

The 2015/2016 Map and Model Challenges

Wah Chiu, Cathy Lawson
challenges.emdatabank.org

NCMI Workshop
November 1, 2015

EMDataBank

- Unified global portal for deposition and retrieval of 3DEM density maps, atomic models, and associated metadata
- Resource for news, events, software tools, data standards, validation methods for the 3DEM community



Supported by NIH National Institute of General Medical Sciences



**EMDataBank**

Unified Data Resource for 3DEM

<http://emdatbank.org>

One-stop shop for 3DEM deposition and retrieval

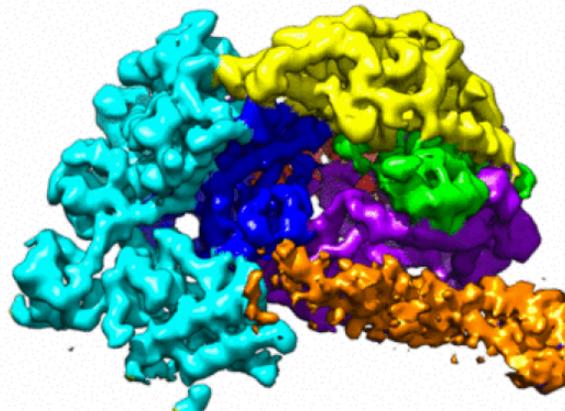
[Home](#)[About](#) ▾[Deposit](#)[Search](#)[Tools](#) ▾[Events](#) ▾[News](#)[Links](#)[Help](#) ▾

Unified Data Resource for 3-Dimensional Electron Microscopy

EMDataBank is a unified global portal for deposition and retrieval of 3DEM density maps, atomic models, and associated metadata, as well as a resource for news, events, software tools, data standards, validation methods for the 3DEM community.

For up-to-date information about map and model challenges, visit challenges.emdatbank.org.

Recently released entries

[All recent entries](#)

EMD-6404

[PDBe](#) | [RCSB](#)

Aug. 1, 2015 RELEASED ON Oct. 28, 2015 singleParticle 4.0Å **NEW**

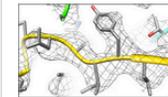
In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus

Zhang X, Ding K, Yu XK, Chang W, Sun JC, Zhou ZH

News

[All news](#)

Announcing the 2015 EMDDataBank Model Challenge



All members of the Scientific Community--at all levels of experience--are

invited to participate as Challengers, and/or as Assessors.

[Read more...](#)

Announcing the 2015 EMDDataBank Map Challenge



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

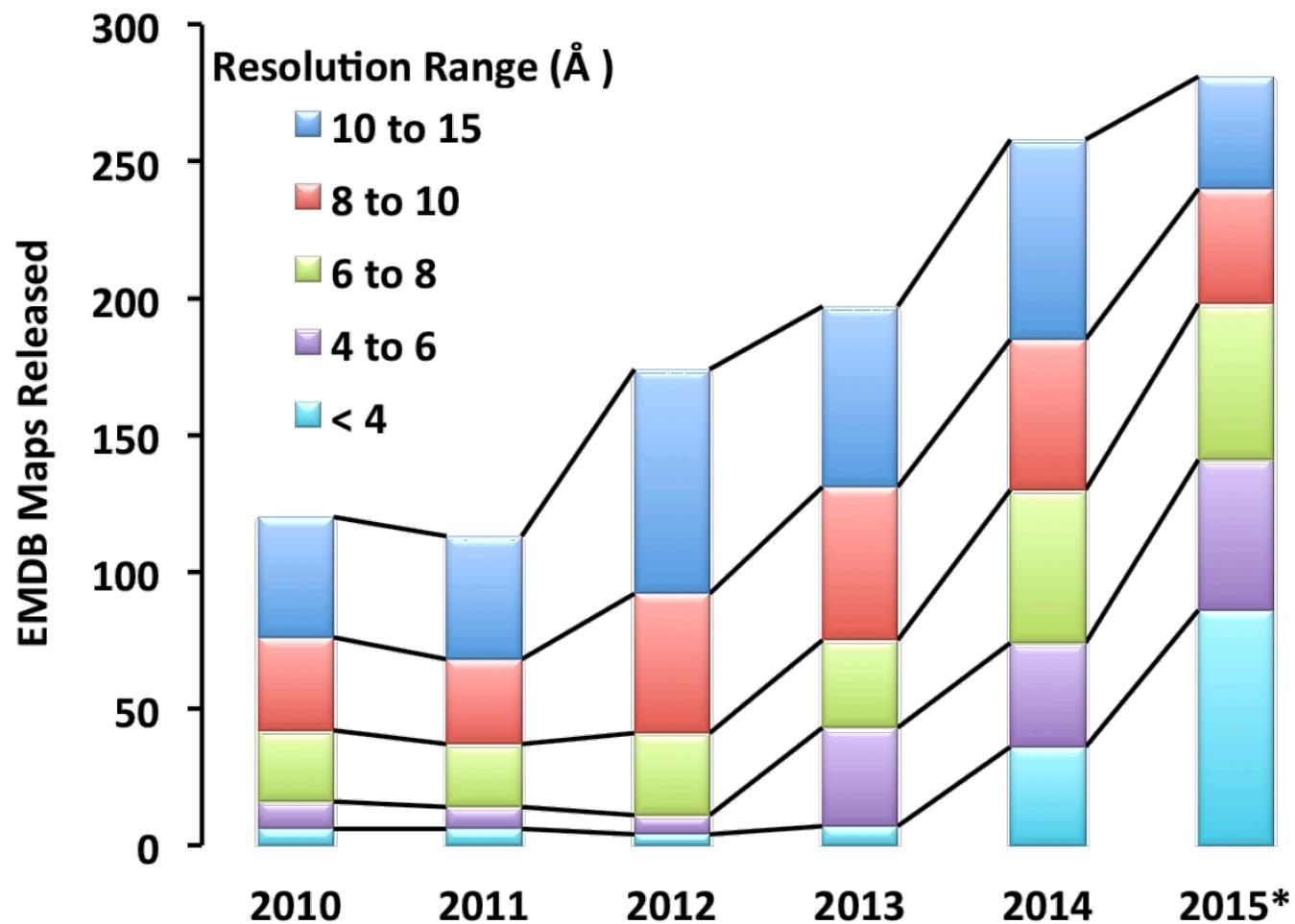
[Read more...](#)[RSS](#)

Supported by
National Institutes of Health
National Institute of General Medical Sciences

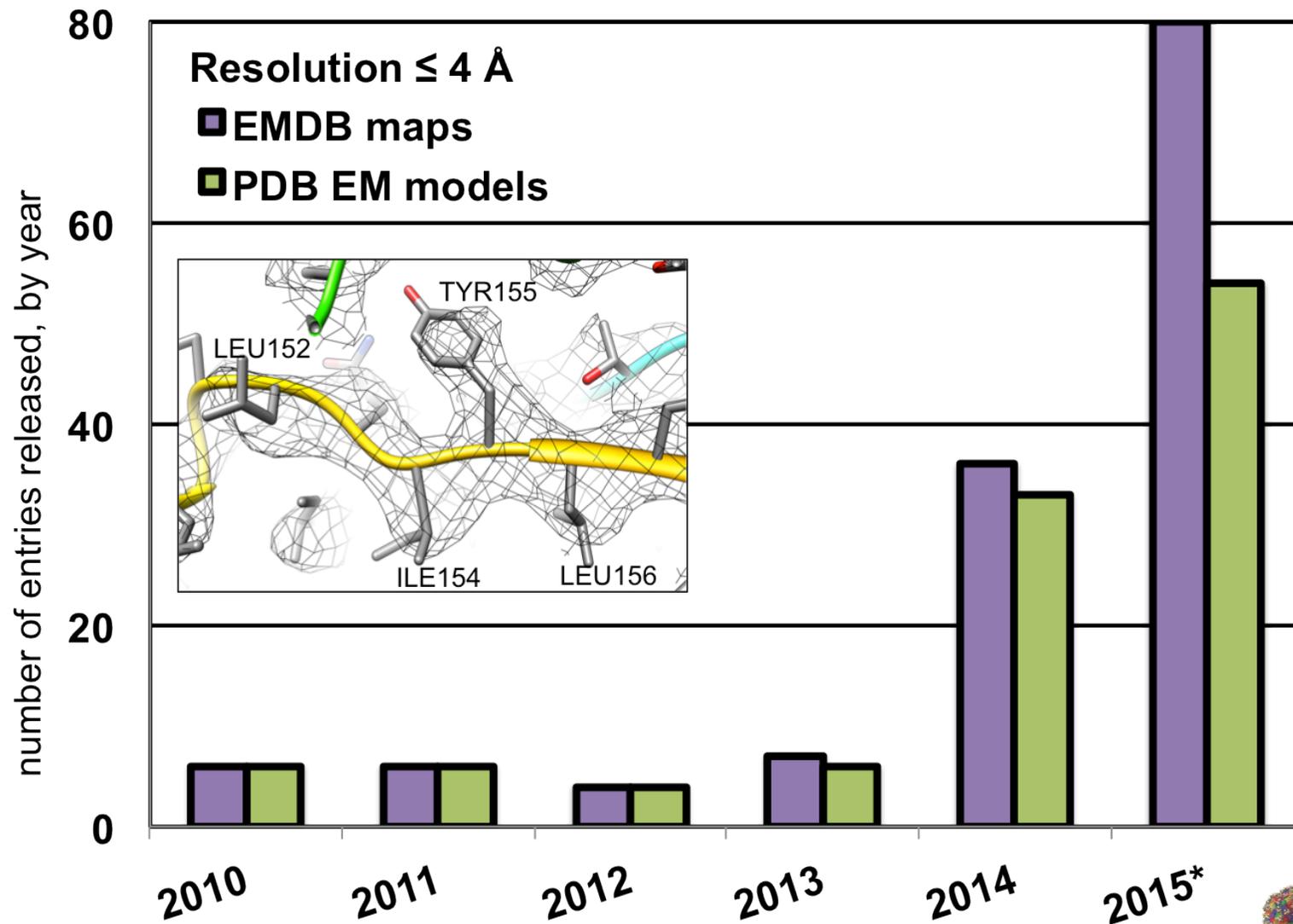


© 2015 EMDDataBank.org

EMDB Map entries vs Resolution

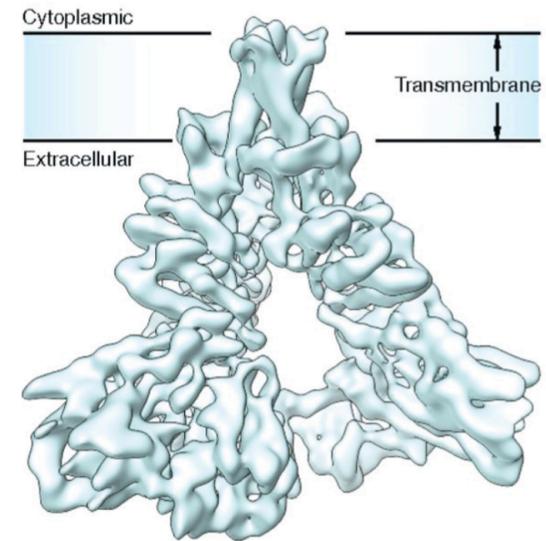


EM Structures @ 4 Å or better



Importance of Validation

- **J. Cohen, Is High-Tech View of HIV Too Good to Be True?** *Science* 341, 443-444 (2013)
- **R.M. Glaeser, Replication and validation of cryo-EM structures** *J. Struct. Biol.* 184, 379-380 (2013)
- **R. Henderson, Avoiding pitfalls of single particle cryo-electron microscopy: Einstein from noise,** *PNAS* 110, 18037-41 (2013)
- **M. van Heel , Finding trimeric HIV-1 envelope glycoproteins in random noise,** *PNAS* 110, E4175-7 (2013)
- **S. Subramaniam, Structure of trimeric HIV-1 envelope glycoproteins,** *PNAS* 110, E4172-4 (2013)



EMD-5418 Y Mao, JG Sodroski *et al.*
Molecular architecture of the uncleaved
HIV-1 envelope glycoprotein trimer *PNAS*
110, 12438-12443 (2013)

Validation Development

through research collaborations
with the community, workshops, challenges:

Assess 3DEM map and map-derived model validation methods

Develop data exchange and archiving standards

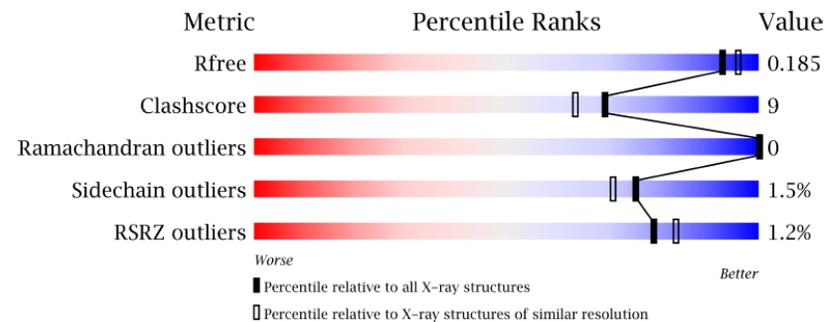
Integrate validation methods, standards into 3DEM validation pipeline



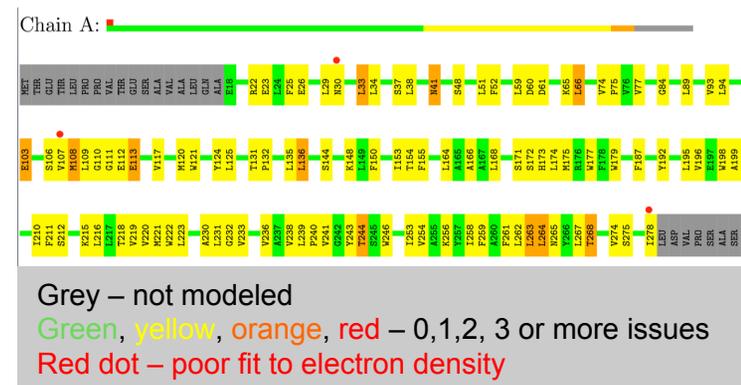
Validation Report: X-ray

- Overall quality at-a-glance
- "Table 1" with key data & refinement statistics
- Component diagnostics for all macromolecules & ligands
- Depositor also receives detailed XML report
- PDF can be uploaded with manuscript submission to a journal

Overall Quality



Residue Plots



Validation for 3DEM

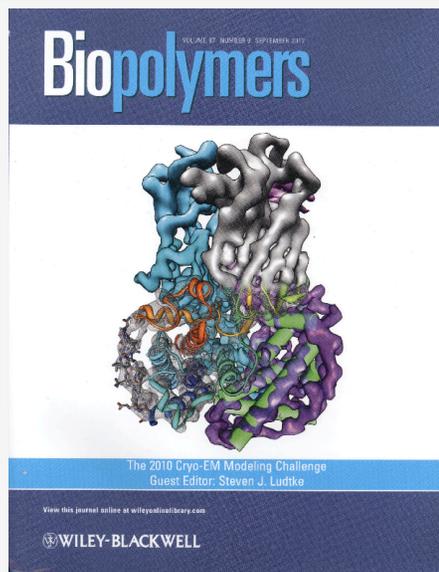


EM Validation Task Force

Henderson *et al.* (2012) ***Structure*** **20**, 205-214

Maps: Standards for assessing resolution and accuracy need to be developed

Models: Criteria needed for model only, fit to map, and fit to additional structural data



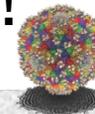
2010 CryoEM Modeling Challenge

Collected papers in a special issue of **Biopolymers** September 2012

- 13 target maps
- 58 participants
- 10 research groups
- 136 submitted models
- 13 software packages

EM VTF Recommendations

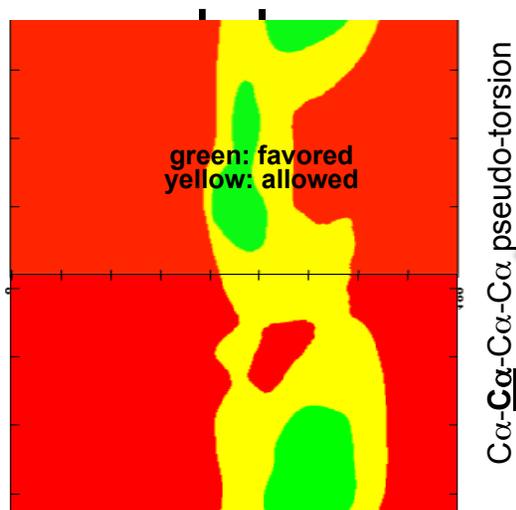
- Main recommendations for EM maps
 - Standards for assessing resolution and accuracy of a map need to be developed
 - Structural features in a map should be in accordance with the claimed resolution
- Main recommendations for models fitted into EM maps
 - Criteria for assessing models need to be developed
 - Capability to archive coarse-grained representations of models is needed
- More research and development needed!



EM Validation Reports

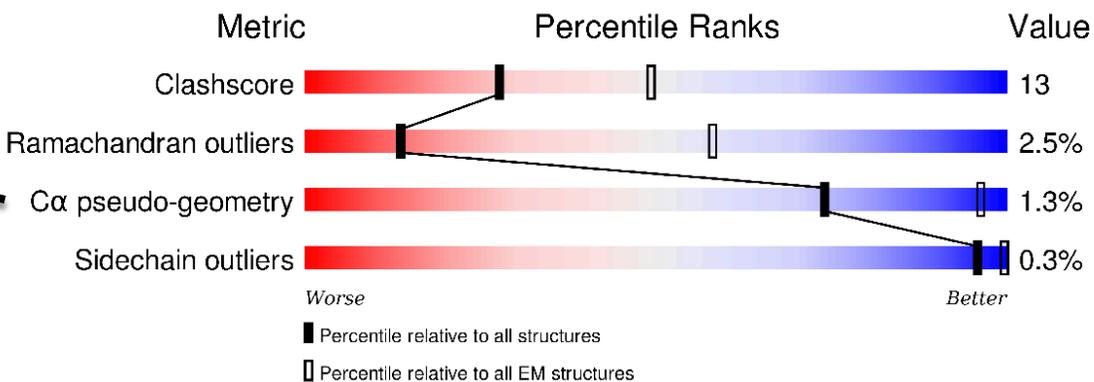
- “Table 1” for EM
- Metrics relevant for EM

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	I	Depositor
Number of images	30000	Depositor
Resolution determination method	FSC 0.143	Depositor
CTF correction method	Not provided	Depositor
Microscope	JEOL 3200FSC	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	Not provided	Depositor
Minimum defocus (nm)	500	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	50000	Depositor
Image detector	DIRECT ELECTRON DE-12 (4k x 3k)	Depositor



Cα-Cα-Cα pseudo-angle

Cα-Cα-Cα pseudo-torsion



2015/2016 Map, Model Challenges

WANTED

Challengers and **Assessors**



FOR MAPS

create/evaluate single particle reconstructions from seven benchmark datasets

FOR MODELS



create/evaluate coordinate models from moderate to high resolution 3DEM reconstructions

Watch EMDatabank News for details

- Each challenge formulated by a community-based committee
- Targets selected from recently deposited maps, models, 2.2-4.5 Å resolution



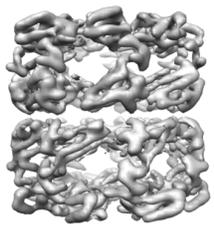
EMDatabank
Unified Data Resource for 3DEM

2015/2016 Map, Model Challenges

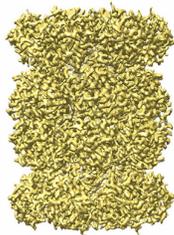
- Goals: Develop benchmarks, encourage development of best practices in 3DEM reconstruction and model fitting, evolve criteria for validation, compare and contrast different approaches
- Results Discussion via Participant Workshops/Journal Special Issues

Benchmark Datasets

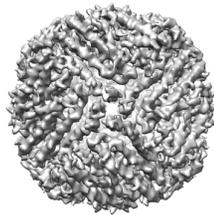
Map Challenge Targets: Raw Images @ EMPIAR



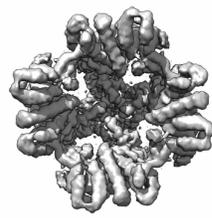
GroEL



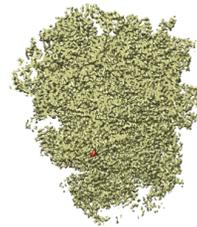
T20S
Proteasome



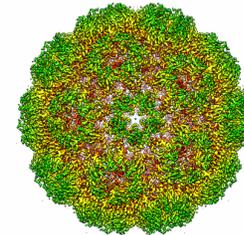
Apo-
Ferritin



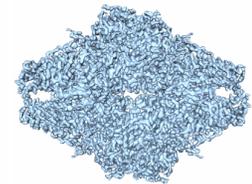
TrpV1
channel



80S
Ribosome

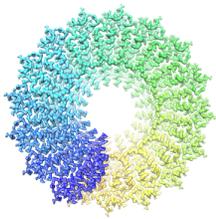


Brome
Mosaic Virus

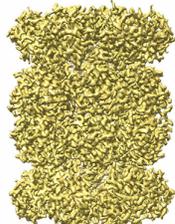


β -galacto-
sidase

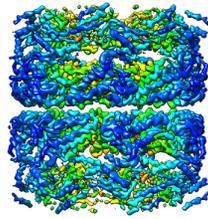
Model Challenge Targets: Maps @ EMDB



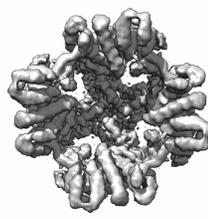
Tobacco
Mosaic Virus



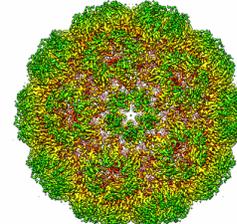
T20S
Proteasome



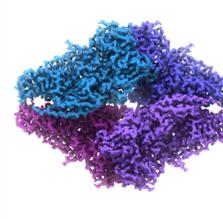
GroEL



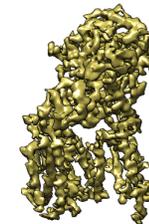
TrpV1
channel



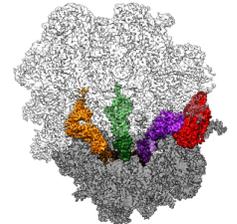
Brome
Mosaic Virus



β -galacto-
sidase



γ -Secretase



70S
Ribosome



Map Challenge

■ **Timing: Registration NOW OPEN**

- Challengers: reconstruction submissions open **August** thru **December**
- Assessors: open data assessment period commences early 2016
- Results Workshop mid-2016

■ **Committee:** Bridget Carragher (Chair), Jose-Maria Carazo, Wen Jiang, John Rubinstein, Peter Rosenthal, Fei Sun, Janet Vonck



Model Challenge

■ **Timing: Registration NOW OPEN**

- Challengers: model submissions open **November 2015** thru **April 2016**

- Assessors: open data assessment period **Summer 2016**

- Results Workshop **Fall 2016**

■ **Committee:** Paul Adams (Chair), Axel Brunger, Randy Read, Torsten Schwede, Maya Topf, Gerard Kleywegt

CPU Availability

Late
Breaking
News

- The San Diego Supercomputing Center (SDSC) has generously offered supercomputing resources to support the ongoing EMDataBank Map Challenge:
 - 1 Million core-hours on SDSC Gordon
 - 20TB of sandbox data storage on Gordon's parallel file system Data Oasis