Deposition and Retrieval of Cryo-EM Data

RCSB PDB
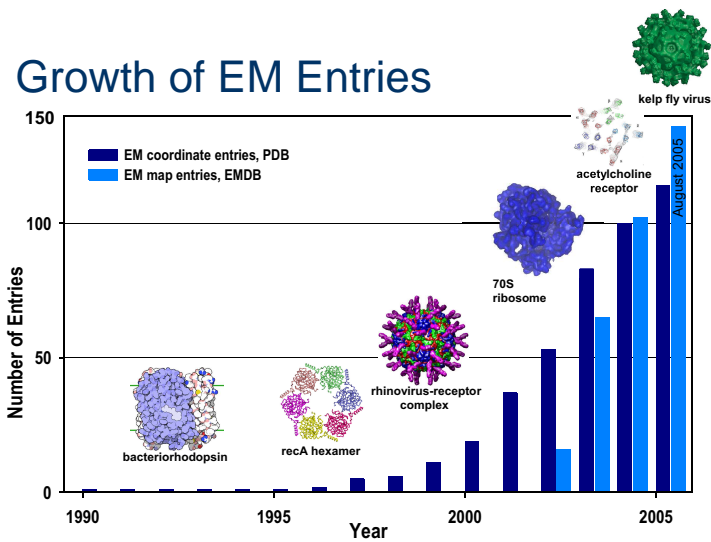PROTEIN DATA BANK
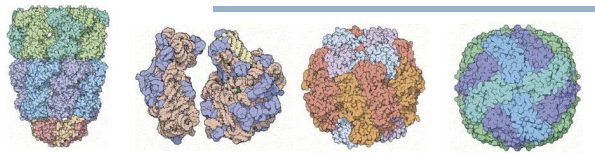
cathy.lawson@rutgers.edu
November 9, 2005
NRAMM, TSRI

Growth of Coordinate Entries

Growth of EM Entries

- EM coordinate entries, PDB
- EM map entries, EMDB

Labels: kelp fly virus, acetylcholine receptor, 70S ribosome, rhinovirus-receptor complex, recA hexamer, bacteriorhodopsin, August 2005

Y-axis: Number of Entries
X-axis: Year
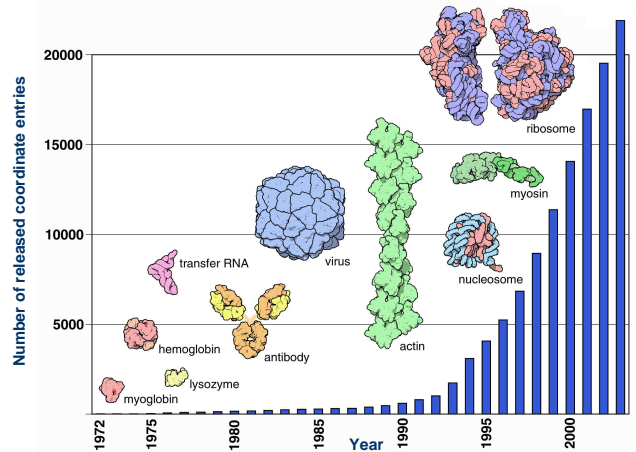
How to deposit/retrieve EM data

RCSB PDB — **EMBL-EBI** — **PDBj** Protein Data Bank Japan

- PDB Archive @wwPDB centers
  - Coordinates
  - Structure Factors
  - Information about the experiments (meta data)

- EM Database @MSD-EBI
  - Maps +
  - Slices, Masks, Structure Factors, Layerlines, Images, Fourier Shell Correlation Curve
  - Information about the experiments (meta data)
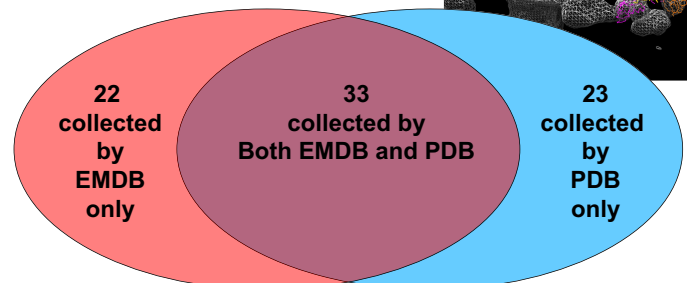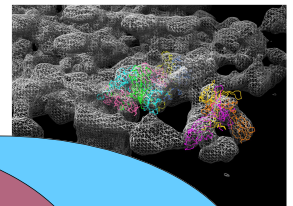

Two deposition steps?

I could DO more experiments
if it were easier to
ARCHIVE them…

Archiving Meta Data

Example:
- T4 bacteriophage baseplate
- PDB 1TJA, EMD 1086

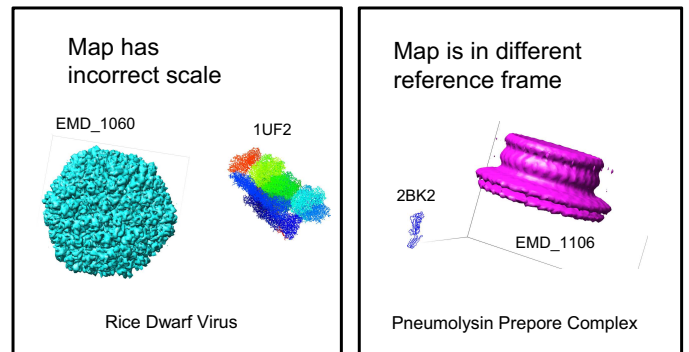| 22 collected by EMDB only | 33 collected by Both EMDB and PDB | 23 collected by PDB only |

Total EM-related meta data items = 78

Two retrieval steps?

How can I look at the EM map AND coordinates of this molecule together?

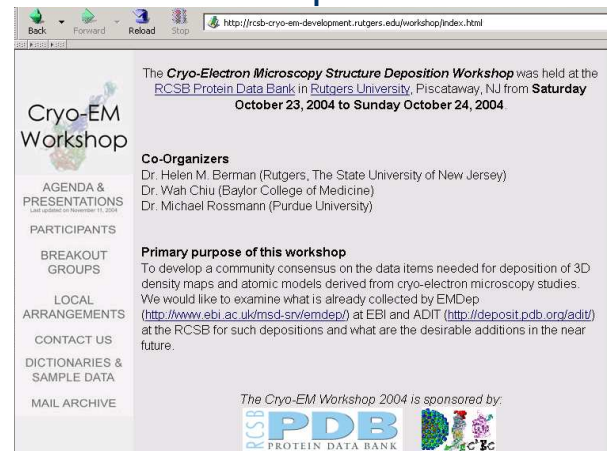## Typical EM data Viewing Problems

Map has incorrect scale

EMD_1060   1UF2

Rice Dwarf Virus

Map is in different reference frame

2BK2   EMD_1106

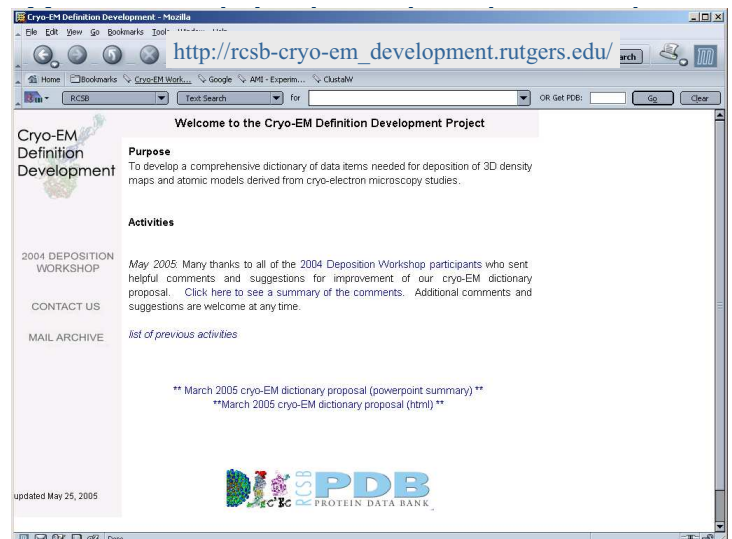Pneumolysin Prepore Complex

Improving the situation

- Cryo-EM Deposition Workshop @ Rutgers, 2004
- Develop comprehensive meta data dictionary
- Create "One-stop-shop" for deposition/retrieval of cryo-EM data

Oct 2004 Workshop

Cryo-EM Workshop

AGENDA & PRESENTATIONS
PARTICIPANTS
BREAKOUT GROUPS
LOCAL ARRANGEMENTS
CONTACT US
DICTIONARIES & SAMPLE DATA
MAIL ARCHIVE

The *Cryo-Electron Microscopy Structure Deposition Workshop* was held at the RCSB Protein Data Bank in Rutgers University, Piscataway, NJ from **Saturday October 23, 2004 to Sunday October 24, 2004**.

Co-Organizers
Dr. Helen M. Berman (Rutgers, The State University of New Jersey)
Dr. Wah Chiu (Baylor College of Medicine)
Dr. Michael Rossmann (Purdue University)

Primary purpose of this workshop
To develop a community consensus on the data items needed for deposition of 3D density maps and atomic models derived from cryo-electron microscopy studies. We would like to examine what is already collected by EMDep (http://www.ebi.ac.uk/msd-srv/emdep/) at EBI and ADIT (http://deposit.pdb.org/adit/) at the RCSB for such depositions and what are the desirable additions in the near future.

The Cryo-EM Workshop 2004 is sponsored by:

March 2005 Cryo-EM Dictionary

Sample Description
em_assembly
em_entity_assembly
em_entity_assembly_list
em_virus_entity
em_icos_virus_shells
em_single_particle
em_filaments
em_2d_crystal

Biochemical Preparation
em_sample_preparation
em_sample_support
em_array_formation
em_solution_composition

EM Specimen Preparation
em_vitrification
em_stain
em_cryo_stain
em_embedding_agent

EM Data Collection
em_imaging
em_detector
em_image_scans
em_microscope
em_micrographs
em_electron_diffraction
em_electron_diffraction_phase
em_electron_diffraction_pattern

Image Processing
em_singleparticle_selection
em_3d_reconstruction
em_particle_picking
em_particle_picking_list
em_filament_selection
em_filament_reconstruction

Structure Analysis
em_3d_fitting
em_3d_fitting_list
em_classes
em_refinement
em_fsc_curve

New categories recommended at the Oct 2004 workshop are in pink

(Cryo-EM Definition Development)

http://rcsb-cryo-em_development.rutgers.edu/

Cryo-EM Definition Development

Welcome to the Cryo-EM Definition Development Project

Purpose
To develop a comprehensive dictionary of data items needed for deposition of 3D density maps and atomic models derived from cryo-electron microscopy studies.

Activities

May 2005: Many thanks to all of the 2004 Deposition Workshop participants who sent helpful comments and suggestions for improvement of our cryo-EM dictionary proposal. Click here to see a summary of the comments. Additional comments and suggestions are welcome at any time.

list of previous activities

2004 DEPOSITION WORKSHOP
CONTACT US
MAIL ARCHIVE

** March 2005 cryo-EM dictionary proposal (powerpoint summary) **
**March 2005 cryo-EM dictionary proposal (html) **

updated May 25, 2005

cryoEM Dictionary Examples

| TYPE	EM_SAMPLE	EM_ENTITY	EM_ASSEMBLY(s)/MAP(s)
single particle	GroEL	<-- GroEL-->	GroEL
	Rice Dwarf Virus	• <--Rice Dwarf Virus--> • outer shell (T=13) • inner shell (T=1)	Rice Dwarf Virus
	Poliovirus-Receptor Complex	• <--Poliovirus-Receptor Complex--> • poliovirus • CD155 receptor	Poliovirus-Receptor Complex
	Pneumolysin-Liposome Complex	• <--Pneumolysin-Liposome cpx • Pneumolysin pore--> • Liposome	Pneumolysin pore
	T4 bacteriophage	• <-- T4 bacteriophage • head (5-fold symmetry)--> • tail (6-fold symmetry)--> • portal • baseplate • other components	• head • tail
helicalV filament	helical crystals, acetylcholine receptor	• <--helical crystal • acetylcholine receptor-->	acetylcholine receptor
	helical crystals, HIV CA-NC	• <--helical crystal--> • HIV CA-NC	helical crystal, HIV CA-NC
	tobacco mosaic virus	• <--tobacco mosaic virus--> • outer coat • nucleic acid	tobacco mosaic virus
2D crystal	2D crystals, aquaporin	• <--2d crystal • aquaporin tetramer -->	aquaporin tetramer
tomogram	embedded mouse skin slice	• <--embedded mouse skin slice • desmosome network-->	desmosome network
	spirochete whole cell	• <--spirochete whole cell--> • ribosome • proteasome • other components...	spirochete whole cell

---

Cryo-EM Dictionary

• 521 data items
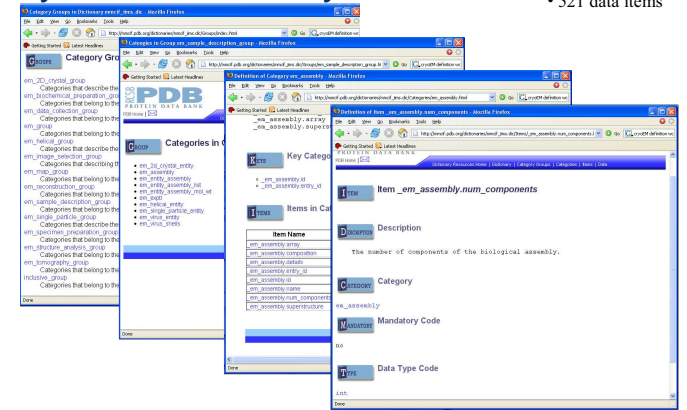
http://mmcif.pdb.org/dictionaries/mmcif_em.dic/Index

EM data representation issues

- Coordinate Format
- Symmetry
- Visualization
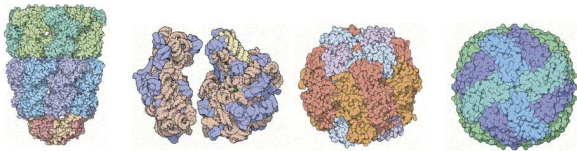
Coordinate Format

- PDB format
  - maximum of 99,999 atoms, 62 chains
  - larger structures represented in multiple files

- mmCIF/PDBML formats
  - no restrictions on size
  - mmCIF recognized by many crystallography applications
  - use is strongly encouraged for current/future software applications

http://mmcif.rcsb.org/, http://pdbml.rcsb.org/

Format Examples

| mmCIF	PDBML

loop_
_atom_site.id
_atom_site.label_atom_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_seq_id
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
1 O5* G A 1 -3.897 61.994 -24.841
2 C5* G A 1 -5.016 62.932 -24.76
	<PDBx:atom_site id="2168">
	<PDBx:group_PDB>ATOM</PDBx:group_PDB>
	<PDBx:type_symbol>C</PDBx:type_symbol>
	<PDBx:label_atom_id>C</PDBx:label_atom_id>
	   <PDBx:label_alt_id xsi:nil="true" />
	
	<PDBx:label_comp_id>PRO</PDBx:label_comp_id>
	  <PDBx:label_asym_id>F</PDBx:label_asym_id>
	
	<PDBx:label_entity_id>2</PDBx:label_entity_id>
	  <PDBx:label_seq_id>5</PDBx:label_seq_id>
	<PDBx:Cartn_x>-9.306</PDBx:Cartn_x>
	<PDBx:Cartn_y>-17.809</PDBx:Cartn_y>
	<PDBx:Cartn_z>14.947</PDBx:Cartn_z>
	<PDBx:occupancy>0.50</PDBx:occupancy>
	<PDBx:B_iso_or_equiv>25.40</PDBx:B_iso_or_equiv>
	   <PDBx:auth_seq_id>6</PDBx:auth_seq_id>
	  <PDBx:auth_comp_id>PRO</PDBx:auth_comp_id>
	   <PDBx:auth_asym_id>B</PDBx:auth_asym_id>
	   <PDBx:auth_atom_id>C</PDBx:auth_atom_id>
	
	<PDBx:pdbx_PDB_model_num>1</PDBx:pdbx_PDB_model_num>
	 </PDBx:atom_site>

Symmetry

Asymmetric Unit → Biological Unit

- Non-trivial problem to provide a correct set of transformations and a procedure for applying them
- We are investigating ways to better standardize this process
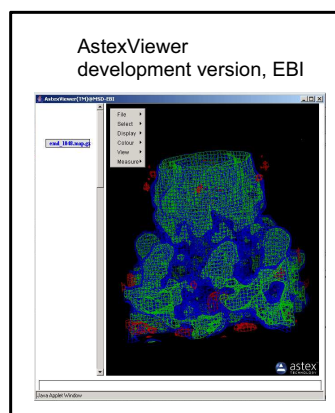- Full biological units are available from the RCSB-PDB

**Rice
Dwarf
Virus**

biological unit

asymmetric unit
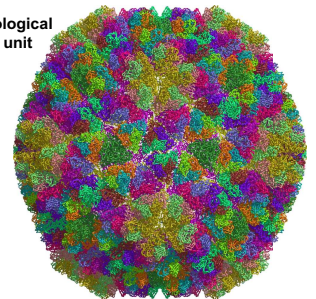
Visualization

For Non-experts:
- free software
- multiple platforms
- no browser dependence
- easy to install and use
- user-friendly interface

AstexViewer
development version, EBI

Acknowledgements