

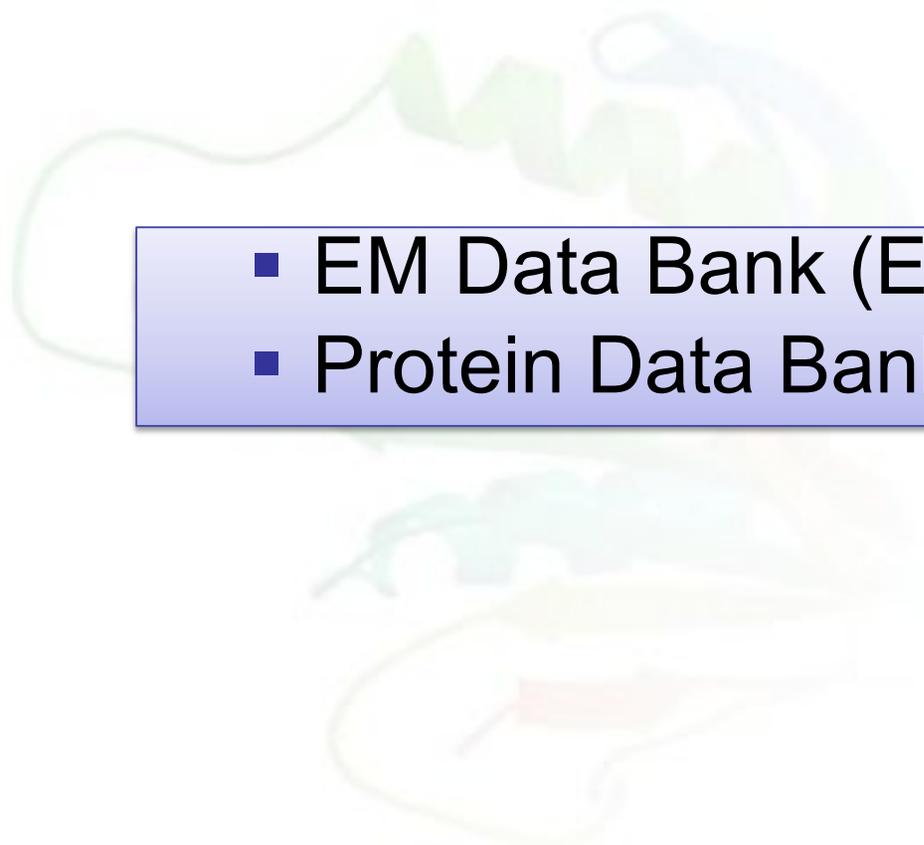
Data Archiving for EM Structures

Cathy Lawson

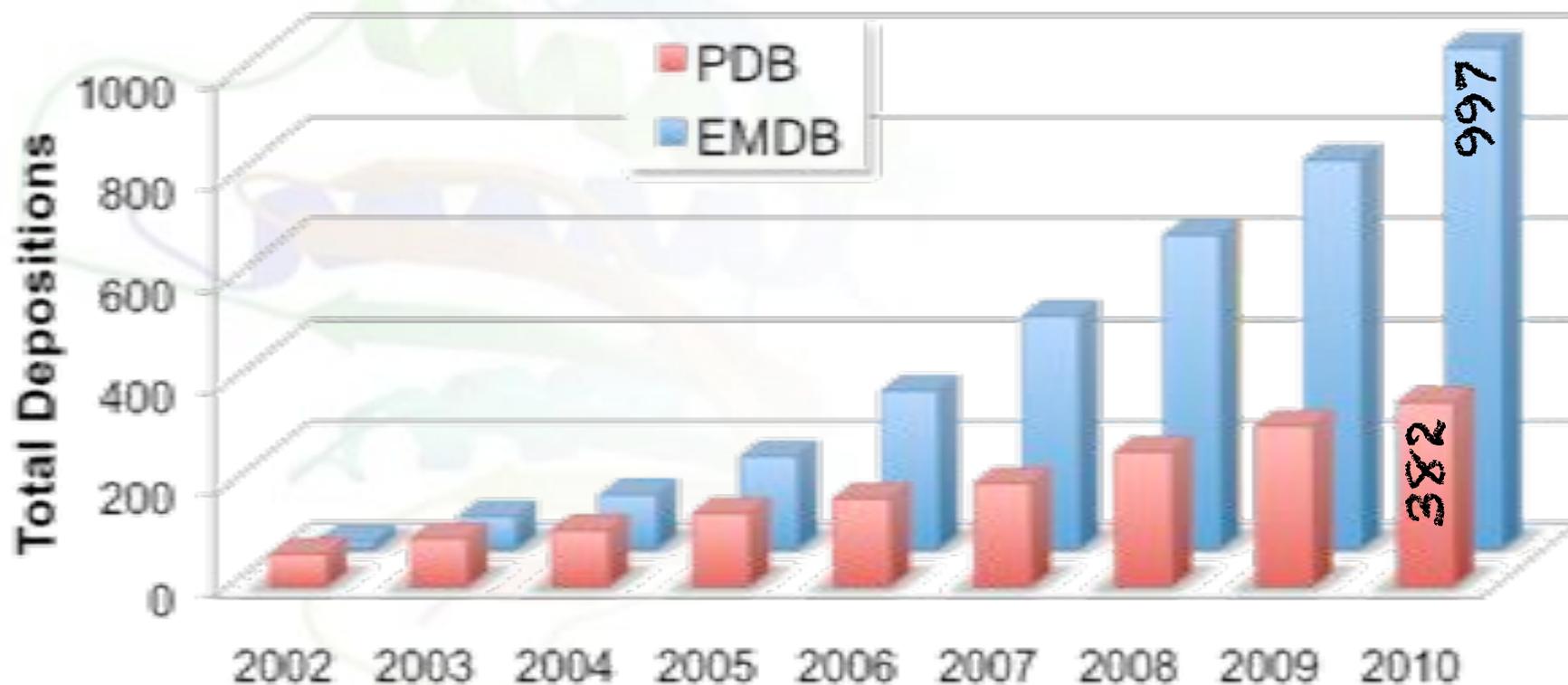
Rutgers University / RCSB PDB

cathy.lawson@rutgers.edu

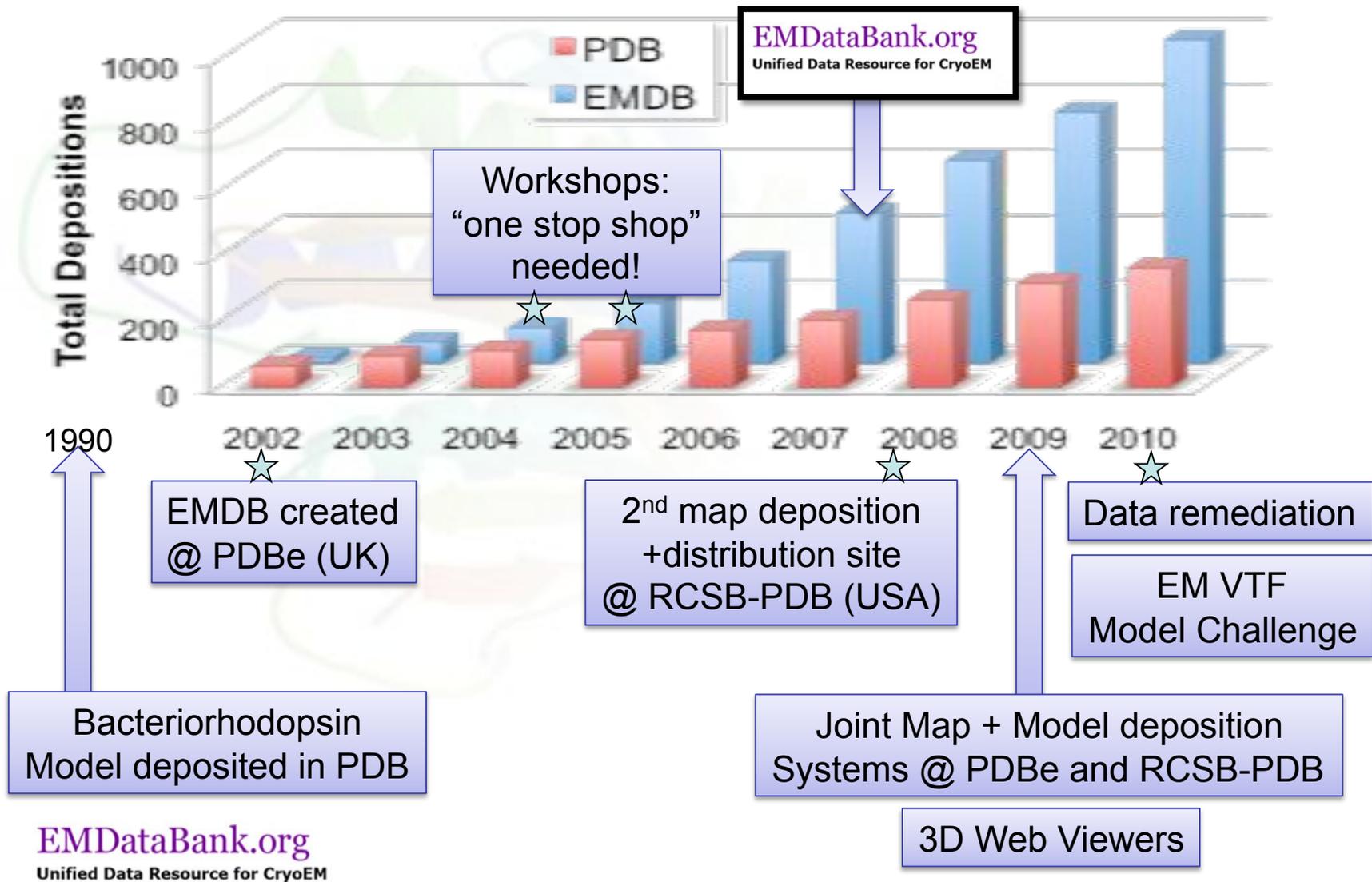
March 16, 2011 NCMI Workshop
UT Houston

- 
- A faded, multi-colored protein structure is visible in the background, rendered in shades of green, blue, yellow, and red. It appears to be a ribbon diagram of a protein molecule.
- EM Data Bank (EMDB): Maps
 - Protein Data Bank (PDB): Models

Growth of EM entries



History



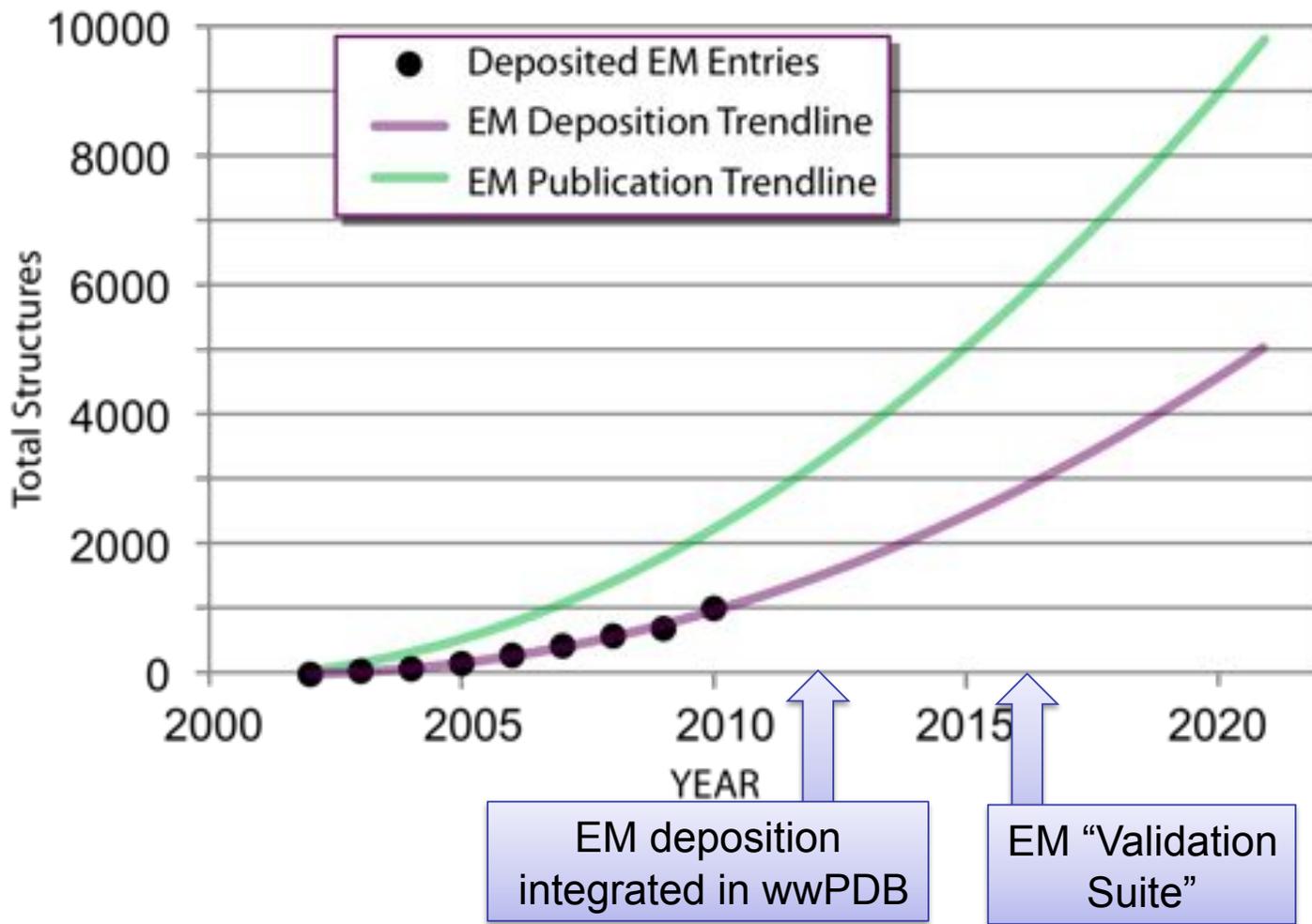
EMDB Reaches 1000 entries!

February 2011: Congratulations are in order to the cryoEM community for reaching a significant milestone: more than 1000 maps representing a wide variety of biological assemblies are now archived in the EM DataBank. **This achievement comes just nine years since EMDb began in 2002** and reflects the rapid growth of cryoEM reconstruction methods and their increasingly wide application in structural biology research. A record 238 maps were deposited in 2010, including ten with reported resolution between 5 and 3.3 Å. **In comparison, it took twenty years for the Protein Data Bank, founded in 1971, to reach 1000 entries.** The increased interest of the community in archiving map volumes resulting from EM reconstruction experiments signals the potential for EMDb to develop into a major structural biology resource over the next decade.



Predicted Growth (maps)

2010: 10^3 entries
2020: 10^4 entries



EMDataBank.org: Unified Data Resource for CryoEM

The screenshot shows the EMDataBank.org website homepage. At the top left is the logo, a colorful sphere of dots, and the text "EMDataBank.org Unified Data Resource for CryoEM". To the right, a tagline reads "One stop shop for cryoEM deposition and retrieval" and a date/status indicator says "2011-03-02: 1012 EMD8 map entries, PDB coordinate entries". A navigation menu includes "Home", "About", "Deposit", "Search", "Tools", "News", "Links", and "Help".

The main content area is divided into three columns. The left column, titled "The Unified Data Resource for Cryo Electron Microscopy", contains a paragraph describing the site as a joint effort of the Protein Data Bank in Europe (PDBe), the Research Collaboratory for Structural Bioinformatics (RCSB), and the National Center for Macromolecular Imaging (NCMI). Below this is a "Recently Released entries" section with three items, each featuring a small image of a protein structure and its title and authors.

The middle column, titled "News", features a headline "EMDB reaches 1000 entries!" and a sub-headline "February 2011: Congratulations are in order to the cryoEM community for reaching a significant milestone: more than 1000 maps representing a wide variety of biological assemblies are now archived in the EM DataBank." Below the text is a small graphic of colored dots and a "Read more" link.

The right column, titled "Announcement - Change in EMD8 Hold Policy for Maps", contains text stating that effective January 2011, the option to hold EM map volumes for two years before release to the public will no longer be available. It also includes a "Read more" link.

At the bottom of the page, there are logos for NCMI, PDBe, and PDB, along with the copyright notice "© 2011 EMDataBank.org" and a "Supported by:" logo for NC2011.

The Unified Data Resource for Cryo Electron Microscopy

EMDataBank.org is a joint effort of the Protein Data Bank in Europe (PDBe), the Research Collaboratory for Structural Bioinformatics (RCSB), and the National Center for Macromolecular Imaging (NCMI) to create a global deposition and retrieval network for cryoEM map, model and associated metadata, as well as a portal for software tools for standardized map format conversion, map, segmentation and model assessment, visualization, and data integration.

Recently Released entries

1823 [PDBe | RCSB] **2010-11-18** RELEASED ON 2011-02-15 singleParticle 35Å
Molecular Architecture and Structural Transitions of a Clostridium thermocellum Mini-Cellulosome.
García-Alvarez B, Melero R, Dias FM, Prates JA, Fontes CM, Smith SP, Romão MJ, Carvalho AL, Llorca O

5255 [PDBe | RCSB] **2010-12-15** RELEASED ON 2011-02-15 Icosahedral 24Å
Seeing the portal in herpes simplex virus type 1 B capsids.
Rochat RH, Liu X, Murata K, Nagayama K, Rixon FJ, Chiu W

1873 [PDBe | RCSB] **2011-02-09** RELEASED ON 2011-02-09 singleParticle
Domain movements of HAP2 in the cap-filament complex formation and growth process of the bacterial flagellum.
Maki-Yonekura S, Yonekura K, Namba K

[View All Recent Entries](#)

News

EMDB reaches 1000 entries!

February 2011: Congratulations are in order to the cryoEM community for reaching a significant milestone: more than 1000 maps representing a wide variety of biological assemblies are now archived in the EM DataBank.

[Read more](#)

Announcement - Change in EMD8 Hold Policy for Maps

Effective January 2011, the option to hold EM map volumes for two years before release to the public will no longer be available.

[Read more](#)

[Read All News](#)

NCMI PDBe PDB © 2011 EMDataBank.org Supported by: NC2011

Portal to Deposit and Search Services

Deposit EM Maps

Deposit 3D EM map reconstructions to the EM Databank:

- USA: <http://emdep.rutgers.edu/>
- UK: <http://www.ebi.ac.uk/pdbe-emdep/emdep/>

Search EMDB

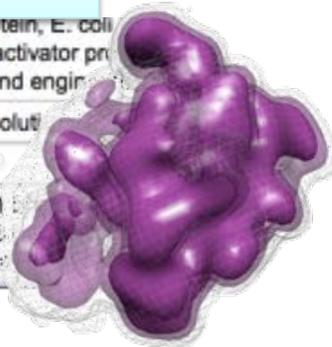
Browse/Search the EMDB:

- USA: <http://emsearch.rutgers.edu/>
- UK: <http://www.ebi.ac.uk/msd-srv/emsearch/>
- EMDB Full Text Search (alpha):
<http://www.ebi.ac.uk/emdb-srv/search/?advsearch=1>

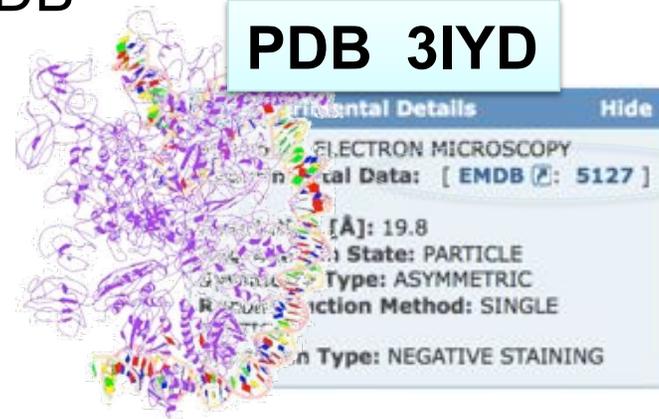
Links maintained between
EMDB and PDB

EMD-5127

Source:	Natural protein, E. coli catabolite activator protein promoter and engineering
Resolution:	19.8 Å (resolution)
Fitted PDB id:	3iyd
Oligomeric Details:	Assembly 1: holoenzyme



PDB 3IYD



Visualization

Java (web) viewers:

- EM Viewer (map only)
- OpenAstexViewer (map +model)

Programmatic Access:

- SOAP-based web service (implemented in Chimera)

The image displays three different visualization tools for cryo-EM data. The top left shows 'EM Viewer' with a green 3D map. The top right shows 'Astex Viewer' with a 3D map and a ribbon model, accompanied by a control panel for map and model settings. The bottom section shows the 'UCSF Chimera -- EMDatabank.org Web Service' interface, which includes a 3D map, a search panel for selecting databases and structures, and a table of search results.

ID	Sample Name	Reference	Resolution	Release Date	Fit PDBs
5001	Native unliganded GroEL residual ADP	De Novo Backbone Trace of GroEL from Single Particle Electron Cryomicroscopy Ludtke SJ, Baker ML, Chen DH, Song JL, Chuang DT, Chiu W	4.2	2008-02-12	3cau

EM Map → Model Joint Deposition

The screenshot shows the EmDep Deposition Tool interface. The top navigation bar includes "EMDB home > depositions > EmDep", "EmDep Deposition Tool", and "Contact EmDep". The left sidebar contains a list of steps: "Initialisation", "Instruction to EmDep", "File Deposition", "Title & Author", "Reference", "Sample Details", "Sample Components", "Experimental Details", "Verification", "Imaging", "Image Processing", "Reconstruction", "Fitting", "Map Details", "Submit", and "Deposit Coordinates at the PDB". The "Deposit Coordinates at the PDB" link is circled in blue. The main content area displays "Submission of data" with a policy notice, "Submission Successful" with a thank-you message and contact information, and "EMDB submission policy" with details on data release schedules and public availability. A "Close Session" button is at the bottom left, and "Done" is at the bottom left of the window.

EMDB home > depositions > EmDep EmDep Deposition Tool Contact EmDep

Deposition reference id:
20019

EMDB Accession issued: 3004

[Initialisation](#)
[Instruction to EmDep](#)
[File Deposition](#)
[Title & Author](#)
[Reference](#)
[Sample Details](#)
[Sample Components](#)
[Experimental Details](#)
[Verification](#)
[Imaging](#)
[Image Processing](#)
[Reconstruction](#)
[Fitting](#)
[Map Details](#)
[Submit](#)
[Deposit Coordinates at the PDB](#)

Submission of data

Please read the submission policy. You must agree to this policy before you can complete the submission of this entry to the EMDB.

Submission Successful

Thank you for depositing your structure to EMDB, you will receive an id code via email shortly. Your deposition will now be processed by the EMDB curation team who will get back to you shortly with their comments and annotated EMDB entry.

If you do not receive an email with your EMDB id code please get in touch with the EMDB team at emdep@rcsb.rutgers.edu.

As part of the new version of EmDep you are now able to begin a PDB deposition from this deposition session. Click the "Deposit Coordinates at the PDB" link in the sidebar to proceed with deposition to the PDB.

EMDB submission policy

By submitting the entry to the EMDB, you are agreeing that your entry may be loaded on the EMDB servers.

This will be done following completion of the submission process, at the time of publication (if you chose the option "release on publication"), or when your entry comes off hold.

Please note, the EMDB release set is updated weekly, scheduled for Wednesday mornings shortly after midnight Eastern Standard Time. Files become available from all official EMDB mirror sites within a day or so. Entries received Wednesday through Sunday are loaded on the EMDB servers the following Wednesday. Entries received Monday and Tuesday are loaded on Wednesday of the next week. Entries on hold will be released with the load following the hold release date according to the above schedule.

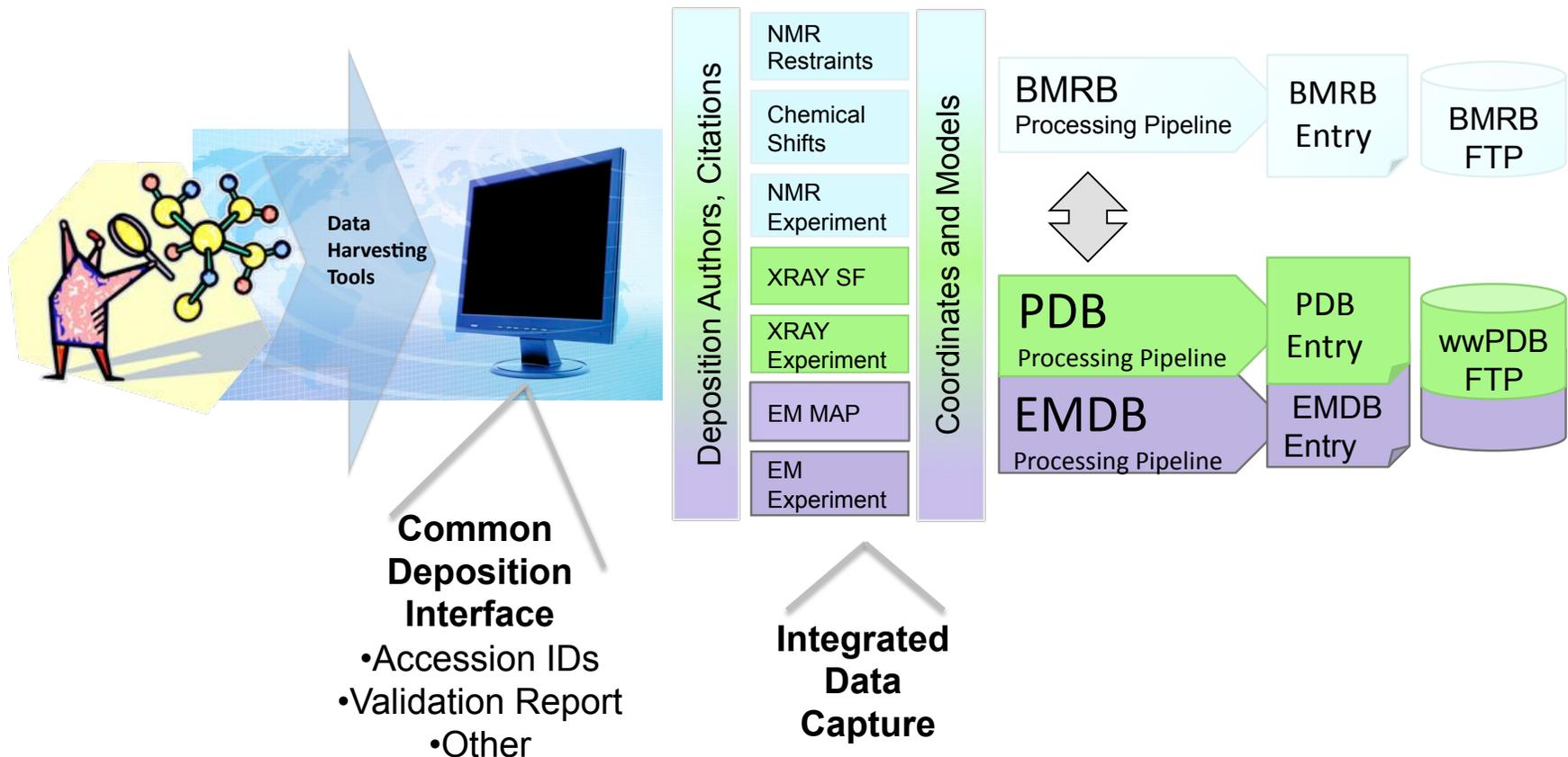
The EMDB strongly encourages authors to make their data available to the public as soon as possible. In the event that you wish to delay release of your data, our Release Policy appears below.

EMDB Release Policy:

Close Session
Done

wwPDB Common Deposition & Annotation Tool

In the works: single interface for deposition and processing of all structural biology data, including EM



Content of EMDB entries

- EM map (ccp4 format)
- Experimental details (xml)
 - sample information
 - specimen preparation
 - imaging parameters
 - reconstruction details
 - fitting details
- Rarely: FSC resolution curves, images, layer lines or structure factors, map slices, masks

Content of PDB entries for EM structures

- Atomic coordinate models fitted to EM maps
- Matrix representations and parameters for point and helical symmetries
- Structure factors
- Experimental details
 - sample information
 - specimen preparation
 - imaging parameters
 - reconstruction details
 - fitting details

Annotation: What is checked?

- Metadata checked for consistency
- EMDB map:
 - converted to redistribution format (CCP4)
 - visually compared to original uploaded map (spider, mrc, ccp4)
 - Map scale, map origin, author-recommended contour level, map statistical profile, hand
- PDB model:
 - Polymer Chemistry
 - Geometrical checks for inter and intra molecular features
 - Biological assembly
- Visual inspection of model fit in map

Some Issues

- CCP4 map format not ideal for all circumstances
- Minimal map, model-fit-to-map validation
- Geometry checks may fail because of modeling procedures used at low resolution
- Number of atoms and chains larger than can be accommodated in current PDB format

Open Questions

- What other data should be collected?
- Are there data that should not be collected?
- How should the maps be validated?
- How should the models be validated?
- Who should make the recommendations?

September 2010 EM Validation Task Force

Co-Chairs:

Richard Henderson
Andrej Sali

Charge:

Advise EMDDataBank.org
on approaches to validate
maps deposited to EMDB
and models deposited to
PDB that are obtained
from cryoEM data



Project Team & Funding



**Baylor College of
Medicine**

Wah Chiu, PI
Matthew Baker
Ian Rees
Greg Pintilie
Steven Ludtke
Matt Dougherty

Rice University

Powei Feng
Joe Warren



Rutgers University

Helen Berman, co-PI
Catherine Lawson
Batsal Devkota
Raul Sala
John Westbrook

Chunxiao Bi



**European Bioinformatics
Institute**

Gerard Kleywegt, co-PI
Arjan Patwardhan
Eduardo Sanz Garcia
Ingvar Lagerstedt
Gaurav Sahni
Glen van Ginkel
Tom Oldfield

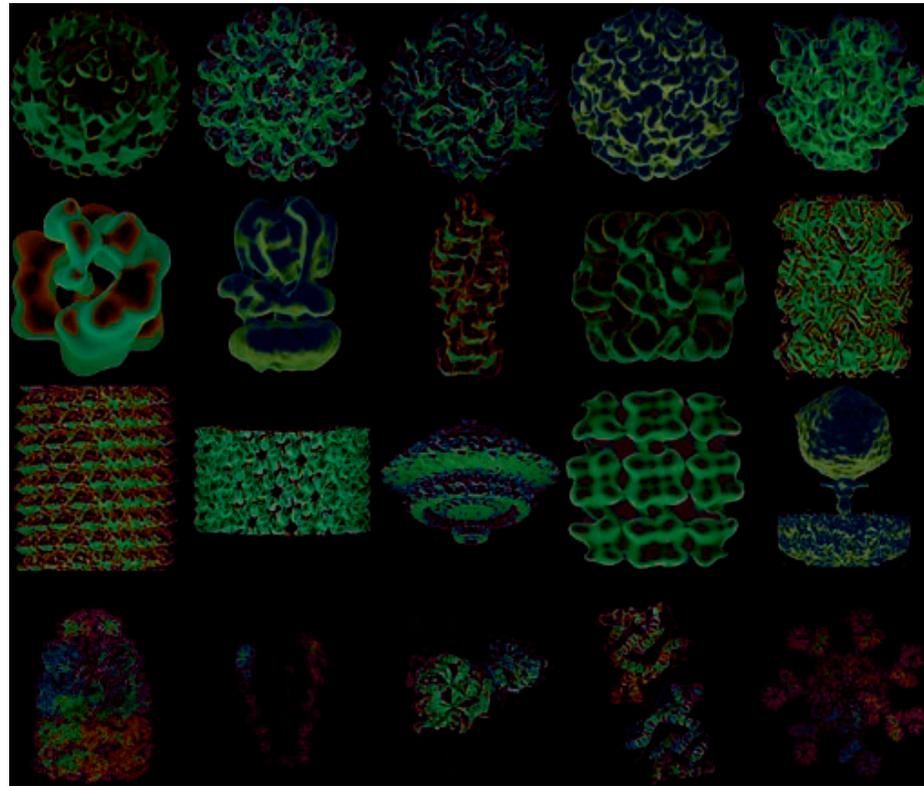
Kim Henrick
Richard Newman
Christoph Best
Antonio Suarez
Sameer Velankar

PRINT ISSN: 0305-1048
ONLINE ISSN: 1362-4962

Nucleic Acids Research

VOLUME 39 DATABASE ISSUE JANUARY 1, 2011

www.nar.oxfordjournals.org



EMDataBank.org
Unified Data Resource for CryoEM

OXFORD
UNIVERSITY PRESS

Open Access

No barriers to access – all articles freely available online

 **Mixed Sources**
Product group from well-managed
forests and other controlled sources
www.fsc.org Cert. no. DNW-COC-000079
© 1996 Forest Stewardship Council

EM Deposition & Retrieval Tutorial and Guide

Workshop participants can view a completed deposition and can create practice depositions at the demo site: <http://emdep-staging.rutgers.edu>

See separate Tutorial/Guide (2nd link for this talk in the interactive workshop agenda)