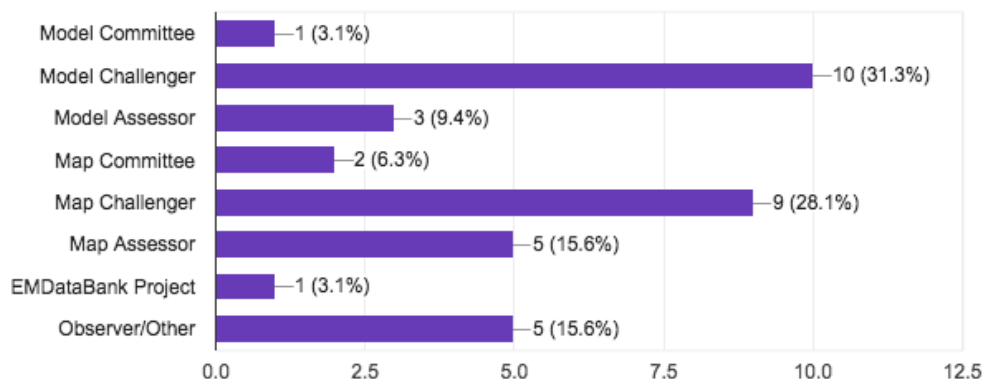


## Oct 2017 Joint Challenges Workshop -- Participants Perspectives: Survey Results

### Role

32 responses



### Map Challenge: Your Synopsis

27 responses

Fantastic

Open the site for 'revisions' of earlier maps, whether they are evaluated or not. I believe just the process of participating in the challenge taught us important things about refinement, and caused us to improve our software.

FSC using independent half maps is an excellent measure of map quality Metrics for agreement between map and raw images are needed

We need another metric besides FSC.

1. I want to propose an appropriate mask generation protocol. 2. I would like to distinguish the different operations currently grouped together under map filtering.

Conclude this challenge with a group paper asap; plan for the next challenge.

Model Challenge #1: Model building is not just local map fit, but needs to use more outside information such as sequence, secondary structure, prior conformational probabilities, topology rules, related folds, and idiosyncrasies of the specific molecule.

It was great having the map challenge and I would very much prefer another round with more diverse quality data (resolution ranges, symmetries etc.)

Good Particles yield good maps. Bad Particles yield bad maps. More Good Particles yield better maps than fewer Good Particles. More Good + Bad Particles probably does not help. But how can you tell what a Good Particle is?

It seems we haven't got an agreed way of prepare map for deposit. One particular issue is filtered or unfiltered?

Time commitment still a major bottleneck for participants; reduce number of datasets? Shame that metadata (defocus, Euler angles, class memberships) were not part of the analysis. Need shorter time window. Most metrics depend on model built in, so we need models to evaluate maps.

N/A

The challenge was educational and offered unique development opportunities; however, the process required too much time and resources, and efforts were split among too many maps. Future challenges to meaningfully resolve heterogeneity would be very useful.

Very good and productive meeting. Maybe it's interesting to also analyze the refined particle parameters for the map challenge submissions.

Great opportunity for scientists outside the community to understand/experiment how to implement a specific method/scheme into the global pipeline.

Need to make sense of all the different assessments. This is a task for the Map Committee as a whole.

This first version of this report is to be shared with Challengers for feedback and be a joint document

Need to re.open map submissions, but only for very limited cases.

Awesome! I would desire to get more information or thoughts on 1. how people design data collection (in the context of target resolution), 2. parameters to select images later on (considering particle distribution may not be clean).

As a model challenger I appreciate the use of "model ability" as a metric to evaluate maps. That seems like it could be one of the key ways to evaluate maps going further. It may be interesting to also test "segmentability" as that is one of the major challenges for modeling.

\*With minimal manual intervention and decision making using default Appion, I could generate a map that was almost as good as the best map. \*For summary paper, given the time elapsed, it would be unfair to publish map rankings. \*Challenge sparked new tools amongst developers

I think we should have more automated and orthogonal map validation tools

On the assessment side, I was struck most by the use of model-guided statistics as supplements to the increasingly sophisticated suite of FSC metrics. For the map challenge itself, I was surprised at how similarly successful different software was for expert users.

Great to have so much data, varying quality, varying resolutions; useful for testing new methods.

Most of the software, when used by experts, seemed to give very similar results and are capable of producing near-atomic resolution results. I am not sure if the results are user dependent.

I think this was extremely useful for map challengers. As for map assessment, the tools and ideas developed are helpful. I think the best way to assess maps in the future will be to give them to the modelers.

A good snapshot of current status of single particle image processing/3D reconstruction in terms of capability of software packages and the variability of user experience/expertise. Most software has the capability but different user experiences can result in variable results.

This Map Challenge was the first one and has been very useful in figuring out some initial questions about software etc. It will be more helpful for future Map Challenges to have better defined targets, preferably blind.

wasn't able to attend this remotely

## Model Challenge: Your Synopsis

24 responses

Great

Much more mature than maps, but more de-novo should be encouraged in future.

Rapid progress is being made in automated map interpretation Combination of methods from different groups could accelerate progress

There are many metrics. We need to tabulate them and evaluate the usefulness of each one

Future model challenges need to separately treat ab initio modeling versus refining from a deposited or homologous structure.

Models sometimes can be derived automatically. Sometimes not. Sometimes you just have to suck it up and do it by hand. We need to do refinement like the cool kids (the Crystallographers).

The model assessment should consist of three components: statistics based knowledge such as Ramachandran/EMRinger, agreement with the map such as cross-correlation, and local/residue-level variation such as Z-score, B-factor or RMSF per residue.

It would be great if unreleased maps are available for the challenge. Evaluation metric that considers uncertainty of reference structures is needed.

The "success" rate is very high. Phenix made a very strong showing, but what about the UK crew? I suggest looking for a way to "blind" the entrants as to the correct answer (perhaps don't give them the best possible map). Very impressive web server for validation - I want it!

I felt that the definition of "good ab initio model" is not clear. Some absolute metrics, such as  $rmsd < 5.0$ , are useful to see the absolute quality of models.

Models can be generated rapidly, and there are lots of validation procedures, though each has some bias. Our ability to model a structure depends heavily on map quality, so we need better maps to produce more accurate models.

Backbone tracing (MAINMAST, PATHWALKER) with real space refinement (Phenix, MDFF, Rosetta) in combination with on-the-fly map smoothing/re-sharpening provides a powerful and reproducible mean for model determination.

Very interesting to have joint meetings of the two Challenges, at least the final evaluation meeting, like it has happened here. The two parts need to know each other quite well, but not necessarily be a master at the two. So, still different Challenges, but some joint meetings

It is amazing to see progress (and utilization of multidimensional approaches) on model building. I would look forward to have a weighting scheme based validation of model that would include both correlation (local and global) with the map and stereochemistry.

I agree that de novo modeling and model refinement should be treated separately as the metrics used to evaluate each are highly different. I think a challenge like this is good for probing the overall state of the field but future challenges should have more defined targets.

\*Excited about new modeling tools that were presented \*This is an area that is ripe for more new developments and I'm excited to see results of new challenges \*Next challenge idea: Can you use modeling to identify maps with issues like noise bias or misclassified particles?

Best practices for interpreting 3-4 Å resolution cryoEM density maps have been validated. Modeling such maps would involve manual adjustment using well-established modeling tools such as Coot and refinement with Phenix in an iterative manner. Phenix promises automation of this.

We need to figure out a way to go back from model to image to improve the data processing pipeline as suggested without over-fitting

For Assessment, a preliminary standard combining a correlation metric, whether CC or FSC, EMRinger and other model-directed methods, and model geometry information from Molprobity is a good starting point. We need ways to refine atomic B factors.

Challengers seemed to approach the challenge in a variety of ways, i.e. de novo modeling, model refinement, software evaluation. A number of evaluation metrics were presented, most of which are very useful but not necessarily universally appropriate for all models/approaches.

This challenge was very helpful. I think in the future there should be smaller, more specific challenges. Exciting progresses in de novo, automated modeling capability.

The Model Challenge has also been quite useful in figuring out where we stand currently in cryo-EM modeling. It will be more useful for future Model Challenge to have lower resolution targets.

I'm a bit unsure about how globally to go about assessing vs. using our specific tool. There are a few areas that I didn't see addressed as much as I would have expected: 1) Bfactor modeling and the role of local resolution 2) the use of crystallography-like R-factors

## Any other Comments you'd like to share (no char limit)

14 responses

I'm afraid I'm not entirely convinced by arguments for "collaborative" work rather than competitive work. If the goal is to learn more about a specific target, this is good. If the goal is to improve methods, eventually some competition is important.

Thank you and the other organizers and participants.

Even though I only just started to work with EM, I have tried almost all the modeling tools which was presented today. In the core, the modeling strategy of different software is not that different: Start with something confident then slowly build upon it. The confident thing can be a C-alpha trace, a secondary structure element, a stretch of sequence clearly identified or a good homology model. Most software right now uses a single confident thing. My gut feeling is at the resolution around 3.5 Angstrom, it probably requires combining all the informations to build the best ab initial model possible.

Ideas for future challenges: keep map computation and modelling joined at the hip for next few years. The cross-pollination of ideas is very valuable. Ideas for future map challenges: (1) have a special "server" category (a la CASP), i.e. where only default settings are allowed - software not expected to do as well, but could be tested on more datasets; (2) give metadata (particle coordinates, CTF, Euler, class occupancies) to assessors; (3) encourage development and refinement of existing dictionary to describe experimental & processing metadata; (4) I like the idea of the hackathon with fellow developers. Ideas for future modeling challenges: (1) also do a server category separate from manual /semiautomatic submissions (2) work towards ranking a la CASP so as to hasten the spreading of new best practices - as a novice to modelling, it'd be great to know what is this year's "best"

15min of Presentation time is to enough. To discuss more detail, poster sessions might be good for us.

My take away from this challenge is that map quality will improve with our ability to address specimen heterogeneity. Since most single particle analysis software leads to the same outcome given a consistent set of particle data, it is now important that we identify what makes particles good/bad, how to optimally subdivide input data to separate discrete heterogeneities without bias from human intervention, and how to properly interpret continuous heterogeneities that cannot easily be addressed by classification.

RMSF-type criteria provides information on model uncertainty and ones determined from MD can distinguish it from map uncertainty.

MAP CHALLENGE: Need to make sense of all the different assessments. Is there some level of consensus in all these reports, even explicitly considering the limitations of each approach?. This is a task for the Map Committee as a whole, even if internally dividing certain roles. This first version of this report is to be shared with Challengers for feedback. Hopefully, a final consensus analysis document could be drafted among Challengers/Assessors/Comm . This would be a "joint paper". Additionally, I think that a number of different assessment reports are deep enough to be additional papers. Need to re.open map submissions, but only for very limited cases. The two softwares (EMAN and XMIPP/HighRes) that have made the case that there are substantial differences between then and now are obvious candidates, but maybe not many more Another very nice document that could result from the Challenge could be the following: "New ideas for validation", that essentially amounts to introducing pieces of the different assessment analysis as part of the reconstruction process in the software packages that were willing to collaborate WAY AHEAD More Challenges are needed, but each with less targets and trying to achieve more submissions. They should also be faster, but realistically faster. The Assessment phase will probably be also faster, because the assessor groups will make better their algorithms each time, but they will not be radically new. Probably a joint paper is the most that should be written on these new Challenges To me the next Challenge should be on low resolution maps (around 6A, at least in pieces of the map). There are many excellent biological systems in this resolution range, and they probably reflect a continuous flexibility issue that it is not well handled (and probably it will not be well handled by the time of the Challenge). Still, the biological and technical current case is strong and compelling DATA STANDARDIZATION: Now we have the need!. We want to have compilations of results, and they must

be under some standards (there are already existing elements in the field, there is no need to re-discover Mare Nostrum/Mediterranean Sea). We should aim to data harvesting, with the different software creating specific files (currently json files, in some cases) with workflow parameters, to be all deposited at the time of map submission

Wanted to add on, 1. Data processing: Can we keep aside a portion of the dataset (5-10%) and develop tools/measures for independent validation in a way that does not require generation of a 3D model. 2.

Workshop: Next challenges in a hierarchical organization to complete specific goals one at a time but it all connects to achieve a larger goal.

One of the outstanding questions in 3D refinement is how many structures there are in solution, and how to determine that N. I think one target for the next map challenge should be a dataset with some undefined number of states.

Deposition of multiple models from the same map is desirable.

I am really excited about the future of automated model building in EM, and I think that a server to allow such automated techniques to automatically validate their results would be a great utility for benchmarking. The tools Andriy built, if they could be turned into a server, would serve this purpose very well.

1. it will be great if the user practice in masking has been assessed. Are the masks too tight or too loose? A noise-substituted, masked map FSC will be useful to assess tightness of the mask. This assessment will help formulate recommendation for paper submission/journal requirement to help reviewers judge if the claimed resolution is fair. An automated approach to search for optimal mask (such as that used by cryoSPARC) might be useful for the community. 2. it will be great if the user practice in sharpening has been assessed. Are most submitted map "optimally" sharpened? How have the different sharpening levels affected the assessments based on atomic models? Tom's phenix.atuosharpen method seems a good candidate as a reference tool for sharpening. 3. The global approach by Jose Carazo that also includes p-value significance test is an excellent assessment. Masking and the scoring function (FSC resolution range or weight) can be further explored. 4. Using model spread to judge map quality should be performed very cautiously. How does masking/segmentation and sharpening affect the spread? Has a positive control using the same map but at different sharpening levels been done to confirm invariant model spread? Has de novo modeling reached a robustness/quality level to be used as a "ruler"? A "ruler" needs to have much higher accuracy than that of the subjects.

eager to be involved in the future!