

# The Electron Microscopy Data Bank and OME

*Rich data, quality assessment,  
and cloud computing*

*Christoph Best  
European Bioinformatics Institute,  
Cambridge, UK*

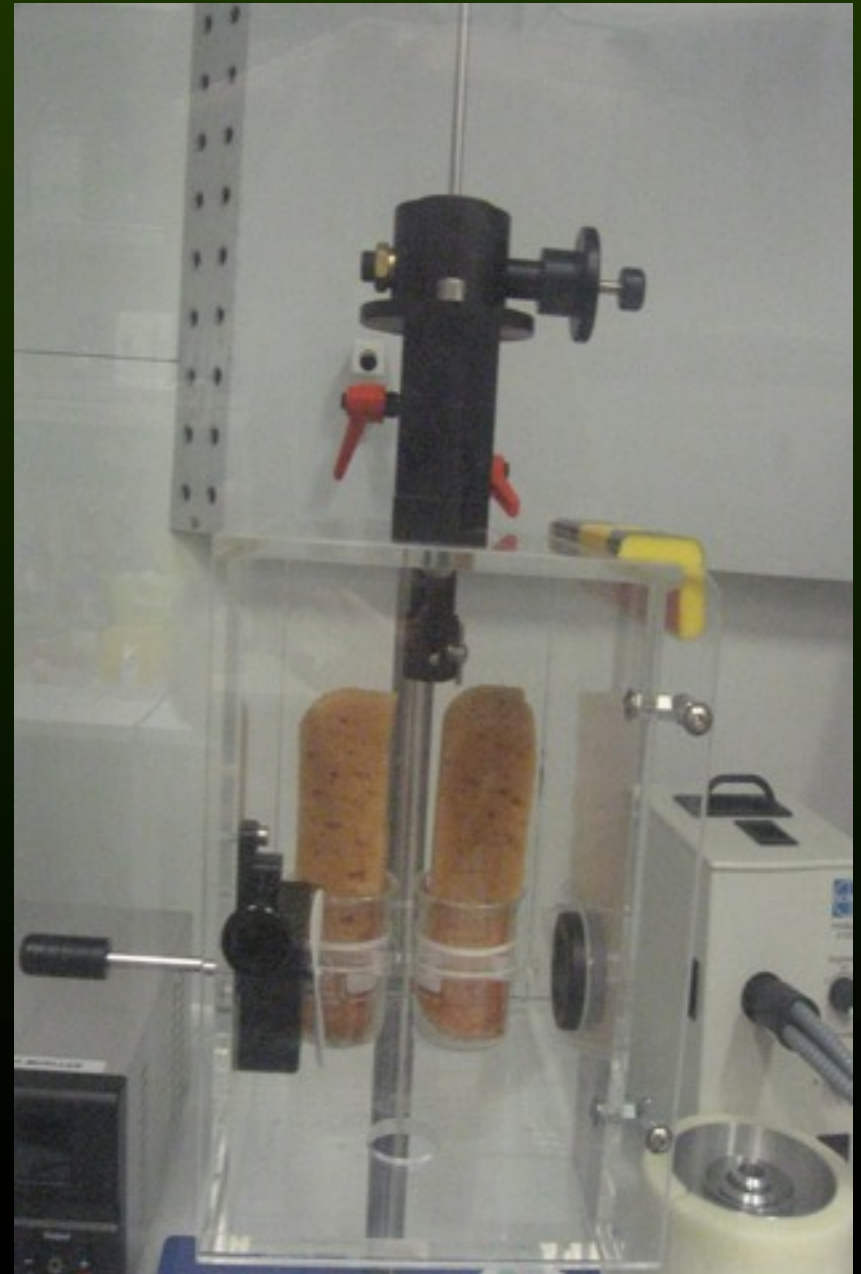
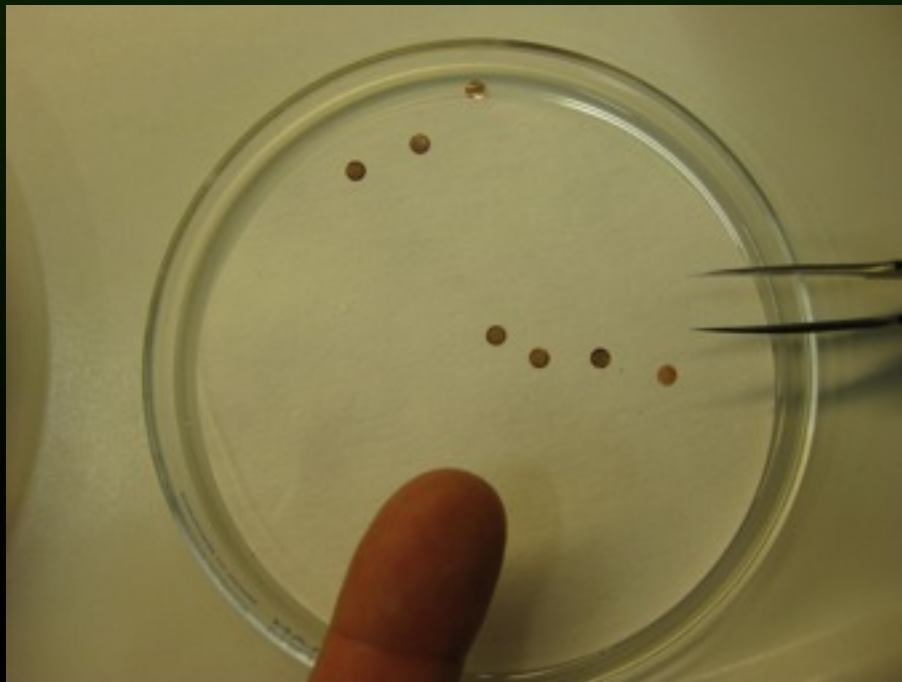
# Transmission Electron Microscope

- **ADVANTAGES**
- Short wavelength
- High cross section
- **DISADVANTAGES**
- Destroys specimen
- Low contrast, high noise
- No genetically engineerable markers
- Thin specimens only (<500 nm)



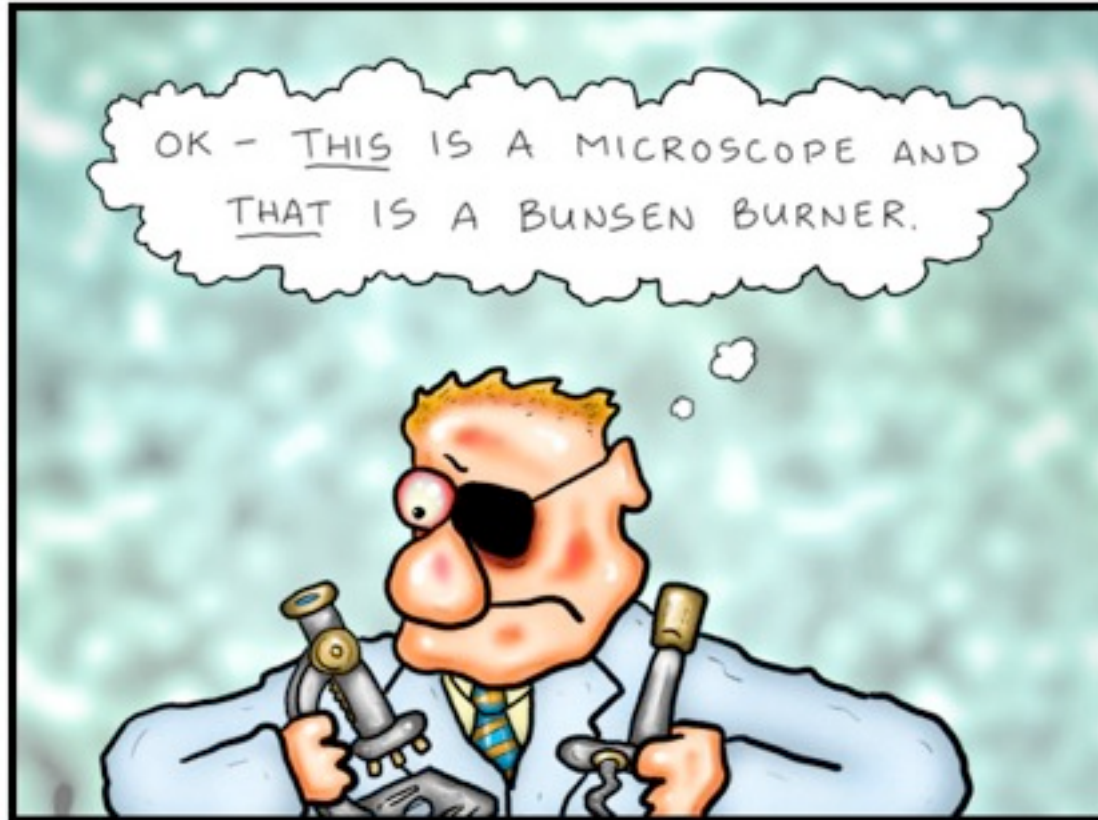
# Specimen preparation

Cryo-Electron Microscopy  
Fixation using amorphous ice  
Thin ice sheet (500 nm) in  
holey carbon on copper grid



# DOCTOR FUN

26 May 2005



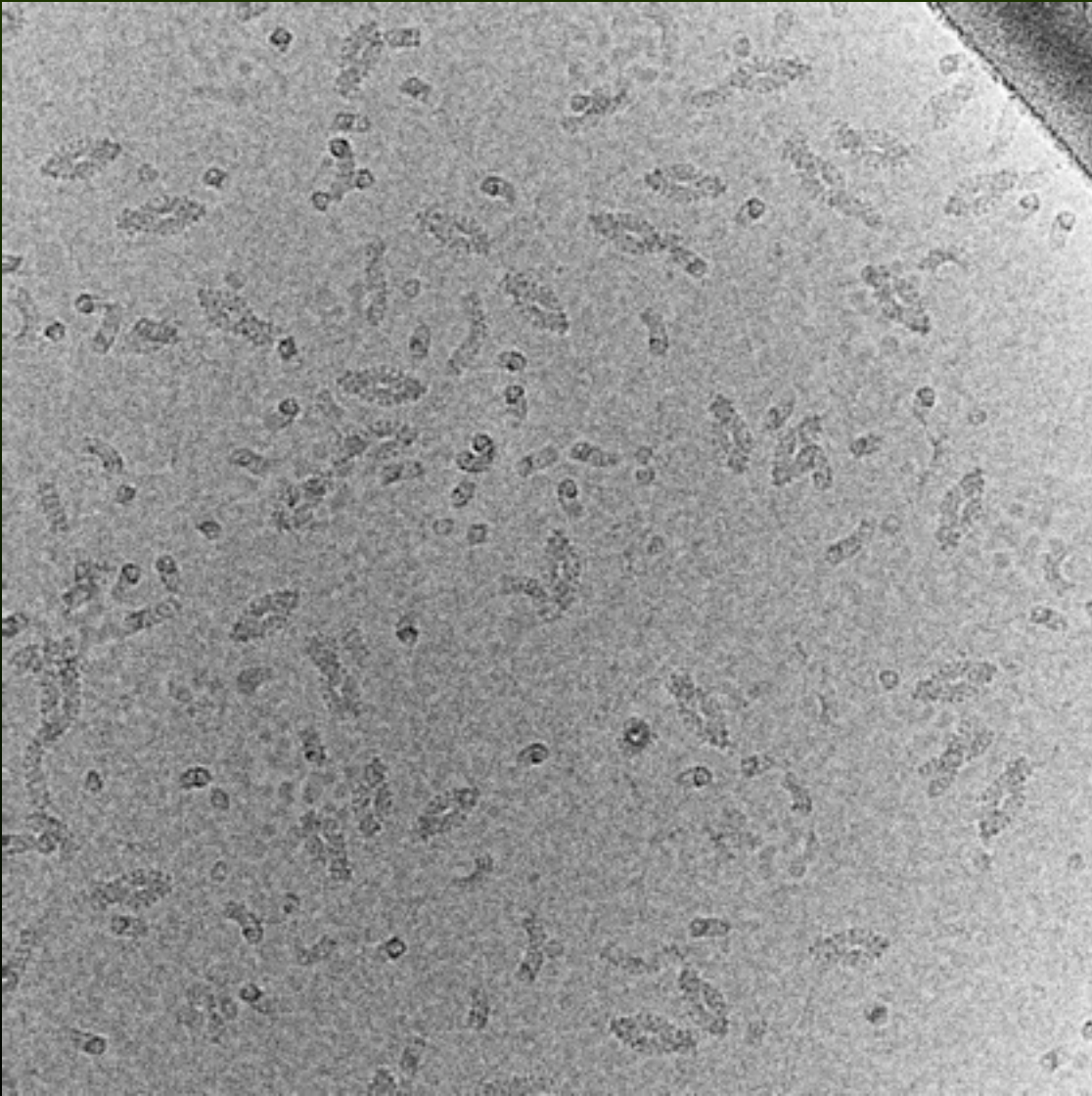
Science lab 101

Copyright © 2005 David Farley, d-farley@ibiblio.org  
<http://ibiblio.org/Dave/drfun.html>

This cartoon is made available on the Internet for personal viewing only. Opinions expressed herein are solely those of the author.



# Single-particle method

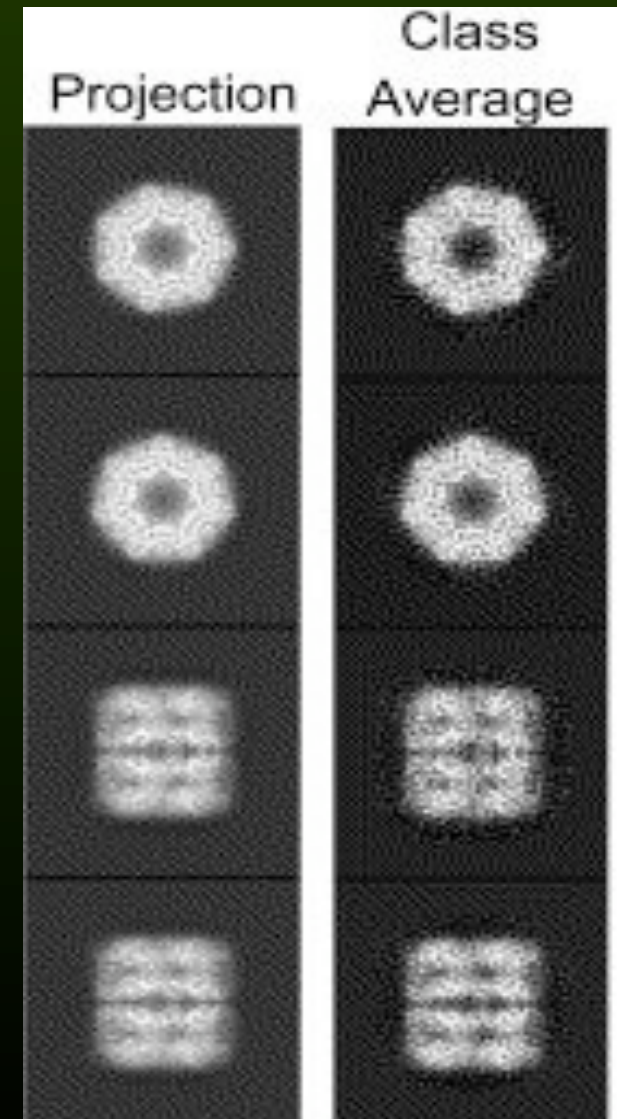
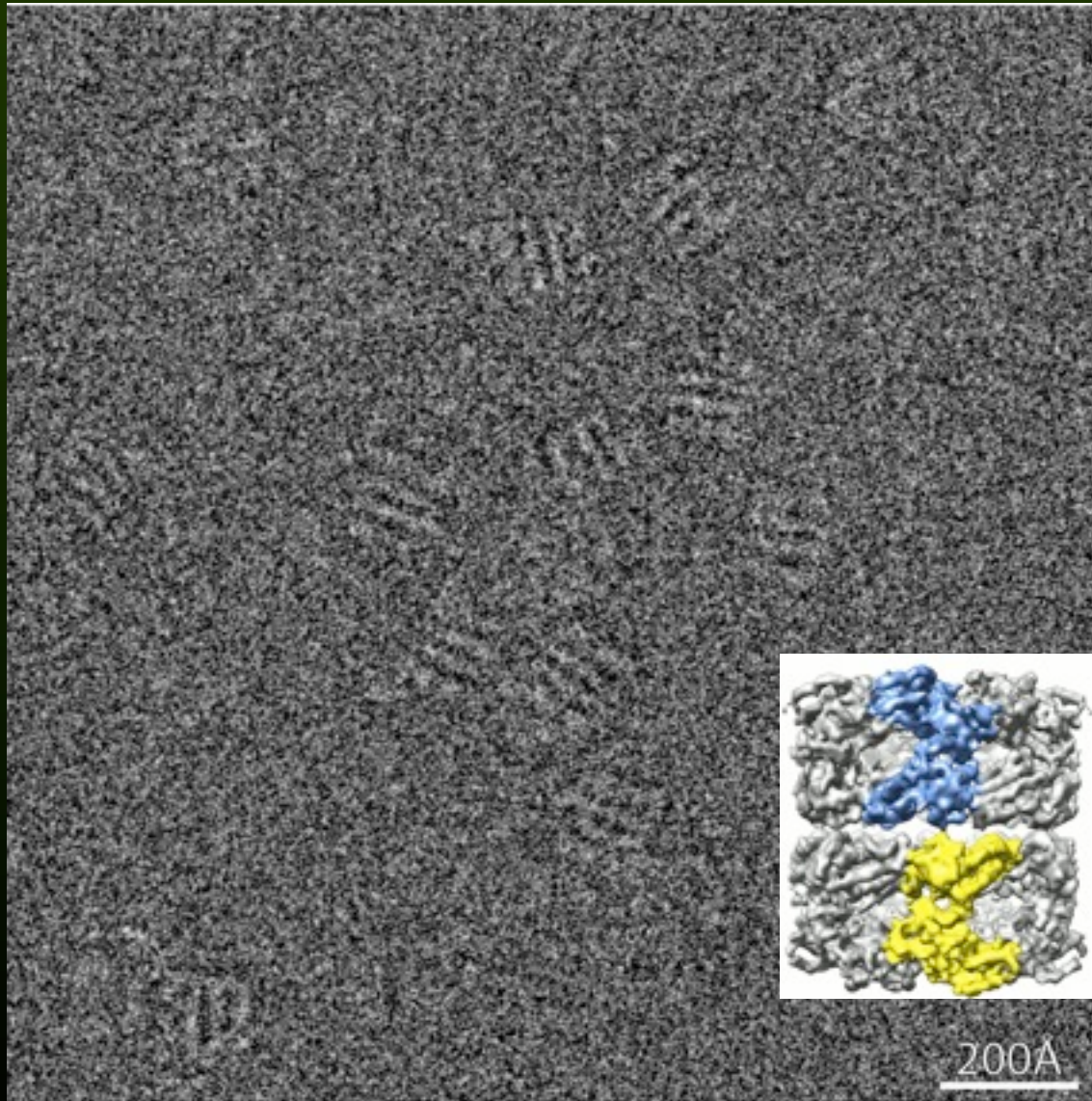


Tripeptidyl-peptidase II  
(TPP II)

courtesy of B. Rockel, Martinsried



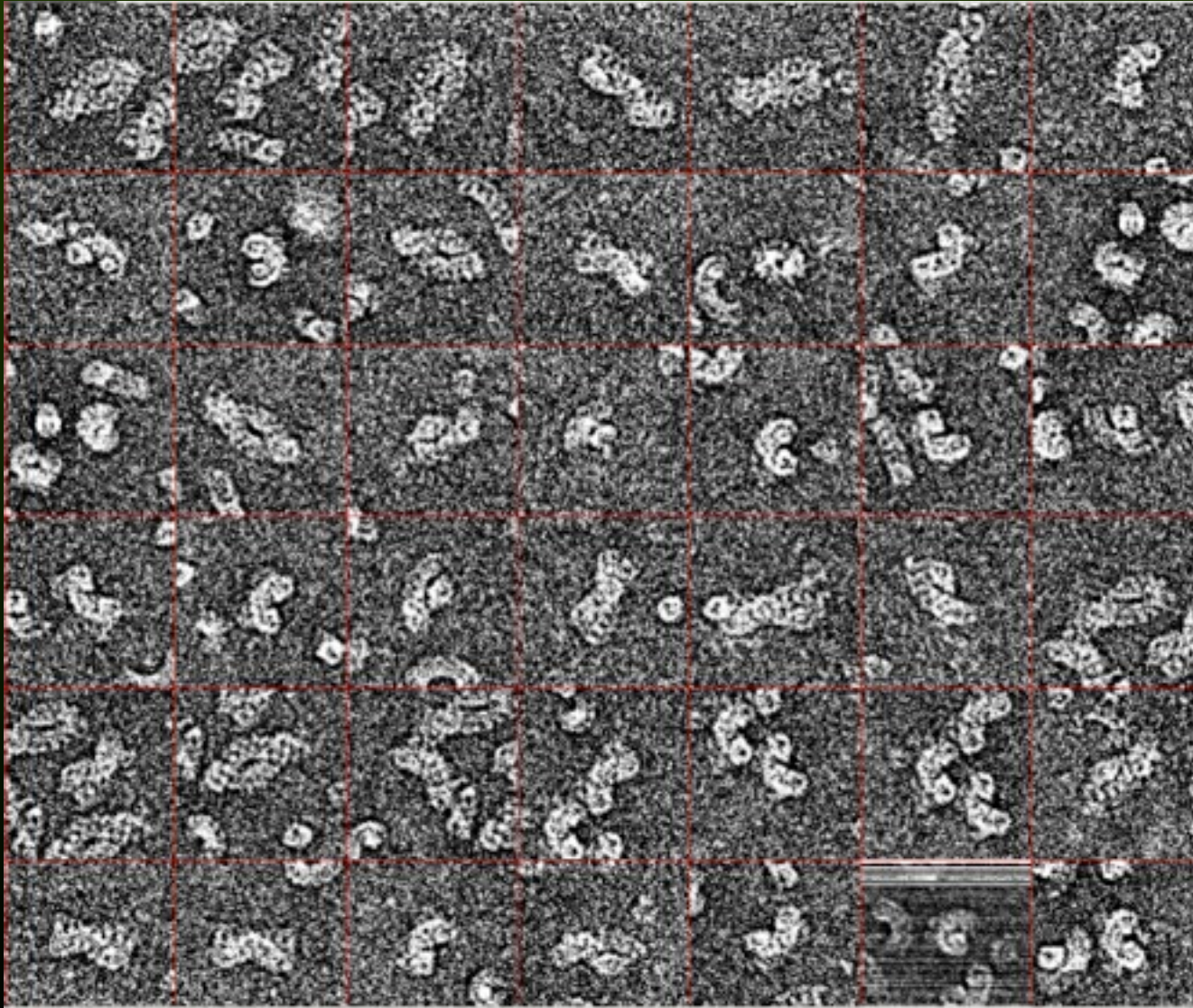
# Single-particle analysis: GroEL to 4A



Ludtke et al, Structure 2008



# Single particle stack

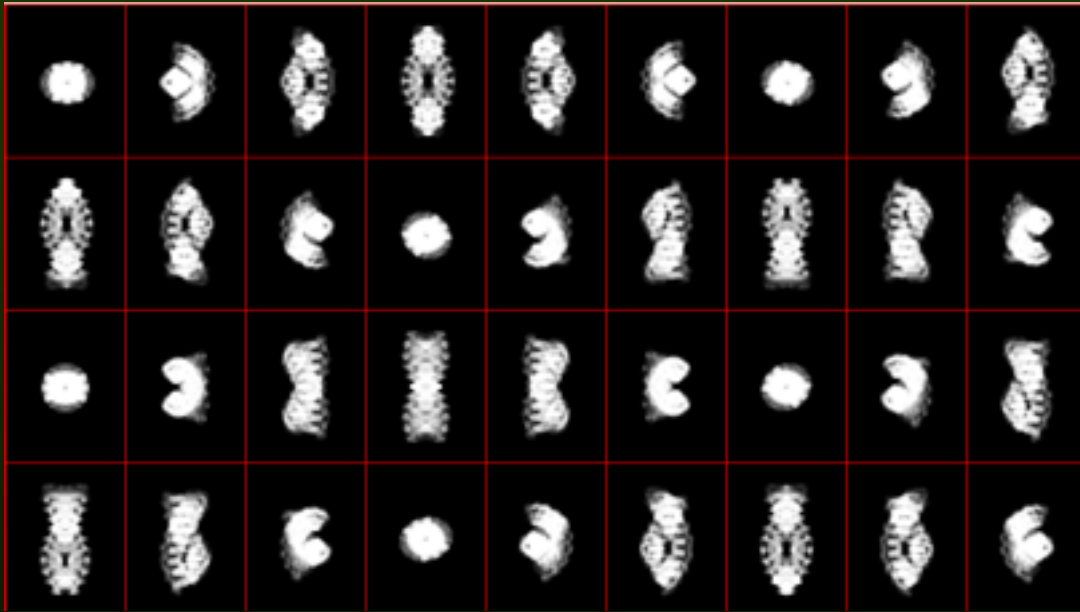


Typically several 1000 to 50000 individual images

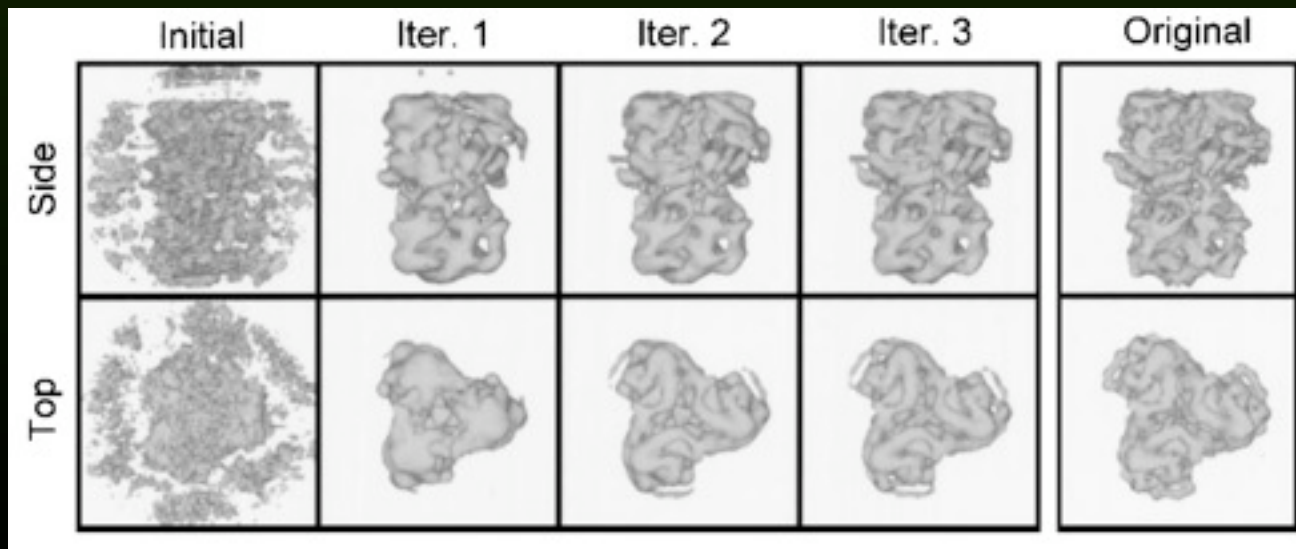
Problems:

- Automatic selection of particles
- Alignment
- Classification
- Angular assignment
- 3D reconstruction

# Classification and iterative refinement



- Classify images
- Assign angles
- 3D reconstruction

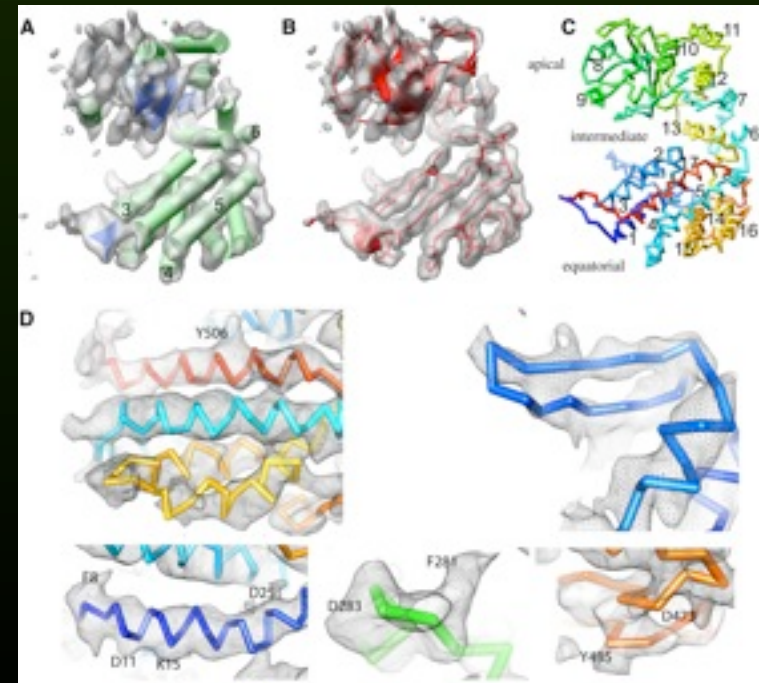
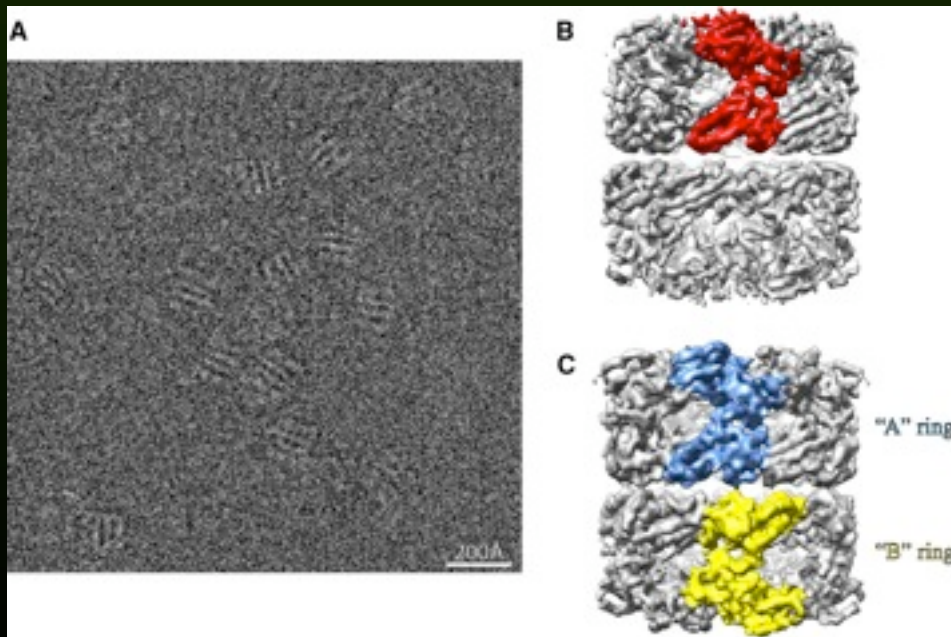




# High-resolution structures

- High resolution single particle averaging reached 4Å  
how much longer till atomic resolution?

Ludtke et al, Structure 2008



# Electron tomography

- 3D reconstruction by taking a series of images from different angles
- Difficulty:  
Nanometer accuracy

- Problems:

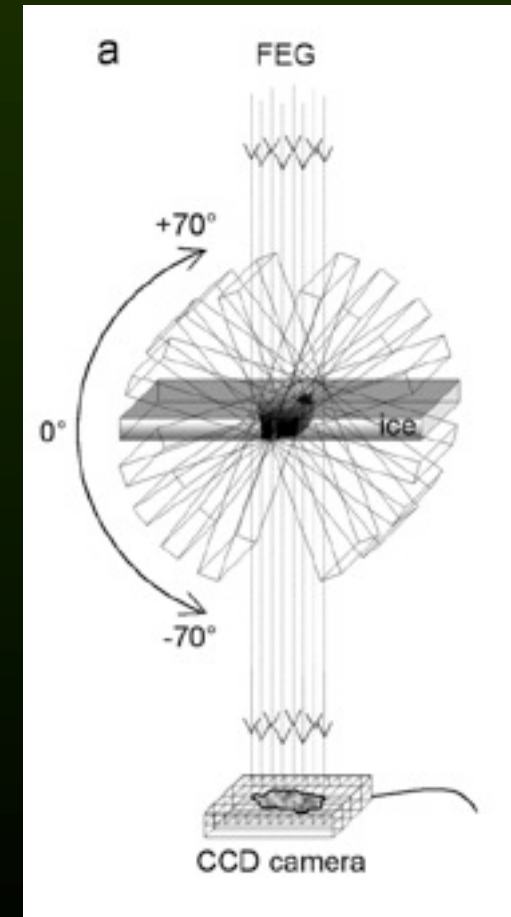
Limited tilt range

↔ **missing wedge**

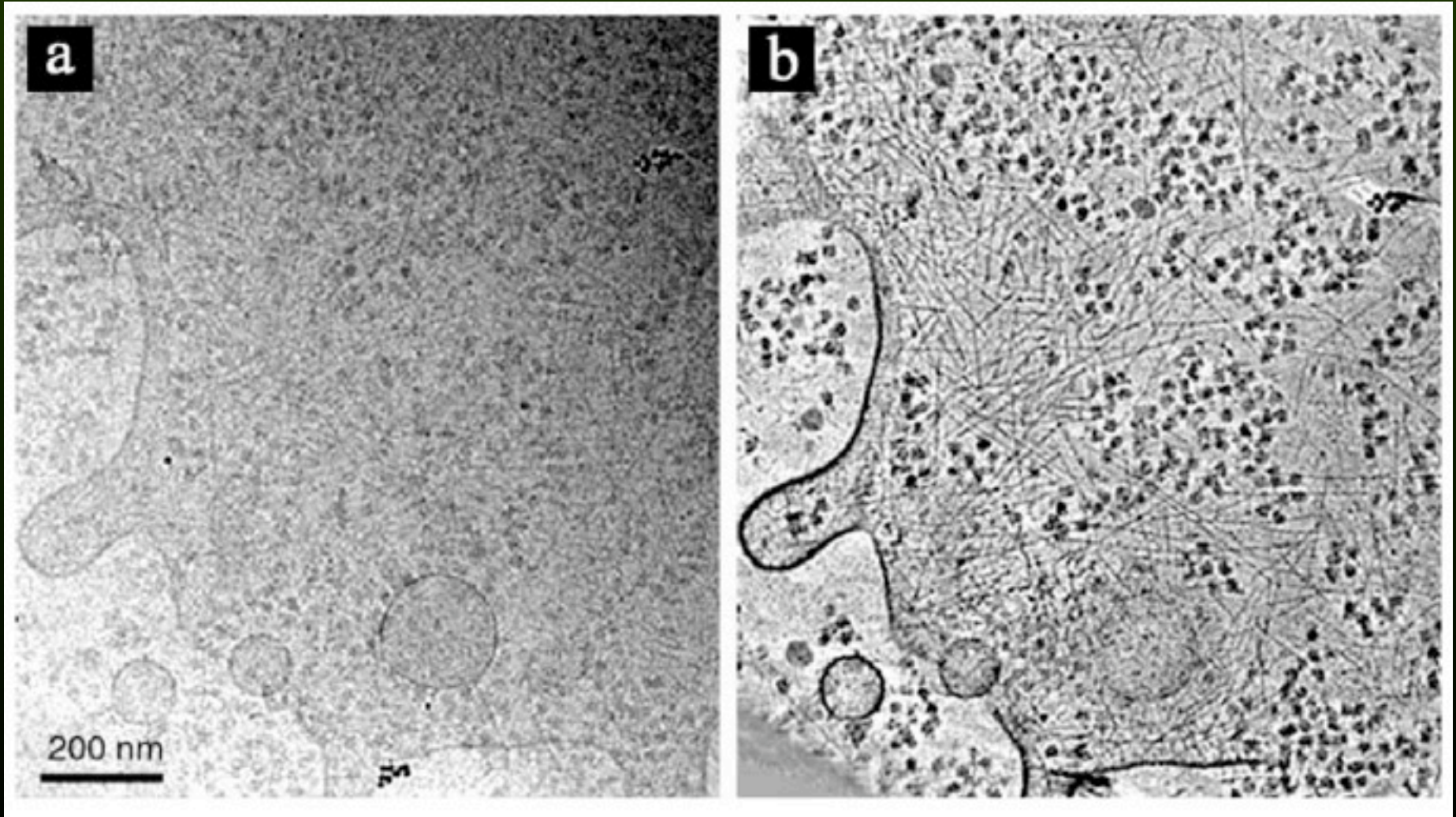
⇒ distortion

Imperfections of the tilt

↔ **alignment**



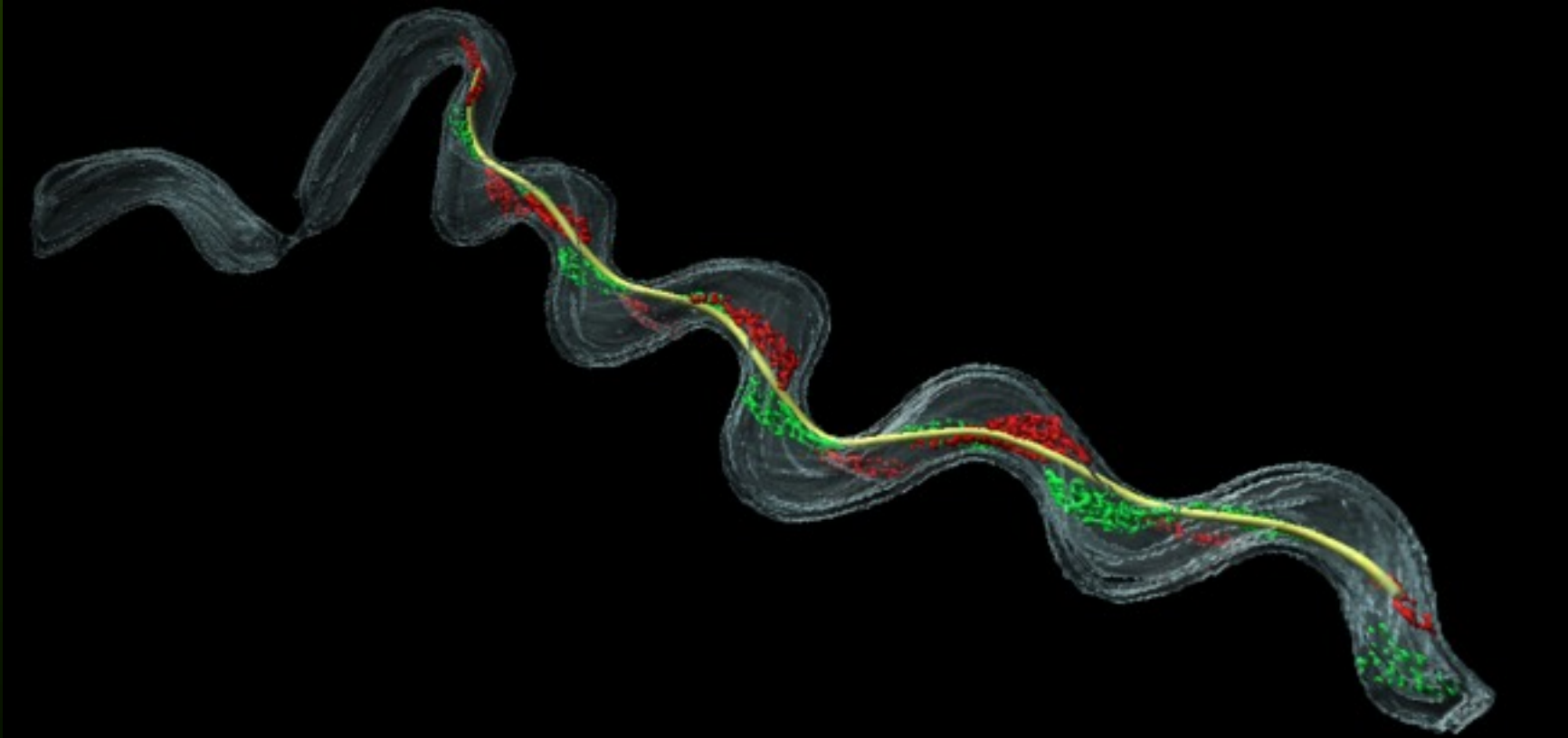
# Tomography of eukaryotic cells



*Dictyostelium discoideum*



# The cytoskeleton of *S. melliferum*

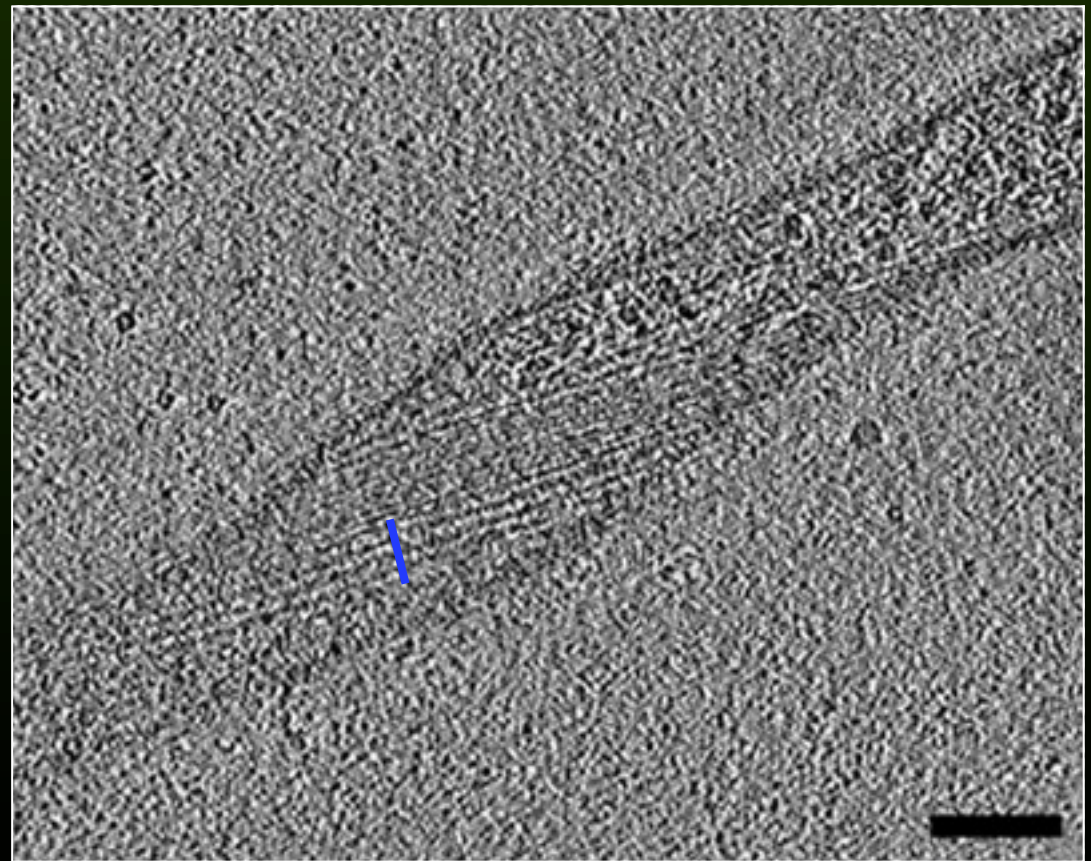
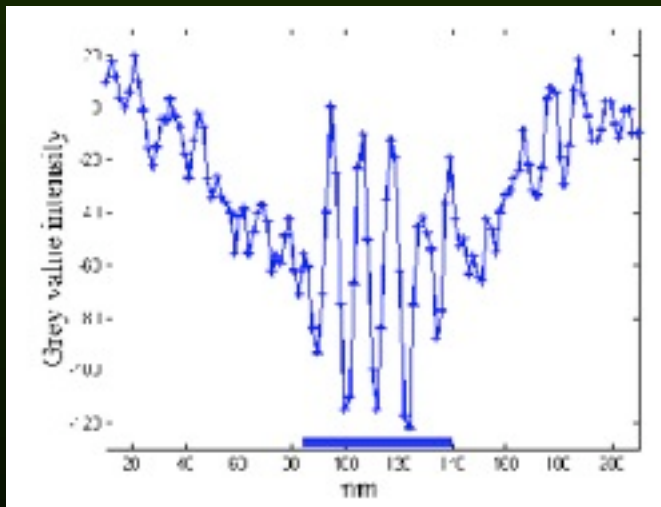


yellow: geodetic line

J. Kürner et al., Science, 2005

# Cytoskeleton

## Cytoskeleton of *Spiroplasma melliferum*



J. Kürner et al., Science, 2005

# Viruses

HIV

Vaccinia

H. simplex



projection

slice

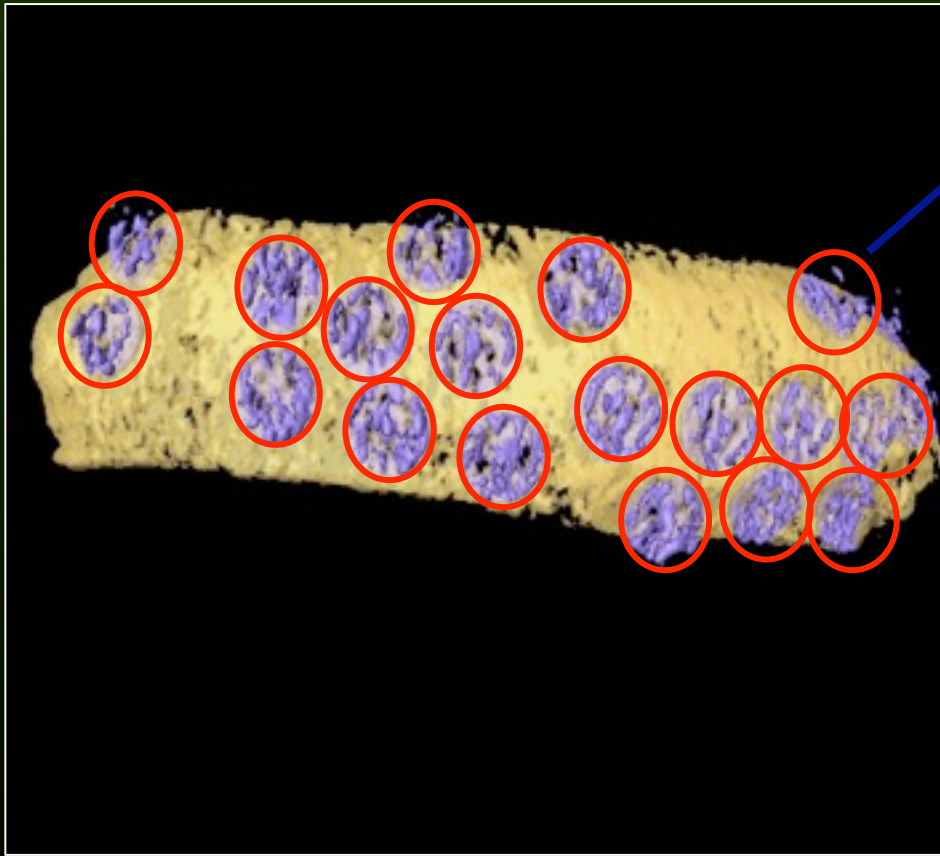
3D  
segmentation

Grünwald & Czirklaff,  
Curr. Opin. Microbiol., 2006



# 3D averaging: The Nuclear Pore Complex

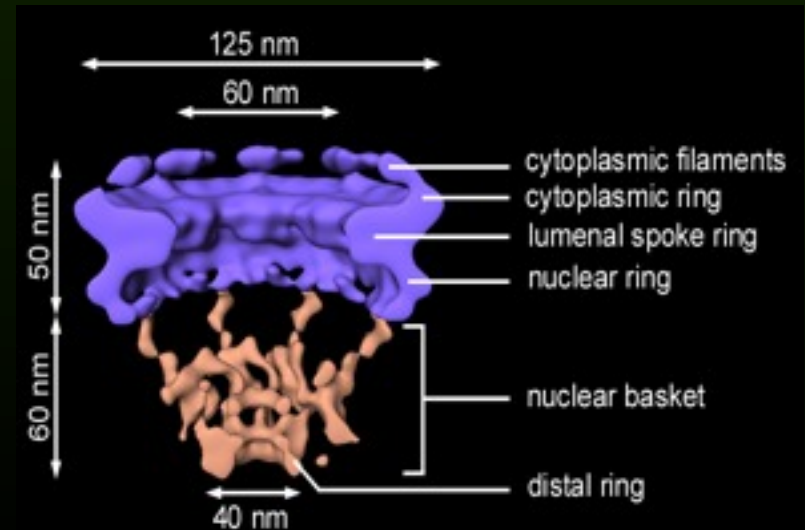
Nuclear pore complex



Matching

Aligning

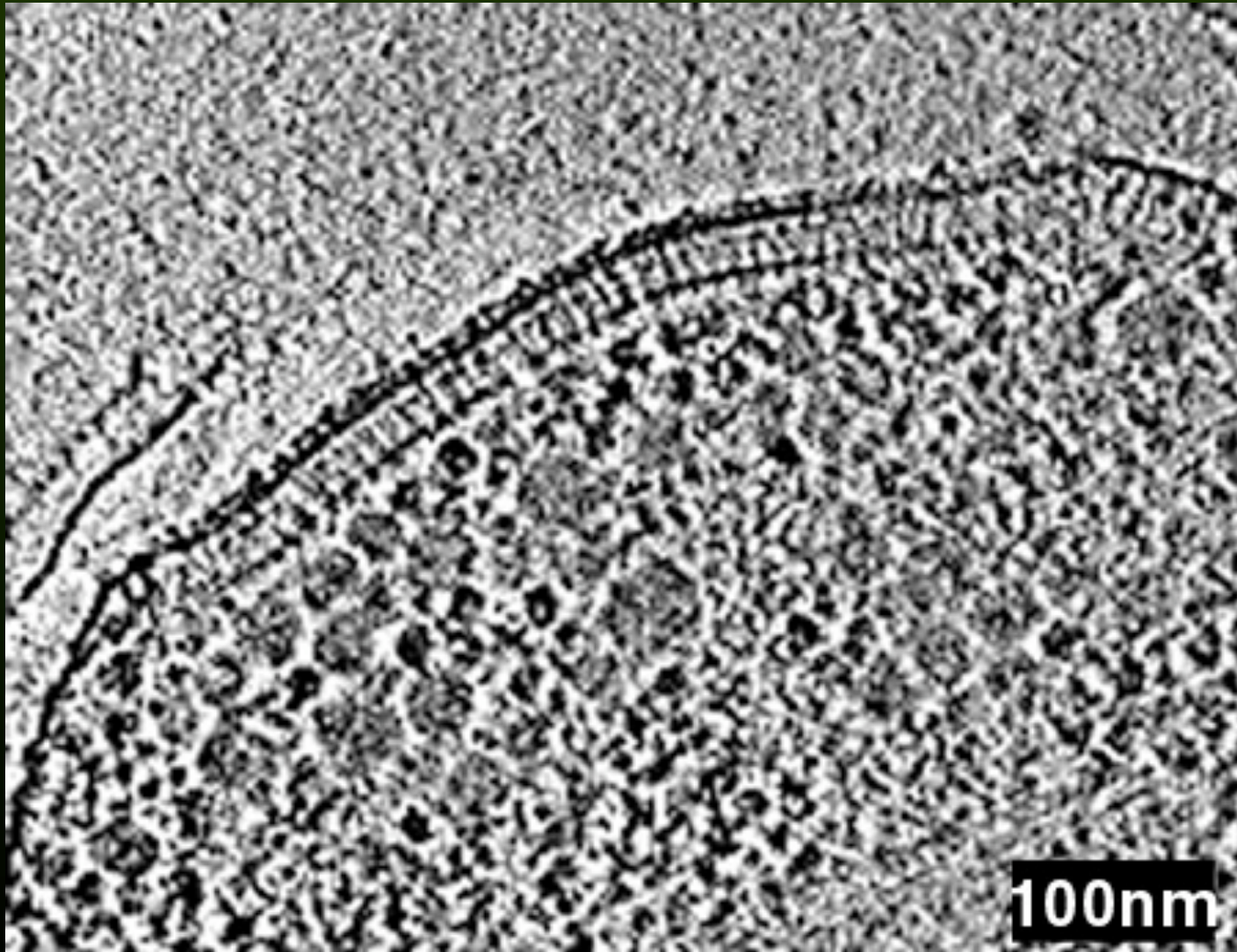
Averaging



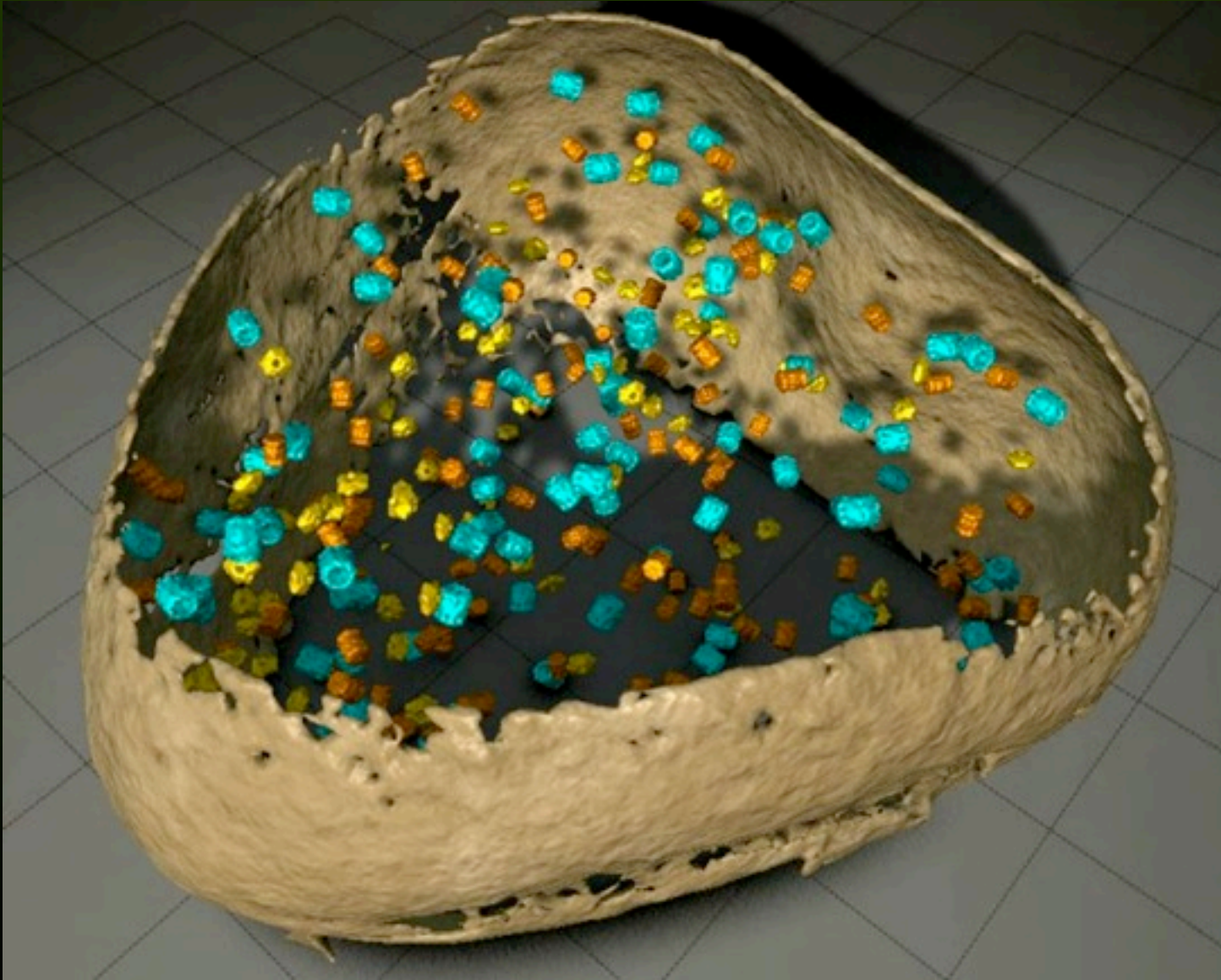
Beck et al., Science, 2004

3D reconstruction

# Macromolecular crowding



# “Visual proteomics”

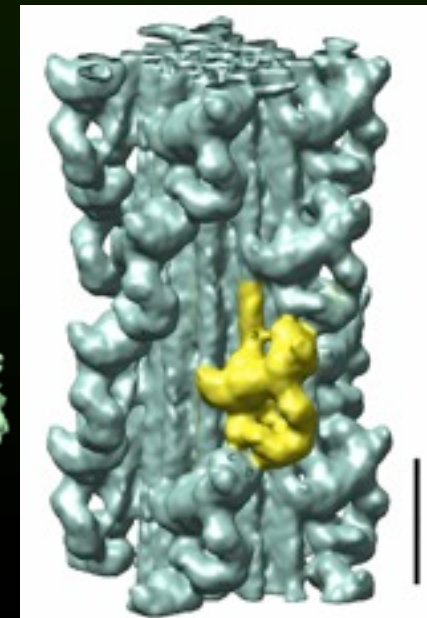
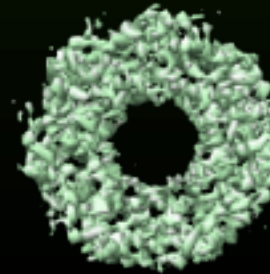
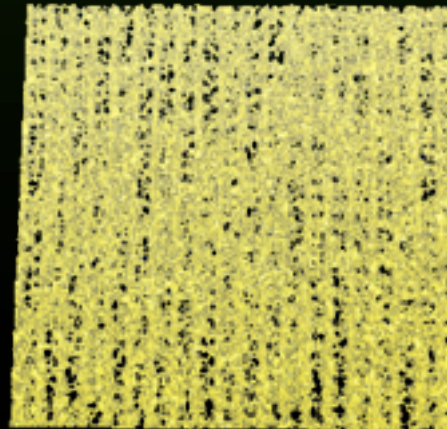
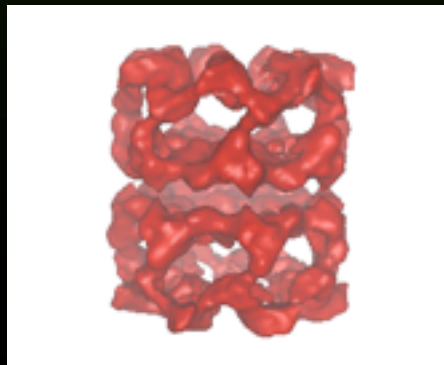
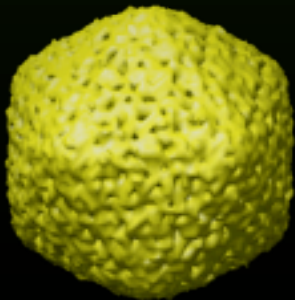
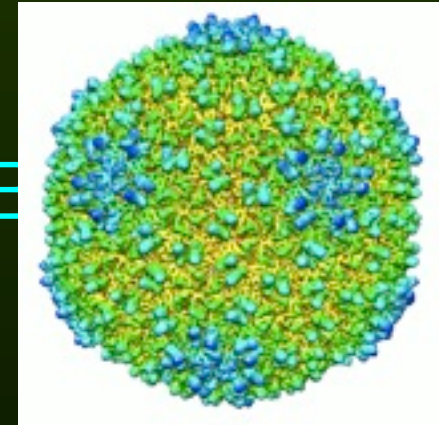




# The Electron Microscopy Data Bank

- contains EM-derived density maps
- complementary to coordinate sets in [PDB](#)
- established 2002 @ EBI (Kim Henrick)
- web-based submission and retrieval
- [hand-curated](#) (R. Newman)

a bit like Ebay – and you won't make any money, either





# EMBL - EBI

- Outstation of **European Molecular Biology Laboratory**

European inter-governmental organization with 20 members

- provides bioinformatics research and services
- EMDB is part of the Protein Databank in Europe

wwPDB worldwide repository for protein structures ("PDB" coordinate files) since 1973

EMDB operated jointly with Rutgers University under an NIH grant

- Joint BBSRC grant between EBI and Dundee to develop EMDB using OME technology in



# EMDep deposition system

- 600 entries, current rate approx. 15-20/month
- Contents of an entry:
  - Metadata (XML header) → experimental metadata/procedure
  - Map (any format, converted to CCP4/MRC)
  - Additional files

The screenshot displays the EMDep Deposition Tool interface. The top navigation bar includes 'EMDB home > depositions > EmDep', 'EmDep Deposition Tool', and 'Contact EmDep'. The left sidebar shows a navigation menu with 'Deposition reference id: 1002' and various sections: 'Initialisation', 'Instruction to EmDep', 'File Deposition', 'Title & Author', 'Reference', 'Sample Details', 'Sample Components', 'Experimental Details', 'Vitrification', 'Imaging', 'Image Processing', 'Reconstruction', 'Fitting', 'Map Details', 'Subunit', and 'Change Password'. The 'Sample Details' section is active, showing a form for entering sample information. The 'Vitrification' section is also visible, showing a form for entering vitrification parameters.

**Sample Details**  
Description of the sample used in the experiment  
CHANGING THE SAMPLE DETAILS ONCE SELECTED WILL RESULT IN A LOSS OF ALL DATA IN THE NON-MANDATORY SECTIONS AND REMOVAL OF ALL FILES UPLOADED

**Sample Details**  
Changing the sample details types once selected will remove all uploaded files and reset the data

✓	Select aggregation state	SINGLE PARTICLE	e.g. SINGLE PARTICLE
✓	Enter name of sample	Fab Fragment of MAb1-1a	e.g. Fab Fragment of MAb1-1a monoclonal antibody to Human Rhinovirus1h Nim-1a Site
✓	Enter the oligomeric state of the sample	n/a	N/A e.g. One hexamer of DnaB binds to one hexamer of DnaC
✓	Enter the number of unique components in the sample	2	e.g. 2 (for the case of antibody bound to ribosome)
✓	Enter the experimental molecular weight of the sample	n/a	N/A e.g. 0.8 (in MegaDaltons)
✓	Enter theoretical molecular weight of the sample	n/a	N/A e.g. 0.8 (in MegaDaltons)
✓	Enter method of determination of molecular weight of sample	n/a	N/A e.g. Sedimentation
✓	Enter any additional sample details	n/a	N/A e.g. The sample was monodisperse

**Vitrification**  
Details of the vitrification process

You may duplicate this section with or without the data by selecting the proper radio button followed by the **Save** button.  
 duplicate  duplicate with data  leave as is

Select cryogen	METHANE	e.g. METHANE
Enter humidity	100	N/A e.g. 90 (percent)
Enter cryogen temperature (kelvin)	4.5	N/A e.g. 4.5 (in degrees Kelvin)
Vitrification apparatus	FEI vitrobot	N/A e.g. Reichert plunger
Vitrification Method	Blot for 2 seconds before plunging	N/A e.g. Blot for 2 seconds before plunging
Time-Resolved State	Vitrified 30 msec after spray	N/A e.g. Vitrified 30 msec after spraying with effector
Vitrification Details	Vitrification carried out in argon atmosphere	N/A e.g. Vitrification carried out in argon atmosphere

Save Global N/A



# EMDB search system


home > search > **EmSearch** EMDB Search Tool [Contact EmDep](#)



Main		Results
<b>ID code:</b>	<input type="text"/> EMDB <input type="button" value="v"/>	<input type="checkbox"/> EMDB title
<b>Author Last Name:</b>	<input type="text"/>	<input type="checkbox"/> Sample Name
<b>Title:</b>	<input type="text"/>	<input type="checkbox"/> Aggregation Type
<b>Sample Name:</b>	virus	<input type="checkbox"/> Resolution
<b>Citation Abstract:</b>	<input type="text"/>	<input type="checkbox"/> Resolution Method
<b>Aggregation Type:</b>	Any Type <input type="button" value="v"/>	<input type="checkbox"/> Microscope Type
<b>Resolution:</b>	high: <input type="text"/> low: <input type="text"/>	<input type="checkbox"/> Reconstruction Software
<b>Date:</b>	Release <input type="button" value="v"/> From <input type="text"/> To <input type="text"/>	<input type="checkbox"/> Interpro ID
<input type="button" value="Start search"/>		<input type="checkbox"/> GO ID
		<input type="checkbox"/> ICTVDB Id
		<input type="checkbox"/> PDB ID
		<input type="checkbox"/> Model PDB ID
		<input type="checkbox"/> Pubmed ID
		<b>Results per page:</b> <input type="text"/> 20 <input type="button" value="v"/>
		<input type="button" value="Start search"/> <input type="button" value="Reset form"/>
		<a href="#">New search</a>

primary developers: Sumeer Velankar

# EMDB search system

home > services > EmSearch EMDB Search Tool [Contact EmDep](#)

1 2 3  
  
 Found 55 hits (3 pages). Showing hits 1 to 20.  
[Start a new search](#) [Refine the results of this search](#)

Download all results  
   
[Text](#) [XML](#)

EMDB Accession code ▲▼	EMDB title ▲▼	Resolution ▲▼	Sample Name ▲▼
<a href="#">1368</a>	DNA Packaging is Similar in a Filamentous Archaeal Virus and Eukaryotic Nucleosomes	18.00Å	Acidianus Filamentous Virus 1
<a href="#">1177</a>	Structure of Broadhaven virus by cryoelectron microscopy	23.00Å	Broadhaven virus
<a href="#">1005</a>	The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes	13.40Å	Bromelain-treated PM2 virus
<a href="#">1118</a>	Structural polymorphism of the major capsid protein of a double stranded RNA virus, an amphipathic helix as a molecular switch	15.00Å	chimeric infectious Bursal Disease Virus capsid
<a href="#">1165</a>	A 9 Å Single Particle Reconstruction From CCD Captured Images On A 200 kV Electron Cryomicroscope	9.00Å	Cytoplasmic Polyhedrosis Virus
<a href="#">1166</a>	Cryo-EM reconstruction of dengue virus in complex with the carbohydrate recognition domain of DC-SIGN	25.00Å	Dengue virus complexed with CRD domain of DC-SIGN
<a href="#">1167</a>	Cryo-EM reconstruction of dengue virus in complex with the carbohydrate recognition domain of DC-SIGN	0.00Å	Dengue virus complexed with CRD domain of DC-SIGN
<a href="#">1246</a>	Distribution and three-dimensional structure of AIDS virus envelope spikes	32.00Å	Envelope Spike on the surface of SIVmac239 virus with truncated cytoplasmic tail
<a href="#">1410</a>	Neutralization of dengue virus by a serotype cross-reactive antibody elucidated by cryoelectron microscopy and x-ray crystallography	24.00Å	Fab Fragment of MAb 1A1D-2 complexed with Dengue 2 virus
<a href="#">1399</a>	Abstract Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions.	0.00Å	Hepatitis B virus
<a href="#">1400</a>	Abstract Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions.	22.00Å	Hepatitis B virus
<a href="#">1401</a>	Abstract Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions.	22.00Å	Hepatitis B virus
<a href="#">1402</a>	Abstract Cryo-electron microscopy of hepatitis B virions reveals variability in	0.00Å	Hepatitis B virus

# EMDB Atlas pages

EMD - entry 1020 - Mozilla  
http://www.ebi.ac.uk/msd-srv/ensearch/atlas/1020\_summary.html

MSD Menu Search MSD by Text Search for

MSD EMBL-EBI

Home > services > atlas pages

**Macromolecular Structure Database**

Entry EMD-1020

**Title:** Structure of a viral DNA gatekeeper at 10 resolution by cryo-electron microscopy  
**Authors:** E.V.Orlova, B.Gowen, A.Droegge, A. Sbege, F.Weise, R.Lutz, M. van Heel, P. Tavares  
**Source:** Bacteriophage SPP1 portal protein gp5  
**Aggregation State:** singleParticle, (resolution 10 Angstroms)

Summary

Source:	Natural protein
Primary Citation:	E.V.Orlova, B.Gowen, A.Droegge, A. Sbege, F.Weise, R.Lutz, M. van Heel, P. Tavares, Structure of a viral DNA gatekeeper at 10 resolution by cryo-electron microscopy, <i>Nature</i> 438, 179-183 (2005)
Release Date:	31-Mar-2003
Resolution:	10 Å (resolution)
Oligomeric Details:	Assembly 1

• Summary  
• Visualisation  
• Sample  
• Experiment  
• Processing  
• Downloads

**Beta**  
This is a BETA service

Primary developers: Joël Filion & John Tate & Richard Newman

EMD - entry 1020 - Mozilla  
http://www.ebi.ac.uk/msd-srv/ensearch/atlas/1020\_visualization.html

MSD Menu Search MSD by Text Search for

MSD EMBL-EBI

Home > services > atlas pages

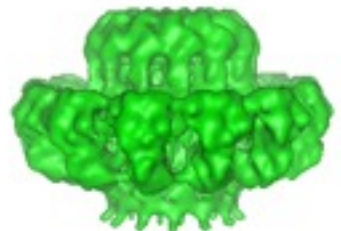
**Macromolecular Structure Database**

Entry EMD-1020

**Title:** Structure of a viral DNA gatekeeper at 10 resolution by cryo-electron microscopy  
**Authors:** E.V.Orlova, B.Gowen, A.Droegge, A. Sbege, F.Weise, R.Lutz, M. van Heel, P. Tavares  
**Source:** Bacteriophage SPP1 portal protein gp5  
**Aggregation State:** singleParticle, (resolution 10 Angstroms)

Summary

Visualisation

BIOLOGICAL CONTEXT	EMDB SNAPSHOT
In tailed bacteriophages and herpes viruses, the viral DNA is packaged through the portal protein channel. Channel closure is essential to prevent DNA release after packaging. This is the connector structure from bacteriophage SPP1.	

The EMD images on this site were generated using [CHIMERA](#)

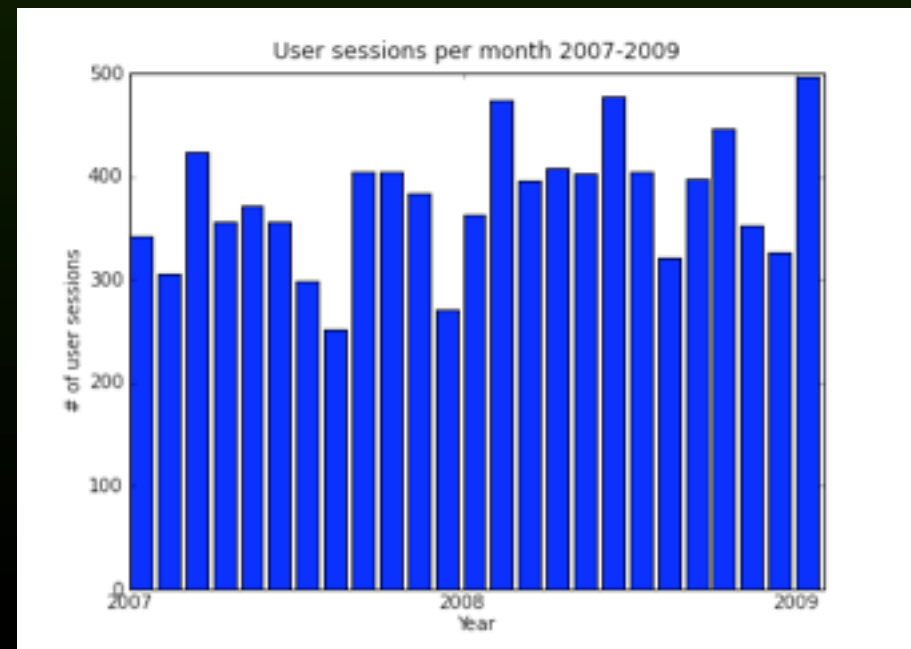
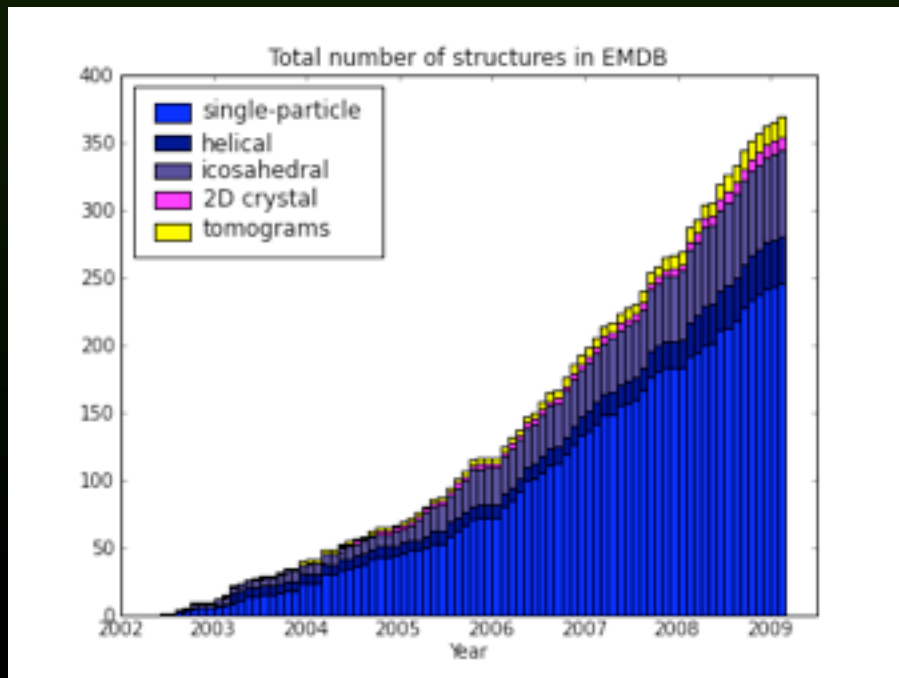
The images in these atlas pages are generated by Richard Newman from the MSD group.

Primary developers: Joël Filion & John Tate & Richard Newman



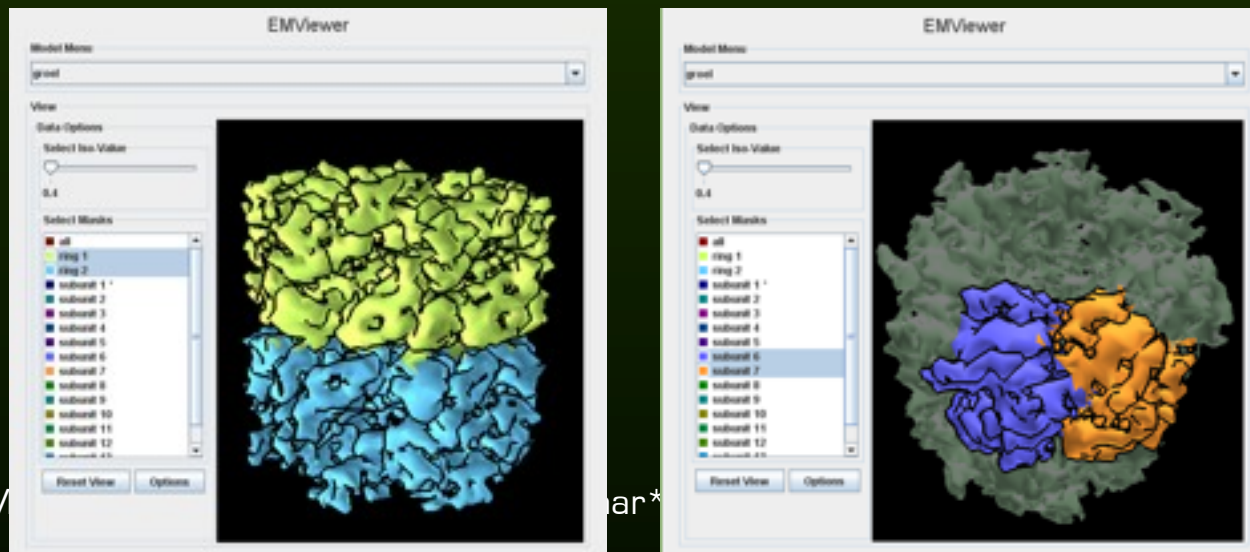
# Deposition statistics

- about 300 structures now, growing fast
- deposition mandatory in major journals
- mostly single-particle structure
- tomograms welcome, but little infrastructure



# Online visualization

- Improved visualization tools:  
Java applet embedded in page



data dictionary  
change:

- contour level becomes a mandatory data item

- OpenAstexViewer (now O/S)
- OMERO visualization server

# Rich data sets

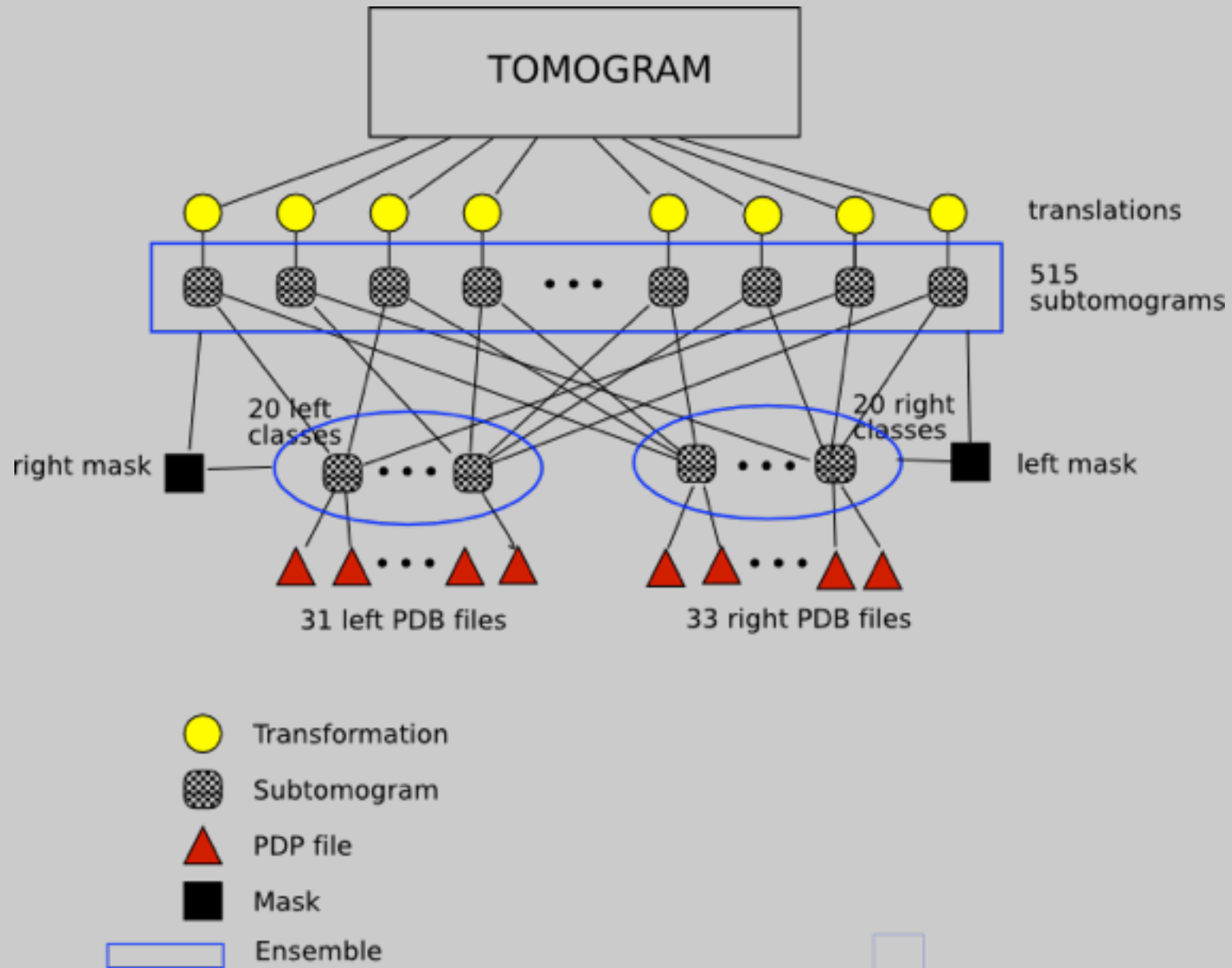
- Submissions consist of
  - maps (increasingly more than one)
  - relations between data sets → unexpressed
- XML-based standards for representing **relationships between data:**

Subject-predicate-object relationships (RDF framework)

- Harvesting interface to EM processing software
- Web-based **visualization** for sub-mission and retrieval, complex sub-missions assembled interactively (**AJAX**)



# Rich data submissions



# Possible XML representation

```
<map id="map1" src="map1.map" fileType="mrc" name="MAP1">
  <description>The primary tomographic map in this entry</description>
</map>
<atomicCoordinates id="pdb1" src="lpma.pdb" name="PROTEASOME">
  <description>Proteasome template</description>
<atomicCoordinates/>
<group id="group1" name="PROTEASOMES">
  <description>Putative proteasomes identified by template matching</description>
  <contains id="rel1" about="map1" target="pdb1">
    <transformMatrix>
      1 0 0 0
      0 1 0 0
      0 0 1 0
    </transformMatrix>
  </contains>
  <contains id="rel2" about="map1" target="pdb1">
    <comment>Low p-value</comment>
    <transformMatrix>
      1 0 0 50
      0 1 0 50
      0 0 1 20
    </transformMatrix>
  </contains>
</group>
```

# Quality assessment

- Quality assessment requires
- Should we store original data? And how much?
- Disk/network limitations can be overcome
- Allows quality assessment, future algorithms? Data mining?

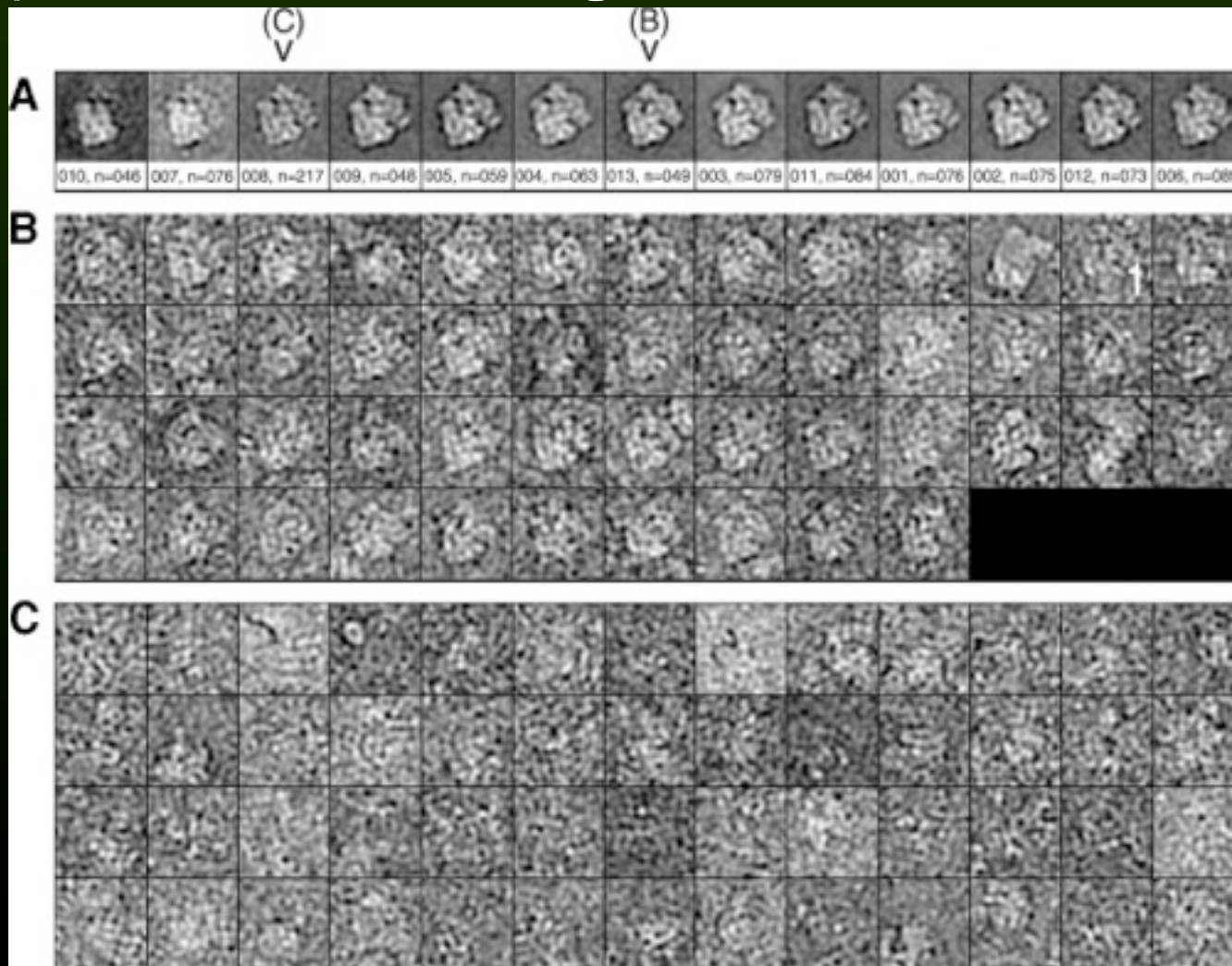




# QA for single particle

- Example: Characterizing classes

Shaikh et al., 2008



# Grid computing for EMDB: OMERO.grid?

- Data upload/distribution

Background data upload tool

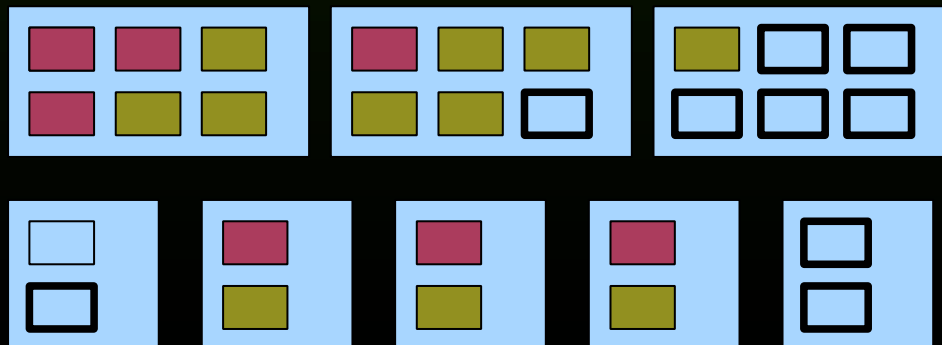
Remote filesystem/image server access

- Distributed computing resources

Access to cloud computing resources

Dynamically instantiated virtual machines

Quality assessment processing



# Conclusions

- Electron Microscopy is an example of a mature bioimaging field
  - Community with similar experimental/computational protocols
  - Many images/data sets combined to yield single result
- Where OME can improve it
  - Improved user experience
  - In-house database (EMDB is depository only)
- Future directions with OME
  - Rich data sets
  - Quality assessment tools, data set browsing
  - Grid/cloud computing