CASP-covid

(or CASP-Commons 2020 experiment)

https://predictioncenter.org/caspcommons

Register for CASP_Commons

Your CASP_Commons groups /predictions

Released Target List

Participating Groups

Submit a prediction

ARCHIVE: submitted predictions (CASP-covid, 2020)

Model Accuracy Estimates [2018, NMR nominated targets]

COVID-19 MODELING RESULTS [2020]

Andriy Kryshtafovych:

introduction of what the initiative was, what we collected, what analyses we did

assessment of models on two structurally determined CASP-covid targets

Chaok Seok:

EMA results on the solved CASP-covid targets

Kliment Olechnovic:

EMA-jury system for CASP-covid

Panelists:

what we learned (or could do better)

what we do further with the results

how to attract more attention of structural biology community

next steps

CASP-covid

(or CASP-Commons 2020 experiment)

https://predictioncenter.org/caspcommons

Register for CASP_Commons

Your CASP_Commons groups /predictions

Released Target List

Participating Groups

Submit a prediction

ARCHIVE: submitted predictions (CASP-covid, 2020)

Model Accuracy Estimates [2018, NMR nominated targets]

COVID-19 MODELING RESULTS [2020]

Ambitious goal:

Utilize the strength of CASP community to generate SARS-Cov2 structures that will be useful to biologists for

- gaining further insight into the virus' structure and function
- identifying possible epitopes for vaccine development
- evaluating possible drug targeting strategies

CASP-covid

Other modeling efforts:

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 group: https://www.ipd.uw.edu/2020/02/rosettas-role-in-fighting-coronavirus/
Zhang group: https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/2019-nCov/
Michael Feig: https://github.com/feiglab/sars-cov-2-proteins
Jinbo Xu group

Strength of CASP:

- collect 3D models from a wide range of methods taking part in CASP
- employ our EMA community to generate accuracy estimates of the models
- identify the most promising models

CASP-covid

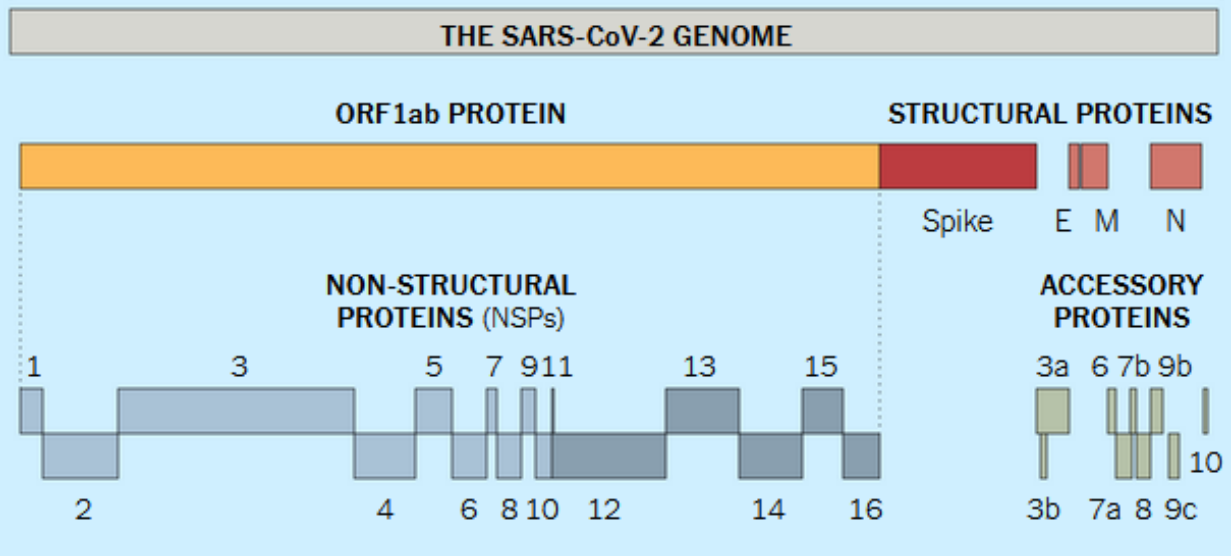
Working plan:

- identify targets
  (our efforts would be the most effectively used by concentrating on targets where there was no experimental structure available and where comparative modeling techniques cannot be used)
- collect initial models (round 1)
- run accuracy estimations
- compare models and discuss results
- call for revised models (round 2)
- refinement

CASP-covid

January 2020:

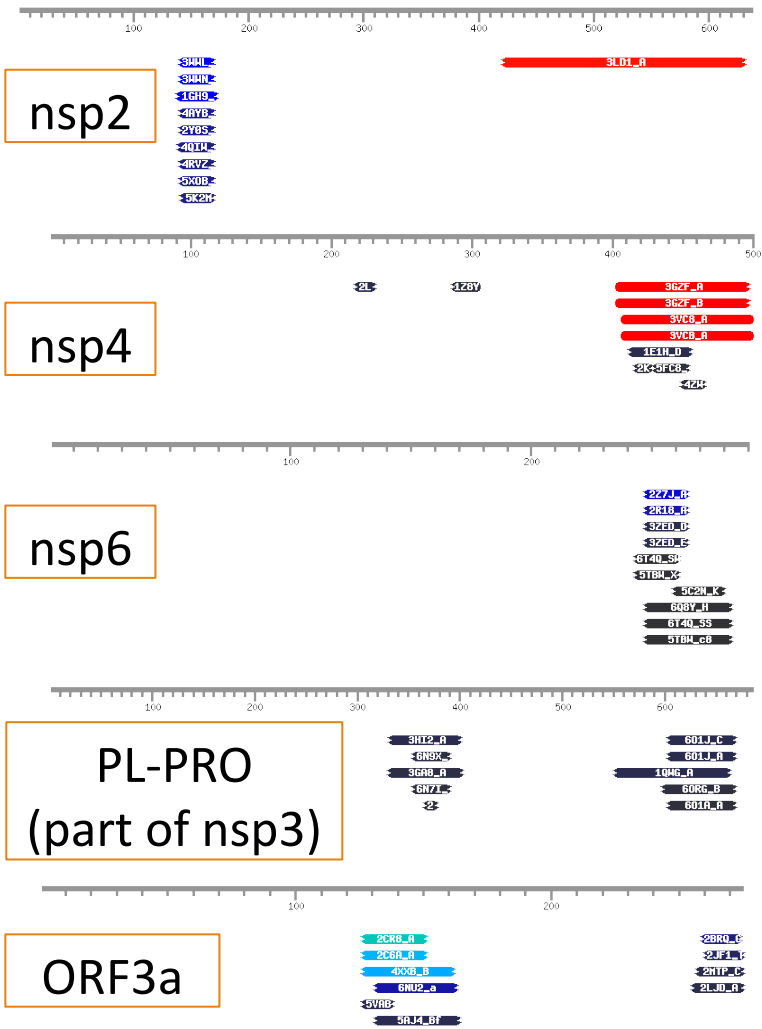
SARS-CoV2 sequence released

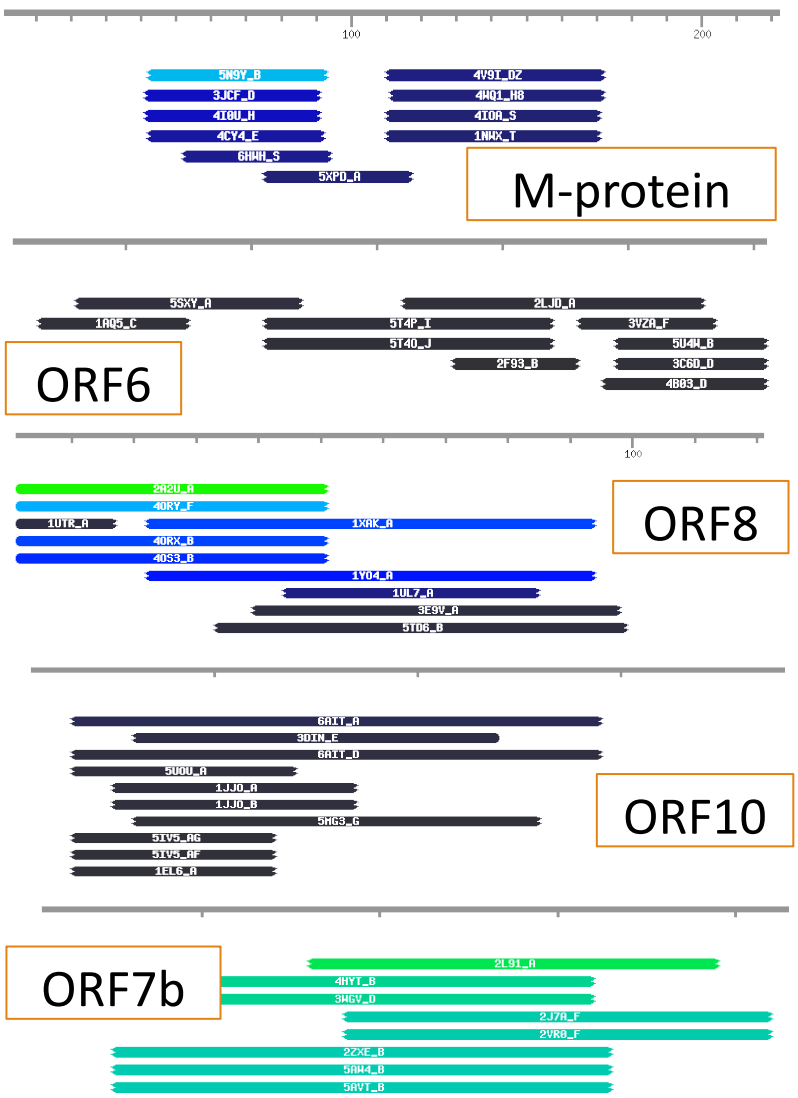


February 2020:

Analyzed the sequence and identified the most difficult for modeling proteins

CASP-covid





HHsearch results

CASP-covid


Target analysis
Model consensus
Model accuracy estimates
Target: C1905 (ORF3a)
C1905
Summary
Sequence templates
Signal peptide
Transmembrane regions
Secondary structure
Protein disorder
Domains
Templates
0
100
200
Position
Template
Not found
Found
Signal peptide
0
100
200
Position
Signal peptide
FALSE
Transmembrane regions
0
100
200
Position
Topology
Inside
outside
TMhelix
Secondary structure
0
100
200
Position
Type
E
C
H
Disordered region
0
100
200
Position
Disordered region
FALSE
TRUE
Domains
0
100
200
Position
Domain 1


Target analysis
Model consensus
Model accuracy estimates
Target: C1908 (ORF8)
C1908
Summary
Sequence templates
Signal peptide
Transmembrane regions
Secondary structure
Protein disorder
Domains
Templates
0
25
50
75
100
125
Position
Template
Not found
Found
Signal peptide
0
25
50
75
100
125
Position
Signal peptide
FALSE
TRUE
Transmembrane regions
0
25
50
75
100
125
Position
Topology
outside
Secondary structure
0
25
50
75
100
125
Position
Type
E
C
H
Disordered region
0
25
50
75
100
125
Position
Disordered region
FALSE
TRUE
Domains
0
25
50
75
100
125
Position
Domain 1

CASP-covid

March 13, 2020: CASP-covid experiment kicked off

https://predictioncenter.org/caspcommons/targetlist.cgi

#	Tar-id	Type	Res	Oligo State	Entry Date	Server Expiration	QA Expiration	Human Expiration	Description
1.	C1901	All groups	638		2020-03-13	2020-04-06	2020-04-12	2020-04-06	nsp2
2.	C1902	All groups	500		2020-03-13	2020-04-06	2020-04-12	2020-04-06	nsp4
3.	C1903	All groups	290		2020-03-13	2020-04-06	2020-04-12	2020-04-06	nsp6
4.	C1904	All groups	686		2020-03-13	2020-04-06	2020-04-12	2020-04-06	PL-PRO
5.	C1905	All groups	275		2020-03-13	2020-04-06	2020-04-12	2020-04-06	ORF3a PDB code 6xdc
6.	C1906	All groups	222		2020-03-13	2020-04-06	2020-04-12	2020-04-06	Membrane protein
7.	C1907	All groups	61		2020-03-13	2020-04-06	2020-04-12	2020-04-06	ORF6
8.	C1908	All groups	121		2020-03-13	2020-04-06	2020-04-12	2020-04-06	ORF8 PDB code 7jtl
9.	C1909	All groups	38		2020-03-13	2020-04-06	2020-04-12	2020-04-06	ORF10
10.	C1910	All groups	43		2020-03-13	2020-04-06	2020-04-12	2020-04-06	ORF7b

April 6, 2020: collected first-round models (>1,500 predictions from 52 groups)
3D predictions were immediately posted at the CASP Archive web place:
www.predictioncenter.org/download_area/CASPCOMMONS/2020_COVID-19/TS_predictions

April 12, 2020: collected 300 EMA from 30 methods
QA predictions were immediately posted at the CASP Archive web place:
www.predictioncenter.org/download_area/CASPCOMMONS/2020_COVID-19/QA_predictions

CASP-covid

Target analysis | Model consensus | Model accuracy estimates

Target: C1905 (ORF3a)

local score scale: (0.8; 1.0) (0.6; 0.8) (0.4; 0.6) (0.2; 0.4) (0.0; 0.2)

#	Model	Predictor	LDDT_cons	GDTTS_cons
1.	C1905TS413_1	TFold-server	0.275	0.201
2.	C1905TS213_3	McGuffin	0.287	0.200
3.	C1905TS213_4	McGuffin	0.285	0.200
4.	C1905TS213_5	McGuffin	0.287	0.200
5.	C1905TS213_2	McGuffin	0.286	0.199
6.	C1905TS213_1	McGuffin	0.287	0.199
7.	C1905TS401_1	FEIGLAB-R	0.298	0.198

Target: C1906 (Membrane protein)

local score scale: (0.8; 1.0) (0.6; 0.8) (0.4; 0.6) (0.2; 0.4) (0.0; 0.2)

#	Model	Predictor	LDDT_cons	GDTTS_cons
1	C1906TS301_1	FEIGLAB-S	0.438	0.327
2	C1906TS152_3	MULTICOM	0.434	0.327
3	C1906TS044_1	FEIGLAB	0.438	0.327
4	C1906TS213_4	McGuffin	0.438	0.324
5	C1906TS438_3	Destini	0.430	0.324
6	C1906TS213_3	McGuffin	0.439	0.323
7	C1906TS438_5	Destini	0.430	0.323
8	C1906TS213_1	McGuffin	0.437	0.323
9	C1906TS438_1	Destini	0.433	0.323
10	C1906TS152_2	MULTICOM	0.428	0.323
11	C1906TS213_2	McGuffin	0.439	0.323
12	C1906TS438_2	Destini	0.430	0.323
13	C1906TS152_1	MULTICOM	0.427	0.322
14	C1906TS413_1	TFold-server	0.418	0.322
15	C1906TS413_2	TFold-server	0.413	0.321
16	C1906TS413_3	TFold-server	0.424	0.321
17	C1906TS213_5	McGuffin	0.438	0.320
18	C1906TS438_4	Destini	0.428	0.320
19	C1906TS413_4	TFold-server	0.421	0.319
20	C1906TS299_1	FALCON-DeepFolder	0.441	0.317
21	C1906TS278_1	FALCON	0.441	0.317
22	C1906TS299_4	FALCON-DeepFolder	0.445	0.317
23	C1906TS299_5	FALCON-DeepFolder	0.442	0.317
24	C1906TS278_4	FALCON	0.445	0.317
25	C1906TS228_1	DellaCorteLab	0.439	0.317
26	C1906TS278_5	FALCON	0.442	0.317
27	C1906TS299_2	FALCON-DeepFolder	0.440	0.315
28	C1906TS278_2	FALCON	0.440	0.315

CASP-covid

Target analysis | Model consensus | Model accuracy estimates

Target: C1905 (ORF3a)

EMA method: 204 (ModFOLD8_rank)

Model: --all--

--all--
035 (angleQA)
131 (BAKER)
182 (Bhattacharya)
206 (Bhattacharya-QDeep)
451 (Bhattacharya-QDeepEx)
490 (bioinsilico_sbi)
377 (bioinsilico_sbi_PAIR)
438 (Destini)
037 (Elofsson)
363 (GCMQA)
210 (Kiharalab)
405 (Kiharalab_Z)

Ca-Ca distances: [0.0; 2.0) [2.0; 4.0) [4.0; 6.0) [6.0; 8.0) [8.0; ∞) N/A

#	TS Model	EMA Method	MQAS
1	C1905TS213_1	204(ModFOLD8_rank)	0.439202
2	C1905TS213_2	204(ModFOLD8_rank)	0.438264
3	C1905TS213_3	204(ModFOLD8_rank)	0.437396
4	C1905TS213_4	204(ModFOLD8_rank)	0.436867
5	C1905TS213_5	204(ModFOLD8_rank)	0.43496
6	C1905TS405_2	204(ModFOLD8_rank)	0.433448
7	C1905TS210_4	204(ModFOLD8_rank)	0.432677

Target: C1905 (ORF3a)

EMA method: 452 (VoroCNN)

Model: --all--

451 (Bhattacharya-QDeepEx)
490 (bioinsilico_sbi)
377 (bioinsilico_sbi_PAIR)
438 (Destini)
037 (Elofsson)
363 (GCMQA)
210 (Kiharalab)
405 (Kiharalab_Z)
227 (MESHI)
216 (MESHI_concensus)
056 (MESHI_EMA)
204 (ModFOLD8_rank)
389 (MULTICOM-CONSTRUCT)
114 (Ornate)
274 (ProQ3D)
307 (qa3dcnn_prof)
092 (QMEANDisCo)
081 (SBROD)
273 (Takeda-Shitaka-Lab)
452 (VoroCNN)

Ca-Ca distances: [0.0; 2.0) [2.0; 4.0) [4.0; 6.0) [6.0; 8.0) [8.0; ∞) N/A

#	TS Model	EMA Method	MQAS
1	C1905TS215_1	452(VoroCNN)	0.521
2	C1905TS215_2	452(VoroCNN)	0.515
3	C1905TS215_3	452(VoroCNN)	0.51
4	C1905TS405_3	452(VoroCNN)	0.508
5	C1905TS210_2	452(VoroCNN)	0.508
6	C1905TS210_4	452(VoroCNN)	0.496
7	C1905TS405_5	452(VoroCNN)	0.496
8	C1905TS405_1	452(VoroCNN)	0.496
9	C1905TS044_1	452(VoroCNN)	0.491
10	C1905TS301_1	452(VoroCNN)	0.491
11	C1905TS213_1	452(VoroCNN)	0.49
12	C1905TS230_2	452(VoroCNN)	0.486
13	C1905TS215_5	452(VoroCNN)	0.486

CASP-covid

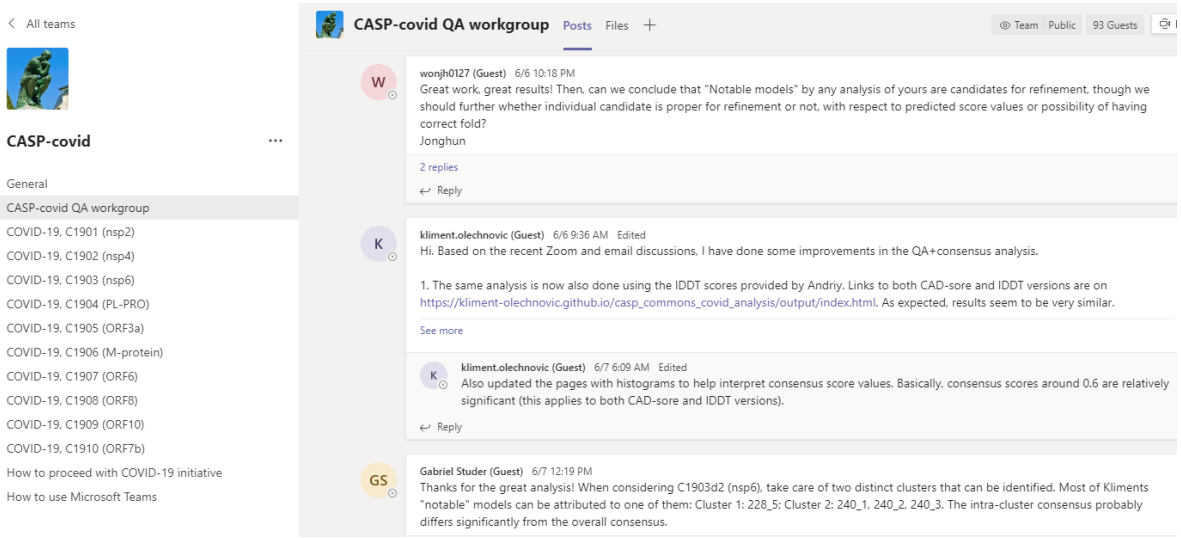
Conclusions from the preliminary analysis:

- disagreement between EMA methods
- often considerable variation between model structures
- visual comparison shows some consistent domain boundaries
- domain-based prediction may be beneficial

CASP-covid

Engaged the community in the discussion (2 weeks, second half of April):

- set up a Microsoft Teams site to discuss the results



- conducted two zoom sessions for the participants
- summarized in MS Teams notes on tentative domain boundaries, membrane regions, signal peptides, and under-utilized templates (John Moult)
- decided to run the second round of CASP-covid modeling on domains and selected targets where additional useful information was revealed (May 3-17)
- decided to select some of the second-round targets as CASP14 refinement targets (or even regular targets – used ORF8 as T1064)

CASP-covid
(2nd round)

Collected >1,500 models from 33 TS groups
>400 model accuracy estimates from 23 QA groups

The Round 2 models (TS and QA) were made available for download at
https://predictioncenter.org/download_area/CASPCOMMONS/2020_COVID-19/

The initial analysis (similar to round 1) was posted on the CASP-covid site:
https://predictioncenter.org/caspcommons/models_consensus2.cgi

Altogether in two rounds:

collected over 3,500 3D models and
700 accuracy estimates

CASP-covid
(2nd round)

Outreach:

Posted information about the available resource at the experimentalists' forums

Krzysztof Fidelis (Prediction Center) presented information at the SARS journal Club led by Guy Montelione

Engaged Randy Read, who is one of the leading scientists in the crystallography PHENIX community in advertising our results and connecting to exper. community

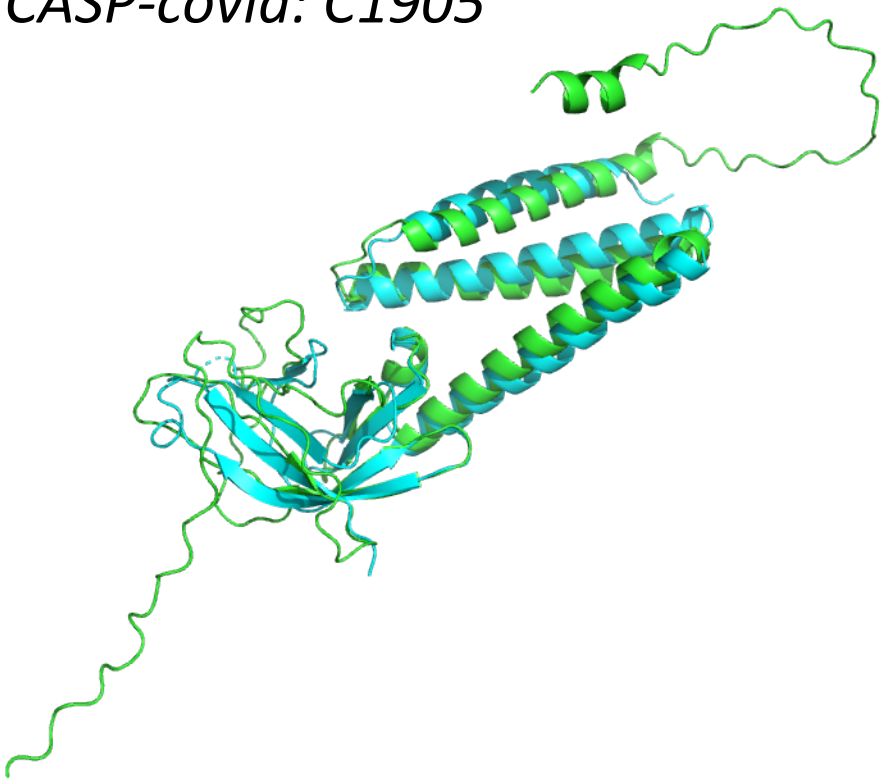
However the dream "If you build it, they will come" did not materialize. Is it because our data were unapproachable for biologists, they did not need them, or they did their own modeling?

Regardless, there are good news: two of CASP-covid targets got solved by now

CASP-covid
(evaluation)

Orf3a
(PDB: 6xdc, Brohawn lab at UC Berkeley)

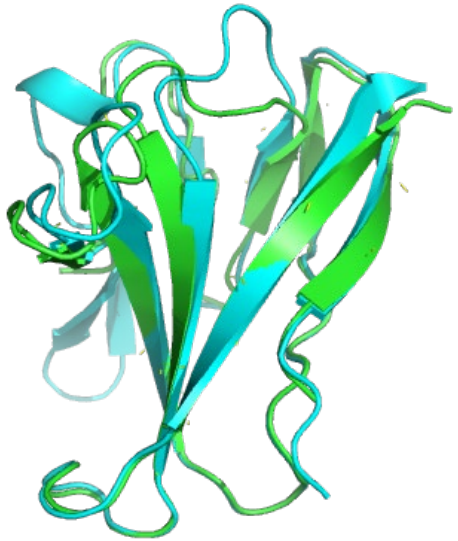
CASP-covid: C1905



Orf8
(PDB: 7jtl, Hurley lab at UC Berkeley)

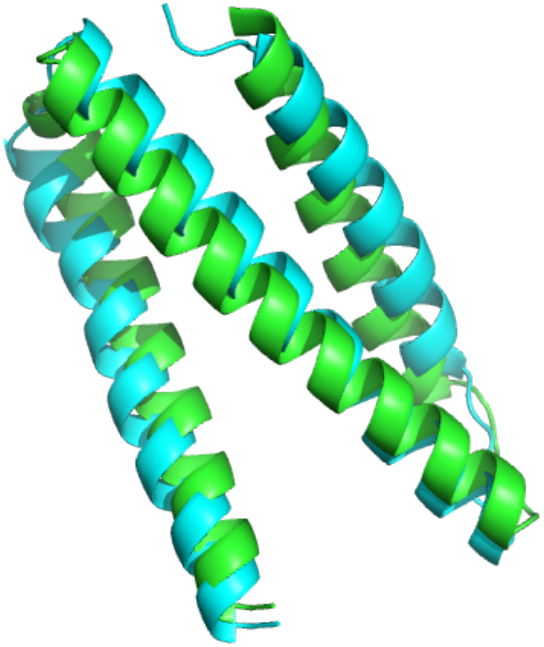
CASP-covid: C1908

CASP14: T1064



Orf3a
CASP-covid: C1905; C1905d1; C1905d2; C1905x2

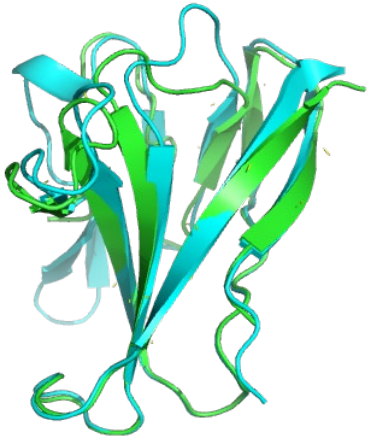
Best model on domain 1:
C1905TS156_2, GDT_TS=74.7



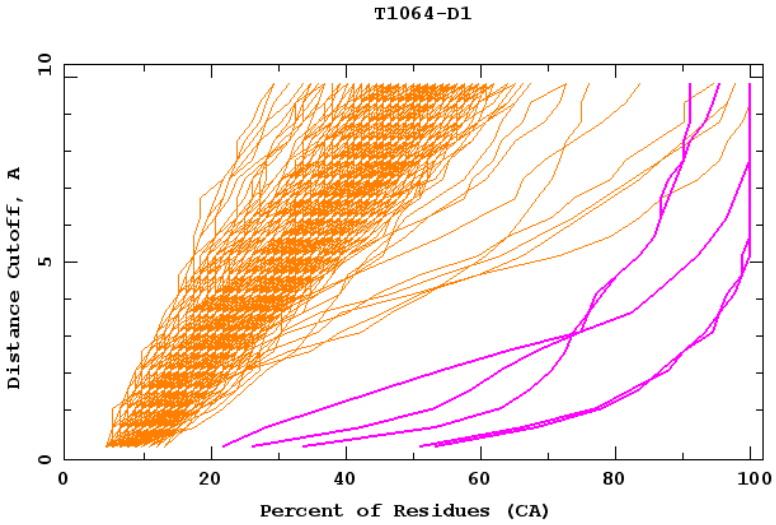
Best model on domain 2:
C1905TS156_3, GDT_TS=63.0

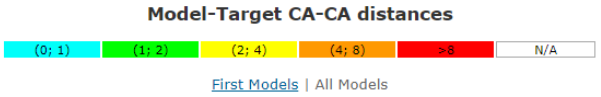
Group 156 in CASP-Commons = AlphaFold

Orf8
CASP-covid: C1908; CASP14: T1064



Best CASP14 model: T1064TS427_1
GDT_TS=87.0
(runner-up GDT_TS=42.9)





First Models | All Models

#	Model	gdt_ts	gdt_ha	gdc_sc	rmsd
1.	T1064TS427_1-D1	86.96	75.27	59.91	1.84
2.	T1064TS427_2-D1	86.14	73.91	57.25	1.93
3.	T1064TS427_3-D1	71.74	57.61	42.32	8.15
4.	T1064TS427_4-D1	66.85	50.82	34.59	5.37
5.	T1064TS427_5-D1	64.13	44.56	27.04	3.52
6.	T1064TS014_1-D1	42.94	21.20	8.55	5.32
7.	T1064TS014_3-D1	39.13	19.57	8.16	5.92
8.	T1064TS014_2-D1	38.86	19.02	5.00	6.20
9.	T1064TS170_4-D1	36.96	20.65	6.49	10.26
10.	T1064TS050_4-D1	36.41	17.12	2.36	7.33
11.	T1064TS131_1-D1	33.15	21.74	11.34	13.44
12.	T1064TS288_5-D1	31.52	17.12	3.90	10.55
13.	T1064TS304_2-D1	31.25	15.76	0.82	9.88
14.	T1064TS483_2-D1	30.71	19.29	8.27	12.63
15.	T1064TS483_1-D1	29.62	18.21	7.51	12.70
16.	T1064TS024_1-D1	29.35	18.21	5.46	14.30
17.	T1064TS024_5-D1	29.35	18.75	5.39	14.28
18.	T1064TS435_5-D1	29.35	19.02	7.77	14.86
19.	T1064TS024_4-D1	29.08	19.02	5.67	14.28
20.	T1064TS050_3-D1	28.80	15.22	4.96	10.96
21.	T1064TS024_2-D1	28.80	19.29	6.43	14.31
22.	T1064TS024_3-D1	28.80	18.21	4.76	14.26
23.	T1064TS324_2-D1	28.80	17.66	4.42	14.29
24.	T1064TS337_4-D1	27.99	17.12	5.89	13.64
25.	T1064TS435_2-D1	27.99	17.39	3.35	13.89
26.	T1064TS435_4-D1	27.99	18.75	7.32	14.96
27.	T1064TS200_2-D1	27.99	17.12	5.89	13.64
28.	T1064TS301_4-D1	27.72	17.12	6.38	14.65
29.	T1064TS061_4-D1	27.72	17.66	6.19	14.52
30.	T1064TS324_5-D1	27.72	18.20	6.71	15.79

Questions to panelists:

What did we learn?

What could we do better?

What do we do further with the results?

How to attract more attention of structural biology community?

Next steps?