

EMDB data deposition

Image processing for cryo-electron microscopy, Birkbeck EMBO practical cryo-EM course
September 15, 2017

Ardan Patwardhan



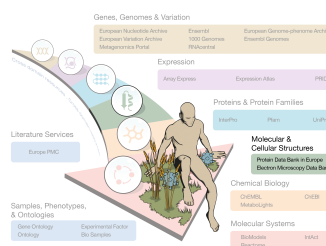
Plan

- Introduction to EMD and EMPIAR
- Trends in the field
- Overview of web resources for searching visualisation and validation
- Planned initiatives and developments
- How to deposit to EMD and EMPIAR

EMBL-EBI

Molecular and Cellular Structure Cluster

- Maintain and manage archives
 - PDB for atomic coordinate models
 - EMD for 3DEM reconstructions
 - EMPIAR for 3DEM raw data
- Develop and maintain web-services – searching, visualisation and validation
- Facilitate community-wide initiatives
- Key themes – integration (with other bioinformatics resources and imaging scales), validation, data-mining



EMBL-EBI

Cellular Structure and 3D Bioimaging Team



Sanja Abbott



Andrii Iudin



Paul Korir



Sriram Somasundharam

- PDB/EMDB annotation staff (Alice Clark...)
- Former members: E Sanz Garcia, C Lugo, J Salavert Torres (UPV), I Lagerstedt (Lilly)
- Former students: M Holmdahl, V Lysenkov, I Solanes Valero
- 1 positions available:
 - validation

EMBL-EBI

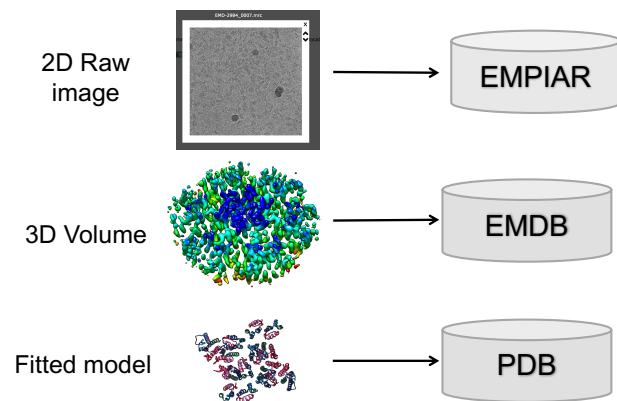
Collaborations & acknowledgements

- Birkbeck College (Elena Orlova, Maya Topf, Helen Saibil)
- Baylor (Wah Chiu)
- Chinese Academy of Sciences (Fei Sun)
- CNB Madrid (Jose Maria Carazo)
- Dundee (Jason Swedlow)
- EMBL-EBI (Helen Parkinson, Alvis Brazma, Ugis Sarkans)
- Francis Crick (Lucy Collinson, Raffa Carzaniga, Peter Rosenthal)
- Osaka University (Genji Kurisu)
- MRC-LMB (John Briggs, Paula da Fonseca, Wanda Kukulski, Garib Murshudov)
- RCSB (Cathy Lawson)
- STFC (Martyn Winn)
- University of Manchester (Alan Roseman)
- And many more...!

Websites

- emdb-empir.org
- empir.org
- pdbe.org
- emdatbank.org

Data: What data is found where...



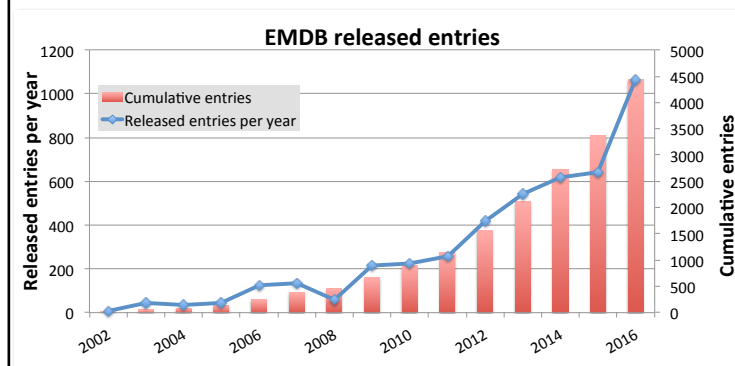
Comparison of archives

	PDB	EMDB	EMPIAR
Inception	1971	2002	2014
Number of entries	133589 (2297 EM)	5163 (1068 in 2016)	84
Archive size	1 GB	489 GB	48 TB
Mandatory deposition	Coordinates (1989) Structure factors (2008)	Single particle and sub-tomogram averaged structures (2012) Representative tomogram strongly recommended	-

EMDB Deposition Discussion at 3DEM GRC 2017

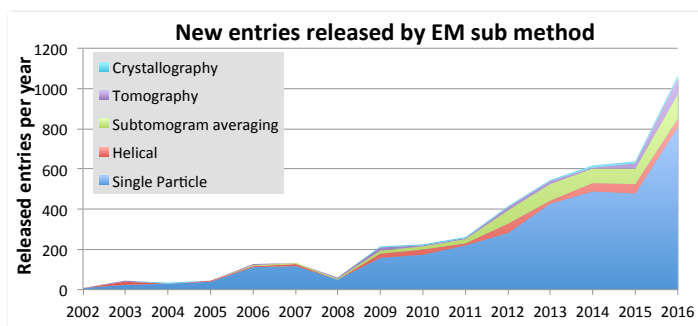
- Following changes were voted on with overwhelming majority (few abstentions and none against)
 - For all tomography-based studies, the deposition of representative tomographic reconstructions should be made mandatory. All tomographic reconstructions that are depicted in published figures (as orthoslices, renderings or similar), should be deposited
 - The deposition of un-modified half-maps should be mandatory in cases where half-maps have been used to assess resolution

EMDB trends: released entries



EMDB trends: EM sub-methods

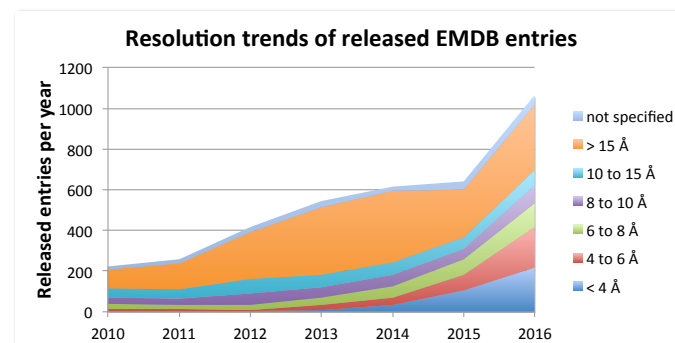
- ~20% from tomography in 2016 - will growth in tomography be sustained?



EMBL-EBI

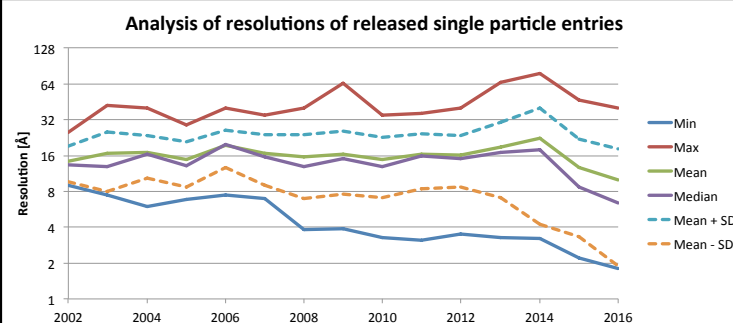
EMDB trends: resolution

- Strong growth of better than 4Å structures



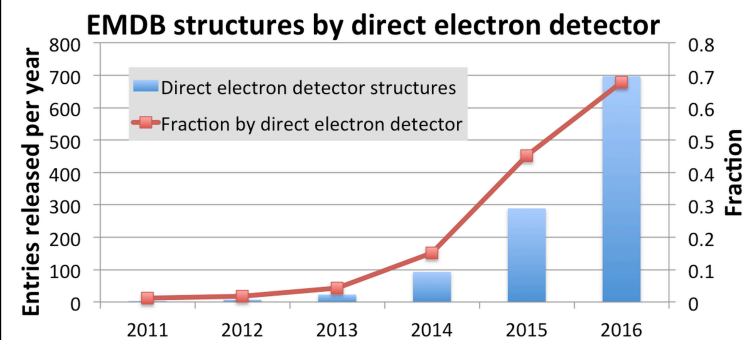
EMBL-EBI

EMDB trends: resolution of single-particle entries



EMBL-EBI

EMDB trends: direct electron detectors



EMBL-EBI

EMDB trends: geographic spread

- Analysis of author affiliations of EMDB-associated publications
- USA, Germany, UK continue to lead
- Strong growth in China
- Many more countries up and coming
- Patwardhan A., *Trends in the Electron Microscopy Data Bank (EMDB)*, Acta D 73(6), 503-508

Country	2010	2011	2012	2013	2014	2015	2016	Sum
USA	55	50	86	116	139	135	162	743
Germany	19	25	20	22	42	53	53	234
UK	14	10	27	16	42	51	62	222
China	3	4	3	10	13	29	39	101
France	8	4	6	8	16	16	23	81
Japan	6	7	4	6	11	11	20	65
Spain	4	7	11	7	8	7	10	54
Canada	1	5	3	2	10	14	17	52
Switzerland	2	2	8	8	6	15	11	52
The Netherlands	0	1	1	3	9	15	16	45

EMBL-EBI

EMDB trends: <4A structures by organization

- Stats on author affiliations of EMDB related publications with high res structures (<4Å)
- Organisations with 5 or more structures in past 3 years shown

Institution	2014	2015	2016	Sum
MRC Laboratory of Molecular Biology	8	14	15	37
Tsinghua University	-	7	10	17
Howard Hughes Medical Institute	1	6	5	12
University of California San Francisco	-	5	4	9
Chinese Academy of Sciences	1	3	4	8
Harvard University/Medical School	2	4	2	8
National Cancer Institute	1	2	5	8
University of California Los Angeles	-	7	1	8
University of Munich	2	3	3	8
Purdue University	1	1	5	7
University of Texas	3	3	1	7
Columbia University	-	2	4	6
ETH	1	3	2	6
University of California Berkeley	1	1	4	6
University of Virginia	2	3	1	6
Max Planck Institute for Biochemistry	-	-	5	5
Max Planck Institute for Biophysical Chemistry	-	1	4	5
The Rockefeller University	-	-	5	5
University of Oxford	-	3	2	5
University of Washington	-	3	2	5

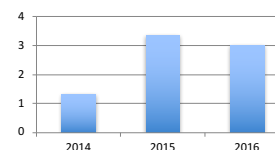
EMPIAR Raw image data archive for EM

- Started in 2014 – strong community support
- Control growth by requiring deposition of 3D to EMDB
- Usage: validation, development, testing, teaching and...
 - Safe storage of your data!
 - Was source for data in EMDatabank Map Validation Challenge
- Multi-frame micrographs, averaged micrographs, particle-stacks, tilt series
- Uses Aspera, Globus, ftp, http for data transfers

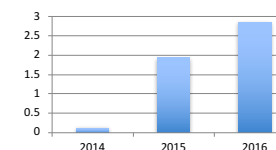
EMPIAR metrics

- Number of entries: 84 (~50TB; average size ~ 600GB)
- 12 datasets greater than 1 TB; one 10TB+ dataset
- Transfer speed: uploads 1-2 TB/24h (Europe, US, Australia)
- “empiar” cited 50+ times in full-text open-access papers
- Nature Methods publication (Iudin et al., 2016)

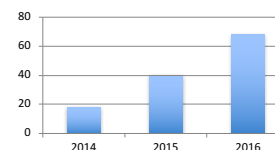
Aspera uploads/month (users)



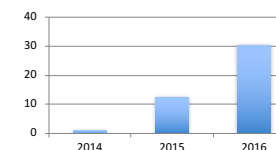
Aspera uploads/month (TB)



Total downloads (users)

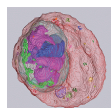


Total downloads (data)

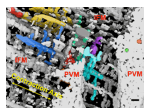


Beyond molecular cryo-EM

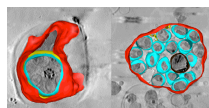
- We have already expanded EMPIAR beyond molecular cryo-EM
 - EMPIAR-10052/3/4/5: SBF-SEM
 - EMPIAR-10070: FIB-SEM
 - EMPIAR-10087: soft X-ray tomography
- Discussion with integrative hybrid modelling community on deposition of related EM data
- EMBL-EBI Bioimaging Workshop (Jan 2017)
- Expert workshop on data models for EMPIAR for cellular imaging in December 2017



EMPIAR-10054



EMPIAR-10070



EMPIAR-10087

Searching EMDb

Latest entries

Quick access

The Electron Microscopy Data Bank (EMDB) at PDBe

Quick access

Click on one of these categories:

Ribosome Virus Phage GroEL Microtubule Polymerase Helicase

Human HIV Entries with Single particle Tomography Helical <3Å resolution fitted models reconstruction

or enter 4-digit EMDb entry number: [1001] [Entry summary](#) [Visual analysis of map](#) [Volume viewer](#)

Form search/browse

Introduction

The Electron Microscopy Data Bank (EMDB) is a public repository for electron microscopy density maps of macromolecular complex techniques, including single-particle analysis, electron tomography, and electron (2D) crystallography. The EMDb was founded at EBI in 2002, under the leadership of Kim Henrick. Since 2007 it has been operated jointly by the PDBe Bioinformatics (RCSB PDBe) as a part of EMDatabank which is funded by a joint NIH grant to PDBe, the RCSB and the National Center

emdb-empiar.org

EMBL-EBI

Searching EMDb – browse the archive

Filter results 2377 entries

EM method	Component type	Organism	Author	EMDB
Single-particle (1852 / 1852)	Protein (1449 / 1449)	Homo sapiens (532 / 532)	Chiu W (143 / 143)	REL (22)
Subtomogram averaging (253 / 253)	Virus (578 / 578)	Escherichia coli (360 / 360)	Carragher B (107 / 107)	HPUB (1)
Helical (173 / 173)	Cell component (220 / 220)	Saccharomyces cerevisiae (232 / 232)	Potter CS (104 / 104)	HOLD1
Tomography (69 / 69)	Prokaryotic ribosome (208 / 208)	Bos taurus (88 / 88)	Rossmann MG (90 / 90)	
2D crystallography (30 / 30)	Nucleic acid (189 / 189)	Mus musculus (78 / 78)	Saibil HR (82 / 82)	
	Eukaryotic ribosome (101 / 101)	Synthetic construct (78 / 78)	Nogales E (78 / 78)	
	Ligand (60 / 60)	Unidentified (75 / 75)	Frank J (77 / 77)	
	EM label (1 / 1)	Rattus norvegicus (64 / 64)	Beckmann R (67 / 67)	
		Human immunodeficiency virus 1 (55 / 55)	Liu X (59 / 59)	
		Dryctolagus cuniculus (49 / 49)	Butcher SJ (56 / 56)	

Reset filters Update search results

Sort by: [Map release date] Download as text Download as XML

Showing results 1 - 25 of 2377

EMD-2389: The limits of structural plasticity in a picornavirus capsid revealed by a massively expanded equine rhinitis A virus particle (Bakker SE, Groppe E, Pearson AR, Stockley PG, Rowlands DJ, Ranson NA)

Deposition date	Header release	Map release	Latest update	Status	EM method	Symmetry	Resolution	Molecular weight
30 May 2013	3 Jul 2013	2 Apr 2014	2 Apr 2014	REL	Single-particle	1	17Å	N/A

Sample: Equine rhinitis A virus (ERAV, virus from *Equus caballus*)

Microscope model: FEI Tecnai F20 Electron source: FIELD EMISSION GUN Acceleration voltage: 200kV

emdb-empiar.org/embrowse

EMBL-EBI

Form based search

- Enumerations for software and microscopy recently updated
- Search by detector model

Software

2dx Amira Appion AUTO3DEM

Bsoft CCP4 CNS Chimera

CTFFIND3 CTFFIND4 CTFFIND Dynamo

EM3DR EMAN EMAN 1 EMAN 2

FREALIGN IHRSR Imagic IMOD

JSPR Jsubtomo Legimon MRC

Relion ResMap Scipion SIMPLE

SPARX Spider Spire Spring

TIGRIS TOM XMIPP XPLORE

Specify

Detector

Falcon Direct Electron Gatan K2

http://emdb-empiar.org/emsearch

EMBL-EBI

EMStats – EMDB statistics

- Dynamically generated charts based on the current state of the archive

Charts

General

Map releases
Map size statistics for released maps
Distribution of maps released based on EM technique used
Trends and distribution of microscope usage for released maps
Trends and distribution of software package usage for released maps
Molecular weight statistics of single-particle released maps
Sample taxonomy statistics for released maps

Resolution

Trend of highest resolutions achieved annually for released maps
Number of released maps achieving given resolution levels
Single-particle released maps – resolution trends
Single-particle released maps – resolution versus acceleration voltage and source type
Single-particle released maps – resolution versus microscope type and software package used
Single-particle released maps – resolution versus number of projections
Tomography – resolution trends for released maps

Publications

Trend for publications associated with released maps
Journal distribution of articles associated with released maps
Publication trends and distribution for different microscope types
Publication trends and distribution for different software packages

FTP downloads

Monthly FTP downloads
Top 10 FTP downloads

EMPIAR

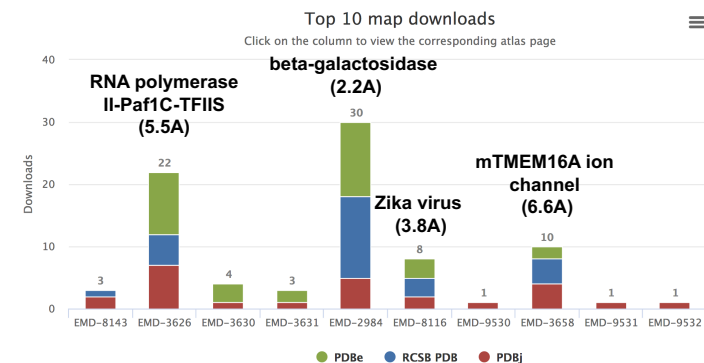
Entry releases
Entry size statistics for released entries
EMPIAR transfer statistics

emdb-empiar.org/emstats

EMBL-EBI

EMStats – map downloads

- Downloads stats for the past two years



EMBL-EBI

EMDB entry pages

- Displays summary info about entry
- Download header file (extensive meta-data), map, FSC etc
- Visual analysis pages and visualisation options

EMDB › EMD-2852

Electron cryo-microscopy of mitochondrial ATP synthase dimers

Source organism: *Polytomella* [3049]

3DbioNotes: available for this entry

Related EMPIAR entry: [EMPIAR-10023](#)

Primary publication:

[1] Horizontal membrane-intrinsic alpha-helices in the stator α -subunit of an F-type ATP synthase

Allegretti M, Klusach N, Mills DJ, Vonck J, Kuehlinbrandt W, Davies KM

NATURE 521 237-240 (2015)

PMID: 25707805

Function and Biology

Sample name: Polytomella ATP-synthase

Protein: Mitochondrial F-type ATP-synthase

Experimental Information

Resolution: 7.0Å

Resolution method: FSC 0.143, gold-standard

Applied symmetry: C2

Single particle reconstruction
7.0Å resolution

Map released:

2015-03-04

Last modified:

2015-05-13



Quick links

EMD-2852 overview

Function and Biology

Experiments and Validation

View

Downloads

Volume viewer

Volume slicer

Visual analysis

Related entries

By authors

By sample

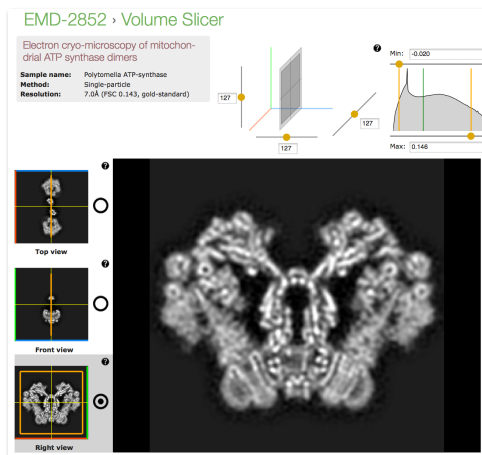
By organism

emdb-empiar.org/emd-2852

EMBL-EBI

Volume slicer

- Orthogonal slices in three directions
- Interactive 3D navigation

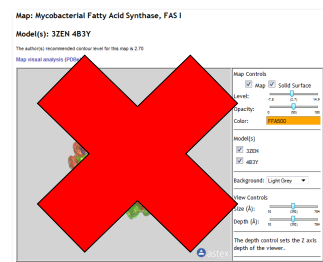


emdb-empiar.org/emd-2852/browse

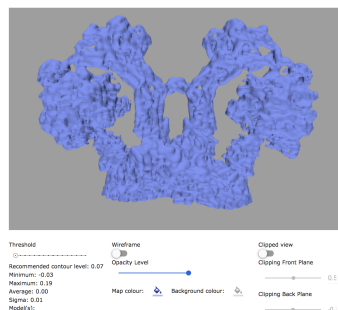
EMBL-EBI

3D interactive viewer

- Web GL based 3D viewer on the way...



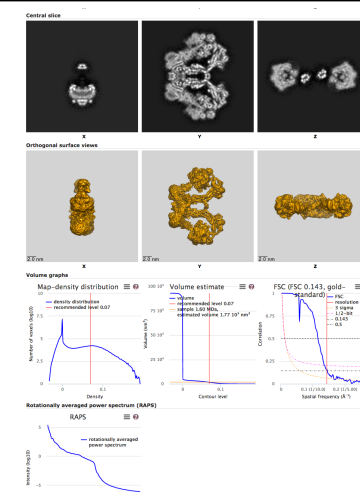
EMD-2852 - Volume viewer



EMBL-EBI

Visual analysis

- Basic sanity checking info
- FSC curves shown if available



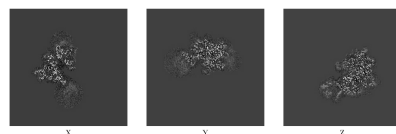
emdb-empir.org/emd-2852/analysis

EMBL-EBI

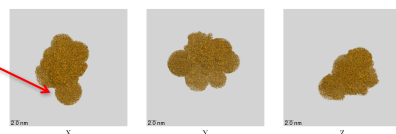
Visual analysis: images

- Surface rendering shows that contour level is too low

5.3 Central slices



5.4 Orthogonal surface views

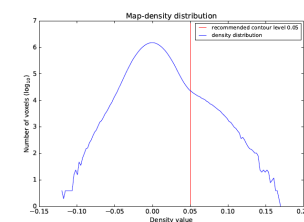


Contour level defines the mask!!

Visual analysis: map density distribution

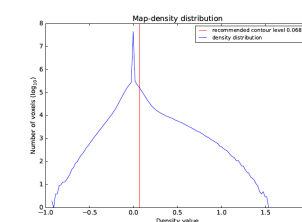
- Masking comes up as sharp spike ~0
- Check for artefacts

5.5 Map-density distribution



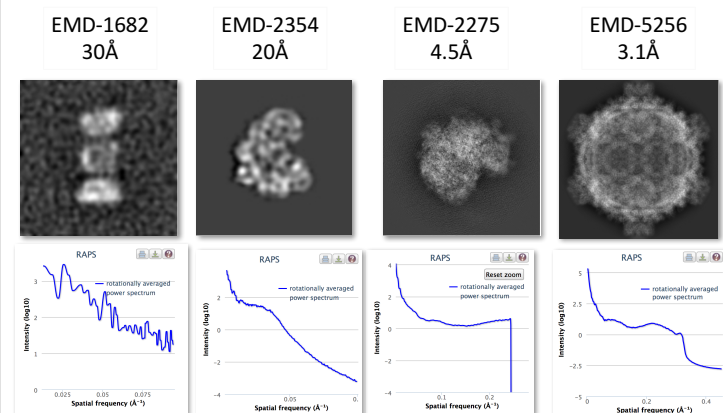
Unmasked map
beta galactosidase
EMD-2984

5.5 Map-density distribution



Masked map
ribosome
EMD-2566

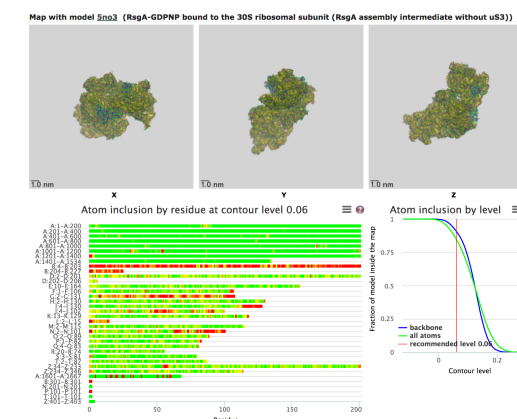
Analysis: Rotationally averaged power spectra (RAPS)



EMBL-EBI

Visual analysis: map-model fit

- How well does model fit into map density?



RsgA-GDPNP bound to the 30S ribosomal subunit (RsgA assembly intermediate without uS3)

EMD-3662/5no3

EMBL-EBI

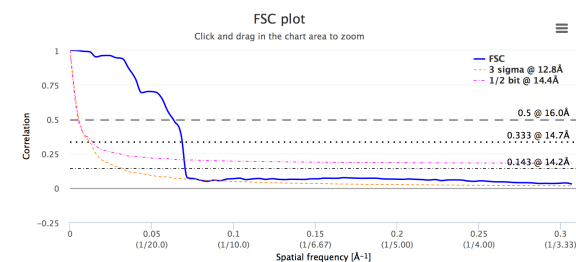
Stand-alone validation servers

- Convenience
- Promote use and reporting
- Work with software developers to develop standards for capturing validation data

EMBL-EBI

Fourier shell correlation (FSC) server

- User uploads two maps -> server calculates FSC curve
- FSC curve can be deposited to EMDB
- Automated calculation of thresholds
- Bsoft, Relion and EMAN2 generate FSC curves that can be deposited



emdb-empair.org/fsc

EMBL-EBI

Tilt-pair validation server

- Developed with Rosenthal lab
- Deposition of tilt-pair data to EMDB still to be developed
- Requested features – read Imagic files, Relion angles

Tilt pair validation server (beta)

Welcome to the PDBe tilt pair validation server!
Tilt-pair validation analysis (Rosenthal and Henderson, 2003) can be used to assess the accuracy of initial a you need to collect two corresponding sets of particle images - one untilted and the other tilted, then upo untilted images. This server is based on the Tilt-pair server developed at MRC National Institute for Medical Research and Peter Rosenthal for their help in developing and testing the current server.
You may upload map files in MRC or CCP4 format, and parameter files (containing Euler angles for individual p you can use to try out the service [here](#). We are still developing the server and appreciate your feedback!

Map (3D volume) No file selected.

Untilted stack No file selected.

Orientation parameters for stack 1 No file selected.

Tilted stack No file selected.

Pixel size (Å)

Mask radius (pixels)

Tilt search range (degrees)

Resolution range (low to high; Å)

Email address

Job name

Perform CTF correction? ☒

Microscope voltage (kV)

Defocus (Å)

Spherical aberration (mm)

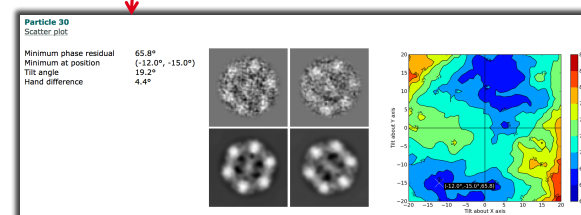
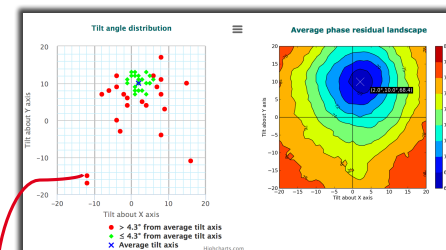
Amplitude contrast

pdbe.org/tiltpair

EMBL-EBI

Tilt-pair results

- Summary result
- Per particle result
- Interactive scatter plot



EMBL-EBI

ResMap local resolution server

- Developed and hosted by Carazo group in Madrid
- Linked to from our validation page

Resmap Online

This page allows you to execute ResMap (<http://resmap.sourceforge.net/>) from the web. ResMap is an easy to use software package for computing the local resolution of 3D density maps studied in structural biology, primarily electron cryo-microscopy (cryo-EM).

Choose your project

- Create an empty project
- Create a project with test data
- Project with Test data (read-only)

PROJECT

Import volumes finished

resmap - local resolution saved

Input

Use half volumes? ☒ Yes ☐ No

Volume half 1

Volume half 2

Mask input volume? ☐ Yes ☒ No

It is strongly recommended to use the pre-whitening wizard.

Pre-whitening Angstroms Ramp

Extra parameters

Step size (Ang):

Resolution Range (Å) Min Max

Confidence level:

Run

local: scipion - import volumes

Use queue? ☐ Yes ☒ No

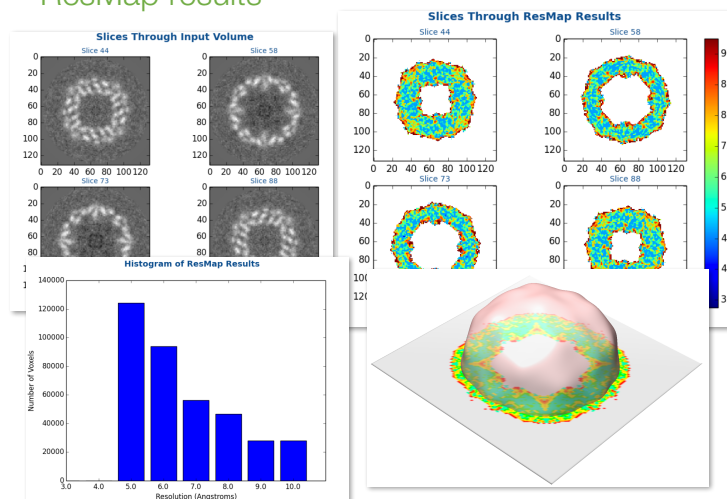
Import

Files directory

Pattern

Pixel size ("sampling rate") (Å/pix)

ResMap results



EMDB API

- Programmatically access EMDb meta-data
- Search API – will be discussed later in tutorial
- e.g., ebi.ac.uk/pdbe/api/emdb/entry/all/EMD-6007

```

    },
    "imaging": [
      {
        "electronSource": "FIELD EMISSION GUN",
        "imagingMode": "BRIGHT FIELD",
        "illuminationMode": "SPOT SCAN",
        "specimenHolder": "Eucentric",
        "microscope": "FEI/PHILIPS CM200FEG",
        "specimenHolderModel": "GATAN LIQUID NITROGEN",
        "acceleratingVoltage": {
          "value": 160.0,
          "units": "kV"
        }
      }
    ],
    "imageAcquisition": [
      {

```

EMPIAR website - empiar.org

- Search, browse EMPIAR entries and download and deposit data

EMPIAR Electron Microscopy Public Image Archive




EMPIAR home | Deposit | REST API | FAQ | About EMPIAR

EMPIAR, the Electron Microscopy Public Image Archive, is a public resource for raw, 2D electron microscopy images. Here, you can browse, upload, and download and reprocess the thousands of raw, 2D images used to build a 3D structure. [More ...](#)

Deposit your data in EMPIAR to share it with the structural biology community.

Browse and [download](#) EMPIAR datasets using the table below.

Show 50 entries

Dataset	Title	Authors	Related EMDs/PDB entries	Size
	The Structure of the Yeast Mitochondrial Ribosome (stack of 2524 particles in HR23 format)	Desai N, Brown A, Amunts A, Ramakrishnan V [PubMed: 28154081] [DOI: 10.1126/science.124115]	EMD-3551, EMD-3552, EMD-3553, EMD-3554, EMD-3555, EMD-3556, Smr, Smr, Smr	138.6 GB
	40 Degree Tilted Single-Particle CryoEM of Highly Preferred Orientated Influenza Hemagglutinin Trimer (multiple data sets in HRC and ASSORTED formats)	Tan YZ, Lyumkis D		1.8 TB
	Untilted Single-Particle CryoEM of Highly Preferred Orientated Influenza Hemagglutinin Trimer (multiple data sets in HRC and ASSORTED formats)	Tan YZ, Lyumkis D		1.2 TB

EMPIAR citations

[Parasitophorous vacuole poration precedes its rupture and rapid host erythrocyte cytoskeleton collapse in Plasmodium falciparum egress.](#)
Hale VL et al. (2017)

[Accurate model annotation of a near-atomic resolution cryo-EM map.](#)
Hryc CF et al. (2017)

[Using the volta phase plate with defocus for cryo-EM single particle analysis.](#)
Danev R et al. (2017)

[Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2.](#)
Kimanius D et al. (2016)

[Structural basis for dynamic regulation of the human 26S proteasome.](#)
Chen S et al. (2016)

[See all citations?](#)

EMBL-EBI

EMPIAR entry pages

- Browse more info on entry
- View thumbnail images before downloading
- Select images to download

EMPIAR-10030
Cryo-EM micrographs of microtubules in GDP-state obtained by cryopolymerization with ES3

Publication
Mechanistic Origin of Microtubule Dynamic Instability and Its Modulation by ES Proteins
Zhang G, Alvarado G, Brown A, Nogales E
Cell (Cambridge, Mass.) 162 849-859 (2015)
PMID: 26234125
DOI: 10.1016/j.cell.2015.07.022

Related PDB entry:
5J51

Released:
9 Jul 2015

Released:
12 Aug 2015

Last modified:
12 Aug 2015

Dataset size:
426.3 GB

Dataset DOI:
10.6026/EMPIAR-10030

Image sets

Unaligned multi-frame micrographs of microtubules in GDP-state

Category:
micrographs - multiframe

Image format:
HDF5

No. of images or IFF series:
383

Frames per image:
32

Image size:
(3715, 3715)

Pixel type:
16-bit float

Pixel spacing:
(1.32 Å, 1.32 Å)

Dataset:
Unaligned, movies 426.3 GB

Download

Thumbnail image of a micrograph showing microtubules.

empiar.org/empiar-10030

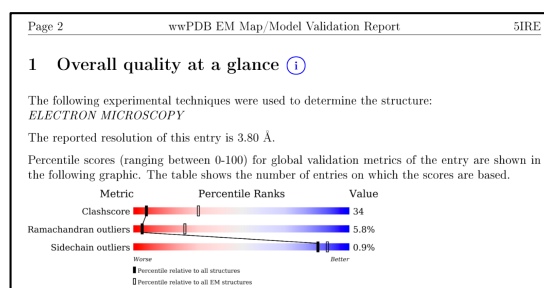
EMBL-EBI

Planned developments

EMBL-EBI

EM validation reports

- In production: mainly info relating to EM models
- Validation report for PDB EM models calculated and available from ftp sites

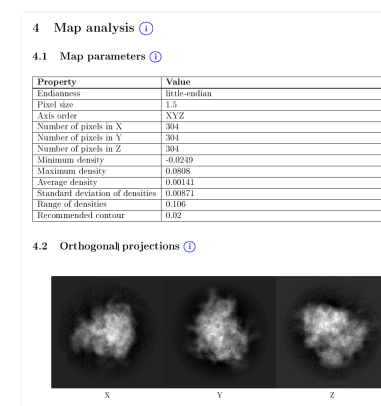


pdbe.org/5ire

EMBL-EBI

EM validation reports – map validation

- Based on EMDB Visual analysis pages (available for released entries)
- Limited by data available – mostly sanity checking
- In testing

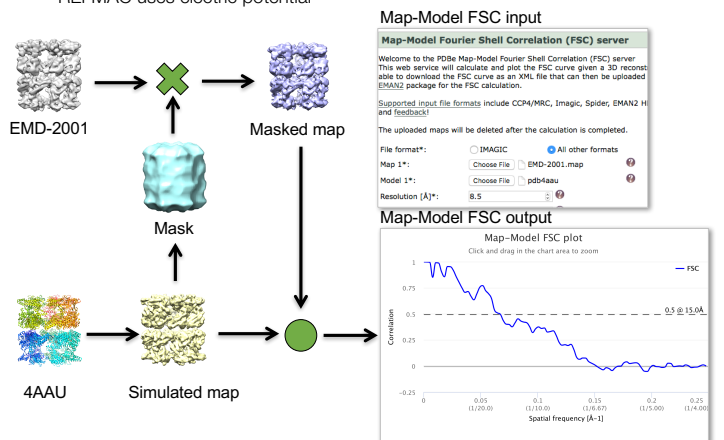


emd-5801

EMBL-EBI

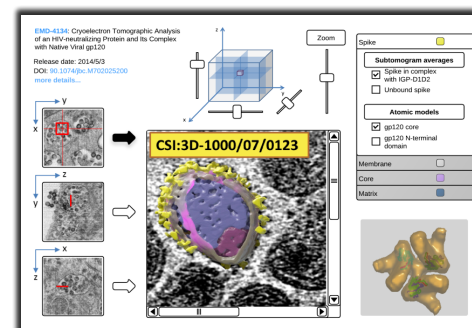
Map/Model FSC server using EMAN2/REFMAC

- EMAN2 generates map from model using Gaussian blurring function
- REFMAC uses electric potential

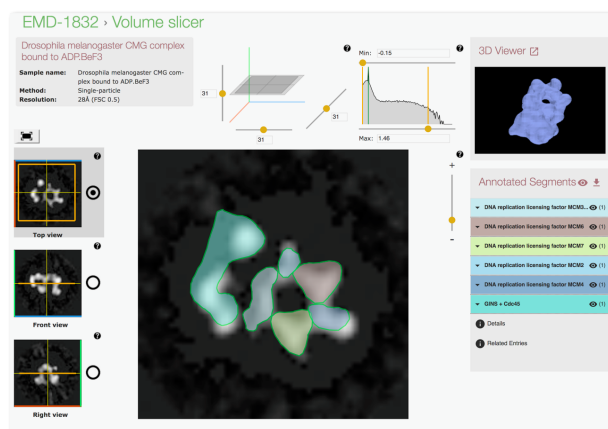


Volume browser – the idea

- Integrated visualisation of structural data
- Spanning scales from cells to molecules



Volume browser – the reality (almost there!)



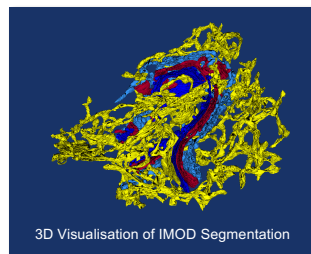
EMDB-SFF (Segmentation File Format)

- Expert workshop on “3D segmentations and transformations - building bridges between cellular and molecular structural biology”, Dec 2015
- Manuscript submitted to eLife
- Working group to advise on development
- EMDB-SFF supports structured biological annotation linking segments to ontologies and other bioinformatics resources
- Implementations in XML, HDF5 and JSON
- Also capable of supporting transforms between subtomogram averages and tomograms



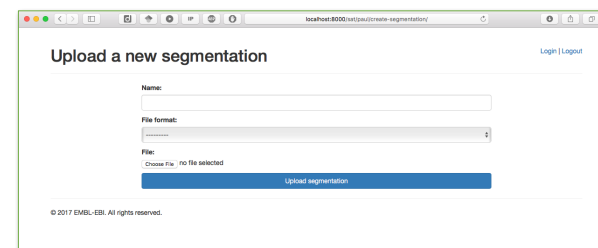
EMDB-SFFTK (toolkit)

- Available from CCP-EM SVN: <http://bit.ly/2oQnvxd>
- Contains code for manipulating EMDB-SFF
 - Translators support external formats
 - IMOD
 - Segger
 - Amira (AmiraMesh and Amira HxSurface)
 - STL
 - EMDB MAP masks
 - Annotation utilities
 - Search ontologies (OLS hosted at EBI)
 - View, add, edit, delete annotations
- Output formats: XML, HDF5, JSON
- Create and load ROIs into OMERO

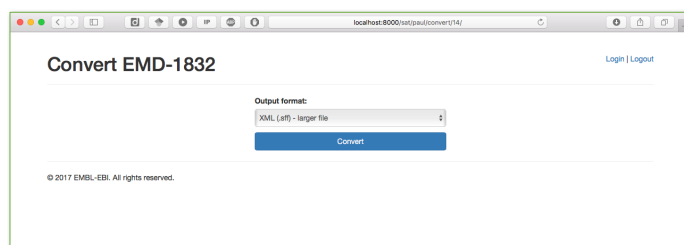


Segmentation annotation tool (SAT)

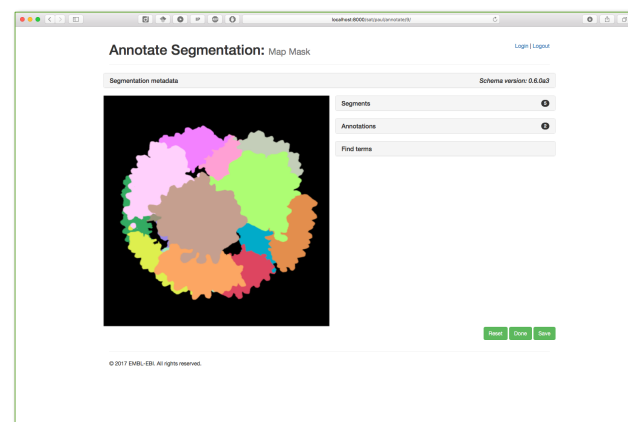
- Web based service
- Desktop tool



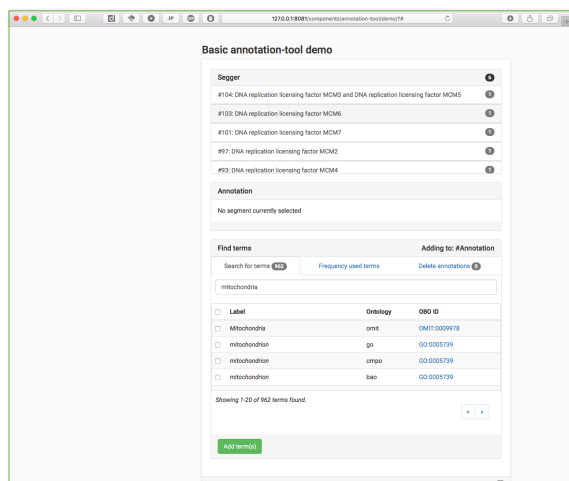
SAT – conversion to EMDB-SFF



SAT



SAT – annotation via OLS (Ontology Lookup Service)



Future outlook

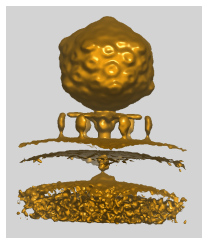
- Crowd sourced annotation of segmentation
- Archiving for related imaging modalities including – 3DSEM, SXT and CLEM
- Data-harvesting and submission pipelines to EMDB and EMPIAR
- Visualisation of and standards for correlative imaging and for relating sub-tomogram averages to tomograms
- Validation
- Automated segmentation – deep learning
- Fast archive-wide sub-structure volumetric (or shape-based) searches

References

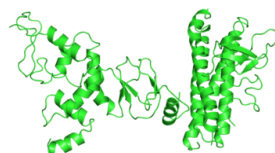
1. Patwardhan A., *et al.* "Data management challenges in three-dimensional EM". *Nat Struct Mol Biol*, **19**, 1203-1207 (2012).
2. Gutmanas A., Oldfield T., Patwardhan A., Sen S., Velankar S. and Kleywegt G. "The role of structural bioinformatics resources in the era of integrative structural biology". *Acta Cryst*, **D69**, 710-721 (2013).
3. Lagerstedt I., Moore W.J., Patwardhan A., Sanz-Garcia E., Best C., Swedlow J.R. and Kleywegt G.J. "Web-based visualisation and analysis of 3D electron-microscopy data from EMDB and PDB". *J Struct Biol*, **184**, 173-181 (2013).
4. Patwardhan A., *et al.* "A 3D cellular context for the macromolecular world". *Nat Struct Mol Biol*, **21**, 841-845 (2014).
5. Wood C., Burnley T., Patwardhan A., Scheres S., Topf M., Roseman A. and Winn M. "Collaborative computational project for electron cryo-microscopy". *Acta Crystallogr D Biol Crystallogr*, **71**, 123-126 (2015).
6. Iudin A., Korir P.K., Salavert-Torres J., Kleywegt G.J. and Patwardhan A. "EMPIAR: A public archive for raw electron microscopy image data". *Nature Methods*, **13**, 387-388 (2016).
7. Patwardhan A. and Lawson C.L. "Databases and Archiving for CryoEM". *Methods Enzymol*, **579**, 393-412 (2016).
8. Salavert-Torres J., Iudin A., Lagerstedt I., Sanz-Garcia E., Kleywegt G.J. and Patwardhan A. "Web-based volume slicer for 3D electron-microscopy data from EMDB". *J Struct Biol*, **194**, 164-170 (2016).
9. Patwardhan A. "Trends in the Electron Microscopy Data Bank (EMDB)". *Acta Crystallographica Section D*, **73**, (2017).

How to deposit?

EMDBspeak 1



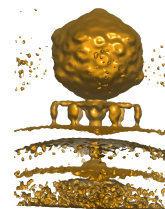
Map: 3D EM reconstruction
(EM volume)



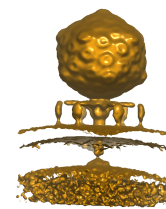
Model: description of atomic coordinates

EMDBspeak 2

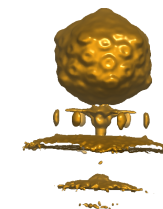
- **Contour level:** Density level at which the surface of molecule is defined
- Important for visual interpretation
- Important for many calculations, e.g., weight, volume, atom inclusion



0.03



0.25



1.0

EMDBspeak 3

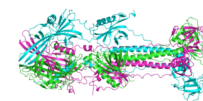
- **EMDB header file:** XML file with metadata describing how map was obtained

```
<Imaging>
  <astigmatism>objective lens astigmatism was corrected at 120,000 times magnification</astigmatism>
  <electronSource>FIELD EMISSION GUN</electronSource>
  <electronDose>units="e-/Å²" 16</electronDose>
  <ImagingMode>BRIGHT FIELD</ImagingMode>
  <nominalDefocusMin>units="nm" 1000</nominalDefocusMin>
  <nominalDefocusMax>units="nm" 3000</nominalDefocusMax>
  <illuminationMode>FLOOD BEAM</illuminationMode>
  <detector>Direct Electron DE-12</detector>
  <nominalCs>units="mm" 2</nominalCs>
  <tiltAngleMin>units="degrees" 0</tiltAngleMin>
  <calibratedMagnification>42134</calibratedMagnification>
</microscope> FEI TECNAI F20</microscope>
```

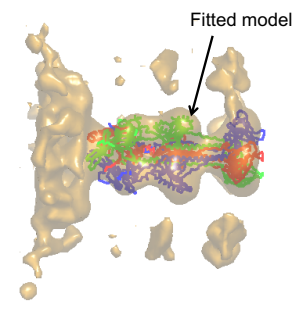
- Each EMD entry consists (at least) of 1 header file + 1 map file
- Map files in **EMDB map format** (EMDB interpretation of CCP4 format)
- Formats: emdb-empiar.org/emschema

EMDBspeak 4

- **Fitting model:** model that will be shifted, rotated and possibly flexed to fit into EM map
- **Fitted model:** model that lies in frame with the map



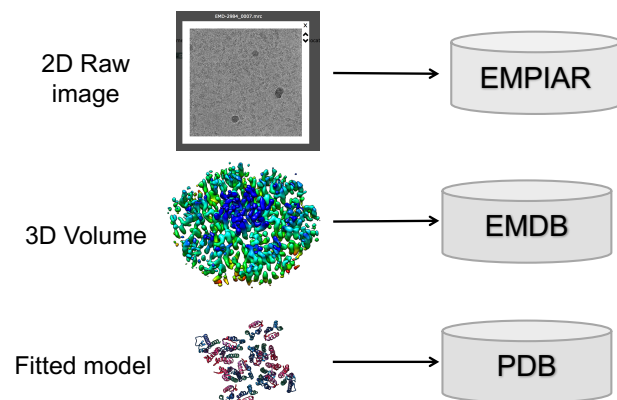
Fitting model



Fitted model

Data: What data goes where...

- Final single-particle and sub-tomogram average maps must go to EMDB (tomograms strongly recommended)
- Fitted/built models must go to PDB



From deposition to release

- You start by filling out a form and uploading the map and other data
- When you submit your deposition, it is locked and forwarded to the annotation team; you are assigned a deposition ID and an accession code
- The annotation team will check your deposition and will communicate with you regarding changes and clarifications
- Once annotation is completed, and you agree -> EMDb header file is released
- Map can be released immediately, kept on-hold until publication, or on-hold for 1 year
- “Immediate” is not immediate – weekly release cycle!

How to deposit?

- Some preparations
 - which EM maps? you can upload compressed maps
 - which fitted models? check that they overlap with the maps
 - note: each deposition can only accommodate as a maximum 1 map and 1 model
- contour level
- voxel size
- what is the sample composition?
- dig out your lab notebooks so that experimental info is readily available
- Check out the 12 golden rules and other help documents on emdatabank.org -> deposit
- Go to the OneDep system (<https://deposit.wwpdb.org/deposition/>)
 - When you select your country you will be directed to the correct site (PDBe, RCSB PDB or PDBj)

Deposition: submission options

- Chose EM sub-method carefully as this will determine options presented in deposition
- Tell the annotator if this is a part of a multipart deposition

The screenshot shows the deposition form with the following fields and annotations:

- Your e-mail address:** ardan@ebi.ac.uk
- Password (optional, or we will provide one):** This is a shared "group password" (8 to 16 alphanumeric characters)
- Country:** United Kingdom (with a "Reset country" button)
- Experimental method:**
 - ☒ X-Ray Diffraction
 - ☒ Electron Microscopy
 - ☐ Helical
 - ☒ Single particle (Annotated with a red box and arrow: **EM sub-method**)
 - ☐ Subtomogram averaging
 - ☐ Tomography
 - ☐ Solution NMR
 - ☐ Neutron Diffraction
 - ☐ Electron Crystallography
 - ☐ Solid-state NMR
 - ☐ Fiber Diffraction
- Are you depositing coordinates with this submission?**
 - ☐ No, experimental data only
 - ☒ Yes (Annotated with a red box and arrow: **Map only or map+model?**)
- Has the associated map been deposited previously?**
 - ☐ No
 - ☒ Yes
- Requested accession codes:**
 - ☐ PDB
 - ☒ EMDb
 - ☐ BMRB

Deposition: file upload I

- Maps may be compressed: '.zip', '.gzip' or '.bz2'
- Maps larger than 1.5Gb -> ftp
- All maps converted to EMDB map format -> check that converted map and fitted models overlap!!
- Allowed formats: CCP4/MRC
- You are also asked to upload representative image (mandatory) that will be used on the website etc

The screenshot displays the EMAP GUI with two map configuration panels. The top panel, titled 'Primary map (final published map: mandatory)', shows a pixel spacing of 1.7, a contour level of 2, and a short description of 'final reconstruction, filtered and masked'. The bottom panel, titled 'Additional maps (e.g. unfiltered or different minor states; optional)', shows a pixel spacing of 1.6, a contour level of 3, and a short description of 'Unmasked and unfiltered map'.

Deposition: file upload II

- Half maps, masks, fsc curves can be uploaded as specific file types (optional but strongly recommended)

	File Name	Size	Properties / Description
Half-maps	<code>emcd122_Ferritin_even.mrc</code>	9.20 MB	EM half map (MRC/CCP4 format) Pixel spacing (\AA): 1.7 Contour level*: 2 Short description: Half map even
	<code>emcd122_Ferritin_odd.mrc</code>	9.20 MB	EM half map (MRC/CCP4 format) Pixel spacing (\AA): 1.7 Contour level*: 2 Short description: Half map odd
Masks	<code>emcd122_Ferritin_mask.mrc</code>	9.20 MB	EM mask (MRC/CCP4 format) Pixel spacing (\AA): 1.7 Contour level*: 1 Short description: Mask applied to obtain final map
	<code>emd_2530_fac.xml</code>	3.95 KB	FSC file (XML format)

FSC curves

Check output of file upload I

- Download converted maps/models and check overlays etc

Upload file summary

You uploaded 7 files to the system.

Number	Used	File name	Size	File type	File header check
1	*	e2map.mrc	8389632	Main volume data	The file has correct format
2	*	stack1.mrc	3277824	Additional volume data	The file has correct format
3	*	emd122_ferritin_xen.mrc	9200896	Half volume data	The file has correct format
4	*	emd122_ferritin_sdb.mrc			Correct format
5	*	emd122_ferritin_msa.mrc			Correct format
6	*	emd_2530_fsc.xml			Join an FSC XML
7	*	fsc1.tif			Join an image for EMD8

Contour level must lie within density range

Map conversion report -- e2map.mrc

Errors:

- Provided contour level is outside the density range of the map.

Check output of file upload II

- Again conversion may shift map etc. so download converted maps/models and check overlays etc

Half map conversion report -- emc122_Ferritin_even.mrc

Warnings:

- Map origin has been automatically shifted:
 - X: from 0.01899113 to 0 pixel units
 - Y: from 0.01899113 to 0 pixel units
 - Z: from 0.01899113 to 0 pixel units

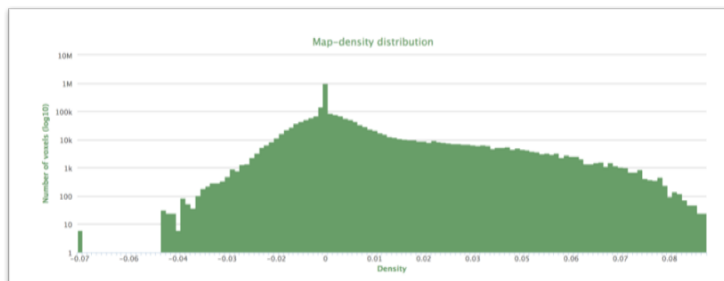
In order to avoid loss of precision in map position upon conversion to EMDataBank format, MRC origin position values (in Angstrom) should be clearly divisible by the map pixel sizes (in Angstrom/pixel).

It is recommended that you download and inspect the converted map. Click on 'Communication' in the navigation menu to inform curators about map conversion problems.

Map title:	Converted map	Original map
Map ends/naness:	Big endian	2015-16 EMDataBank Map Challenge-label removes for blind assessment EMAN 4/28/2016 11:16
Map mode:	Image stored as floating point number (4 bytes)	Big endian
Fast, medium and slow axes:	X, Y, Z	X, Y, Z
Grid sampling on x, y, and z:	132, 132, 132	132, 132, 132
Pixel sampling on x, y, and z:	1.7, 1.7, 1.7	1.346, 1.346, 1.346
Cell dimensions (x, y, z, alpha, beta, gamma):	224.40001, 224.40001, 224.40001, 90.0, 90.0, 90.0	177.672, 177.672, 177.672, 90.0, 90.0, 90.0
Number of columns, rows, and sections:	132, 132, 132	132, 132, 132
Start points on columns, rows, and sections:	0, 0, 0	0, 0, 0
Origin in MRC format:	0.0, 0.0, 0.0	0.01899113, 0.01899113, 0.01899113
Space group number:	1	1
Minimum density:	-0.068754	-0.068754
Maximum density:	0.09492525	0.09492525
Average density:	3.4052873E-4	3.4052873E-4

Check output of file upload III

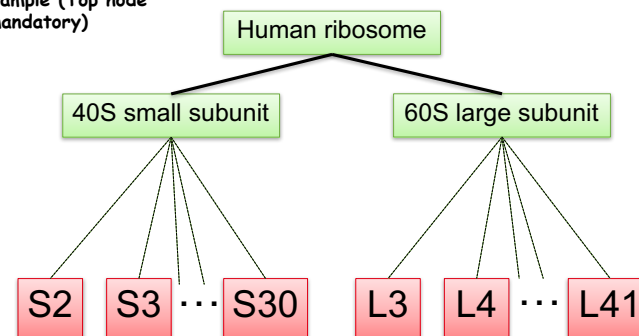
- Check that histogram looks like similar to that e.g., in Chimera
- Weird histograms indicate byte-swapping issues etc



Sample description

- Two tiered description to accommodate molecular and cellular EM

Sample (Top node mandatory)

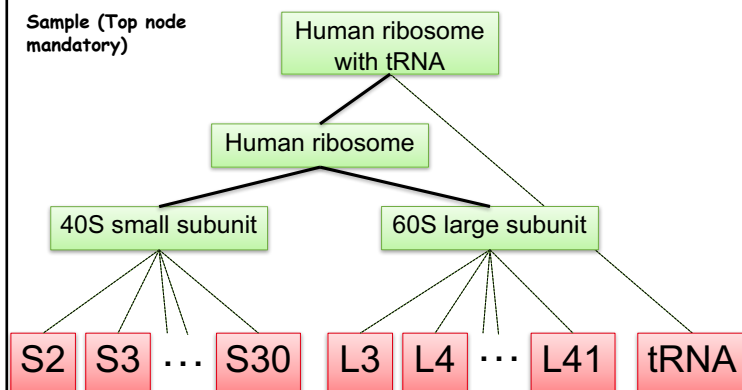


Macromolecules (optional for map-only depositions but recommended)

Sample description II

- Links can be from any sample node to macromolecules

Sample (Top node mandatory)



Macromolecules (optional for map-only depositions but recommended)

Software usage

- During the deposition there are several places where you can provide info about software used
- You can specify more than one package
- This information is very important for software developers

Software used to collect images		
Name	Version	Details
APPION	9.03	Used for initial data collec
SerialEM	1.2	Used for second round

EMPIAR deposition

- Go to empiar.org/deposition
- Read the deposition manual
- User based login – one user can have multiple depositions and users can share depositions
- Create login account or login using facebook/google/orcid

EMPIAR deposition system

[Deposition manual](#)

[Log in](#)

Login to the EMPIAR deposition system


Username:


Password:


[Register a new user](#)

[Can't log in?](#)

Sign in with:

 Google

 Facebook

 ORCID

EMPIAR deposition – user roles

- PI can, for example, grant access to student to fill out form but not submit

Deposition ID: 29

Show users

Search:

First name	Last name	E-mail	Organization	Permissions
TetsFNT	TestLNT	empiar.test2@gmail.com	Department of Crystallography, Institute of	<input type="radio"/> Transfer ownership <input type="radio"/> Choose rights to grant <input type="button" value="Apply"/>
TestU	Test	empiar.test4@gmail.com	Test org test	<input type="radio"/> Transfer ownership <input type="radio"/> Choose rights to grant <input type="button" value="Apply"/>
Helen	Salbil	empiar.test3@gmail.com	Department of Crystallography, Institute of Structural and Molecular Biology, Birkbeck College, University of London	<input type="radio"/> Transfer ownership <input checked="" type="radio"/> Choose rights to grant <input type="radio"/> View only <input type="radio"/> View + Edit <input type="radio"/> View + Edit + Submit <input type="button" value="Apply"/>

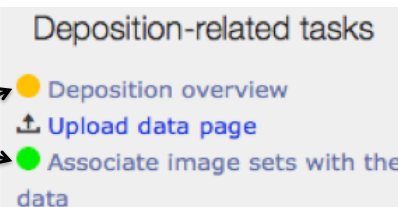
Showing 1 to 3 of 3 users

Previous Next

EMPIAR deposition workflow

- providing the general metadata about the deposition – EMDB accession code, title authors, etc.;
- uploading the data – the transfer can take some time, so the next step most likely would not be undertaken immediately;
- associating the uploaded data with the corresponding image sets – that is, identifying the image sets present and describing them.

Traffic lights indicate if mandatory items on page have been completed



Deposition overview

- For single particle EM and electron tomography, a link to EMDB entry is mandatory
 - This also means that we can extract some info from the related EMDB entry, e.g., authors
- Red star(*) indicates mandatory items
- For fields with an N/A button, you have to explicitly press the button if you have no value for the field

Surname: *	<input type="text" value="Smith"/> Example: Smith
Initials: *	<input type="text" value="JW"/> Example: JW
ORCID:	<input type="text" value="0000-0001-7663-9028"/> Example: 0000-0001-7663-9028

Release instruction

- **IMMEDIATELY:** As soon as the processing has finished the release procedure will be initiated. For datasets in the TB range this could take several days
- **EMDBPUB:** Wait till the associated EMD entry has been released before releasing the EMPIAR entry. This would require the public release by EMD of both the header and the map
- **HPUB:** Wait until the primary citation for the associated EMD entry has been published or max 1 year (whichever comes first)
- **HOLD1:** Wait for 1 year

Release instruction: *

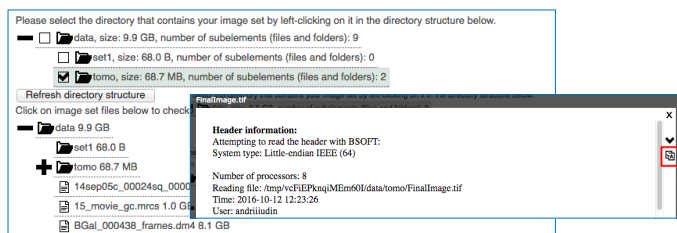
IMMEDIATELY

File upload page

- Three options — Aspera via command line, Aspera via web-client or Globus
- The Aspera web client allows you to upload max ~ 300 files at a time and best for < 400GB datasets
- Data-transfers commonly proceed at 50 - 200 GB per hour so expect TB+ sized datasets to take days in some cases
- Structure your data so that each raw image set is in a separate directory
- Multiple image sets are fine
- Each image set should be consistent (images of the same format/conditions/sizes etc)
- Use of common file formats is encouraged, e.g., MRC
- You can include auxiliary data such as particle picking coordinates
 - Put them in directories but do not mark them as "image sets"
 - Use free text "details" field to describe them for now

Associate image sets with data page

- You need to tell us how your data is organized
- **Checking the uploaded data**
 - System will highlight zero sized files
 - there is a downloadable script for md5 checksum verification
- **Associate datasets with directories**
- **File parameters can be harvested from files**



Invite reviewers

- Give anonymous access to reviewers to browse and download your data prior to submission
- Generate link and pass the link on to the journal
- The link can only be used once

Entry ID:

Invitation link: