Data Archiving for EM Structures

Cathy Lawson

Rutgers University / RCSB PDB
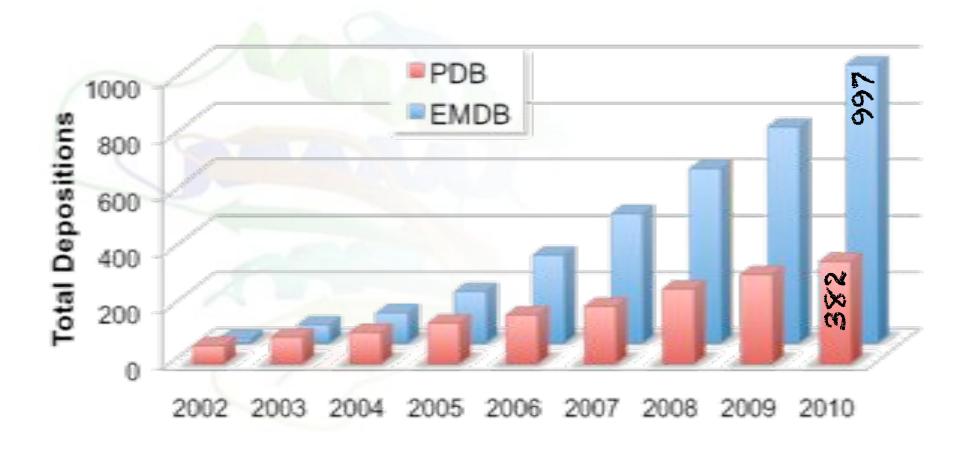
cathy.lawson@rutgers.edu

March 16, 2011 NCMI Workshop
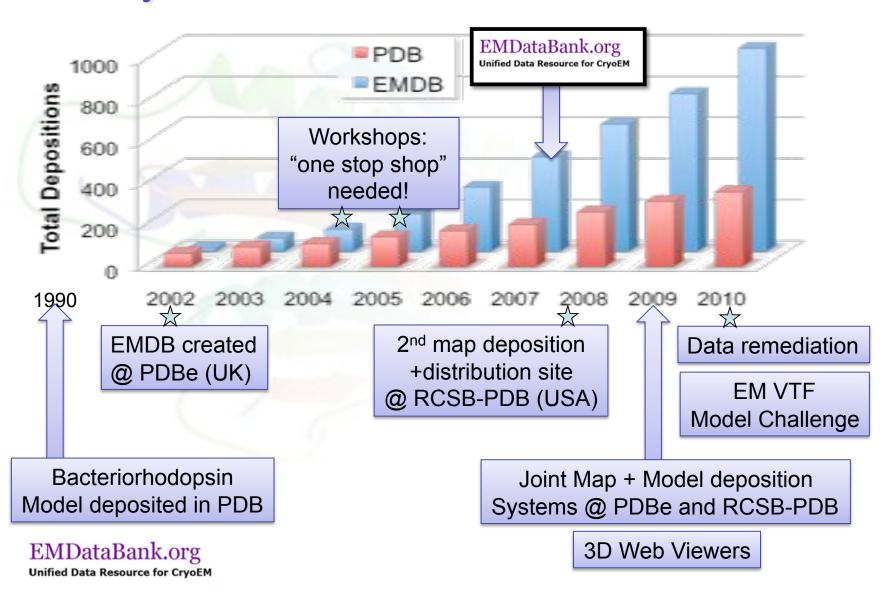UT Houston

EMDataBank.org
Unified Data Resource for CryoEM

- EM Data Bank (EMDB): Maps
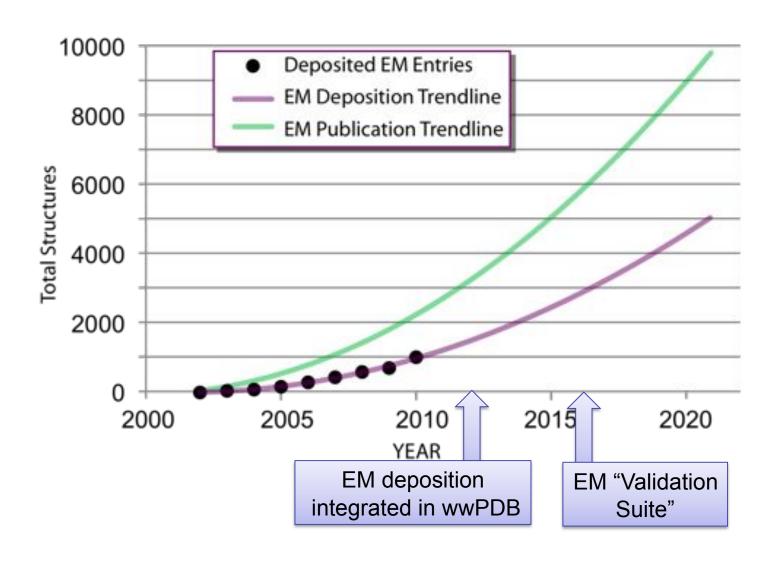- Protein Data Bank (PDB): Models

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

Legend:
- ● Deposited EM Entries
- — EM Deposition Trendline
- — EM Publication Trendline

Y-axis: Total Structures (0, 2000, 4000, 6000, 8000, 10000)
X-axis: YEAR (2000, 2005, 2010, 2015, 2020)

EM deposition integrated in wwPDB

EM "Validation Suite"

EMDataBank.org: Unified Data Resource for CryoEM

Portal to Deposit and Search Services

Deposit EM Maps

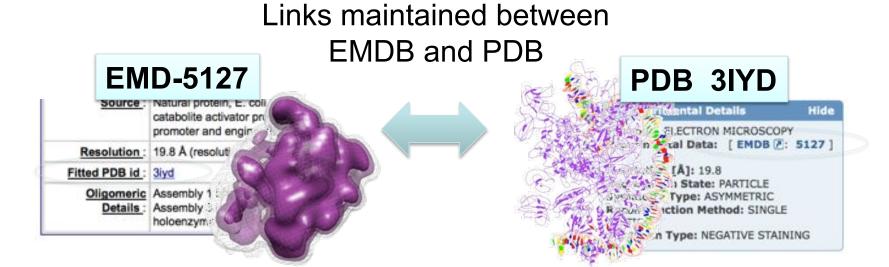*Deposit 3D EM map reconstructions to the EM Databank:*

- 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

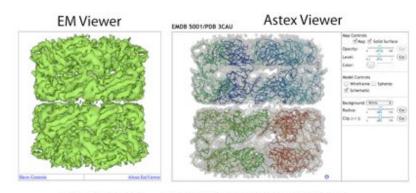Links maintained between
EMDB and PDB

EMD-5127

Source : Natural protein, E. coli
catabolite activator pr
promoter and engin

Resolution : 19.8 Å (resolut

Fitted PDB id : 3iyd

Oligomeric   Assembly 1
Details :   Assembly
holoenzym

PDB 3IYD

rimental Details   Hide

ELECTRON MICROSCOPY
tal Data: [EMDB : 5127]
[Å]: 19.8
State: PARTICLE
Type: ASYMMETRIC
ction Method: SINGLE
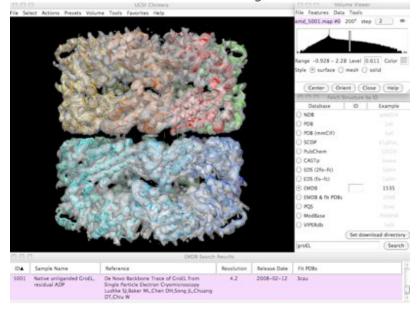Type: NEGATIVE STAINING

Visualization

Java (web) viewers:

– EM Viewer (map only)

– OpenAstexViewer (map +model)

Programmatic Access:

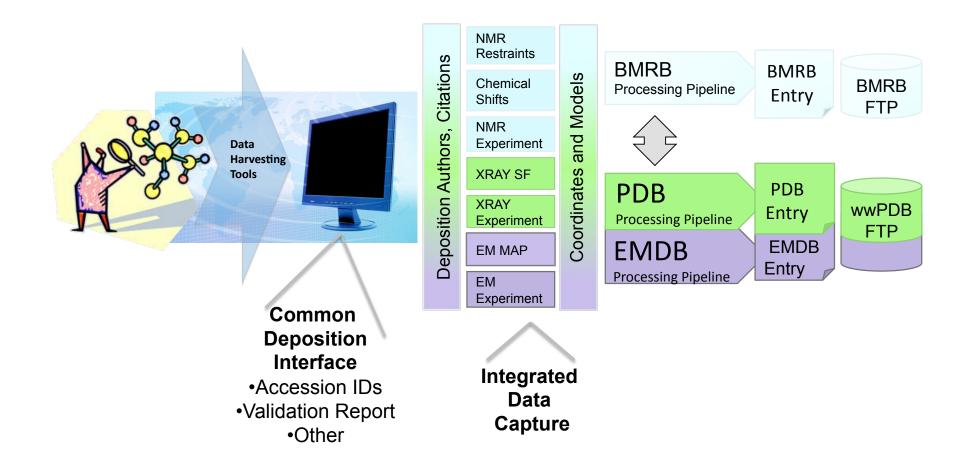– SOAP-based web service (implemented in Chimera)

EM Map → Model Joint Deposition

wwPDB Common Deposition & Annotation Tool

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

Data Harvesting Tools

Common Deposition Interface
- Accession IDs
- Validation Report
- Other

Deposition Authors, Citations

NMR Restraints

Chemical Shifts

NMR Experiment

XRAY SF

XRAY Experiment

EM MAP

EM Experiment

Coordinates and Models

Integrated Data Capture

BMRB Processing Pipeline

BMRB Entry

BMRB FTP

PDB Processing Pipeline

PDB Entry

EMDB Processing Pipeline

EMDB Entry

wwPDB FTP

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

EMDataBank.org
Unified Data Resource for CryoEM

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?

EMDataBank.org
Unified Data Resource for CryoEM

September 2010
EM Validation Task Force

Co-Chairs:

Richard Henderson

Andrej Sali

Charge:

Advise EMDataBank.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
Ardan Patwardhan
Eduardo Sanz Garcia
Ingvar Lagerstedt
Gaurav Sahni
Glen van Ginkel
Tom Oldfield

Kim Henrick
Richard Newman
Christoph Best
Antonio Suarez
Sameer Velankar

EMDataBank.org
Unified Data Resource for CryoEM

Funding: National Institute of General Medical Sciences   BBSRC bioscience for the future

Nucleic Acids Research

EMDataBank.org
Unified Data Resource for CryoEM

OXFORD
UNIVERSITY PRESS

Open Access
No barriers to access – all articles freely available online

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)