CASP-covid
(or CASP-Commons 2020 experiment)

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

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
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
CASP-covid

March 13, 2020: CASP-covid experiment kicked off

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

April 6, 2020: collected first-round models (>1,500 predictions from 52 groups)
3D predictions were immediately posted at the CASP Archive web place:

April 12, 2020: collected 300 EMA from 30 methods
QA predictions were immediately posted at the CASP Archive web place:
CASP-covid
CASP-covid

### Target Analysis

**Target:** C1905 (ORF3a)  
**EMA method:** 204 (ModFOLD8_rank)  
**Model:**
- 035 (angleQA)  
- 131 (BAKER)  
- 182 (BhatCharyya)  
- 206 (BhatCharyya-QDeep)  
- 451 (BhatCharyya-QDeepEx)  
- 490 (bioniko_sbi)  
- 377 (bioniko_sbi_PAIR)  
- 438 (Destini)  
- 037 (Elofsson)  
- 363 (GMQA)  
- 210 (Kharab)  
- 405 (Kharab_2)  

### Model Accuracy Estimates

<table>
<thead>
<tr>
<th>#</th>
<th>TS Model</th>
<th>EMA Method</th>
<th>Co-Co distances</th>
<th>MQAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C1905TS213.1</td>
<td>204 (ModFOLD8_rank)</td>
<td>0.430202</td>
<td><img src="image1" alt="Heatmap" /></td>
</tr>
<tr>
<td>2</td>
<td>C1905TS213.2</td>
<td>204 (ModFOLD8_rank)</td>
<td>0.438284</td>
<td><img src="image2" alt="Heatmap" /></td>
</tr>
<tr>
<td>3</td>
<td>C1905TS213.3</td>
<td>204 (ModFOLD8_rank)</td>
<td>0.437356</td>
<td><img src="image3" alt="Heatmap" /></td>
</tr>
<tr>
<td>4</td>
<td>C1905TS213.4</td>
<td>204 (ModFOLD8_rank)</td>
<td>0.430687</td>
<td><img src="image4" alt="Heatmap" /></td>
</tr>
<tr>
<td>5</td>
<td>C1905TS213.5</td>
<td>204 (ModFOLD8_rank)</td>
<td>0.43486</td>
<td><img src="image5" alt="Heatmap" /></td>
</tr>
<tr>
<td>6</td>
<td>C1905TS405.2</td>
<td>204 (ModFOLD8_rank)</td>
<td>0.433448</td>
<td><img src="image6" alt="Heatmap" /></td>
</tr>
<tr>
<td>7</td>
<td>C1905TS210.4</td>
<td>204 (ModFOLD8_rank)</td>
<td>0.432677</td>
<td><img src="image7" alt="Heatmap" /></td>
</tr>
</tbody>
</table>

### Target Analysis

**Target:** C1905 (ORF3a)  
**EMA method:** 451 (BhatCharyya-QDeepEx)  
**Model:**
- 451 (BhatCharyya-QDeepEx)  
- 490 (bioniko_sbi)  
- 377 (bioniko_sbi_PAIR)  
- 438 (Destini)  
- 037 (Elofsson)  
- 363 (GMQA)  
- 210 (Kharab)  
- 405 (Kharab_2)  
- 227 (MESH)  
- 216 (MESH_concensus)  
- 056 (MESH_EMA)  
- 204 (ModFOLD8_rank)  
- 389 (MULTICOM-CONSTRUCT)  
- 114 (Ornate)  
- 274 (ProQ30)  
- 307 (q2somm_prof)  
- 092 (QM AND 600)  
- 081 (SRDOM)  
- 273 (Tsakda-Shitalo Lab)  
<table>
<thead>
<tr>
<th>#</th>
<th>TS Model</th>
<th>EMA Method</th>
<th>Co-Co distances</th>
<th>MQAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C1905TS215.1</td>
<td>452 (VoroCNN)</td>
<td>0.521</td>
<td><img src="image8" alt="Heatmap" /></td>
</tr>
<tr>
<td>2</td>
<td>C1905TS215.2</td>
<td>452 (VoroCNN)</td>
<td>0.515</td>
<td><img src="image9" alt="Heatmap" /></td>
</tr>
<tr>
<td>3</td>
<td>C1905TS215.3</td>
<td>452 (VoroCNN)</td>
<td>0.61</td>
<td><img src="image10" alt="Heatmap" /></td>
</tr>
<tr>
<td>4</td>
<td>C1905TS405.3</td>
<td>452 (VoroCNN)</td>
<td>0.585</td>
<td><img src="image11" alt="Heatmap" /></td>
</tr>
<tr>
<td>5</td>
<td>C1905TS210.2</td>
<td>452 (VoroCNN)</td>
<td>0.508</td>
<td><img src="image12" alt="Heatmap" /></td>
</tr>
<tr>
<td>6</td>
<td>C1905TS210.4</td>
<td>452 (VoroCNN)</td>
<td>0.496</td>
<td><img src="image13" alt="Heatmap" /></td>
</tr>
<tr>
<td>7</td>
<td>C1905TS105.5</td>
<td>452 (VoroCNN)</td>
<td>0.496</td>
<td><img src="image14" alt="Heatmap" /></td>
</tr>
<tr>
<td>8</td>
<td>C1905TS405.1</td>
<td>452 (VoroCNN)</td>
<td>0.496</td>
<td><img src="image15" alt="Heatmap" /></td>
</tr>
<tr>
<td>9</td>
<td>C1905TS344.1</td>
<td>452 (VoroCNN)</td>
<td>0.491</td>
<td><img src="image16" alt="Heatmap" /></td>
</tr>
<tr>
<td>10</td>
<td>C1905TS310.1</td>
<td>452 (VoroCNN)</td>
<td>0.491</td>
<td><img src="image17" alt="Heatmap" /></td>
</tr>
<tr>
<td>11</td>
<td>C1905TS213.4</td>
<td>452 (VoroCNN)</td>
<td>0.49</td>
<td><img src="image18" alt="Heatmap" /></td>
</tr>
<tr>
<td>12</td>
<td>C1905TS230.2</td>
<td>452 (VoroCNN)</td>
<td>0.486</td>
<td><img src="image19" alt="Heatmap" /></td>
</tr>
<tr>
<td>13</td>
<td>C1905TS215.6</td>
<td>452 (VoroCNN)</td>
<td>0.486</td>
<td><img src="image20" alt="Heatmap" /></td>
</tr>
<tr>
<td>14</td>
<td>C1905TS215.1</td>
<td>452 (VoroCNN)</td>
<td>0.487</td>
<td><img src="image21" alt="Heatmap" /></td>
</tr>
</tbody>
</table>
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
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)
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

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
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)
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?