence. In the spring of 2020, CASP launched a community project to compute the structures of the most structurally challenging proteins coded for in the SARS-CoV2 genome. Forty-seven research groups submitted over 3,000 three-dimensional models and 700 sets of accuracy estimates on ten proteins. The resulting models were released to the public. CASP community members also worked together to provide estimates of local and global accuracy and identify structure-based domain boundaries for some proteins. Subsequently, two of these structures (ORF3a and ORF8) have been solved experimentally, allowing assessment of both model quality and the accuracy estimates. Models from the AlphaFold2 group were found to have good agreement with the experimental structures, with main chain GDT_TS accuracy scores ranging from 63 (a correct topology) to 87 (competitive with experiment).

Introduction.

The advent of the COVID-19 crisis spurred major efforts to combat the disease from biologists all over the world. Key to understanding many aspects of the disease mechanism is knowledge of protein structure. Experimental research groups have devoted major effort to this task, but progress has been necessarily slow and more than 2,300 amino acids in the SARS2 proteins still have no experimental structural coverage. Computed protein structure, while until recently not as accurate as experiment [1-5], can nevertheless provide models that may aid in the choice of drug targets, development of vaccine strategies, and insights into viral mechanisms. Early in the pandemic, a number of leading structure modeling research groups, including SWISSMODEL https://swissmodel.expasy.org/repository/species/2697049; AlphaFold https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19; Baker https://www.ipd.uw.edu/2020/02/rosettas-role-in-fighting-coronavirus; Zhang https://zhanglab.ccmb.med.umich.edu/COVID-19; Feig https://github.com/feiglab/sars-cov-2-proteins; and the Xu group, produced sets of computed structures of SARS-CoV-2 proteins. Because of earlier experimental work on other viruses, particularly SARS, there are homologous structures available for the majority of SARS-CoV-2 proteins, so that useful models can be produced with straightforward template-based methods [6-11]. The CASP initiative engaged the broader modeling community with the aim of producing the best possible structures for the more demanding cases, those without detectable homology to experimentally determined structures, where a community effort was likely to have the most impact. The strategy for this CASP-COVID experiment was to collect models from as many modeling groups as possible and to also solicit community input on evaluating the accuracy of those models, so as to provide the scientific community with the most accurate structures currently possible. The strategy built on three things - the existence of a closely knit CASP modeling community, extensive previous CASP results on the reliability of modeling and accuracy estimation methods [10-17], and the CASP infrastructure [18-22].

The CASP-COVID experiment was started on March 9, 2020. The experiment proceeded through six stages, followed by the discussion of the results at the CASP14 conference in December 2020. The stages were as follows: 1) Selection of targets and their analysis, 2) Call for 3D models, 3) Call for accuracy estimates of the models, 4) Community discussion of the initial results, 5) Call for revised and refined models and accuracy estimates, and 6) Re-release of some targets in CASP14, allowing thorough comparison of models with new experimental data. In addition, there was a post-CASP follow-up to further assess effectiveness of EMA (estimates of model accuracy) methods.

There was a strong community response to the call for CASP-COVID participation, with 47 research groups submitting models using a total of 53 3D modeling approaches and 30 accuracy estimation approaches. All groups who submitted at least five models to CASP-COVID and submitted an abstract to CASP14 Abstract book (or had a documented history of participation in CASP) were invited to contribute their method description to this paper.

Results.

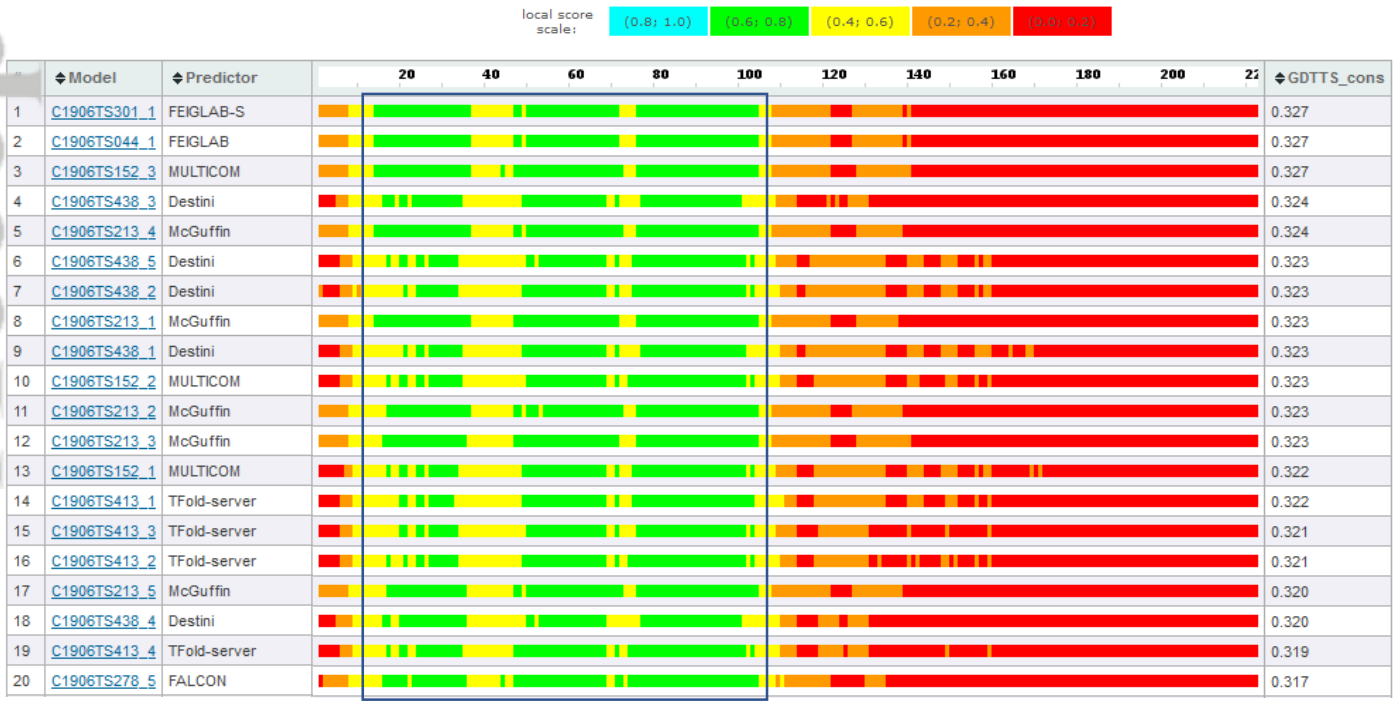
1. Selection of targets and their analysis

The CASP organizers analyzed 29 proteins coded for by the SARS-CoV-2 genome[23] and identified 10 for which part or all of the sequence did not have reliable homologs in the structural database [24]. These were selected as CASP-COVID targets. Supplementary Table S1 shows graphical representations of the HHsearch [25] sequence searches against the structural database for the selected targets. The targets were analyzed to identify the predicted secondary structure and domain composition [26], disorder regions [27], trans-membrane regions [28] and signal peptides [29]. The results of the analysis were posted on the CASP-Commons website https://predictioncenter.org/caspcommons/target_analysis.cgi. Target sequence information was also posted at https://predictioncenter.org/caspcommons/targetlist.cgi. Participants were asked to return their models in three weeks.

2. 3D Structures

Over 1,500 3D models were submitted in the first CASP-COVID round. Those included models from the most capable research groups as previously assessed in CASP [30-37]. Methods descriptions provided by authors of this paper are in available in the Supplementary Material ('TS methods' file). The full list of participants and associated statistics are at https://predictioncenter.org/caspcommons/groups_info.cgi.

All collected models were posted at the Prediction Center Data Archive site https://predictioncenter.org/download_area/CASPCOMMONS/2020_COVID-19/ immediately after closing the first round of submissions. The models were analyzed for structural consensus based on the average pair-wise global and local LDDT [38] and GDT_TS [39,40] scores. The results of the analysis allowed identification of consensus regions of structure and of groups with structurally similar models. For example, for the SARS-CoV2 M-protein (target C1906), high local consensus scores in region 1-105 (marked with the black box in Fig. 1) suggested the protein has two domains, and that a split into two domain level targets in round 2 of the experiment might assist modeling.



	Model	Predictor	GDTTS_cons
1	C1906TS301_1	FEIGLAB-S	0.327
2	C1906TS044_1	FEIGLAB	0.327
3	C1906TS152_3	MULTICOM	0.327
4	C1906TS438_3	Destini	0.324
5	C1906TS213_4	McGuffin	0.324
6	C1906TS438_5	Destini	0.323
7	C1906TS438_2	Destini	0.323
8	C1906TS213_1	McGuffin	0.323
9	C1906TS438_1	Destini	0.323
10	C1906TS152_2	MULTICOM	0.323
11	C1906TS213_2	McGuffin	0.323
12	C1906TS213_3	McGuffin	0.323
13	C1906TS152_1	MULTICOM	0.322
14	C1906TS413_1	TFold-server	0.322
15	C1906TS413_3	TFold-server	0.321
16	C1906TS413_2	TFold-server	0.321
17	C1906TS213_5	McGuffin	0.320
18	C1906TS438_4	Destini	0.320
19	C1906TS413_4	TFold-server	0.319
20	C1906TS278_5	FALCON	0.317

Figure 1. Partial screenshot showing part of the consensus table (https://predictioncenter.org/caspcommons/models_consensus2.cgi) for the SARS-CoV2 M-

protein (target C1906) showing local structural agreement along the sequence of the selected model (second column) with the remaining models. The black box shows the region where many models agree, suggesting a relatively easy to model domain.

3. Community-wide discussion of the results and second round of modeling

Following the first round of modeling, the community discussed the results in two Zoom conferences and group chat using the Microsoft teams. Consensus analyses helped identify consistent domain boundaries within the targets, used in the second modeling round. Community members also discussed possible features of models such as membrane regions and signal peptides, that could help guide the next stage of modeling.

The second round ran for two weeks in May 2020, immediately before the start of the regular CASP14 experiment. The round consisted of 15 domain-level targets derived from the round 1 analysis, and 7 first-round targets re-released for prediction. 33 groups submitted over 1,500 3D models, which were again made public immediately after the deadline.

Second round models underwent the same evaluation procedure as those from round 1.

4. Accuracy estimates

Each of the submitted models in both rounds of modeling was evaluated by accuracy estimation methods developed by the CASP community. Overall, 32 EMA methods were used. The list of participated methods and brief descriptions are provided in the 'EMA methods' Supplementary file. All submitted accuracy estimates are available at https://predictioncenter.org/caspcommons/models_QAresults.cgi .

The overall goal of this step was to identify the best models for each target and to estimate their accuracy. This was the first time CASP has addressed this non-trivial task in a real-life situation. Previous regular CASP experiments have shown that EMA methods are overall effective at ranking models by accuracy, but even the best-performing methods cannot identify the most accurate models for all targets [41-43]. The CASP-COVID results

showed surprisingly high variation in model rankings: for no target was there unanimous agreement on the best 3D model. Rather, for most targets over ten distinct models were selected as the best (Supplementary table STQA1), creating a problem in recommending which model should be used. To address this issue, the Venclovas group devised a new EMA-jury algorithm that identifies which models were most favored by the EMA methods. The algorithm is described in detail in the Supplement. Briefly, the method pools the top 1, top 2, …, top 10 models selected by each EMA ranking into ten corresponding supersets. If a model is selected by more than one EMA method, it is included multiple times, thus receiving more weight. A consensus structural similarity score is calculated for every model in each superset as an average of CAD-scores[44] from the model's pairwise comparisons with other models in the superset (Supplementary figure SFQA1). The maximum of superset-specific consensus scores for a model is recorded as the EMA-jury consensus score. Note that the EMA-jury consensus score quantifies how typical the structure of a model is among the top selections made by the EMA methods rather than the expected level of its structural similarity to the native structure (as individual EMAs do). The EMA-jury scores together with two additional refinement criteria described in the Supplementary Material are used for the final selection of models that are most strongly supported by the EMA methods (Supplementary table STQA2).

Comparison of the EMA-jury scores with the overall consensus scores computed on full sets of models for each CASP-COVID target shows that the EMA-jury method always selects a subset of models that are more structurally similar within the subset than overall (Fig.2). This indicates that individual EMA rankings are not random and often agree in favoring some structural features.

The EMA-jury algorithm was also run using the LDDT scoring function (instead of CAD-score). The results are presented in the Supplementary Material (figures SFQA2-3, and table STQA3). They are very similar to the CAD-score based results with 84% of selected CASP-COVID models being the same, and at least one model in common for every target.

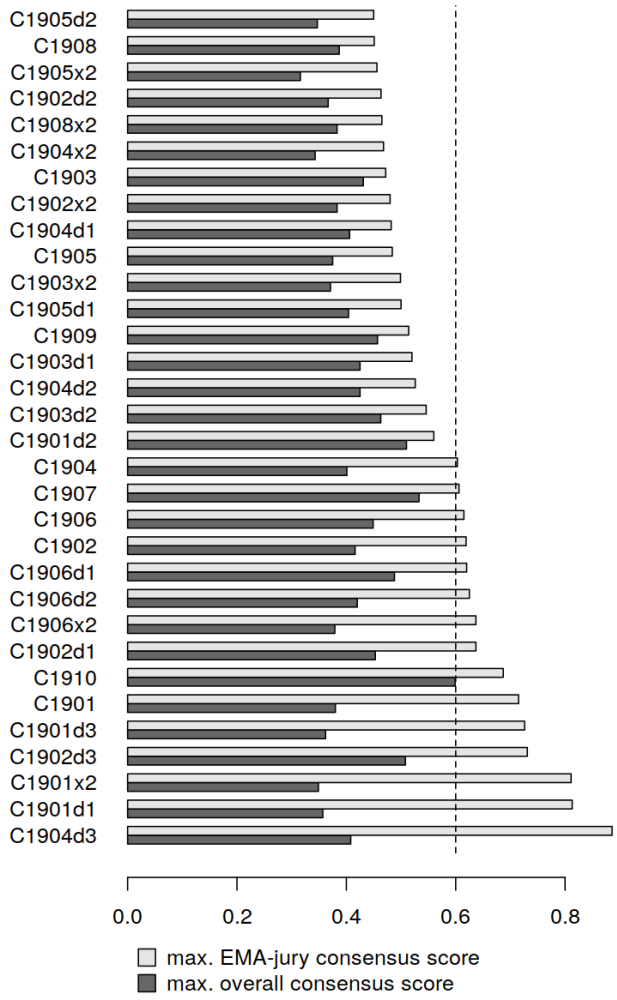


Figure 2. Maximum consensus scores on CASP-COVID targets (EMA-jury - grey bars; overall consensus - black). Targets are ordered by increasing EMA-jury values. The grey bars are always longer than black ones, indicating that the EMA-jury method successfully selects subsets of models that are more structurally consistent. The vertical dashed line corresponds to the consensus level of 0.6, which represents 100th percentile of overall consensus scores for all models (Supplementary figure SFQA4).

To assess the effectiveness of the EMA-jury method, we evaluated its ability to select the best available model from a set of models. Such an analysis requires knowing actual accuracy of models with respect to the target structure. Since only two CASP-COVID targets have been solved so far, we tested the EMA-jury on CASP13 set of server models (almost 11,000 models on 80 targets). Fig. 3 shows that the EMA-jury very often picks the best or nearly the best model, and that the EMA-jury selection is better than simple consensus-based selection. The mean score of the EMA-jury-selected models (0.622) is just slightly behind the mean of the maximum CAD-scores of CASP13 models (0.640) and better than the mean score of

models selected with simple-consensus (0.574). The average Z-score (calculated from the distribution of individual EMA scores) of Jury-selected models stands at 1.67, almost twice the value of the average simple-consensus Z-score (0.87). Of interest is also the fact that the relative performance of the EMA-jury with respect to simple consensus becomes even more dominant on harder modeling targets. For example, the average EMA-jury Z-score grows from 1.67 on all CASP13 targets to 2.02 on FM targets, while the corresponding numbers for simple consensus are trending downward: 0.87 → 0.75. Similar tendencies in scores are observed when analyzing the LDDT-based results (Supplementary figure SFQA5).

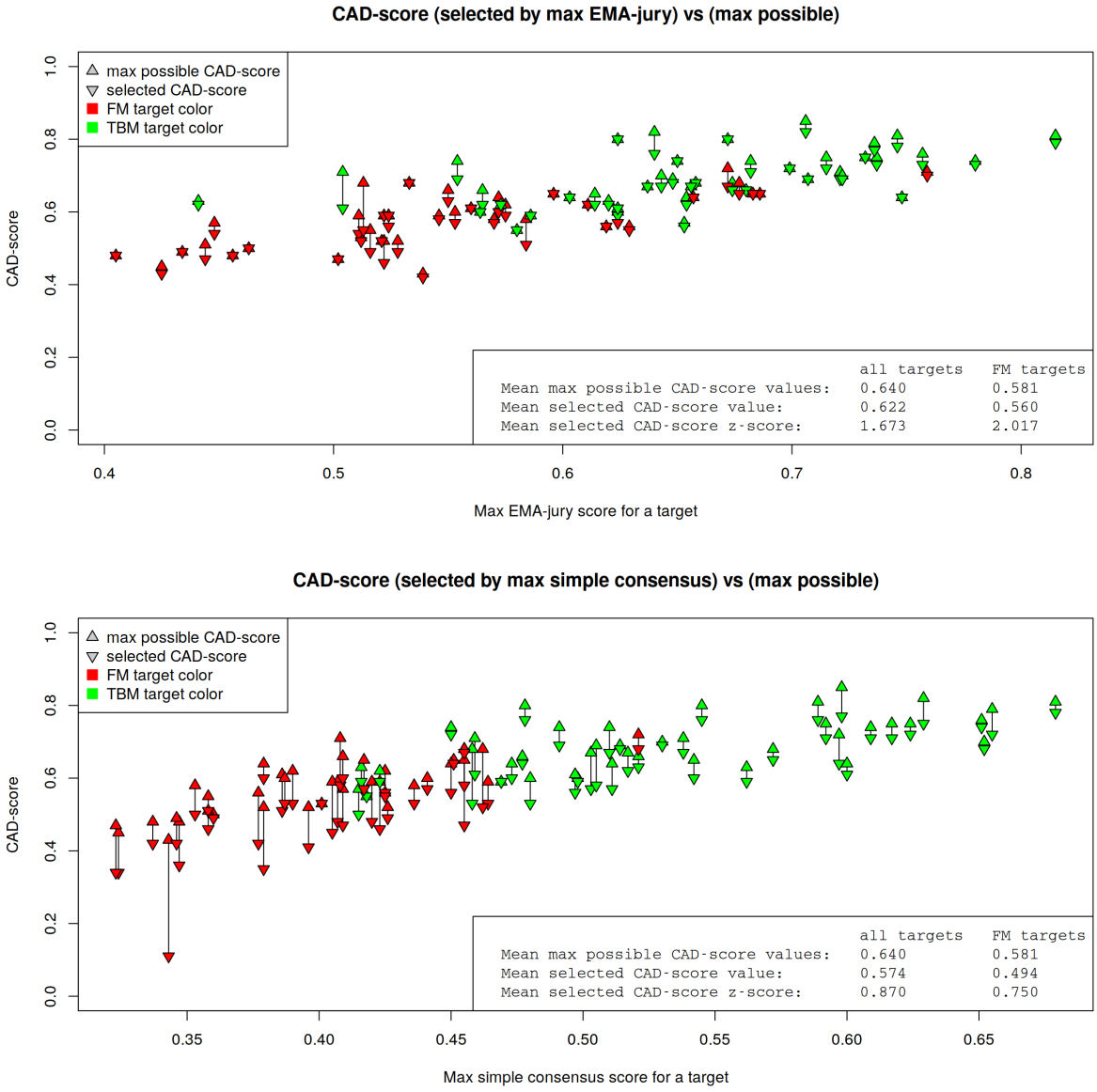


Figure 3. Selection of the top model by the EMA-jury (top panel) and simple structural consensus (bottom panel) on 80 CASP13 targets. Maximum per-target CAD-scores are shown as pointing up triangles; the CAD-scores of models selected by the EMA-jury

approach (top) and simple structural consensus method (bottom) are shown as pointing down triangles. The hardest to predict targets (FM) are in red, others in green. Vertical lines between the corresponding triangles represent the error in the selection process. Comparison of the top and bottom panels demonstrates that the EMA-jury method selects models closer to the best absolute value more often than the simple consensus.

5. Evaluation of ORF3a and ORF8 models.

Structures of two CASP-COVID proteins - ORF3a (Target ID: C1905) and ORF8 (Target ID: C1908) - were experimentally solved by the start of CASP14 conference allowing full CASP evaluation of accuracy of the corresponding models against experimental structures.

Full-length sequences of both solved targets were released for modeling in both rounds of CASP-COVID, and ORF3a was additionally released in the second round as domain targets C1905-D1 and C1905-D2. Independently, ORF8 was also released in the CASP14 experiment as target T1064. The number of 3D models and EMA estimates collected in the CASP-COVID experiment are summarized in **Table 1**.

Table 1. The number of 3D models and accuracy estimates in the CASP-COVID experiment for ORF3a and ORF8. Numbers in parentheses show the number of high-accuracy models. ORF3a was treated as one target in the first round of CASP-COVID (C1905) and as two separate domains in Round 2 (C1905-D1, C1905-D2).

CASP-COVID Target ID	ORF3a			ORF8
	C1905	C1905-D1	C1905-D2	C1908
No. 3D models (GDT_TS ≥ 40)	153 (6)	83 (38)	79 (0)	181 (0)
No. EMA submissions in CASP-COVID	30	19	19	29

Since there was no significant accuracy improvement in models submitted on full-length targets in the second round, we report only the first-round results for those.

5.1.Post-CASP experiment

From the CASP-COVID and CASP14 evaluation of ORF3a and ORF8 targets, it was immediately apparent that models from DeepMind's AlphaFold2 group were by far the most accurate, consistent with the broader CASP14 results. An interesting question to check was

whether accuracy assessment methods can recognize the high accuracy of these models. However, it was impossible to answer this question only with the available data at hand: AlphaFold2 did not submit models in the second round of CASP-COVID (thus no domain-based models for ORF3a), nor did they submit ORF8 models to CASP-COVID (only to CASP14). To adjust for that, we added five AlphaFold models to each of the three CASP-COVID model sets. For ORF8, we added AlphaFold2 (AF2) models submitted on the CASP14 T1064 target. For ORF3a domains, we added AlphaFold models submitted to CASP-COVID (AF-COV) and a-posteriori split into domains.

Additional accuracy estimates were solicited on the added AlphaFold models from the authors of ten established in CASP EMA methods. We discuss here the results for four (out of these ten) that participated both in CASP-COVID and CASP14: ModFOLD8_rank, ProQ3D, VoroMQA-dark, and QMEANDisCo. The overall conclusions do not change by including all ten post-CASP EMA methods.

This analysis, aimed at determining whether accuracy estimation methods were able to recognize high accuracy of AlphaFold models of the two CASP_COVID targets, is referred to here as the post-CASP EMA.

5.2.Results for ORF3a (C1905)

5.2.1. Round 1 results: models of the full structure

Among the first-round 3D models of the full structure, only six models have GDT_TS scores above 40 (green crosses in **Fig. 4**a). Five of these models are from AlphaFold (with accuracy ranging from 45 to 59 GDT_TS), and the sixth is from FEIGLAB-R, who attempted to refine an AlphaFold model resulting in a lower (worse) GDT_TS score of 42. The six top models are all monomeric, while the experimental ORF3a structure is dimeric. Overall, the best AlphaFold model (AF-COV_2, GDT_TS=59) correctly reproduces ORF3a's fold (Supplementary figure SFQA6a), but loops and orientation of helixes around the dimeric interface are less accurate: the average per-residue distance error (as calculated from the

optimal LGA model-target superposition) is 3.9 Å for the whole structure, and 4.6 Å for the interface region.

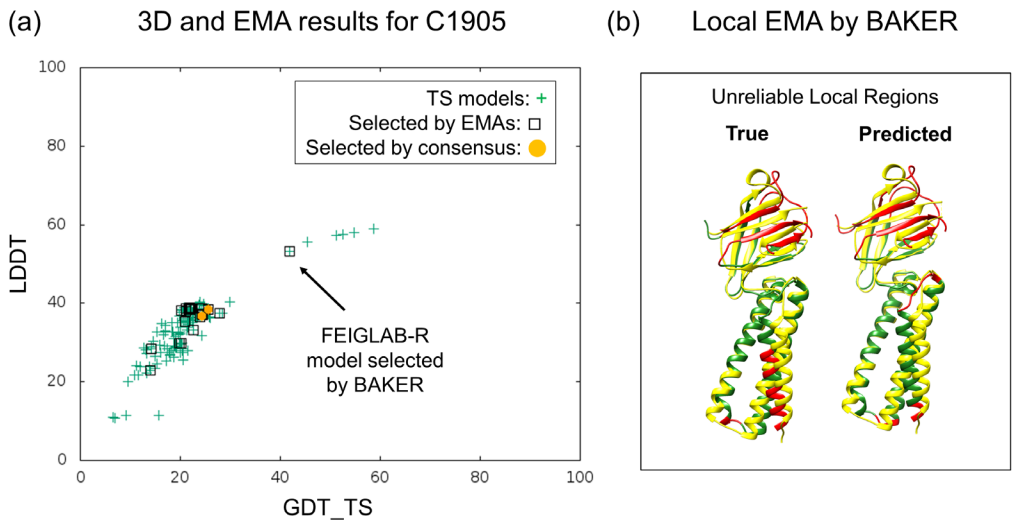


Figure 4. Round 1 3D and accuracy estimation results for SARS2 ORF3a (C1905). (a) Each green cross represents a 3D model, black squares indicate models selected as high accuracy by accuracy estimation methods, and orange circles indicate models selected by the EMA-Jury method. 3D model accuracy is shown in terms of LDDT (Y axis) and GDT_TS (X-axis). Only one accuracy estimation method selected a higher accuracy model. (b) Locally inaccurate regions of the highest-scoring model, AF-COV_2, according to the ULR definition (left) and as predicted for the same model by the BAKER EMA method (right). The superpositions are identical; the crystal structure is in yellow, ULRs and predicted inaccurate regions are in red and the rest of the model in green.

In terms of global EMA, BAKER was the only group who selected a reasonable model (GDT_TS > 40) as top1. However, it was the sixth-ranked model with the GDT_TS of 42 rather than the most accurate model with GDT_TS of 59. Other EMA methods selected a number of much less accurate models (black squares at low LDDT and GDT_TS), including the EMA-Jury method (orange circles), which by its nature selects models preferred by the majority of individual EMAs.

In the evaluation of local accuracy in the post-CASP EMA, the ProQ3D group had the best average results, with the ASE score of 85.4 (ASE – Assessment of S-function Errors, see the EMA assessment paper [41]), AUC of 0.86 (AUC - Area Under the ROC Curve of the

prediction of accurate/inaccurate residues), and the ULR-F1 score of 0.4 (ULR-F1 – the F1-score on Unreliable Local Regions, see papers [41,45]) for the best submitted model AF_2 (C1905TS156_2). AlphaFold's self-estimate of per-residue distance errors was worse than the results of ProQ3D, scoring ASE of 72.7, AUC of 0.78 and ULR-F1 of 0.0. The BAKER local EMA method was able to identify some part of the ULRs in the beta sheet domain (actual ULRs = 163-198 and 219-235; predicted ULRs = 163-199 and 214-238), but the ULRs in the alpha helix domain were identified less precisely (actual ULRs = 40-48, 51-55, and 102-104; predicted ULRs = 40-43, 62-68, and 99-101), as illustrated in **Fig. 4**b. ULRs are defined as regions consisting of three or more sequential model residues deviating by more than 3.8 Å from the corresponding target residues in the optimal superposition on the crystal structure.

5.2.2. Round 2 results: prediction of the domain structures

Fig. 5 shows the accuracy distribution of CASP-COVID second round models for the two domains of ORF3a separately. The domain structures of the AF-COV models submitted in the first round are included (pink stars), and are substantially more accurate that those from other groups, especially for domain 2.

In the post-CASP experiment, three and two out of four EMA groups picked an AF-COV model as top1 for domains 1 and 2, respectively (pink squares in **Figs 5**a and b). Although some EMA groups could discriminate AF-COV models from the others, no group was successful in predicting the correct ranking within the five AF-COV models, although these models are very close.

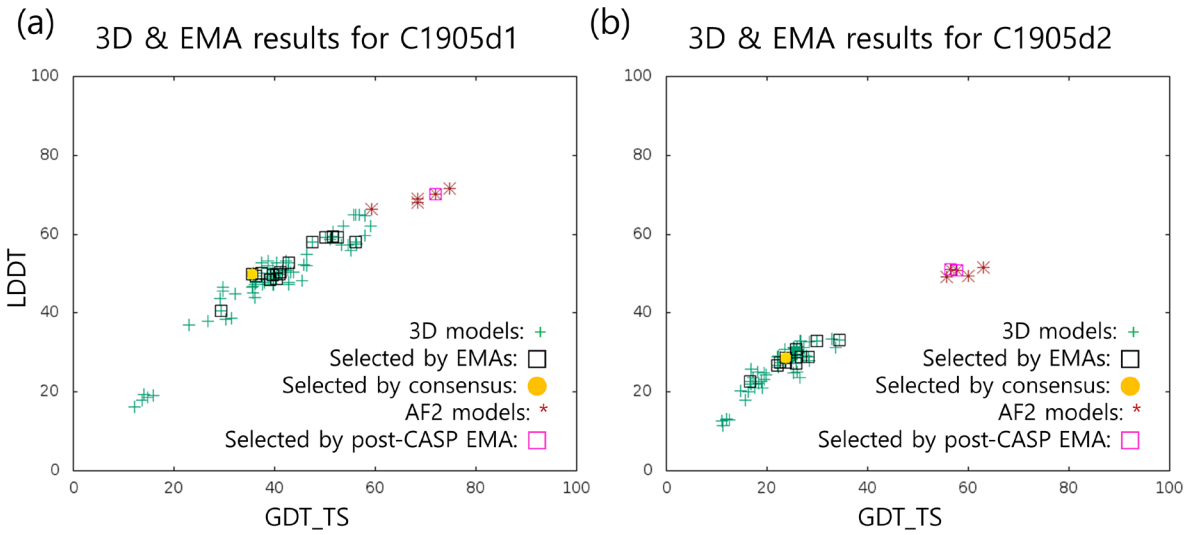


Figure 5. Round 2 3D and accuracy estimation results for two domains of SARS-CoV-2 ORF3a protein (a) C1905-D1 and (b) C1905-D2. 3D model accuracy is shown in terms of LDDT (Y axis) and GDT_TS (X-axis) (green crosses). The panels show both models from CASP-COVID and AF-COV models added in the post-CASP EMA experiment (pink stars). The models selected by EMA methods as top1 during CASP-COVID are shown as black hollow squares; models selected in the post-CASP experiment are in pink hollow squares. For domain 1, three out of four EMA groups selected one of the higher accuracy AlphaFold models, with many low accuracy models also selected. There is a similar pattern for domain 2, where two of four methods picked two different AlphaFold models.

5.3.Results for ORF8 (C1908)

For ORF8, no high-accuracy models were submitted during CASP-COVID (maximum GDT_TS=26, AlphaFold not participating) (see green crosses in **Figure 6**). The protein was re-released in the regular CASP14 experiment as target T1064 (without 15 N-term residues corresponding to a signal peptide, a feature which almost all CASP-COVID participants ignored, one cause of poor models). The AlphaFold2 group submitted five high-accuracy predictions for this target. These models (ranging from 64 to 87 GDT_TS) were added to the pool of models for the post-CASP analysis (pink stars in **Fig. 6**). The crystal structure of ORF8 was solved as a covalent dimer, while AlphaFold models were monomeric. Despite this, the best monomeric model possesses some important structural features needed for forming the dimeric assembly. In particular, the model correctly reproduces the sidechain orientation of the cysteine involved in covalent chain linkage (Supplementary figure

SFQA6b). The average per-residue distance error is similar for the whole structure (1.25 Å) and for the interface region (1.46 Å).

In global accuracy estimation, only VoroMQA-dark could identify AF2 models as superior to others (pink squares in **Fig. 6**). However, this method did not predict the big difference in absolute model quality (as quantified by GDT_TS). For example, VoroMQA-dark assigned the best AF2 model (AF2_1, GDT_TS=87) a global EMA score of 67 (on the 0-100 scale), while some models by other groups with the GDT_TS<20 were assigned a relatively high EMA score of 50+ (all scores are for ORF8 without the signal peptide). It should be noted that VoroMQA-dark has a narrow range of values so that a difference of 10+ may indicate substantial difference in model accuracy.

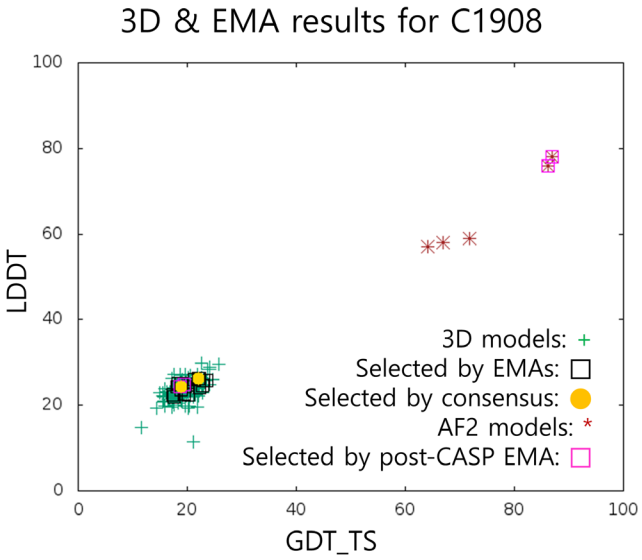


Figure 6. Round 1 3D modeling and Accuracy Estimation (EMA) results for SARS-CoV-2 protein ORF8 (C1908). 3D model accuracy for submissions in terms of LDDT (Y axis) and GDT_TS (X-axis) (green crosses) and EMA selections (black squares for CASP-COVID, pink squares for post-CASP experiment, orange circles for EMA-Jury). Five AF2 models added in the post-CASP experiment are shown as pink stars. Two of the AF2 models are impressively accurate. Two post-CASP EMA methods succeeded in selecting those models as best.

In the evaluation of local accuracy in the post-CASP EMA, the best results were shown again by the ProQ3D, with ASE of 88.5, AUC of 0.89, and the perfect ULR-F1 score of 1.0

for the AlphaFold2 model AF2_1 (CASP14 id: T1064TS427_1). AlphaFold2's self-estimate of per-residue distance errors was comparable or better than the results of the best EMA method, scoring ASE of 92.7, AUC of 0.96 and ULR-F1 of 1.0.

All AlphaFold2 models showed local structural differences to experiment near residues 60-86, which are involved in a crystal contact (Supplementary figure SFQA6c), and residues 104-110 which have high crystallographic B-factor of ~70 (Supplementary figure SFQA6d). ProQ3D could identify the structural deviations in these two loop regions of AF2 models with high accuracy, scoring 0.78 with ULR-F1 measure. GraphQA also showed a high performance with average ULR-F1 score of 0.68, while AlphaFold2's self-assessment scored 0.47. On the other hand, it is not clear that the models have errors in either of these regions rather than being a crystal artifact and a crystallographic error respectively. It's possible that the EMA methods are predicting relatively flexible regions of polypeptide, rather than model errors.

Discussion.

The central goal of CASP is to make assessment of both 3D modeling methods and accuracy estimation methods as rigorous possible, by using a blind prediction system and comparison with experiment. In doing so over 14 rounds, CASP has built a strong community. Further, recent advances in modeling methods show the field has advanced to the point [46-48] where taking on the most challenging structures should yield useful results. In the past, CASP has also found that properly balanced consensus models can achieve higher accuracy than any of the contributing models [49]. So, there was an obvious appeal to drawing on this community resource to address one aspect of the COVID-19 emergency. Indeed, there was very enthusiastic response and participation from the CASP community.

From a more pragmatic point of view, the CASP-COVID modeling initiative also provided a different, real-world application of the modeling methods. Although CASP strives

to be as realistic as possible, assessment is done with knowledge of the experimental answers. What can be done when the goal is to generate useful information from models?

Since we do not yet know most of the experimental structures of the target proteins, conventional CASP analysis is limited to just two targets. In both cases, correct folds were produced by just one group, AlphaFold2. Based on the most recent CASP14 results [46,48,50], we expect better performance overall, with at least the majority of the folds correctly predicted by multiple groups. We will have to wait for more experimental results to see if that is true.

The most difficult task in generating recommended models turned out to be estimating relative accuracy and, beyond that, absolute accuracy of the submissions. CASP has nurtured the development of accuracy estimation methods for more than a decade, and assessment against experiment has shown impressive progress, with apparently very useful outcomes [41-43,45,51-53]. However, in the absence of experimental ground truth, initial focus was on agreement between methods and this was low. In turn, this prompted the development of a new method for obtaining consensus accuracy estimates.

In spite of these limitations, overall, we regard the experiment as a success, both in terms of bringing the community together to tackle an urgent problem, and in producing a set of potentially useful models. As noted above it was also valuable in drawing attention to issues in real world use that were not apparent in the standard CASP environment. It also once again demonstrated the value of community science. In particular, the experiment was particularly impactful for undergraduate students just beginning in the field, as they were able to better understand the role of their research in a broader scientific context and its potential for benefiting society at large.

Acknowledgements and contributions

Major contributions to the main text of the paper:

- Idea, organization, paper concept, coordination and editing - AK, JM, KF.
- Abstract, Introduction, Sections 1, 2, 3, 4, 5 and Discussion – AK, JM.
- Section 4 – KO, ČV.
- Section 5 – SK, JW, CS.
- Introduction, Discussion – WB, DDC.

All other authors contributed to the paper by providing description of their methods for the Supplementary Material.

The CASP experiment is supported by the US National Institute of General Medical Sciences (NIGMS/NIH), grant number GM100482.

CS was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (Nos. 2020M3A9G7103933 and 2019M3E5D4066898).

ČV and KO were in part supported by the Research Council of Lithuania (grants S-MIP-17-60 and S-MIP-21-35).

The predictions made by MULTICOM predictors were partially supported by two NSF grants (DBI 1759934 and IIS1763246), one NIH grant (GM093123), and two DOE grants (DE-SC0020400 and DE-SC0021303) to JC.

KT, YT, YY were partially supported by Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP20am0101110. Computational resource of AI Bridging Cloud Infrastructure (ABCI) provided by National Institute of Advanced Industrial Science and Technology (AIST) was used.

NFF, AM, JAP acknowledge support received from the UK Biotechnology and Biological Science Research Council (BBS/E/W/0012843D)

The work by LJM and RA was supported by the Biotechnology and Biological Sciences Research Council (BBSRC), grant number BB/T018496/1.

CC, MK, AL, EL are supported by grants UMO-2017/26/M/ST4/00044 (to CC), UMO-2017/25/B/ST4/01026 (to AL) from the National Science Centre (NCN), Poland. UNRES group used computer resources: CI TASK, Technical University of Gdańsk; ICM, University of Warsaw (grant: GA76-11); Cyfronet, AGH University of Science and Technology, Cracow (grant: unres19). The authors thank Anna Antoniak, Artur Giełdoń, Sergey A. Samsonov, Adam K. Sieradzan, Rafał Ślusarz (Faculty of Chemistry, University of Gdańsk) for assistance in solving part of the targets and background work.

DK is partially supported by the National Institutes of Health (R01GM133840 and R01GM123055) and the National Science Foundation (CMMI1825941, MCB1925643, and DBI2003635). CC is supported by the National Institute of General Medical Sciences-funded predoctoral fellowship to C.C. (T32 GM132024).

Figure captions

Figure 1. Screenshot of the model consensus table (https://predictioncenter.org/caspcommons/models_consensus2.cgi) showing local and global structural agreement of models according to the LDDT (left) and GDT_TS (right) scores. Local consensus scores are presented as color-coded bars, while global scores are presented by a number to the right of the corresponding local bar.

Figure 2. Maximum consensus scores on CASP-COVID targets (EMA-jury - grey bars; overall consensus - black). Targets are ordered by increasing EMA-jury values. The grey bars are always longer than black ones, indicating that the EMA-jury method successfully selects subsets of models that are more structurally consistent. The vertical dashed line corresponds to the consensus level of 0.6, which represents 100^{th} percentile of overall consensus scores for all models (Supplementary figure SFQA4).

Figure 3. Selection of the top model by the EMA-jury (top panel) and simple structural consensus (bottom panel) on 80 CASP13 targets. Maximum per-target CAD-scores are shown as pointing up triangles; the CAD-scores of models selected by the EMA-jury approach (top) and simple structural consensus method (bottom) are shown as pointing down triangles. The hardest to predict targets (FM) are in red, others in green. Vertical lines between the corresponding triangles represent the error in the selection process. Comparison of the top and bottom panels demonstrates that the EMA-jury method selects models closer to the best absolute value more often than the simple consensus.

Figure 4. Round 1 3D and accuracy estimation results for SARS2 ORF3a (C1905). (a) Each green cross represents a 3D model, black squares indicate models selected as high accuracy by accuracy estimation methods, and orange circles indicate models selected by the EMA-

Jury method. 3D model accuracy is shown in terms of LDDT (Y axis) and GDT_TS (X-axis). Only one accuracy estimation method selected a higher accuracy model. (b) Locally inaccurate regions of the highest-scoring model, AF-COV_2, according to the ULR definition (left) and as predicted for the same model by the BAKER EMA method (right). The superpositions are identical; the crystal structure is in yellow, ULRs and predicted inaccurate regions are in red and the rest of the model in green.

Figure 5. Round 2 3D and accuracy estimation results for two domains of SARS-CoV-2 ORF3a protein (a) C1905-D1 and (b) C1905-D2. 3D model accuracy is shown in terms of LDDT (Y axis) and GDT_TS (X-axis) (green crosses). The panels show both models from CASP-COVID and AF-COV models added in the post-CASP EMA experiment (pink stars). The models selected by EMA methods as top1 during CASP-COVID are shown as black hollow squares; models selected in the post-CASP experiment are in pink hollow squares. For domain 1, three out of four EMA groups selected one of the higher accuracy AlphaFold models, with many low accuracy models also selected. There is a similar pattern for domain 2, where two of four methods picked two different AlphaFold models.

Figure 6. Round 1 3D modeling and Accuracy Estimation (EMA) results for SARS-CoV-2 protein ORF8 (C1908). 3D model accuracy for submissions in terms of LDDT (Y axis) and GDT_TS (X-axis) (green crosses) and EMA selections (black squares for CASP-COVID, pink squares for post-CASP experiment, orange circles for EMA-Jury). Five AF2 models added in the post-CASP experiment are shown as pink stars. Two of the AF2 models are impressively accurate. Two post-CASP EMA methods succeeded in selecting those models as best.

Table caption

Table 1. The number of 3D models and accuracy estimates in the CASP-COVID experiment for ORF3a and ORF8. Numbers in parentheses show the number of high-accuracy models. ORF3a was treated as one target in the first round of CASP-COVID (C1905) and as two separate domains in Round 2 (C1905-D1, C1905-D2)

References

1. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)-Round XIII. *Proteins.* 2019;87(12):1011-1020.
2. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)-Round XII. *Proteins.* 2018;86 Suppl 1:7-15.
3. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins.* 2016;84 Suppl 1:4-14.
4. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)--round x. *Proteins.* 2014;82 Suppl 2:1-6.
5. Moult J, Fidelis K, Kryshtafovych A, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)--round IX. *Proteins.* 2011;79 Suppl 10:1-5.
6. Cozzetto D, Kryshtafovych A, Fidelis K, Moult J, Rost B, Tramontano A. Evaluation of template-based models in CASP8 with standard measures. *Proteins.* 2009;77 Suppl 9:18-28.

7. Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T. Assessment of template based protein structure predictions in CASP9. *Proteins.* 2011;79 Suppl 10:37-58.
8. Huang YJ, Mao B, Aramini JM, Montelione GT. Assessment of template-based protein structure predictions in CASP10. *Proteins.* 2014;82 Suppl 2:43-56.
9. Modi V, Xu Q, Adhikari S, Dunbrack RL, Jr. Assessment of template-based modeling of protein structure in CASP11. *Proteins.* 2016;84 Suppl 1:200-220.
10. Kryshtafovych A, Monastyrskyy B, Fidelis K, Moult J, Schwede T, Tramontano A. Evaluation of the template-based modeling in CASP12. *Proteins.* 2018;86 Suppl 1:321-334.
11. Croll TI, Sammito MD, Kryshtafovych A, Read RJ. Evaluation of template-based modeling in CASP13. *Proteins.* 2019;87(12):1113-1127.
12. Lafita A, Bliven S, Kryshtafovych A, et al. Assessment of protein assembly prediction in CASP12. *Proteins.* 2018;86 Suppl 1:247-256.
13. Read RJ, Sammito MD, Kryshtafovych A, Croll TI. Evaluation of model refinement in CASP13. *Proteins.* 2019;87(12):1249-1262.
14. Hovan L, Oleinikovas V, Yalinca H, Kryshtafovych A, Saladino G, Gervasio FL. Assessment of the model refinement category in CASP12. *Proteins.* 2018;86 Suppl 1:152-167.
15. Guzenko D, Lafita A, Monastyrskyy B, Kryshtafovych A, Duarte JM. Assessment of protein assembly prediction in CASP13. *Proteins.* 2019;87(12):1190-1199.
16. Abriata LA, Tamo GE, Monastyrskyy B, Kryshtafovych A, Dal Peraro M. Assessment of hard target modeling in CASP12 reveals an emerging role of alignment-based contact prediction methods. *Proteins.* 2018;86 Suppl 1:97-112.
17. Abriata LA, Tamo GE, Dal Peraro M. A further leap of improvement in tertiary structure prediction in CASP13 prompts new routes for future assessments. *Proteins.* 2019;87(12):1100-1112.
18. Kryshtafovych A, Prlic A, Dmytriv Z, et al. New tools and expanded data analysis capabilities at the Protein Structure Prediction Center. *Proteins.* 2007;69 Suppl 8:19-26.
19. Kryshtafovych A, Monastyrskyy B, Fidelis K. CASP11 statistics and the prediction center evaluation system. *Proteins.* 2016;84 Suppl 1:15-19.
20. Kryshtafovych A, Monastyrskyy B, Fidelis K. CASP prediction center infrastructure and evaluation measures in CASP10 and CASP ROLL. *Proteins.* 2014;82 Suppl 2:7-13.
21. Kryshtafovych A, Milostan M, Szajkowski L, Daniluk P, Fidelis K. CASP6 data processing and automatic evaluation at the protein structure prediction center. *Proteins.* 2005;61 Suppl 7:19-23.
22. Kryshtafovych A, Krysko O, Daniluk P, Dmytriv Z, Fidelis K. Protein structure prediction center in CASP8. *Proteins.* 2009;77 Suppl 9:5-9.
23. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* 2020;579(7798):265-269.
24. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. *Methods Mol Biol.* 2017;1607:627-641.
25. Steinegger M, Meier M, Mirdita M, Vohringer H, Haunsberger SJ, Soding J. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics.* 2019;20(1):473.
26. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic acids research.* 2019;47(W1):W402-W407.
27. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. *Bioinformatics.* 2004;20(13):2138-2139.
28. Kall L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic acids research.* 2007;35(Web Server issue):W429-432.
29. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019;37(4):420-423.
30. Cheng J, Choe MH, Elofsson A, et al. Estimation of model accuracy in CASP13. *Proteins.* 2019;87(12):1361-1377.
31. Dapkunas J, Olechnovic K, Venclovas C. Structural modeling of protein complexes: Current capabilities and challenges. *Proteins.* 2019;87(12):1222-1232.

32. Heo L, Arbour CF, Feig M. Driven to near-experimental accuracy by refinement via molecular dynamics simulations. *Proteins.* 2019;87(12):1263-1275.
33. Park H, Lee GR, Kim DE, Anishchenko I, Cong Q, Baker D. High-accuracy refinement using Rosetta in CASP13. *Proteins.* 2019;87(12):1276-1282.
34. Ovchinnikov S, Park H, Kim DE, DiMaio F, Baker D. Protein structure prediction using Rosetta in CASP12. *Proteins.* 2018;86 Suppl 1:113-121.
35. Hou J, Wu T, Cao R, Cheng J. Protein tertiary structure modeling driven by deep learning and contact distance prediction in CASP13. *Proteins.* 2019;87(12):1165-1178.
36. Zheng W, Li Y, Zhang C, Pearce R, Mortuza SM, Zhang Y. Deep-learning contact-map guided protein structure prediction in CASP13. *Proteins.* 2019;87(12):1149-1164.
37. Senior AW, Evans R, Jumper J, et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins.* 2019;87(12):1141-1148.
38. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics.* 2013;29(21):2722-2728.
39. Zemla A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* 2003;31(13):3370-3374.
40. Zemla A, Venclovas, Moult J, Fidelis K. Processing and evaluation of predictions in CASP4. *Proteins.* 2001;Suppl 5:13-21.
41. Won J, Baek M, Monastyrskyy B, Kryshtafovych A, Seok C. Assessment of protein model structure accuracy estimation in CASP13: Challenges in the era of deep learning. *Proteins.* 2019;87(12):1351-1360.
42. Kryshtafovych A, Monastyrskyy B, Fidelis K, Schwede T, Tramontano A. Assessment of model accuracy estimations in CASP12. *Proteins.* 2018;86 Suppl 1:345-360.
43. Kryshtafovych A, Barbato A, Monastyrskyy B, Fidelis K, Schwede T, Tramontano A. Methods of model accuracy estimation can help selecting the best models from decoy sets: Assessment of model accuracy estimations in CASP11. *Proteins.* 2016;84 Suppl 1:349-369.
44. Olechnovic K, Kulberkyte E, Venclovas C. CAD-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins.* 2013;81(1):149-162.
45. Kwon S, Won J, Kryshtafovych A, Seok C. Assessment of protein model structure accuracy estimation in CASP14: Old and new challenges. *Proteins.* 2021; This issue, doi: 10.1002/prot.26192.
46. Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. High-accuracy protein structure prediction in CASP14. *Proteins.* 2021; This issue, doi: 10.1002/prot.26171.
47. Jumper J, et al. AlphaFold2 in CASP14. *Proteins.* 2021;This issue, Prot-00211-2021, in review.
48. Kinch LN, Pei J, Kryshtafovych A, Schaeffer RD, Grishin NV. Topology evaluation of models for difficult targets in the 14th round of the critical assessment of protein structure prediction. *Proteins.* 2021;This issue, doi: 10.1002/prot.26172.
49. Wallner B, Larsson P, Elofsson A. Pcons.net: protein structure prediction meta server. *Nucleic Acids Res.* 2007;35(Web Server issue):W369-374.
50. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical Assessment of Methods of Protein Structure Prediction (CASP) – Round XIV. *Proteins.* 2021;This issue, Prot-00250-2021, accepted.
51. Cozzetto D, Kryshtafovych A, Tramontano A. Evaluation of CASP8 model quality predictions. *Proteins.* 2009;77 Suppl 9:157-166.
52. Kryshtafovych A, Barbato A, Fidelis K, Monastyrskyy B, Schwede T, Tramontano A. Assessment of the assessment: evaluation of the model quality estimates in CASP10. *Proteins.* 2014;82 Suppl 2:112-126.
53. Kryshtafovych A, Fidelis K, Tramontano A. Evaluation of model quality predictions in CASP9. *Proteins.* 2011;79 Suppl 10:91-106.