EMDB & EMPIAR: Archiving of Electron Microscopy Data

Jack Turner

EMDB

EMBL-EBI

Acknowledgements


Kyle Morris


Sanja Abbott


Neli Fonsecca


Lucas Carrijo de Oliveria


Zhe Wang


Amudha Duraisamy


Osman Salih

EMDB
Electron Microscopy Data Bank

Acknowledgements

WORLDWIDE PDB PROTEIN DATA BANK

EMDB Electron Microscopy Data Bank

RCSB PDB PROTEIN DATA BANK

EMPIAR Electron Microscopy Public Image Archive

PDBe Protein Data Bank in Europe

PDBj Protein Data Bank Japan

PDBc

BMRB Biological Magnetic Resonance Data Bank

Overview of session

- Introduction to Structural archives
- EMDB Website
- Depositing data

Archiving across scales

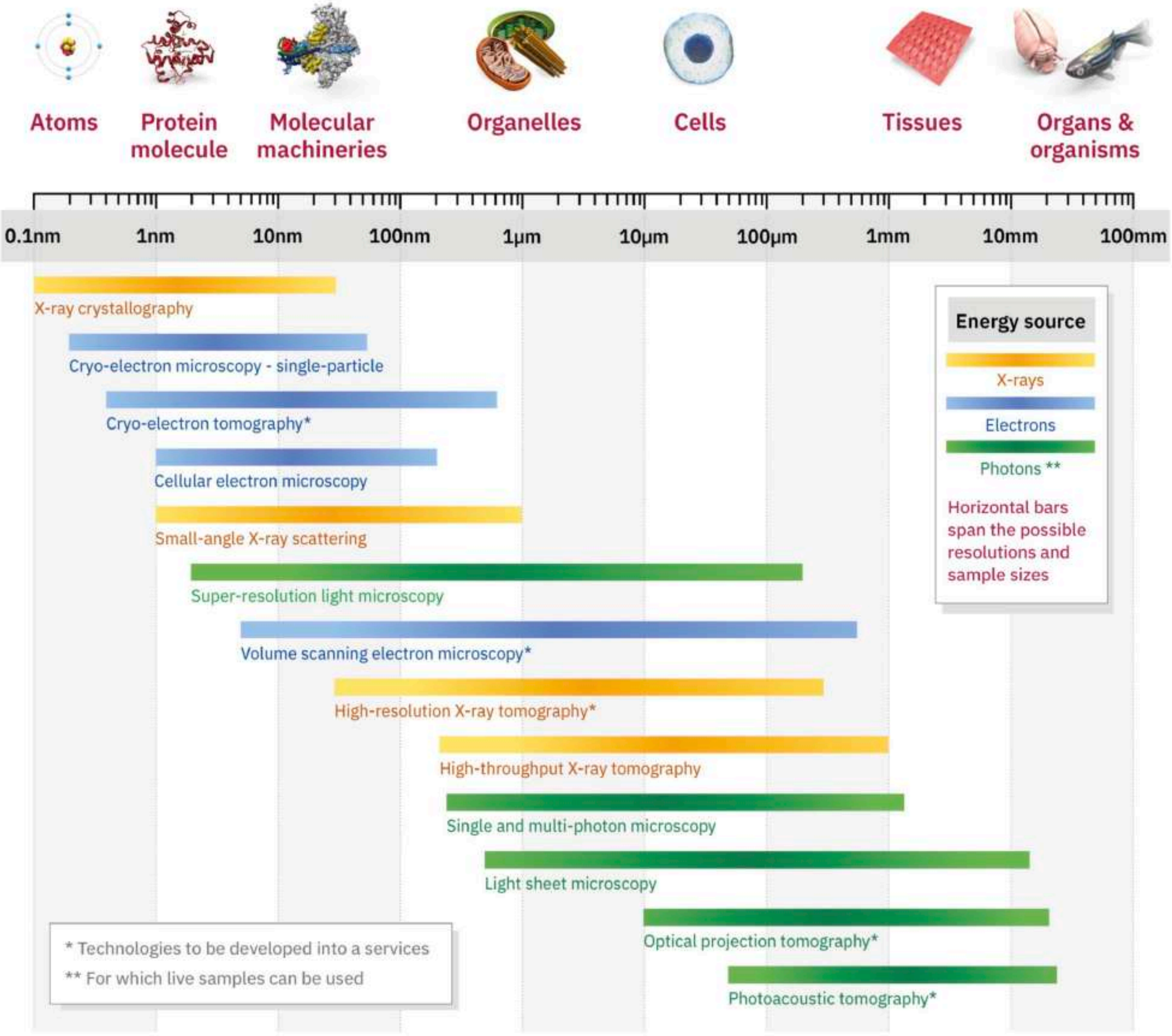


Image credit: EMBL Indicative Scheme pg150 (https://www.embl.org/documents/wp-content/uploads/2022/04/embl-programme-2022-2026.pdf)

Archiving across scales

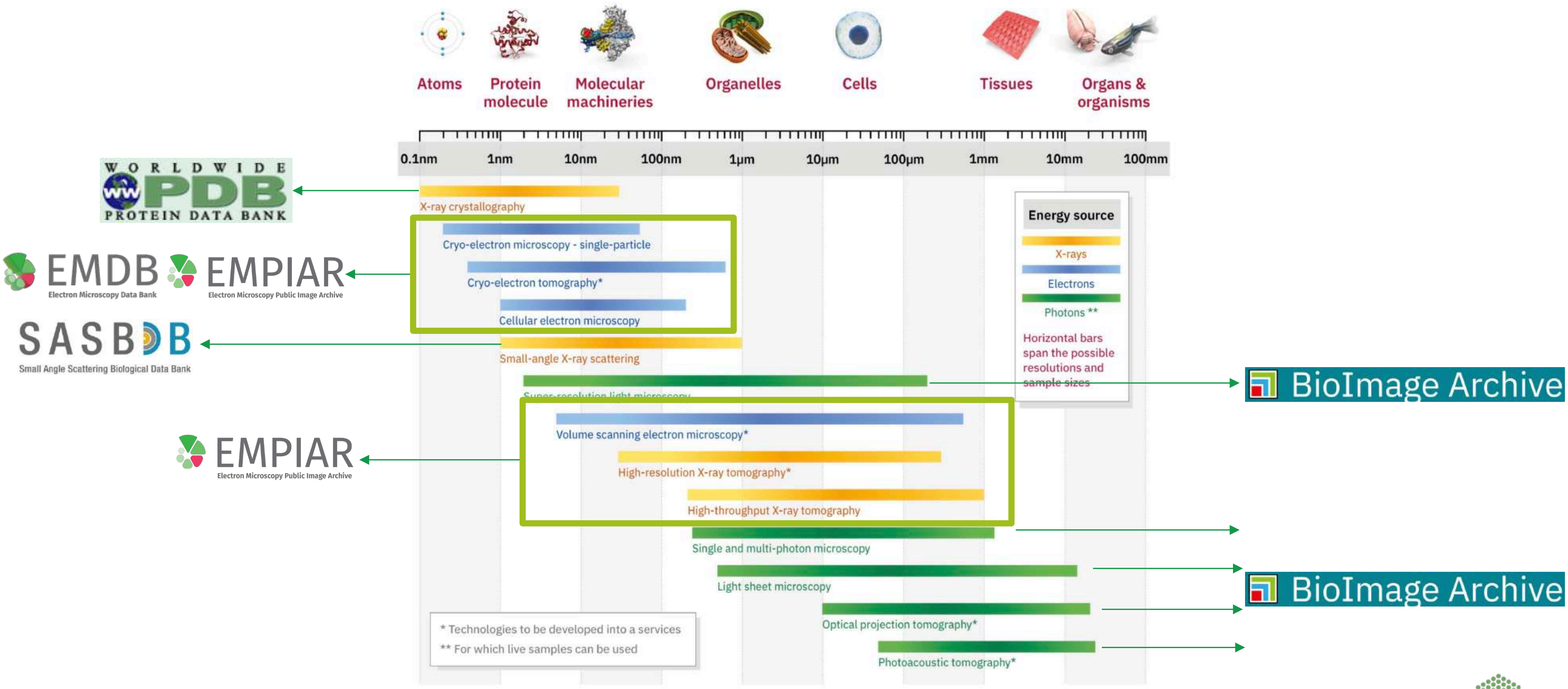
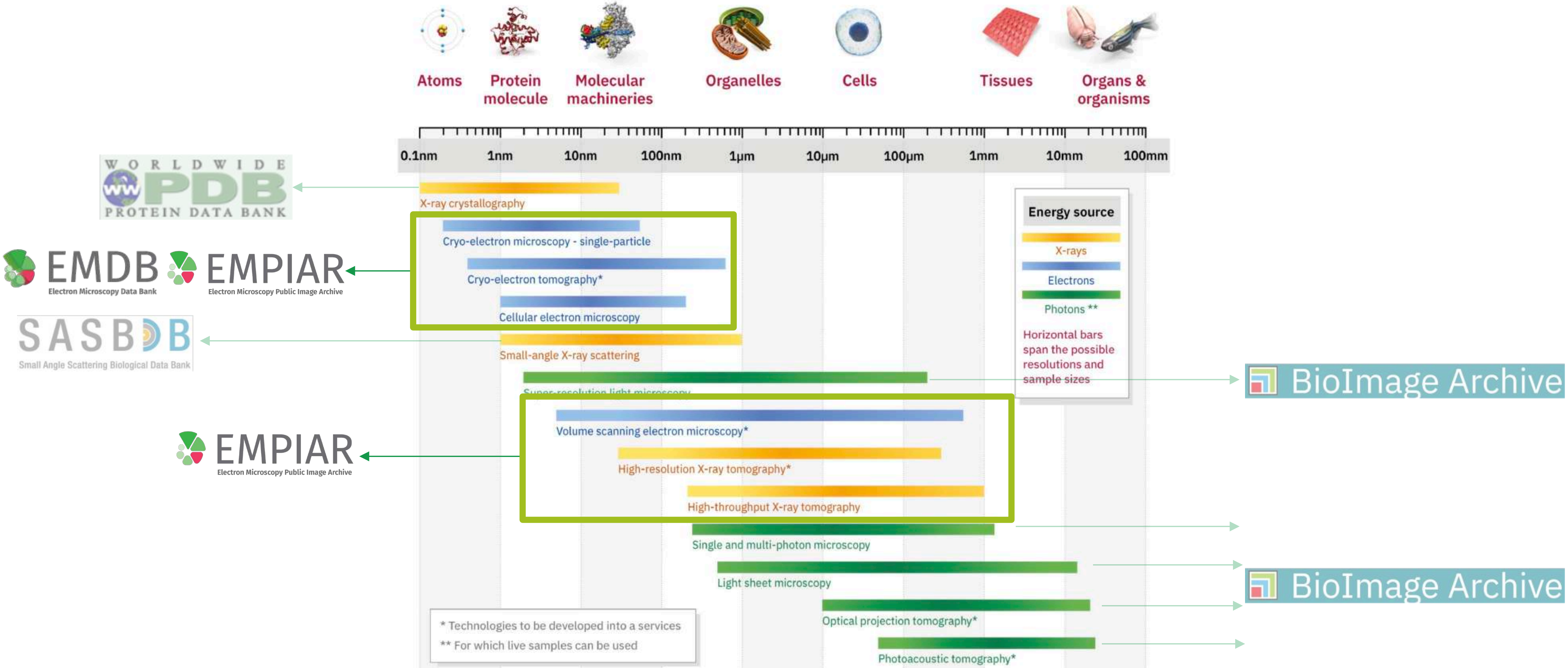


Image credit: EMBL Indicative Scheme pg150 (https://www.embl.org/documents/wp-content/uploads/2022/04/embl-programme-2022-2026.pdf)

EMBL-EBI

Archiving across scales



Atoms | Protein molecule | Molecular machineries | Organelles | Cells | Tissues | Organs & organisms

0.1nm | 1nm | 10nm | 100nm | 1µm | 10µm | 100µm | 1mm | 10mm | 100mm

WORLDWIDE PDB PROTEIN DATA BANK ← X-ray crystallography

EMDB Electron Microscopy Data Bank
EMPIAR Electron Microscopy Public Image Archive ← Cryo-electron microscopy - single-particle, Cryo-electron tomography*, Cellular electron microscopy

SASBDB Small Angle Scattering Biological Data Bank ← Small-angle X-ray scattering

Super-resolution light microscopy → BioImage Archive

EMPIAR Electron Microscopy Public Image Archive ← Volume scanning electron microscopy*, High-resolution X-ray tomography*, High-throughput X-ray tomography

Single and multi-photon microscopy, Light sheet microscopy, Optical projection tomography*, Photoacoustic tomography* → BioImage Archive

Energy source
- X-rays
- Electrons
- Photons **

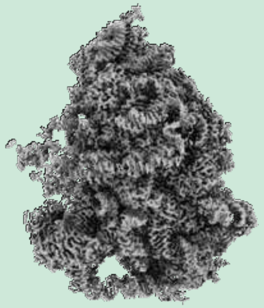
Horizontal bars span the possible resolutions and sample sizes

* Technologies to be developed into a services
** For which live samples can be used

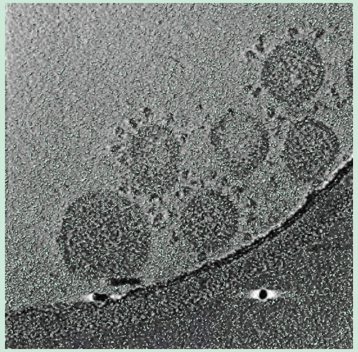
Image credit: EMBL Indicative Scheme pg150 (https://www.embl.org/documents/wp-content/uploads/2022/04/embl-programme-2022-2026.pdf)

The Electron Microscopy Data Bank - EMDB

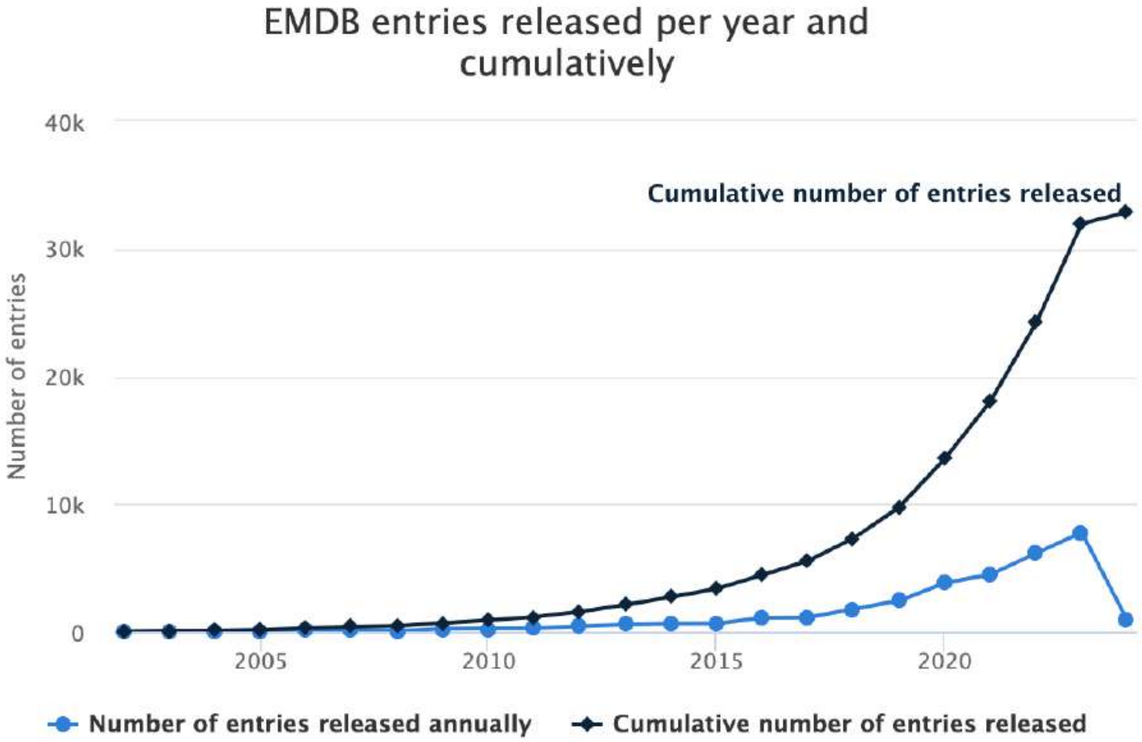
- Established in 2002 at EMBL-EBI
- wwPDB Partner
- Electron microscopy volume maps and Tomograms
- All data is freely and openly available under a CC0 license



EMD-18382



EMD-19668

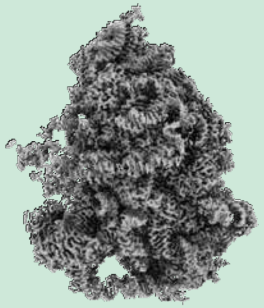


as of 2024-02-08

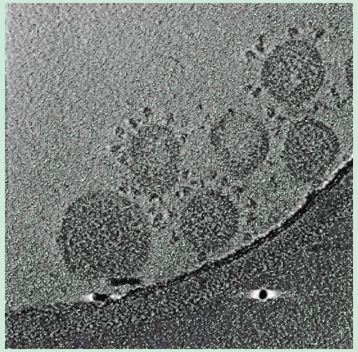
EMDB: emdb-empiar.org
EMPIAR: empiar.org

The Electron Microscopy Data Bank - EMDB

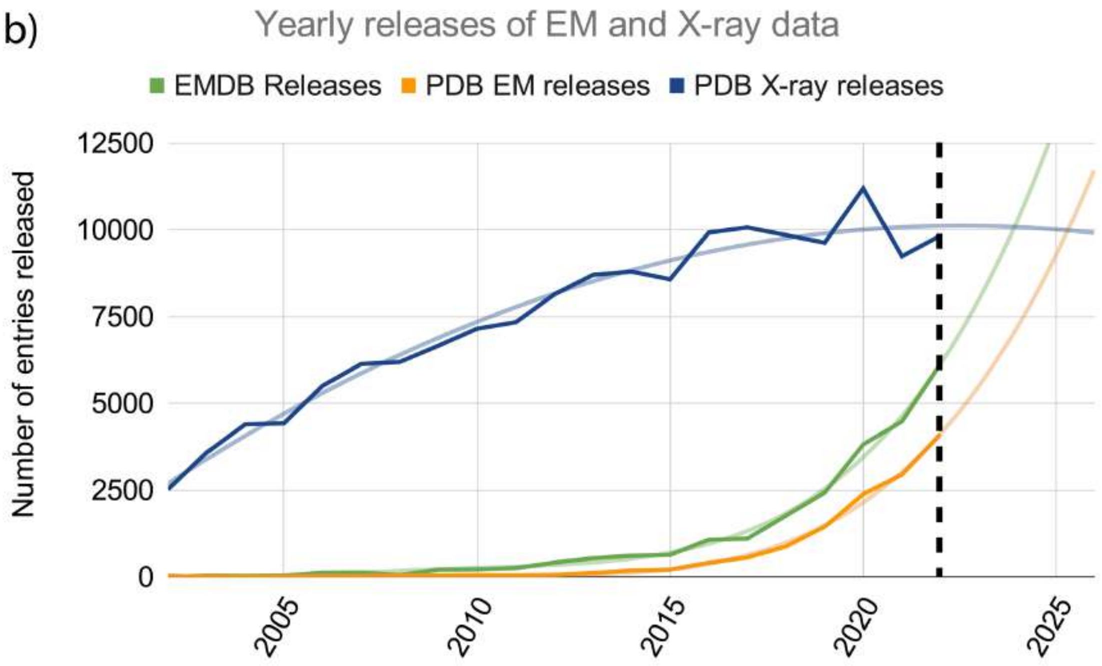
- Established in 2002 at EMBL-EBI
- wwPDB Partner
- Electron microscopy volume maps and Tomograms
- All data is freely and openly available under a CC0 license



EMD-18382



EMD-19668



The wwPDB Consortium , EMDB—the Electron Microscopy Data Bank, *Nucleic Acids Research*, Volume 52, Issue D1, 5 January 2024, Pages D456–D465, https://doi.org/10.1093/nar/gkad1019

EMDB
Electron Microscopy Data Bank

EMPIAR
Electron Microscopy Public Image Archive

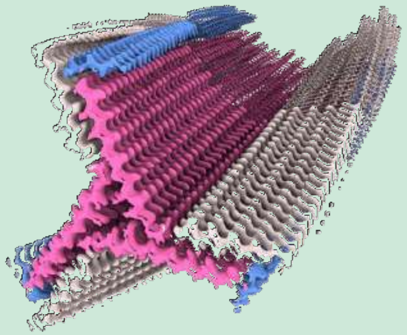
WORLDWIDE PDB
PROTEIN DATA BANK

EMDB: emdb-empiar.org
EMPIAR: empiar.org

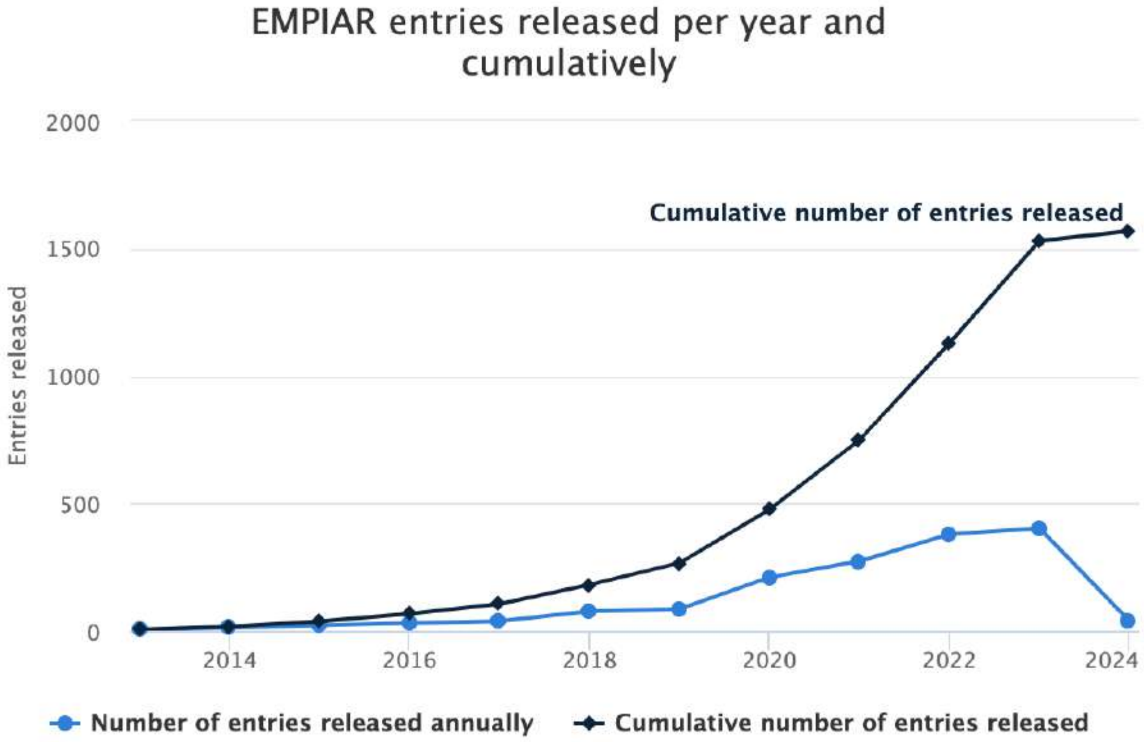
EMBL-EBI

Electron Microscopy Public Image Archive - EMPIAR

- Raw imaging data used to produce maps found in the EMDB
- Volume EM techniques and X-ray tomography and microscopy
- All data is freely and openly available under a CC0 license



EMPIAR-11278/EMD-14167

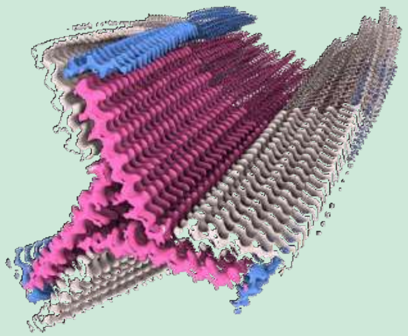


as of 2024-02-08

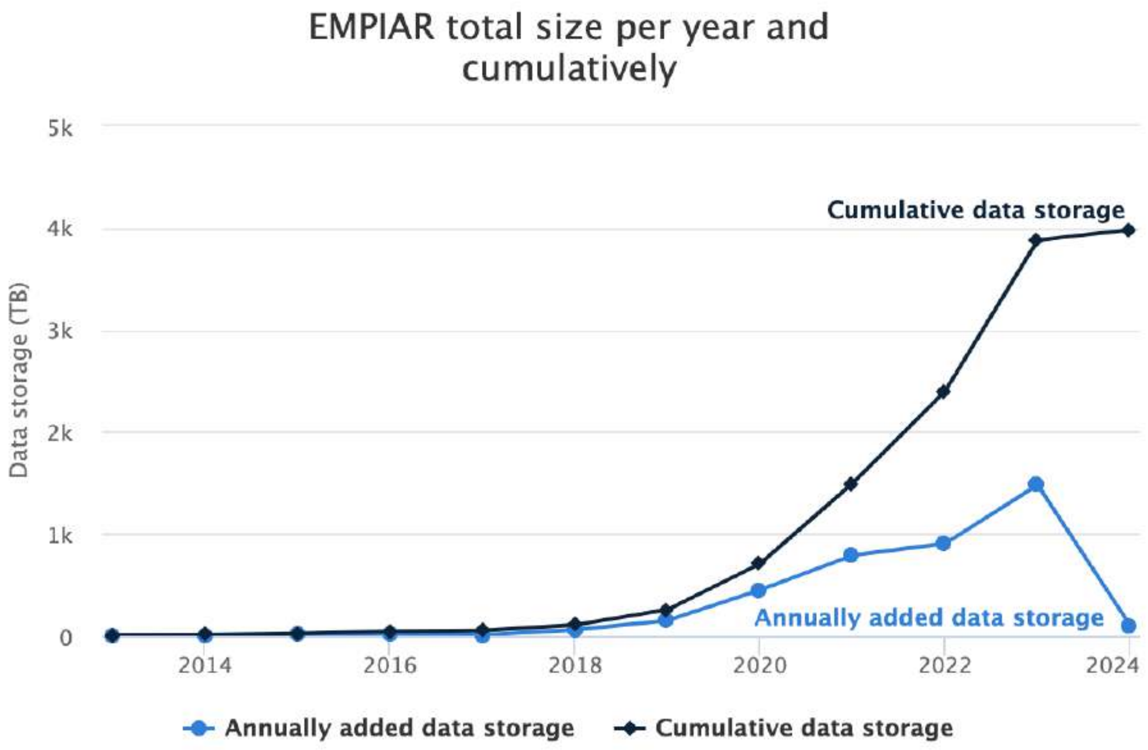
EMDB: emdb-empiar.org
EMPIAR: empiar.org

Electron Microscopy Public Image Archive - EMPIAR

- Raw imaging data used to produce maps found in the EMDB
- Volume EM techniques and X-ray tomography and microscopy
- All data is freely and openly available under a CC0 license



EMPIAR-11278/EMD-14167

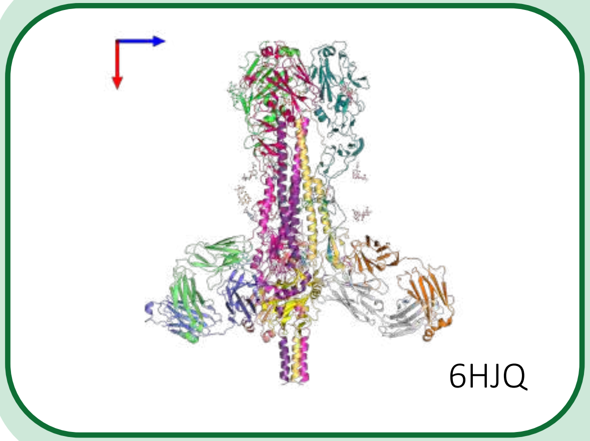


as of 2024-02-08

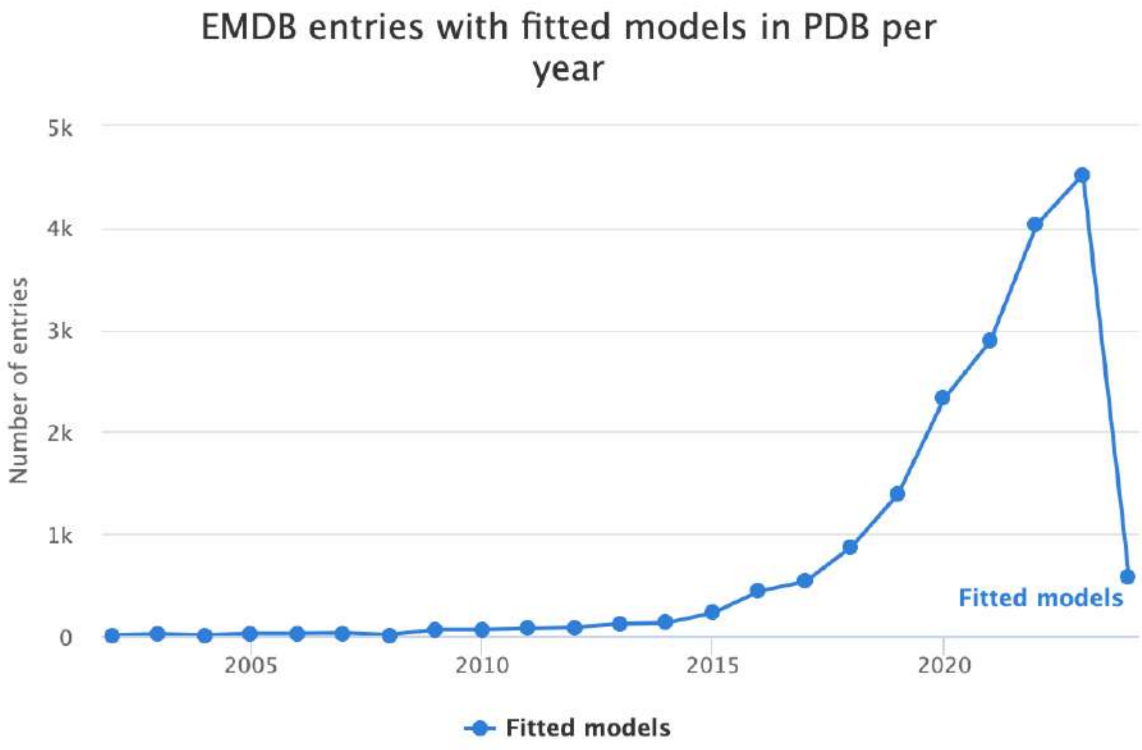
EMDB: emdb-empiar.org
EMPIAR: empiar.org

Protein Data Bank - PDB

- Archiving Coordinate models from EM, X-ray and NMR
- All data is freely and openly available under a CC0 license


6HJQ


7Z4B


EMDB entries with fitted models in PDB per year

as of 2024-02-16

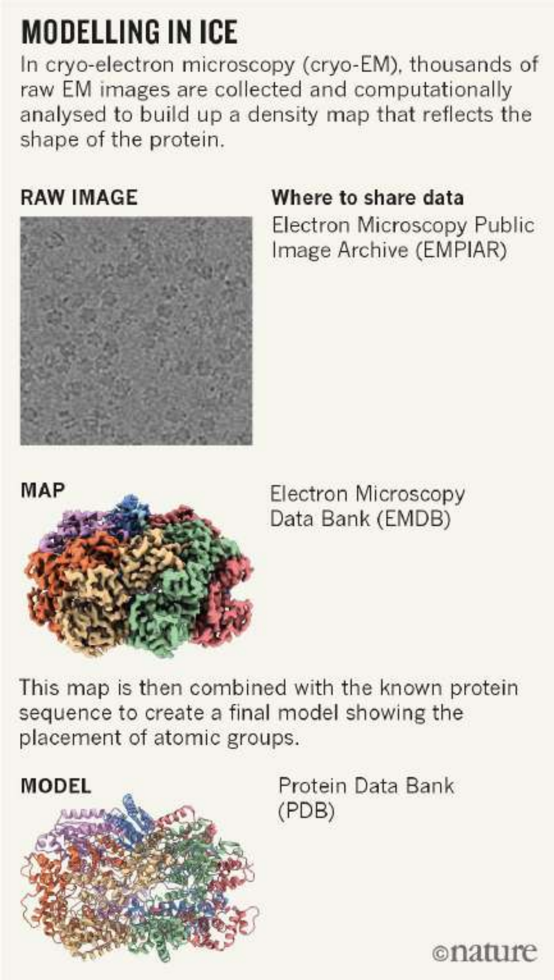
~18,500 entries

~15,000 entries last year

www.ebi.ac.uk/pdbe/ - www.rcsb.org - www.pdbj.org

EMDB, EMPIAR & PDB Summary

- Archive the full electron microscopy (EM) processing workflow
- EMPIAR – Supporting development of software and scientists
- EMDB – Refined EM structures and tomograms
- PDB - Coordinate models from X-ray, EM and NMR



MODELLING IN ICE

In cryo-electron microscopy (cryo-EM), thousands of raw EM images are collected and computationally analysed to build up a density map that reflects the shape of the protein.

RAW IMAGE

Where to share data

Electron Microscopy Public Image Archive (EMPIAR)

MAP

Electron Microscopy Data Bank (EMDB)

This map is then combined with the known protein sequence to create a final model showing the placement of atomic groups.

MODEL

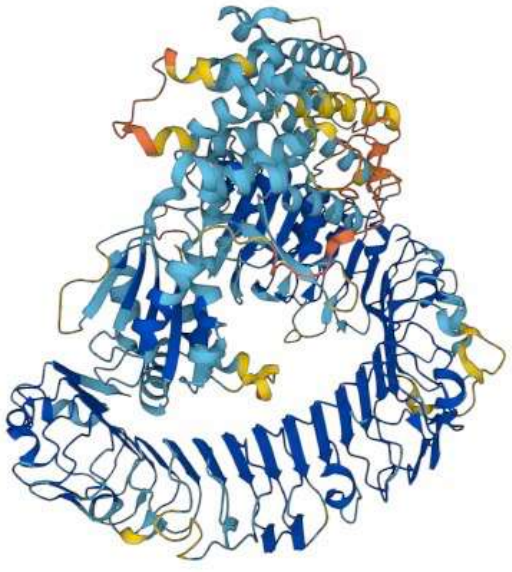
Protein Data Bank (PDB)

©nature

Image credit: *Nature* **561**, 565-567 (2018)

AlphaFold DB

- AlphaFold DB is a collaboration between EBI and Google Deepmind
- AlphaFold DB contains >200 million protein structure predictions
  - PDB contains ~216,000 structures



AF-Q8W3K0-F1

Course

https://www.ebi.ac.uk/training/online/courses/alphafold/

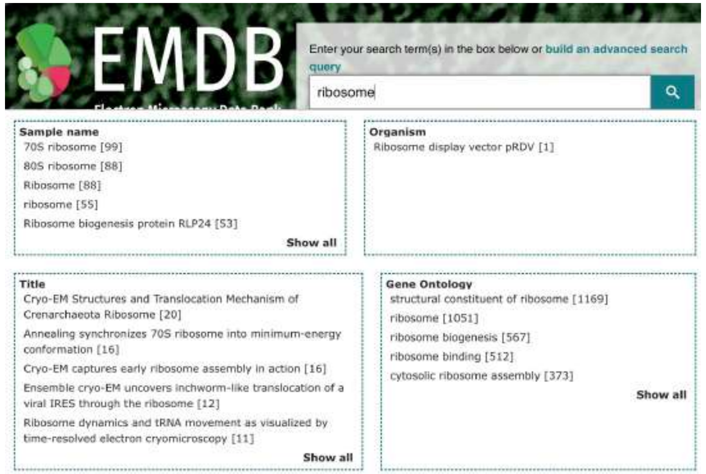
EMBL-EBI

EMDB Website

EMBL-EBI

EMDB Website

EMDB Search



EMDB

Enter your search term(s) in the box below or build an advanced search query

ribosome

Sample name
- 70S ribosome [99]
- 80S ribosome [88]
- Ribosome [88]
- ribosome [55]
- Ribosome biogenesis protein RLP24 [53]

Show all

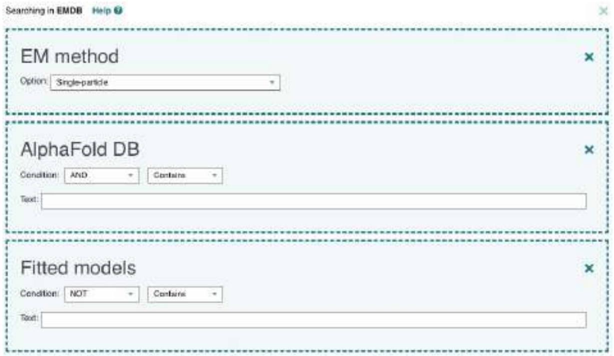
Organism
- Ribosome display vector pRDV [1]

Title
- Cryo-EM Structures and Translocation Mechanism of Crenarchaeota Ribosome [20]
- Annealing synchronizes 70S ribosome into minimum-energy conformation [16]
- Cryo-EM captures early ribosome assembly in action [16]
- Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome [12]
- Ribosome dynamics and tRNA movement as visualized by time-resolved electron cryomicroscopy [11]

Show all

Gene Ontology
- structural constituent of ribosome [1169]
- ribosome [1051]
- ribosome biogenesis [567]
- ribosome binding [512]
- cytosolic ribosome assembly [373]

Show all



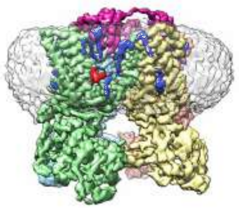
Searching in EMDB Help

EM method
Option: Single-particle

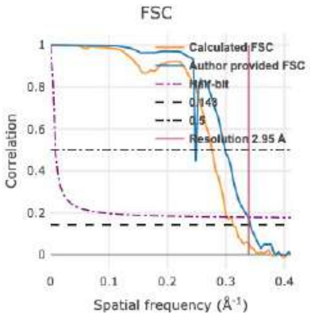
AlphaFold DB
Condition: AND Contains
Text:

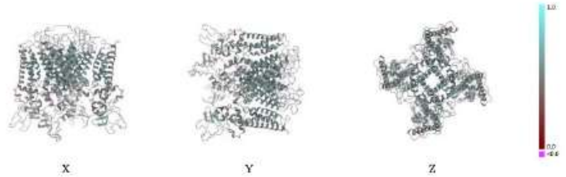
Fitted models
Condition: NOT Contains
Text:

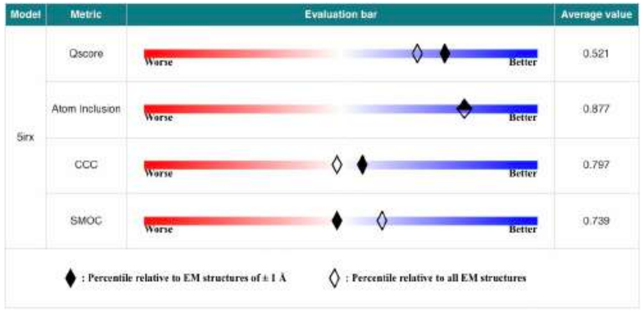
EM Validation



EMD-8117



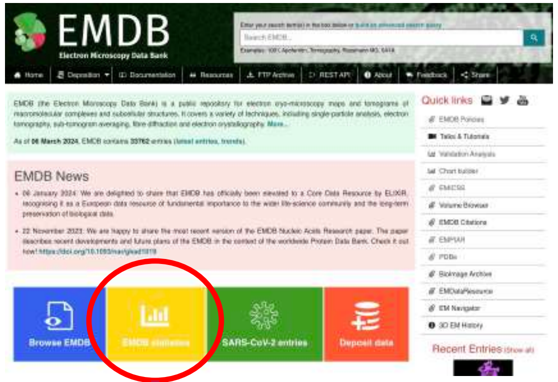


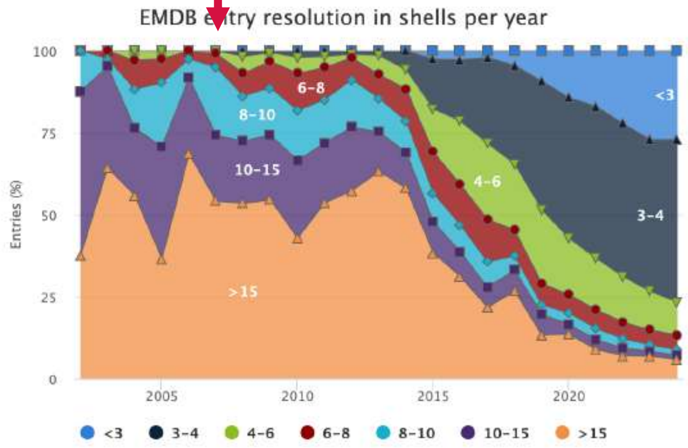


Model	Metric	Evaluation bar	Average value
5irx	Qscore	Worse Better	0.521
	Atom inclusion	Worse Better	0.877
	CCC	Worse Better	0.797
	SMOC	Worse Better	0.739

◆ : Percentile relative to EM structures of ± 1 Å ◇ : Percentile relative to all EM structures

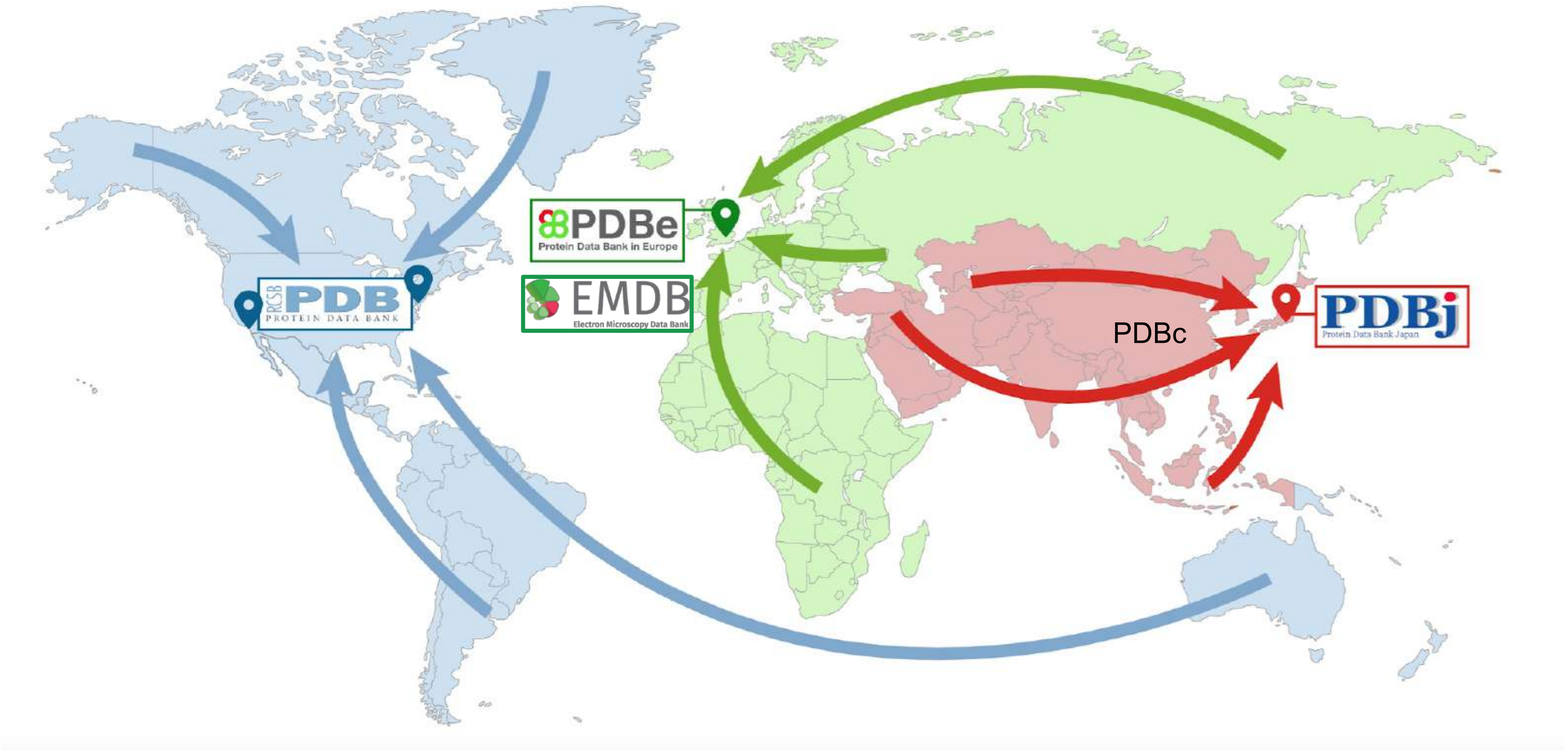
EM Statistics





EMDB/PDB Deposition

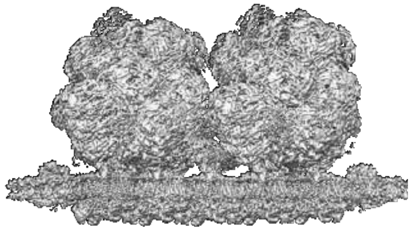
OneDep – A unified deposition system for EM, X-ray and NMR



EMDB - Methodologies

Single Particle

Double-PBS-PSII-PSI-LHCs megacomplex



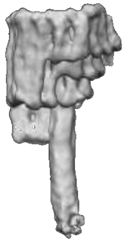
Author List: You X, Zhang X, Cheng J, Xiao YN, Sun S, Sui SF

EMD-33658

[1]

Subtomogram Averaging

DNA Origami Signpost



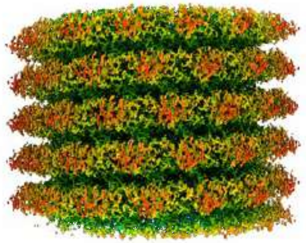
Author List: Silvester E, Vollmer B, Prazak V, Vasishtan D, Machala EA, Whittle C, Black S, Bath J, Turberfield AJ, Gruenewald K, Baker LA

EMD-12188

[2]

Helical Reconstruction

Tobacco Mosaic Virus

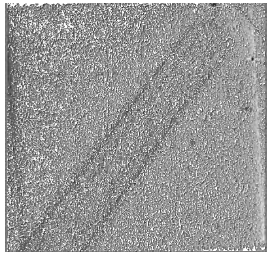


Author List: Bhella D, Love AJ, Streetley J, Taliansky M, McGeachy K, Bukharova T

EMD-16572

Tomography

Syncytial Virus Filamentous Virion



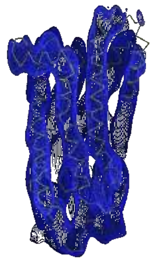
Author List: Conley MJ , Vijayakrishnan S , Bhella D

EMD-13856

[3]

Electron Crystallography

Metarhodopsin I



Author List: Ruprecht JJ, Mielke T, Vogel R, Villa C, Schertler GFX

EMD-1079

[4]

---

[1] You, X., Zhang, X., Cheng, J. et al. In situ structure of the red algal phycobilisome–PSII–PSI–LHC megacomplex. Nature 616, 199–206 (2023). https://doi.org/10.1038/s41586-023-05831-0

[2] Silvester, E., Vollmer, B., Pražák, V., Vasishtan, D., Machala, E. A., Whittle, C., Black, S., Bath, J., Turberfield, A. J., Grünewald, K., & Baker, L. A. (2021). DNA origami signposts for identifying proteins on cell membranes by electron cryotomography. Cell, 184(4), 1110-1121.e16. https://doi.org/https://doi.org/10.1016/j.cell.2021.01.033

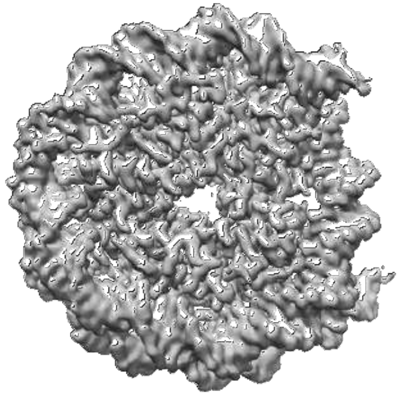
[3] The EMBO Journal (2022) 41: e109728 https://doi.org/10.15252/embj.2021109728

[4] Ruprecht, J.J., Mielke, T., Vogel, R., Villa, C. and Schertler, G.F. (2004), Electron crystallography reveals the structure of metarhodopsin I. The EMBO Journal, 23: 3609-3620. https://doi.org/10.1038/sj.emboj.7600374

EMDB Deposition - Types

Map Only

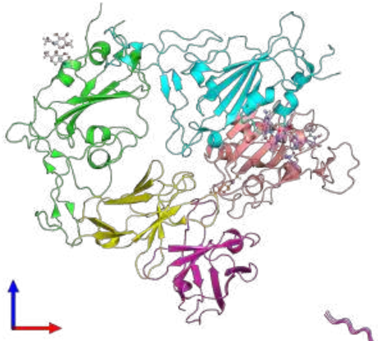
EMD-16202



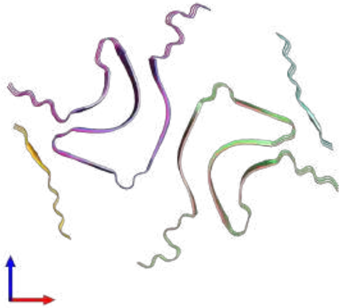
EMD-29854

All Methodologies

Model Only



7X93

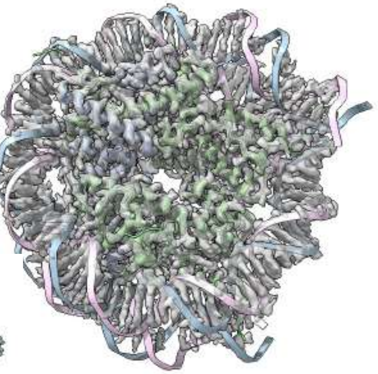


7YK8

All methodologies
except Tomography

Must have associated map!

Map + Model



EMD-29837

EMD-14492

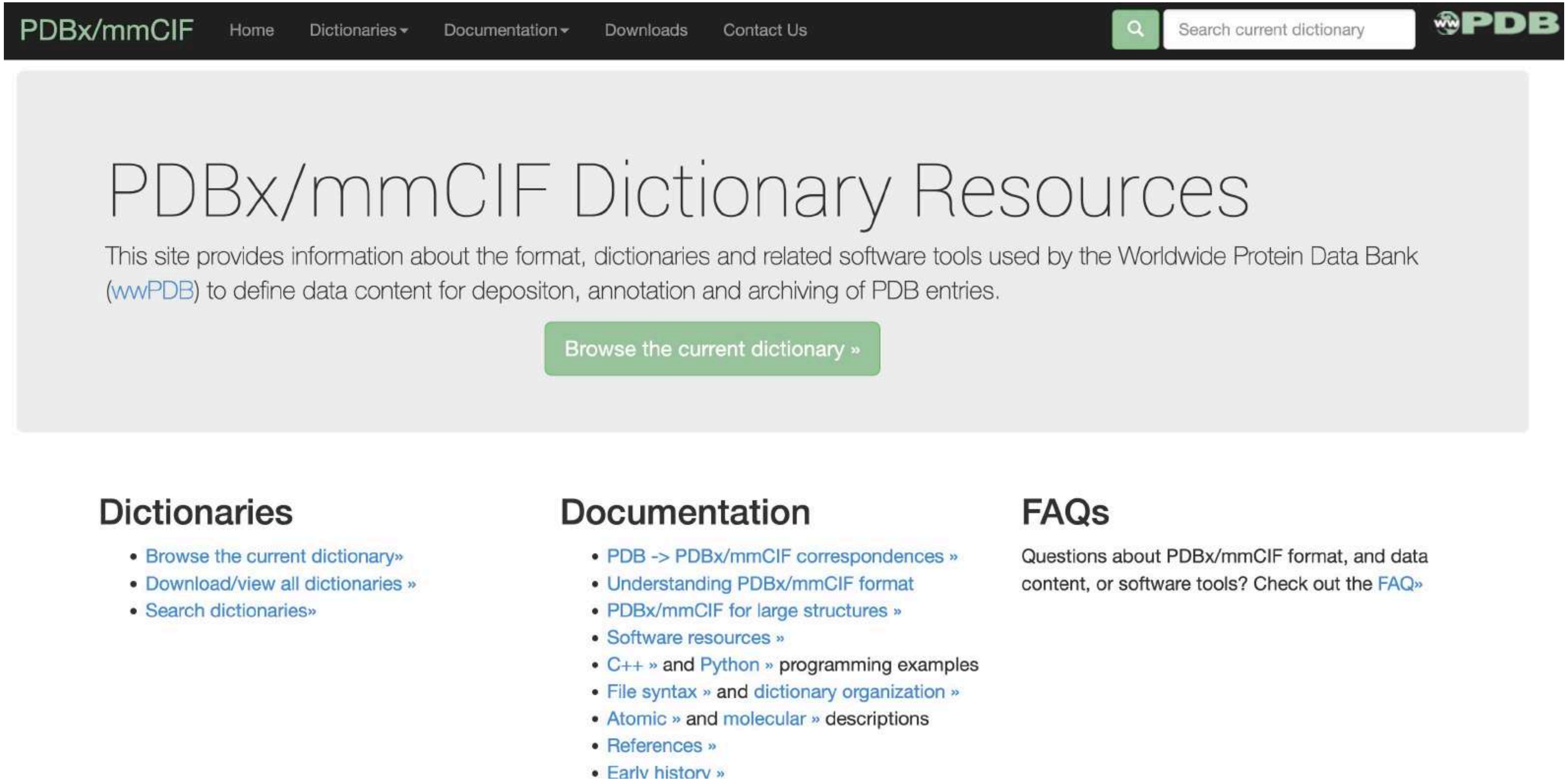
All methodologies
except Tomography

Depositing Data

mmCIF files

- PDB format now considered a legacy format
- mmCIF files are both model and metadata files
- mmCIF files also contain EM metadata and can be used in the absence of a model

mmCIF files



PDBx/mmCIF Dictionary Resources

This site provides information about the format, dictionaries and related software tools used by the Worldwide Protein Data Bank (wwPDB) to define data content for depositon, annotation and archiving of PDB entries.

Browse the current dictionary »

Dictionaries

- Browse the current dictionary»
- Download/view all dictionaries »
- Search dictionaries»

Documentation

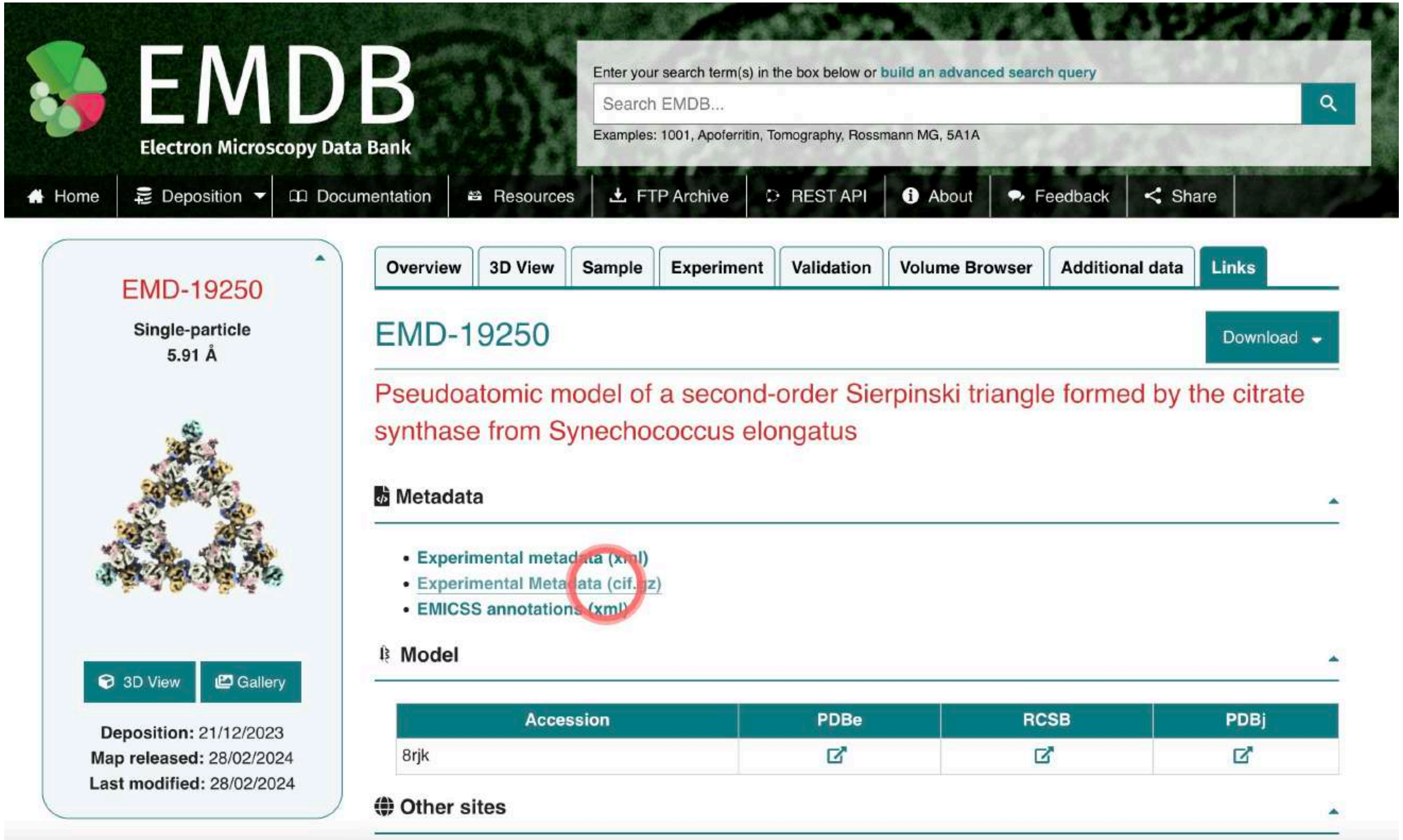
- PDB -> PDBx/mmCIF correspondences »
- Understanding PDBx/mmCIF format
- PDBx/mmCIF for large structures »
- Software resources »
- C++ » and Python » programming examples
- File syntax » and dictionary organization »
- Atomic » and molecular » descriptions
- References »
- Early history »

FAQs

Questions about PDBx/mmCIF format, and data content, or software tools? Check out the FAQ»

https://mmcif.wwpdb.org/

mmCIF files



EMD-19250

Single-particle
5.91 Å

Deposition: 21/12/2023
Map released: 28/02/2024
Last modified: 28/02/2024

EMD-19250

Pseudoatomic model of a second-order Sierpinski triangle formed by the citrate synthase from Synechococcus elongatus

Metadata

- Experimental metadata (xml)
- Experimental Metadata (cif.gz)
- EMICSS annotations (xml)

Model

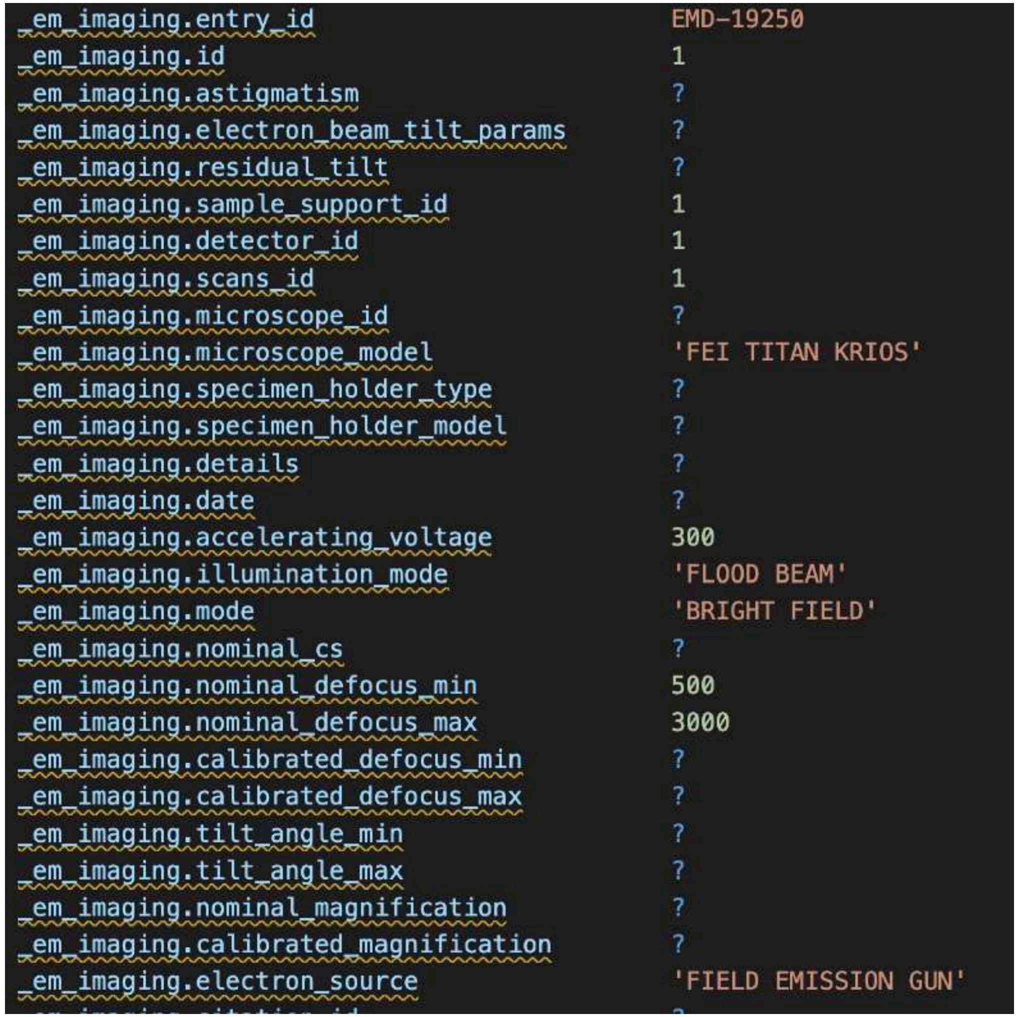
Accession	PDBe	RCSB	PDBj
8rjk	↗	↗	↗

Other sites

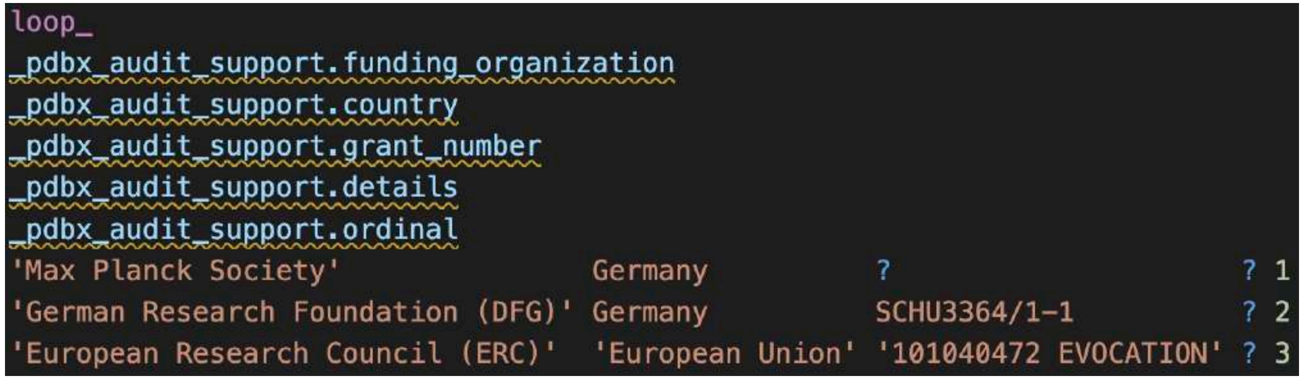
https://mmcif.wwpdb.org/

EMBL-EBI

mmCIF files



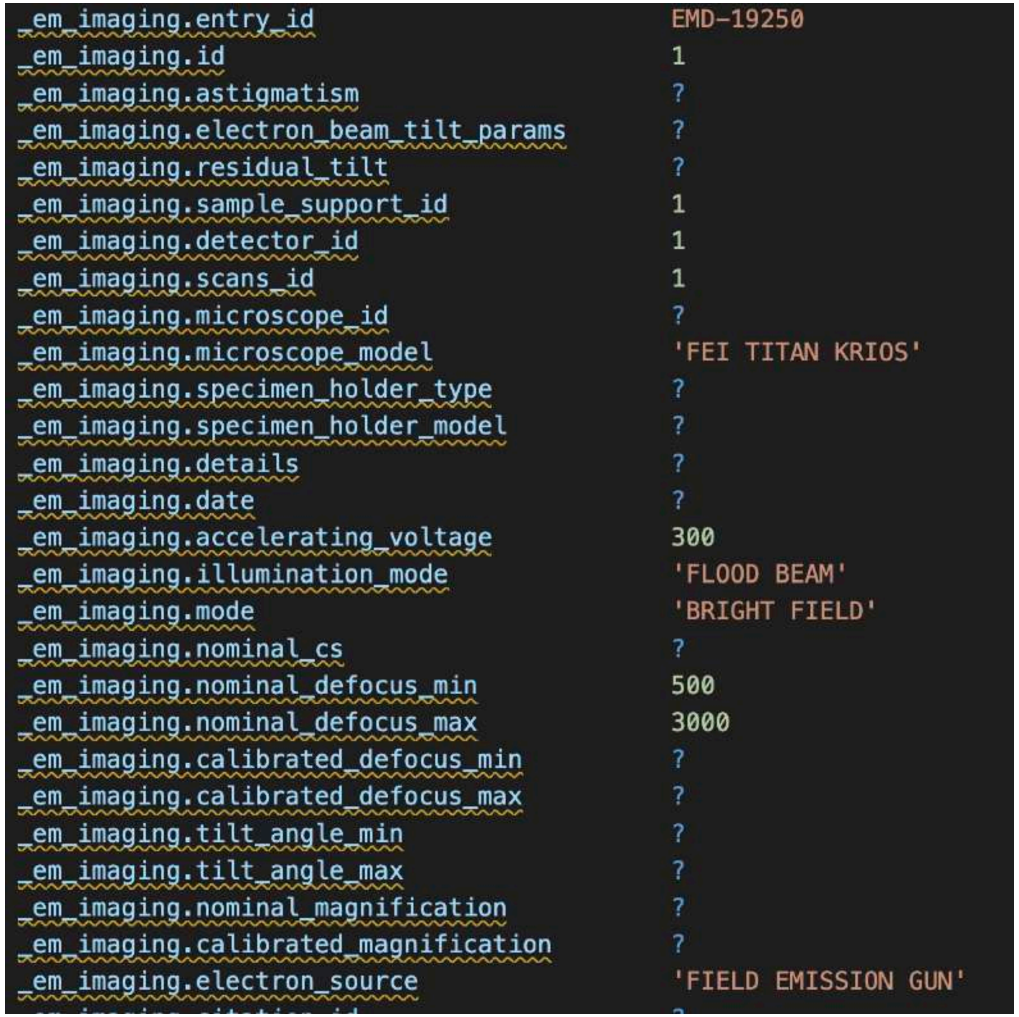
```
_em_imaging.entry_id                        EMD-19250
_em_imaging.id                              1
_em_imaging.astigmatism                     ?
_em_imaging.electron_beam_tilt_params       ?
_em_imaging.residual_tilt                   ?
_em_imaging.sample_support_id               1
_em_imaging.detector_id                     1
_em_imaging.scans_id                        1
_em_imaging.microscope_id                   ?
_em_imaging.microscope_model                'FEI TITAN KRIOS'
_em_imaging.specimen_holder_type            ?
_em_imaging.specimen_holder_model           ?
_em_imaging.details                         ?
_em_imaging.date                            ?
_em_imaging.accelerating_voltage            300
_em_imaging.illumination_mode               'FLOOD BEAM'
_em_imaging.mode                            'BRIGHT FIELD'
_em_imaging.nominal_cs                      ?
_em_imaging.nominal_defocus_min             500
_em_imaging.nominal_defocus_max             3000
_em_imaging.calibrated_defocus_min          ?
_em_imaging.calibrated_defocus_max          ?
_em_imaging.tilt_angle_min                  ?
_em_imaging.tilt_angle_max                  ?
_em_imaging.nominal_magnification           ?
_em_imaging.calibrated_magnification        ?
_em_imaging.electron_source                 'FIELD EMISSION GUN'
```



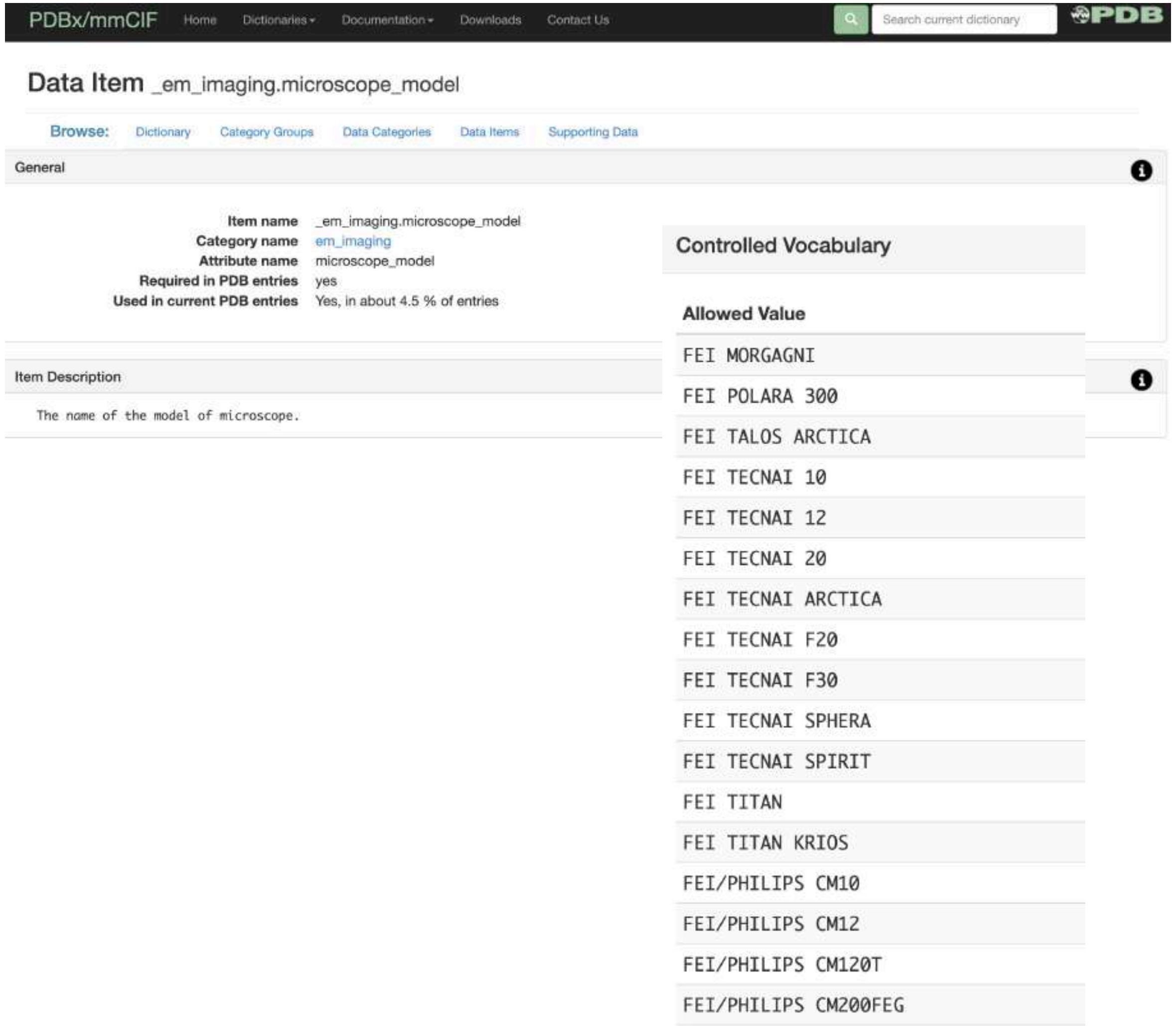
```
loop_
_pdbx_audit_support.funding_organization
_pdbx_audit_support.country
_pdbx_audit_support.grant_number
_pdbx_audit_support.details
_pdbx_audit_support.ordinal
'Max Planck Society'                Germany          ?                     ? 1
'German Research Foundation (DFG)'  Germany          SCHU3364/1-1          ? 2
'European Research Council (ERC)'   'European Union' '101040472 EVOCATION' ? 3
```

https://mmcif.wwpdb.org/

mmCIF files



```
_em_imaging.entry_id                            EMD-19250
_em_imaging.id                                  1
_em_imaging.astigmatism                         ?
_em_imaging.electron_beam_tilt_params           ?
_em_imaging.residual_tilt                       ?
_em_imaging.sample_support_id                   1
_em_imaging.detector_id                         1
_em_imaging.scans_id                            1
_em_imaging.microscope_id                       ?
_em_imaging.microscope_model                    'FEI TITAN KRIOS'
_em_imaging.specimen_holder_type                ?
_em_imaging.specimen_holder_model               ?
_em_imaging.details                             ?
_em_imaging.date                                ?
_em_imaging.accelerating_voltage                300
_em_imaging.illumination_mode                   'FLOOD BEAM'
_em_imaging.mode                                'BRIGHT FIELD'
_em_imaging.nominal_cs                          ?
_em_imaging.nominal_defocus_min                 500
_em_imaging.nominal_defocus_max                 3000
_em_imaging.calibrated_defocus_min              ?
_em_imaging.calibrated_defocus_max              ?
_em_imaging.tilt_angle_min                      ?
_em_imaging.tilt_angle_max                      ?
_em_imaging.nominal_magnification               ?
_em_imaging.calibrated_magnification            ?
_em_imaging.electron_source                     'FIELD EMISSION GUN'
```



Data Item _em_imaging.microscope_model

General

Item name	_em_imaging.microscope_model
Category name	em_imaging
Attribute name	microscope_model
Required in PDB entries	yes
Used in current PDB entries	Yes, in about 4.5 % of entries

Item Description

The name of the model of microscope.

Controlled Vocabulary

Allowed Value
FEI MORGAGNI
FEI POLARA 300
FEI TALOS ARCTICA
FEI TECNAI 10
FEI TECNAI 12
FEI TECNAI 20
FEI TECNAI ARCTICA
FEI TECNAI F20
FEI TECNAI F30
FEI TECNAI SPHERA
FEI TECNAI SPIRIT
FEI TITAN
FEI TITAN KRIOS
FEI/PHILIPS CM10
FEI/PHILIPS CM12
FEI/PHILIPS CM120T
FEI/PHILIPS CM200FEG

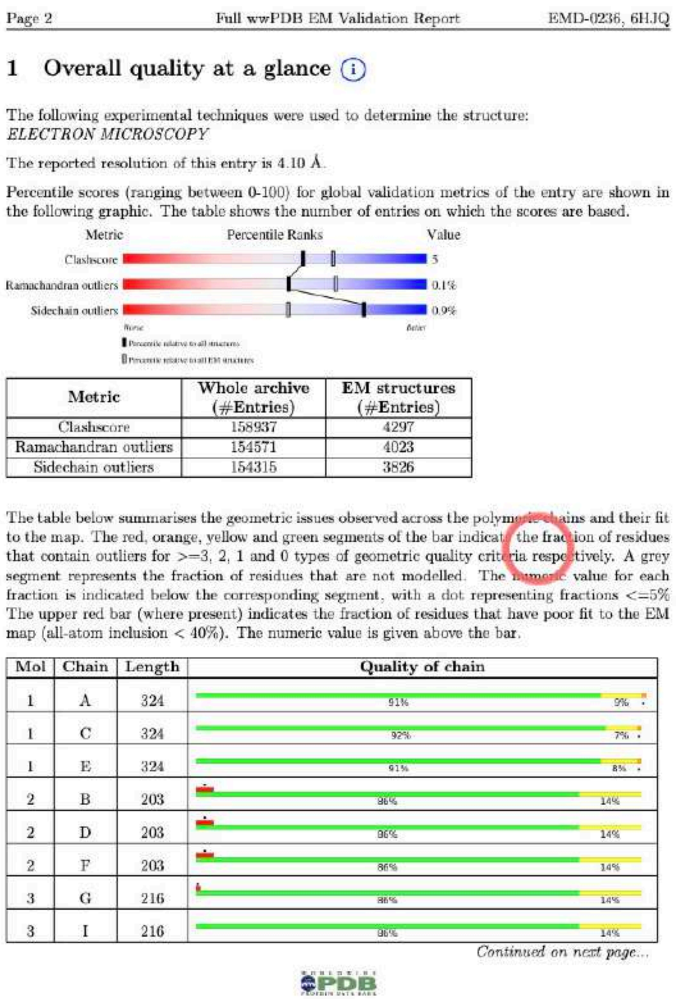
https://mmcif.wwpdb.org/

Entry annotation

- Transmod
- Ligmod
- Seqmod
- Annmod
- Reviewmod

- -> output files and a wwPDB validation report

wwPDB validation report



1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 4.10 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

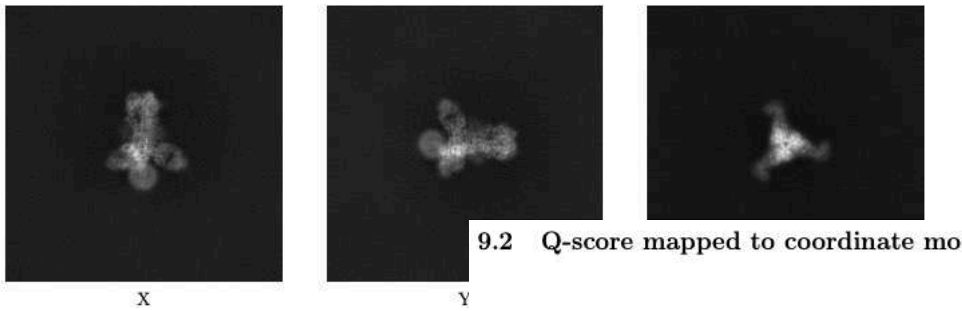
The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	324	91% 9% •
1	C	324	92% 7% •
1	E	324	91% 8% •
2	B	203	• 86% 14%
2	D	203	• 86% 14%
2	F	203	• 86% 14%
3	G	216	86% 14%
3	I	216	86% 14%

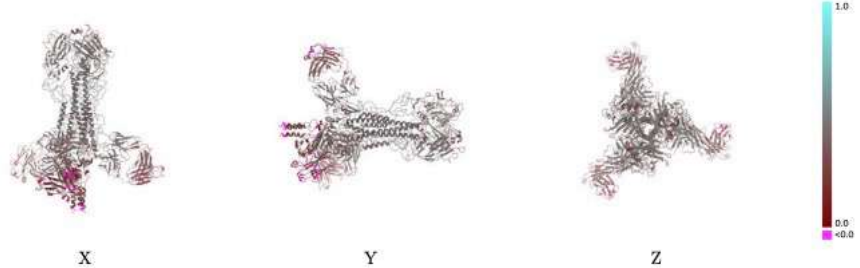
Continued on next page...

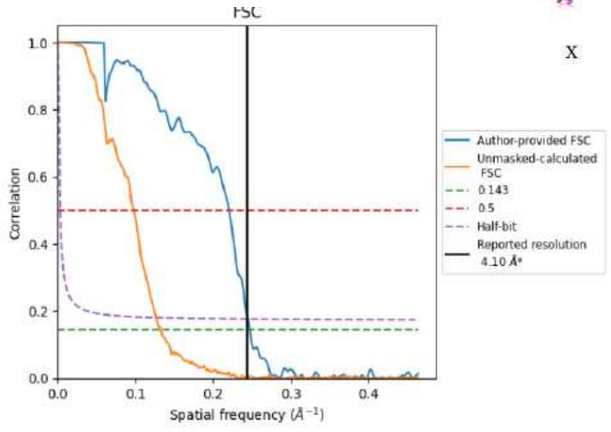
6.1 Orthogonal projections ⓘ

6.1.1 Primary map



9.2 Q-score mapped to coordinate model ⓘ





*Reported resolution corresponds to spatial frequency of 0.244 Å^{-1}

EMD-0236

EMPIAR deposition

EMBL-EBI

EMPIAR Deposition Step-By-Step

1. Register as a new user on EMPIAR’s deposition system
2. Sign-in & Create a New Deposition
3. Complete the Deposition Overview Page
4. Granting rights or transferring ownership of an entry
5. Main EMPIAR Data Upload page
    a. Globus
    b. Aspera Command Line
    c. Aspera Web Interface
6. Complete the Associate image sets with the data page & Submit

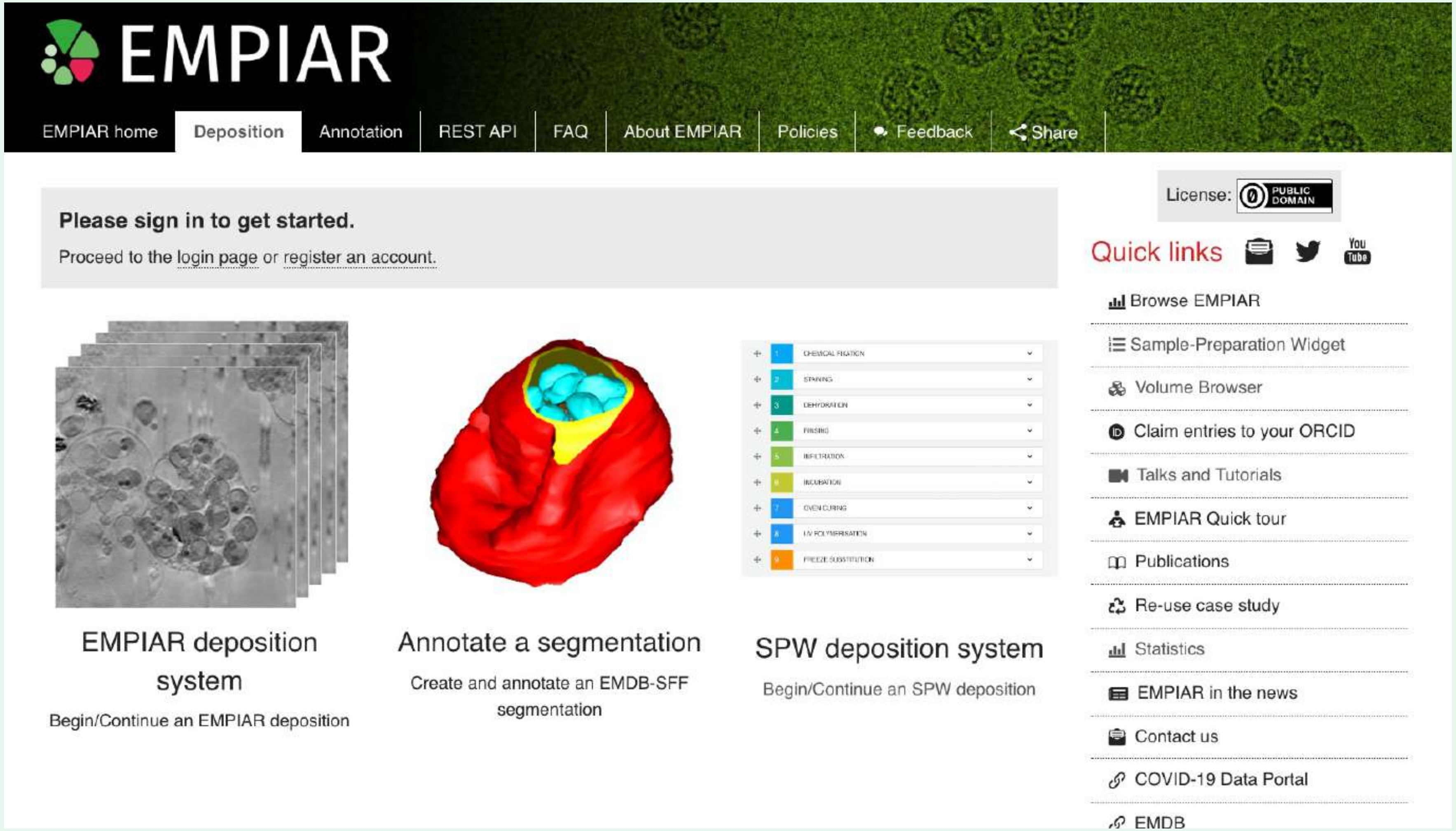
EMPIAR Deposition Policies

- EMPIAR is relatively unstructured compared to EMDB
- Polices pages should be visited to understand accepted file types
- Supported Methods (more info on the policies pages):

i. EMDB - raw image data relating to structures deposited to the EMDB
ii. SBF-SEM - image data collected using serial block-face scanning electron microscopy (like the Gatan 3View system)
iii. SXT - image data collected using soft x-ray tomography
iv. FIB SEM - image data collected using focused ion beam scanning electron microscopy
v. IHM - integrative hybrid modelling data
vi. CLEM - correlative light-electron microscopy
vii. CLXM - correlative light X-ray microscopy
viii. MicroED - microcrystal electron diffraction
ix. ATUM-SEM - Automated Tape-collecting Ultramicrotome Scanning Electron Microscopy
x. Hard X-ray/X-ray microCT - Hard X-ray/X-ray micro-computed tomography
xi. ssET - serial section electron tomography

https://www.ebi.ac.uk/empiar/policies/

EMPIAR Deposition Setup



Please sign in to get started.

Proceed to the login page or register an account.

EMPIAR deposition system

Begin/Continue an EMPIAR deposition

Annotate a segmentation

Create and annotate an EMDB-SFF segmentation

SPW deposition system

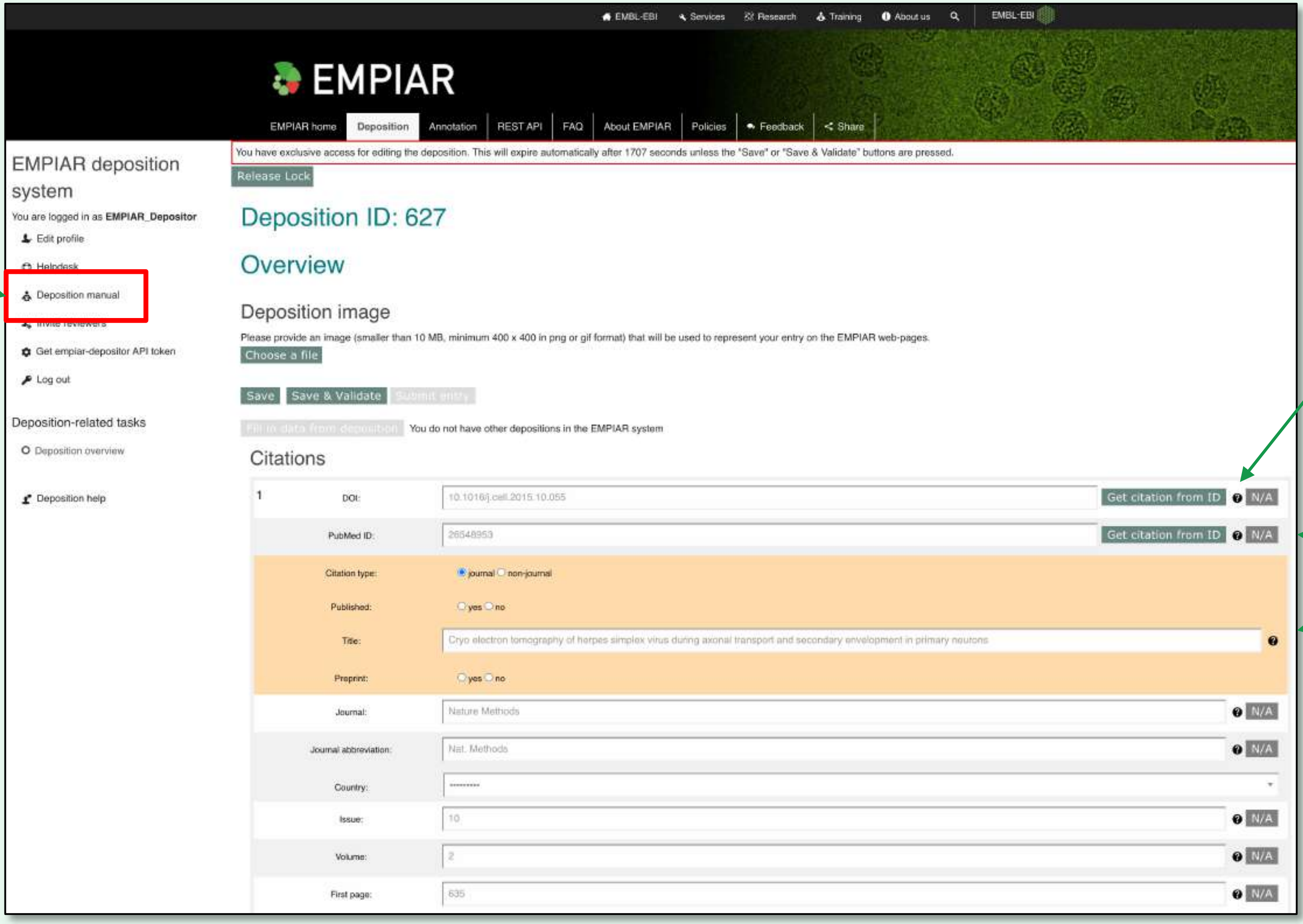
Begin/Continue an SPW deposition

https://www.ebi.ac.uk/empiar/deposition/choose-action/

EMBL-EBI

EMPIAR Deposition Overview Page

Deposition manual available if you get stuck



You have exclusive access for editing the deposition. This will expire automatically after 1707 seconds unless the "Save" or "Save & Validate" buttons are pressed.

EMPIAR deposition system

You are logged in as **EMPIAR_Depositor**

Deposition ID: 627

Overview

Deposition image

Please provide an image (smaller than 10 MB, minimum 400 x 400 in png or gif format) that will be used to represent your entry on the EMPIAR web-pages.

You do not have other depositions in the EMPIAR system

Citations

1

DOI:	10.1016/j.cell.2015.10.055
PubMed ID:	26548953
Citation type:	● journal ○ non-journal
Published:	○ yes ○ no
Title:	Cryo electron tomography of herpes simplex virus during axonal transport and secondary envelopment in primary neurons
Preprint:	○ yes ○ no
Journal:	Nature Methods
Journal abbreviation:	Nat. Methods
Country:	---------
Issue:	10
Volume:	2
First page:	635

Citation can be pulled from DOI or PubMed ID

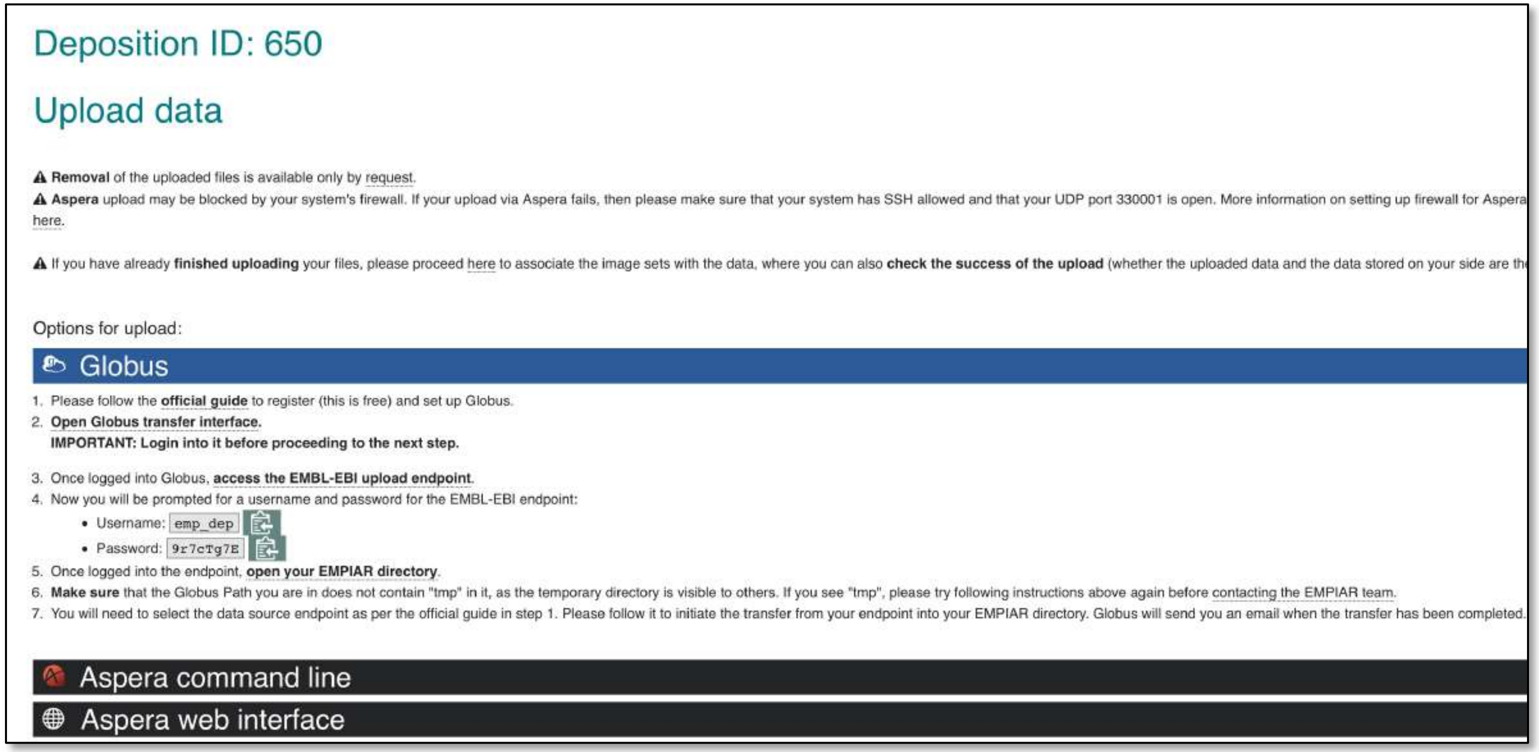
Grey = Optional
Orange = Mandatory
(N/A is clickable for optional fields)

https://www.ebi.ac.uk/empiar/deposition/

EMBL-EBI

EMPIAR Deposition - File Upload

- Aspera
- Globus
- FTP (Not recommended)



Deposition ID: 650

Upload data

⚠ **Removal** of the uploaded files is available only by request.
⚠ **Aspera** upload may be blocked by your system's firewall. If your upload via Aspera fails, then please make sure that your system has SSH allowed and that your UDP port 330001 is open. More information on setting up firewall for Aspera here.

⚠ If you have already **finished uploading** your files, please proceed here to associate the image sets with the data, where you can also **check the success of the upload** (whether the uploaded data and the data stored on your side are th

Options for upload:

Globus

1. Please follow the **official guide** to register (this is free) and set up Globus.
2. **Open Globus transfer interface.**
   IMPORTANT: Login into it before proceeding to the next step.
3. Once logged into Globus, **access the EMBL-EBI upload endpoint**.
4. Now you will be prompted for a username and password for the EMBL-EBI endpoint:
   - Username: `emp_dep`
   - Password: `9r7cTg7E`
5. Once logged into the endpoint, **open your EMPIAR directory**.
6. **Make sure** that the Globus Path you are in does not contain "tmp" in it, as the temporary directory is visible to others. If you see "tmp", please try following instructions above again before contacting the EMPIAR team.
7. You will need to select the data source endpoint as per the official guide in step 1. Please follow it to initiate the transfer from your endpoint into your EMPIAR directory. Globus will send you an email when the transfer has been completed.

Aspera command line

Aspera web interface

EMPIAR Deposition - Overview

1. Gather Information
   - EMPIAR accepts various types of data, including multi-frame micrographs, frame-averaged micrographs, particle stacks and tilt series, as well as auxiliary files (*e.g.* particle selection coordinates)
     - Make a note of dataset details (*e.g.* # of images (or tilt-series), raw or processed, single or multi-frame, image format and dimensions, pixel spacing and type)
2. Organise your Data!
   - Each data-type must be located in a separate directory (*e.g.* one for micrographs and one for particle stacks) and please sub-divide directories possessing more than 10,000 files
3. Setup Data Transfer Programs
   - Aspera (http://asperasoft.com) and Globus (https://www.globus.org) efficiently and robustly transfer large data volumes (empiar.org/faq)
4. Deposit your Data and Metadata (empiar.org/deposition)

Deposition Summary

- EMDB and PDB deposition occurs through OneDep system
- Please use **Login with ORCID!**
- **mmCIF file are also metadata files**
- EMPIAR deposition occurs through a separate system

Questions and links

EMDB Website

https://www.ebi.ac.uk/emdb/

Alphafold Course

https://www.ebi.ac.uk/training/online/courses/alphafold/

OneDep Validation Server

https://validate-rcsb-2.wwpdb.org/

EMDB API

Google Colab notebook