PDBe
PROTEIN DATA BANK EUROPE

OMERO: A resource for electron microscopy

Ingvar Lagerstedt

EMBL-EBI

EBI is an Outstation of the European Molecular Biology Laboratory.

Contents

1. Overview of single-particle electron microscopy and electron tomography
2. What does EMDataBank.org do?
3. Work that has been completed
4. Work in progress
5. Ideas for the future

Single-particle EM

- Projection images of macromolecules

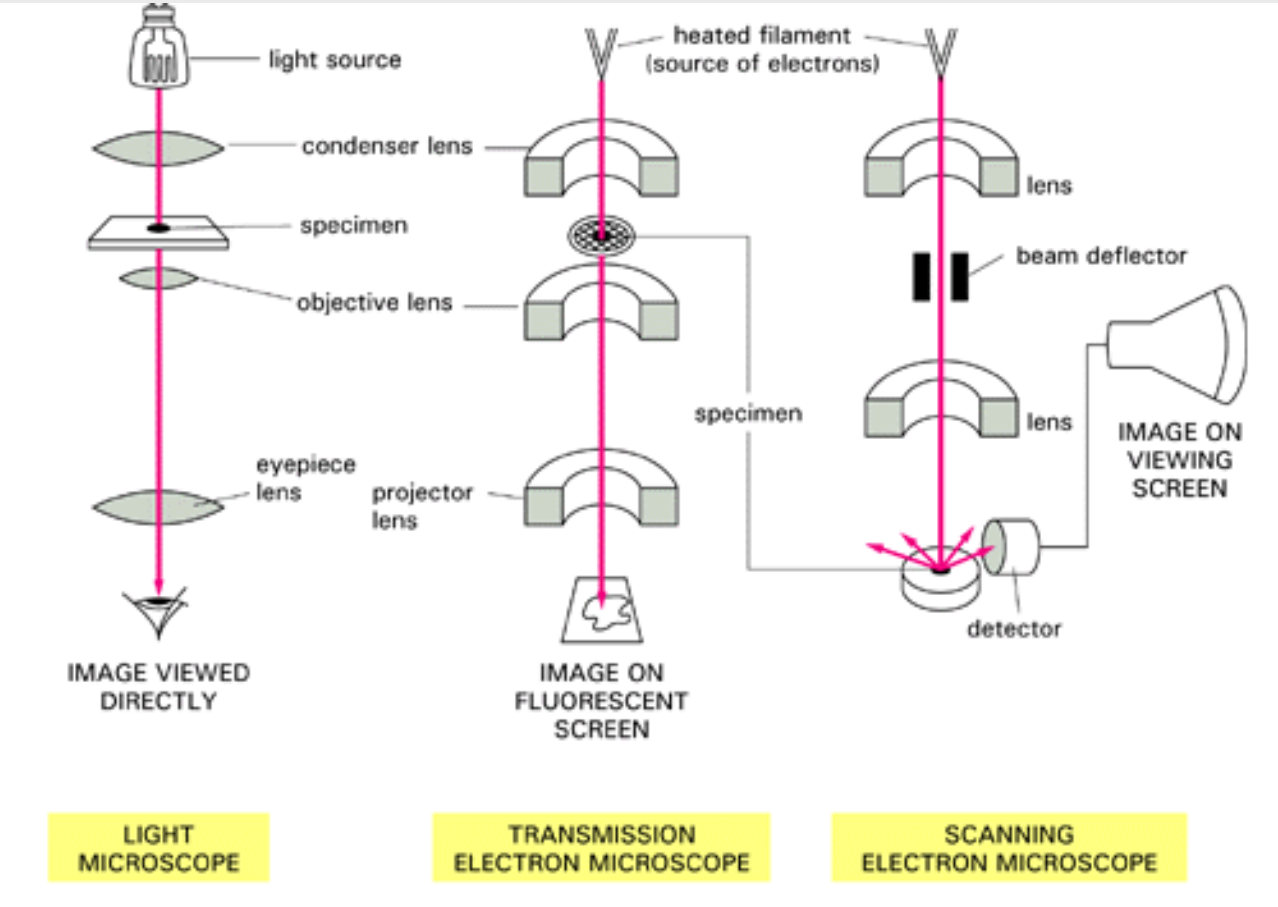
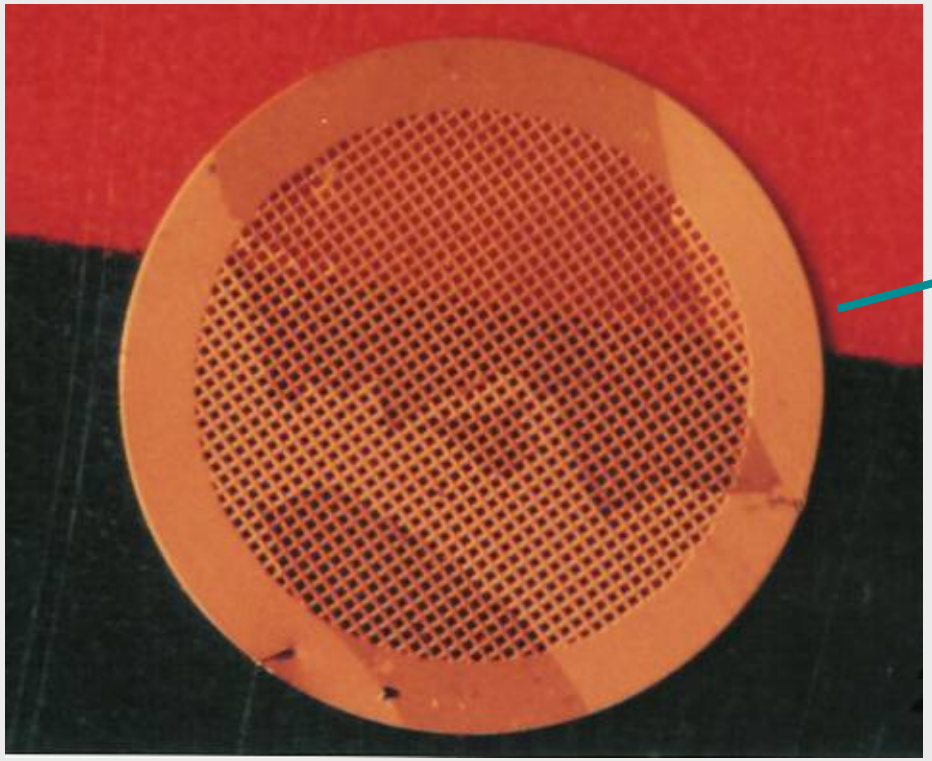


Image taken from the microscopy course at the Univ. of Basel / Biozentrum

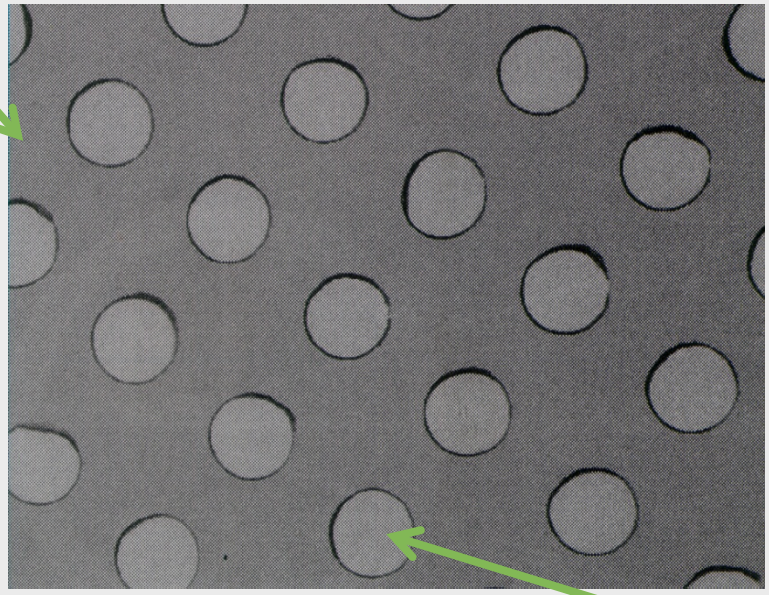
Cryo EM sample preparation

- Small drop of solution of purified macromolecule (e.g. protein complex) put on grid
- Grid blotted to form thin film
- Snap frozen



Holey carbon support film: Quantifoil

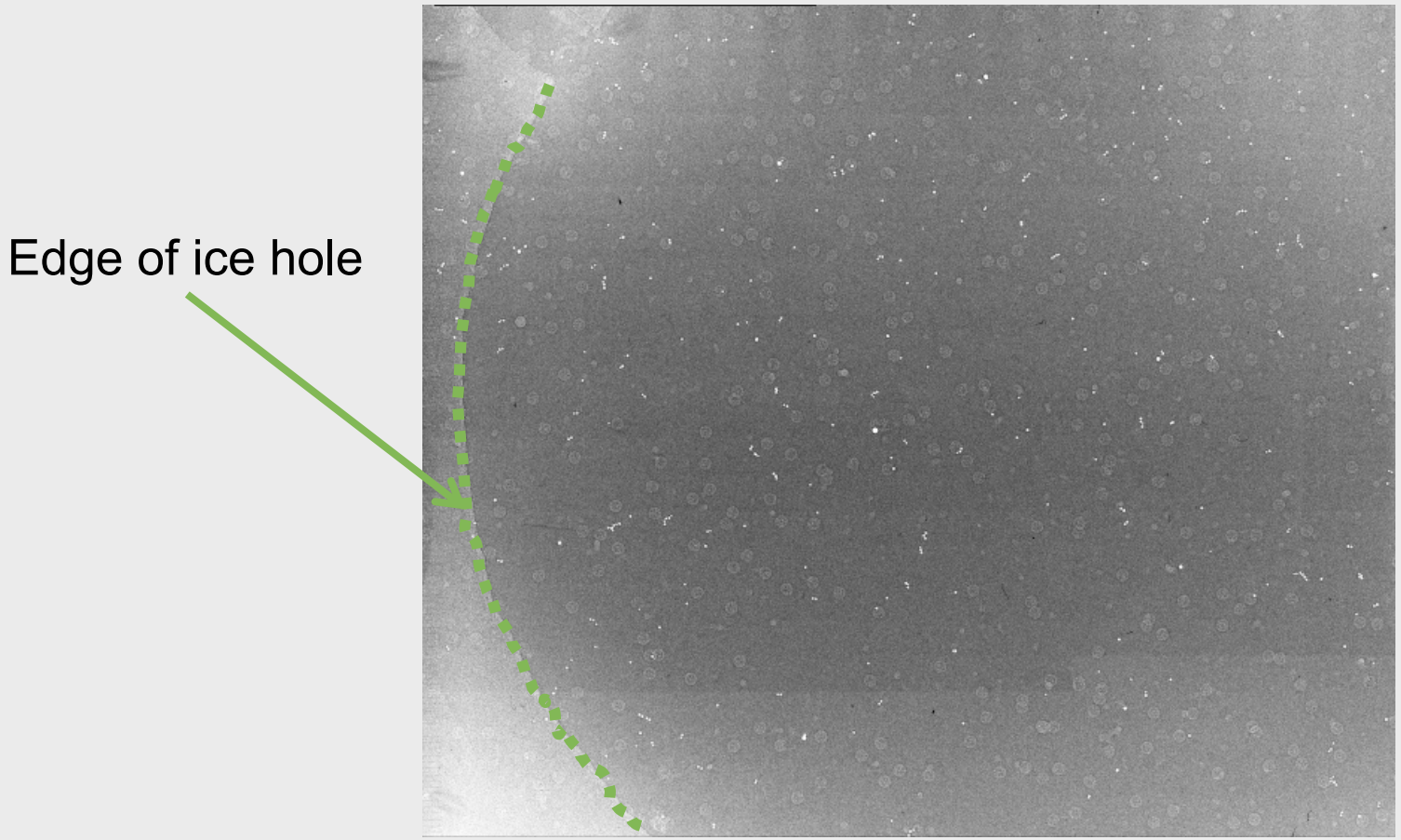
Carbon support film



1μm holes with thin layer of amorphous ice

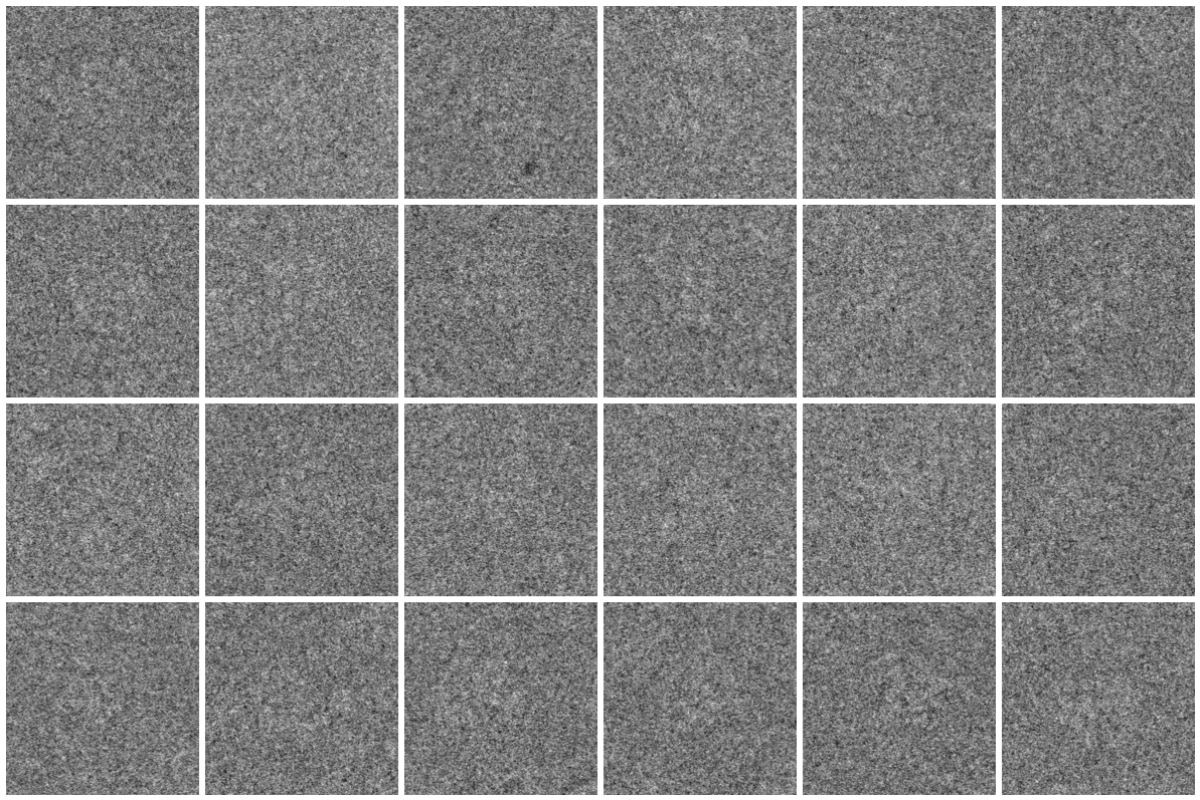
An ice hole

- Particles are randomly positioned and oriented

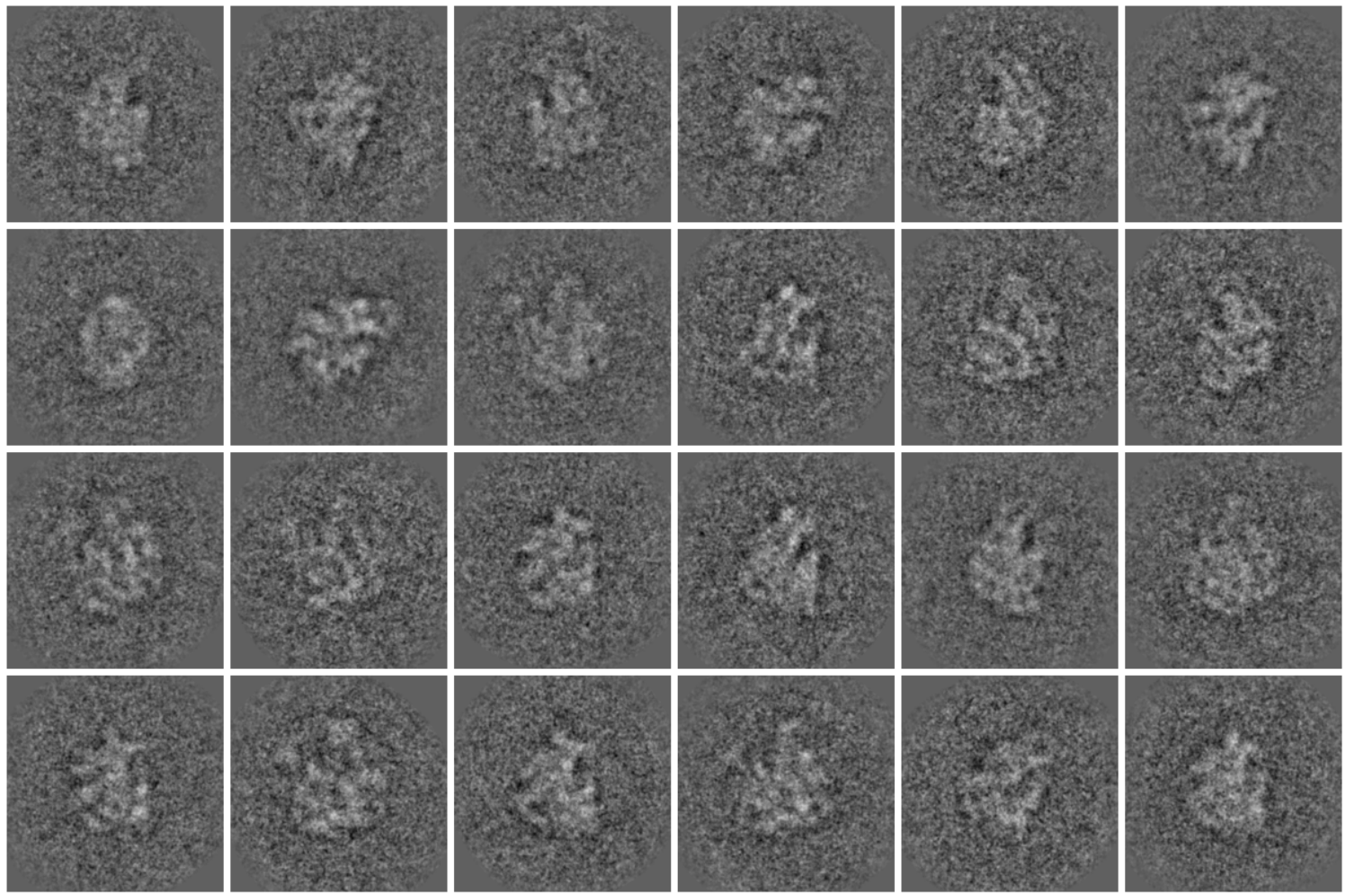


Single particle image processing problem

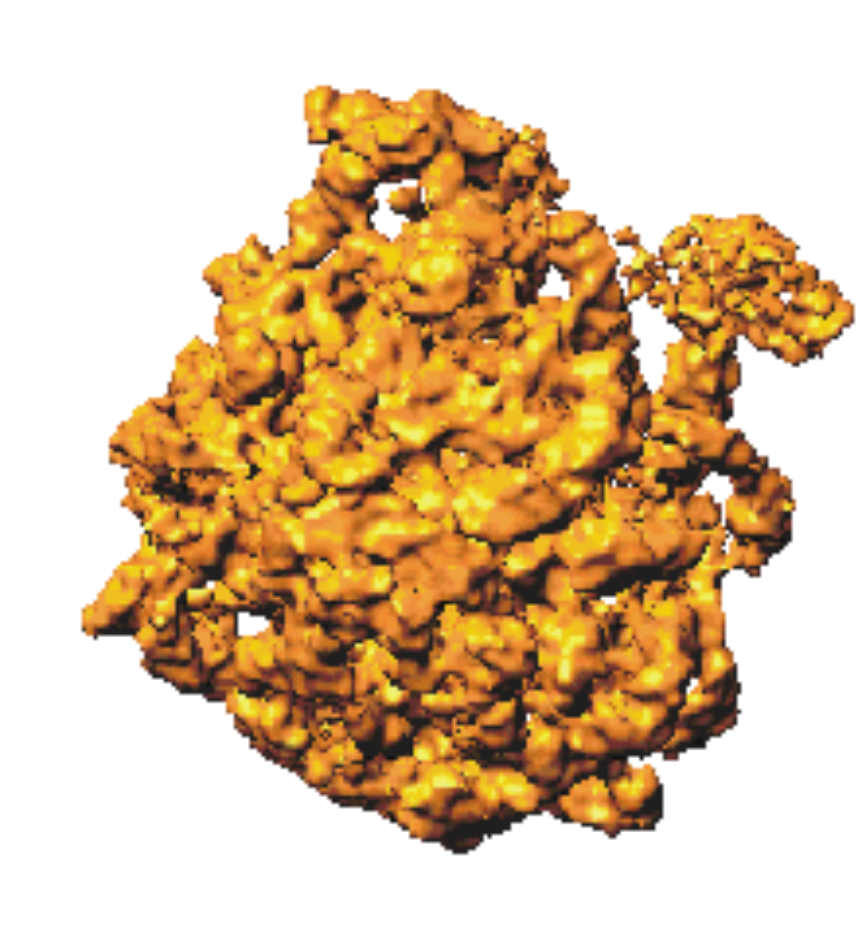
- Combine the information from a large number of random projections of a particle, in order to generate a 3D reconstruction of it
- Pick particles! Example images of 50S ribosomal subunits of *E. coli*.



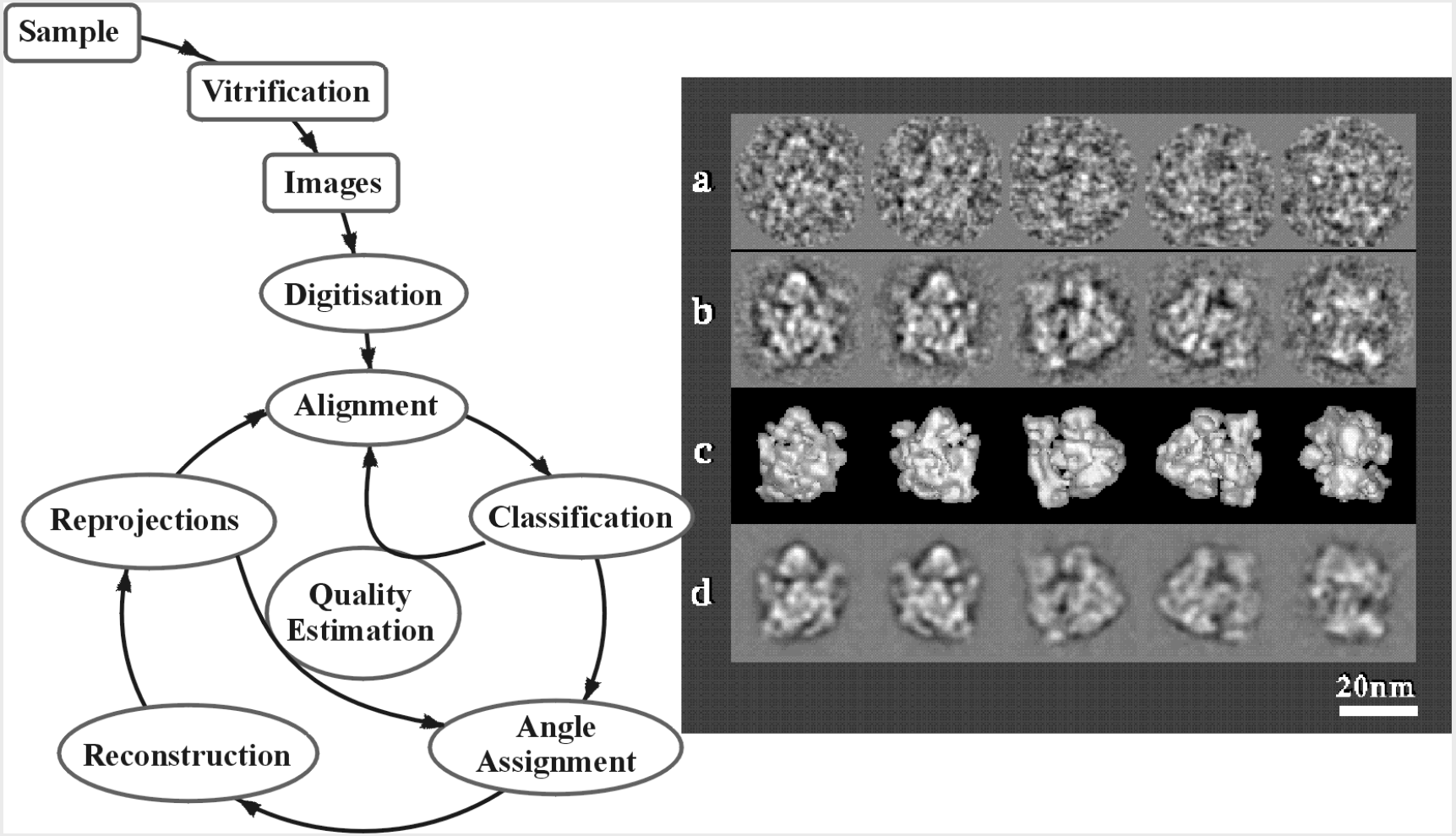
Classify and average images to improve signal to noise ratio



Calculate angles and reconstruct the molecule



Repeat *ad nauseam*...

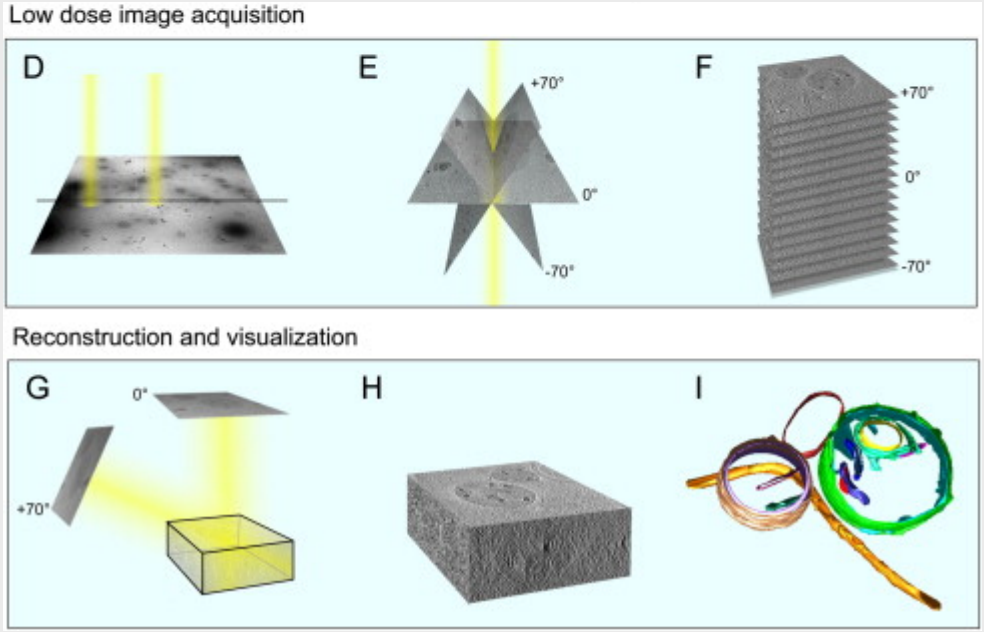


Single-particle specific issues

- No unified or consistent way of describing the workflow involved in obtaining final reconstruction
- How well individual steps can be tracked is down to individual note-taking!
- Example image: real densities 8 bytes/voxel, 512^2= 2MB per image
- Data set: 100K – 1M images - 200GB+ dataset
- Typical final reconstructions may involve 10 – 100 iterations
- Easily 1M+ metadata entries to track the whole process

Electron Tomography

- The specimen is tilted incrementally and a series of images are acquired
- 3D reconstruction of a specimen instead of an ensemble of macromolecules
- Highest resolution ~2nm
- Figure from: Koning RI, Koster AJ, Cryo-electron tomography in biology and medicine, *Ann Anat* **191** (2009) 427-445



Tomography specific issues

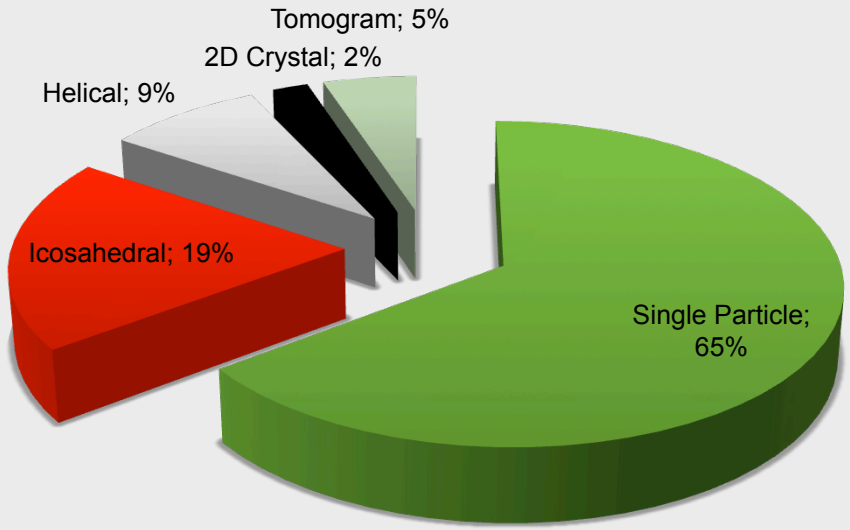
- Example image – 4k x 4k x 8 bpp= 128 MB per image
- Tilt series: 100 – 200 images – 10 GB+ raw data sets
- Reconstruction: 500GB
- What is biologically interesting are the segmentations associated with the tomograms
- Segmentations can be stored efficiently as bitmaps, surfaces or vectors

EMDataBank.org

- EMDB (Electron Microscopy Data Bank) was started at EBI in 2002 as a repository of 3D EM maps of biologically relevant macromolecules
- Since 2008 EMDB is run jointly by the PDBe, RCSB and NCMI as a part of EMDataBank.org
- Aim: Create a global “one-stop shop” for the deposition and retrieval of cryoEM map, model and associated metadata, and to provide a portal for software tools for standardized map format conversion, map segmentation and model assessment, visualization, and data integration
- We now have over 1000 released maps in the EMDB
- Total depositions are expected to rise 5-10 fold by 2020

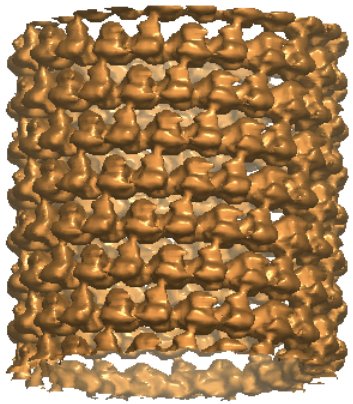
EMDB map distribution

- Currently most entries are single-particle data
- Tomography is expected to grow in prominence in the future



Examples – Map only

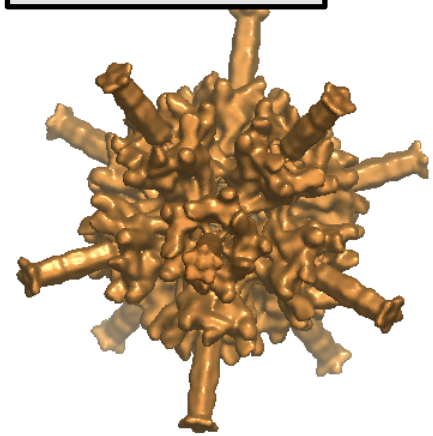
HELICAL



GDP tubulin
EMD-1129 12 Å

Wang, Nogales (2005)

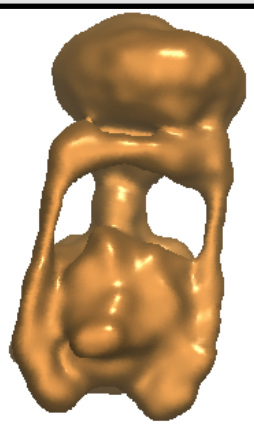
ICOSAHEDRAL



Human Adenovirus type 3 EMD-1179 16.5 Å

Fuschiotti, Schoen, Fender, Fabry, Hewat, Chroboczek, Ruigrok, Conway (2005)

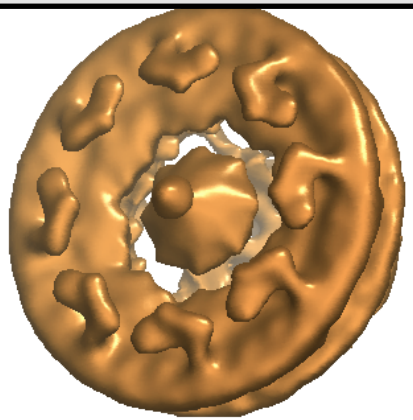
SINGLE PARTICLE



V/A ATPase Thermus Thermophilus EMD-1888 16 Å

Lau, Rubinstein (2011)

TOMOGRAM AVERAGE

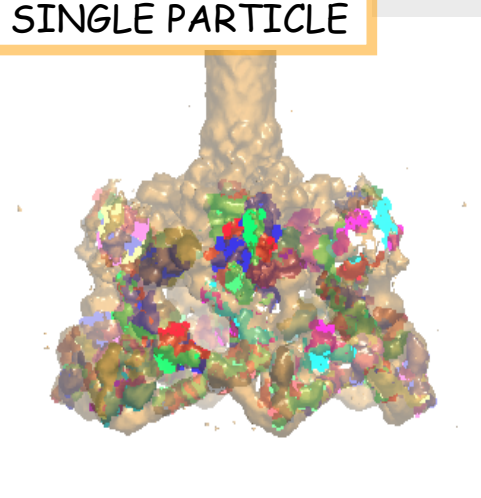


Nuclear pore complex EMD-1097 85 Å

Beck, Forster, Ecke, Plitzko, Melchior, Gerisch, Baumeister, Medalia (2004)

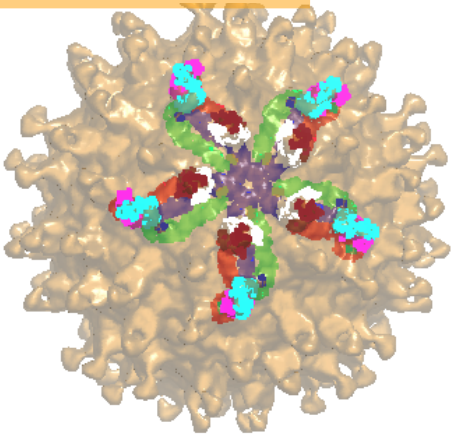
Examples – Map + Model

SINGLE PARTICLE



T4 phage baseplate
EMD-1048 12 Å
PDB: 1pdf 1pdi 1pdj 1pdl 1pdm 1pdp 2fl8
Kostyuchenko, Leiman, Chipman, Kanamaru, van Raaij, Arisaka, Mesyanzhinov, Rossmann (2003)

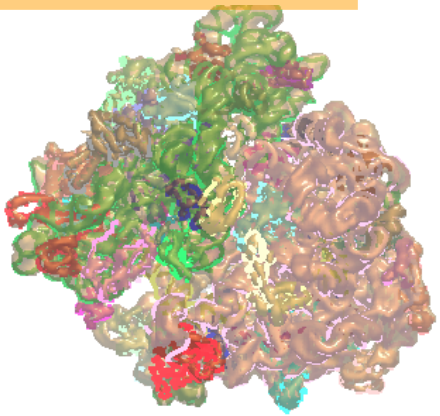
ICOSAHEDRAL



West Nile virus NY99
EMD-5190 13.7 Å
PDB: 3iyw

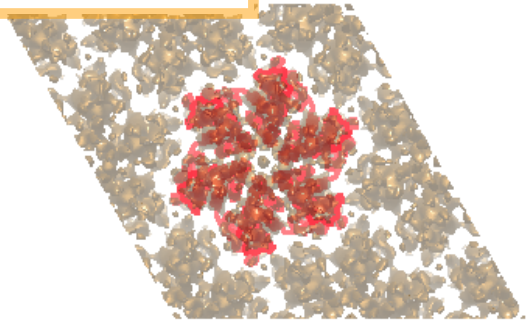
Kaufmann, Vogt, Goudsmit, Holdaway, Aksyuk, Chipman, Kuhn, Diamond, Rossman, (2010)

SINGLE PARTICLE



70S E. coli ribosome
EMD-5036 6.7 Å
PDB: 3fih 3fik
Villa, Sengupta, Trabuco, LeBarron, Baxter, Shaikh, Grassuci, Nissen, Ehrenberg, Schulten, Frank (2009)

2D CRYSTAL



HIV-1 CA R18L
EMD-1529 9 Å

Ganser-Pornillos, Yeager, Sundquist(2009)

PDBe PROTEIN DATA BANK EUROPE

EMBL-EBI

EMDB metadata

- Specified using the XML dictionary: ftp://ftp.ebi.ac.uk/pub/databases/emdb/doc/XML-schema/emdb_v1_5.xsd
- A snippet!

```
- <xs:complexType name="imgType">
  - <xs:all>
      <xs:element name="astigmatism" type="xs:string" minOccurs="0" />
      <xs:element name="electronSource" type="eSourceType" />
      <xs:element name="electronDose" type="eDoseType" minOccurs="0" />
      <xs:element name="energyFilter" type="xs:string" minOccurs="0" />
      <xs:element name="imagingMode" type="imgModeType" />
      <xs:element name="nominalDefocusMin" type="defocusType" minOccurs="0" />
      <xs:element name="nominalDefocusMax" type="defocusType" minOccurs="0" />
      <xs:element name="illuminationMode" type="illumType" />
      <xs:element name="specimenHolder" type="xs:string" />
      <xs:element name="details" type="xs:string" minOccurs="0" />
      <xs:element name="detector" type="xs:string" minOccurs="0" />
      <xs:element name="nominalCs" type="csType" minOccurs="0" />
      <xs:element name="tiltAngleMin" type="tiltType" minOccurs="0" />
      <xs:element name="calibratedMagnification" type="xs:float" minOccurs="0" />
      <xs:element name="tiltAngleMax" type="tiltType" minOccurs="0" />
      <xs:element name="temperature" type="tempType" minOccurs="0" />
      <xs:element name="temperatureMin" type="tempType" minOccurs="0" />
      <xs:element name="temperatureMax" type="tempType" minOccurs="0" />
      <xs:element name="microscope" type="xs:string" />
      <xs:element name="date" type="xs:string" minOccurs="0" />
      <xs:element name="specimenHolderModel" type="xs:string" />
      <xs:element name="acceleratingVoltage" type="accVoltType" minOccurs="0" />
      <xs:element name="nominalMagnification" type="xs:float" minOccurs="0" />
      <xs:element name="energyWindow" type="eWindowType" minOccurs="0" />
      <xs:element name="detectorDistance" type="xs:string" minOccurs="0" />
      <xs:element name="electronBeamTiltParams" type="xs:string" minOccurs="0" />
    </xs:all>
  </xs:complexType>
```

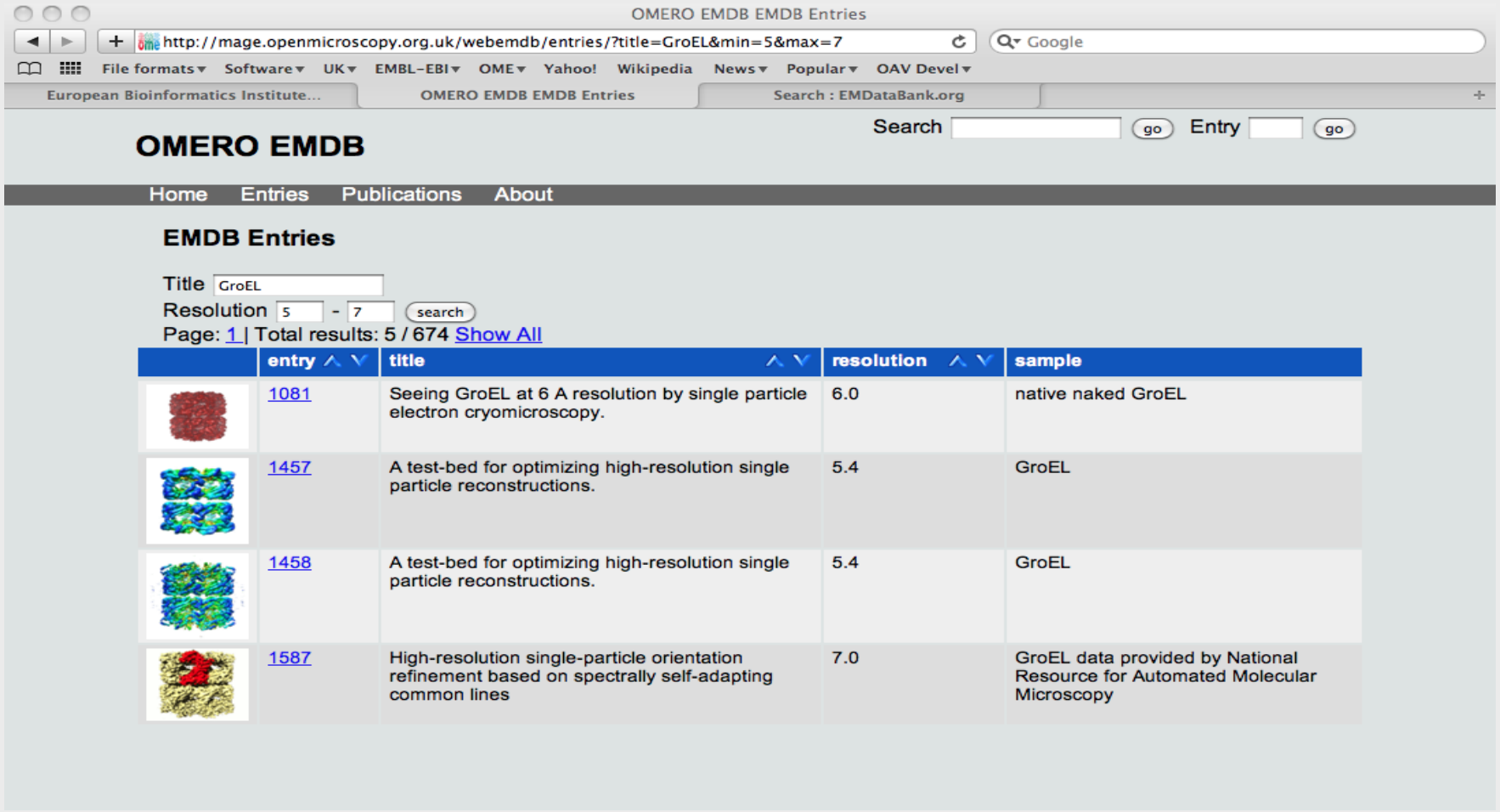
Constraints on the EMDB

- Cost of storage: $3000/TByte
- Without validation, and quality assurance, errors and inconsistencies accumulate over time rendering the stored data less useful
- Most validation currently requires manual input
- Increased efficiency with automation will be offset by increased deposition rate!
- We limit ourselves to storing the end results (3D maps) and their interpretations (segmentations)
- **Regardless of our constraints, the community needs tools for the integrated management of 3D and 2D EM data!**

OMERO EMDB objectives

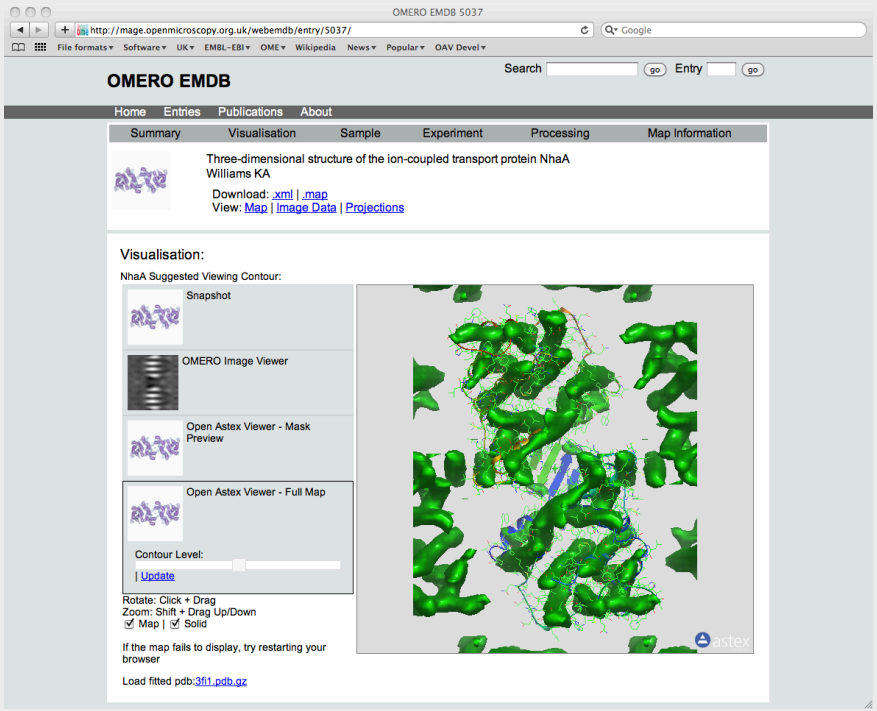
- Use the experience, data model, and data from the EMDB to prototype an integrated system for managing 2D and 3D EM data based on OMERO
- Extend BioFormats so that relevant EM formats can be handled by OMERO
- Provide web based visualisation tools for viewing
  - slices from tomograms (Slice Viewer)
  - cryo EM maps (Open Astex Viewer)
  - segmentations
  - quality assessment map/model fit, map statistics
- Manage meta data
- Access other applications such as EMAN2, Spider, and IMOD for tasks such as basic processing and validation

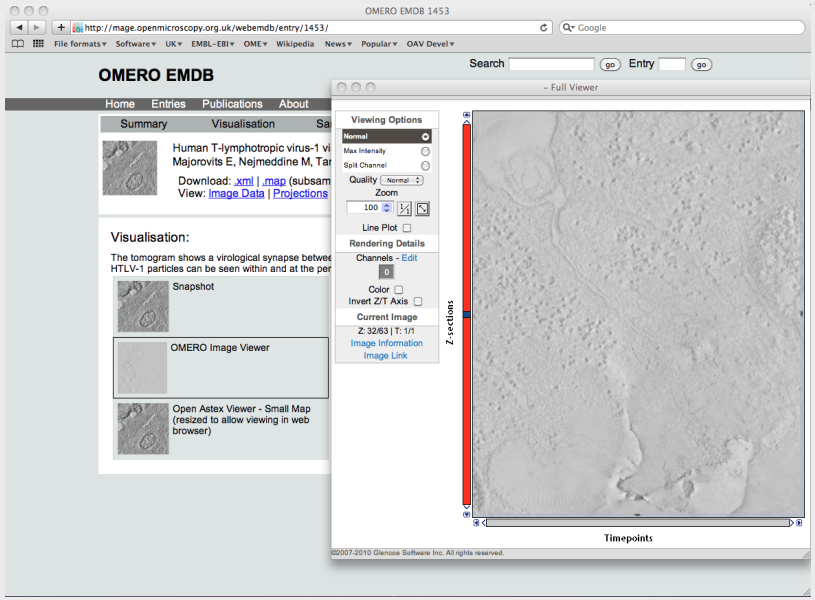
Working prototype: emdb.openmicroscopy.org.uk



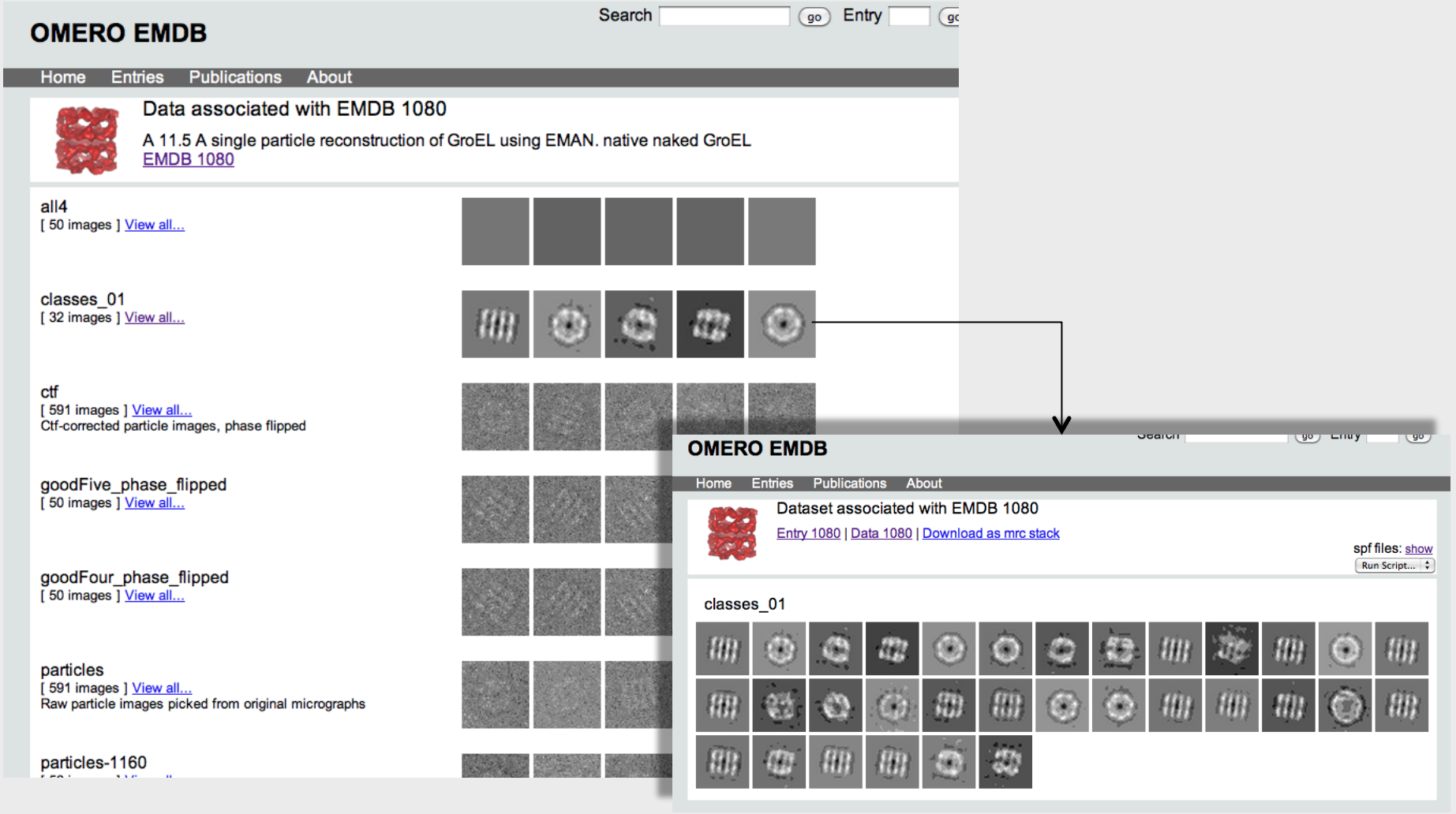
2D and 3D viewers

- Open Astex Viewer for 3D (entry 5037)
- Slice viewer for 2D (entry 1453)

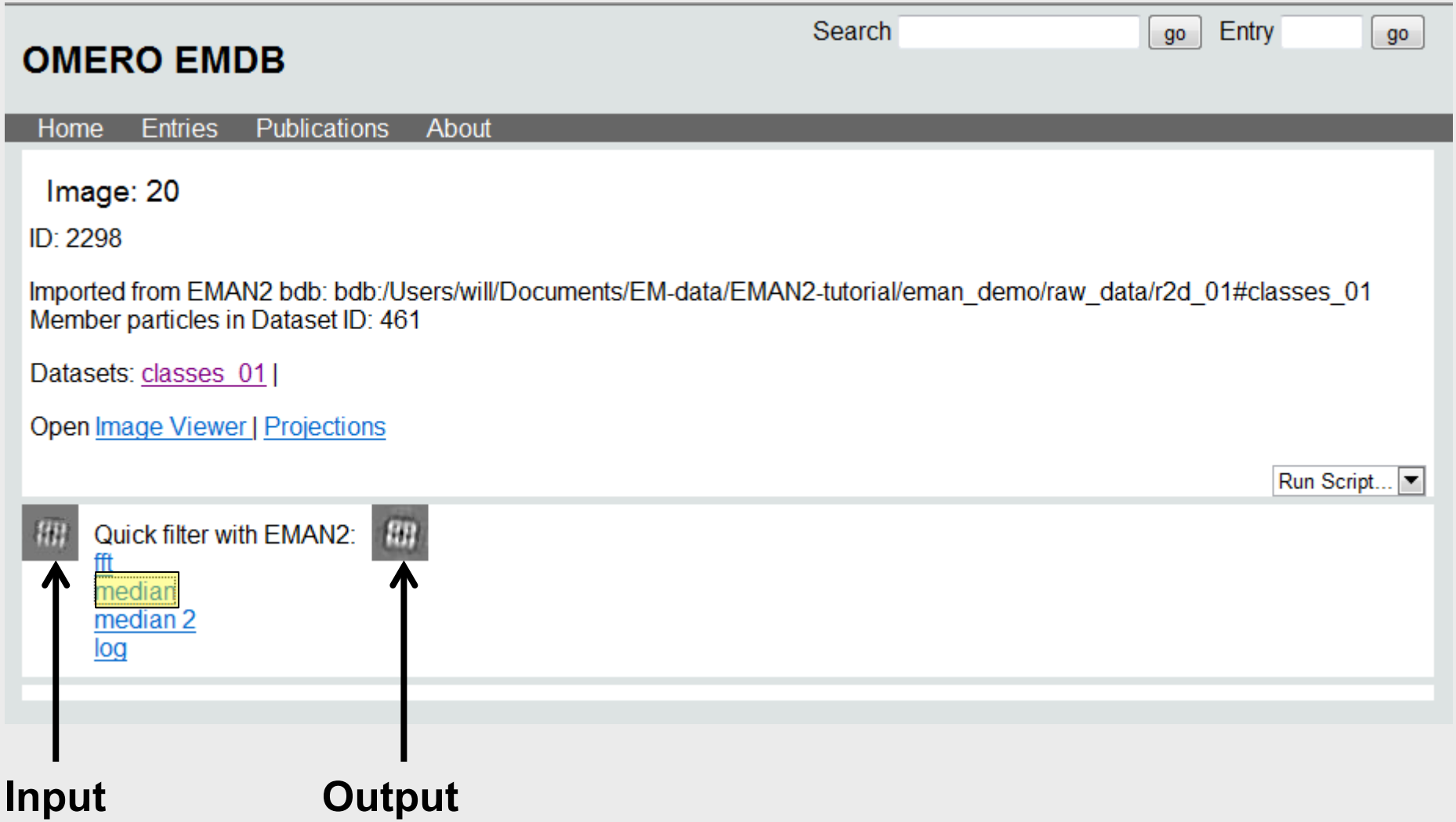




Associated 2D data



Access to other applications



Database for EM test data

- Providing easy access to test EM data is vital to support the development of software, e.g. validation tools, by the EM community
- At present there is no organized database for this purpose and the test data is offered through links from scattered sites
- The OMERO EMDB prototype would be an excellent tool for this purpose!
- The curation of information would be up to the community rather than EMDataBank.org
- Encourage user to upload test data sets
- Support for EM image formats: MRC, Spider and Imagic

Stand-alone slice viewer for tomography (mock-up)

EMDB Home | EMDB Entry EMD-1051 | Contact EMDB

Title: Untangling desmosomal knots with electron tomography.
Authors: He W, Cowin P, Stokes DL
Aggregation State: individualStructure, (resolution 30 Angstroms)

- Summary
- Visualisation
- Sample
- Experiment
- Processing
- Map Information
- Downloads

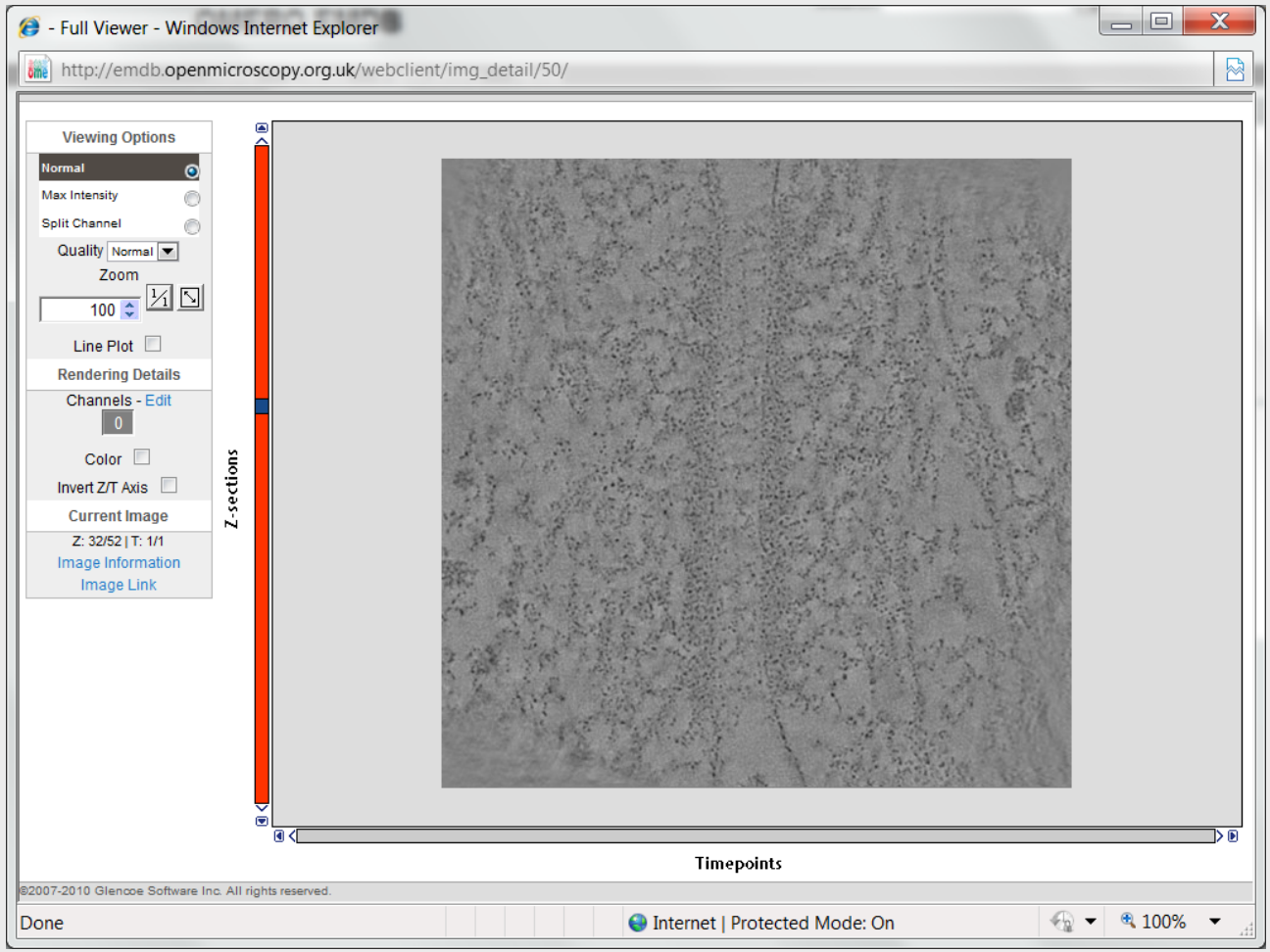
Visualisation

BIOLOGICAL CONTEXT	EMDB SNAPSHOT
This is a 3D reconstruction map of an desmosome based on high pressure freezing/freeze-subsitution electron tomography	
Suggested contour level for viewing the map: 1.95e+03	**OMERO Slice Viewer**

The EMDB images on this site were either supplied by the depositor or generated using CHIMERA.

OMERO Slice Viewer

- Pressing button re-directs user to Slice Viewer connected to OMERO server



Ideas for the future

- We invite suggestions from the audience and the EM community!
- Particularly interested in how to visualize very large datasets within the constraints of a browser app
- 3D wavelet transforms for dynamic multi-scale visualization of maps

Summary

- OMERO EMDB prototype completed
- Bioformats support for MRC, Spider and Imagic available
- Database for test data in development

Acknowledgements

- EBI
  Ardan Pathwardhan
  Gerard Kleywegt (PDBe team leader)

- RCSB
  Cathy Lawson

- University of Dundee
  Jason Swedlow
  Will Moore

- Funding
  BBSRC
  NIH
  EMBL


EMBL-EBI