

This is the author's final, peer-reviewed manuscript as accepted for publication (AAM). The version presented here may differ from the published version, or version of record, available through the publisher's website. This version does not track changes, errata, or withdrawals on the publisher's site.

Outcomes of the EMDataResource cryo-EM Ligand Modeling Challenge

Catherine L. Lawson, Andriy Kryshtafovych, Grigore D. Pintilie, Stephen K. Burley, Jiří Černý, Vincent B. Chen, Paul Emsley, Alberto Gobbi, Andrzej Joachimiak, Sigrid Noreng, Michael G. Prisant, Randy J. Read, Jane S. Richardson, Alexis L. Rohou, Bohdan Schneider, Benjamin D. Sellers, Chenghua Shao, Elizabeth Sourial, Chris I. Williams, Christopher J. Williams, Ying Yang, Venkat Abbaraju, Pavel V. Afonine, Matthew L. Baker, Paul S. Bond, Tom L. Blundell, Tom Burnley, Arthur Campbell, Renzhi Cao, Jianlin Cheng, Grzegorz Chojnowski, K. D. Cowtan, Frank DiMaio, Reza Esmaeeli, Nabin Giri, Helmut Grubmüller, Soon Wen Hoh, Jie Hou, Corey F. Hryc, Carola Hunte, Maxim Igaev, Agnel P. Joseph, Wei-Chun Kao, Daisuke Kihara, Dilip Kumar, Lijun Lang, Sean Lin, Sai R. Maddhuri Venkata Subramaniya, Sumit Mittal, Arup Mondal, Nigel W. Moriarty, Andrew Muenks, Garib N. Murshudov, Robert A. Nicholls, Mateusz Olek, Colin M. Palmer, Alberto Perez, Emmi Pohjolainen, Karunakar R. Pothula, Christopher N. Rowley, Daipayan Sarkar, Luisa U. Schäfer, Christopher J. Schlicksup, Gunnar F. Schröder, Mrinal Shekhar, Dong Si, Abhishek Singharoy, Oleg V. Sobolev, Genki Terashi, Andrea C. Vaiana, Sundeep C. Vedithi, Jacob Verburt, Xiao Wang, Rangana Warshamanage, Martyn D. Winn, Simone Weyand, Keitaro Yamashita, Minglei Zhao, Michael F. Schmid, Helen M. Berman & Wah Chiu

Published version information

Citation: Lawson, C.L., Kryshtafovych, A., Pintilie, G.D. et al. Outcomes of the EMDataResource cryo-EM Ligand Modeling Challenge. *Nat Methods* 21, 1340–1348 (2024).

DOI: <https://doi.org/10.1038/s41592-024-02321-7>

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at:

<https://doi.org/10.1038/s41592-024-02321-7>

This version is made available in accordance with publisher policies. Please cite only the published version using the reference above. This is the citation assigned by the publisher at the time of issuing the AAM. Please check the publisher's website for any updates.

1 **Outcomes of the EMDDataResource Cryo-EM Ligand Modeling Challenge**

2 Catherine L. Lawson[1*], Andriy Kryshtafovych[2], Grigore D. Pintilie[3], Stephen K. Burley[1,4-6], Jiří
3 Černý[7], Vincent B. Chen[8], Paul Emsley[9], Alberto Gobbi[10,a], Andrzej Joachimiak[11], Sigrid
4 Noreng[12,b], Michael Prisant[8], Randy J. Read[13], Jane S. Richardson[8], Alexis L. Rohou[12], Bohdan
5 Schneider[7], Benjamin D. Sellers[10,c], Chenghua Shao[1], Elizabeth Sourial[14], Chris I. Williams[14],
6 Christopher J. Williams[8], Ying Yang[12], Venkat Abbaraju[1], Pavel V. Afonine[15], Matthew L.
7 Baker[16], Paul S. Bond[17], Tom L. Blundell[18], Tom Burnley[19], Arthur Campbell[20], Renzhi Cao[21],
8 Jianlin Cheng[22], Grzegorz Chojnowski[23], Kevin D. Cowtan[17], Frank DiMaio[24], Reza Esmaeeli[25],
9 Nabin Giri[22], Helmut Grubmüller[26], Soon Wen Hoh[17], Jie Hou[27], Corey F. Hryc[16], Carola
10 Hunte[28], Maxim Igaev[26], Agnel P. Joseph[19], Wei-Chun Kao[28], Daisuke Kihara[29,30], Dilip
11 Kumar[31,d], Lijun Lang[25,e], Sean Lin[32], Sai R. Maddhuri Venkata Subramaniya[30], Sumit
12 Mittal[33,34], Arup Mondal[25], Nigel W. Moriarty[15], Andrew Muenks[24], Garib N. Murshudov[9], Robert
13 A. Nicholls[9], Mateusz Olek[17,35], Colin M. Palmer[19], Alberto Perez[25], Emmi Pohjolainen[26],
14 Karunakar R. Pothula[36], Christopher N. Rowley[37], Daipayan Sarkar[29,33,f], Luisa U. Schäfer[36],
15 Christopher J. Schlicksup[15], Gunnar F. Schröder[36,38], Mrinal Shekhar[20,33], Dong Si[32], Abhishek
16 Singharoy[33], Oleg V. Sobolev[15], Genki Terashi[29], Andrea C. Vaiana[26,39], Sundeep C. Vedithi[18],
17 Jacob Verburgt[29], Xiao Wang[29], Rangana Warshamanage[9], Martyn D. Winn[19], Simone
18 Weyand[18], Keitaro Yamashita[9], Minglei Zhao[40], Michael F. Schmid[41], Helen M. Berman[4,42],
19 Wah Chiu[3,41*]

20 Primary Affiliations: [1]Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey,
21 Piscataway, NJ, USA, [2]Genome Center, University of California, Davis, CA, USA, [3]Departments of
22 Bioengineering and of Microbiology and Immunology, Stanford University, Stanford, CA, USA,
23 [4]Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey,
24 Piscataway, NJ, USA, [5]Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New
25 Jersey, New Brunswick, NJ USA, [6]San Diego Supercomputer Center, University of California San Diego,
26 La Jolla, CA USA, [7]Institute of Biotechnology, Czech Academy of Sciences, Vestec, CZ, [8]Department
27 of Biochemistry, Duke University, Durham NC, USA, [9]MRC Laboratory of Molecular Biology, Cambridge,
28 UK, [10]Discovery Chemistry, Genentech Inc, South San Francisco, USA, [11]Structural Biology Center, X-
29 ray Science Division, Argonne National Laboratory, Argonne, IL, USA, [12] Structural Biology, Genentech
30 Inc, South San Francisco, USA, [13]Department of Haematology, Cambridge Institute for Medical
31 Research, University of Cambridge, Cambridge, UK, [14]Chemical Computing Group, Montreal, Quebec,
32 CA, [15]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory,
33 Berkeley, CA, USA, [16]Department of Biochemistry and Molecular Biology, The University of Texas Health
34 Science Center at Houston, Houston, TX, USA, [17]York Structural Biology Laboratory, Department of
35 Chemistry, University of York, York, UK, [18]Department of Biochemistry, University of Cambridge,
36 Cambridge, UK, [19]Scientific Computing Department, UKRI Science and Technology Facilities Council,

37 Research Complex at Harwell, Didcot, UK, [20]Center for Development of Therapeutics, Broad Institute of
38 MIT and Harvard, Cambridge, MA, USA, [21]Department of Computer Science, Pacific Lutheran University,
39 Tacoma, WA, USA, [22]Department of Electrical Engineering and Computer Science, University of
40 Missouri, Columbia, MO, USA, [23]European Molecular Biology Laboratory, Hamburg Unit, Hamburg,
41 Germany, [24]Department of Biochemistry and Institute for Protein Design, University of Washington,
42 Seattle, WA, USA, [25]Department of Chemistry and Quantum Theory Project, University of Florida,
43 Gainesville, FL, USA, [26]Theoretical and Computational Biophysics Department, Max Planck Institute for
44 Multidisciplinary Sciences, Göttingen, Germany, [27]Department of Computer Science, Saint Louis
45 University, St. Louis, MO, USA, [28]Institute of Biochemistry and Molecular Biology, ZBMZ, Faculty of
46 Medicine and CIBSS - Centre for Integrative Biological Signalling Studies, University of Freiburg, 79104
47 Freiburg, Germany, [29]Department of Biological Sciences, Purdue University, West Lafayette, IN, USA,
48 [30]Department of Computer Science, Purdue University, West Lafayette, IN, USA, [31]Verna and Marrs
49 McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX,
50 USA, [32]Division of Computing & Software Systems, University of Washington, Bothell, WA, USA,
51 [33]Biodesign Institute, Arizona State University, Tempe, AZ, USA, [34]School of Advanced Sciences and
52 Languages, VIT Bhopal University, Bhopal, India, [35]Electron Bio-Imaging Centre, Diamond Light Source,
53 Harwell Science and Innovation Campus, Didcot, UK, [36]Institute of Biological Information Processing (IBI-
54 7: Structural Biochemistry) and Jülich Centre for Structural Biology (JuStruct), Forschungszentrum Jülich,
55 Jülich, Germany, [37]Department of Chemistry, Carleton University, Ottawa, ON, Canada , [38]Physics
56 Department, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [39]Nature's Toolbox (NTx), Rio
57 Rancho, NM, USA, [40]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago,
58 IL, USA, [41]Division of Cryo-EM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Menlo
59 Park, CA, USA, [42]Department of Quantitative and Computational Biology, University of Southern
60 California, Los Angeles, CA 90089, USA

61 Current Addresses: [a]Berlin, Germany, [b]Protein Science, Septerna, South San Francisco, USA,
62 [c]Computational Chemistry, Vilya, South San Francisco, USA, [d]Trivedi School of Biosciences, Ashoka
63 University, Sonipat, India, [e]Chinese University of Hong Kong, Hong Kong [f]MSU-DOE Plant Research
64 Laboratory, Michigan State University, East Lansing MI, USA and School of Molecular Sciences, Arizona
65 State University, Tempe, AZ, USA

66 *Contact Authors: Catherine L. Lawson: cathy.lawson@rutgers.edu; Wah Chiu: wahc@stanford.edu

67 **Abstract**

68 The EMDataResource Ligand Model Challenge aimed to assess the reliability and reproducibility of
69 modeling ligands bound to protein and protein/nucleic-acid complexes in cryogenic electron microscopy
70 (cryo-EM) maps determined at near-atomic (1.9-2.5 Å) resolution. Three published maps were selected as
71 targets: *E. coli* beta-galactosidase with inhibitor, SARS-CoV-2 RNA-dependent RNA polymerase with
72 covalently bound nucleotide analog, and SARS-CoV-2 ion channel ORF3a with bound lipid. Sixty-one
73 models were submitted from 17 independent research groups, each with supporting workflow details. We
74 found that (1) the quality of submitted ligand models and surrounding atoms varied, as judged by visual
75 inspection and quantification of local map quality, model-to-map fit, geometry, energetics, and contact
76 scores, and (2) a composite rather than a single score was needed to assess macromolecule+ligand model
77 quality. These observations lead us to recommend best practices for assessing cryo-EM structures of
78 liganded macromolecules reported at near-atomic resolution.

79 Cryogenic electron microscopy (Cryo-EM) has rapidly emerged as a powerful method for determining
80 structures of macromolecular complexes. It is complementary to macromolecular crystallography in its
81 ability to visualize macromolecules, and complexes thereof, of varying sizes and extents of structural
82 heterogeneity in 3D at near to full atomic resolution. The number of new structures determined by cryo-EM
83 has been steadily increasing, and with improved resolution (**Figure 1a**). Macromolecular complexes may
84 contain, in addition to larger components (*i.e.*, proteins or nucleic acids), smaller components such as
85 enzyme cofactors, substrates, analogs or inhibitors, medically relevant drug discovery candidates or
86 approved drugs, glycans, lipids, ions, or water molecules. Accurate modeling of ligands within their
87 macromolecular environment is important, as they can substantially influence larger-scale structure and
88 functions. As the number of novel ligands in cryo-EM-derived structures continues to increase rapidly
89 (**Figure 1b**), it becomes important to investigate how best to validate them to ensure optimal modeled ligand
90 quality using various measures such as fit of model-to-map, geometry scores of the ligand, and local
91 interactions with ions, waters, protein or nucleic acid components.

92 An international workshop on validation of ligands in crystallographic PDB depositions¹ held in 2015
93 identified several common problems, including weak experimental density, ligand atoms poorly placed,
94 incorrectly defined or misinterpreted chemical species, and inclusion of atoms not directly supported by
95 experimental evidence. The main outcome was a set of best practice recommendations for PDB depositors
96 and for the PDB archive. For PDB depositors, recommendations included providing unambiguous chemical
97 definitions for all ligands present in a structure, including hydrogen atoms, providing ligand geometry and
98 refinement restraints, clearly identifying atoms not supported by experimental evidence, providing the
99 experimental map used for modeling, and including comments explaining outliers. Recommendations for
100 PDB validation included providing informative images of ligands in their density; providing stick figure
101 diagrams indicating geometry outliers; identifying atoms not supported by experimental evidence; providing
102 quality assessment metrics for each identified ligand; and identifying possible protonation states. Most of
103 the workshop validation recommendations have been implemented in PDB validation reports, with ligand
104 geometric assessments implemented for all experimental methods²⁻⁴.

105 Since 2010, EMDataResource (EMDR) has organized multiple Challenge activities
106 (<https://challenges.emdataresource.org>) with the aim of bringing the cryo-EM community together to
107 address important questions regarding the reconstruction and interpretation of maps and map-derived
108 atomic coordinate models⁵. For each Challenge, a committee consisting of prominent experts is invited to
109 recommend targets and set goals. Each event has been conducted with the operational principles of
110 fairness, transparency, and openness, using modeler-blind assessments and open results, with a major
111 goal of promoting innovation.

112 In 2016, paired Map and Model Challenges invited participants to apply their novel algorithms/software to
113 reconstruct maps and to evaluate models at resolutions of 2.9-4.5 Å. The results were published in a 19-
114 article special journal issue⁶. By 2018, most participating groups had improved their pipelines, eliminating

115 many identified mistakes. The unique EMRinger map metric for sidechain-mainchain consistency⁷ was first
116 tested systematically in the 2016 Challenge and is now standard.

117 The 2019 Model Metrics Challenge evaluated models, while also evaluating the effectiveness of many
118 different coordinate-only and map-model fit metrics for 4 targets at 1.7-3.3 Å resolution. The results were
119 published in a single joint paper⁸. To streamline the challenge process, input of data from participants and
120 initial assessment pipelines were automated, and comprehensive statistics, visualizations of scores and
121 comparisons were made available. The CaBLAM multi-residue mainchain metric⁹, introduced in 2016, was
122 shown in the 2019 Challenge to be the score most highly correlated with measures of match-to-target. The
123 Q score¹⁰, inspired and introduced by the 2019 Challenge, has now been adopted by the wwPDB Validation
124 System used at deposition as well as in the detailed validation report¹¹.

125 The 2021 Ligand Model challenge brought together research and industry groups to evaluate and discuss
126 available measures and tools for ligand quality assessment. Many of the issues identified for
127 crystallographic structures in the 2015 ligand workshop were also expected to occur in cryo-EM structures
128 with modeled ligands, but with additional considerations unique to cryo-EM. Targets were chosen from
129 publicly available maps with sufficient resolution to theoretically allow de-novo ligand modeling, include
130 diverse components such as protein and RNA, and have current interest and relevance. The objectives set
131 out were to identify 1) methods for modeling such ligands and 2) metrics to evaluate map-model fit,
132 stereochemical geometry, and chemically sensible interactions between the ligand and protein or RNA
133 component. We describe here the overall design and outcomes of the EMDR Ligand Challenge,
134 recommendations for the cryo-EM community based on currently available assessment methods, and what
135 is needed for the future.

136 Results

137 Challenge Design

138 Three Cryo-EM map targets were chosen based on the following criteria: recently published with resolution
139 better than 3 Å, maps released in the Electron Microscopy Databank (EMDB), associated coordinates in
140 the Protein Data Bank (PDB), small molecules present (ligands, water, metal ions, detergent, and/or lipid),
141 and having current topical relevance (**Figure 2 panels A-C**):

- 142 • **Target 1:** 1.9 Å *E. coli* β-Galactosidase (**β-Gal**) in complex with inhibitor 2-phenylethyl 1-thio-beta-
143 D-galactopyranoside (PETG) with PDB Chemical Composition Dictionary (CCD) id **PTQ**, EMDB
144 map entry EMD-7770, PDB reference model 6CVM¹²
- 145 • **Target 2:** 2.5 Å SARS-CoV-2 RNA-dependent RNA polymerase (**RNAP**) with the
146 pharmacologically active, nucleotide form of the prodrug remdesivir (CCD id **F86**) covalently-bound
147 to RNA, EMD-30210, PDB reference model 7BV2^{13 14}

- 148 • **Target 3:** 2.1 Å SARS-CoV-2 Open Reading Frame 3a (**ORF3a**) putative ion channel in complex
149 with 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine phospholipid (CCD id **PEE**), EMD-22898,
150 PDB reference model 7KJR¹⁵

151 Next, modeling teams were solicited via emails to multiple bulletin board lists and were asked to register,
152 generate and upload optimized models for each Target, following provided guidelines (see **Online**
153 **Methods**). A total of 61 independently determined models were contributed by seventeen teams from
154 different institutions (ids **EM001-EM017**), with workflow details collected for each (see summary in **Table 1**
155 and **Supplementary Data S1, S2** for details).

156 **Model Assessments**

157 Submitted and PDB reference models for each target were evaluated by passing them through the EMDR
158 Model Challenge validation pipeline^{8,16}. Individual scores were obtained for many different sets of metrics,
159 with a new Ligand analysis track added to the existing Fit-to-Map, Coordinates-only, Comparison-to-
160 Reference, and Comparison-among-Models tracks.

161 Global Fit-to-Map metrics included Map-Model Fourier shell correlation (FSC)¹⁷, Atom Inclusion¹⁸,
162 EMRinger⁷, density-based correlation scores from TEMPy¹⁹, Phenix²⁰ and Q-score¹⁰.

163 Overall Coordinates-only quality was evaluated using Clashscore, Rotamer outliers, Ramachandran
164 outliers, and CaBLAM from MolProbity^{9,21}, as well as standard geometry measures (e.g., bond, chirality,
165 planarity) from Phenix²². Davis-QA, a measure used in critical assessment of protein structure prediction
166 (CASP) competitions, was used to assess similarity among submitted models²³.

167 Assessment teams contributed a wide variety of ligand-specific assessments (**Table 2**, ids **AT01-AT07**)
168 including ligand, ligand environment, solvent, and RNA-specific analyses. AT01 used Mogul²⁴ to evaluate
169 ligand covalent geometry as implemented in the wwPDB validation process^{2,4}, with inclusion of a novel
170 composite ligand geometry ranking score²⁵. AT02 evaluated model ligands using Coot²⁶ and AceDRG²⁷.
171 AT03 evaluated RNA conformation with DNATCO^{28,29} and solvent atom placement around protein residues
172 using water distributions^{30,31}. AT04 analyzed ligand all-atom contacts with Molprobity Probescore⁹, and ion
173 and water placements using UnDowser³². AT05 scored ligand placements using density fields derived from
174 pharmacophore consensus field analysis³³, a method utilized in computer-aided drug design to identify and
175 extract possible interactions between a ligand–receptor complex based on steric and electronic features³⁴.
176 AT06 examined ligand strain energies using both molecular mechanics and neural net potential energy
177 strategies^{35–37}, where strain energy is the calculated difference in energy between the modeled
178 conformation and the lowest energy conformation in solution. AT07 prepared Q-score analyses¹⁰ for model-
179 fit-to-map of whole models, protein, ligands, and water, as well as ligand plus protein and/or nucleic acid
180 polymer atoms in the immediate vicinity of the ligand (LIVQ). Assessor scores are available online at [model-
181 compare.emdataresource.org](http://model-compare.emdataresource.org); results are briefly outlined below.

182 Outcomes

183 The modeled ligands from each of the submissions are shown superimposed with their corresponding map
184 density in **Figure 2 panels D-F**; selected ligand and whole-model score distributions are shown for all three
185 targets in **Figure 3**. The full set of pipeline and assessment team scores and their definitions are provided
186 in **Supplementary Data S3**. The superimposed views and score distributions demonstrate that the methods
187 utilized by the modeling teams produced a range of ligand positions and conformations.

188 *Overall model scoring.* With regards to overall Fit-to-Map evaluation, the majority of submitted models
189 scored very similarly to PDB reference models for all targets, both in terms of the overall map-model FSC¹⁷
190 and protein Q-score¹⁰ (**Figure 3, rows 9 and 11**). For Targets 2 and 3, several teams modestly improved
191 upon EMRinger score⁷ (**Figure 3, columns 2 and 3, row 10**). With regards to overall Coordinates-only
192 evaluation, many teams were able to improve upon PDB reference models for all targets in terms of
193 Clashscore³² and CaBLAM³², metrics that identify steric clashes and evaluate protein backbone geometry,
194 respectively (**Figure 3, rows 6, 7**).

195 *Ligand and ligand environment scoring.* Ligand and ligand environment evaluation methods were
196 challenged by missing atoms in some submissions, the covalently bound ligand (Target 2), and presence
197 of charged ligands (Targets 2 and 3). In terms of ligand-specific Fit-to-Map (Ligand Q-score), many teams
198 made improvements relative to the PDB reference model of Target 1, but scored similarly or worse than
199 the PDB reference of Targets 2 and 3 (**Figure 3, row 1**). In terms of covalent geometry (Mogul)^{24,25}, many
200 ligands in the submitted models were improved relative to references for Targets 1 and 3, while results
201 were mixed for Target 2 (**Figure 3, row 5**). With respect to calculated ligand strain energy and
202 pharmacophore ligand environment modeling, many of the submitted models were improved relative to
203 references for Targets 1 and 2, but some poses were less favorable (**Figure 3, rows 3-4**). Ligand strain
204 energy qualitatively should be less than 3 kcal/mol with minor relaxation using the sampling and scoring as
205 described in **Online Methods**. Only a subset of submitting groups carefully considered treatment of ions
206 (**Extended Data Figure 5**).

207 *Nucleic Acid scoring.* Target 2's RNA (a typical A-form double helix, with two unpaired nucleotides at the 5'
208 end of the template strand) had close to expected geometries for most submitted models as assessed by
209 DNATCO nucleic acid Confal scores²⁹ (**Figure 3, row 8**). Values of torsion angles in the dinucleotide units
210 assigned to DNATCO NtC classes agreed with expected distributions including sugar ring torsions that
211 define pucker. Note that prior to running this Challenge, Target 2's reference model (PDB 7bv2) had been
212 re-versioned by the deposition authors and re-released by the PDB with several corrections to sequence,
213 RNA conformation, and CaBLAM outliers³⁸, thus limiting scope for model improvement.

214 *Submitted Model rankings.* To evaluate and rank quality of ligand Fit-to-Map *within the context of the*
215 *macromolecular complex*, we developed a novel score, the **Ligand + Immediate Vicinity Q-score (LIVQ)**,
216 which averages Q-scores of non-hydrogen atoms of the ligand together with all non-hydrogen polymer
217 atoms in the immediate vicinity of ligand. A distance cutoff of 5 Å was chosen to define the immediate

218 vicinity of the ligand for model ranking purposes (**LIVQ5, Figure 4A-C**); extension to 10 Å yielded similar
219 results (**LIVQ10, Extended Data Figure 2**). The results of the analysis show that for each target there are
220 several models that exhibit very good model-to-map fit comparable to that of reference PDB-deposited
221 models (**Figure 4A-C**, blue bars). Nine, two and three submitted models respectively on Targets 1-3 score
222 better than the corresponding deposited reference model.

223 *Group rankings.* Overall ranking of participating groups (**Figure 4D**) employed a combination of LIVQ5 and
224 MolProbity score, itself a weighted function of clashes, Ramachandran favored, and rotamer outliers⁹.
225 LIVQ5 was weighted higher than stereochemical plausibility, similar to the approach customarily used in
226 CASP³⁹:

$$227 \quad rank = \sum_{target=1\dots3} (0.8 * z.LIVQ5_{target} + 0.2 * z.MolProbity_{target})$$

228 where *z.metric* is the number of standard deviations relative to the mean of the score distribution for all
229 models from each group on the selected target according to the selected metric. Overall, group EM003
230 (DiMaio) had the best relative performance by this ranking criterion, being the only group that outscored all
231 deposited reference PDB models (**Figure 4A-C**).

232 *Alternate group rankings.* The model-compare website Group Ranking calculator enables users to explore
233 other possible ranking formulas: z-scores of up to 40 different individual metrics can be selected for
234 inclusion with adjustable weighting. **Extended Data Figure 3** illustrates an alternate ranking method based
235 upon thirteen different metrics including ligand, ligand+environment, full model coordinates-only and full
236 model fit-to-map. By this alternate method, five groups ranked higher than PDB reference models: EM010
237 (Chojnowski), EM008 (Emsley), EM012 (Palmer), EM003 (DiMaio), and EM009 (Moriarty), and one
238 performed very close to reference, EM011 (Igaev).

239 *Ligand Quality.* The ligand environment for the reference models and the best submitted models is
240 compared for each target in **Figure 5**.

241 For Target 1 (β-Gal, **Fig 5 A,D**), the PTQ ligand O5 atom connected to the sugar ring is situated at the
242 bottom of the binding pocket in the reference model and in eight submitted models, whereas in the top-
243 scoring model, as well as five other submitted models, the sugar ring is flipped with oxygen O5 situated at
244 the top. The flipped ligand fits the density better and has more optimal interatomic distances to water and
245 protein atoms for hydrogen-bonding, with O5 H-bonded to a coordinated water of the nearby magnesium
246 ion (see Supplemental section S5). The density shape does not preclude the possibility that both original
247 and flipped conformations are present, each with partial occupancy, and probescores for the two states are
248 nearly identical (**Extended Data Fig 4A**).

249 For Target 2 (RNAP; **Fig 5 B,E**), the F86 ligand is very similar for the deposited and top-scoring model,
250 though distances to base-paired U10 are slightly different. F86 probescores varied greatly across models,
251 with the reference at 10.1, model EM008_1 at 39.9, and the worst model at -106.9 (**Extended Data Figure**

252 **4).** Many models did not correctly create the RNA polymer – F86 (remdesivir) covalent bond. In addition,
253 only five models indicated partial occupancy for F86, yet the map density for F86 and its paired base is
254 almost exactly half that of adjacent base pairs (**Extended Data Fig 4B**), indicating 50% occupancy.

255 In the case of Target 3 (ORF3 ion channel; **Fig 5 C,F**), the PEE ligand has similar interactions to nearby
256 atoms and placed water molecules, though with slightly different interatomic distances. The head-group
257 amino N atom (which has no close contacts within 4Å) points up in the deposited model but away from the
258 camera view in the top-scoring model (**Fig 5F**). The long lipid tails of PEE have lower density, with
259 confusingly interlaced and gapped connectivity that indicates disorder; the ensemble of all PEE ligand
260 models shown in **Fig 2F** may be a more meaningful representation than any one individual model.

261 **Discussion**

262 The selected targets for the Ligand Challenge are some of the first structures deposited and released into
263 PDB that contain ligands modeled into cryo-EM maps with resolution of 3 Å or better. Our Challenge results
264 revealed that a deposited PDB model's ligand and local ligand environment may not be fully optimal in
265 terms of concurrent Fit-to-Map and Coordinates-only measures. For all three targets and especially for
266 Target 1, adjustments in the ligand and/or ligand environment could be made to the deposited reference
267 model that improved one or more validation criteria, as demonstrated by several modeler groups. Most of
268 the submitted models were in the “better” range, where tiny differences in measured scores become
269 inconsequential. In our previous Challenge, we showed that overall Fit-to-Map and Coordinates-only
270 metrics are orthogonal measures⁸; here we see that at a local level, ligand/ligand-environment Fit-to-Map
271 and Coordinates-only metrics are similarly independent (**Figure 3, Extended Data Figure 3B,**
272 **Supplementary Data S3**). In other words, ligands that fit quite well into density may not be optimized with
273 respect to ligand coordinates-only validation criteria, and vice versa.

274 Based on our analyses and experiences running the Challenge, we make the following recommendations.

275 **Recommendation 1, regarding validation of the macromolecular models:** *For ligand-macromolecular*
276 *complexes, the macromolecular model should be subject to standard geometric checks as done for X-ray*
277 *crystallographic based models¹. These include standard covalent geometry checks and MolProbity*
278 *evaluation, including CaBLAM, clashscore^{9,21,32}. Sugar pucker and DNATCO conformational analysis^{28,29}*
279 *should be checked for nucleic acid components. The macromolecular model-map fit should be evaluated*
280 *by EM Ringer⁷, Q score¹⁰, and FSC¹⁷. Serious local outliers (which usually indicate an incorrect local*
281 *conformation) should be emphasized, rather than overall average scores.*

282 The individual MolProbity scores, CaBLAM and clashscore have more utility for validation of protein
283 conformation than overall MolProbity score which incorporates Ramachandran and side-chain rotamer
284 quality, since cryo-EM model refinement includes these as restraints.

285 **Recommendation 2, regarding validation of ligand models:** *Ligands in macromolecular complexes*
286 *should conform to standard covalent geometry measures (bond lengths, angles, planarity, chirality) as*

287 recommended by the wwPDB validation report^{2,4}. Additional checks that should be applied to ligands
288 include fit to density using methods applicable to cryo-EM such as Q-score, occupancy (density strength,
289 both absolute and relative to surroundings), and identification of missing atoms, including any surrounding
290 ions.

291 Ligand energetics should also be examined. Ligand models should be assessed for their strain energy (the
292 calculated difference in energy between the modeled conformation and the lowest energy conformation in
293 solution) to identify improbable model geometries and lower energy alternatives^{35,36}. Other methods can be
294 used but may have different thresholds due to variation in absolute energy values. Strain energy
295 calculations using neural net potentials offer speed close to force fields with the accuracy of QM calculations
296 and are predicted to play a primary role in identifying accurate strain energies in the future. More research
297 is needed to evaluate the overall utility of these deep learning novel methods.

298 **Recommendation 3, regarding validation of ligand environment:** *The detailed interaction of the ligand*
299 *with its binding site is of great importance and should be assessed by several independent metrics.*
300 *Pharmacophore modeling³³ is an optimized and time-tested energetic measure for how well the site would*
301 *bind the specific ligand. LIVQ scores, introduced here, measure the density fit of the surrounding residues*
302 *as well as the ligand itself. Probescore³² both quantifies and identifies specific all-atom contacts of H-bond,*
303 *clash, and van der Waals interactions. All three types of measures should be taken into account. If the*
304 *ligand model shows only weak interaction with its environment, the model is not right.*

305 During the virtual wrap-up workshop, modelers and assessors shared their experiences and strategies to
306 identify/assess the correct pose for the ligand based on the cryo-EM density maps. It was noted that the
307 local map resolution for a ligand can be worse than the overall map resolution. As one objective measure,
308 Q-scores were found to be lower for ligands in the best submitted models than for the nearby environment
309 (**Table 3**). Factors that may affect resolvability of local ligand map features include incomplete occupancy,
310 multiple conformations/poses present, regions of ligand flexibility or disorder, chemical modifications, and
311 radiation damage.

312 **Recommendation 4, regarding organization of future Challenges:** *Future cryo-EM Model Challenges*
313 *should be organized similarly to the well-established CASP and CAPRI challenge events of the X-ray*
314 *crystallography and prediction communities²³, with incorporation of automated checks and immediate*
315 *author feedback on all model submissions.*

316 **Recommendation 5, regarding topics for future Challenges:** *For future Challenge topics, consider*
317 *validation of RNA models, including identification of RNA-associated ions, owing to the rapidly rising*
318 *numbers of RNA-containing cryo-EM structures⁴⁰⁻⁴². We also recommend maps determined in the 3.5-to-*
319 *10 Å resolution range be considered as future targets to reflect the rapid rise in depositions of maps from*
320 *subtomogram averaging of components in cell tomograms⁴³⁻⁴⁵. There are very few validation tools for that*
321 *resolution range.*

322 Online Methods

323 Challenge process and organization

324 The Ligand Model Challenge process closely followed the streamlined procedure adopted in the previous
325 Model Metrics Challenge⁸. In the fall of 2020, a panel of advisors with expertise in cryo-EM methods, ligand
326 modeling and/or ligand model assessment was recruited (J. Černý, P. Emsley, A. Joachimiak, J.
327 Richardson, R. Read, A. Rohou, B. Schneider). The panel worked with EMDR team members to develop
328 the challenge goals and guidelines, to identify suitable ligand-containing reference models from the PDB
329 with cryo-EM map targets from EMDB, and to recommend metrics to be calculated for each submitted
330 model.

331 The main stated goal was to identify metrics most suitable for evaluating and comparing fit of ligands in
332 atomic coordinate models into cryo-EM maps with 3.0 Å or better reported overall resolution. The specific
333 focus areas for assessor teams suggested by the expert panel were: (1) Geometry and fit to map of small
334 molecules including ligands, water, metal ions, detergent, lipid, nanodiscs. (2) Model geometry (including
335 backbone and side-chain conformations, clashes) in the neighborhood surrounding the small molecules.
336 (3) Local model Fit-to-Map density per residue and per atom. (4) Resolvability at residue or atom-level. (5)
337 Atomic Displacement parameters (B-factors) recommended optimization practice. A key question to be
338 answered: How reliable are ligands/waters/ions built into cryo-EM maps? Can they be placed automatically
339 or is manual intervention needed?

340 Modeling teams were tasked with creating and uploading their optimized model for each Target Map. The
341 challenge rules and guidance were as follows: (1) Submitted models should be as complete and as accurate
342 as possible (i.e., close to publication-ready), with atomic coordinates and atomic displacement parameters
343 for all model components. (2) Submitted models must use the deposited PDB Reference Model's residue,
344 ligand, and chain numbering/labeling for all shared model components. (3) Ligands should ideally be
345 deleted and refitted independently. (4) Additional polymer residues should be labeled according to the
346 Reference Model's sequence/residue numbering/chain ids. (5) If additional waters/ions/ligands are
347 included, they should be labeled with unique chain ids. (6) If predicted hydrogen atom positions are part of
348 the modeling process, hydrogens should be included in the submitted coordinates. (7) Models are expected
349 to adhere to the reconstruction's point symmetry (D2 for Target 1, C1 for Target 2, C2 for Target 3).

350 Members of cryo-EM and modeling communities were invited to participate in February 2021 and details
351 were posted at challenges.emdataresource.org. Models were submitted by participant teams between
352 March 1 and April 15. For each submitted model, metadata describing the full modeling workflow were
353 collected via a Drupal webform (see **Supplementary Data S1, S2**), and coordinates were uploaded and
354 converted to PDBx/mmCIF format using PDBextract⁴⁶. Model coordinates were then processed for
355 atom/residue ordering and nomenclature consistency using PDB annotation software (Feng Z., [https://sw-
356 tools.rcsb.org/apps/MAXIT](https://sw-tools.rcsb.org/apps/MAXIT)) and additionally checked for sequence consistency, ligand atom naming, and

357 correct position relative to the designated target map. Models were then evaluated as described below (see
358 Model evaluation system).

359 In mid-April 2021, models, workflows and initial calculated scores were made publicly available for
360 evaluation, blinded to modeler team identity and software used. In the period mid-April to mid-May,
361 evaluators discovered several problems with the submitted models that blocked assessment software from
362 completing calculations. The primary issue identified was inconsistent ligand atom naming. Approximately
363 half of all submitted models had to be revised to make atom names consistent with the deposited reference
364 models (see Challenge rule (2) above). Corrected coordinate files were provided by the submitting modeler
365 teams, which were then re-processed as described above and re-released to evaluators.

366 A virtual 3-day (~4 hours/day) workshop was held in mid-July 2021 to review the Challenge results. All
367 modeling participants were invited to attend remotely and present overviews of their modeling processes
368 and/or assessment strategies. Recommendations were made for additional evaluations of the submitted
369 models as well as for future challenges. Modeler teams, workflows and software were unblinded during the
370 workshop.

371 Data sources and Modeling

372 Target maps were obtained from EM Data Bank⁴⁷. **Target 1** E. coli β -Galactosidase/PETG¹²: EMD-7770,
373 **Target 2** SARS-CoV-2 RNA-dependent RNA polymerase/Remdesivir¹³: EMD-30210, **Target 3** SARS-CoV-
374 2 ORF3a putative ion channel/phospholipid in nanodisc¹⁵: EMD-22898.

375 **Table 1** summarizes the approach and lists the software used by each modeling team. Further details for
376 each model can be found in **Supplement S2**. Modeling teams categorized their polymer modeling type as
377 either *ab initio* (followed by optimization), optimized, or not optimized. Non-*ab initio* approaches made use
378 of polymer coordinates from the following PDB entries. **Target 1**: 6cvm, 1jz7, 6tte. **Target 2**: 7bv2, 7b3d,
379 6x71, 3ovb. **Target 3**: 7kjr.

380 Submitted models were further categorized by ligand modeling type, either independently refit or optimized.
381 Initial ligand coordinates and restraints were obtained from the PDB Chemical Component Dictionary
382 (CCD)⁴⁸, Crystallography Open Database (COD)⁴⁹, or from a PDB entry. Ligand restraint generation
383 software included BUSTER Grade (Global Phasing Ltd., Cambridge, UK), Phenix eLBOW⁵⁰, CCP4
384 AceDRG⁵¹, PyRosetta⁵², AMBER Antechamber⁵³, OpenBabel⁵⁴, CHARMM CGenFF⁵⁵, LigPrep
385 (Schrödinger LLC, New York, USA), and CCP4 monomer library⁵⁶. Restraints were not applied by teams
386 using MD-based approaches.

387 *Ab initio* modeling software included ARP/wARP⁵⁷, Mainmast⁵⁸, Mainmastseg⁵⁹, Pathwalker⁶⁰, Rosetta⁶¹,
388 Modeller⁶², and DeepTracer^{63,64}. Model optimization software included CDMD⁶⁵, Phenix²², REFMAC⁶⁶,
389 Servalcat⁶⁷, ProSMART⁶⁸, MDFF⁶⁹, CryoFold⁷⁰, Amber⁵³, MELD^{71,72}, Schrödinger (Schrödinger LLC, New
390 York, USA). The program doubleHelix⁷³ was used to assign RNA sequence and refinement restraints.

391 Atomic displacement parameters (B-factors) were optimized for 32 of 61 models, with 23 applying individual
392 atomic B-factors.

393 Participants made use of VMD⁷⁴, Chimera⁷⁵, ChimeraX⁷⁶, Coot²⁶, ISOLDE⁷⁷, EMDA⁷⁸ and PyMOL for visual
394 evaluation and/or manual model improvement of map-model fit. Manipulation of map densities was carried
395 out using CCP-EM⁷⁹, EMDA, and LAFTER⁸⁰.

396 **Model evaluation system**

397 The evaluation pipeline for the 2021 challenge (model-compare.emdataresource.org) was built upon the
398 basis of the 2019 Model Challenge pipeline^{8,16}. Submitted models were evaluated for >70 individual metrics
399 in four established tracks: Fit-to-Map, Coordinates-only, Comparison-to-Reference and Comparison-
400 among-Models, plus a new Ligand track, created for comparison of ligand-specific scores (See
401 **Supplementary Data S3**). Ligand and Nucleic-acid specific scores provided by Assessor teams (**Table 2**)
402 were integrated into data tables alongside scores from the evaluation pipeline to enable comparisons and
403 composite score generation.

404 **Pharmacophore Modeling**

405 The Molecular Operating Environment platform (MOE) was used to score the placement of ligands. Starting
406 from the model coordinates submitted by each group, the MOE QuickPrep application was used to prepare
407 all-atom structures with hydrogens and atomic partial charges. For each target, an ensemble of structures
408 consisting of all submitted models was input into the db_AutoPH4 application to produce pharmacophore
409 consensus fields based on the ensemble. The pharmacophore consensus fields were then used to score
410 the ligand poses of each submission. Additional details are provided in **Supplementary Data S4**.

411 **Strain energy calculations**

412 Preparation: ligands were extracted from model files. For the T2 F86 ligand, strain energy was measured
413 after deleting the covalent bond to the RNA polymer
414 (SMILES:Nc(ncn1)c2n1c([C@]3(C#N)O[C@@H]([C@H]([C@H]3O)O)COP([O-])([O-]=O)cc2). For the T3 PEE ligand, all
415 models were truncated to just the head group (SMILES:CCC(OC[C@@H](OC(CC)=O)CO[P@]([O-]
416])(OCC[NH3+]=O)=O). Hydrogens were added using MOE/Protonate3D from the Chemical Computing Group.

417 Molecular Mechanics (MM) Forcefield Strain Energy: *predicted ligand energy* was calculated by minimizing
418 each ligand structure using OpenEye/SZYBKI (MMFF94S with Sheffield solvation model) with a maximum
419 RMSD of 0.6 Angstroms. Predicted *global minimum energy* was identified by sampling conformations using
420 OpenEye/Omega and then minimizing each conformer structure using OpenEye/SZYBKI (MMFF94S with
421 Sheffield solvation model) with no restraints, then selecting the conformer with the lowest minimized energy.

422 Neural Net Potential (NNP) Energy: *predicted ligand energy* was calculated by minimizing each ligand
423 structure in an in-house implementation of the ANI neural net potential³⁷ with a maximum RMSD of 0.6
424 Ångstroms. Predicted *global minimum energy* was identified by sampling conformations using

425 OpenEye/Omega and then minimizing each conformer structure using the same in-house implementation
426 of the ANI neural net potential with no restraints.

427 Reported scores are predicted strain energy as (*predicted ligand energy - global minimum energy*) in
428 kcal/mol. NNP was only calculated for the T1 ligand as the method currently does not support atomic
429 charges.

430 **Molecular Graphics**

431 Molecular graphics images were generated using UCSF Chimera (Figures 2, 5, Extended Data Figure 1).

432 **Classification of unique ligands in PDB introduced by Cryo-EM**

433 Search of the Protein Data Bank via RCSB PDB's data API⁸¹ identified 981 unique non polymer
434 ligands/PDB Chemical Component Dictionary (CCD) ids in EM-derived PDB structures released through
435 December 2021. Next, for each ligand, the PDB entry that first introduced the ligand/CCD id was identified.
436 The 403 unique non-polymer ligands that were found to be introduced in structures determined by cryo-EM
437 were then manually classified as enzyme modulators (substrates, inhibitors, agonists, co-factors), medically
438 relevant drugs, lipids, photochemicals (e.g. carotenoids), peptides (amino-acid-based), reagents (buffers
439 or labels), nucleotides, or steroids (fused rings).

440 **Acknowledgements**

441 EMDDataResource (CLL, AK, GP, HMB, WC) was supported by the US National Institutes of Health
442 (NIH)/National Institute of General Medical Science (R01GM079429).

443 The following additional grants are acknowledged for participant support.

444 JSR, CJW, VBC, and DCR acknowledge support from the US National Institutes of Health (P01GM063210,
445 R01GM073919, R35GM131883)

446 NWM, PVA, CJS and OVS gratefully acknowledge the financial support of NIH/NIGMS through grants
447 P01GM063210, R01GM071939, R24GM141254 and the Phenix Industrial Consortium. This work was
448 supported in part by the US Department of Energy under Contract DE-AC02-05CH11231.

449 JC and BS acknowledge support to the Institute of Biotechnology of the Czech Academy of Sciences RVO
450 86652036.

451 RJR was supported by a Principal Research Fellowship from the Wellcome Trust (209407/Z/17/Z).

452 DK acknowledges support from the US National Institutes of Health (R01GM133840).

453 AP acknowledges support from an NSF-CAREER award (CHE-2235785).

454 HG and MI acknowledge support by the German Science Foundation (DFG, RTG 2756).

455 JC and NG acknowledge the support from the US National Institutes of Health (grant #: R01GM146340).

456 CH and W-CK were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research
457 Foundation) under Germany's Excellence Strategy (CIBSS – EXC-2189 – Project ID 390939984).

458 SCV was supported by the American Leprosy Missions (grant number G88726) at the Department of
459 Biochemistry, University of Cambridge.

460 PSB was supported by the Biotechnology and Biological Sciences Research Council (grant number
461 BB/S005099/1).

462 SWH was supported by the Biotechnology and Biological Sciences Research Council (grant number
463 BB/T012935/1).

464 SM acknowledges support from SERB (Project No. CRG/2022/002761).

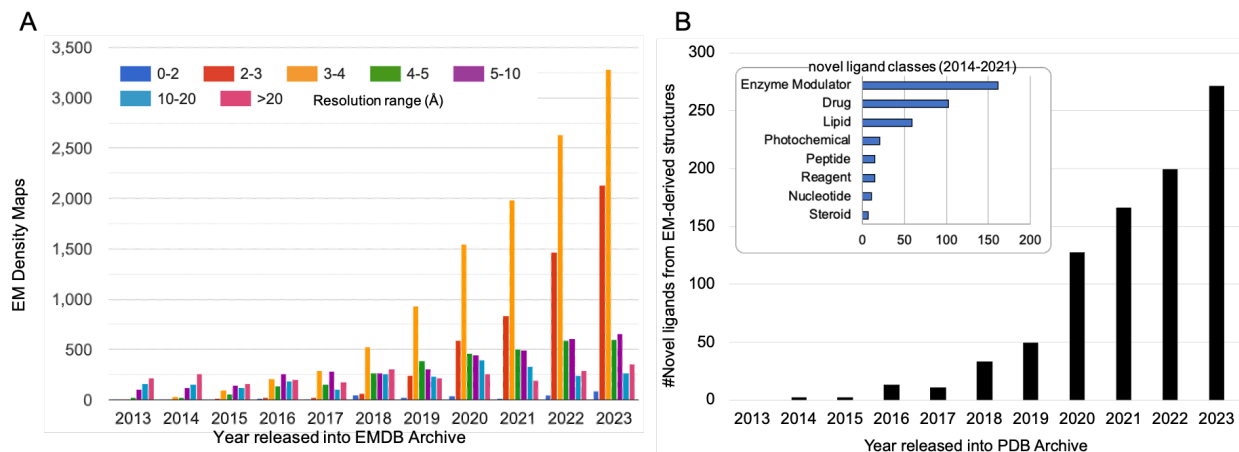
465 CMP, TB, APJ and MW were supported by the Medical Research Council (grant MR/V000403/1).

466 FD and AMuenks acknowledge support from the National Institute of General Medical Sciences
467 (1R01GM123089-01). AMuenks acknowledges support by the NSF Graduate Research Fellowship (DGE-
468 1762114).

469 Author Contributions: HB and WC conceived the project; CL and AK organized the Challenge with the
470 assessors and modelers; GP and MFS assisted in the analysis.

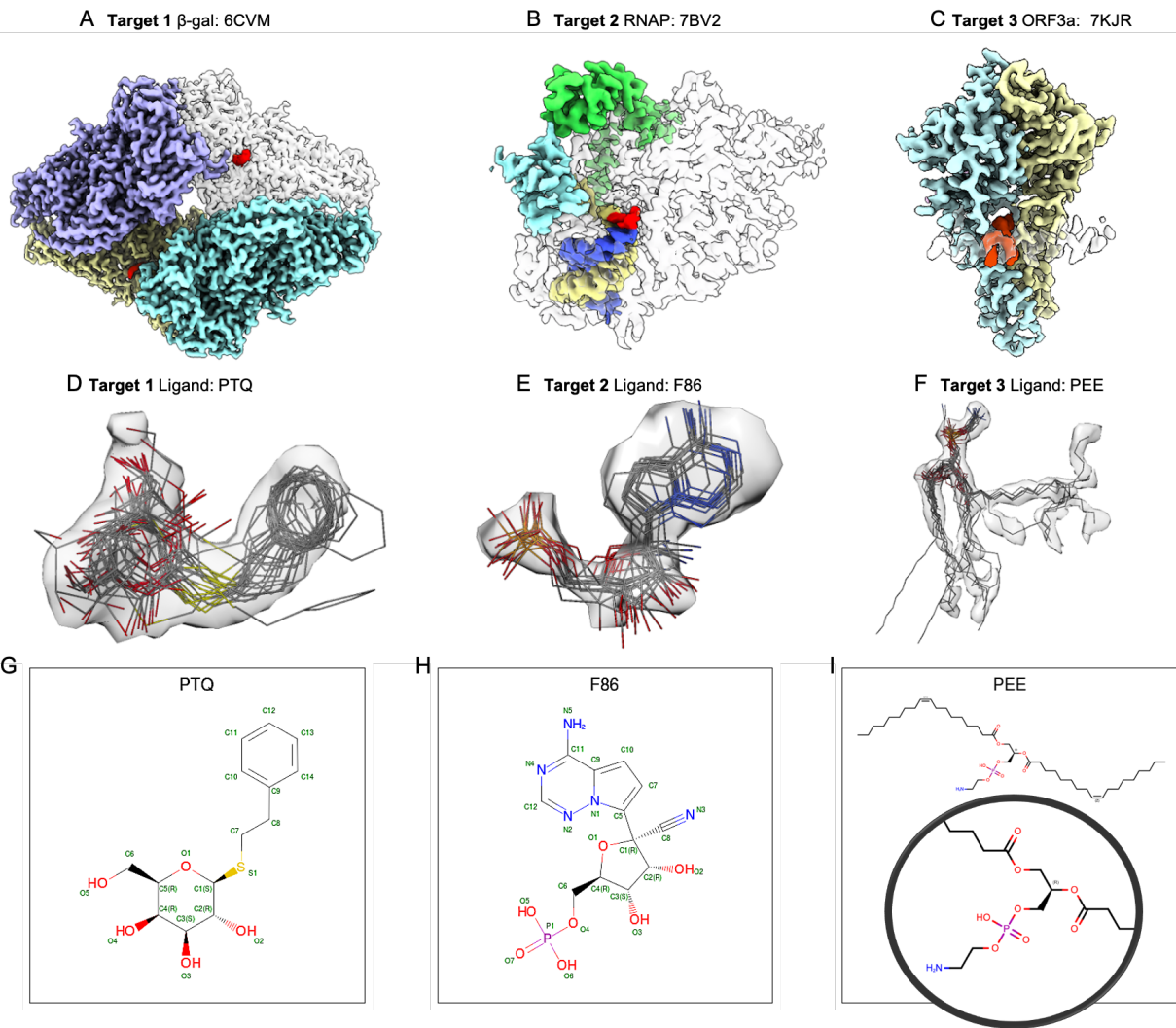
471 **Competing Interests**

472 S. Noreng, A. Gobbi, A. Rohou, B. Sellers and Y. Yang are current or former employees of Genentech. The
473 remaining authors declare no competing interests.



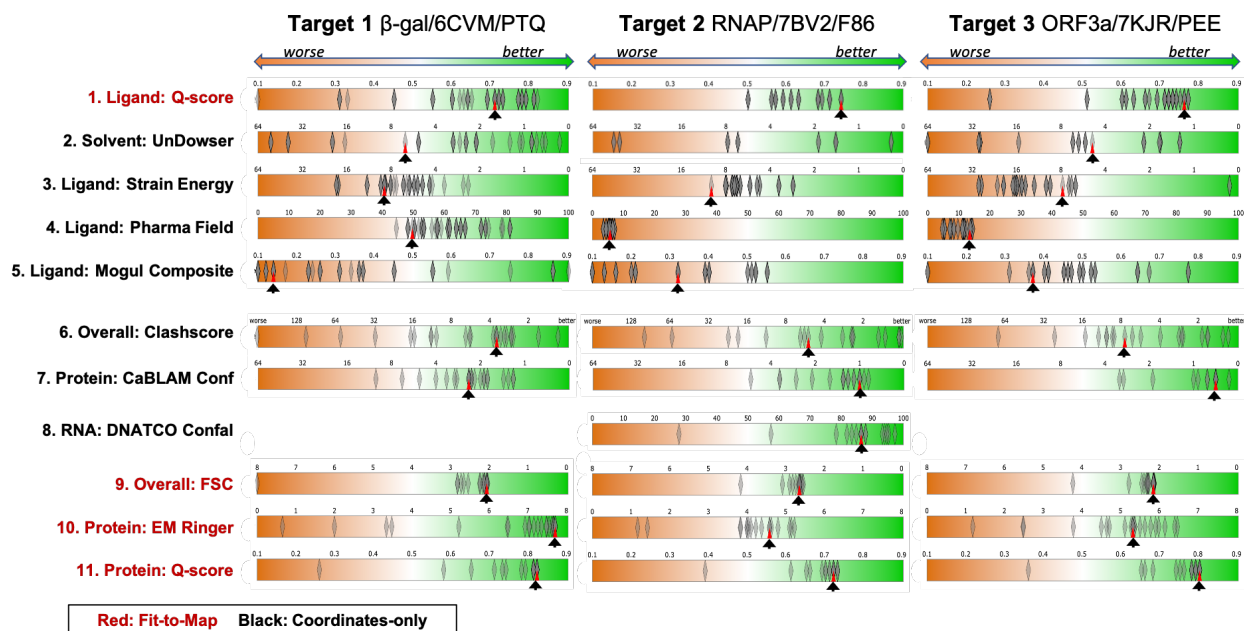
474

475 **Figure 1.** Growth of cryo-EM structures and novel ligands derived from them. (A) Cryo-EM maps released
 476 into the EM Data Bank (EMDB) archive by year and resolution range (source: www.emdataresource.org)
 477 up to the end of 2023. (B) Novel non-polymer ligands included in cryo-EM structures by year of release into
 478 the Protein Data Bank (PDB) through 2023. Inset: major categories of novel ligands found in cryo-EM-
 479 derived models (through 2021). See **Online Methods** for details.



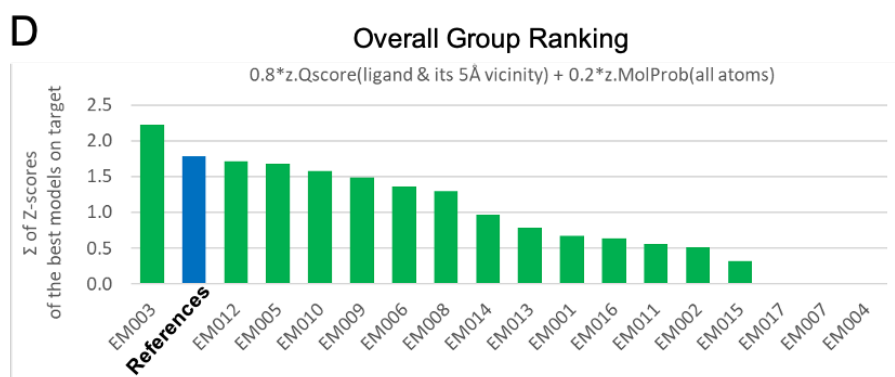
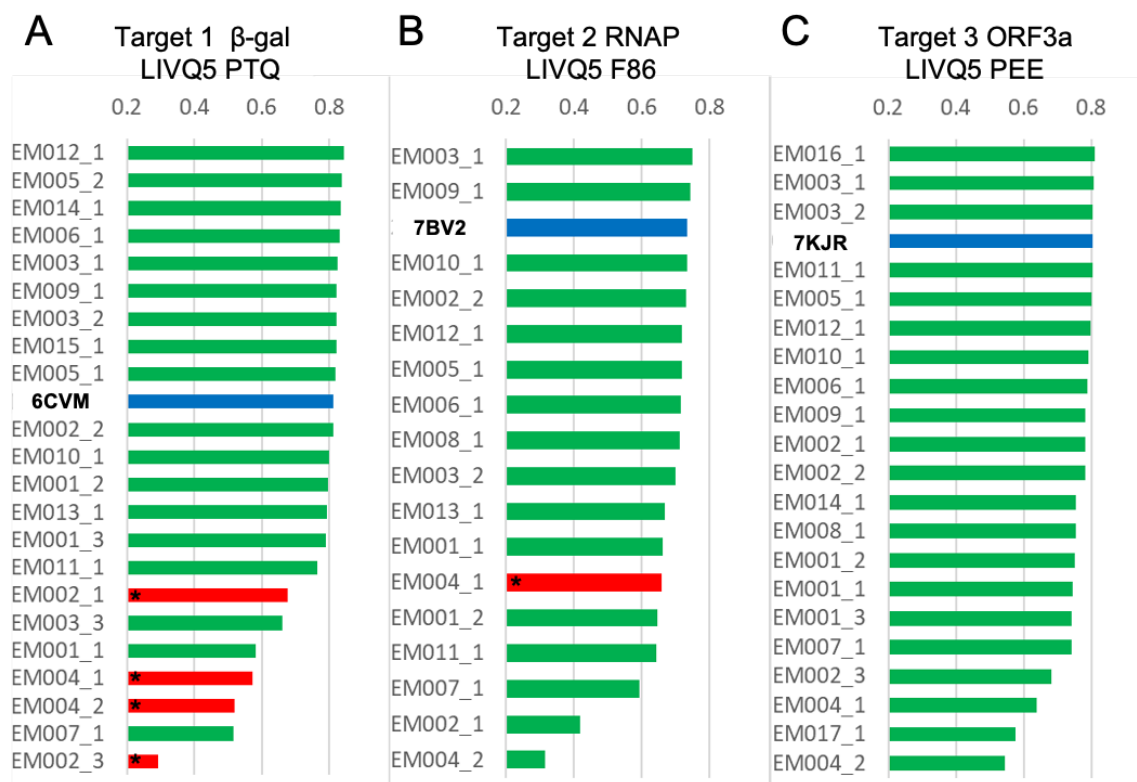
480

481 **Figure 2. Ligand Challenge targets and ligands from submitted models.** In (A-C), Targets 1-3 are
 482 shown, with each polymer/nucleic acid chain rendered as a separate surface with a different color, in some
 483 cases semi-transparent. Target ligands are shown in red. In (D-F), segmented density representing each
 484 target ligand is shown with a semi-transparent surface, with submitted ligand models overlaid. Map contour
 485 levels are 0.35 (2.3σ), 0.036 (2.6σ), 0.25 (3.7σ) respectively (sigma values were calculated from the full
 486 unmasked map to capture variation in background noise). (G-I) Chemical sketches for each of the target
 487 ligands (source: PDB). Selected individual ligand poses from submitted models superimposed on target
 488 map densities are shown in [Extended Data Figure 1](#).

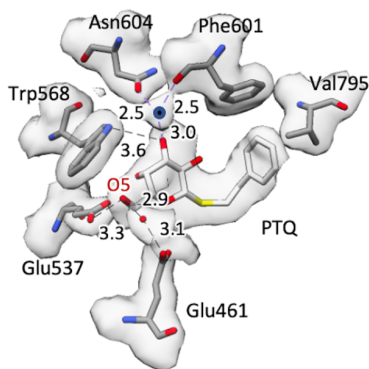
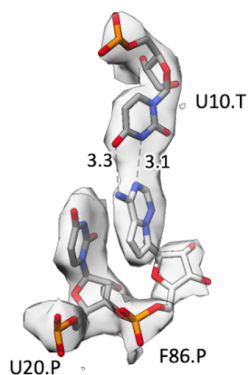
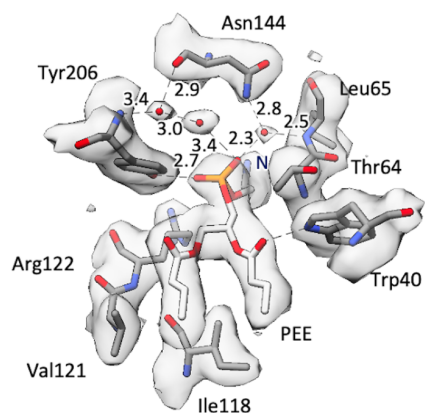
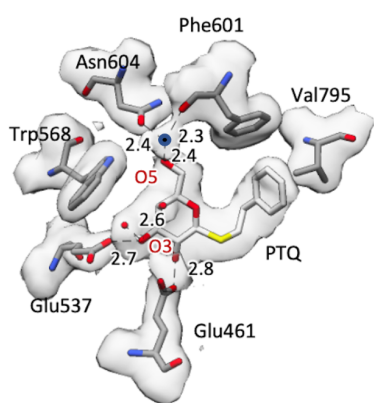
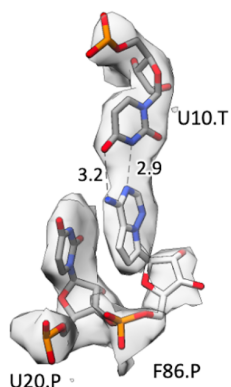
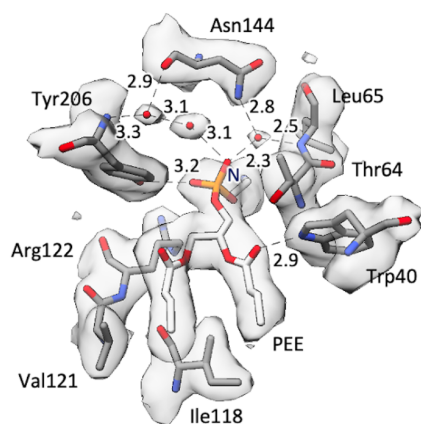


489

490 **Figure 3 Model score distributions of selected assessments for Targets 1-3.** Top 5 rows: ligand and
 491 solvent scores, bottom 6 rows: overall and protein-specific scores. Fit-to-Map based metrics have red
 492 labels; Coordinates-only metrics have black labels. Diamonds indicate individual scores of submitted
 493 models; red triangles (with supporting black arrows) indicate the scores of the reference models; in a few
 494 cases no score is available for the reference model. Each score distribution is plotted against an
 495 orange(left)-white-green(right) color gradient with orange indicating poorer scores, and green indicating
 496 better scores⁸.



497
 498 **Figure 4. Model and modeling group rankings.** (A-C) LIVQ5 (Ligand + Immediate Vicinity Q-score
 499 $\leq 5\text{\AA}$) is plotted according to rank for each submitted model (labeled as participant group id, see [Table 1](#),
 500 followed by model number) and for each reference model (labeled as PDB id). Models with good overall
 501 MolProbity (MP) scores (< 3.0) are shaded green; those with poor MP scores (> 3.0) are shaded red and
 502 starred; reference models are shaded blue and labeled in bold. Immediate vicinity includes all non-hydrogen
 503 model atoms $\leq 5\text{\AA}$ from any ligand non-hydrogen atom. Model rankings with extended vicinity (LIVQ10) are
 504 provided in **Extended Data Figure 2**. (D) Ranking of Challenge participant groups based on the Fit-to-Map
 505 accuracy of ligands as shown in (A-C), and stereochemical plausibility, as described in the main text.
 506 Overall rank is calculated as the all-target sum of weighted z-scores for the best per-target models from the
 507 group (see equation in main text).

A Target 1 β -gal/6CVM/PTQ**B Target 2 RNAP/7BV2/F86****C Target 3 ORF3a/7KJR/PEE****D Target 1 Model EM012_1****E Target 2 Model EM003_1****F Target 3 Model EM016_1**

508

509

510

511

512

513

Figure 5. Visualization of ligands and surrounding atoms in deposited reference models and best-scoring submitted models (A,B,C) deposited reference models for Targets 1-3 as described in the main text. (D,E,F) best-scoring submitted models for each target. Modeled solvent atoms are shown as red spheres; a modeled ion in panels A,D is shown as a dark blue sphere. Numerical labels with dashed lines indicate atom-to-atom distances in Ångstroms.

514 **Table 1.** Modeling teams with number of models per target, approaches and software used.

ID	Modeling Team	T1	T2	T3	Polymer Modeling	Ligand Modeling	Ligand Restraints Software	Automation level	Modeling Software
EM001	D. Kihara, G. Terashi, D. Sarkar, J. Verburgt	3	2	3	ab initio or optimized	refit or optimized	MD Force Field	partial	Mainmast, Mainmastseg, Rosetta PyMOL, Schrodinger, VMD, Chimera, MDFF
EM002	D. Si, S. Lin, M. Zhao, R. Cao, J. Hou	3	2	3	ab initio or none	refit	Phenix eLBOW	full	DeepTracer, Phenix
EM003	A. Muenks, F. DiMaio	3	2	2	optimized	refit	Phenix eLBOW, Open Babel	partial	Rosetta, Chimera
EM004	J. Cheng, N. Giri	2	2	2	ab initio	refit	PyRosetta	partial	Rosetta, Chimera, DeepTracer
EM005	G. Pintilie, M. Schmid, W. Chiu	2	1	1	none	refit	Phenix eLBOW	partial	Chimera
EM006	M. Baker, C. Hryc	1	1	1	ab initio	refit	Phenix eLBOW	partial	Pathwalker, Phenix
EM007	A. Perez, A. Mondal, R. Esmaeeli, L. Lang	1	1	1	optimized	optimized	PyRosetta, Antechamber, MD Force Field	partial	MELD, Amber, VMD
EM008	P. Emsley	1	1	1	optimized	refit	CCP4 AceDRG	partial	Coot, REFMAC
EM009	N.W. Moriarty, P. V. Afonine, C.J. Schlicksup, O.V. Sobolev	1	1	1	optimized	refit	Phenix eLBOW	partial	Coot, Chimera, ChimeraX, Phenix
EM010	G. Chojnowski	1	1	1	ab initio	refit	CCP4 mon lib	partial	ARP/wARP, ChimeraX, Coot, Isolde, Phenix, doubleHelix
EM011	M. Igaev, H. Grubmüller, . Pohjolainen, A. Vaiana	1	1	1	ab initio	optimized	MD Force Field	partial	Chimera, Modeller, VMD, CDMD
EM012	C. Palmer, R. Nicholls, R. Warshamanage, K. Yamashita, G. Murshudov, P. Bond, S. Hoh, M. Olek, K. Cowtan, A. Joseph, T. Burnley, M. Winn	1	1	1	optimized	refit or optimized	CCP4 AceDRG	partial	CCP-EM, Coot, EMDA, LAFTER, ProSMART, REFMAC, Servalcat
EM013	A. Singharoy, S. Mittal, A. Perez, D. Kihara, M. Shekhar, D. Sarkar, G. Terashi, C. Rowley, R. Esmaeeli, L. Lang, A. Mondal, A. Campbell	1	1		optimized	refit or optimized	CGENFF	partial	MDFF, CryoFold, MELD
EM014	W.-C. Kao, C. Hunte	1		1	optimized	refit	Grade (BUSTER), Phenix eLBOW	manual	ChimeraX, Coot, Isolde, Phenix
EM015	G. Schröder, L. Schäfer, K. Pothula	1			optimized	refit	MD Force Field	partial	CDMD

EM016	D. Kumar			1	optimized	refit	Phenix eLBOW	partial	Coot, Phenix
EM017	S. Weyand, S.C. Vedithi, T. Blundell, S. Brohawn			1	optimized	refit	Schrödinger Ligprep	full	Schrödinger
Totals		23	17	21					

515

516 **Table 2.** Ligand assessment teams and methods

Assessment Team ID	Team members	Assessment method
AT01	C. Shao	wwPDB validation report pipeline (Mogul)
AT02	P. Emsley	Coot Tools
AT03	B. Schneider, J. Černý	Nucleic acid conformations, protein hydration analysis
AT04	J.S. Richardson, C.J. Williams, V. Chen, D. Richardson	Contact analysis, probescore, occupancy, UnDowser, CaBLAM, visual examination
AT05	C.I. Williams, Chemical Computing Group Support Team	Pharmacophore density fields (PH4)
AT06	B. Sellers, A. Gobbi, S. Noreng, Y. Yang, A. Rohou	Molecular Mechanics Force Field Strain Energy (MM), Neural Net Potential Energy (NNP)
AT07	G. Pintilie, M. Schmid, W. Chiu	Q-score analysis

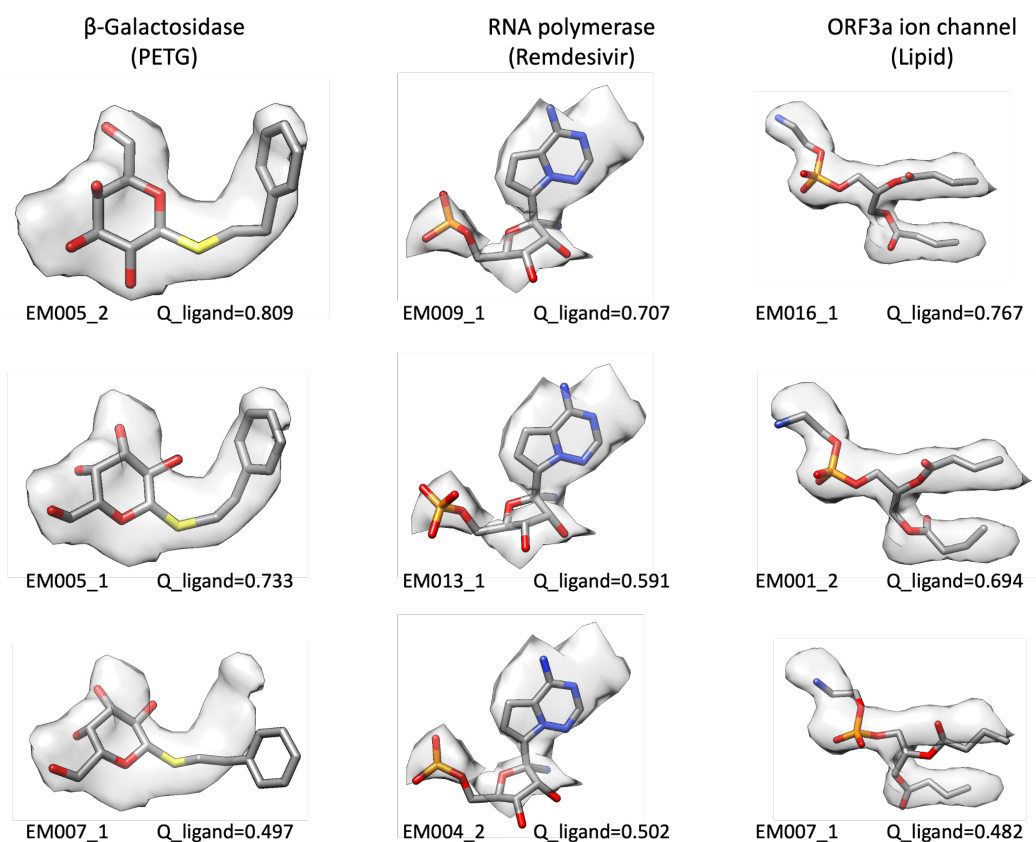
517

518 **Table 3.** Ligand and Ligand+environment Q-scores for submitted models with highest ligand Q-scores.
 519 Expected_Q is the expected Q-score for well-fitted models in maps at similar resolutions, based on
 520 analysis of a subset of publicly archived maps and models⁸². Q-scores well below the expected value
 521 indicate either that the map is not as well resolved as other maps at similar resolution, e.g. due to
 522 heterogeneity, or that the model is not optimally fitted to the map.

Target Map (Reported Resolution)	Model with highest ligand Q-score	Q_ligand (ligand atoms)	Q_near (atoms $\leq 5\text{\AA}$ of ligand)	LIVQ5 (ligand +atoms $\leq 5\text{\AA}$ of ligand)	Expected_Q at reported map resolution
T1 β -gal (1.9 \AA)	EM005_2	0.809	0.849	0.845	0.846
T2 RNAP (2.5 \AA)	EM009_1	0.707	0.735	0.731	0.690
T3 ORF3a (2.1 \AA)	EM016_1	0.767	0.819	0.812	0.791

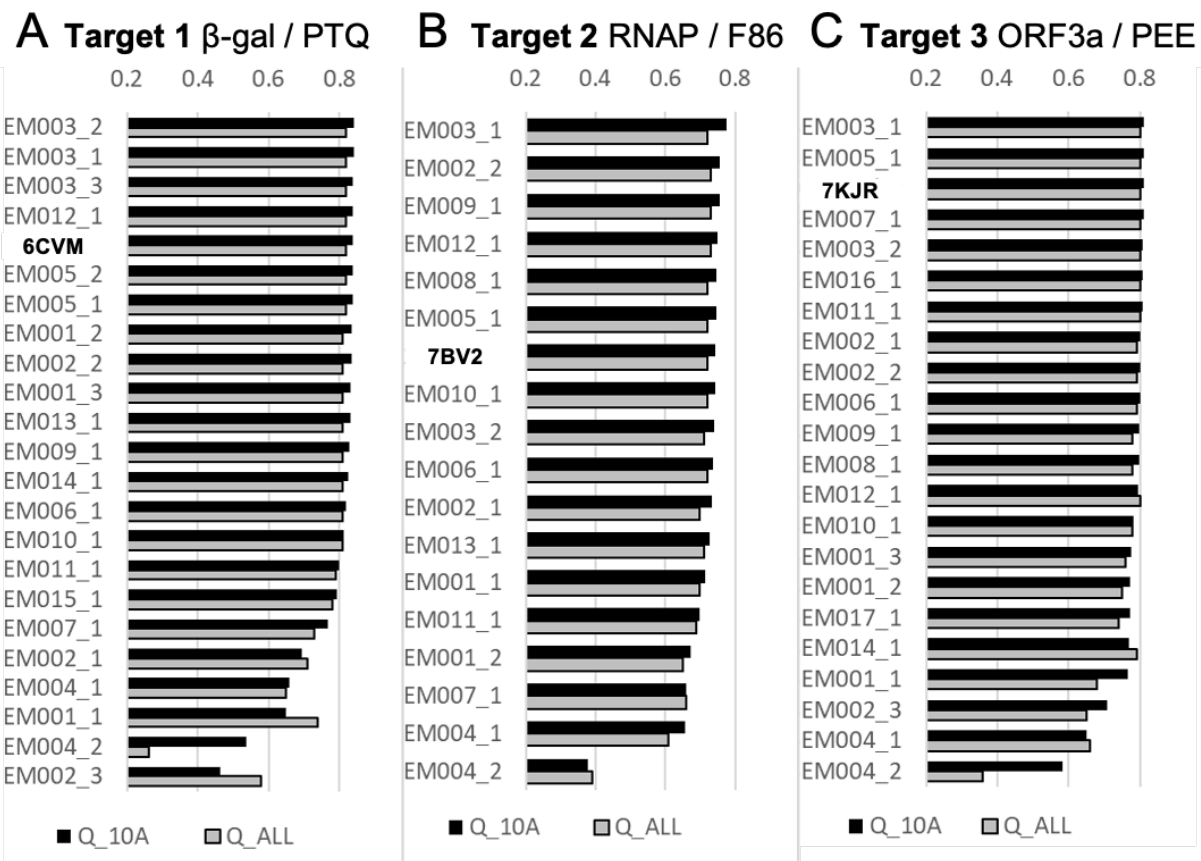
523

524



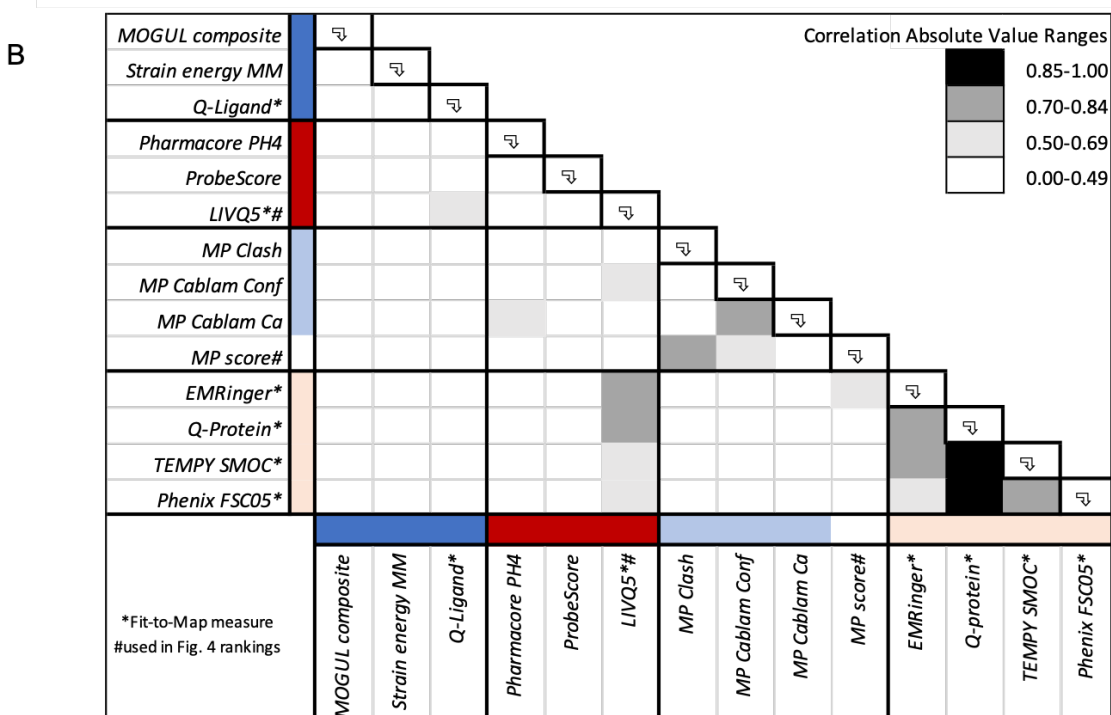
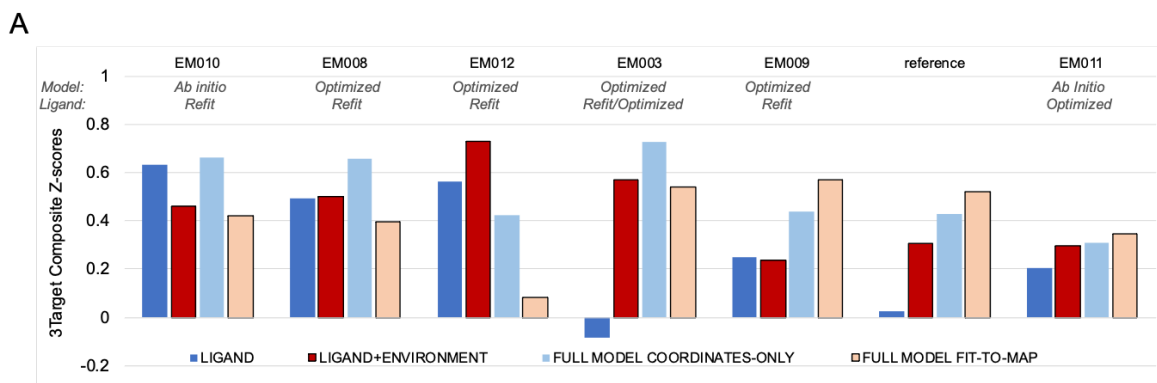
526

527 **Extended Data Figure 1.** Selected submitted ligand models for each of the Challenge targets, labeled by
528 team ID and model # (see Table 1), in order of decreasing ligand Q-scores (see Figure 3, row 1) from top
529 to bottom. The portion of the map corresponding to the ligand is shown as a semi-transparent surface,
530 along with the model of the ligand. Ligand Q-score is the average Q-score of all non-H atoms in the ligand.
531 For each atom, the Q-score is measured by correlation of map density to the expected gaussian function,
532 at points within 2 Å of the atom and closer to the atom than any other non-H atom¹⁰. Higher-
533 scoring ligand models fit better in the cryo-EM density than lower-scoring models.



534

535 **Extended Data Figure 2.** Q-score rankings for ligand + extended vicinity and for full models. **(A-C)** LIVQ10
 536 (Ligand + extended vicinity ≤ 10 Å) Q-scores (black bars) and full model Q-scores (gray bars) are plotted
 537 for each submitted model and each reference model, with order according to ligand + extended vicinity
 538 rank. Reference model positions are highlighted with red arrows. Target/reference labels are as defined in
 539 the Figure 4 legend.



540

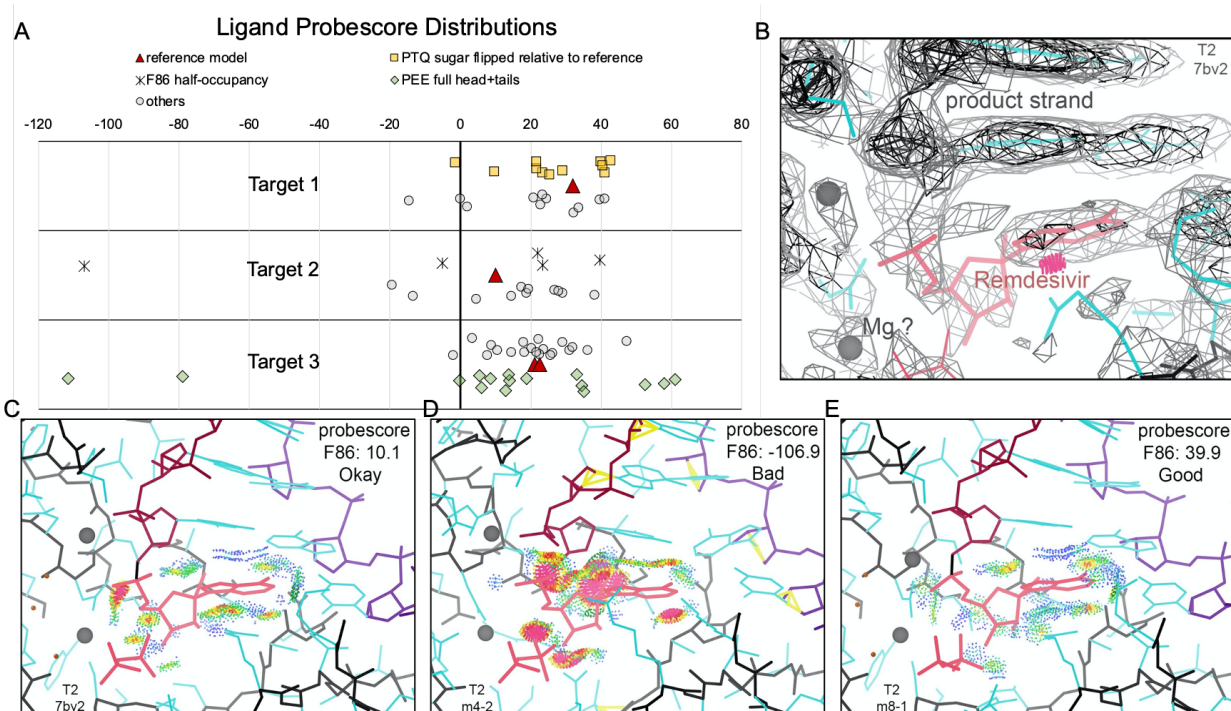
541 **Extended Data Figure 3.** Alternative Group Ranking by sum of Ligand, Ligand+Environment, Full Model
 542 Coordinates-only, Full Model Fit-to-Map composite scores. (A) Group ranking (left-to-right) according to the
 543 sum of four composite z-scores, as described below. Only groups that submitted models for all 3 targets
 544 and have rank similar to or better than PDB reference models are shown. (B) Correlation table (n=64) of
 545 scores used to create z-scores and rankings in panel (A) and/or Figure 4. Group composite scores were
 546 calculated per team as follows. For each submitted model, and for each score type, a composite z-score
 547 was calculated. For each target (T1, T2, T3), the model submitted by that group with maximum composite
 548 z-score was selected for inclusion in the final average score over all targets.

549 Ligand: $z = (0.33 * z.MogulComposite + 0.33 * z.StrainEnergyMM + 0.33 * z.Q-ligand)$

550 Ligand+environment: $z = (0.33 * z.Pharmacore + 0.33 * z.Probescore + 0.33 * z.LIVQ5)$

551 Full model coordinates-only: $z = (0.25 * z.Clash + 0.5 * z.CablamConf + 0.25 * z.CablamCa)$

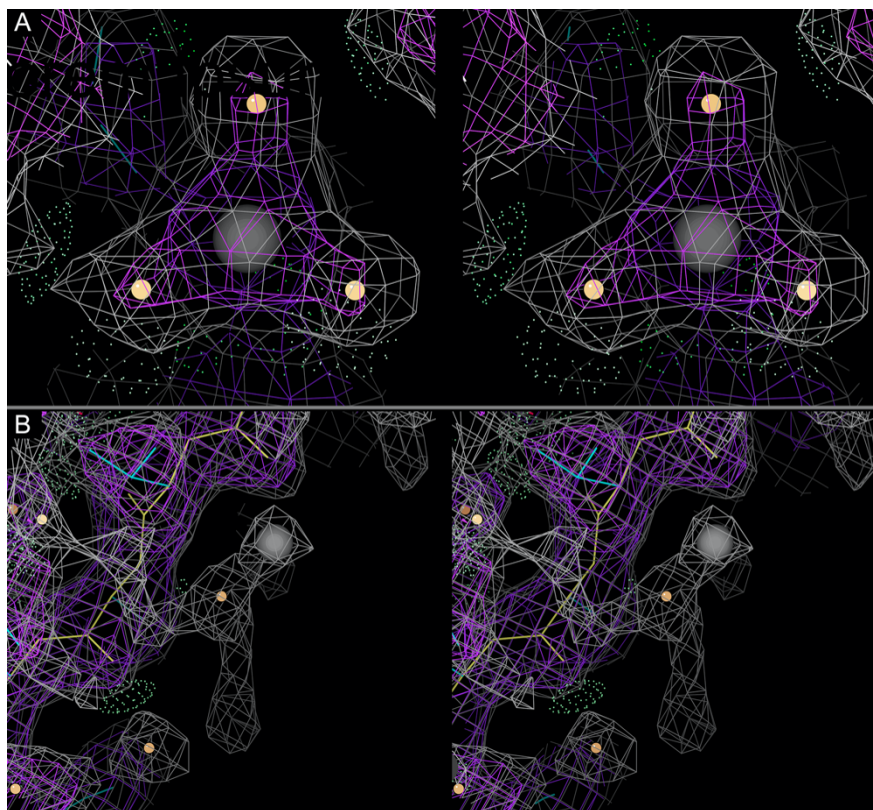
552 Full model fit-to-map: $z = (0.25 * z.EMRinger + 0.25 * z.Q-Protein + 0.25 * z.TEMPYSMOC + 0.25 * z.PhenixFCS05)$



554

555 **Extended Data Figure 4.** Ligand/Ligand Environment Proboscres. **(A)** Molprobit Proboscres³²
 556 distributions for ligands in Targets 1-3 (reference models: red triangles; submitted model scores are plotted
 557 as gray circles with following exceptions: Target 1, yellow boxes if PTQ sugar ring position was flipped
 558 relative to reference; Target 2, asterisk if F86 was set to half-occupancy; Target 3, blue diamonds if PEE
 559 was modeled as head-group+tails). Scores are plotted in horizontal axis lanes with small random vertical
 560 shifts to visually separate clustered points. Notably, score distributions have wide spreads independent of
 561 noted model features: PTQ sugar orientation, F86 occupancy, or PEE inclusion of tails—although for PEE
 562 the score distribution is noticeably broader when the larger and more variable tails are included. **(B)** T2
 563 density map with reference model in the region of the F86 ligand³⁸, showing half-strength density for the
 564 remdesivir ligand, implying that only half the molecules have covalently bound inhibitor. **(C-E)** T2 F86 +
 565 pyrophosphate ligand environments for the reference model (PDBid 7BV2), model EM004_2, and model
 566 EM008_1, respectively. All-atom contact dots are from Proboscres, with all-atom clashes in hot pink and
 567 favorable H-bonds and vdW contacts in green and blue. Molecular graphics are shown in KiNG⁸³.

568



569

570 **Extended Data Figure 5:** Evaluation of ions in submitted models (stereo images). (A) Target 1 6cvm
 571 reference model Mg A2002 (gray sphere) with water ligands (orange spheres), located near the PETG
 572 ligand, with density for classic octahedral coordination. Only six of 23 submitted Target 1 models included
 573 the Mg²⁺ and all three coordinating waters. Others had either only Mg²⁺, Mg²⁺ plus one or two waters, Mg²⁺
 574 plus waters with zero occupancy, no atoms modeled, or atoms significantly displaced. (B) Some groups
 575 placed metal ions with weak justification, as exemplified by the Na⁺ (grey sphere) shown here in model
 576 EM005_1 for Target 3.

577

578 **Supplementary Information**

579 S1: Ligand Challenge Model Submission Statistics and Form (.pdf)

580 S2: Ligand Challenge Submitted Model Metadata (.xlsx)

581 S3: Ligand Challenge Scores (.xlsx)

582 S4: MOE Pharmacore Assessment Summary (.pdf)

583 References

- 584 1. Adams, P. D. *et al.* Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop. *Structure*
585 **24**, 502–508 (2016).
- 586 2. Gore, S. *et al.* Validation of Structures in the Protein Data Bank. *Structure* **25**, 1916–1927 (2017).
- 587 3. Smart, O. S. *et al.* Validation of ligands in macromolecular structures determined by X-ray
588 crystallography. *Acta Crystallogr D Struct Biol* **74**, 228–236 (2018).
- 589 4. Feng, Z. *et al.* Enhanced validation of small-molecule ligands and carbohydrates in the Protein Data
590 Bank. *Structure* **29**, 393–400.e1 (2021).
- 591 5. Lawson, C. L., Berman, H. M. & Chiu, W. Evolving data standards for cryo-EM structures. *Struct*
592 *Dyn* **7**, 014701 (2020).
- 593 6. Lawson, C. L. & Chiu, W. Comparing cryo-EM structures. *J. Struct. Biol.* **204**, 523–526 (2018).
- 594 7. Barad, B. A. *et al.* EMRinger: side chain-directed model and map validation for 3D cryo-electron
595 microscopy. *Nat. Methods* **12**, 943–946 (2015).
- 596 8. Lawson, C. L. *et al.* Cryo-EM model validation recommendations based on outcomes of the 2019
597 EMDataResource challenge. *Nat. Methods* **18**, 156–164 (2021).
- 598 9. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure
599 validation. *Protein Sci.* **27**, 293–315 (2018).
- 600 10. Pintilie, G. *et al.* Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods*
601 **17**, 328–334 (2020).
- 602 11. Wang, Z., Patwardhan, A. & Kleywegt, G. J. Validation analysis of EMDB entries. *Acta Crystallogr D*
603 *Struct Biol* **78**, 542–552 (2022).
- 604 12. Bartesaghi, A. *et al.* Atomic Resolution Cryo-EM Structure of β -Galactosidase. *Structure* **26**, 848–
605 856.e3 (2018).
- 606 13. Yin, W. *et al.* Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-
607 CoV-2 by remdesivir. *Science* **368**, 1499–1504 (2020).
- 608 14. Kokic, G. *et al.* Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* **12**,
609 279 (2021).
- 610 15. Kern, D. M. *et al.* Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nat. Struct. Mol.*

- 611 *Biol.* **28**, 573–582 (2021).
- 612 16. Kryshchuk, A., Adams, P. D., Lawson, C. L. & Chiu, W. Evaluation system and web
613 infrastructure for the second cryo-EM model challenge. *J. Struct. Biol.* **204**, 96–108 (2018).
- 614 17. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and
615 contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
- 616 18. Lagerstedt, I. *et al.* Web-based visualisation and analysis of 3D electron-microscopy data from
617 EMDB and PDB. *J. Struct. Biol.* **184**, 173–181 (2013).
- 618 19. Joseph, A. P., Lagerstedt, I., Patwardhan, A., Topf, M. & Winn, M. Improved metrics for comparing
619 structures of macromolecular assemblies determined by 3D electron-microscopy. *J. Struct. Biol.*
620 **199**, 12–26 (2017).
- 621 20. Afonine, P. V. *et al.* New tools for the analysis and validation of cryo-EM maps and atomic models.
622 *Acta Crystallogr D Struct Biol* **74**, 814–840 (2018).
- 623 21. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta*
624 *Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
- 625 22. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons:
626 recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).
- 627 23. Kryshchuk, A. *et al.* Challenging the state of the art in protein structure prediction: Highlights of
628 experimental target structures for the 10th Critical Assessment of Techniques for Protein Structure
629 Prediction Experiment CASP10. *Proteins* **82 Suppl 2**, 26–42 (2014).
- 630 24. Bruno, I. J. *et al.* Retrieval of crystallographically-derived molecular geometry information. *J. Chem.*
631 *Inf. Comput. Sci.* **44**, 2133–2144 (2004).
- 632 25. Shao, C. *et al.* Simplified quality assessment for small-molecule ligands in the Protein Data Bank.
633 *Structure* **30**, 252–262.e4 (2022).
- 634 26. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model
635 building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* **29**, 1069–1078 (2020).
- 636 27. Nicholls, R. A. *et al.* Modelling covalent linkages in CCP4. *Acta Crystallogr D Struct Biol* **77**, 712–
637 726 (2021).
- 638 28. Černý, J., Božíková, P., Svoboda, J. & Schneider, B. A unified dinucleotide alphabet describing both

- 639 RNA and DNA structures. *Nucleic Acids Res.* **48**, 6367–6381 (2020).
- 640 29. Černý, J. *et al.* Structural alphabets for conformational analysis of nucleic acids available at
641 dnatco.datmos.org. *Acta Crystallogr D Struct Biol* **76**, 805–813 (2020).
- 642 30. Biedermannová, L. & Schneider, B. Structure of the ordered hydration of amino acids in proteins:
643 analysis of crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 2192–2202 (2015).
- 644 31. Černý, J., Schneider, B. & Biedermannová, L. WatAA: Atlas of Protein Hydration. Exploring
645 synergies between data mining and ab initio calculations. *Phys. Chem. Chem. Phys.* **19**, 17094–
646 17102 (2017).
- 647 32. Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S. & Richardson, D. C. New tools in
648 MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink ‘waters,’ and NGL
649 Viewer to recapture online 3D graphics. *Protein Sci.* **29**, 315–329 (2020).
- 650 33. Jiang, S., Feher, M., Williams, C., Cole, B. & Shaw, D. E. AutoPH4: An Automated Method for
651 Generating Pharmacophore Models from Protein Binding Pockets. *J. Chem. Inf. Model.* **60**, 4326–
652 4338 (2020).
- 653 34. Tyagi, R., Singh, A., Chaudhary, K. K. & Yadav, M. K. Chapter 17 - Pharmacophore modeling and
654 its applications. in *Bioinformatics* (eds. Singh, D. B. & Pathak, R. K.) 269–289 (Academic Press,
655 2022).
- 656 35. Sellers, B. D., James, N. C. & Gobbi, A. A Comparison of Quantum and Molecular Mechanical
657 Methods to Estimate Strain Energy in Druglike Fragments. *J. Chem. Inf. Model.* **57**, 1265–1275
658 (2017).
- 659 36. Lee, M.-L. *et al.* chemalot and chemalot_knime: Command line programs as workflow tools for drug
660 discovery. *J. Cheminform.* **9**, 38 (2017).
- 661 37. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT
662 accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
- 663 38. Croll, T. I., Williams, C. J., Chen, V. B., Richardson, D. C. & Richardson, J. S. Improving SARS-
664 CoV-2 structures: Peer review by early coordinate release. *Biophys. J.* **120**, 1085–1096 (2021).
- 665 39. Modi, V., Xu, Q., Adhikari, S. & Dunbrack, R. L., Jr. Assessment of template-based modeling of
666 protein structure in CASP11. *Proteins* **84 Suppl 1**, 200–220 (2016).

- 667 40. Zhang, K. *et al.* Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat.*
668 *Commun.* **10**, 5511 (2019).
- 669 41. Su, Z. *et al.* Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 Å resolution. *Nature* **596**,
670 603–607 (2021).
- 671 42. Lawson, C. L., Berman, H. M., Chen, L., Vallat, B. & Zirbel, C. L. The Nucleic Acid Knowledgebase:
672 a new portal for 3D structural information about nucleic acids. *Nucleic Acids Res.* (2023)
673 doi:10.1093/nar/gkad957.
- 674 43. Sun, S. Y. *et al.* Cryo-ET of parasites gives subnanometer insight into tubulin-based structures.
675 *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).
- 676 44. Liu, H.-F. *et al.* nextPYP: a comprehensive and scalable platform for characterizing protein
677 variability in situ using single-particle cryo-electron tomography. *Nat. Methods* (2023)
678 doi:10.1038/s41592-023-02045-0.
- 679 45. Chmielewski, D. *et al.* Integrated analyses reveal a hinge glycan regulates coronavirus spike tilting
680 and virus infectivity. *Res Sq* (2023) doi:10.21203/rs.3.rs-2553619/v1.
- 681 46. Yang, H. *et al.* Automated and accurate deposition of structures solved by X-ray diffraction to the
682 Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1833–1839 (2004).
- 683 47. wwPDB Consortium. EMDB-the Electron Microscopy Data Bank. *Nucleic Acids Res.* (2023)
684 doi:10.1093/nar/gkad1019.
- 685 48. Westbrook, J. D. *et al.* The chemical component dictionary: complete descriptions of constituent
686 molecules in experimentally determined 3D macromolecules in the Protein Data Bank.
687 *Bioinformatics* **31**, 1274–1278 (2015).
- 688 49. Gražulis, S. *et al.* Crystallography Open Database (COD): an open-access collection of crystal
689 structures and platform for world-wide collaboration. *Nucleic Acids Res.* **40**, D420–7 (2012).
- 690 50. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization
691 Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol.*
692 *Crystallogr.* **65**, 1074–1080 (2009).
- 693 51. Nicholls, R. A. *et al.* The missing link: covalent linkages in structural models. *Acta Crystallogr D*
694 *Struct Biol* **77**, 727–745 (2021).

- 695 52. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing
696 molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
- 697 53. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a
698 general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
- 699 54. O’Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
- 700 55. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules
701 compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–
702 690 (2010).
- 703 56. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for
704 its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184–2195 (2004).
- 705 57. Chojnowski, G., Sobolev, E., Heuser, P. & Lamzin, V. S. The accuracy of protein models
706 automatically built into cryo-EM maps with ARP/wARP. *Acta Crystallogr D Struct Biol* **77**, 142–150
707 (2021).
- 708 58. Terashi, G. & Kihara, D. De novo main-chain modeling for EM maps using MAINMAST. *Nat.*
709 *Commun.* **9**, 1618 (2018).
- 710 59. Terashi, G., Kagaya, Y. & Kihara, D. MAINMASTseg: Automated Map Segmentation Method for
711 Cryo-EM Density Maps with Symmetry. *J. Chem. Inf. Model.* **60**, 2634–2643 (2020).
- 712 60. Chen, M. & Baker, M. L. Automation and assessment of de novo modeling with Pathwalking in near
713 atomic resolution cryoEM density maps. *J. Struct. Biol.* **204**, 555–563 (2018).
- 714 61. DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of protein structures into
715 low-resolution density maps using rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
- 716 62. Webb, B. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **1137**, 1–15
717 (2014).
- 718 63. Si, D. *et al.* Deep Learning to Predict Protein Backbone Structure from High-Resolution Cryo-EM
719 Density Maps. *Sci. Rep.* **10**, 4282 (2020).
- 720 64. Pfab, J., Phan, N. M. & Si, D. DeepTracer for fast de novo cryo-EM protein structure modeling and
721 special studies on CoV-related complexes. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
- 722 65. Igaev, M., Kutzner, C., Bock, L. V., Vaiana, A. C. & Grubmüller, H. Automated cryo-EM structure

723 refinement using correlation-driven molecular dynamics. *Elife* **8**, (2019).

724 66. Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-
725 microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 136–153 (2015).

726 67. Yamashita, K., Palmer, C. M., Burnley, T. & Murshudov, G. N. Cryo-EM single-particle structure
727 refinement and map calculation using Servalcat. *Acta Crystallogr D Struct Biol* **77**, 1282–1291
728 (2021).

729 68. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent
730 structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D Biol. Crystallogr.* **70**,
731 2487–2499 (2014).

732 69. Singharoy, A. *et al.* Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron
733 microscopy maps. *Elife* **5**, (2016).

734 70. Shekhar, M. *et al.* CryoFold: determining protein structures and data-guided ensembles from cryo-
735 EM density maps. *Matter* **4**, 3195–3216 (2021).

736 71. MacCallum, J. L., Perez, A. & Dill, K. A. Determining protein structures by combining semireliable
737 data with atomistic physical models by Bayesian inference. *Proc. Natl. Acad. Sci. U. S. A.* **112**,
738 6985–6990 (2015).

739 72. Perez, A., MacCallum, J. L. & Dill, K. A. Accelerating molecular simulations of proteins using
740 Bayesian inference on weak information. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11846–11851 (2015).

741 73. Chojnowski, G. DoubleHelix: nucleic acid sequence identification, assignment and validation tool for
742 cryo-EM and crystal structure models. *Nucleic Acids Res.* **51**, 8255–8269 (2023).

743 74. Hsin, J., Arkhipov, A., Yin, Y., Stone, J. E. & Schulten, K. Using VMD: an introductory tutorial. *Curr.*
744 *Protoc. Bioinformatics* **Chapter 5**, Unit 5.7 (2008).

745 75. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis.
746 *J. Comput. Chem.* **25**, 1605–1612 (2004).

747 76. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis.
748 *Protein Sci.* **27**, 14–25 (2018).

749 77. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-
750 density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).

- 751 78. Warshamanage, R., Yamashita, K. & Murshudov, G. N. EMDA: A Python package for Electron
752 Microscopy Data Analysis. *J. Struct. Biol.* **214**, 107826 (2022).
- 753 79. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta*
754 *Crystallogr D Struct Biol* **73**, 469–477 (2017).
- 755 80. Ramlaul, K., Palmer, C. M. & Aylett, C. H. S. A Local Agreement Filtering Algorithm for
756 Transmission EM Reconstructions. *J. Struct. Biol.* **205**, 30–40 (2019).
- 757 81. Rose, Y. *et al.* RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and
758 Efficient Access to Macromolecular Structure Data from the PDB Archive. *J. Mol. Biol.* **433**, 166704
759 (2021).
- 760 82. Burley, S. K. *et al.* Electron microscopy holdings of the Protein Data Bank: the impact of the
761 resolution revolution, new validation tools, and implications for the future. *Biophys. Rev.* **14**, 1281–
762 1301 (2022).
- 763 83. Chen, V. B., Davis, I. W. & Richardson, D. C. KING (Kinemage, Next Generation): a versatile
764 interactive molecular and scientific visualization program. *Protein Sci.* **18**, 2403–2409 (2009).