Supplementary Information for “Theoretical study of the optical spectra of SARS-CoV-2 proteins”

Zhuo Li\textsuperscript{a} and Jonathan D. Hirst\textsuperscript{b}

\textsuperscript{a}School of Pharmaceutical Sciences, Jilin University, Changchun 130021, China
\textsuperscript{b}School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

*Corresponding author: Jonathan.Hirst@nottingham.ac.uk
Table S1 Experimental structures and computational models used. Numbers of total residue and aromatic amino acid are based on the actual coordinates. The secondary structure contents were calculated with DSSP.

<table>
<thead>
<tr>
<th>Protein</th>
<th>PDB ID / model*</th>
<th>Total residue</th>
<th>PHE</th>
<th>TYR</th>
<th>TRP</th>
<th>Helix</th>
<th>Strand</th>
<th>Turn</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental structures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Main protease</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6M03</td>
<td>306</td>
<td>17</td>
<td>11</td>
<td>3</td>
<td>26.1</td>
<td>28.8</td>
<td>19.3</td>
<td></td>
</tr>
<tr>
<td>6Y2E</td>
<td>306</td>
<td>17</td>
<td>11</td>
<td>3</td>
<td>25.8</td>
<td>28.8</td>
<td>19.6</td>
<td></td>
</tr>
<tr>
<td>6Y84**</td>
<td>304</td>
<td>16</td>
<td>11</td>
<td>3</td>
<td>26.0</td>
<td>28.9</td>
<td>19.4</td>
<td></td>
</tr>
<tr>
<td>5R8T</td>
<td>304</td>
<td>16</td>
<td>11</td>
<td>3</td>
<td>26.0</td>
<td>28.9</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>6Y2F</td>
<td>304</td>
<td>17</td>
<td>11</td>
<td>3</td>
<td>25.3</td>
<td>28.0</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>6Y2G</td>
<td>607</td>
<td>32</td>
<td>22</td>
<td>6</td>
<td>25.2</td>
<td>28.8</td>
<td>20.6</td>
<td></td>
</tr>
<tr>
<td><strong>Methyltransferase-nsp10</strong> (nsp16-nsp10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6W61</td>
<td>417</td>
<td>15</td>
<td>17</td>
<td>8</td>
<td>28.1</td>
<td>19.4</td>
<td>19.9</td>
<td></td>
</tr>
<tr>
<td>6W4H</td>
<td>418</td>
<td>15</td>
<td>17</td>
<td>8</td>
<td>27.5</td>
<td>20.3</td>
<td>19.6</td>
<td></td>
</tr>
<tr>
<td>6W75</td>
<td>834</td>
<td>30</td>
<td>34</td>
<td>16</td>
<td>26.5</td>
<td>20.6</td>
<td>19.9</td>
<td></td>
</tr>
<tr>
<td><strong>ADP ribose phosphatase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6W6Y</td>
<td>336</td>
<td>10</td>
<td>14</td>
<td>0</td>
<td>40.8</td>
<td>19.9</td>
<td>16.1</td>
<td></td>
</tr>
<tr>
<td>6W02</td>
<td>333</td>
<td>10</td>
<td>14</td>
<td>0</td>
<td>38.4</td>
<td>19.2</td>
<td>18.6</td>
<td></td>
</tr>
<tr>
<td>6VXS</td>
<td>335</td>
<td>10</td>
<td>14</td>
<td>0</td>
<td>40.3</td>
<td>20.3</td>
<td>17.0</td>
<td></td>
</tr>
<tr>
<td><strong>Papain-like proteinase</strong> (PL-PRO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6W9C</td>
<td>934</td>
<td>36</td>
<td>69</td>
<td>6</td>
<td>25.3</td>
<td>32.9</td>
<td>23.7</td>
<td></td>
</tr>
<tr>
<td><strong>nsp9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6W4B</td>
<td>226</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td>12.4</td>
<td>46.5</td>
<td>17.7</td>
<td></td>
</tr>
<tr>
<td>6W9Q</td>
<td>123</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>12.2</td>
<td>49.6</td>
<td>19.5</td>
<td></td>
</tr>
<tr>
<td><strong>Nucleocapsid phosphoprotein</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6VYO</td>
<td>502</td>
<td>16</td>
<td>28</td>
<td>12</td>
<td>2.4</td>
<td>24.9</td>
<td>30.3</td>
<td></td>
</tr>
<tr>
<td>6M3M</td>
<td>506</td>
<td>16</td>
<td>28</td>
<td>12</td>
<td>2.4</td>
<td>24.1</td>
<td>26.1</td>
<td></td>
</tr>
<tr>
<td><strong>RNA polymerase-nsp7-nsp8</strong> (nsp12-nsp7-nsp8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6M71**</td>
<td>1086</td>
<td>54</td>
<td>58</td>
<td>12</td>
<td>43.2</td>
<td>13.7</td>
<td>21.0</td>
<td></td>
</tr>
<tr>
<td>7BTF</td>
<td>1232</td>
<td>62</td>
<td>64</td>
<td>14</td>
<td>40.1</td>
<td>17.3</td>
<td>22.2</td>
<td></td>
</tr>
<tr>
<td><strong>Spike protein RBD-ACE2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6M0J</td>
<td>792</td>
<td>41</td>
<td>43</td>
<td>22</td>
<td>50.5</td>
<td>10.0</td>
<td>17.9</td>
<td></td>
</tr>
<tr>
<td>6LZG</td>
<td>792</td>
<td>41</td>
<td>43</td>
<td>22</td>
<td>49.4</td>
<td>10.1</td>
<td>17.9</td>
<td></td>
</tr>
<tr>
<td>6M17**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spike glycoprotein</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6VXX**</td>
<td>2951</td>
<td>195</td>
<td>126</td>
<td>18</td>
<td>21.5</td>
<td>33.8</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>6VYB**</td>
<td>2913</td>
<td>182</td>
<td>107</td>
<td>16</td>
<td>22.3</td>
<td>32.9</td>
<td>18.6</td>
<td></td>
</tr>
<tr>
<td><strong>Modelled structures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>M-protein</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feig</td>
<td>222</td>
<td>11</td>
<td>9</td>
<td>7</td>
<td>42.3</td>
<td>24.3</td>
<td>19.4</td>
<td></td>
</tr>
<tr>
<td>AlphaFold</td>
<td>192</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>47.9</td>
<td>30.2</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td><strong>nsp2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feig</td>
<td>638</td>
<td>30</td>
<td>19</td>
<td>7</td>
<td>32.8</td>
<td>10.0</td>
<td>23.8</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Method</td>
<td>AlphaFold</td>
<td>547</td>
<td>26</td>
<td>16</td>
<td>6</td>
<td>27.2</td>
<td>34.4</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>PL-PRO C-terminal</td>
<td>Feig</td>
<td>686</td>
<td>36</td>
<td>29</td>
<td>9</td>
<td>45.9</td>
<td>10.6</td>
<td>25.4</td>
</tr>
<tr>
<td>AlphaFold</td>
<td></td>
<td>357</td>
<td>16</td>
<td>14</td>
<td>3</td>
<td>44.5</td>
<td>20.2</td>
<td>19.3</td>
</tr>
<tr>
<td>Protein_3a</td>
<td>Feig</td>
<td>275</td>
<td>14</td>
<td>17</td>
<td>6</td>
<td>33.5</td>
<td>18.2</td>
<td>22.5</td>
</tr>
<tr>
<td>AlphaFold</td>
<td></td>
<td>196</td>
<td>11</td>
<td>16</td>
<td>6</td>
<td>46.4</td>
<td>16.3</td>
<td>11.7</td>
</tr>
<tr>
<td>nsp4</td>
<td>Feig</td>
<td>500</td>
<td>37</td>
<td>32</td>
<td>6</td>
<td>42.2</td>
<td>10.4</td>
<td>25.0</td>
</tr>
<tr>
<td>AlphaFold</td>
<td></td>
<td>489</td>
<td>37</td>
<td>32</td>
<td>6</td>
<td>43.1</td>
<td>20.9</td>
<td>20.9</td>
</tr>
<tr>
<td>nsp6</td>
<td>Feig</td>
<td>290</td>
<td>22</td>
<td>17</td>
<td>6</td>
<td>69.7</td>
<td>0.0</td>
<td>14.5</td>
</tr>
<tr>
<td>AlphaFold</td>
<td></td>
<td>278</td>
<td>22</td>
<td>17</td>
<td>6</td>
<td>71.9</td>
<td>0.7</td>
<td>10.4</td>
</tr>
<tr>
<td>ORF10</td>
<td>Feig</td>
<td>38</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0.0</td>
<td>26.3</td>
<td>26.3</td>
</tr>
<tr>
<td>ORF6</td>
<td>Feig</td>
<td>61</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>73.8</td>
<td>0.0</td>
<td>13.1</td>
</tr>
<tr>
<td>ORF7b</td>
<td>Feig</td>
<td>43</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>86.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>ORF8</td>
<td>Feig</td>
<td>121</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>8.3</td>
<td>33.1</td>
<td>19.8</td>
</tr>
<tr>
<td>Helicase</td>
<td>SwissModel (5wwp)</td>
<td>590</td>
<td>24</td>
<td>34</td>
<td>3</td>
<td>27.1</td>
<td>23.4</td>
<td>23.7</td>
</tr>
<tr>
<td>nsp14</td>
<td>SwissModel (5c8t)</td>
<td>527</td>
<td>30</td>
<td>25</td>
<td>10</td>
<td>24.5</td>
<td>19.5</td>
<td>26.8</td>
</tr>
<tr>
<td>Protein 7a</td>
<td>SwissModel (1xak)</td>
<td>68</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>0.0</td>
<td>60.3</td>
<td>23.5</td>
</tr>
<tr>
<td>Nucleoprotein (247-364)</td>
<td>SwissModel (2jw8)</td>
<td>118</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>33.9</td>
<td>5.1</td>
<td>28.8</td>
</tr>
<tr>
<td>PL-PRO (413-676)</td>
<td>SwissModel (2w2g)</td>
<td>264</td>
<td>5</td>
<td>14</td>
<td>1</td>
<td>34.8</td>
<td>19.7</td>
<td>22.7</td>
</tr>
<tr>
<td>PL-PRO (678-732)</td>
<td>SwissModel (2kqw)</td>
<td>55</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>21.8</td>
<td>38.2</td>
<td>23.6</td>
</tr>
<tr>
<td>PL-PRO (1089-1203)</td>
<td>SwissModel (2k87)</td>
<td>115</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>26.1</td>
<td>25.2</td>
<td>20.0</td>
</tr>
</tbody>
</table>

* The PDB ID of the homology model are listed in the bracelet next to the name of the method - SwissModel.

** Proteins whose MD simulations snapshots are available.

& These structures were only calculated for its near-UV CD spectra and compared with the computed spectra with the MD snapshots.
Fig S1 Calculated spectra with experimentally determined structures of methyltransferase-nsp10 complex. Left: far-UV CD; middle: near-UV CD; right: IR. PDB ID and colour code: 6W61 (black), 6W4H (blue), 6W75 (red).

Fig. S2 Calculated spectra with experimentally determined structures of ADP ribose phosphatase. Left: far-UV CD; middle: near-UV CD; right: IR. PDB ID and colour code: 6W6Y (black), 6W02 (blue), 6VXS (red).

Fig. S3 Calculated spectra with experimentally determined structure of papain-like proteinase (PDB ID 6W9C). Left: far-UV CD; middle: near-UV CD; right: IR.
Fig. S4 Calculated spectra with experimentally determined structures of nsp9. Left: far-UV CD; middle: near-UV CD; right: IR. PDB ID and colour code: 6W4B (black), 6W9Q (blue).

Fig. S5 Calculated spectra with experimentally determined structures of nucleocapsid phosphoprotein. Left: far-UV CD; middle: near-UV CD; right: IR. PDB ID and colour code: 6VYO (black), 6M3M (blue).

Fig. S6 Calculated spectra with experimentally determined structures of RNA polymerase-nsp7-nsp8 complex. Left: far-UV CD; middle: near-UV CD; right: IR. PDB ID and colour code: 6M71 (black), 7BTF (blue).
Fig. S7 Calculated spectra with experimentally determined structures of spike protein receptor binding domain and ACE2 complex. Left: far-UV CD; middle: near-UV CD; right: IR. PDB ID and colour code: 6M0J (black), 6LZG (blue).

Fig. S8 Calculated spectra of highly helical proteins of SARS-CoV-2. Models are from Feig group. Left: far-UV CD; middle: near-UV CD; right: IR. Colour code: ORF6 (black) and ORF7b (blue).

Fig. S9 Calculated spectra of over-40%-helical content proteins of SARS-CoV-2. Models are from Feig group (solid line) or AlphaFold (dashed line). Left: far-UV CD; middle: near-UV CD; right: IR. Colour code: PL-PRO C-terminal (black) and Protein_3a (blue).
Fig. S10 Calculated spectra of non-helical proteins of SARS-CoV-2. Models are from Feig group (solid line) or SwissModel (dotted line). Left: far-UV CD; middle: near-UV CD; right: IR. Colour code: ORF10 (black) and Protein 7a (blue).

Fig. S11 Calculated spectra of mix-α,β proteins from SARS-CoV-2. Models are from Feig group (solid line), AlphaFold (dashed line) or SwissModel (dotted line). Left: far-UV CD; middle: near-UV CD; right: IR. Color code: nsp2 (black), ORF8 (blue), helicase (green), nsp14 (red), nucleocapsid (orange), PL-PRO (413-676, brown), PL-PRO (678-732, cyan) and PL-PRO (1089-1203, magenta).
Fig. S12 Comparison of the calculated near-UV CD spectra of spike glycoprotein (left), spike protein receptor binding domain and ACE2 complex (middle) and nsp12-nsp7-nsp8 complex (right). The solid lines are calculated with the crystal structures while the dashed lines are calculated with MD simulation snapshots. PDB identifier of each protein includes: (left) 6VXX (red) is the closed state while 6VYB (black) is the opened state; (middle) 6M17; and (right) 6M71.

Fig. S13 6Y2F has only one Phenylalanine 137 within the binding pocket.