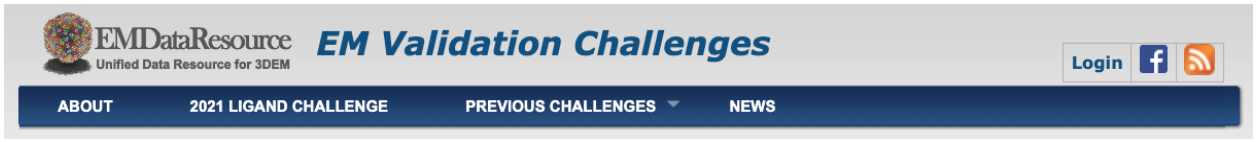
EMDataResource 2021 Model Ligand Challenge Website Archive



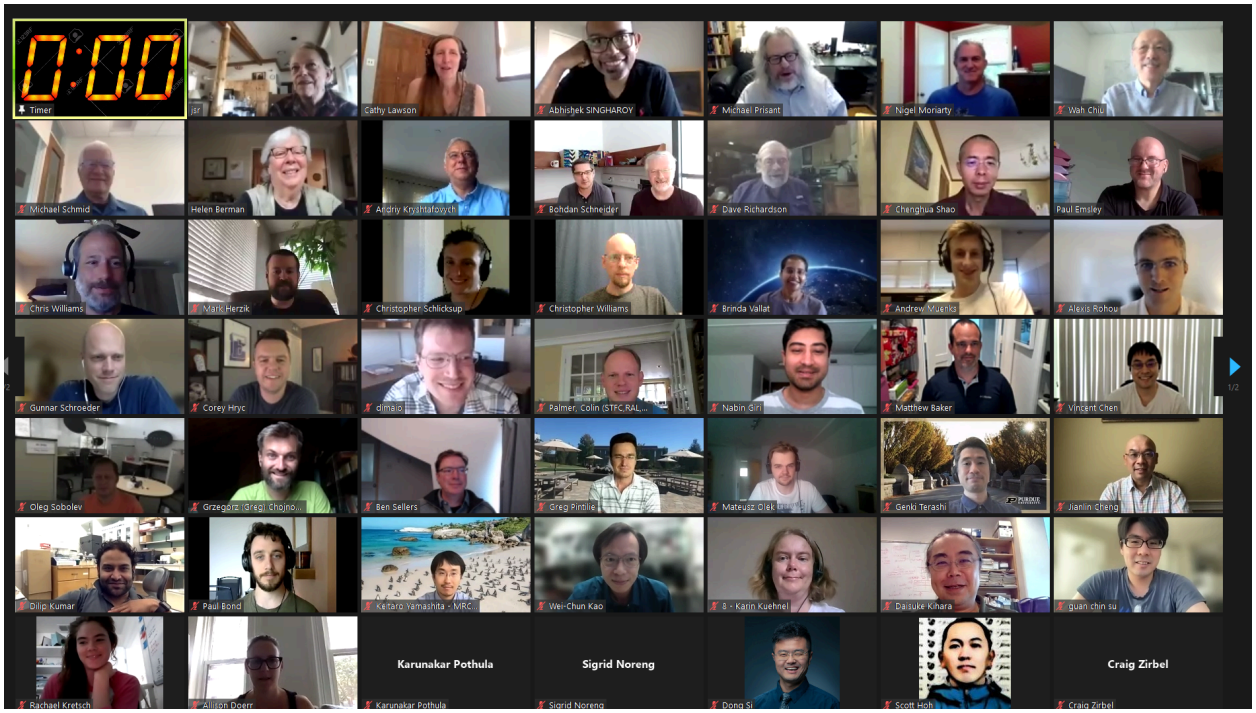


Image from July 26-28 2021 Wrap-up Meeting

News

Announcing the 2021 Ligand Model Challenge

February 1, 2021: Our next model challenge launches today: the 2021 EMDataResource Ligand Model Challenge. The task for modeler teams in this round is to optimize reference models against selected target maps containing protein or protein+RNA, ligands, ions and solvent (1.9-2.5 Å resolution range).

The three targets are β-Galactosidase (1.9 Å), SARS-CoV-2 Polymerase/remdesivir (2.5 Å), and SARS-CoV-2 ORF3a putative ion channel in nano disc (2.1 Å). Evaluation/model comparison in this round will strongly focus on the ligands and their local environment, and will include additional ligand-specific assessments.

The deadline for submission of completed models is **April 1 ~~March 15~~ (3 PM US Eastern)**. All interested in participating should fill out the **registration form**.

In preparation for the evaluation phase of the Challenge, we welcome recommendations from the structural biology community regarding ligand-specific evaluation tools. Please send suggestions to **challenges@emdataresource.org**.

Overview

A major goal of the **EMDataResource team** is to work with the 3DEM community to establish data validation methods that can be used in the structure determination process, define key indicators of a well-determined structure that should accompany every structure deposition, and implement appropriate validation procedures into a 3DEM validation pipeline. Following recommendations of the EM Validation Task Force (**Henderson et al 2012**), we are regularly hosting new benchmark challenges, with the aim to stimulate further community discussions about validation procedures for 3DEM maps and map-derived models. Prior community-organized 3DEM challenge activities have included a Particle Picking Challenge (**Zhu et al 2004**), CryoEM Modeling Challenge (**Ludtke et al 2012**), CTF Challenge (**Marabini et al 2015**), Map and Model Cryo-EM Challenges (**Lawson & Chiu 2018**), and most recently the **2019 Model Metrics Challenge** (**Lawson et al in press**).

Goals for Challenge

This Challenge will focus on ligands and water with the goal of identifying metrics most suitable for evaluating and comparing fit of atomic coordinate models into cryo-EM maps for specimens in the <3.0 Å reported overall resolution range. The specific focus areas are:

1. Small molecules including ligands, water, metal ions, detergent, lipid, nanodiscs -- geometry and fit to map
2. Model geometry (including Rama, rotamers, clashes, EMringer, CaBLAM) in the neighborhood surrounding the small molecules
3. Local fit of model into map density per residue and per atom
4. Resolvability at residue or atom-level
5. Atomic Displacement parameters (B-factors) recommended optimization practice

Questions: How reliable are ligands/waters/ions built into cryoEM maps? Can they be placed automatically or is manual intervention needed?

Model Evaluation

Submitted models will be scored as in prior Challenges using global and local Model-only, Fit-to-Map and Comparison-to-Reference metrics. Assessments will be blinded with repect to modeler identity and software used. New in this round: evaluation/model comparison will strongly focus on the ligands and their local environment, and will include additional ligand-specific assessments.

Timeline

- February 1 : Challenge Announced, **Model Team Registration opens**
- February 8 @ noon US Eastern: Virtual Information Session for Participants--please register to receive connection details
- February 26: **Model Submission site opens**
- **SUBMISSION DEADLINE UPDATED:** ~~March 15~~ April 1 @ 3 PM US Eastern: Submissions close
- July 14: Model Compare Site Demo (Andriy)
- July 26-28: Ligand Challenge Virtual Wrap Up Discussions for all participants

Targets

Specimen/Reference	Target Map	EMDB	Symmetry	Reference Model	Model Components	Resol (Å)
E. coli β-Galactosidase (**EMPIAR 10061** reprocessed) **doi**		**EMD-7770**	D2	**6cvm**	Protein (1 unique chain, A,B,C,D), Water, Mg, Na ions,PETG: 2-phenylethyl 1-thio-beta-D-galactopyranoside (**PTQ**)	1.9
SARS-CoV-2 RNA-dependent RNA polymerase **doi**		**EMD-30210**	P1	**7bv2**	Protein (3 unique chains A|B|C), RNA (2 unique chains D|E), Water, Zn, Mg ions, pyrophosphate, **F86 (remdesivir**, covalent inhibitor)	2.5
SARS-CoV-2 ORF3a putative ion channel in nanodisc **doi**		**EMD-22898**	C2	**7kjr**	Protein (2 unique chains A,B|C,D), Water, 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (**PEE**)	2.1

How to Participate

1. Modeling teams create and upload their optimized model for each **Target Map** (target map = the primary map of the EMDB entry).
2. Submitted models should be as complete and as accurate as possible (i.e., close to publication-ready), with atomic coordinates and atomic displacement parameters for all model components.
3. Submitted models must use the **Reference Model**'s residue, ligand, and chain numbering/labeling for all shared model components.
4. Ligands should ideally be deleted & refit independently.
5. If additional polymer residues are included in the optimized model: label them according to the Reference Model's sequence/residue numbering/chain ids.

6. If additional waters/ions/ligands are included in the optimized model: label them with unique chain ids (e.g., chain "X").
7. If predicted hydrogen atom positions are part of the modeling process, hydrogens should be included in the submitted coordinates.
8. The unique point asymmetric unit or full complex may be uploaded. In either case the model is expected to adhere to the reconstruction's point symmetry as shown above.

Q&A Summary: Monday Feb 8 Discussion

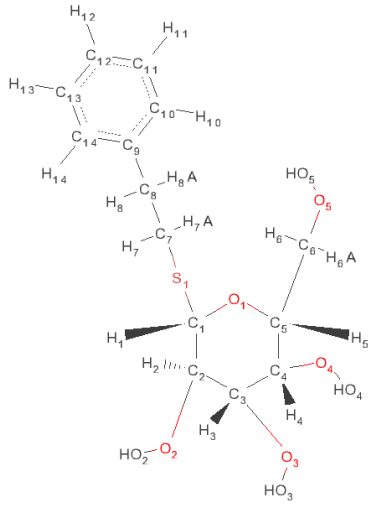
1. **Polymer Chains (protein, RNA): should modelers re-determine *ab initio* and/or re-refine polymer chain(s) in the reference structure?** The decision is up to each modeling team, and should be based on careful evaluation of the reference model. On submission you will be asked to identify whether your process involved *ab initio*+optimization, optimization, or no optimization of polymer chains.
2. **Ligands: should we trust the overall geometry of the ligand in the reference structure?** Again this should be based on your team's evaluation. If the ligand position requires independent refitting, and/or if you wish to test an automated ligand fitting procedure you can certainly do so. Categories: ligand independently refit, reference ligand optimized.
3. **Is it required to maintain the point group symmetry for models against targets 1 and 3?** Yes.
4. **Is there a "better" reference model for each target?** No. We will be comparing each submitted model to the current reference and to all other submitted models. (there are several beta-galactosidase models available in PDB. Modelers are welcome to make use of any available model for optimization, but the other models won't be included in final comparisons).
5. **How to determine the charge on the ligand when parameterizing small molecule force fields?** Charge parameterization, if used for optimization, is at the discretion of the modeling team. The submission form will request details about this.
6. **How many models can each team submit?** Preferred: one model per target; maximum: 3 per target. Each model must be uploaded as a separate submission.
7. **Could occupancy be refined?** If there is a reasonable justification, yes (e.g. part or all of the ligand, water in close proximity to a symmetry axis, or low density). Otherwise refinement should be limited to atomic coordinate positions (x,y,z) and isotropic B-factor.
8. **How can I participate in assessment of submitted models/ligands?** Two options: (1) Download submitted models when they become available in ~~late March~~ early April (submitter/software blinded) and evaluate them using your own tools. (2) suggest a validation tool to include in our analysis pipeline. In either case please send us an **email** to let us know of your interest.

Advisory Committee

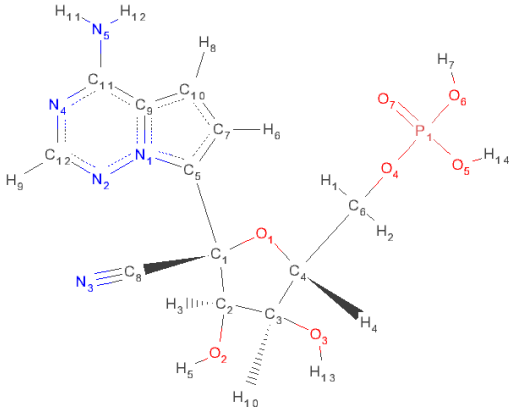
- Paul Emsley, Andrej Joachimiak, Randy Read, Jane Richardson, Alexis Rohou, Bohdan Schneider, Jiri Cerny, EMDR team.

Ligand Images

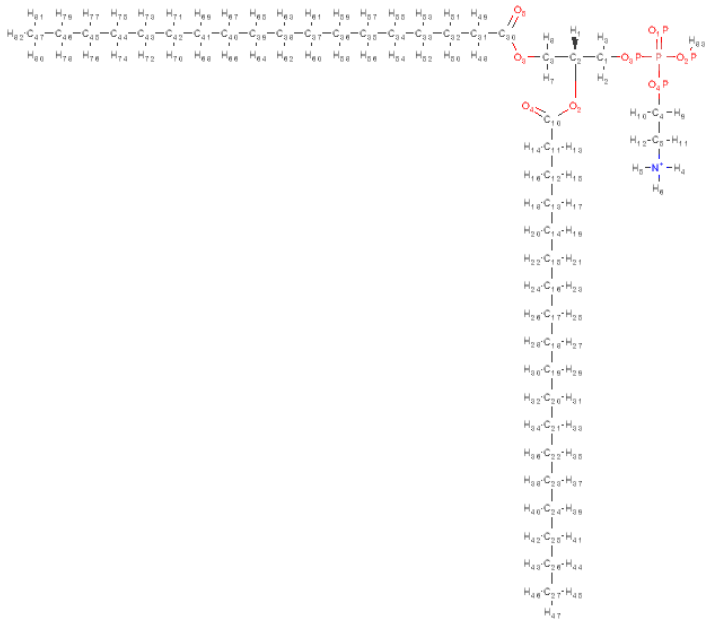
PTQ: 2-phenylethyl 1-thio-beta-D-galactopyranoside



F86: remdesivir, covalent inhibitor



PEE: 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine



EMDataResource Validation Challenges are supported by NIH **National Institute of General Medical Sciences**