EMDataBank 2015/2016 Map Challenge Website Archive

This document collates the information provided at the website challenges.emdatabank.org for the 2015/2016 Map Challenge. Some information/links may be outdated.

Table of Contents:

News Archive

EMDataBank Challenges

published Tue, 07/14/2015

EMDataBank is hosting community-wide challenges to critically evaluate 3DEM methods that are coming into use, with the ultimate goal of developing recommendations for validation criteria associated with every 3DEM map deposited to the EM Data Bank (EMDB) and map-derived model deposited to Protein Data Bank (PDB).

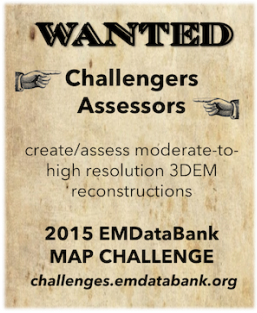
Committees comprised of respected 3DEM community members are charged to formulate the details for each challenge, including:

- choosing challenge reference data
- deciding what information participants will need to submit
- deciding on criteria for validation and comparison of results
- deciding on the timeline for challenge events
- promoting worldwide participation
- emphasizing the challenge as a collaborative and constructive activity
- evaluating the results and producing a report

In 2015/2016 we are hosting two challenges that focus, respectively, on reconstruction and modelling at moderate to high resolution, with the goals of establishing benchmarks, comparing current practices, and evolving criteria for evaluation of results. Click here to view a mini-poster about the challenges that we have presented at recent meetings. In the future we plan to host additional challenges for reconstruction and interpretation at lower resolution.

Announcing the 2015 EMDataBank Map Challenge

published Wed, 07/22/2015



EMDataBank/Unified Data Resource for 3DEM is pleased to announce the **2015 Map Challenge**.

All members of the Scientific Community--at all levels of experience--are invited to participate as **Challengers**, and/or as **Assessors**.

Seven benchmark raw image datasets have been selected for the challenge. Six are selected from recently described single particle structure determinations with image data collected as multi-frame movies; one is based on simulated (*in silico*) images.

Challengers are sought to create single particle reconstructions from the targets, and then to upload their results with associated details.

Assessors are sought to participate in evaluating submitted reconstructions.

Registration is now open for all interested participants. Challengers may submit maps between August and December. Before submissions open, all are encouraged to provide

feedback on submission requirements. An open assessment period will commence in early 2016.

To learn more about this challenge and to register, please visit http://challenges.emdatabank.org and click on "MAP CHALLENGE" in the menu bar.

The map challenge is the first of two community-wide challenges being sponsored by EMDataBank this year to critically evaluate 3DEM methods that are coming into use, with the ultimate goal of developing validation criteria associated with every 3DEM map and map-derived model. The second challenge, focused on creating coordinate models from 3DEM maps, will be announced later this year.

SDSC Offers Supercomputer Resources; Map Challenge Submission Deadline Extended

published Mon, 11/09/2015

Nov 9, 2015: We have two major announcements regarding the ongoing EMDataBank Map Challenge.

First, the San Diego Supercomputing Center (SDSC) is generously offering supercomputing resources to support the Map Challenge. The resources, made available through an SDSC Director's Discretionary Award, include 1 Million core-hours on SDSC Gordon and 20TB of sandbox data storage on Gordon's parallel file system Data Oasis.

Second, we are extending the deadline for submission of completed reconstructions from benchmark data from December 31, 2015 to **March 31, 2016**. Thus far more than 30 scientists have registered to participate in the challenge. We anticipate that availability of this supercomputer resource, as well as the additional time, will better enable these scientists to complete their calculations and may also encourage additional scientists to participate.

The Map Challenge is one of two scientific community-wide challenges being sponsored by EMDataBank to critically evaluate cryo-electron microscopy methods that are coming into use. In the current challenge phase, participants are tasked to create 3D reconstructions (maps) of several different macromolecular complexes from benchmark raw images, and to then submit their results with a full description of their method. In the subsequent assessment phase, uploaded results will be compared/contrasted by various methods in a positive spirit.

Registered map challenge participants who wish to use SDSC Gordon may apply here. All applications must be received by November 20, 2015. Allocations will be made shortly thereafter.

SDSC Gordon Info

published Sun, 11/22/2015

Everyone who applied for SDSC Gordon compute time for the map challenge will receive an initial allocation of 22,000 service units per proposed target. Altogether 14 users plan to perform a total of 45 reconstructions. Some useful links:

- create an xsede portal account
- read the SDSC Gordon user guide

Other things to be aware of:

Other things to be aware of:

- Our shared 20TB storage space area is /oasis/projects/nsf/ddp235. All of the benchmark data (~13TB) for the challenge has been copied from EMPIAR to the subfolder /oasis/projects/nsf/ddp235/benchmarks and should be accessible to you. Please avoid making additional copies!
- Each user can control the permissions on their subdirectories using the chmod commands. You may want to create areas to be shared with the group (e.g. for sharing compiled software) and private areas.
- Keep in mind that any publications that arise from this work should acknowledge that the research was supported by a Director's Discretionary Award on the Gordon supercomputer at the San Diego Supercomputer Center, University of California, San Diego

updates:

Dec 18: Users who have been allocated time on Gordon can also request time on the Comet system (GPU hardware).

Feb 17: Inquiry was made regarding increasing the 48 hour wall-time for jobs on Gordon, but it is difficult to break this policy. For Relion users should make use of the restart option.

Feb 29: SDSC has installed Relion, Frealign and EMAN2 on Comet and Gordon. Users can add the software to their paths using the following commands:

- module load relion
- module load frealign
- module load eman2

New Publications

published Mon, 11/23/2015

New open access articles about EMDataBank and EMDB access are now available online, in advance of publication in the upcoming January 2016 Nucleic Acids Research Database Issue.

- EMDataBank unified data resource for 3DEM provides an overview of the rapidly growing 3DEM structural data archives, which include maps in EM Data Bank and map-derived models in the Protein Data Bank. Also, discussion of progress and approaches toward development of validation protocols and methods, working with the scientific community, in order to create a validation pipeline for 3DEM data.
- PDBe: improved accessibility of macromolecular structure data from PDB and EMDB describes PDBe's website redesign, API access to the PDB and EMDB archives, and value-added annotations.

Challenge Submissions Update

published Thu, 01/28/2016

The following updates have been made to the challenges site this week:

1. Model challenge submissions are now open.

2. Challenge submissions (both map and model) now require login to the challenges site. Emails have been sent out to all challenger registrants with their login information.
3. Challenge news is now available via rss feed.

Map Submission Deadline Extended (Again)
published Tue, 03/29/2016

By popular request, we are extending the map challenge submission deadline to April 15, 2016 @ 21:00 UTC (5 PM US ET). Click here to see the new deadline time in your location.

Assessment Phases are Coming Soon
published Fri, 10/28/2016 - 11:20

Following two amazing challenge submission finishes in April (66 maps!) and in June (106 models!), we have been working with our respective committees to prepare and organize the data for blinded assessments and to perform preliminary analyses. This process has taken more time than originally anticipated, so we have been making adjustments to assessment phase timelines. For the map challenge, we plan to announce the beginning of the assessment phase in early November. Watch this space for more details!

Map Challenge Assessment Phase is Now Open
published Mon, 11/07/2016

The Assessment Phase of the EMDataBank Map Challenge is now officially open, and all are welcome to participate.

The Challenge Phase (July 2015-Apr 2016) was a tremendous success, with 66 submitted maps spread across the 7 image data benchmark targets. **We recognize the significant time commitment required and are grateful to the 27 challengers who contributed their efforts and made their results available for analysis.**

Following review, the challenge data and files are now publicly available to assessors. The authors of the entries and the software used have been suppressed to promote a blind assessment. We welcome input from the assessors as to how to most usefully evaluate these results. For example, assessment/comparison methods could include statistical analyses, resolution estimation, or fitting of atomic models. Some suggestions are provided on the website but these are not meant to be prescriptive.

Map Challenge Webinar
published Thu, 11/10/2016

As a reminder the Blind Assessment phase of the EMDataBank Map Challenge is now open. **A 1 hour webinar** held on November 17 @15:00 UTC (9:00 US-CT, 10:00 US-ET, 15:00 UK, 16:00 Eur, 23:00 China) covered these topics:

- map challenge submissions overview
- how the submitted data and files were prepared for blind assessment
- how to download/access the data
- what preliminary analyses are already available

- how to participate in the assessment phase
- Q/A period

You can view/listen to the recorded webinar here: https://youtu.be/Eq_-ZH-olUg

---

Map Challenge Assessment Update
published Thu, 04/13/2017

Map Challenge Assessors:

- If you have not yet completed your analysis, the deadline for submitting assessment reports is now end of April (April 30). After that time we will be releasing the masked software/submitter info.
- We are working on the planning for a two day face-to-face discussion of assessment results, and are considering several possible dates/locations either over the summer or (more likely) in September.
- Following the workshop each assessor/assessment group will have the opportunity to publish their findings in a journal special issue.

Map Challenge: Data Released!
published Tue, 06/06/2017

Several updates for the map challenge:

(1) The method-blinded Assessment Phase is now over and we have six excellent assessor reports which rank the submissions based on a variety of criteria. These are available to view/download here: http://challenges.emdatabank.org/?q=map-assessment.

(2) The full set of metadata collected for each Map Challenge submission is now available for download. To access this info please go to http://challenges.emdatabank.org/?q=maps2016-download and click on the file link "map-challenge-metadata-full.xslx."

(3) We are aiming to hold a joint EMDataBank Map and Model challenge face-to-face meeting for all participants, to take place in Stanford, CA. More info to follow -- for now please reserve Oct 5-8, 2017 on your calendar (includes travel days).

Many thanks to all who have contributed to this successful community effort!

addendum June 7: Particle parameter files for each submission are also now publicly accessible here.

Joint Challenges Wrap-Up Meeting Oct 6-8
published Wed, 08/16/2017

In 2016 EMDataBank ran two community challenges in parallel to create awareness of the need for cryoEM structure validation as a routine process in research studies and publications, and to expedite development of quantitative tools for assessment. The Map and Model Challenges were developed by cryoEM and modeling community experts, respectively, who have been charged with developing challenge tasks, promoting worldwide participation, evaluating the results, and producing a report. In each case, benchmark datasets (i.e. raw single particle images and 3D density maps) have been assembled for molecular machines of varying size and

complexity, based on current state-of-the-art detectors and processing methods, in the resolution range 2-5 Å. Challenge tasks are designed to be suitable for all levels of expertise.

The cryoEM and modelling scientific communities have responded enthusiastically : a grand total of 83 scientists have participated as committee members, challengers, and/or assessors. There were 66 submissions to the Map Challenge, and 107 submissions to the Model Challenge, each with supporting details about workflow from benchmark data to final result. Analyses of all of these depositions is now nearing completion, making use of both currently available as well as novel procedures, conducted by volunteers and experts.

In order to share and fully explore the results and analyses of both challenges with the community, we plan to hold a joint Challenges Wrap-Up Workshop October 6-8, 2017 at the Conference Center of SLAC National Accelerator Laboratory, Stanford University, Menlo Park, California. We are inviting all of the participants from both challenges to present and discuss their findings, providing a unique opportunity for two somewhat separate communities (3DEM reconstruction and molecular modelling) to come together to review the challenge results, to address the need for robust validation procedures for maps and models, and to make recommendations for future challenge events for increasingly complex data with high compositional and/or conformational heterogeneity.

The format of the meeting is to have the first day devoted to density map generation from raw single particle images and the second day devoted to modeling from 3D density maps. Each of these two sessions will have presentations from assessors and challengers on their chosen computational approaches and their rationales of adoptions. The session discussion leaders will be drawn from our Committee experts. The third day will be devoted to the necessary metrics of cryoEM structure validation report for structures archived in EMDB and PDB, discucssion on integration of map and model validation, and possible topics and formats for future challenge events. After the workshop, we plan to organize a journal special issue that will be contributed by the assessors and challengers so that the outcomes will be disseminated freely to the entire scientific community.

If you are interested, please join us! The workshop registration site is here: http://ncmi.bcm.edu/ncmi/events/workshops/workshops_163

Joint Challenges Wrap Up Workshop: Thanks to Our Participants!

published Thu, 10/12/2017

The Oct 6-8, 2017 Joint Challenges Wrap-Up workshop at Stanford/SLAC was a tremendous success.

With more than 60 scientists attending, participants of the 2016 Map and Model Challenges, including challengers, assessors, committee members presented and discussed their findings, and to help to develop recommendations for future challenge events. More outcome details will be posted soon.

JSB Special Issue on Outcomes of the Map and Model Challenges

published Mon, 11/13/2017

We are pleased to announce that the *Journal of Structural Biology* has agreed to produce a special issue on the 2016 Map and Model Challenges.

For those planning to submit a manuscript, here are the particulars:

Submission Format and Guidelines

All submitted papers must be clearly written in excellent English and contain only original work which has not been published by or is currently under review for any other journal or conference. A detailed submission guideline is available as "Guide to Authors" at: https://www.elsevier.com/journals/journal-of-structural-biology/1047-8477/guide-for-authors

All manuscripts and any supplementary material should be submitted through Elsevier Editorial System (EES). Select **VSI:2016 CryoEM Challenges** when you reach the **Article Type** step in the submission process. This will ensure that all manuscripts are correctly identified for inclusion into the special issue.

We have been advised that there will be no publication charges to authors, and use of color figures will be free. In addition, Elsevier has agreed to give the entire issue promotional free access during the 1st 6 months following publication.

The earliest submission date will be **February 1, 2018**. The final submission deadline is **March 1, 2018.**

The EES submission site is located at: https://ees.elsevier.com/jsb/default.asp

Please refer to the journal's Guide for Authors for specific advice on how to prepare your paper.

All papers will be peer-reviewed by three independent reviewers.

Requests for additional information should be addressed to the guest editors, Wah Chiu and Cathy Lawson.

Goals

- Establish a benchmark set of single particle raw image datasets suitable for high resolution cryoEM, suitable for both software developers and beginners
- Encourage developers of 3DEM software packages and biological end users to analyze these datasets and present results with the best practice
- Evolve criteria for evaluation and validation of the results of the reconstruction and analysis
- Compare and contrast the various reconstruction approaches in a positive spirit, to achieve high efficiency and accuracy

Map Committee

Bridget Carragher (Chair), Jose-Maria Carazo, Wen Jiang, John Rubinstein, Peter Rosenthal, Fei Sun, Janet Vonck, Wah Chiu, Cathy Lawson, Ardan Patwardhan

How to Participate

All members of the Scientific Community--at all levels of experience--are invited to participate as **Challengers,** and/or as **Assessors**.

In the Challenge Phase (July 2015-Apr 2016), 27 participants created and submitted 66 reconstructions of the challenge targets using the supplied raw image data. Challengers were encouraged to perform their own movie frame alignment, frame summation, and particle picking. Alternately, they could begin with pre-aligned, summed images and/or original author-provided particle positions. For each submission, challengers filled out a questionnaire and provided the following data:

- final unmasked map with filtering/sharpening
- final unmasked map without filtering or sharpening
- half-maps and mask used for FSC calculation
- CTF, coordinates, euler angles for each particle image used in the reconstruction

In the Blinded Assessment Phase (Nov 2016-Apr 2017), six groups have contributed reports. Following initial review period by the map committee, the challenge data and files are now publicly available (entry authorship and software suppressed) for anyone to assess. Assessment methods could include statistical analyses, resolution estimation, or coordinate model fitting. A few suggestions gathered from software developers are summarized below. The intention is to enable comparisons of the various packages available and their options in a positive spirit. During this period, assessors are strongly encouraged to share their plans and

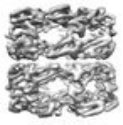
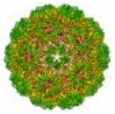
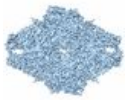
short result summaries on this website using the Assessment Registration Form. Assessment results will be more fully presented and discussed via a workshop (early October 2017) as well as via manuscript submissions to a Journal special issue.

Timeline

JAN-JUN 2015	**DEVELOPMENT PHASE**
Jan-May	Map Committee meets monthly to identify challenge targets, goals, and parameters
Mar-May	Requests to 3DEM community members for public deposition of raw image Datasets; website development
July 1	Raw image data for all targets available for download at EMPIAR
JUL 2015-APR 2016	**CHALLENGE PHASE**
July	Pre-Challenge Announcement, Challenger and Assessor Registration Opens
August 1	Map Submission Opens
Nov 9-20	Registered Participants may apply for SDSC Gordon Supercomputer Usage
April 1-2	Map Challenge Committee satellite discussion @ International CryoEM Image Analysis Symposium (Lake Tahoe, CA)
Apr 15	Map Submission Closes @ 21:00 UTC
2016-2017	**ASSESSMENT PHASE**
May-August	Challenge Data initial assessments, metadata extraction, preparation for release (Map Committee)
Sept-Oct	Review Period (Map Committee)
4 Nov 2016-30 Apr 2017	**Assessors invited to perform analyses and comment on Released Data (Blinded)**
6 June 2017	Map Submission Data UnBlinded
June - Sept 2017	Analysis of Assessments with full metadata
Oct 6-8 2017	2 day Workshop for all challenge participants -- Committee, Challengers, Assessors
Post-workshop	Challenge Writeups (multiple articles) for a Journal Special Issue

Map Challenge Targets

Six challenge targets are based on recently described 3DEM single particle structure determinations with data collected as multiple-frames-per-second movies, using the latest generation of detectors. One additional target is based on simulated (in silico) images. For each experimental target, the original raw micrograph movie frames are data available for download at EMPIAR, PDBe's raw 3DEM image data archive. Summed image data are also available, either as full micrographs or as picked particle stacks. In one case aligned frames are also deposited. Particle positions and defocus values from the raw data depositors are also available for download and may optionally be used by challengers in their reconstructions.

target	**1. GroEL *in silico***	**2. T20S Proteasome**	**3. Apo-Ferritin**	**4. TRPV1 Channel**	**5. 80S Ribosome**	**6. Brome Mosaic Virus**	**7. β-Galactosidase**
Reference EMDB map entry	 --	 **EMD-6287**	 **EMD-2788**	 **EMD-5778**	 **EMD-2660**	 **EMD-6000**	 **EMD-5995**
Primary Citation	**Vulovic et al**	**Campbell et al**	**Russo & Passmore**	**Liao et al**	**Wong et al**	**Wang et al**	**Bartesaghi et al**
Reported Resolution (Å)	~3	2.8	4.7	3.3	3.2	3.8	3.2
Reference Model PDB entry	4hel **RCSB-PDB** / **PDBe** / **PDBj**	1yar **RCSB-PDB** / **PDBe**/ **PDBj**	4v1w **RCSB-PDB** /**PDBe** / **PDBj**	3j5p **RCSB-PDB** / **PDBe** / **PDBj**	3j79/3j7a RCSB PDB: **LS**, **SS** PDBe: **LS**, **SS** PDBj: **LS**, **SS**	3j7l **RCSB-PDB** / **PDBe**/ **PDBj**	5a1a **RCSB-PDB** / **PDBe**/ **PDBj**
Benchmark Storage Size	2 GB	2000 GB	181 GB	6300 GB	2000 GB	460 GB	550 GB
target	**1. GroEL *in silico***	**2. T20S Proteasome**	**3. Apo-Ferritin**	**4. TRPV1 Channel**	**5. 80S Ribosome**	**6. Brome Mosaic Virus**	**7. β-Galactosidase**
*EMPIAR ID(s) data can also be downloaded from **Chinese Academy of Sciences***	**EMPIAR -10029**	**EMPIAR -10025**	**EMPIAR -10026**	**EMPIAR -10005**	**EMPIAR-10028**	**EMPIAR -10010FAQ EMPIAR -10011**	**EMPIAR -10013 EMPIAR -10012**
Raw Frames	n.a.	✔	✔	✔	✔	✔	✔
Aligned Frames		✔					
Summed Micrographs		✔		✔	✔		✔
Summed Particle Stacks	✔		✔		✔	✔	
Initial Particle Coordinates (directory link)				**spider FAQ**		**eman-box**	**eman-box**
Final Particle Coordinates (direct file link)		**relion-star**	**relion-star**		**relion-star**		
*Particle coordinates in **EMX format**; python script used for conversion files contributed by Roberto Marabini and Jose Maria Carazo*		**10025.emx 10025.py**	**10026.emx 10026.py**	**10005.emx 10005.py**	**10028.emx 10028.py**	**10011.emx 10011.py**	**10013.emx 10013.py**
target	**1. GroEL *in silico***	**2. T20S Proteasome**	**3. Apo-Ferritin**	**4. TRPV1 Channel**	**5. 80S Ribosome**	**6. Brome Mosaic Virus**	**7. β-Galactosidase**
Imposed Symmetry	Dihedral (D7) None (C1)	Dihedral (D7)	Octahedral (O)	Cyclic (C4)	None (C1)	Icosahedral (I)	Dihedral (D2)

	FAQ						
Sample MW (MDa)	0.8	0.7	0.44	0.3	4.2	4.6	0.47
Unique MW (kDa)	57	50	20	80	4200	80	120
Microscope	--	Titan Krios	Polara 300	Polara 300	Polara 300	JEOL3200FSC	Titan Krios
Voltage(kV)	300	300	300	300	300	300	300
Cs (mm)	2.7	2.7	2.7 **FAQ**	2.0	2.0	4.1	2.7
Detector	*Falcon I*	K2	Falcon II	K2	Falcon II	DE12	K2
Frame Sampling (Å/pixel)	1.42	0.6575	1.346	1.22 **FAQ**	1.34	0.99	0.64
total dose ($e^-/Å^2$)	50	53	16	41	20	52	45
dose per frame ($e^-/Å^2$)	--	1.4	0.95	1.37	1	1.4	1.2
frame rate (f/s)	--	5	17	5	16	25	2.5
frame alignment method	--	**UCSF**	not performed	**UCSF**	**Statistical**	DE script	cross-correlation script
Particle selection method	--	**Appion -FindEM**	**EMAN2**	SamViewer	**EMAN2**	**EMAN2**	Gaussian correlation
Number of Particles	10000	49954	483	35645	105247	30000	11726
Particle/Map Sampling (Å/pixel)	1.42	0.98	1.346	1.22	1.34	0.99	0.64
Raw Data Contributors (Thank You!)	Yuchen Deng, Fei Sun	Melody Campbell, Bridget Carragher	Chris Russo, Lori Passmore	Jean-Paul Armache, Maofu Liao, Yifan Cheng	Xiaochen Bai, Sjors Scheres	Zhao Wang, Wah Chiu	Alberto Bartesaghi, Sriram Subramaniam

Assessment Guidance

We've gathered some suggestions here about how to proceed with comparisons of the map submissions.* These are not meant to be prescriptive; results from other approaches are also welcome.

FSC Analysis

FSC curves based on provided half-maps and masks have been prepared for each map challenge entry. In most cases the results are consistent with or very close to the submitter-reported resolution, but this initial analysis cannot be used to directly compare submissions, because of differences in masking and map sizes, and thus convolution effects. FSC is a fundamental similarity metric, but its use in standard cryoEM practice has been problematic because of the maps being compared. Many suggestions were made on how to carry out follow-up analyses:

- Apply a single, common mask to all entries belonging to a target (e.g. 15-20 A average of all entries, with soft-edges or low-pass filtered).
- Employ other methods/techniques such as:
    - post-process phase randomization (e.g., to investigate effects of different masking on FSC)
    - mask artifact compensation
    - determine FSC error
    - Calculate map-model FSCs

Map Density Analysis

Images of each map both by itself and aligned to a common model are provided for reference (link), but further investigation is warranted, as variations in density appearance may be due to differences in power spectra and/or filtering/sharpening schemes. Some suggestions:

- Both overall images and close-up views are desirable; for comparison it is best to have the exact same view
- Both well-ordered regions and not-so well ordered regions should be investigated
- Views containing slices (slabs or grey-scale planes) could be useful
- Apply a common filtering/sharpening scheme to the unfiltered (raw) map entries for a target, bringing power spectra to a “common denominator” for density comparison
- Along this line, view density across maps attenuated at a common low resolution, and then walk the attenuation towards higher resolution
- Density quality could be investigated by fitting defined portions of each map using modeling tools (e.g. compare rmsd's of multiple models).

*suggestion credits: Maya Holmdahl, Roberto Marabini, Sjors Scheres, Bernard Heymann, Niko Grigorieff, Pawel Penczek, Ed Egelman, Steve Ludtke, Scott Stagg, Marin van Heel

FAQ

Below are compiled questions/answers regarding the map challenge that may be of interest to all of the participants. We'll update this FAQ as needed.

Challenge Phase

1. Is the Cs listed in the target table (2.7) correct for the Apoferritin data? No. There was an error in the script to generate the metadata requested for the challenge. The manufacturer specified Cs value for the Polara is 2.0 mm, but this has never been measured accurately for the instrument. In practice, this is an somewhat arbitrary fitting parameter that can be input during analysis of the raw data or refined during fitting. **Followup question: What Cs do the defocus values provided in the particle stack refer to?** The provided defocus values assume Cs = 2.7. *posted Nov 20, 2015, thanks to Niko Grigorieff for questions, and Chris Russo for answers*

2. Why are more images provided in the newly added gain-corrected subdirectory vs original non-gain-corrected subdirectory for BMV (EMPIAR-10010)? The gain-corrected data include the whole collected dataset, including images with lower resolution; the non-gain-corrected images correspond to those selected by the deposition authors for processing. *posted Nov 20, 2015, thanks to Ardan Patwardhan for question and Zhao Wang for answer. Please also see **note below** from Ben Bammes regarding the method used for gain-correction in the original work.*

A note about BMV image frame data *published Tue, 12/22/2015 - 11:18*

Some map challenge participants may be encountering difficulty processing the BMV individual frames data (not the original boxed out particle images). One has to work out the image processing scheme which is a part of the challenge. To help challengers understand the data, the BMV raw data providers have provided the following note:

Note on the BMV image frame data -- provided by Benjamin Bammes *(Wang, Z, Hryc, CF, Bammes, B, Afonine, PV, Jakana, J, Chen, DH, Liu, X, Baker, ML, Kao, C, Ludtke, SJ, Schmid, MF, Adams, PD, & Chiu, W (2014) An atomic model of brome mosaic virus using direct electron detection and real-space optimization. **Nature communications 5:4808**) :*

"We collected dark and gain reference images for each day of data collection except 2013-01-12, for which we used the dark reference image from the following day (2013-01-13). Note that the dark and gain reference images from each day were rotated and/or flipped if necessary to match the orientation of the raw frames from that day.

In order to improve statistics of the gain reference image, we averaged all the dark-subtracted gain reference images to create one average gain reference image to apply to the data from all days of data collection. The average gain reference image was converted to 32-bit float and then divided by its mean intensity, so that the mean intensity of the average gain reference image was normalized to one. We then applied a threshold to the average gain reference image such that all pixel values less than 0.01 were set to 0.01. Finally, we inverted the average gain reference image by calculating the reciprocal value of each pixel.

All raw frames were processed by first applying dark correction, and then applying gain correction. Dark subtraction was performed simply by integer subtraction of the day's dark reference image from each raw frame. Gain correction was then performed by multiplying each frame by the average gain reference image to produce the final flat-field corrected, 32-bit float stacks of frames.

Note that we did not yet apply any sigma filter to remove X-ray pixels and/or detector noise from each frame."

3. The TRPV1 challenge dataset lists the pixel size as 1.22 A, but after downloading the movies we found that they are stored in super-resolution and the pixel size is therefore 0.61 A. Could this be clarified? The movie stacks in the section of "TRPV1 raw multi-frame micrographs" (6.4G each) are super-resolution images, and should have a pixel size of 0.61 A (or 0.6078 A). *posted Dec 3, 2015, thanks to Niko Grigorieff for the question, and Maofu Liao for the answer.*

4. We cannot find particle positions for TRPV1 micrographs in EMPIAR-10005, are they missing? Yes!, at least up until the writing of this FAQ item. Following consultation with the data contributors, particle coordinates were added and are located in a new subdirectory of the EMPIAR entry micrographs section (**picked_coordinates**). Please note that these are SPIDER coordinates for all picked particles (not the final particle set). Also, the coordinates are for 3x binned images, so values will need to be multiplied by 3 to correspond to unbinned images. The 88915 folder reflects the initial particles that were selected by Maofu Liao for further processing. The 35645 one relates to the final particles he used in his processing to obtain the high-resolution map (meaning: he initially extracted 88915 particles, then post 2d and 3d classification he proceeded to minimize the stack by subselecting the final group of particles from his "star" file -> these 35645 particles). [Previously, we erroneously pointed to **a relion star file** in the "final particle coordinates" row of the target table, but that file provides shift information for the picked particle stack.] *posted January 18, 2016, thanks to Jose Maria Carazo and Carlos Oscar Sorozano for the question and Yifan Cheng and Jean Paul Armache for coordinates and answers. Picked coordinates link updated March 21, 2016, thanks to John Rubenstein for reporting that the link was broken.*

5. Are submissions being looked at currently by anyone who would also participate in the challenge? Or are they all 'confidential' until the deadline? We are setting up a process that copies the submissions to an ftp area with all submitter-related data removed. The content/structure of the ftp is currently under review by the map committee. We will open up ftp access to everyone during the assessment period, but submissions will remain anonymous until the end of the assessment period. *posted January 28, 2016, thanks to Sjors Scheres for the question.*

6. What is the correct symmetry for the GroEL simulated data? It is C1, not C7, please see **clarification below**. *posted Feb. 19, 2016*

Clarification about symmetry of the GroEL simulated data *published Fri, 02/19/2016 - 12:07*

The authors of the GroEL simulated dataset have asked us to let all map challenge participants know that 4HEL, the model from which the images were generated, is non-symmetrical, owing to unmatched crystallographic symmetry ($P2_12_12_1$). They have just become aware of this issue following discussion with two challenge participants.

To try to reach high resolution it will therefore more appropriate to perform reconstructions assuming C1 symmetry instead of D7 symmetry. (D7 resolution limit is ~ 4 angstrom). The full hemisphere of projections were generated by the authors, so in principle enough information exists to reconstruct in the lower symmetry. However, since there are only 10K particles it will be very interesting to see if it is possible.

Thanks for Steve Ludtke and Sjors Scheres for initiating the discussion, and Fei Sun for the clarification.

7. Can I fill in multiple submissions simultaneously? This is not recommended. If you have multiple maps resulting from a single benchmark, these need to be submitted separately, and we suggest to work on one at a time. Once you have completed one submission you will get an email with your answers. You can use then use email text to cut/paste answers as appropriate into your subsequent submissions. Also, please note that you can save drafts while you are working, and further edit your submission after you submit. posted *March 29. 2016, thanks to Shabih Shakeel for the question.*

Blind Assessment Phase

8. Are CTF parameters, coordinates, and Euler angles of particle images of the submitted maps available for assessment of the maps? We have not been able to find this information. Assessment of the submitted 3D maps without additional information is the main goal at this stage. We have held-back the CTF/coordinate/euler information during the blind assessment phase because the file formats provided (e.g., .star, .spi files) in many cases do "give away" the identity of software used for particular map submissions. Even so, we want to encourage anyone interested to pursue analyses that require this information at a later stage when these data become available. We appreciate that such work will likely yield insights of benefit to the community. *posted December 2, 2016, thanks to Mohammadreza Paraan/Stagg Lab for the question.*

Final Assessment

9. I see that the blind assessment phase has finished ;-). Should we start the not-so blind assessment phase in which more detailed information on how the 3D map has been obtained is available? Questions that would be of interest to answer:

1. **Did the groups that picked the particles perform better than those that use the particles supplied by default ?**
2. **Should we -always- dose filter the movie frames?**
3. **Regarding movie alignment: do global methods perform consistently worse than local ones or is it the other way around (at least for some specimens)?**
4. **Is it better to be very selective picking particles or is it better to pick as many as possible?**
5. **If CTF information can be obtained, it will be very interesting to check up to which frequency ctf computed by the different participant for the same data set are equivalent.**
6. **Particle consensus. For those groups that picked the particles it may be interesting to analyze the intersection and difference between the selected datasets.**
7. **There are many question related with 2D and 3D classification.**

8. **Finally the more important question is what software should I use.**

Answer: Yes!!! With the newly released data it should be possible to further analyze the results of the blinded assessments, and these are the kinds of questions we hope that the challenge can ultimately address. Please have a look at the **list of metadata collected along with each reconstruction**, now available in the **newly released spreadsheet**. Particle parameter sets are also available for download (**ftp://public.emdatabank.org/maps2016/particle_params**). *posted Jun 8, 2017, many thanks to Roberto Marabini for the question.*

Submission Overview

This document provides an overview of the 2015 map challenge web submission form, so that participants can become familiar with the requirements beforehand.

Files Needed for Upload

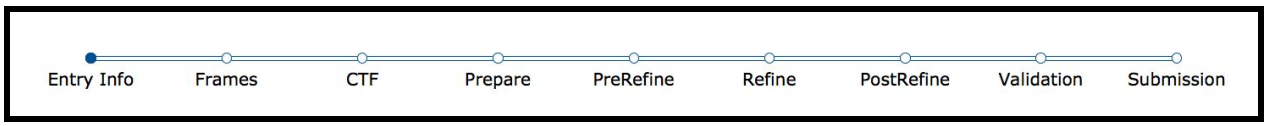
Have these files at-hand before you begin a submission:

- File or files with your final particle parameters, including particle positions, euler angles, and CTF. If your software package allows you to export a file in Electron Microscopy Exchange format, please do so (.emx, preferred format). Otherwise, please create a tar archive file to upload the parameter files (.tar).
- The final unfiltered map -- no masking, filtering, or sharpening (.mrc).
- The final filtered map -- no masking, with filtering and/or sharpening (.mrc).
- The even and odd half-maps used for FSC calculation (.mrc).
- Mask (if any) used for FSC calculation (.mrc).

Questionnaire

The map committee has designed the challenge submission questionnaire with the goals of enabling comparison of how different groups solve structures by EM, providing an educational tool for less-experienced microscopists and establishing reasonable expectations for computational resource requirements.

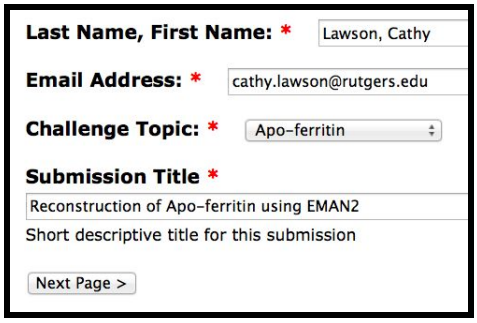
The submission form consists of nine pages in the following order:



Please fill out the requested information as completely as possible. You can navigate back and forth through the pages using the "previous page" and "next page" buttons. (**warning: using browser navigation may reset your submission**). On the final page there is a "submit" button to complete your entry.

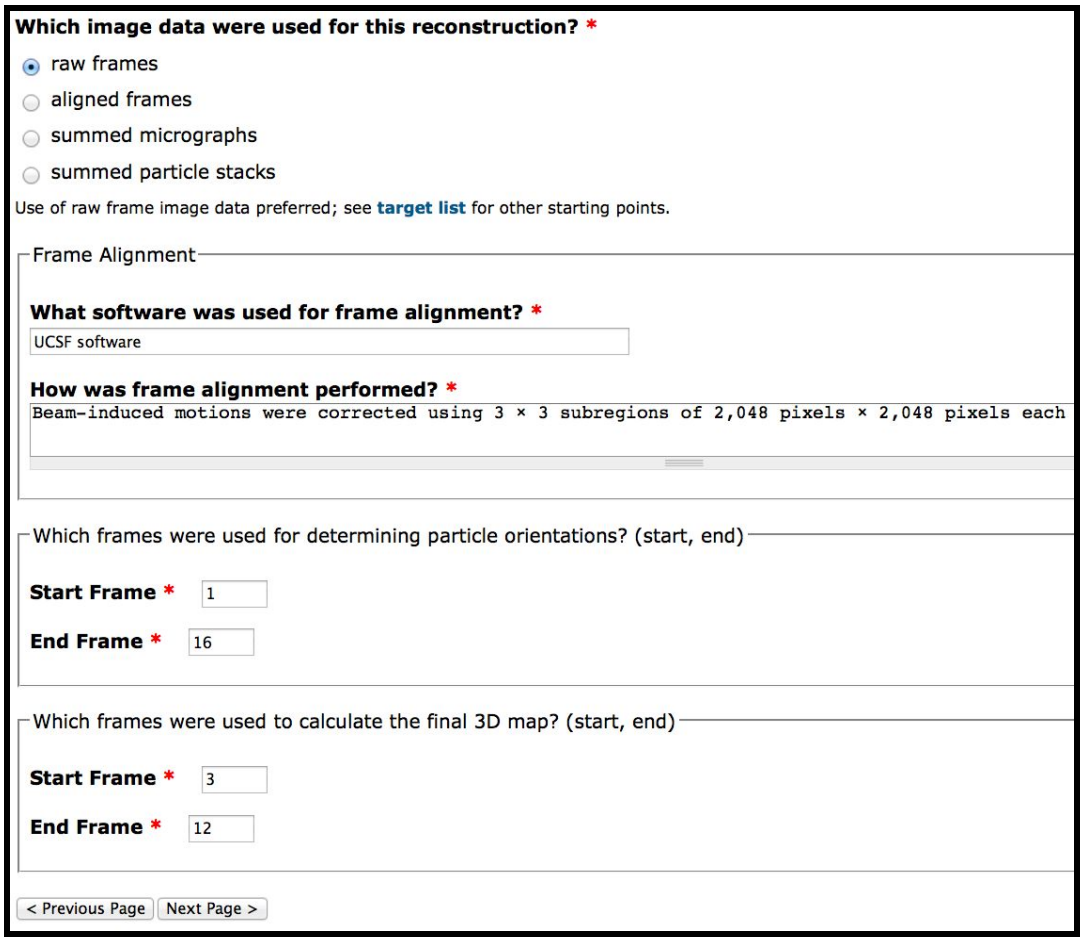
Screenshots of each page are provided below.

Entry Info



Frames

The source image data for the reconstruction and any initial movie frame processing are described here. Frame-related questions are only asked for raw frames and aligned frames. This page is skipped for the *in silico* target.



CTF

How were CTF parameters determined? *

CTF correction was performed on boxed particles using e2ctf.py --autofit

Indicate software name and any relevant settings

How was CTF correction performed? *

phases: particles were phase-flipped (e2ctf.py); amplitudes: corrected during refinement (e2refine.py ctf.auto)

Indicate software name and any relevant settings

Were any other corrections performed on the images? *

no

For instance, beam tilt, anisotropic magnification correction. If yes, indicate software name and any relevant settings

< Previous Page Next Page >

Prepare

Did you use particle coordinates that were provided with the image data? *

- ○ yes (initial particle coordinates)
- ◉ yes (final particle coordinates)
- ○ no

How many particle images were initially selected? *

483

How were the particle images selected? *

e2boxer.py using the provided coordinates as input

Please provide the software name and any relevant parameters.

How was the initial model generated (e.g. from a published map or model)? *

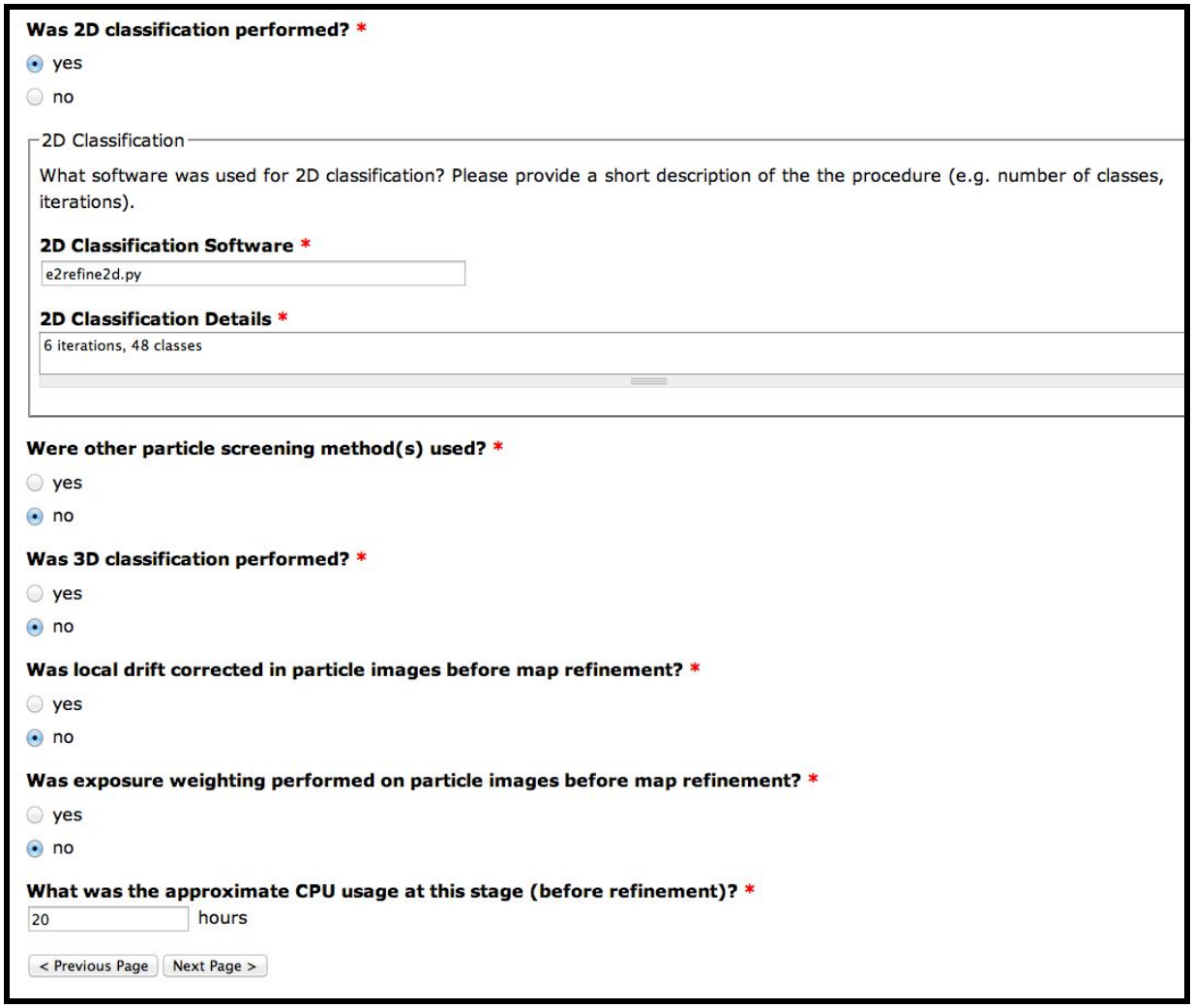
6 dominant reference-free class averages (e2refine2d.py) were manually selected and fed to Simple (v.2.0).

Please provide database name/id, software name and any relevant parameters.

< Previous Page Next Page >

Prerefine

Additional information is requested for yes answers.



Was 2D classification performed? *

(•) yes
( ) no

2D Classification

What software was used for 2D classification? Please provide a short description of the the procedure (e.g. number of classes, iterations).

2D Classification Software *

e2refine2d.py

2D Classification Details *

6 iterations, 48 classes

Were other particle screening method(s) used? *

( ) yes
(•) no

Was 3D classification performed? *

( ) yes
(•) no

Was local drift corrected in particle images before map refinement? *

( ) yes
(•) no

Was exposure weighting performed on particle images before map refinement? *

( ) yes
(•) no

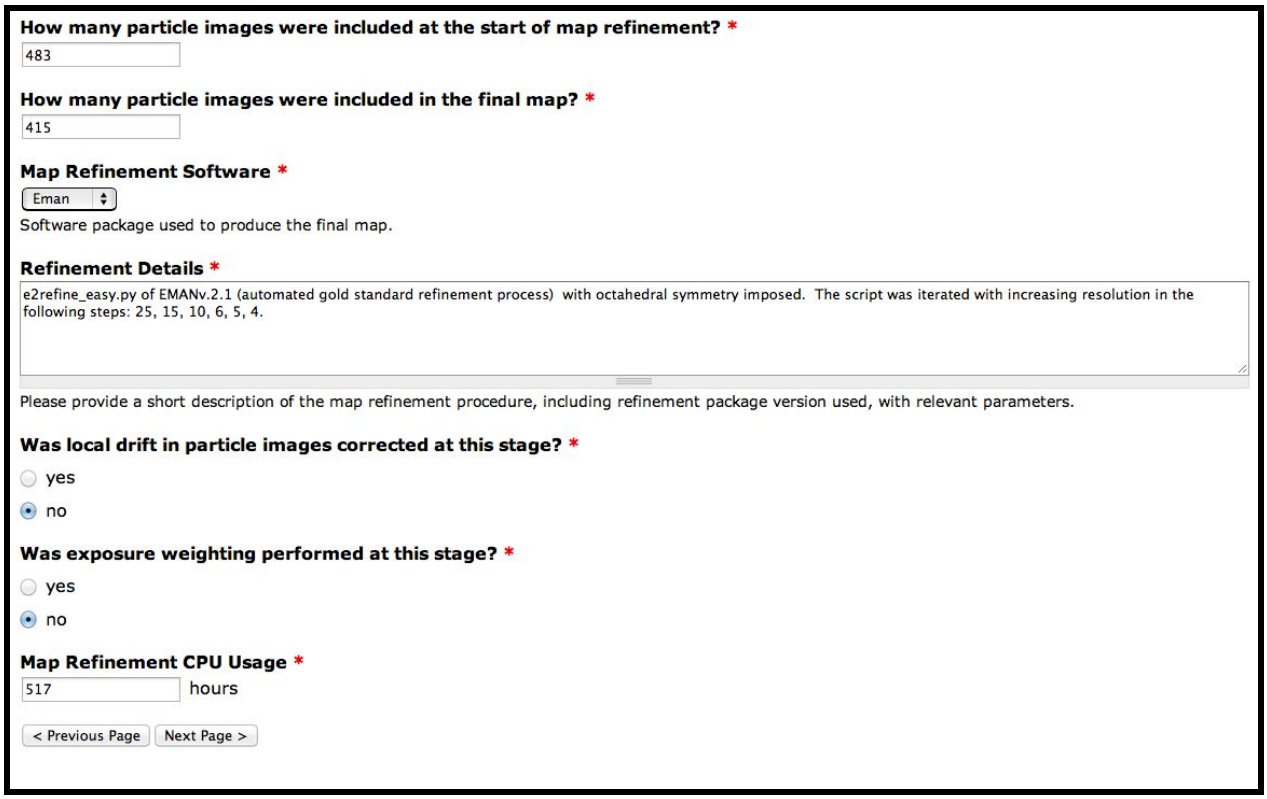
What was the approximate CPU usage at this stage (before refinement)? *

20 hours

< Previous Page Next Page >

Refine

Additional information is requested for yes answers. If “Other” is selected for map refinement software, an additional box appears to collect the software name.



How many particle images were included at the start of map refinement? *

483

How many particle images were included in the final map? *

415

Map Refinement Software *

Eman

Software package used to produce the final map.

Refinement Details *

e2refine_easy.py of EMANv.2.1 (automated gold standard refinement process) with octahedral symmetry imposed. The script was iterated with increasing resolution in the following steps: 25, 15, 10, 6, 5, 4.

Please provide a short description of the map refinement procedure, including refinement package version used, with relevant parameters.

Was local drift in particle images corrected at this stage? *

yes

no

Was exposure weighting performed at this stage? *

yes

no

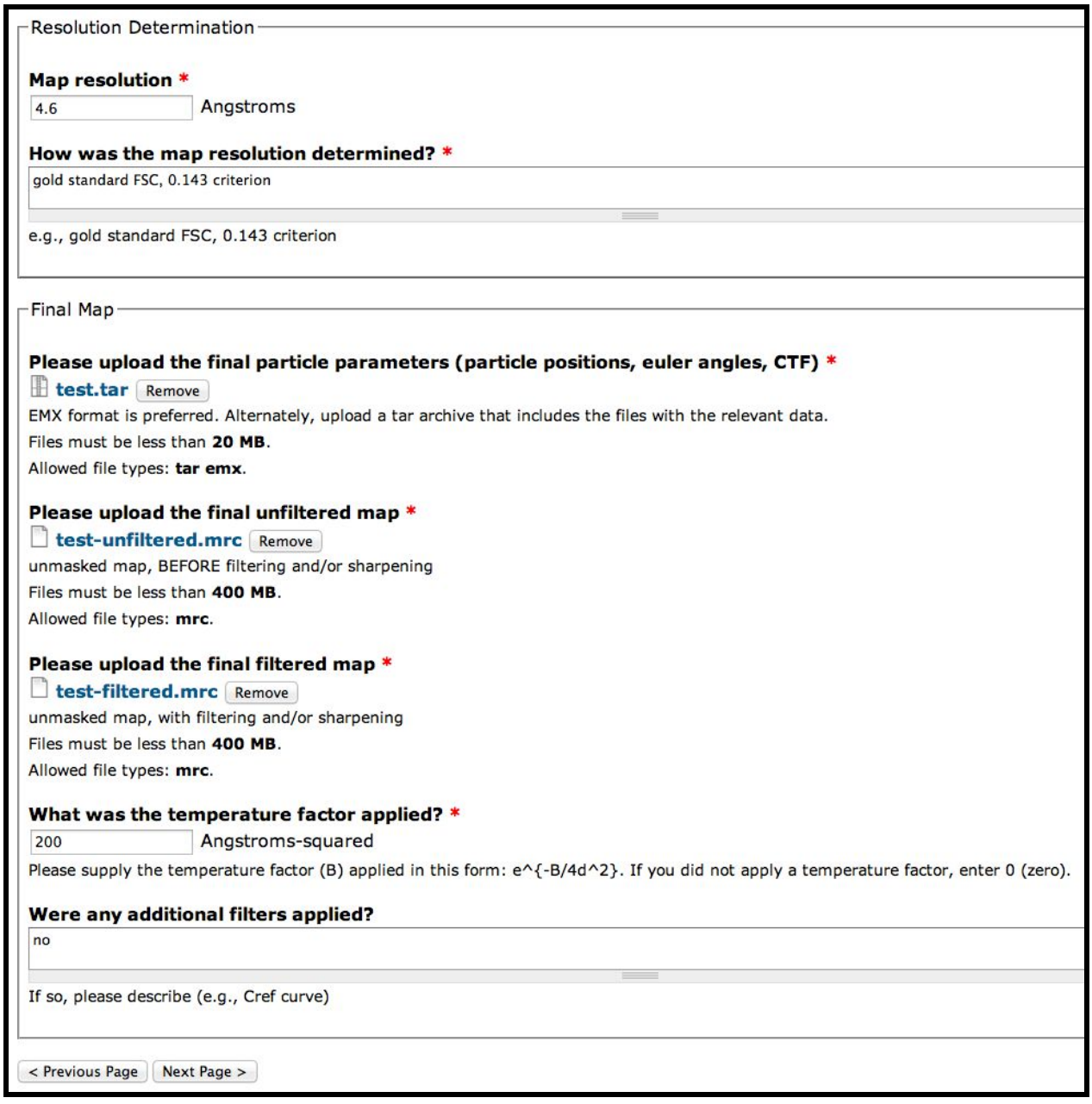
Map Refinement CPU Usage *

517 hours

< Previous Page Next Page >

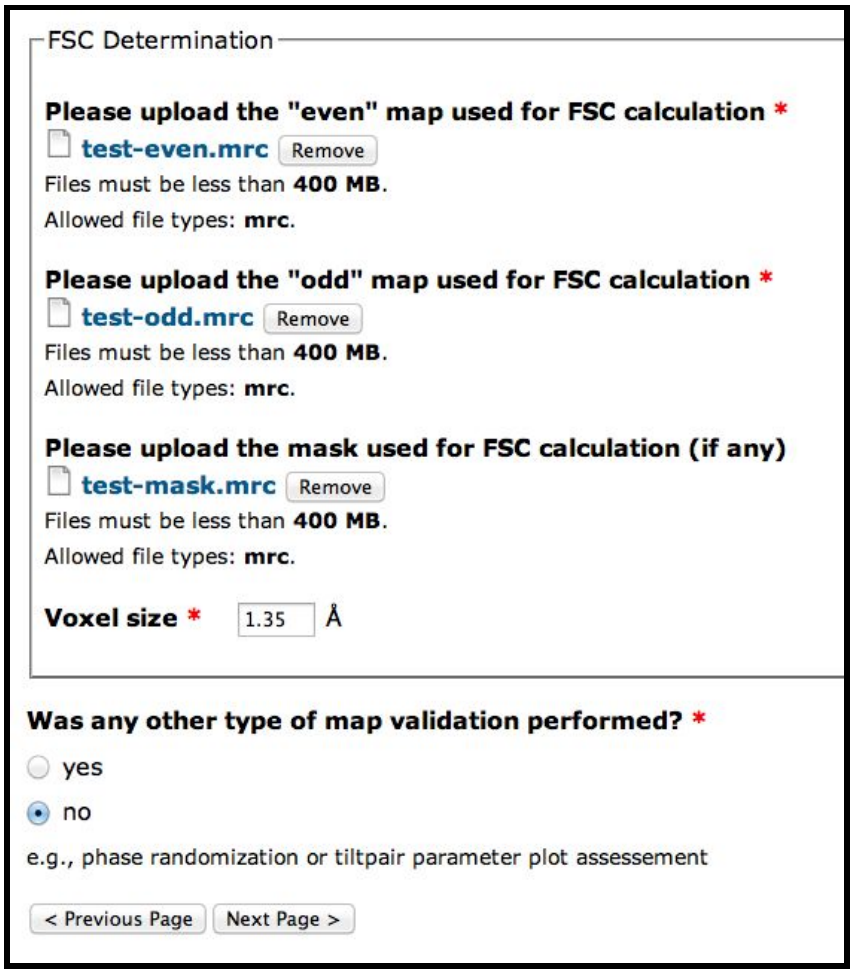
Postrefine

A description of the particle parameter file(s) uploaded is now also requested, including the euler angle convention used (7/21).



Validation

FSC curves from the submitted maps (and optional mask) will be calculated during the assessment phase from the provided data using the FSC validation server at PDBe.



Submission

Two text boxes are provided to request information about (1) computational resources and (2) any additional information not captured by the questions above.

Summary Statistics

- There are 66 total submissions for the 2015/2016 Map Challenge, from 27 registered challengers
- SDSC Gordon Supercomputer service units allocated for the challenge were used for 15 of the 66 submissions, from 4 challengers

Challenge Topic:

Brome Mosaic Virus	7
Apo-ferritin	8
TRPVI Channel	8
80S Ribosome	13
20S Proteasome	9
beta-Galactosidase	12
GroEL Simulated Data	9

Which image data were used for this reconstruction? (n/a for GroEL)

raw frames	45
summed micrographs	4
summed particle stacks	8

Did you use particle coordinates that were provided with the image data? (n/a for GroEL)

yes (initial particle coordinates)	29
yes (final particle coordinates)	10
no	18

Was 2D classification performed?

yes	34
no	32

Were other particle screening method(s) used?

yes	10
no	56

Was 3D classification performed?

yes	19
no	47

Was local drift corrected in particle images before map refinement? (for 45 entries using raw frames)

yes	14
no	31

Was exposure weighting performed on particle images before map refinement?

yes	17
no	49

Was local drift in particle images corrected at [refinement] stage? (for 45 entries using raw frames)

yes	8
no	37

Was exposure weighting performed at [refinement] stage?

yes	13
no	53

Was any other type of map validation performed?

yes	27
no	39

Metadata Handling for Blind Assessment

The table below summarizes the data being released/suppressed for the 2016 map challenge during the blind assessment period.

Released Metadata

- Challenge Topic
- Which image data were used for this reconstruction?
- Start/End Frame for particle orientation
- Start/End Frame for final map
- Did you use particle coordinates that were provided with the image data?
- How many particle images were initially selected?
- Was 2D classification performed?
- Were other particle screening method(s) used?
- Was 3D classification performed?
- Was local drift corrected in particle images before map refinement?
- Was exposure weighting performed on particle images before map refinement?
- What was the approximate CPU usage at this stage (before refinement)?
- How many particle images were included at the start of map refinement?
- How many particle images were included in the final map?
- Was local drift in particle images corrected at this stage?
- Was exposure weighting performed at this stage?
- Map Refinement CPU Usage
- Map resolution
- How was the map resolution determined? (annotated)
- What was the temperature factor applied?
- Any Additional Filters applied to the map? (annotated)
- Was any other type of map validation performed?
- Mask submitted?
- Map Voxel size

Suppressed Metadata

- Submitter name
- Entry title
- What software was used for frame alignment?
- How was frame alignment performed?
- How were CTF parameters determined?
- How was CTF correction performed?
- Were any other corrections performed on the images?
- How were the particle images selected?
- How was the initial model generated (e.g. from a published map or model)?
- 2D Classification Software / Details
- Other Particle Screening Details
- 3D Classification Software / Details
- Local Drift Correction Software / Details (2D)
- Exposure Weighting Details
- Map Refinement Software / Other / Version / Details
- Local Drift Correction Details (3D)
- Exposure Weighting Details (3D)
- How was the map resolution determined?
- Were any additional filters applied?
- Please describe how the map was validated
- Computational Resources
- Any Additional Details
- Final Particle Parameters (uploaded files)
- Map Header Label Field (from uploaded files)

Download Map Challenge Files

Maps submitted to the 2015/2016 Map Challenge [were] available here: ftp://public.emdatabank.org/maps2016. Two subdirectories [held] submitted map data:

- finalmap : final filtered maps and raw unfiltered maps
- even_odd_maps : half-maps and (optional) mask used by the submitter to calculate FSC.

Update June 2017: A third subdirectory holding uploaded particle parameter info is now also available: particle_params

Three additional subdirectories contain supplemental files based on the submitted maps. These are provided to help assessors get started with analyses.

- fitted-models : reference models fitted to each final filtered map
- chimera : UCSF Chimera sessions and matrices with final maps aligned to common reference coordinates (README)
- images : automatically produced images with common views across each target (map-only and map+model)

All map, model, and session files are gzipped. Map files are named according to submission id (emcd###), targetname, and filetype: “emcd###_[targetname]_[filetype].mrc.gz”

- [targetname] = GroEL, Ferritin, Proteasome, TRPVI, BMV, Ribosome, or BetaGal
- [filetype] = filtered, unfiltered, odd, even, mask

Download the full archive (29 Gb) using wget:

wget --mirror -r ftp://public.emdatabank.org/maps2016

Download all files for a specific target:

wget -A "*[targetname]*" -r --mirror ftp://public.emdatabank.org/maps2016/

Rsync access is also available by email request to challenges@emdatabank.org (please include your public ssh key).

Files with reported voxel size for each submission, as well as other associated metadata, are available below. Note that a correction was made to metadata (resolution determination description) for emcd154_BetaGal on July 8, 2016.

Website Footer