

Outcomes of the EMDataResource cryo-EM Ligand Modeling Challenge

Received: 14 January 2024

Accepted: 24 May 2024

Published online: 25 June 2024

 Check for updates

A list of authors and their affiliations appears at the end of the paper

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

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

An international workshop on validation of ligands in crystallographic Protein Data Bank (PDB) depositions¹ held in 2015 identified

several common problems, including weak experimental density, ligand atoms poorly placed, incorrectly defined or misinterpreted chemical species, and inclusion of atoms not directly supported by experimental evidence. The main outcome was a set of best practice recommendations for PDB depositors and for the PDB archive. For PDB depositors, recommendations included providing unambiguous chemical definitions for all ligands present in a structure, including hydrogen atoms, providing ligand geometry and refinement restraints, clearly identifying atoms not supported by experimental evidence, providing the experimental map used for modeling and including comments explaining outliers. Recommendations for PDB validation included providing informative images of ligands in their density, providing stick figure diagrams indicating geometry outliers, identifying atoms not supported by experimental evidence, providing quality assessment metrics for each identified ligand and identifying possible protonation states. Most of the workshop validation recommendations have been implemented in PDB validation reports, with ligand geometric assessments implemented for all experimental methods^{2–4}.

Since 2010, EMDataResource (EMDR) has organized multiple Challenge activities (<https://challenges.emdataresource.org>) with the aim of bringing the cryo-EM community together to address important questions regarding the reconstruction and interpretation of maps

✉ e-mail: cathy.lawson@rutgers.edu; wahc@stanford.edu

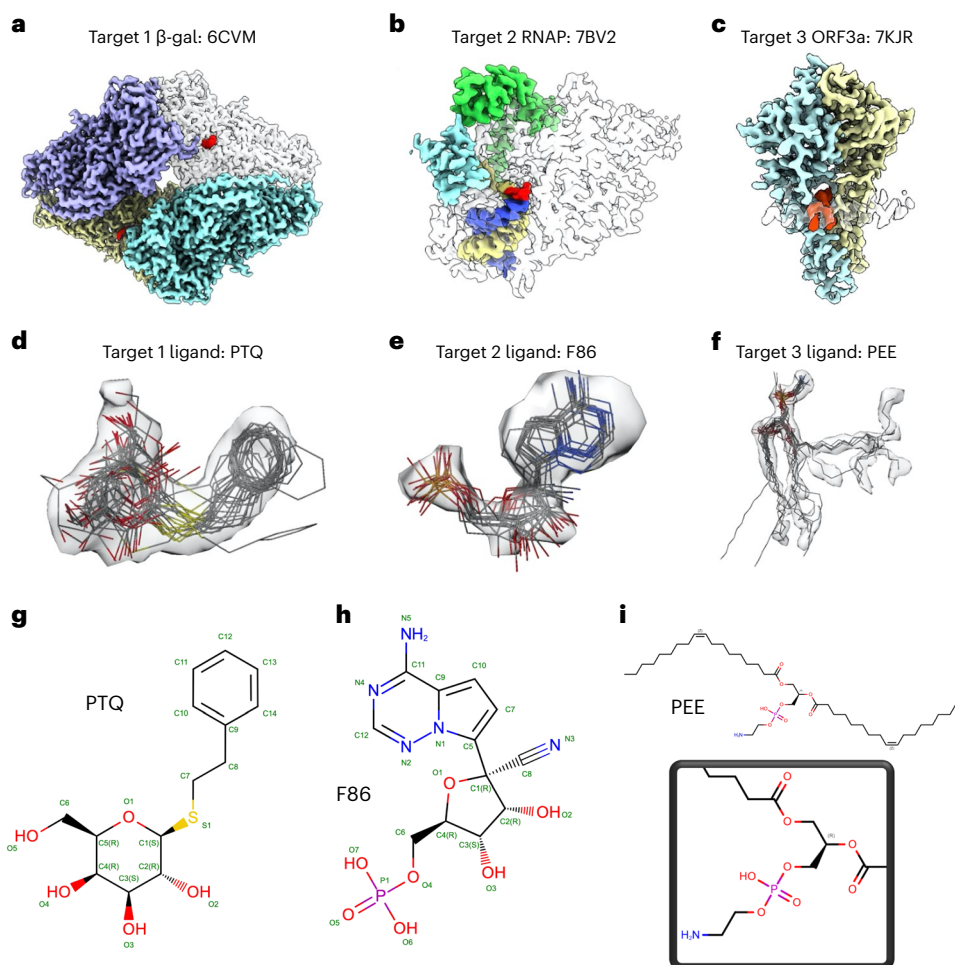


Fig. 1 | Ligand Challenge targets and ligands from submitted models.

a–c, Targets 1–3 are shown, with each polymer and/or nucleic acid chain rendered as a separate surface with a different color, in some cases semitransparent: target 1 (**a**), target 2 (**b**) and target 3 (**c**). Target ligands are shown in red. **d–f**, Segmented density representing each target ligand is shown with a semitransparent surface, with all submitted ligand models overlaid: target 1 (**d**), target 2 (**e**) and target 3 (**f**).

Map contour levels are 0.35 (2.3 σ), 0.036 (2.6 σ) and 0.25 (3.7 σ), respectively (sigma values were calculated from the full unmasked map to capture variation in background noise). **g–i**, Chemical sketches for each of the target ligands PTQ (**g**), F86 (**h**) and PEE (**i**) (source, PDB). Selected individual ligand poses from submitted models superimposed on target map densities are shown in Extended Data Fig. 2.

and map-derived atomic coordinate models⁵. For each Challenge, a committee consisting of prominent experts is invited to recommend targets and set goals. Each event has been conducted with the operational principles of fairness, transparency and openness, using modeler-blind assessments and open results, with a major goal of promoting innovation.

In 2016, paired Map and Model Challenges invited participants to apply their new algorithms and/or software to reconstruct maps and to evaluate models at resolutions of 2.9–4.5 Å. The results were published in a 19-article special journal issue⁶. By 2018, most participating groups had improved their pipelines, eliminating many identified mistakes. The unique EMRinger map metric for side chain–main chain consistency⁷ was first tested systematically in the 2016 Challenge and is now standard.

The 2019 Model Metrics Challenge evaluated models, while also evaluating the effectiveness of many different coordinate-only and map-model fit metrics for four targets at 1.7–3.3 Å resolution. The results were published in a single joint paper⁸. To streamline the challenge process, input of data from participants and initial assessment pipelines were automated, and comprehensive statistics, visualizations of scores and comparisons were made available. The CaBLAM multi-residue main-chain metric⁹, introduced in 2016, was shown in the 2019 Challenge to be the score most highly correlated with measures

of match to target. The *Q* score¹⁰, inspired and introduced by the 2019 Challenge, has now been adopted by the wwPDB Validation System used at deposition as well as in the detailed validation report¹¹.

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

Results

Challenge design

Three cryo-EM map targets were chosen based on the following criteria: recently published with resolution better than 3 Å, maps released in the

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, Schrödinger, 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, OpenBabel	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. Warshamane, 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, LAFER, 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					

T1, target 1; T2, target 2; T3, target 3.

Electron Microscopy Data Bank (EMDB), associated coordinates in the PDB, small molecules present (ligands, water, metal ions, detergent and/or lipid) and having current topical relevance (Fig. 1a–c):

- Target 1: 1.9 Å *Escherichia coli* β-galactosidase (β-gal) in complex with inhibitor 2-phenylethyl 1-thio-β-D-galactopyranoside (PETG) with PDB Chemical Component Dictionary (CCD) ID PTQ, EMDB map entry EMD-7770, PDB reference model 6CVM (ref. 12)
- Target 2: 2.5 Å SARS-CoV-2 RNA-dependent RNA polymerase (RNAP) with the pharmacologically active, nucleotide form of the prodrug remdesivir (CCD ID F86) covalently bound to RNA, EMD-30210, PDB reference model 7BV2 (refs. 13,14)
- Target 3: 2.1 Å SARS-CoV-2 Open Reading Frame 3a (ORF3a) putative ion channel in complex with 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine phospholipid (CCD ID PEE), EMD-22898, PDB reference model 7KJR (ref. 15)

Next, modeling teams were solicited via emails to multiple bulletin board lists and were asked to register, generate and upload optimized models for each target, following provided guidelines (Methods).

A total of 61 independently determined models were contributed by 17 teams from different institutions (IDs EM001–EM017), with workflow details collected for each (see the summary in Table 1, Supplementary Information, pp. 1–7 and Supplementary Data 1 for details).

Model assessments

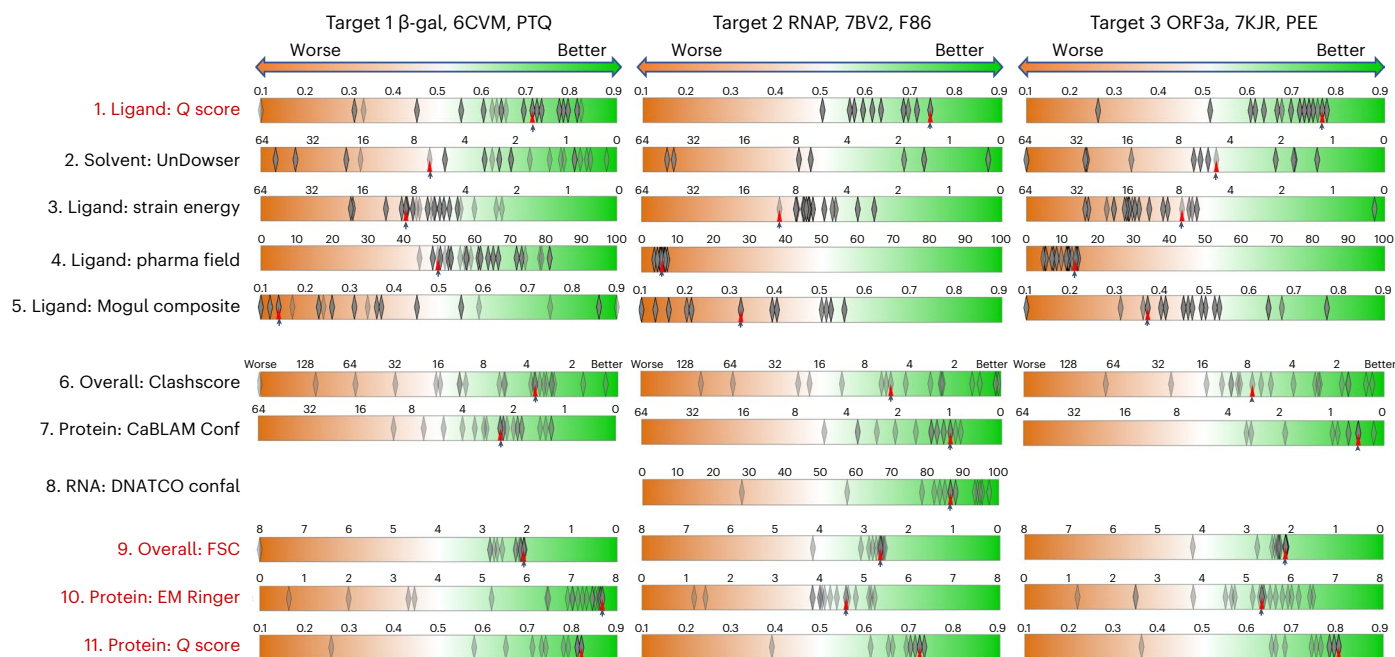
Submitted and PDB reference models for each target were evaluated by passing them through the Model Challenge validation pipeline^{8,16}. Individual scores were obtained for many different sets of metrics, with a new ligand analysis track added to the existing fit-to-map, coordinates-only, comparison-to-reference and comparison-among-models tracks.

Global fit-to-map metrics included map-model Fourier shell correlation (FSC)¹⁷, atom inclusion¹⁸ and EMRinger⁷ and density-based correlation scores from TEMPy¹⁹, Phenix²⁰ and Q score¹⁰.

The overall coordinates-only quality was evaluated using Clash-score, rotamer outliers, Ramachandran outliers and CaBLAM from MolProbity^{9,21}, as well as standard geometry measures (for example, bond, chirality, planarity) from Phenix²². Davis-QA, a measure used in critical assessment of protein structure prediction (CASP) competitions, was used to assess similarity among submitted models²³.

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, NNP
AT07	G. Pintilie, M. Schmid, W. Chiu	Q score analysis

**Fig. 2 | Model score distributions of selected assessments for targets 1–3.**

The top five rows show the ligand and solvent scores, the bottom six rows show overall and protein-specific scores. Fit-to-map-based metrics have red labels. Coordinates-only metrics have black labels. Diamonds indicate individual scores of submitted models (target 1 $n = 23$, target 2 $n = 17$, target 3 $n = 21$); red triangles

(with supporting black arrows) indicate the scores of the reference models; in a few cases no score is available for the reference model. Each score distribution is plotted against an orange (left) to white to green (right) color gradient with orange indicating poorer scores, and green indicating better scores, using a scale appropriate to the metric⁸. Red, fit to map and black, coordinates only.

Assessment teams contributed a wide variety of ligand-specific assessments (Table 2, IDs AT01–AT07) including ligand, ligand environment, solvent and RNA-specific analyses. AT01 used Mogul²⁴ to evaluate ligand covalent geometry as implemented in the wwPDB validation process^{2,4}, with inclusion of a composite ligand geometry ranking score²⁵. AT02 evaluated model ligands using Coot²⁶ and AceDRG²⁷. AT03 evaluated RNA conformation with DNATCO^{28,29} and solvent atom placement around protein residues using water distributions^{30,31}. AT04 analyzed ligand all-atom contacts with MolProbity Probescore⁹, and ion and water placements UnDowser³². AT05 scored ligand placements using density fields derived from pharmacophore consensus field analysis³³; a method used in computer-aided drug design to identify and extract possible interactions between a ligand–receptor complex based on steric and electronic features³⁴. AT06 examined ligand strain energies using both molecular mechanics and neural net potential (NNP) energy strategies^{35–37}; where strain energy is the calculated difference in energy between the modeled conformation and the lowest energy conformation in solution. AT07 prepared Q score analyses¹⁰ for model fit-to-map of whole models, protein, ligands and water, as well as ligand plus protein and/or

nucleic acid polymer atoms in the immediate vicinity of the ligand (ligand + immediate vicinity Q score or LIVQ).

Outcomes

The modeled ligands from each of the submissions are shown superimposed with their corresponding map density in Fig. 1d–f. Selected ligand and whole-model score distributions are shown for all three targets in Fig. 2. Selected individual ligand poses from submitted models superimposed on target map densities are shown in Extended Data Fig. 2. The full set of pipeline and assessment team scores and their definitions are provided in Supplementary Data 1 and online at model-compare.emdataresource.org.

Overall model scoring. With regards to overall fit-to-map evaluation, most submitted models scored very similarly to PDB reference models for all targets, both in terms of the overall map-model FSC¹⁷ and protein Q score¹⁰ (Fig. 2, rows 9 and 11). For targets 2 and 3, several teams modestly improved on EMRinger score⁷ (Fig. 2, columns 2 and 3, row 10). With regards to overall coordinates-only evaluation, many teams were able to improve on PDB reference models for all targets in terms

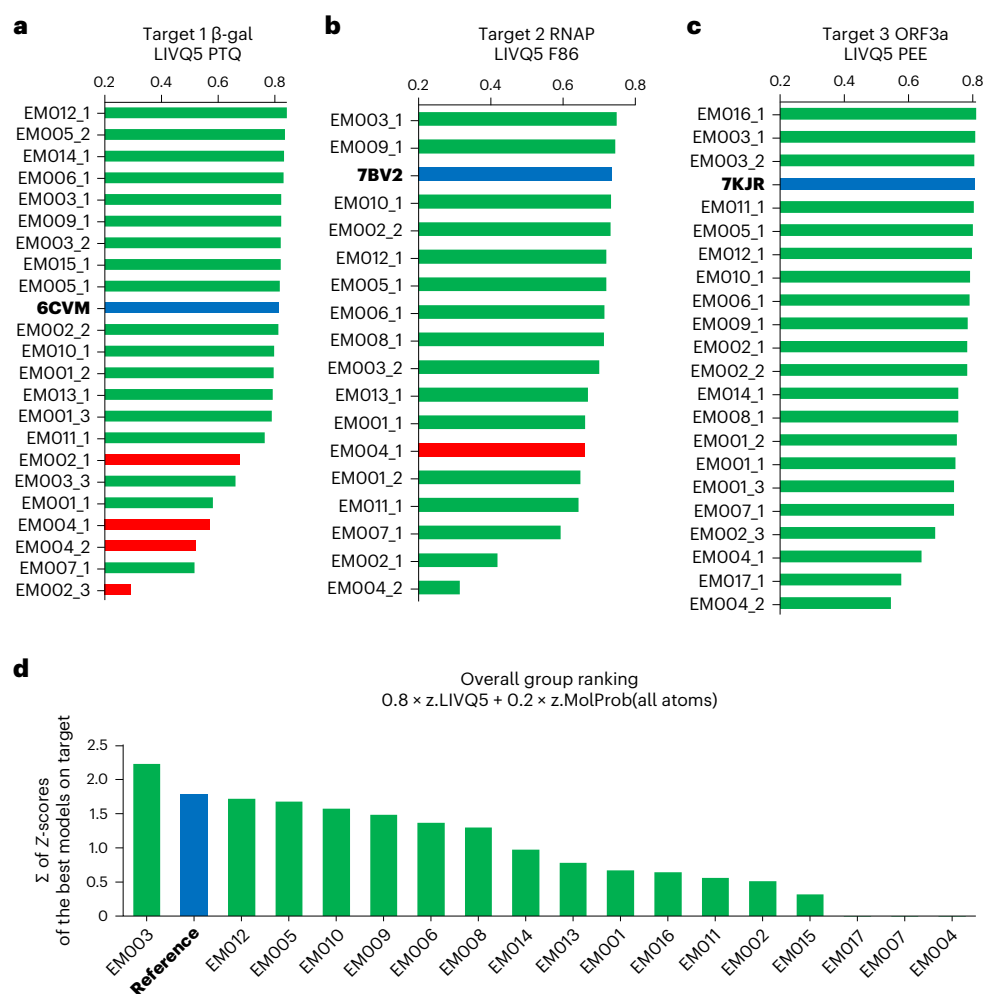


Fig. 3 | Model and modeling group rankings. **a–c**, LIVQ5 is plotted according to rank for each submitted model (labeled as participant group ID, Table 1, followed by model number) and for each reference model (labeled as PDB ID): target 1 (**a**), target 2 (**b**) and target 3 (**c**). Models with good overall MolProbity scores (<3.0) are shaded green; those with poor MP scores (>3.0) are shaded red and starred; reference models are shaded blue and labeled in bold. Immediate vicinity

includes all nonhydrogen model atoms ≤ 5 Å from any ligand nonhydrogen atom. Model rankings with extended vicinity (LIVQ10) are provided in Extended Data Fig. 4. **d**, Ranking of Challenge participant groups based on the fit-to-map accuracy of ligands as shown in **a–c**, and stereochemical plausibility, as described in the main text. Overall rank is calculated as the all-target sum of weighted z scores for the best per-target models from the group (equation in text).

of Clashscore³² and CaBLAM³², metrics that identify steric clashes and evaluate protein backbone geometry, respectively (Fig. 2, rows 6 and 7).

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

Nucleic acid scoring. Target 2's RNA (a typical A-form double helix, with two unpaired nucleotides at the 5' end of the template strand) had

close-to-expected geometries for most submitted models as assessed by DNATCO nucleic acid Confal scores^{28,29} (Fig. 2, column 2, row 8). Values of torsion angles in the dinucleotide units assigned to DNATCO NtC classes agreed with expected distributions including sugar ring torsions that define pucker. Note that before running this Challenge, target 2's reference model (PDB 7BV2) had been reverted by the deposition authors and rereleased by the PDB with several corrections to sequence, RNA conformation and CaBLAM outliers³⁸, thus limiting scope for model improvement.

Submitted model rankings. To evaluate and rank quality of ligand fit-to-map within the context of the macromolecular complex, we developed a new score, LIVQ, which averages *Q* scores of nonhydrogen atoms of the ligand together with all nonhydrogen polymer atoms in the immediate vicinity of ligand (LIVQ, ligand + immediate enVironment *Q* score). A distance cutoff of 5 Å was chosen to define the immediate vicinity of the ligand for model ranking purposes (LIVQ5, Fig. 3a–c); extension to 10 Å yielded similar results (LIVQ10, Extended Data Fig. 4). The results of the analysis show that for each target there are several models that exhibit very good model-to-map fit comparable to that of reference PDB-deposited models (Fig. 3a–c, blue bars). Nine, two and three submitted models

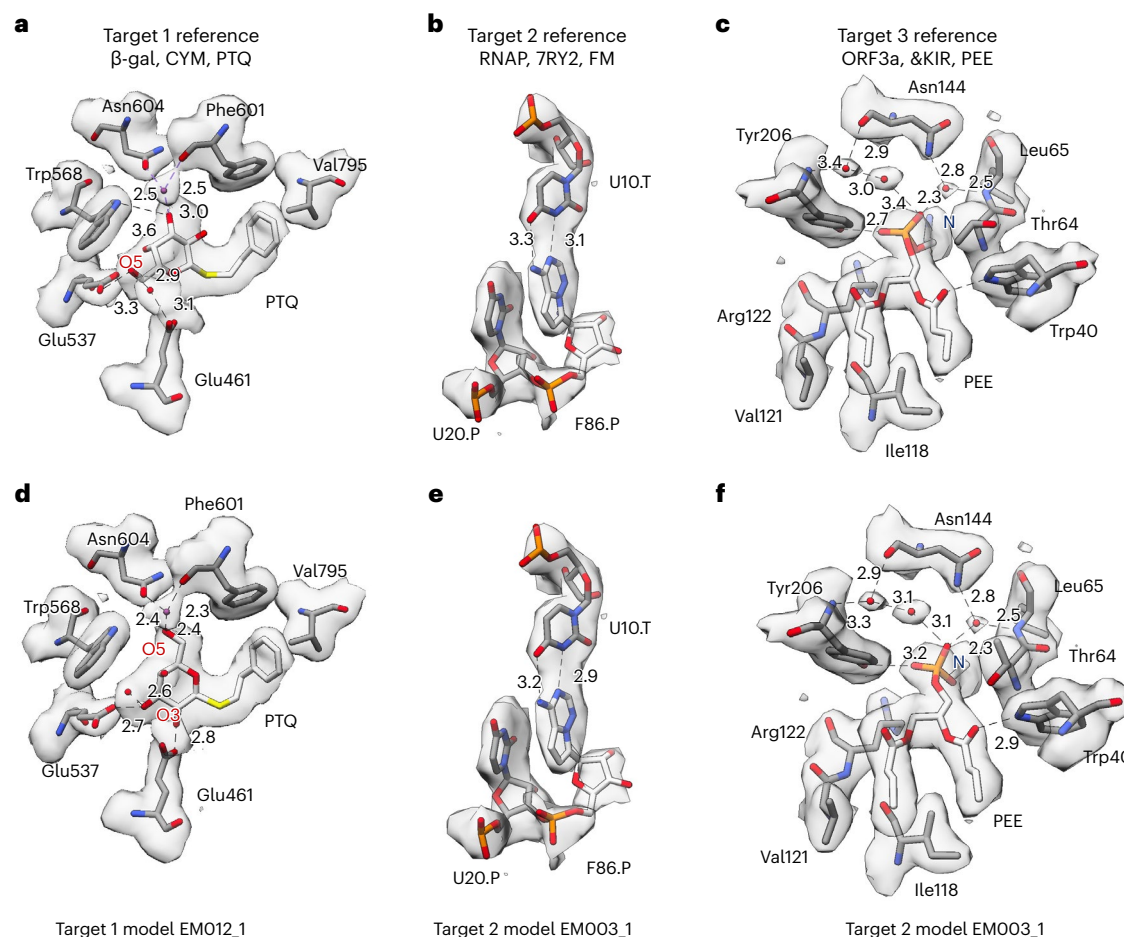


Fig. 4 | Visualization of ligands and surrounding atoms in deposited reference models and best-scoring submitted models. a–c. The deposited reference models for targets 1–3 as described in the main text: target 1 (a), target 2 (b) and target 3 (c). **d–f.** The best-scoring submitted models for each target:

target 1 (d), target 2 (e) and target 3 (f). Modeled solvent atoms are shown as red spheres; a modeled ion in a and d is shown as a dark blue sphere. Numerical labels with dashed lines indicate atom-to-atom distances in Ångströms.

respectively on targets 1–3 score better than the corresponding deposited reference model.

Group rankings. Overall ranking of participating groups (Fig. 3d) used a combination of LIVQ5 and MolProbity score, itself a weighted function of clashes, Ramachandran favored and rotamer outliers⁹. LIVQ5 was weighted higher than stereochemical plausibility, similar to the approach customarily used in CASP³⁹:

$$\text{rank} = \sum_{\text{target}=1\dots3} (0.8 \times z.\text{LIVQ5}_{\text{target}} + 0.2 \times z.\text{MolProbity}_{\text{target}})$$

where the z .metric is the number of standard deviations relative to the mean of the score distribution for all models from each group on the selected target according to the selected metric. Overall, group EM003 (DiMaio) had the best relative performance by this ranking criterion, being the only group that outscored all deposited reference PDB models (Fig. 3a–c).

Alternate group rankings. The model-compare website Group Ranking calculator enables users to explore other possible ranking formulas: z scores of up to 40 different individual metrics can be selected for inclusion with adjustable weighting. Extended Data Fig. 5 illustrates an alternate ranking method based on 13 different metrics including ligand, ligand+environment, full model coordinates-only and full model fit-to-map. By this alternate method, five groups ranked higher than PDB reference models: EM010 (Chojnowski), EM008 (Emsley),

EM012 (Palmer), EM003 (DiMaio) and EM009 (Moriarty), and one performed very close to reference, EM011 (Igaev).

Ligand quality. The ligand environment for the reference models and the best submitted models are compared for each target in Fig. 4.

For Target 1 (β-gal), the PTQ ligand O5 atom connected to the sugar ring is situated at the bottom of the binding pocket in the reference model (Fig. 4a) and in eight submitted models, whereas in the top-scoring model (Fig. 4d), as well as five other submitted models, the sugar ring is flipped with oxygen O5 situated at the top. The flipped ligand fits the density better and has more optimal interatomic distances to water and protein atoms for hydrogen-bonding, with O5 H-bonded to a coordinated water of the nearby magnesium ion. The density shape does not preclude the possibility that both original and flipped conformations are present, each with partial occupancy, and probescores for the two states are nearly identical (Extended Data Fig. 6a).

For Target 2 (RNAP), the F86 ligand is very similar for the deposited and top-scoring model (Fig. 4b,e, respectively), although distances to base-paired U10 are slightly different. F86 probescores varied greatly across models, with the reference at 10.1, model EM008_1 at 39.9 and the worst model at −106.9 (Extended Data Fig. 6). Many models did not correctly create the RNA polymer–F86 (remdesivir) covalent bond. In addition, only five models indicated partial occupancy for F86, yet the map density for F86 and its paired base is almost exactly half that of adjacent base pairs (Extended Data Fig. 6b), indicating 50% occupancy.

In the case of target 3 (ORF3a ion channel), the deposited and top-scoring PEE ligand models (Fig. 4c,f, respectively) have similar interactions to nearby atoms and placed water molecules, although with slightly different interatomic distances. The head-group amino N atom (which has no close contacts within 4 Å) points up in the deposited model but away from the camera view in the top-scoring model. The long lipid tails of PEE have lower density, with confusingly interlaced and gapped connectivity that indicates disorder; the ensemble of all PEE ligand models shown in Fig. 1f may be a more meaningful representation than any one individual model.

Discussion

The selected targets for the Ligand Challenge are some of the first structures deposited and released into PDB that contain ligands modeled into cryo-EM maps with resolution of 3 Å or better. Our Challenge results revealed that a deposited PDB model's ligand and local ligand environment may not be fully optimal in terms of concurrent fit-to-map and coordinates-only measures. For all three targets, and especially for target 1, adjustments in the ligand and/or ligand environment could be made to the deposited reference model that improved one or more validation criteria, as demonstrated by several modeler groups. Most of the submitted models were in the 'better' range, where tiny differences in measured scores become inconsequential. In our previous Challenge, we showed that overall fit-to-map and coordinates-only metrics are orthogonal measures⁸; here we see that at a local level, ligand and/or ligand-environment fit-to-map and coordinates-only metrics are similarly independent (Fig. 2, Extended Data Fig. 5b and Supplementary Data 2). In other words, ligands that fit well into density may not be optimized with respect to ligand coordinates-only validation criteria, and vice versa. The inclusion of environment atoms in the LIVQ5 score explains the difference in group rankings reported here in Fig. 3d versus those reported in ref. 40, based solely on ligand *Q* scores.

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

Recommendation 1, regarding validation of the macromolecular models: for ligand-macromolecular complexes, the macromolecular model should be subject to standard geometric checks as done for X-ray crystallographic based models¹. These include standard covalent geometry checks and MolProbity evaluation, including CaBLAM, Clashscore^{9,21,32}, sugar pucker and DNATCO^{28,29} conformational analysis should be checked for nucleic acid components. The macromolecular model-map fit should be evaluated by EMRinger⁷, *Q* score¹⁰ and FSC¹⁷. Serious local outliers (which usually indicate an incorrect local conformation) should be emphasized, rather than overall average scores.

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

Recommendation 2, regarding validation of ligand models: ligands in macromolecular complexes should conform to standard covalent geometry measures (bond lengths, angles, planarity, chirality) as recommended by the wwPDB validation report^{2,4}. Additional checks that should be applied to ligands include fit to density using methods applicable to cryo-EM such as *Q* score, occupancy (density strength, both absolute and relative to surroundings), and identification of missing atoms, including any surrounding ions.

Ligand energetics should also be examined. Ligand models should be assessed for their strain energy (the calculated difference in energy between the modeled conformation and the lowest energy conformation in solution) to identify improbable model geometries and lower energy alternatives^{35,36}. Other methods can be used but may have different thresholds due to variation in absolute energy values. Strain energy calculations using NNPs offer speed close to force fields with the accuracy of quantum mechanics calculations and are predicted to play a primary role in identifying accurate strain energies in the

future. More research is needed to evaluate the overall utility of these new deep learning methods.

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

During the virtual wrap-up workshop, modelers and assessors shared their experiences and strategies to identify and/or assess the correct pose for the ligand based on the cryo-EM density maps. It was noted that the local map resolution for a ligand can be worse than the overall map resolution. As one objective measure, *Q* scores were found to be lower for ligands in the best submitted models than for the nearby environment (Extended Data Table 1). Factors that may affect resolvability of local ligand map features include incomplete occupancy, multiple conformations and/or poses present, regions of ligand flexibility or disorder, chemical modifications and radiation damage.

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

Recommendation 5, regarding topics for future Challenges: for future Challenge topics, validation of RNA models should be considered, including identification of RNA-associated ions, owing to the rapidly rising numbers of RNA-containing cryo-EM structures^{41–43}. We also recommend maps determined in the 3.5–10 Å resolution range be considered as future targets to reflect the rapid rise in depositions of maps from subtomogram averaging of components in cell tomograms^{44–46}. There are very few validation tools for that resolution range.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-024-02321-7>.

References

- Adams, P. D. et al. Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop. *Structure* **24**, 502–508 (2016).
- Gore, S. et al. Validation of structures in the Protein Data Bank. *Structure* **25**, 1916–1927 (2017).
- Smart, O. S. et al. Validation of ligands in macromolecular structures determined by X-ray crystallography. *Acta Crystallogr. D. Struct. Biol.* **74**, 228–236 (2018).
- Feng, Z. et al. Enhanced validation of small-molecule ligands and carbohydrates in the Protein Data Bank. *Structure* **29**, 393–400.e1 (2021).
- Lawson, C. L., Berman, H. M. & Chiu, W. Evolving data standards for cryo-EM structures. *Struct. Dyn.* **7**, 014701 (2020).
- Lawson, C. L. & Chiu, W. Comparing cryo-EM structures. *J. Struct. Biol.* **204**, 523–526 (2018).
- Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
- Lawson, C. L. et al. Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. *Nat. Methods* **18**, 156–164 (2021).

9. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
10. Pintilie, G. et al. Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* **17**, 328–334 (2020).
11. Wang, Z., Patwardhan, A. & Kleywegt, G. J. Validation analysis of EMDB entries. *Acta Crystallogr. D. Struct. Biol.* **78**, 542–552 (2022).
12. Bartesaghi, A. et al. Atomic resolution cryo-EM structure of β -Galactosidase. *Structure* **26**, 848–856.e3 (2018).
13. Yin, W. et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499–1504 (2020).
14. Kokic, G. et al. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* **12**, 279 (2021).
15. Kern, D. M. et al. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nat. Struct. Mol. Biol.* **28**, 573–582 (2021).
16. Kryshchuk, A., Adams, P. D., Lawson, C. L. & Chiu, W. Evaluation system and web infrastructure for the second cryo-EM model challenge. *J. Struct. Biol.* **204**, 96–108 (2018).
17. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
18. Lagerstedt, I. et al. Web-based visualisation and analysis of 3D electron-microscopy data from EMDB and PDB. *J. Struct. Biol.* **184**, 173–181 (2013).
19. Joseph, A. P., Lagerstedt, I., Patwardhan, A., Topf, M. & Winn, M. Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy. *J. Struct. Biol.* **199**, 12–26 (2017).
20. Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D. Struct. Biol.* **74**, 814–840 (2018).
21. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).
22. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D. Struct. Biol.* **75**, 861–877 (2019).
23. Kryshchuk, A. et al. Challenging the state of the art in protein structure prediction: highlights of experimental target structures for the 10th Critical Assessment of Techniques for Protein Structure Prediction Experiment CASP10. *Proteins* **82**, 26–42 (2014).
24. Bruno, I. J. et al. Retrieval of crystallographically-derived molecular geometry information. *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144 (2004).
25. Shao, C. et al. Simplified quality assessment for small-molecule ligands in the Protein Data Bank. *Structure* **30**, 252–262.e4 (2022).
26. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1069–1078 (2020).
27. Nicholls, R. A. et al. Modelling covalent linkages in CCP4. *Acta Crystallogr. D. Struct. Biol.* **77**, 712–726 (2021).
28. Černý, J., Božíková, P., Svoboda, J. & Schneider, B. A unified dinucleotide alphabet describing both RNA and DNA structures. *Nucleic Acids Res.* **48**, 6367–6381 (2020).
29. Černý, J. et al. Structural alphabets for conformational analysis of nucleic acids available at dnatco.datmos.org. *Acta Crystallogr. D. Struct. Biol.* **76**, 805–813 (2020).
30. Biedermannová, L. & Schneider, B. Structure of the ordered hydration of amino acids in proteins: analysis of crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* **71**, 2192–2202 (2015).
31. Černý, J., Schneider, B. & Biedermannová, L. WatAA: Atlas of Protein Hydration. Exploring synergies between data mining and ab initio calculations. *Phys. Chem. Chem. Phys.* **19**, 17094–17102 (2017).
32. Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S. & Richardson, D. C. New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink ‘waters,’ and NGL Viewer to recapture online 3D graphics. *Protein Sci.* **29**, 315–329 (2020).
33. Jiang, S., Feher, M., Williams, C., Cole, B. & Shaw, D. E. AutoPH4: an automated method for generating pharmacophore models from protein binding pockets. *J. Chem. Inf. Model.* **60**, 4326–4338 (2020).
34. Tyagi, R., Singh, A., Chaudhary, K. K. & Yadav, M. K. in *Bioinformatics* (eds Singh, D. B. & Pathak, R. K.) Ch. 17, 269–289 (Academic, 2022).
35. Sellers, B. D., James, N. C. & Gobbi, A. A comparison of quantum and molecular mechanical methods to estimate strain energy in druglike fragments. *J. Chem. Inf. Model.* **57**, 1265–1275 (2017).
36. Lee, M.-L. et al. chemalot and chemalot_knime: command line programs as workflow tools for drug discovery. *J. Cheminform.* **9**, 38 (2017).
37. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
38. Croll, T. I., Williams, C. J., Chen, V. B., Richardson, D. C. & Richardson, J. S. Improving SARS-CoV-2 structures: peer review by early coordinate release. *Biophys. J.* **120**, 1085–1096 (2021).
39. Modi, V., Xu, Q., Adhikari, S. & Dunbrack, R. L. Jr. Assessment of template-based modeling of protein structure in CASP11. *Proteins* **84**, 200–220 (2016).
40. Giri, N. & Cheng, J. Improving protein-ligand interaction modeling with cryo-EM data, templates, and deep learning in 2021 Ligand Model Challenge. *Biomolecules* **13**, 132 (2023).
41. Zhang, K. et al. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat. Commun.* **10**, 5511 (2019).
42. Su, Z. et al. Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 Å resolution. *Nature* **596**, 603–607 (2021).
43. Lawson, C. L., Berman, H. M., Chen, L., Vallat, B. & Zirbel, C. L. The Nucleic Acid Knowledgebase: a new portal for 3D structural information about nucleic acids. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkad957> (2023).
44. Sun, S. Y. et al. Cryo-ET of parasites gives subnanometer insight into tubulin-based structures. *Proc. Natl Acad. Sci. USA* **119**, e2111661119 (2022).
45. Liu, H.-F. et al. nextPYP: a comprehensive and scalable platform for characterizing protein variability in situ using single-particle cryo-electron tomography. *Nat. Methods* **20**, 1909–1919 (2023).
46. Chmielewski, D. et al. Structural insights into the modulation of coronavirus spike tilting and infectivity by hinge glycans. *Nat. Commun.* **14**, 7175 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

Catherine L. Lawson¹✉, Andriy Kryshchak², Grigore D. Pintilie³, Stephen K. Burley^{1,4,5,6}, Jiří Černý⁷, Vincent B. Chen⁸, Paul Emsley⁹, Alberto Gobbi^{10,43}, Andrzej Joachimiak^{11,12}, Sigrid Noreng^{13,44}, Michael G. Prisant⁸, Randy J. Read¹⁴, Jane S. Richardson¹⁵, Alexis L. Rohou¹³, Bohdan Schneider¹⁶, Benjamin D. Sellers^{10,45}, Chenghua Shao¹, Elizabeth Sourial¹⁵, Chris I. Williams¹⁵, Christopher J. Williams¹⁶, Ying Yang¹³, Venkat Abbaraju¹, Pavel V. Afonine¹⁶, Matthew L. Baker¹⁷, Paul S. Bond¹⁸, Tom L. Blundell¹⁹, Tom Burnley²⁰, Arthur Campbell²¹, Renzhi Cao²², Jianlin Cheng²³, Grzegorz Chojnowski²⁴, K. D. Cowtan¹⁸, Frank DiMaio²⁵, Reza Esmaeeli²⁶, Nabin Giri²³, Helmut Grubmüller²⁷, Soon Wen Hoh¹⁸, Jie Hou²⁸, Corey F. Hryc¹⁷, Carola Hunte²⁹, Maxim Igaev²⁷, Agnel P. Joseph²⁰, Wei-Chun Kao²⁹, Daisuke Kihara^{30,31}, Dilip Kumar^{32,46}, Lijun Lang^{26,47}, Sean Lin³³, Sai R. Maddhuri Venkata Subramaniya³¹, Sumit Mittal^{34,35}, Arup Mondal^{26,48}, Nigel W. Moriarty¹⁶, Andrew Muenks²⁵, Garib N. Murshudov⁹, Robert A. Nicholls^{18,20}, Mateusz Olek^{18,36}, Colin M. Palmer²⁰, Alberto Perez²⁶, Emmi Pohjolainen²⁷, Karunakar R. Pothula³⁷, Christopher N. Rowley³⁸, Daipayan Sarkar^{30,34,49,50}, Luisa U. Schäfer³⁷, Christopher J. Schlicksup¹⁶, Gunnar F. Schröder^{37,39}, Mrinal Shekhar^{21,34}, Dong Si³³, Abhishek Singharoy³⁴, Oleg V. Sobolev¹⁶, Genki Terashi³⁰, Andrea C. Vaiana^{27,40}, Sundeep C. Vedithi¹⁹, Jacob Verburg³⁰, Xiao Wang³¹, Rangana Warshamange⁹, Martyn D. Winn²⁰, Simone Weyand¹⁹, Keitaro Yamashita⁹, Minglei Zhao¹², Michael F. Schmid⁴¹, Helen M. Berman^{4,42} & Wah Chiu^{3,41}✉

¹RCSB Protein Data Bank and Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. ²Genome Center, University of California, Davis, CA, USA. ³Departments of Bioengineering and of Microbiology and Immunology, Stanford University, Stanford, CA, USA. ⁴Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. ⁵Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA. ⁶RCSB Protein Data Bank and San Diego Supercomputer Center, University of California San Diego, La Jolla, CA, USA. ⁷Institute of Biotechnology, Czech Academy of Sciences, Vestec, Czech Republic. ⁸Department of Biochemistry, Duke University, Durham, NC, USA. ⁹MRC Laboratory of Molecular Biology, Cambridge, UK. ¹⁰Discovery Chemistry, Genentech Inc., San Francisco, CA, USA. ¹¹Structural Biology Center, X-ray Science Division, Argonne National Laboratory, Argonne, IL, USA. ¹²Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, USA. ¹³Structural Biology, Genentech Inc., South San Francisco, CA, USA. ¹⁴Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. ¹⁵Chemical Computing Group, Montreal, Quebec, Canada. ¹⁶Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ¹⁷Department of Biochemistry and Molecular Biology, The University of Texas Health Science Center at Houston, Houston, TX, USA. ¹⁸York Structural Biology Laboratory, Department of Chemistry, University of York, York, UK. ¹⁹Department of Biochemistry, University of Cambridge, Cambridge, UK. ²⁰Scientific Computing Department, UKRI Science and Technology Facilities Council, Research Complex at Harwell, Didcot, UK. ²¹Center for Development of Therapeutics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ²²Department of Computer Science, Pacific Lutheran University, Tacoma, WA, USA. ²³Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA. ²⁴European Molecular Biology Laboratory, Hamburg Unit, Hamburg, Germany. ²⁵Department of Biochemistry and Institute for Protein Design, University of Washington, Seattle, WA, USA. ²⁶Department of Chemistry and Quantum Theory Project, University of Florida, Gainesville, FL, USA. ²⁷Theoretical and Computational Biophysics Department, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany. ²⁸Department of Computer Science, Saint Louis University, St. Louis, MO, USA. ²⁹Institute of Biochemistry and Molecular Biology, ZBMZ, Faculty of Medicine and CIBSS—Centre for Integrative Biological Signalling Studies, University of Freiburg, Freiburg, Germany. ³⁰Department of Biological Sciences, Purdue University, West Lafayette, IN, USA. ³¹Department of Computer Science, Purdue University, West Lafayette, IN, USA. ³²Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, USA. ³³Division of Computing & Software Systems, University of Washington, Bothell, WA, USA. ³⁴Biodesign Institute, Arizona State University, Tempe, AZ, USA. ³⁵School of Advanced Sciences and Languages, VIT Bhopal University, Bhopal, India. ³⁶Electron Bio-Imaging Centre, Diamond Light Source, Harwell Science and Innovation Campus, Didcot, UK. ³⁷Institute of Biological Information Processing (IBI-7, Structural Biochemistry) and Jülich Centre for Structural Biology (JuStruct), Forschungszentrum Jülich, Jülich, Germany. ³⁸Department of Chemistry, Carleton University, Ottawa, Ontario, Canada. ³⁹Physics Department, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. ⁴⁰Nature's Toolbox (NTx), Rio Rancho, NM, USA. ⁴¹Division of Cryo-EM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. ⁴²Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA. ⁴³Present address: Berlin, Germany. ⁴⁴Present address: Protein Science, Septerna, South San Francisco, CA, USA. ⁴⁵Present address: Computational Chemistry, Vilya, South San Francisco, CA, USA. ⁴⁶Present address: Trivedi School of Biosciences, Ashoka University, Sonapat, India. ⁴⁷Present address: The Chinese University of Hong Kong, Hong Kong, China. ⁴⁸Present address: National Renewable Energy Laboratory (NREL), Golden, CO, USA. ⁴⁹Present address: MSU-DOE Plant Research Laboratory, East Lansing, MI, USA. ⁵⁰Present address: School of Molecular Sciences, Arizona State University, Tempe, AZ, USA. ✉e-mail: cathy.lawson@rutgers.edu; wahc@stanford.edu

Methods

Challenge process and organization

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

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

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

Members of cryo-EM and modeling communities were invited to participate in February 2021 and details were posted at challenges.emdataresource.org. Models were submitted by participant teams between 1 March and 15 April. For each submitted model, metadata describing the full modeling workflow were collected via a Drupal webform (Supplementary Data 1 and 2), and coordinates were uploaded and converted to PDBx/mmCIF format using PDBextract⁴⁷. Model coordinates were then processed for atom and/or residue ordering and nomenclature consistency using PDB annotation software (Feng, Z., <https://sw-tools.rcsb.org/apps/MAXIT>) and additionally checked for sequence consistency, ligand atom naming and correct position relative to the designated target map. Models were then evaluated as described below (in the 'Model evaluation system' section).

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

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

Data sources and modeling

Target maps were obtained from EMDb⁴⁸: target 1 *E. coli* β -galactosidase-PETG¹², EMD-7770; target 2 SARS-CoV-2 RNAP or Remdesivir¹³, EMD-30210 and target 3 SARS-CoV-2 ORF3a putative ion channel and/or phospholipid in nanodisc¹⁵, EMD-22898.

Table 1 summarizes the approach and lists the software used by each modeling team. Further details for each model can be found in Supplementary Data 1. Modeling teams categorized their polymer modeling type as ab initio (followed by optimization), optimized or not optimized. Non-ab initio approaches made use of polymer coordinates from the following PDB entries: target 1, 6CVM, 1JZ7 and 6TTE; target 2, 7BV2, 7B3D, 6X71 and 3OVB and target 3, 7KJR.

Submitted models were further categorized by ligand modeling type, either independently refit or optimized. Initial ligand coordinates and restraints were obtained from the PDB CCD⁴⁹, Crystallography Open Database⁵⁰ or from a PDB entry. Ligand restraint generation software included BUSTER Grade (Global Phasing Ltd), Phenix eLBOW⁵¹, CCP4 AceDRG⁵², PyRosetta⁵³, AMBER Antechamber⁵⁴, OpenBabel⁵⁵, CHARMM CGenFF⁵⁶, LigPrep (Schrödinger LLC) and CCP4 monomer library⁵⁷. Restraints were not applied by teams using molecular dynamics-based approaches.

Ab initio modeling software included ARP/wARP⁵⁸, Mainmast⁵⁹, Mainmastseg⁶⁰, Pathwalker⁶¹, Rosetta⁶², Modeller⁶³ and DeepTracer^{64,65}. Model optimization software included CDMD⁶⁶, Phenix²², REFMAC⁶⁷, Servcat⁶⁸, ProSMART⁶⁹, MDFF⁷⁰, CryoFold^{71,72}, Amber⁵⁴, MELD^{73,74} and Schrödinger (Schrödinger LLC). The program doubleHelix⁷⁵ was used to assign RNA sequence and refinement restraints. Atomic displacement parameters (*B* factors) were optimized for 32 of 61 models, with 23 applying individual atomic *B* factors.

Participants made use of VMD⁷⁶, Chimera⁷⁷, ChimeraX⁷⁸, Coot²⁶, ISOLDE⁷⁹, EMDA⁸⁰ and PyMOL for visual evaluation and/or manual model improvement of map-model fit. Manipulation of map densities was carried out using CCP-EM⁸¹, EMDA and LAFTER⁸².

Model evaluation system

The evaluation pipeline for the 2021 challenge (model-compare.emdataresource.org) was built on the basis of the 2019 Model Challenge pipeline⁸. Submitted models were evaluated for >70 individual metrics in four established tracks using the software packages as follows: fit-to-map EMDb Cryo-EM Validation Analysis¹¹, TEMPy¹⁹, Phenix²² and UCSF ChimeraX⁷⁸/MapQ¹⁰; coordinates-only, Phenix²² and comparison-to-reference CAD⁸³, HBPLUS⁸⁴, LGA⁸⁵, MMalign⁸⁶, OpenStructure⁸⁷ and Phenix²². A new ligand track was also created for comparison of ligand-specific scores. Ligand and nucleic acid-specific scores provided by Assessor teams (Table 2) were integrated into data tables alongside scores from the evaluation pipeline to enable comparisons and composite score generation (Supplementary Data 2).

Pharmacophore modeling

The Molecular Operating Environment (MOE) platform was used to score the placement of ligands. Starting from the model coordinates submitted by each group, the MOE QuickPrep application was used to prepare all-atom structures with hydrogens and atomic partial charges. For each target, an ensemble of structures consisting of all submitted models was input into the db_AutoPH4 application to produce pharmacophore consensus fields based on the ensemble. The pharmacophore consensus fields were then used to score the ligand

poses of each submission. Additional details are provided in the Supplementary Information.

Strain energy calculations

To prepare, ligands were extracted from model files. For the T2 F86 ligand, strain energy was measured after deleting the covalent bond to the RNA polymer (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 models were truncated to just the head group (SMILES: CCC(OC[C@@H](OC(CC)=O)CO[P@]([O-])(OCC[NH3+])=O)=O). Hydrogens were added using MOE/Protonate3D from the Chemical Computing Group.

To calculate molecular mechanics force field strain energy, predicted ligand energy was calculated by minimizing each ligand structure using OpenEye/SZYBKI (MMFF94S with the Sheffield solvation model) with a maximum root mean-square deviation of 0.6 Å. Predicted global minimum energy was identified by sampling conformations using OpenEye/Omega and then minimizing each conformer structure using OpenEye/SZYBKI (MMFF94S with Sheffield solvation model) with no restraints, then selecting the conformer with the lowest minimized energy.

To calculate NNP energy, predicted ligand energy was calculated by minimizing each ligand structure in an in-house implementation of the ANI NNP³⁷ with a maximum root mean-square deviation of 0.6 Å. Predicted global minimum energy was identified by sampling conformations using OpenEye/Omega and then minimizing each conformer structure using the same in-house implementation of the ANI NNP with no restraints.

Reported scores are predicted strain energy as (predicted ligand energy – global minimum energy) in kcal mol⁻¹. NNP was only calculated for the T1 ligand as the method currently does not support atomic charges.

Molecular graphics

Molecular graphics images were generated using UCSF Chimera (Figs. 1 and 4 and Extended Data Fig. 2) and KING⁸⁸ (Extended Data Figs. 3 and 6).

Classification of unique ligands in PDB introduced by cryo-EM

Search of the PDB via RCSB PDB's data API⁸⁹ identified 981 unique nonpolymer ligands and/or PDB CCD IDs in EM-derived PDB structures released through December 2021. Next, for each ligand, the PDB entry that first introduced the ligand/CCD ID was identified. The 403 unique nonpolymer ligands that were found to be introduced in structures determined by cryo-EM were then manually classified as enzyme modulators (substrates, inhibitors, agonists, cofactors), medically relevant drugs, lipids, photochemicals (for example carotenoids), peptides (amino-acid-based), reagents (buffers or labels), nucleotides or steroids (fused rings).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Cryo-EM map targets were the primary maps of EMD entries EMD-7770, EMD-30210 and EMD-22898 (www.ebi.ac.uk/emdb, emdataresource.org). Reference models were PDB entries 6CVM v.1.3 (target 1), 7BV2 v.3.4 (target 2) and 7KJR v.1.1 (target 3) wwpdb.org. Submitted models, model metadata, result logs and compiled data are available via challenges.emdataresource.org/?q=2021-model-challenge and archived via Zenodo at <https://doi.org/10.5281/zenodo.10551958> (ref. 90). Interactive summary tables, graphical views and spreadsheet downloads of compiled results are available at model-compare.emdataresource.org/2021/cgi-bin. Source Data are provided with this paper.

References

- Yang, H. et al. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 1833–1839 (2004).
- wwPDB Consortium. EMDB—the Electron Microscopy Data Bank. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkad1019> (2023).
- Westbrook, J. D. et al. The Chemical Component Dictionary: complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank. *Bioinformatics* **31**, 1274–1278 (2015).
- Gražulis, S. et al. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res.* **40**, D420–D427 (2012).
- Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. Electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D. Biol. Crystallogr.* **65**, 1074–1080 (2009).
- Nicholls, R. A. et al. The missing link: covalent linkages in structural models. *Acta Crystallogr. D. Struct. Biol.* **77**, 727–745 (2021).
- Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
- O'Boyle, N. M. et al. Open Babel: an open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
- Vanommeslaeghe, K. et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
- Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2184–2195 (2004).
- Chojnowski, G., Sobolev, E., Heuser, P. & Lamzin, V. S. The accuracy of protein models automatically built into cryo-EM maps with ARP/wARP. *Acta Crystallogr. D. Struct. Biol.* **77**, 142–150 (2021).
- Terashi, G. & Kihara, D. De novo main-chain modeling for EM maps using MAINMAST. *Nat. Commun.* **9**, 1618 (2018).
- Terashi, G., Kagaya, Y. & Kihara, D. MAINMASTseg: automated map segmentation method for cryo-EM density maps with symmetry. *J. Chem. Inf. Model.* **60**, 2634–2643 (2020).
- Chen, M. & Baker, M. L. Automation and assessment of de novo modeling with pathwalking in near atomic resolution cryoEM density maps. *J. Struct. Biol.* **204**, 555–563 (2018).
- DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of protein structures into low-resolution density maps using rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
- Webb, B. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **1137**, 1–15 (2014).
- Si, D. et al. Deep learning to predict protein backbone structure from high-resolution cryo-EM density maps. *Sci. Rep.* **10**, 4282 (2020).
- Pfaff, J., Phan, N. M. & Si, D. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proc. Natl Acad. Sci. USA* **118**, e2017525118 (2021).
- Igaev, M., Kutzner, C., Bock, L. V., Vaiana, A. C. & Grubmüller, H. Automated cryo-EM structure refinement using correlation-driven molecular dynamics. *eLife* **8**, e43542 (2019).
- Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D. Biol. Crystallogr.* **71**, 136–153 (2015).

68. Yamashita, K., Palmer, C. M., Burnley, T. & Murshudov, G. N. Cryo-EM single-particle structure refinement and map calculation using Servalcat. *Acta Crystallogr. D. Struct. Biol.* **77**, 1282–1291 (2021).
69. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D. Biol. Crystallogr.* **70**, 2487–2499 (2014).
70. Singharoy, A. et al. Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. *eLife* **5**, e16105 (2016).
71. Shekhar, M. et al. CryoFold: determining protein structures and data-guided ensembles from cryo-EM density maps. *Matter* **4**, 3195–3216 (2021).
72. Chang, L., Mondal, A., MacCallum, J. L. & Perez, A. CryoFold 2.0: cryo-EM structure determination with MELD. *J. Phys. Chem. A* **127**, 3906–3913 (2023).
73. MacCallum, J. L., Perez, A. & Dill, K. A. Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. *Proc. Natl Acad. Sci. USA* **112**, 6985–6990 (2015).
74. Perez, A., MacCallum, J. L. & Dill, K. A. Accelerating molecular simulations of proteins using Bayesian inference on weak information. *Proc. Natl Acad. Sci. USA* **112**, 11846–11851 (2015).
75. Chojnowski, G. DoubleHelix: nucleic acid sequence identification, assignment and validation tool for cryo-EM and crystal structure models. *Nucleic Acids Res.* **51**, 8255–8269 (2023).
76. Hsin, J., Arkhipov, A., Yin, Y., Stone, J. E. & Schulten, K. Using VMD: an introductory tutorial. *Curr. Protoc. Bioinforma.* **Chapter 5**, Unit 5.7 (2008).
77. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
78. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
79. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D. Struct. Biol.* **74**, 519–530 (2018).
80. Warshamanage, R., Yamashita, K. & Murshudov, G. N. EMDA: a Python package for electron microscopy data analysis. *J. Struct. Biol.* **214**, 107826 (2022).
81. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr. D. Struct. Biol.* **73**, 469–477 (2017).
82. Ramlal, K., Palmer, C. M. & Aylett, C. H. S. A local agreement filtering algorithm for transmission EM reconstructions. *J. Struct. Biol.* **205**, 30–40 (2019).
83. Olechnovič, K. & Venclovas, Č. Contact area-based structural analysis of proteins and their complexes using CAD-score. *Methods Mol. Biol.* **2112**, 75–90 (2020).
84. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
85. Zemla, A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **31**, 3370–3374 (2003).
86. Mukherjee, S. & Zhang, Y. MM-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. *Nucleic Acids Res.* **37**, e83 (2009).
87. Biasini, M. et al. OpenStructure: an integrated software framework for computational structural biology. *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 701–709 (2013).
88. Chen, V. B., Davis, I. W. & Richardson, D. C. KING (Kinemage, Next Generation): a versatile interactive molecular and scientific visualization program. *Protein Sci.* **18**, 2403–2409 (2009).
89. Rose, Y. et al. RCSB Protein Data Bank: architectural advances towards integrated searching and efficient access to macromolecular structure data from the PDB Archive. *J. Mol. Biol.* **433**, 166704 (2021).
90. Lawson, C. L. et al. 2021 EMDDataResource Ligand Model Challenge dataset. *Zenodo* <https://doi.org/10.5281/zenodo.10551958> (2024).
91. Burley, S. K. et al. Electron microscopy holdings of the Protein Data Bank: the impact of the resolution revolution, new validation tools, and implications for the future. *Biophys. Rev.* **14**, 1281–1301 (2022).

Acknowledgements

EMDataResource (C.L.L., A.K., G.D.P., H.M.B., W.C.) was supported by the US National Institutes of Health/National Institute of General Medical Science (NIH/NIGMS) grant no. R01GM079429. The following additional grants are acknowledged for participant support. American Leprosy Missions grant no. G88726 to S.C.V. Biotechnology and Biological Sciences Research Council grant no. BB/S005099/1 to P.S.B. and K.D.C. Biotechnology and Biological Sciences Research Council grant no. BB/T012935/1 to S.W.H. Biotechnology and Biological Sciences Research Council grant no. BB/S007083/1 to R.A.N. German Research Foundation grant no. CIBSS – EXC-2189 – 390939984 to C.H. and W.-C.K. Institute of Biotechnology Czech Academy of Sciences grant no. RVO 86652036 to J. Černý and B.S. Max Planck Society and German Research Foundation grant no. RTG 2756 to H.G. and M.I. Medical Research Council grant no. MR/V000403/1 to C.M.P., T.B., A.P.J. and M.D.W. Medical Research Council grant no. MC_UP_A025_1012 to G.N.M., K.Y. and P.E. NIH/NIGMS grant no. P01GM063210 to M.L.B. and C.H. NIH/NIGMS grant nos. P01GM063210, R01GM073919 and R35GM131883 to J.S.R., M.G.P., C.J.W., V.B.C. and D. C. Richardson. NIH/NIGMS grant nos. P01GM063210, R01GM071939 and R24GM141254. US Department of Energy grant no. DE-AC02-05CH11231 to the Phenix Industrial Consortium: N.W.M., P.V.A., C.J.S. and O.V.S. NIH/NIGMS grant no. R01GM133840, National Science Foundation grant no. IIS2211598 to D. Kihara. NIH/NIGMS grant no. R01GM146340 to J. Cheng and N.G. NIH/NIGMS grant no. R01GM123089 F.D. and A. Muenks. National Science Foundation grant no. DGE-1762114 to A. Muenks. National Science Foundation grant no. CHE-2235785 to A.P. National Science Foundation grant no. DBI-1832184, US Department of Energy grant nos. DESC0019749 and NIH/NCI, NIAID, NIGMS R01GM133198 to S.K.B., C.S. and C.L.L. NSERC of Canada, grant no. RGPIN-05795-2016 to C.N.R. SERB grant no. CRG/2022/002761 to S.M. US Department of Energy grant no. DE-AC02-06CH11357 to A.J. UWB Scholarship, Research and Creative Practice grant no. 2023-2024 to D. Si. Wellcome Trust grant no. 209407/Z/17/Z to R.J.R. Wellcome Trust, grant no. 208398/Z/17/Z to R.W. Funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

H.M.B. and W.C. conceived the project. C.L.L. and A.K. organized the Challenge with the assessors and modelers. G.D.P. and M.F.S. assisted in the analysis. J. Černý, P.E., A.J., J.S.R., R.J.R., A.L.R. and B.S. helped to develop Challenge goals and guidelines. Authors listed in Table 1 prepared and submitted models for the Challenge. Authors listed in Table 2 provided assessment results. C.L.L., A.K., G.D.P., M.S., H.M.B. and W.C. wrote the initial draft. All coauthors participated in review and revision of the paper.

Competing interests

S.N., A.G., A.R., B.S. and Y.Y. are current or former employees of Genentech. E.S. and C.I.W. are current employees of

Chemical Computing Group. The other authors declare no competing interests.

Additional information

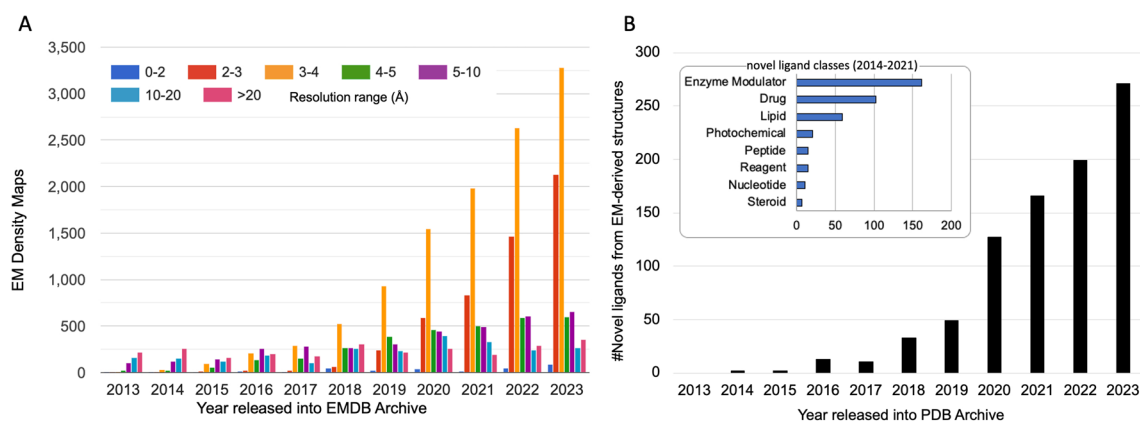
Extended data is available for this paper at <https://doi.org/10.1038/s41592-024-02321-7>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-024-02321-7>.

Correspondence and requests for materials should be addressed to Catherine L. Lawson or Wah Chiu.

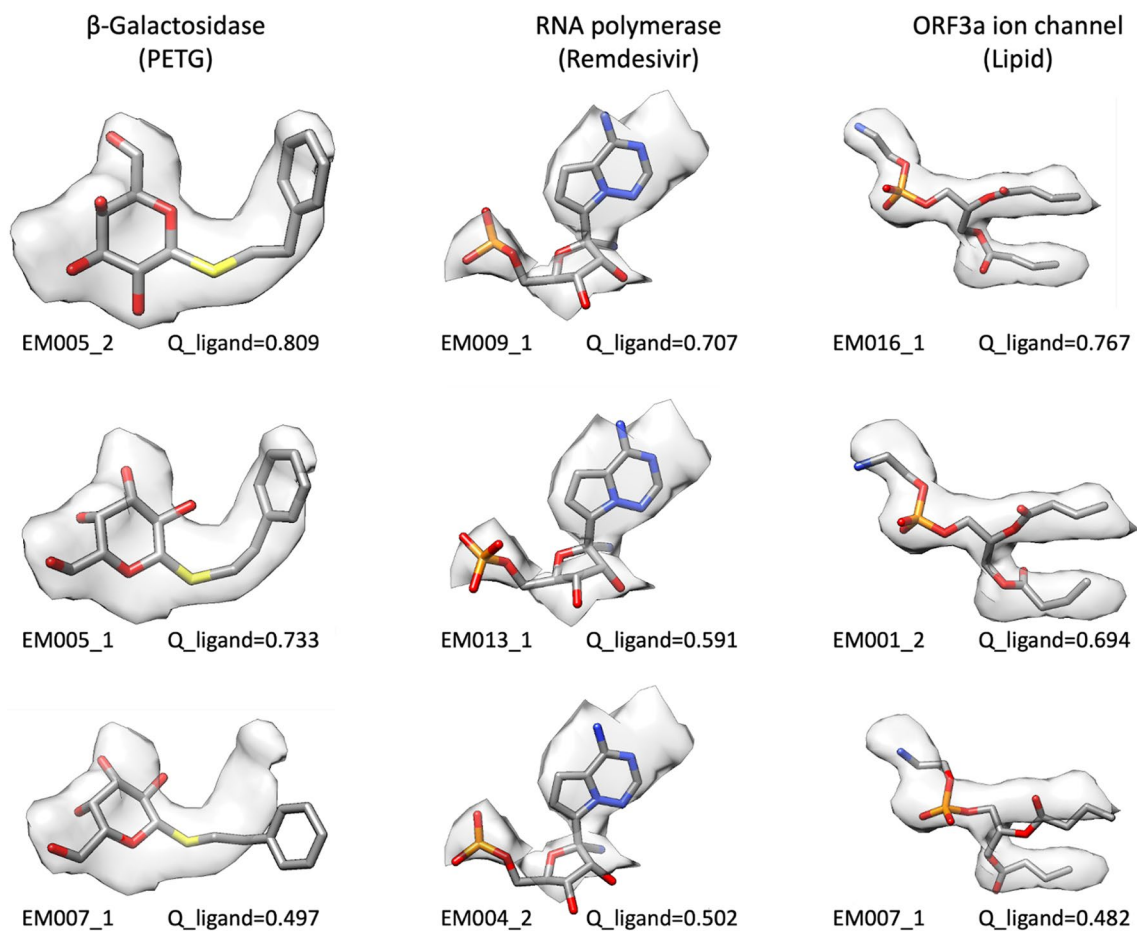
Peer review information *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Arunima Singh, in collaboration with the *Nature Methods* team.

Reprints and permissions information is available at www.nature.com/reprints.



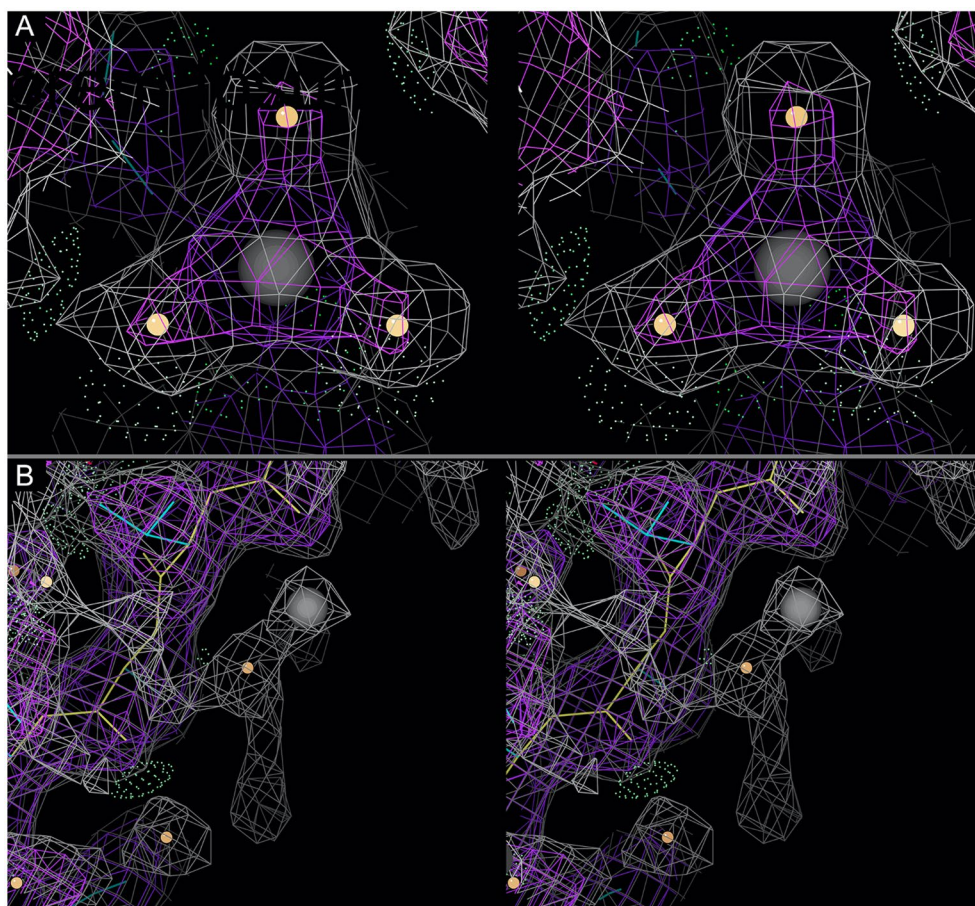
Extended Data Fig. 1 | Growth of cryo-EM structures and novel ligands derived from them. (a) Cryo-EM maps released into the EM Data Bank (EMDB) archive by year and resolution range (source: www.emdataresource.org) up to the end of 2023. **(b)** Novel non-polymer ligands included in cryo-EM structures

by year of release into the Protein Data Bank (PDB) through 2023. Inset: major categories of novel ligands found in cryo-EM-derived models (through 2021). See **Online Methods** for details.



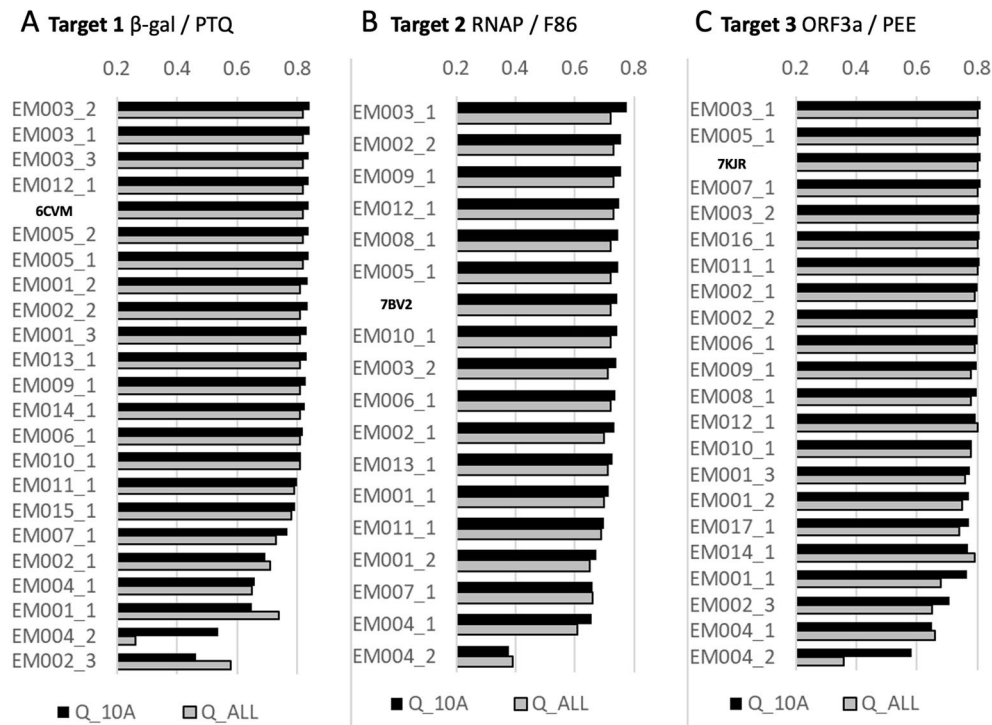
Extended Data Fig. 2 | Selected submitted ligand models for each of the Challenge targets. Panels are labeled by team ID and model # (see Table 1), in order of decreasing ligand Q-scores (see Fig. 3, row 1) from top to bottom. The portion of the map corresponding to the ligand is shown as a semi-transparent surface, along with the model of the ligand. Ligand Q-score is the average Q-score

of all non-H atoms in the ligand. For each atom, the Q-score is measured by correlation of map density to the expected gaussian function, at points within 2 Å of the atom and closer to the atom than any other non-H atom in the model¹⁰. Higher-scoring ligand models fit better in the cryo-EM density than lower-scoring models.

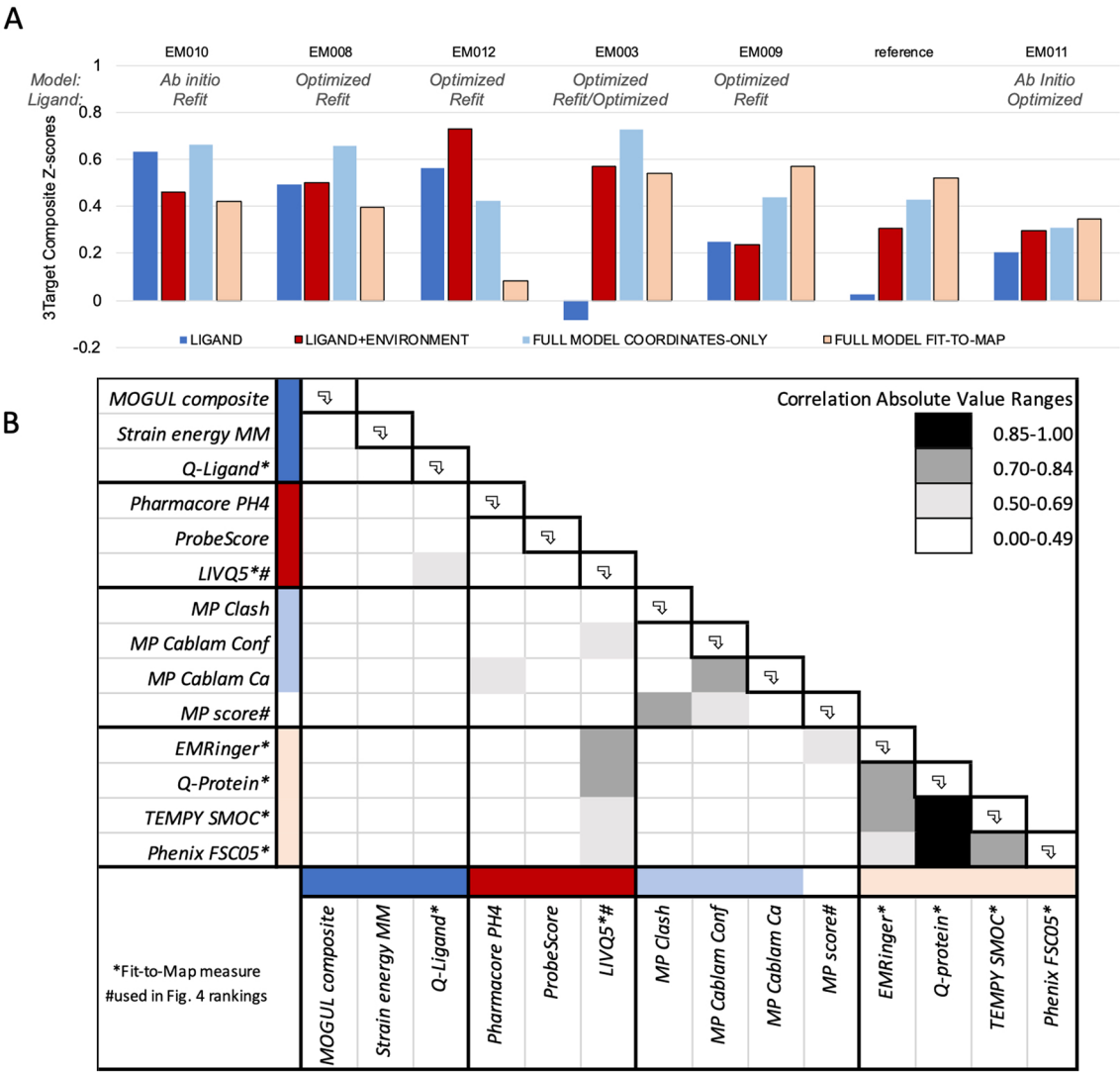


Extended Data Fig. 3 | Evaluation of ions in submitted models (stereo images). (a) Target 1 6cvm reference model Mg A2002 (gray sphere) with water ligands (orange spheres), located near the PETG ligand, with density for classic octahedral coordination. Only six of 23 submitted Target 1 models included the Mg^{2+} and all three coordinating waters. Others had either only Mg^{2+} , Mg^{2+} plus

one or two waters, Mg^{2+} plus waters with zero occupancy, no atoms modeled, or atoms significantly displaced. (b) Some groups placed metal ions with weak justification, as exemplified by the Na^+ (grey sphere) shown here in model EM005_1 for Target 3.

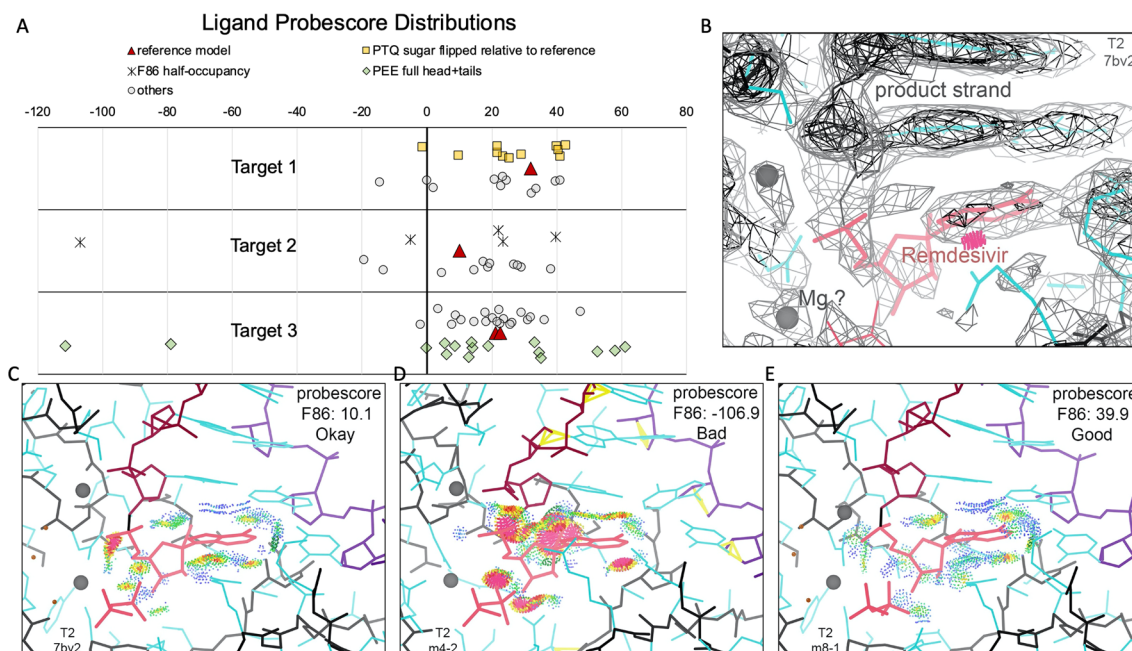


Extended Data Fig. 4 | Q-score rankings for ligand + extended vicinity and for full models. (a-c) LIVQ10 (Ligand + extended vicinity ≤ 10 Å) Q-scores (black bars) and full model Q-scores (gray bars) are plotted for each submitted model and each reference model, with order according to ligand + extended vicinity rank.



Extended Data Fig. 5 | Alternative Group Ranking by sum of Ligand, Ligand+Environment, Full Model Coordinates-only, Full Model Fit-to-Map composite scores. (a) Group ranking (left-to-right) according to the sum of four composite z-scores, as described below. Only groups that submitted models for all 3 targets and have rank similar to or better than PDB reference models are shown. **(b)** Correlation table (n = 61) of scores used to create z-scores and rankings in panel (A) and/or Fig. 4. Group composite scores were calculated per team as follows. For each submitted model, and for each score type, a composite

z-score was calculated. For each target (T1, T2, T3), the model submitted by that group with maximum composite z-score was selected for inclusion in the final average score over all targets. Ligand: $z = (0.33 \times z_{\text{MogulComposite}} + 0.33 \times z_{\text{StrainEnergyMM}} + 0.33 \times z_{\text{Q-ligand}})$. Ligand+environment: $z = (0.33 \times z_{\text{Pharmacore}} + 0.33 \times z_{\text{ProbeScore}} + 0.33 \times z_{\text{LIVQ5}})$. Full model coordinates-only: $z = (0.25 \times z_{\text{Clash}} + 0.5 \times z_{\text{CablamConf}} + 0.25 \times z_{\text{CablamCa}})$. Full model fit-to-map: $z = (0.25 \times z_{\text{EMRinger}} + 0.25 \times z_{\text{Q-Protein}} + 0.25 \times z_{\text{TEMPySMOC}} + 0.25 \times z_{\text{PhenixFCS05}})$.



Extended Data Fig. 6 | Ligand/Ligand Environment Probescores.

(a) Molprobit Probescore³² distributions for ligands in Targets 1–3 (reference models: red triangles; submitted model scores are plotted as gray circles with following exceptions: Target 1, yellow boxes if PTQ sugar ring position was flipped relative to reference; Target 2, asterisk if F86 was set to half-occupancy; Target 3, blue diamonds if PEE was modeled as head-group+tails). Scores are plotted in horizontal axis lanes with small random vertical shifts to visually separate clustered points. Notably, score distributions have wide spreads independent of noted model features: PTQ sugar orientation, F86 occupancy, or PEE inclusion of tails—although for PEE the score distribution is noticeably broader when the

larger and more variable tails are included. (b) T2 density map with reference model in the region of the F86 ligand, showing half-strength density for the remdesivir ligand, implying that only half the molecules have covalently bound inhibitor. Image is reproduced from Figure 6 of reference³⁸ (open access CC-BY license, no permission required for reuse). (c–e) T2 F86 + pyrophosphate ligand environments for the reference model (PDBid 7BV2), model EM004_2, and model EM008_1, respectively. All-atom contact dots are from Probescore, with all-atom clashes in hot pink and favorable H-bonds and vdW contacts in green and blue. Molecular graphics are shown in KiNG³⁸.

Extended Data Table 1 | Ligand and Ligand+environment Q-scores for submitted models with highest ligand Q-scores

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

Expected_Q is the expected Q-score for well-fitted models in maps at similar resolutions, based on analysis of a subset of publicly archived maps and models⁶⁹. Q-scores well below the expected value indicate either that the map is not as well resolved as other maps at similar resolution, for example due to heterogeneity, or that the model is not optimally fitted to the map.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

The following software was used by Ligand Challenge Modelling Teams:

Ligand restraint generation:

AMBER/Antechamber v.20 ambermd.org/antechamber
 BUSTER Grade v.1.2.19 grade.globalphasing.org
 CCP4/AceDRG v.2.23 www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/acedrg/acedrg.html
 CCP4 monomer library v.5.52 www.ccp4.ac.uk/html/mon_lib.html
 CHARMM/CGenFF v.2.4.0 cgenff.silcsbio.com
 LigPrep v21-1 newsite.schrodinger.com/platform/products/ligprep
 OpenBabel v3.1.0 openbabel.org
 Phenix/eLBOW v1.14, v1.19.2 phenix-online.org
 PyRosetta v.4 www.pyrosetta.org

Ab initio modeling:

ARP/wARP v.8.1 arpwarp.embl-hamburg.de
 DeepTracer v1 deepttracer.uw.edu
 Mainmast v.1.04 kiharalab.org/emsuites/mainmast.php
 Mainmastseg v.1 kiharalab.org/emsuites/mainmastseg.php
 Modeller v. 10.2 salilab.org/modeller
 NAMD v.2.14 www.ks.uiuc.edu/Research/namd
 Phenix/Pathwalker v.1.19.1 phenix-online.org

Rosetta v.4 [rosettacommons.org](https://www.rosettacommons.org)

Model optimization:

Amber v.20 ambermd.org

CDMD vGromacs-2018-densfit www.mpinat.mpg.de/grubmueller/densityfitting

CryoFold v2.0 github.com/SingharoyLab/CryoFold_GUI

doubleHelix v.5.0.1 gitlab.com/gchojnowski/doublehelix

MDFF v0.4 www.ks.uiuc.edu/Research/vmd/plugins/mdff

MELD v.0.4.20 github.com/macallumlab/meld

Phenix v.1.18.2 – 1.19.2 phenix-online.org

ProSMART v.0.859 www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/prosmart/prosmart.html

REFMAC v5.8.0272 www.ccpem.ac.uk, www.ccp4.ac.uk

Schrödinger v.21-1 newsite.schrodinger.com

Servalcat development version github.com/keitaroyam/servalcat

Visual evaluation/manual model improvement:

Chimera v.1.1.5 www.cgl.ucsf.edu/chimera

ChimeraX v.1.1 www.cgl.ucsf.edu/chimerax

Coot v.0.8.9, v.0.9.5 www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot

EMDA v.1.1.3 www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/emda/emda.html

ISOLDE v.1.1.0 tristanic.github.io/isolde

KiNG v.2.24 www.biochem.duke.edu/people/richardson-lab

PyMOL v.2.4.0 pymol.org

VMD v.1.9.4 www.ks.uiuc.edu/Research/vmd

Manipulation of map densities:

CCP-EM v.20210312 nightly release www.ccpem.ac.uk/download.php

EMDA v1.1.3 www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/emda/emda.html

LAFTER v.1.1 github.com/StructuralBiology-ICLMedicine/lafter

The following software was used by EMDDataResource to collect Challenge Models/Data:

Model metadata collection:

Drupal v.7.88 webform v.7.x-4.24 drupal.org

Model coordinates collection:

PDB-extract v.4.0 pdb-extract.wwpdb.org

Model coordinates processing:

MAXIT v.11.1 swtools.rcsb.org/apps/MAXIT

Data analysis

The following software was used in the Ligand Challenge Pipeline or by individual Assessment Teams:

Ligand:

UCSF Chimera/MapQ plugin v1.2 github.com/gregdp/mapq (Q-score_Ligand; Q-score_HOH)
 wwPDB report Model statistics 20191225.v01 and Mogul v1.8.5 [validate-rcsb-4.wwpdb.org](https://www.rcsb.org/validate-rcsb-4)
 OpenEye/SZYBK1 v2.7.0 eyesopen.com
 OpenEye/Omega v5.0.0 eyesopen.com
 ANI neural net potential, Genentech in-house version github.com/Genentech/g_ani

Ligand Environment:

UCSF Chimera/MapQ plugin v1.2 github.com/gregdp/mapq (LIVQ5; LIVQ10)
 MOE QuickPrep/db_AutoPH4 MOE v.2020.09 www.chemcomp.com/Products.htm (Pharmacophore)
 Probescore v2.18 <https://github.com/r1abduke/probe>

Fit-to-Map:

EMDB CryoEM Validation Analysis v.0.0.0.dev8 pypi.org/project/va/0.0.0.dev8 (AI_all, AI_BB)
 TEMPpy v.2 tempy.ismb.lon.ac.uk (CCC,SMOC)
 Phenix v.1.19.2-4158 phenix-online.org (EMRinger, boxCC, CCpeaks, CCmask, CCvol, FSC05)
 UCSF Chimera/MapQ plugin v1.2 github.com/gregdp/mapq (Q-score_Protein)

Coordinates-only:

DNATCO v4.1 dnatco.datmos.org (conformer_score, conformer_percentile, RMSD)
 KING v2.23 kinemage.biochem.duke.edu/software (issue visualization)
 Phenix v1.19.2-4158 phenix-online.org (Molprobability: MPscore, Clashscore, Rotamer_out, Rama_out, Rama_favor, CABLAM_conf_out, CABLAM_Calpha_out, RNAsuite-out, RibosePucker-out, UnDowser_HOHclash)

Comparison-to-Reference:

CAD v.1646 bitbucket.org/kliment/voronota/src/master (CAD)
 HBPLUS v.3.06 www.ebi.ac.uk/thornton-srv/software/HBPLUS (HBjaccard, HBPrecision)
 LGA v.04.2019 proteinmodel.org/AS2TS/LGA/lga.html (GDT_TS, GDT_HA, GDC_ALL, GDC_SC, RMSD, DAVIS_QA)
 MMalin v.20210816 zhanggroup.org/MM-align (TM_score)
 OpenStructure v.2.1 www.openstructure.org (LDDT, QS_score)
 Phenix v.1.19.2-4158 phenix-online.org (Nclose, Nfar, CAscore, SeqMatch)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Cryo-EM Map Targets were the primary maps of EMD entries EMD-7770, EMD-30210, and EMD-22898 www.ebi.ac.uk/emdb, emdataresource.org

Reference models were PDB entries 6cvm v1.3 (Target 1), 7bv2 v3.4 (Target 2), and 7kjr v1.1 (Target 3) [wwpdb.org](https://www.pdb.org)

Submitted models, model metadata, result logs and compiled data are available via challenges.emdataresource.org/?q=2021-model-challenge and archived at Zenodo: doi.org/10.5281/zenodo.10551958

Interactive summary tables, graphical views and spreadsheet downloads of compiled results are available at model-compare.emdataresource.org/2021/cgi-bin.

Supplementary source data tables are provided with this paper.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

n/a

Population characteristics

n/a

Recruitment

n/a

Ethics oversight

n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size (n=61) consisted of all of models submitted by modeler co-authors.
Data exclusions	no data were excluded from analysis
Replication	multiple models of each Challenge target were submitted by different teams
Randomization	randomization not applicable to study owing to Challenge design
Blinding	analyses of challenge targets were blinded with respect to team identity

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging