Archiving Electon Microscopy Data – EMDB and EMPIAR

Jack Turner

EMDB

EMBL-EBI

Outline of session

Introduction to EMDB and EMPIAR

Websites and Validation

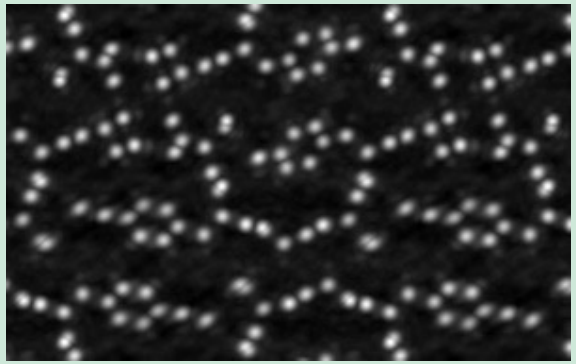
Deposition

EMBL-EBI

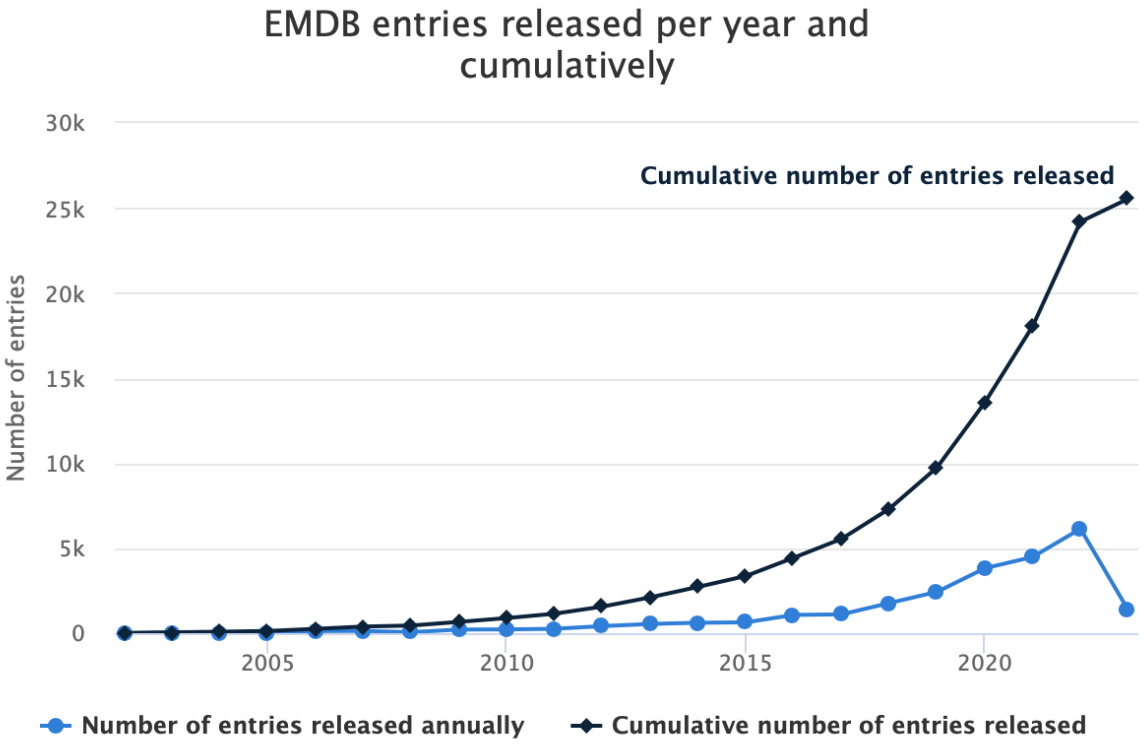
The Electron Microscopy Data Bank - EMDB

- 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-15182



EMD-0698

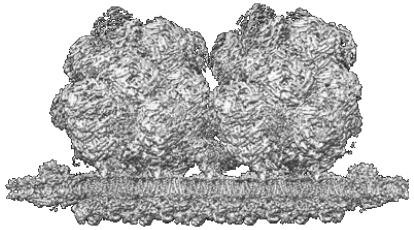


EMDB: emdb-empiar.org
EMPIAR: empiar.org

EMDB - Methodologies

Single Particle

Double-PBS-PSII-PSI-LHCs megacomplex

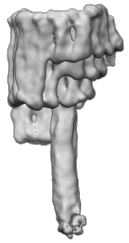


Author List: You X, Zhang X, Cheng J, Xiao YN, Sun S, Sui SF

EMD-33658

Subtomogram Averaging

DNA Origami Signpost

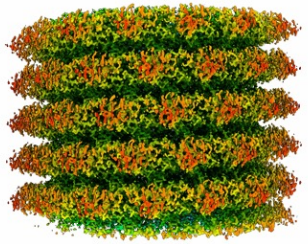


Author List: Silvester E, Vollmer B, Prazak V, Vasishtan D, Machala EA, Whittle C, Black S, Bath J, Turberfield AJ, Gruenewald K, Baker LA

EMD-12188 [1]

Helical Reconstruction

Tobacco Mosaic Virus

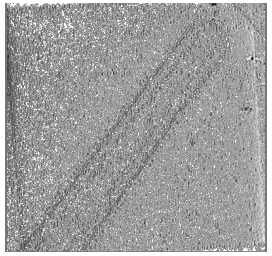


Author List: Bhella D, Love AJ, Streetley J, Taliansky M, McGeachy K, Bukharova T

EMD-16572

Tomography

Syncytial Virus Filamentous Virion



Author List: Conley MJ , Vijayakrishnan S , Bhella D

EMD-13856 [2]

Electron Crystallography

Metarhodopsin I

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

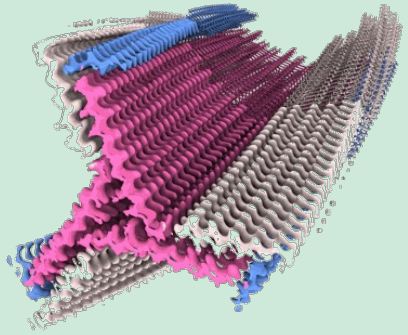
EMD-1079 [3]

[1] 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
[2] The EMBO Journal (2022) 41: e109728 https://doi.org/10.15252/embj.2021109728
[3] 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

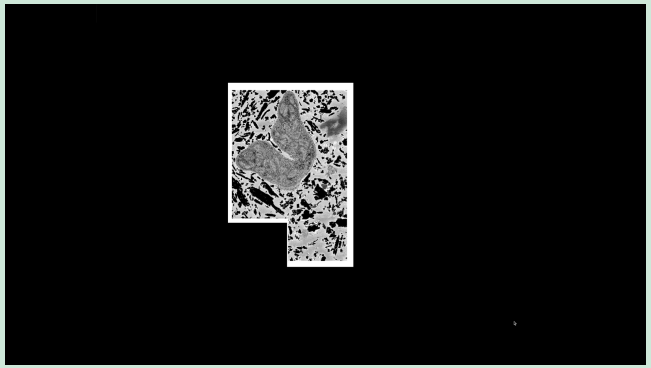
EMBL-EBI

Electron Microscopy Public Image Archive - EMPIAR

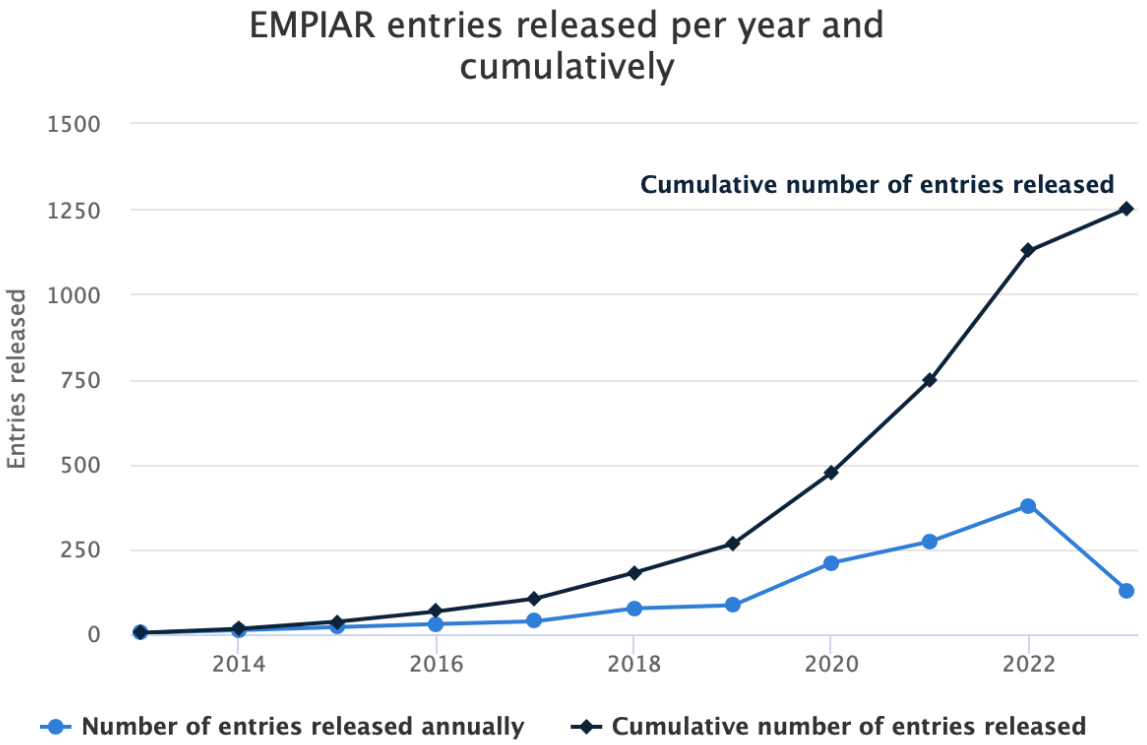
- Raw imaging data used to produce maps found in the EMDB
- Volume EM techniques and X-ray tomography and microscopy
- All data is freely and openly available under a CC0 license



EMPIAR-11278/EMD-14167



EMPIAR-10310



EMDB Electron Microscopy Data Bank

EMPIAR Electron Microscopy Public Image Archive

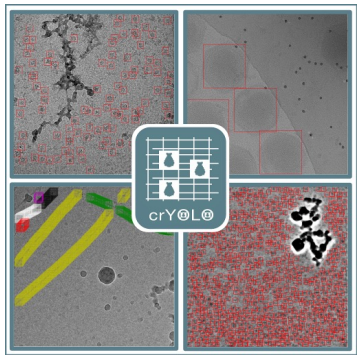
WORLDWIDE PDB PROTEIN DATA BANK

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

EMBL-EBI

EMPIAR & EMDB – Data Re-use – Machine Learning

- **DeepEMhancer:** Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Commun Biol 4, 874 (2021). https://doi.org/10.1038/s42003-021-02399-1
- **Deep Consensus:** Sanchez-Garcia, R., Segura, J., Maluenda, D., Carazo, J. M. & Sorzano, C. O. S. (2018). IUCrJ 5, 854-865.
- **DRPnet:** Nguyen, N.P., Ersoy, I., Gotberg, J. et al. DRPnet: automated particle picking in cryo-electron micrographs using deep regression. BMC Bioinformatics 22, 55 (2021). https://doi.org/10.1186/s12859-020-03948-x
- **crYOLO:** Wagner, T., Merino, F., Stabrin, M. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Commun Biol 2, 218 (2019). https://doi.org/10.1038/s42003-019-0437-z
- **WARP:** Tegunov, D., Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. Nat Methods 16, 1146–1152 (2019). https://doi.org/10.1038/s41592-019-0580-y
- **M:** Tegunov, D., Xue, L., Dienemann, C. et al. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. Nat Methods 18, 186–193 (2021). https://doi.org/10.1038/s41592-020-01054-7



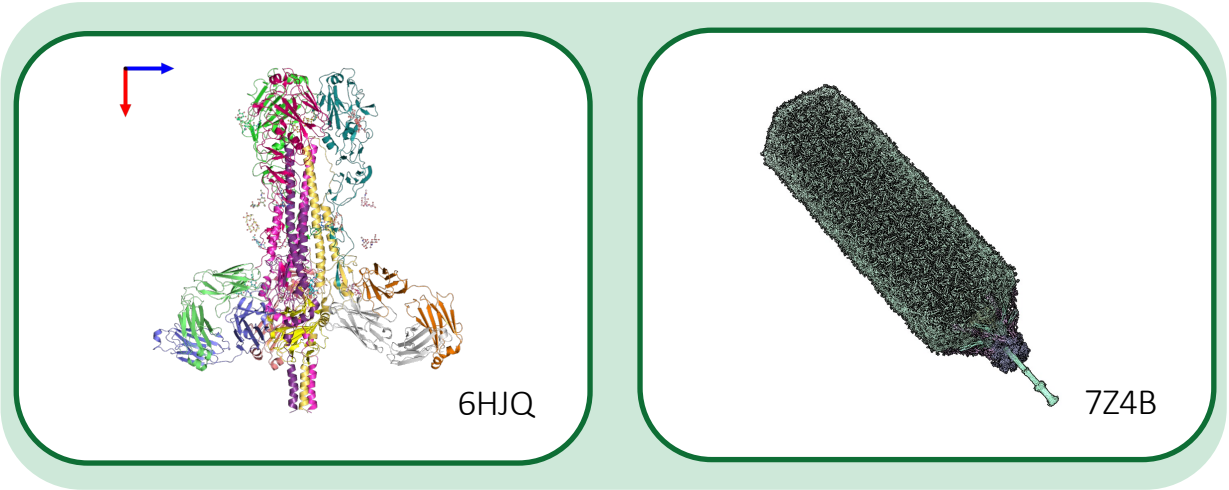
EMBL-EBI

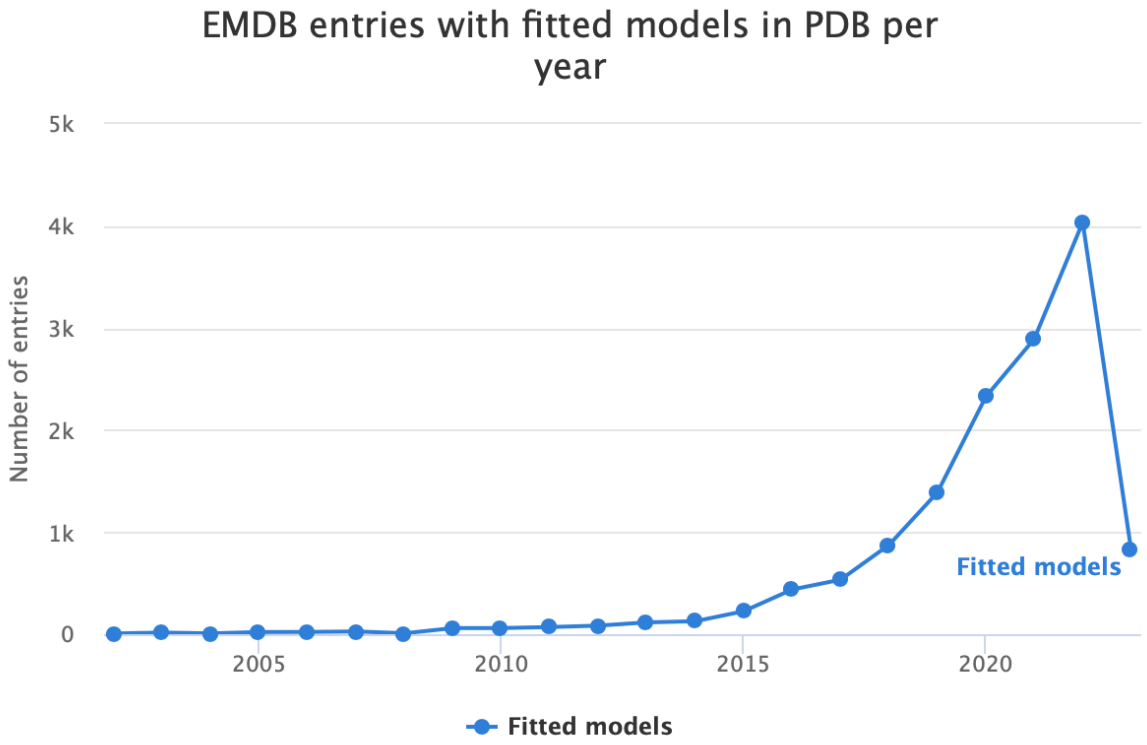
EMPIAR & EMDB – Data Re-use

- **Relion 4:** Dari Kimanius, Liyi Dong, Grigory Sharov, Takanori Nakane, Sjors H. W. Scheres; New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J* 22 December 2021; 478 (24): 4169–4185. doi: https://doi.org/10.1042/BCJ20210708
- **Relion with DeepEMhancer and SIDESPLITTER:** Ramirez-Aportela, E., Carazo, J. M. & Sorzano, C. O. S. (2022). IUCrJ 9, 632-638. https://doi.org/10.1107/S2052252522006959
- **CISTEM:** Timothy Grant, Alexis Rohou, Nikolaus Grigorieff (2018) cisTEM, user-friendly software for single-particle image processing eLife 7:e35383 https://doi.org/10.7554/eLife.35383
- **Cryoem-cloud-tools:** Cianfrocco MA, Lahiri I, DiMaio F, Leschziner AE. cryoem-cloud-tools: A software platform to deploy and manage cryo-EM jobs in the cloud. J Struct Biol. 2018 Sep;203(3):230-235. doi: 10.1016/j.jsb.2018.05.014. Epub 2018 Jun 1. PMID: 29864529; PMCID: PMC6091888.
- **cryoSPARC:** Punjani, A., Rubinstein, J., Fleet, D. et al. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290–296 (2017). https://doi.org/10.1038/nmeth.4169

Protein Data Bank - PDB

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





~14,000 entries

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

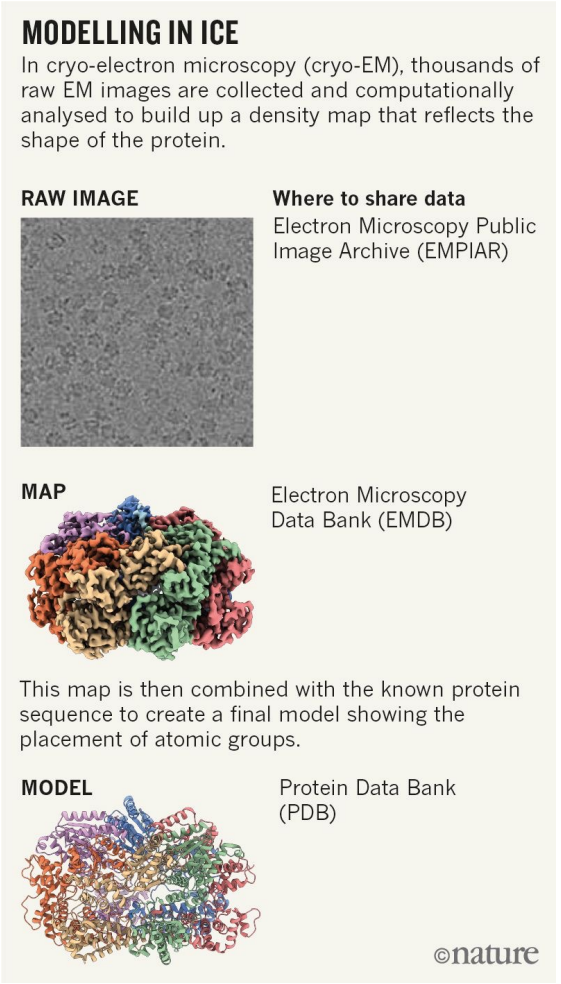


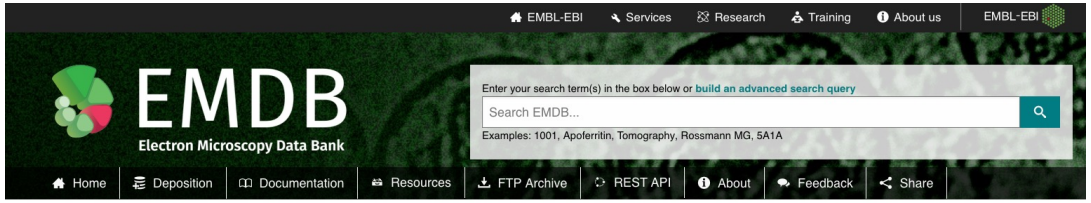
Image credit: *Nature* **561**, 565-567 (2018)

EMBL-EBI

Introduction to EMDB Website and Validation

EMBL-EBI

The EMDB website – Homepage



EMDB (the Electron Microscopy Data Bank) is a public repository for electron cryo-microscopy maps and tomograms of macromolecular complexes and subcellular structures. It covers a variety of techniques, including single-particle analysis, electron tomography, sub-tomogram averaging, fibre diffraction and electron crystallography. **More...**

As of **16 March 2022**, EMDB contains **19125** entries (**latest entries**, **trends**).

EMDB News

- wwPDB is switching to version 3 of the EMDB data model. From 9 February 2022 the old data model (v1.9.6) will no longer be supported. Read the **wwPDB announcement** for more details.
- In July 2021, the EMDB website moved to a new location. **What do I need to know?**

Browse EMDB | EMDB statistics | SARS-CoV-2 entries | Deposit data

Explore EMDB

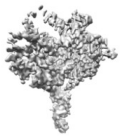
Select a graph from the pull-down menu. Click on any element of a graph to see the corresponding list of EMDB entries.

Sample type

Quick Links

- EMDB Policies
- Talks & Tutorials
- Validation Analysis
- EMDB Citations
- EMPIAR
- PDBe
- BioImage Archive
- EMDataResource
- EM Navigator
- 3D EM History

Recent Entries (Show all)

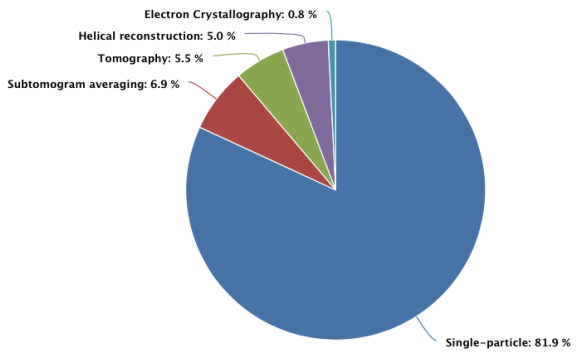


EMD-32390 [1/95]
PlmCasX-sgRNAv1-dsDNA ternary complex at ts loading state

Explore EMDB

Select a graph from the pull-down menu. Click on any element of a graph to see the corresponding list of EMDB entries.

EM Method

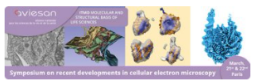


EMD-32390 [1/95]
PlmCasX-sgRNAv1-dsDNA ternary complex at ts loading state

View Entry

Tweets by @EMDB_EMPIAR

EMDB - EMPIAR @EBI
@EMDB_EMPIAR
Where cryo-EM meets volume
EMIdevcellem.sciencesconf.org



29m

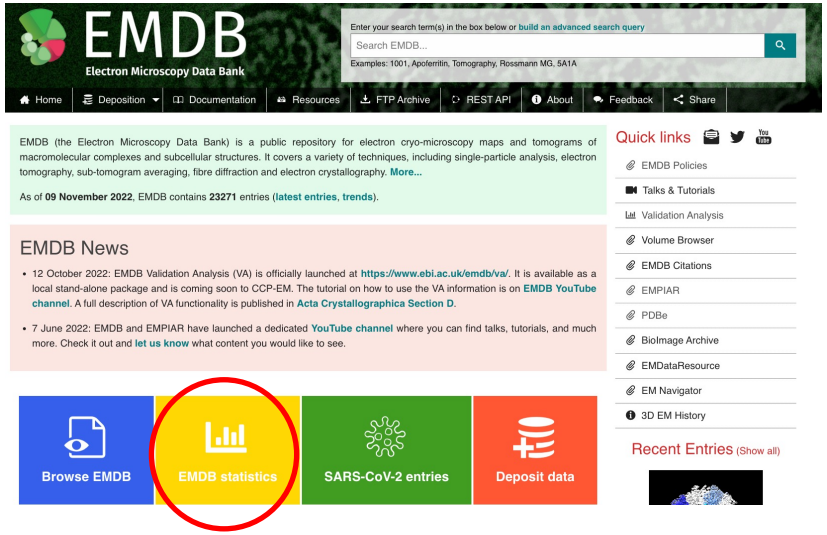
EMDB - EMPIAR @EBI
@EMDB_EMPIAR
This freshly released #EMPIAR entry

Embed View on Twitter

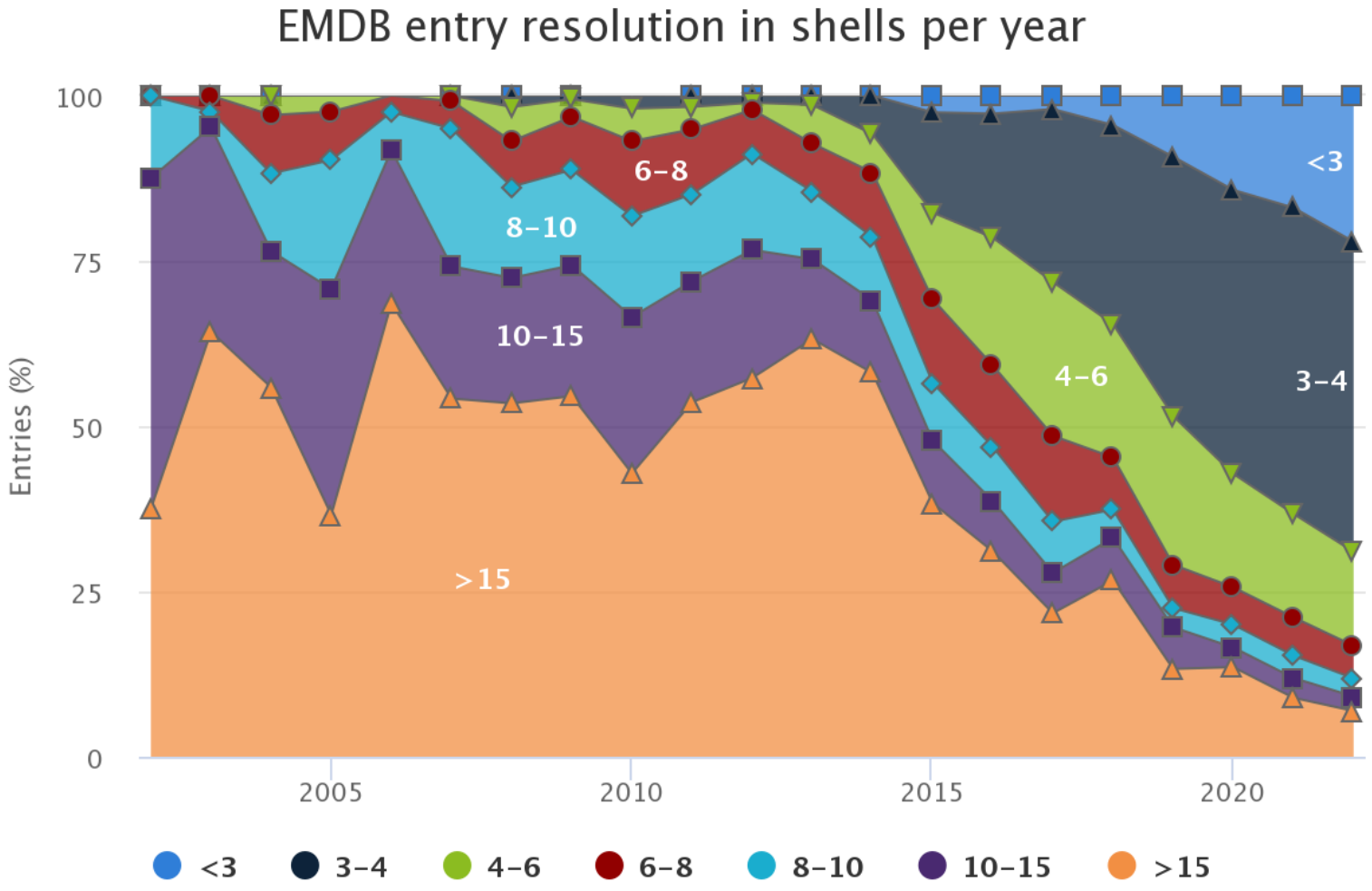
https://www.ebi.ac.uk/emdb/

EMBL-EBI

The EMDB website – Statistics



EMDB: ebi.ac.uk/emdb/



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

EMBL-EBI

The EMDB website – Chart Build

Chart builder

Beta version

Archive: EMDB

Chart type: Line

Data X: Year | Initial year: 2002 | Final year: 2022

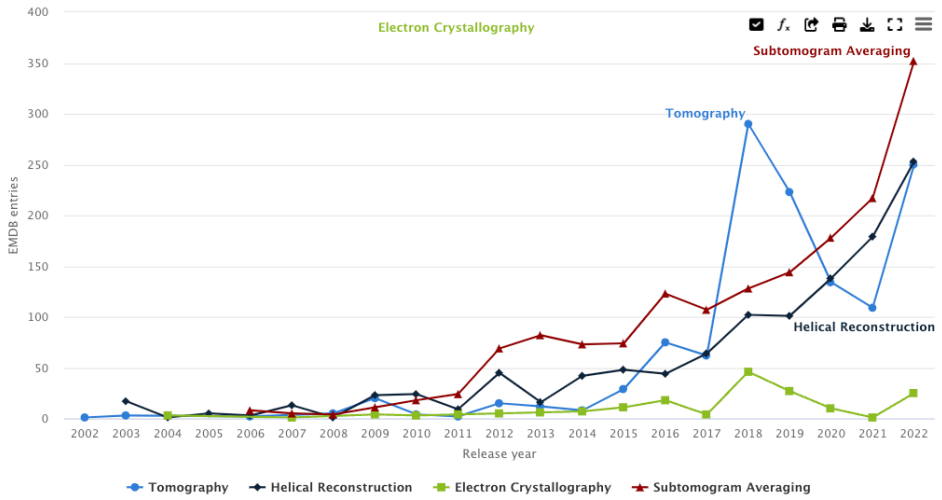
Data Y: EMDB entries

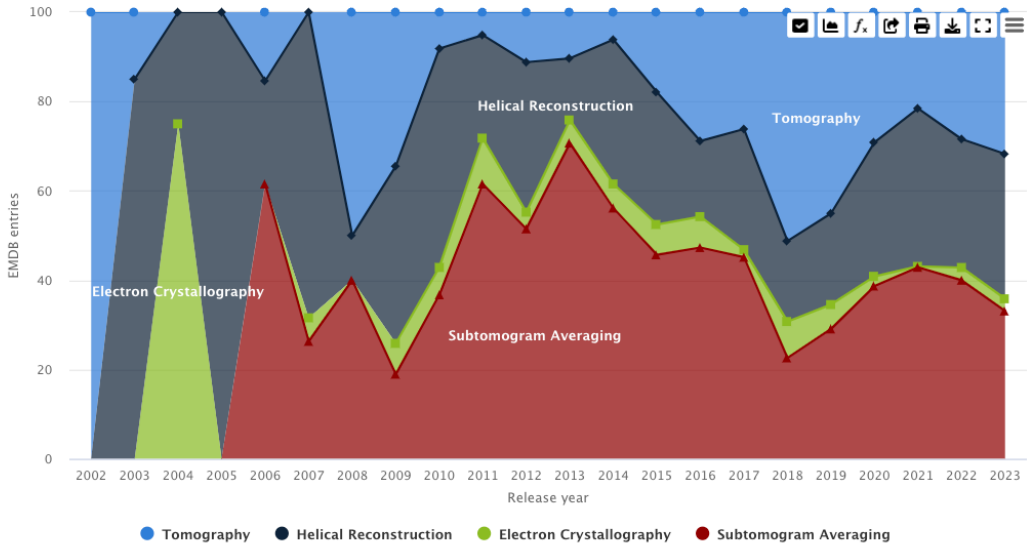
Data series

Color	Operator	Filter query	Label
	Count (unique)	structure_determination_method:"tomography"	Tomography
	Count (unique)	structure_determination_method:"helical"	Helical Reconstruction
	Count (unique)	structure_determination_method:"electroncryst	Electron Crystallography
	Count (unique)	structure_determination_method:"subtomogram	Subtomogram Averaging

Additional options

Generate

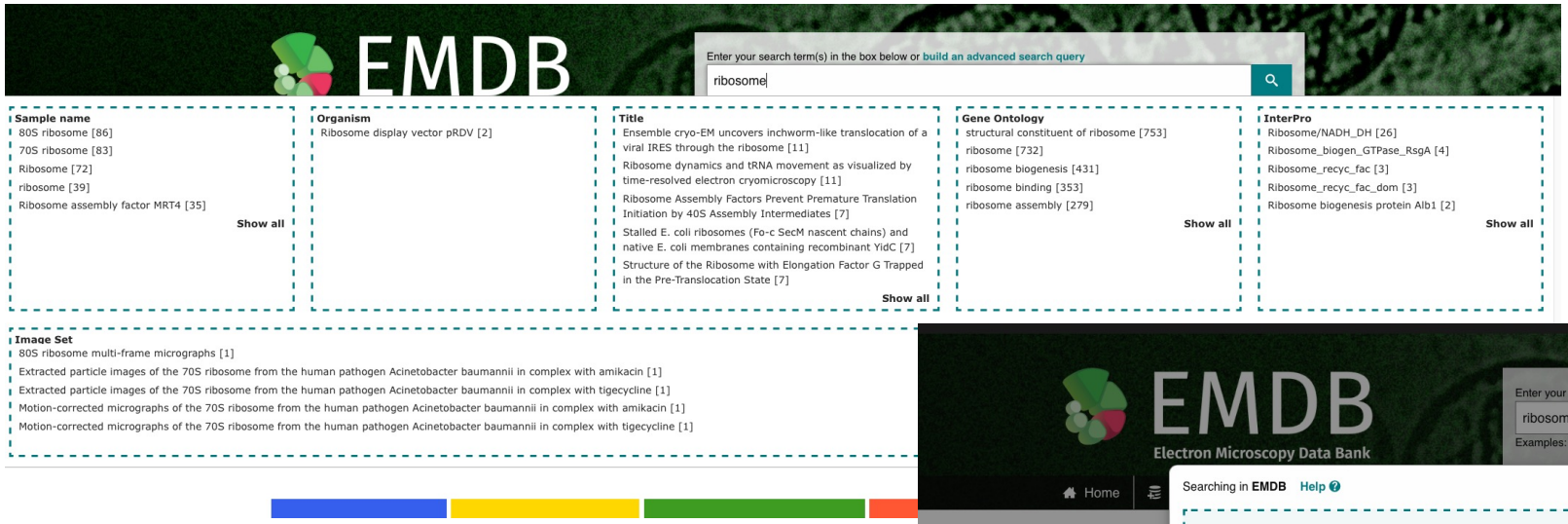


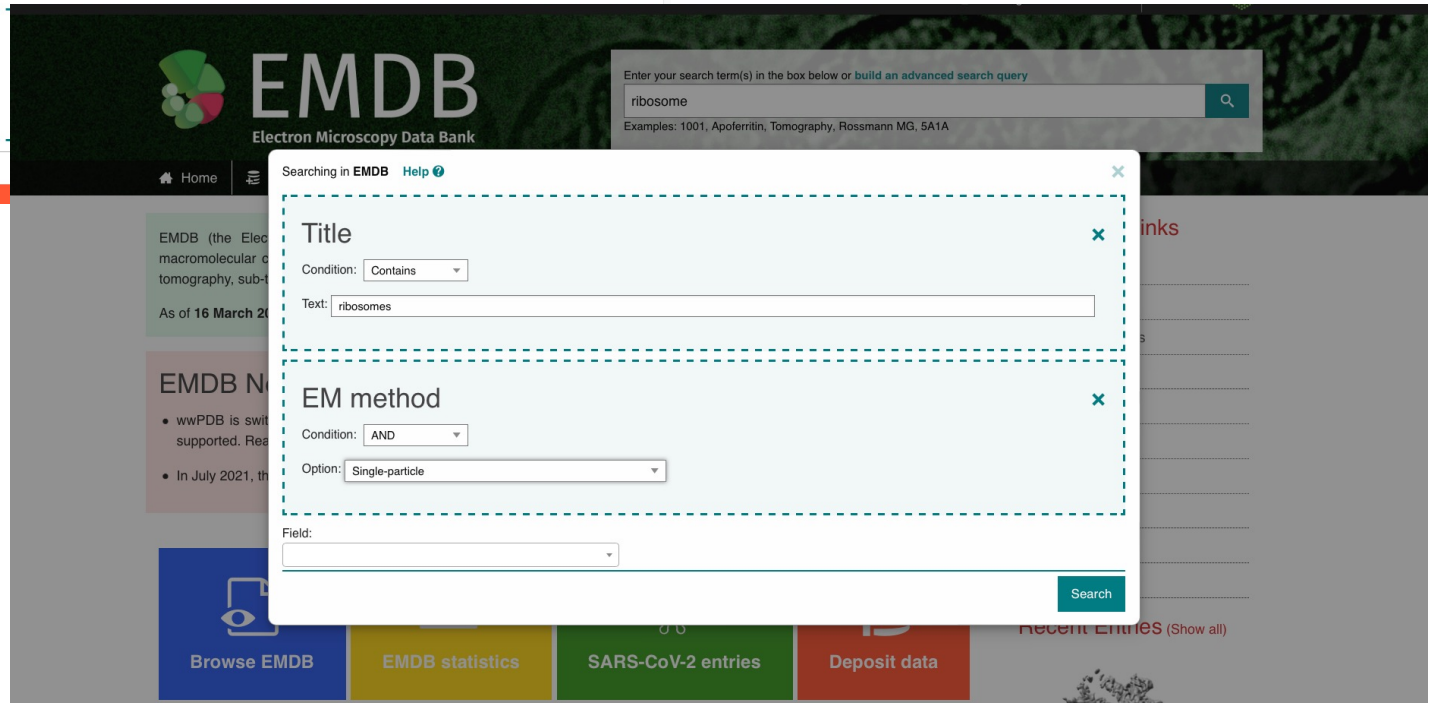


https://www.ebi.ac.uk/emdb/statistics/builder

EMBL-EBI

The EMDB website – Basic and Advanced Search





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

EMBL-EBI

The EMDB website – Filtering & Statistics

Edit and save complex search queries

Edit your search query here and hit Return to execute it:

```
HIV-1 AND structure_determination_method:"tomography"
```

Extensive list of filtering options

Filter By

Current Database
- EMDB [26]

Current status
- REL [26]

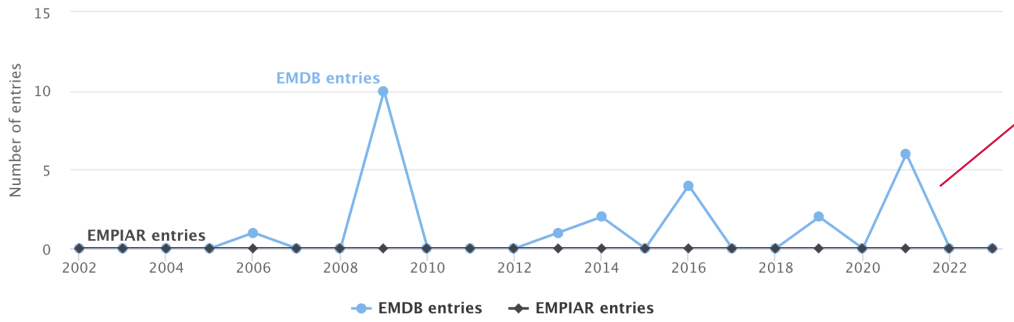
Sample type
- Virus [25]
- Sample [14]
- Cellular component [1]

Organism
- Human immunodeficiency virus 1 [24]
- Mus musculus [1]
- Human immunodeficiency virus [1]

EM Method
- Tomography [26]

Model molecular weight

Number of entries released by year

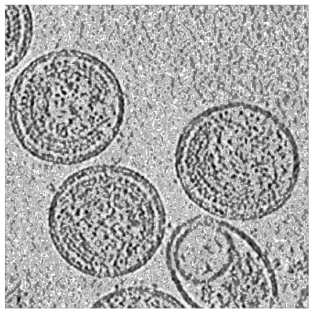


Interactive graph that can be used to filter results

Sort by: Release Date ▲ Download ◀ Showing 1 - 10 of 26. Page 1 of 3 ▶ Display: 10 Statistics

EMD-13085

Download Map



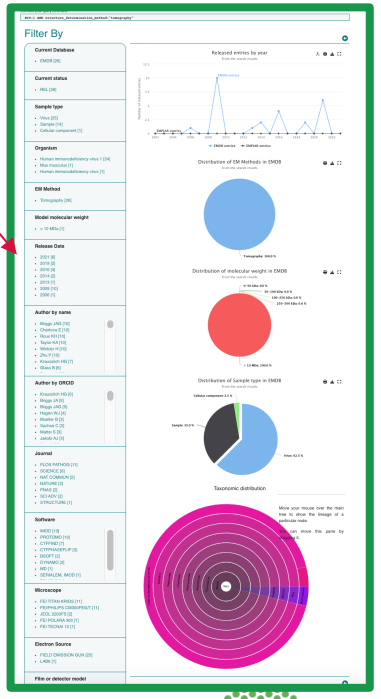
Representative cryo-electron tomogram of immature HIV-1 particles

Release date: Aug. 18, 2021
EM Method: Tomography

Qu K, Ke Z, Zila V, Anders-Osswein M, Glass B, Mucksch F, Muller R, Schultz C, Muller B, Krausslich HG, Briggs JAG

Science (2021) **373** pp. 700-704 [**Pubmed:** 34353956 **DOI:** doi:10.1126/science.abe6821]

Sample:
- **Human immunodeficiency virus 1** (**Virus** from **Human immunodeficiency virus 1**)



Download

Select database:

◉ EMDB [26]

Format: JSON (metada ▼

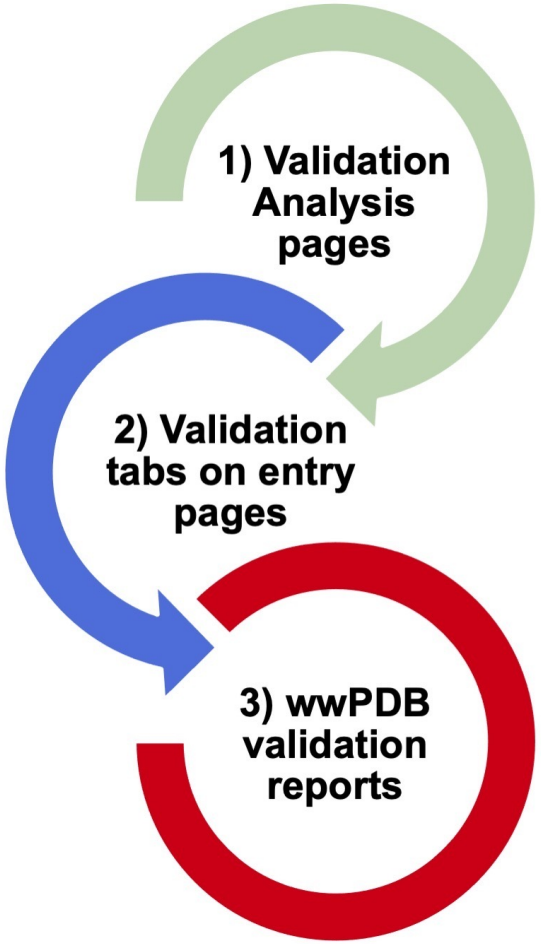
Preview first 10

Download

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

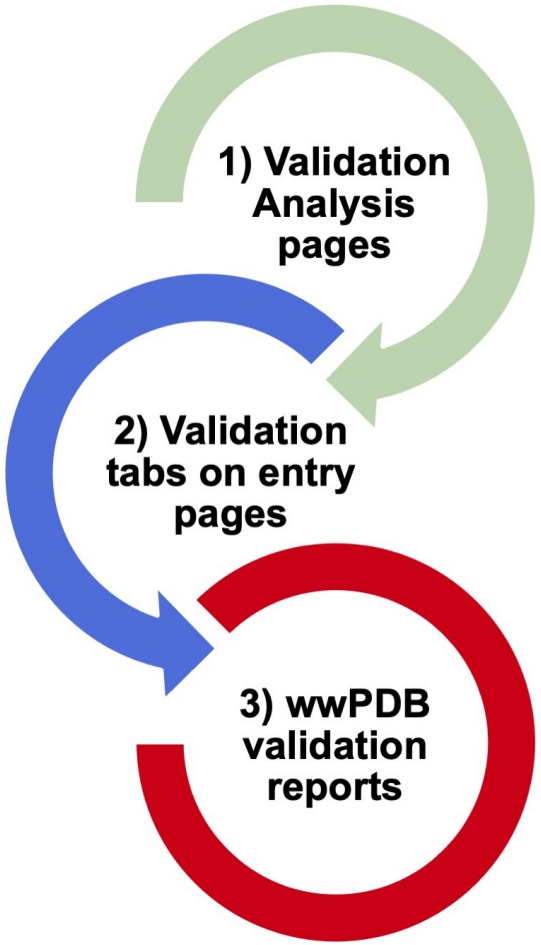
EMBL-EBI

EMDB Validation



EMBL-EBI

EMDB Validation



EMD-11145 › Validation analysis

To see the Validation Analys
entry, please enter the EMD
(e.g., 8117 or 5irx):

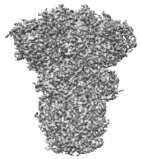
SARS CoV-2 Spike protein, Closed conformation, C3 symmetry

Additional validation information

For more information, please see the wwPDB validation report for this entry with fitted PDB model **6zb5**.

Resolution:	**2.85 Å (FSC 0.143 CUT-OFF, depositor provided)**
Method:	**Single particle reconstruction**
Map released:	**2020-09-30**
Last modified:	**2020-11-18**
Sample name:	SARS CoV-2 Spike protein, Closed conformation, C3 symmetry
Organism:	Severe acute respiratory syndrome coronavirus 2
Fitted atomic model:	6zb5

Map parameters

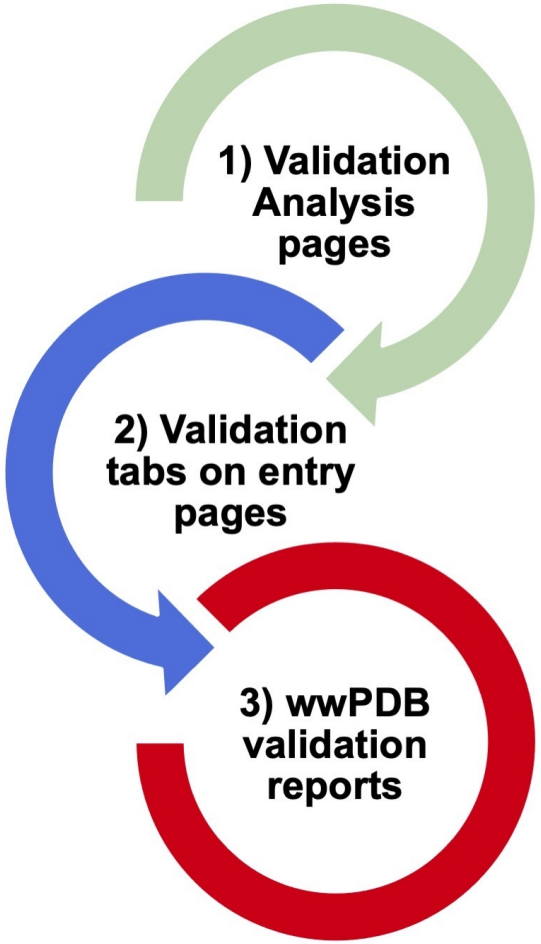


Recommended contour level:	0.006
Number of grid points:	220 × 220 × 220
Voxel size:	1.050 × 1.050 × 1.050 Å
Minimum value:	-0.197
Maximum value:	0.316
Average value:	0.001
Standard deviation:	0.012

ebi.ac.uk/emdb/va/

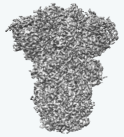
EMBL-EBI

EMDB Validation



EMD-11145

Single-particle
2.85 Å



3D View | Gallery

Deposition: 07/06/2020
Map released: 30/09/2020
Last modified: 18/11/2020

Overview | 3D View | Sample | Experiment | Validation | Volume Browser | Additional data | Links

EMD-11145

Download

SARS CoV-2 Spike protein, Closed conformation, C3 symmetry

Additional validation information

For more information, please see the wwPDB validation report for this entry with fitted PDB model **6zb5**. Cryo-EM specialists may also be interested in the more extensive analysis of this entry in the **EMDB Validation Analysis** resource.

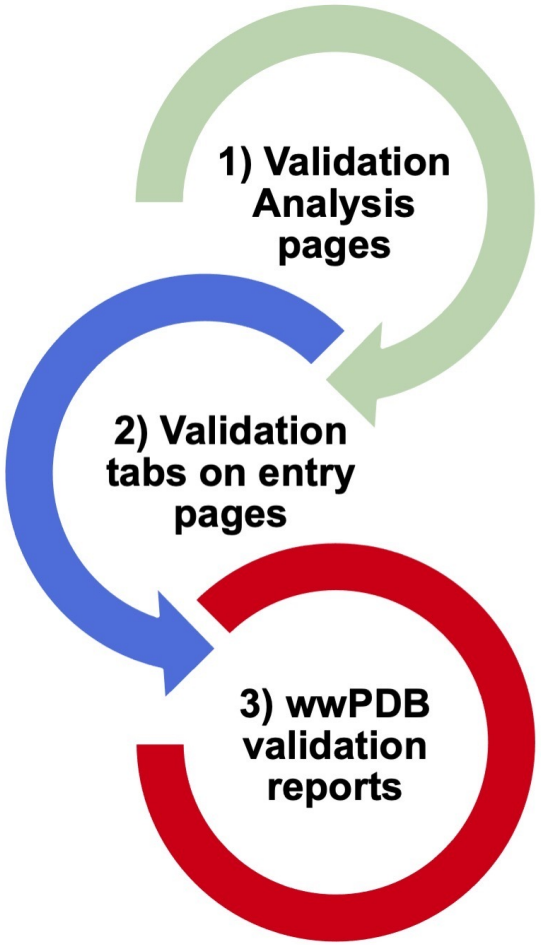
Resolution:	**2.85 Å (FSC 0.143 CUT-OFF, depositor provided)**
Method:	**Single particle reconstruction**
Map released:	**2020-09-30**
Last modified:	**2020-11-18**
Sample name:	SARS CoV-2 Spike protein, Closed conformation, C3 symmetry
Organism:	Severe acute respiratory syndrome coronavirus 2
Fitted atomic model:	**6zb5**

Map parameters

Recommended contour level:	0.006
Number of grid points:	220 × 220 × 220

EMBL-EBI

EMDB Validation

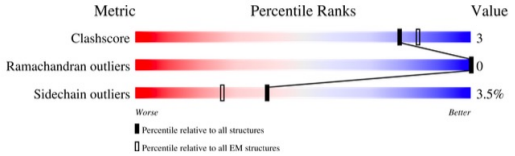


1 Overall quality at a glance

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

The reported resolution of this entry is 2.85 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.

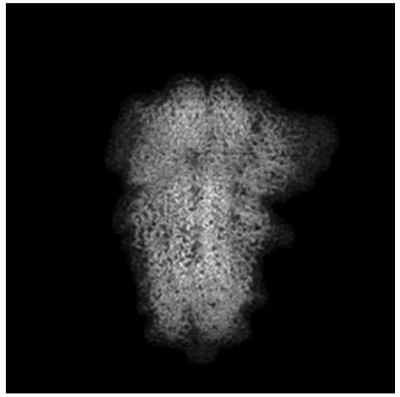
Mol	Chain	Length	Quality of chain
1	A	1259	76% 6% 18%
1	B	1259	76% 6% 18%
1	C	1259	76% 6% 18%
2	D	2	50% 50%
2	E	2	100%
2	F	2	100%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

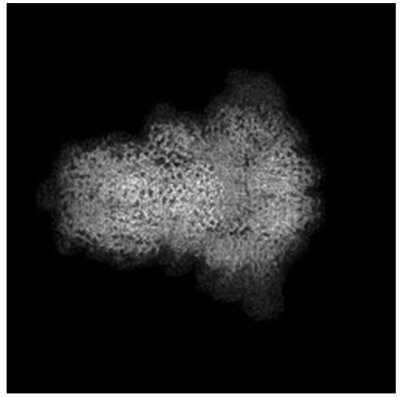
EMBL-EBI

EMDB Validation

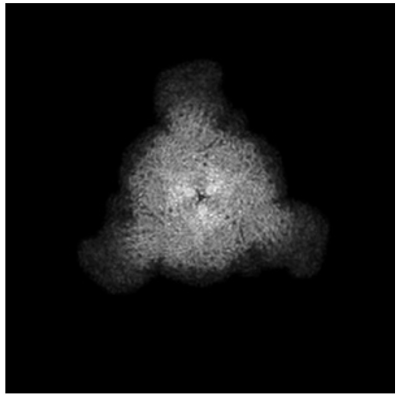
Qualitative



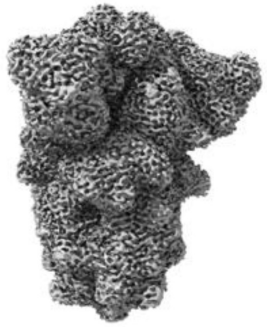
X



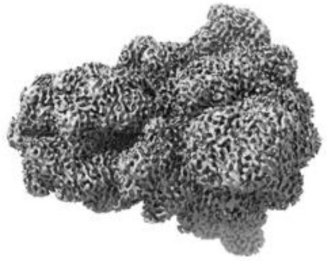
Y



Z



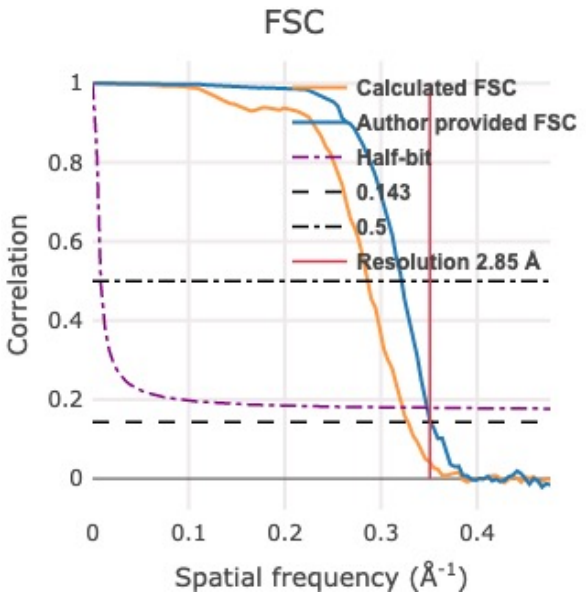
X

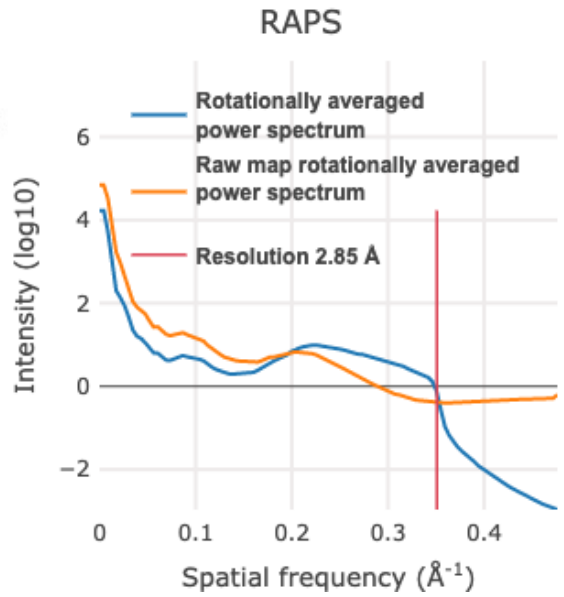


Y

Z

Quantitative



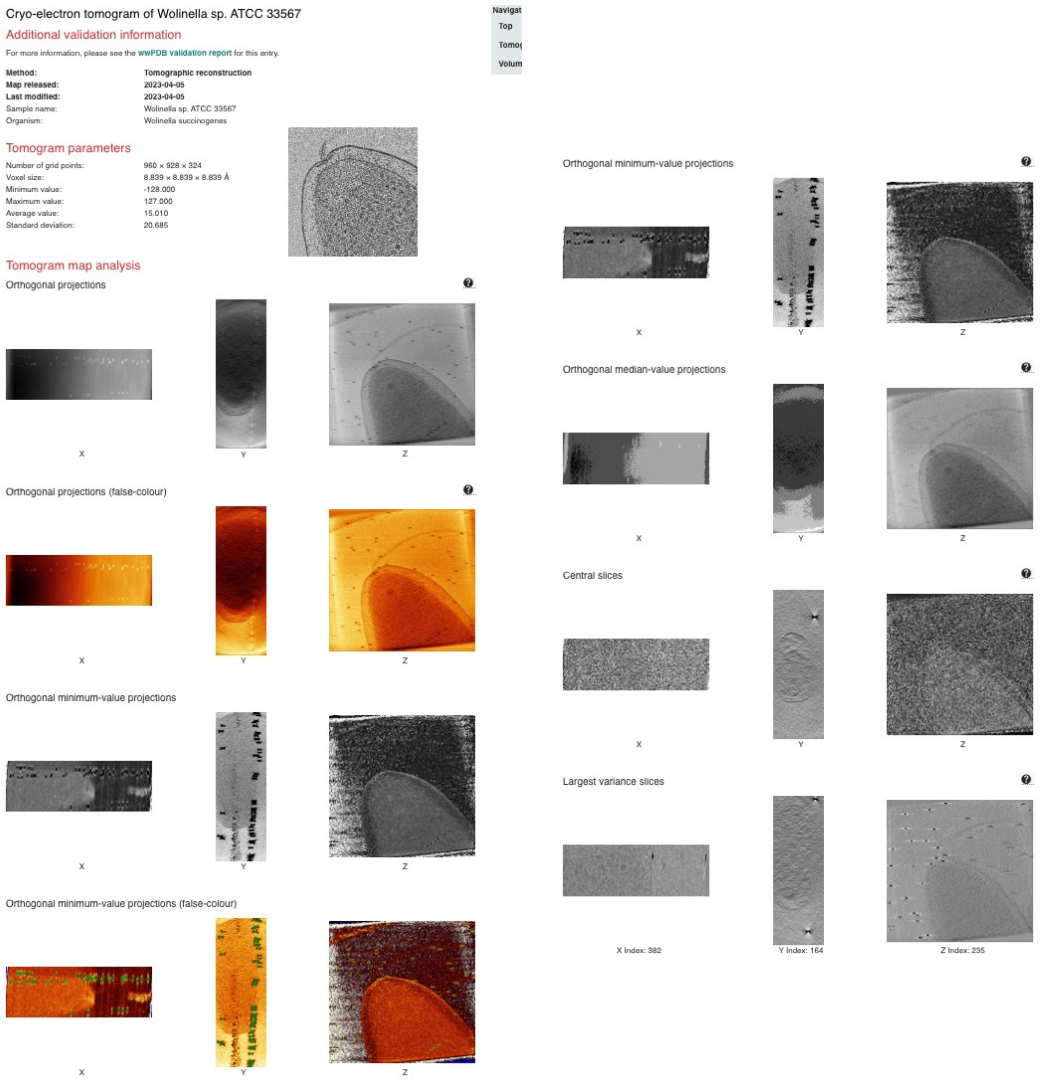


EMD-11145

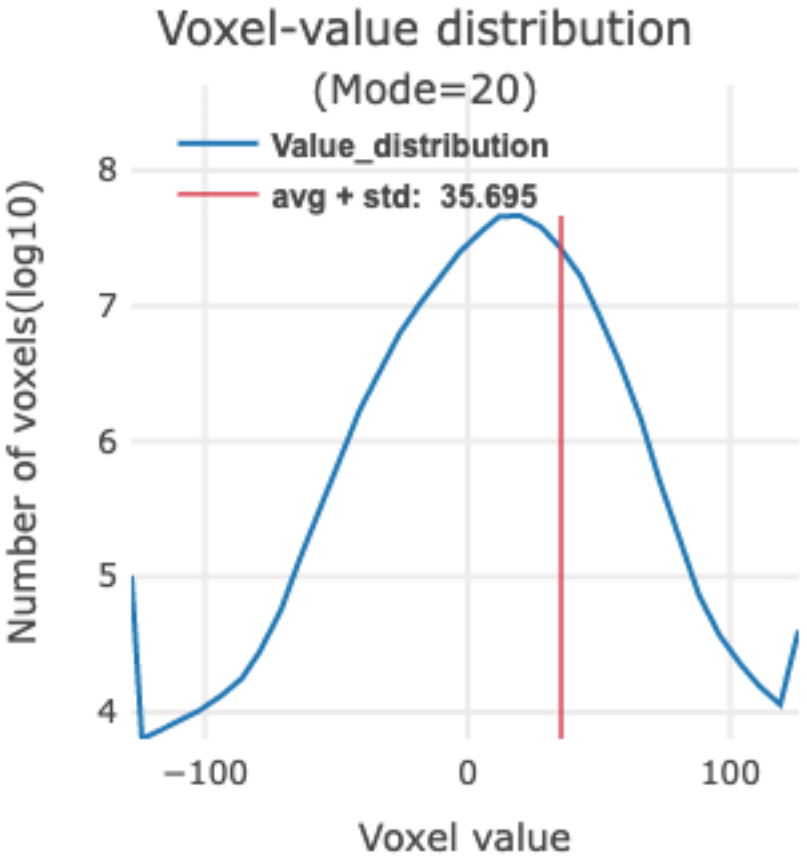
EMBL-EBI

EMDB Validation - Tomography

Qualitative



Quantitative



EMD-15449

EMBL-EBI

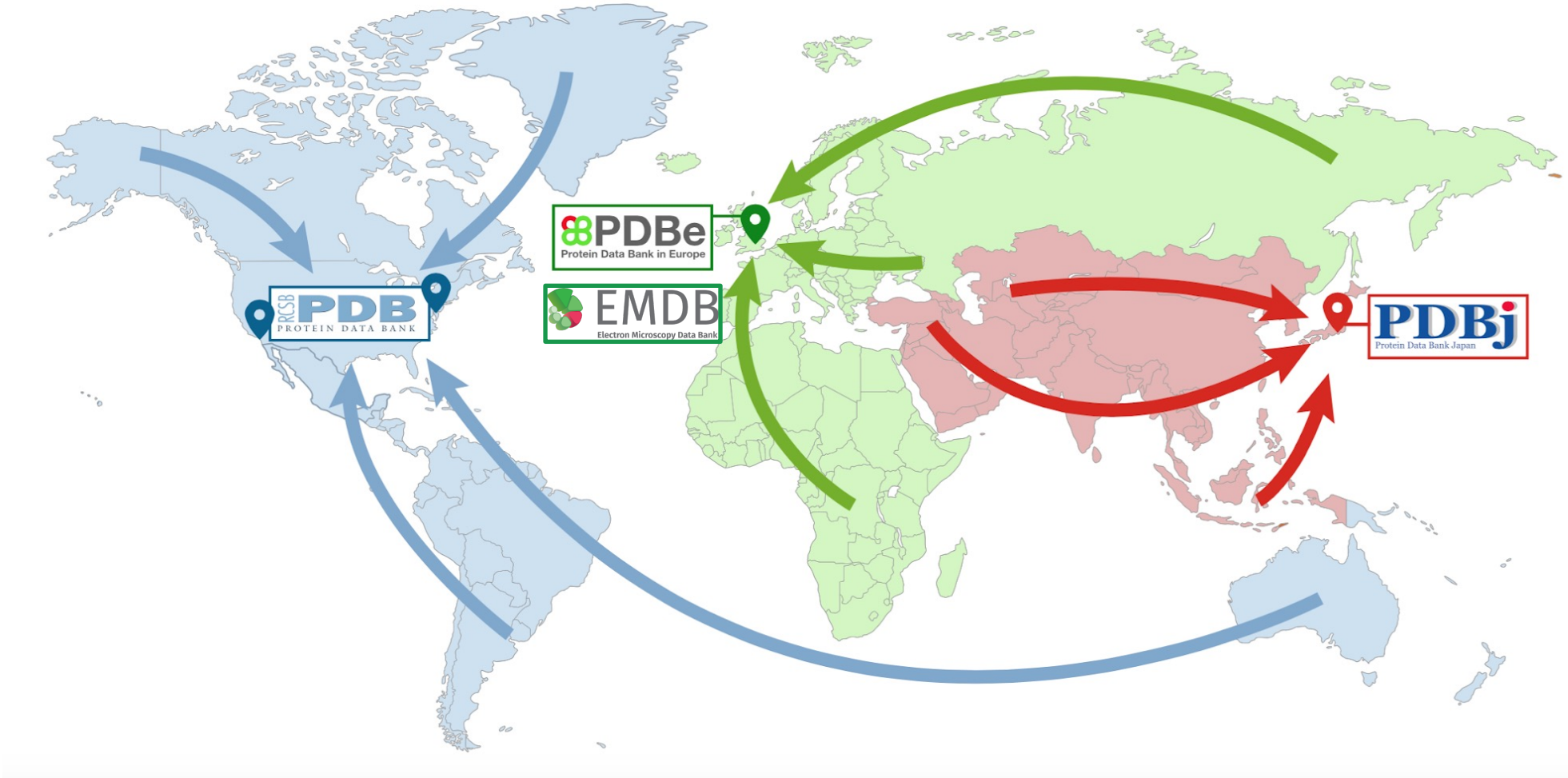
EMDB Validation – Full List

- EM Map view
  - Orthogonal projections x 7
  - Central Slices
  - Largest variance slices
  - Orthogonal surface views
  - EM map with masks
- Volume Graphs
  - Voxel-value distribution
  - Volume estimate
  - RAPs
  - FSC
- Fitted Model Analysis
  - Primary map with model
  - Atom inclusion
  - Map-model FSC
  - Q-score

Deposition

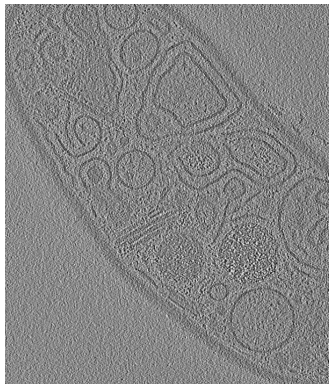
EMBL-EBI

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

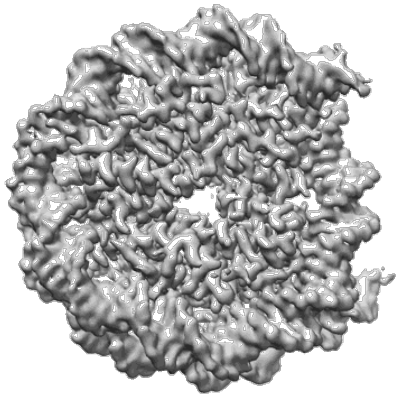


EMDB Deposition - Types

Map Only



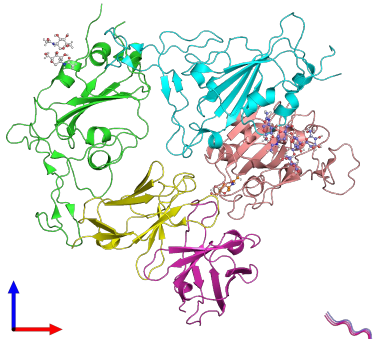
EMD-16202



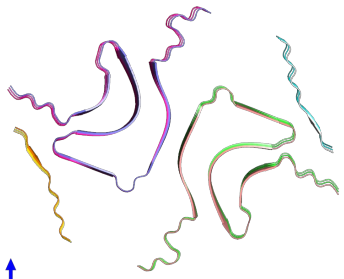
EMD-29854

All Methodologies

Model Only



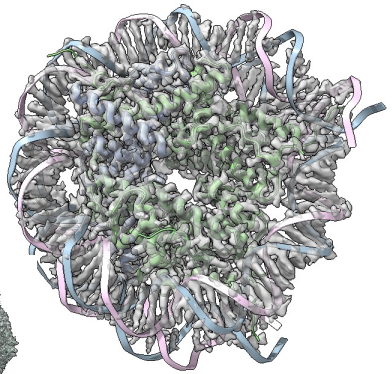
7X93



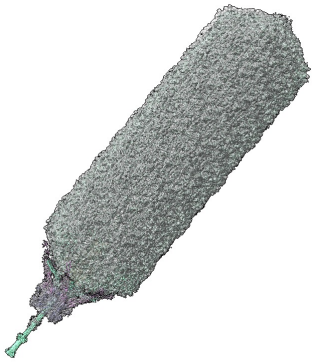
7YK8

All methodologies
except Tomography

Map + Model



EMD-29837



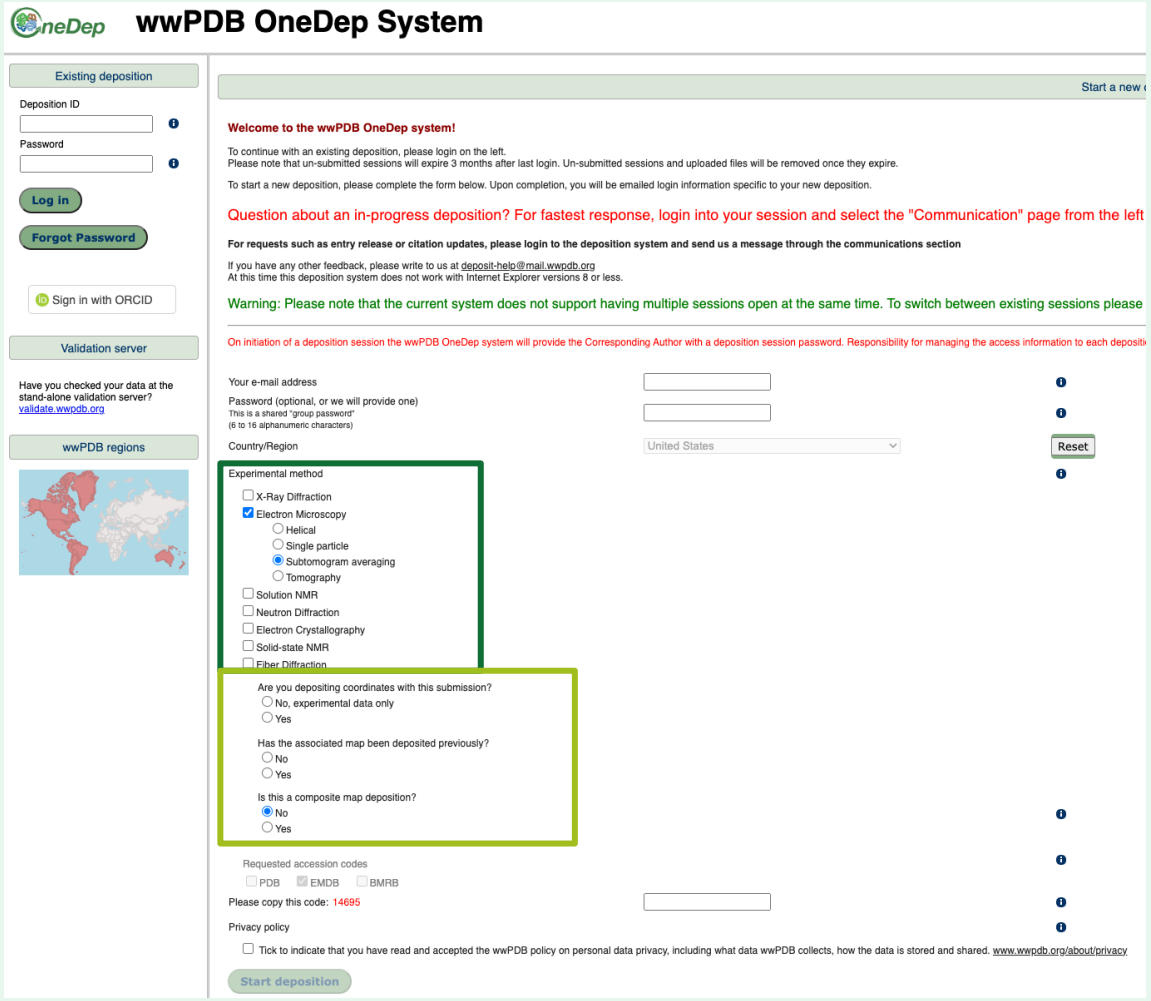
EMD-14492

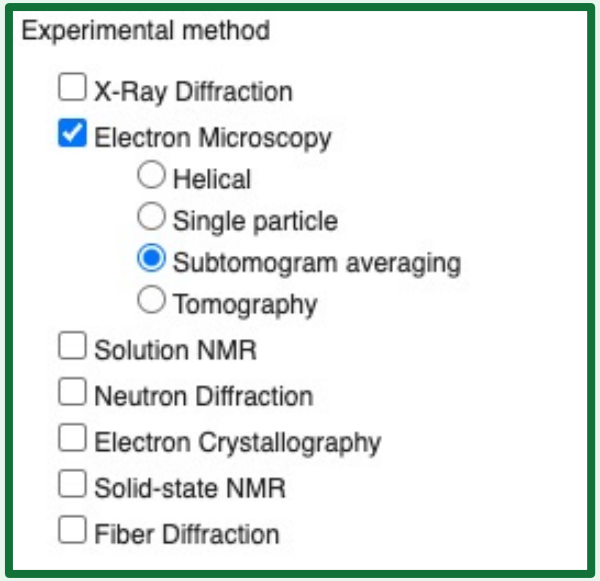
All methodologies
except Tomography

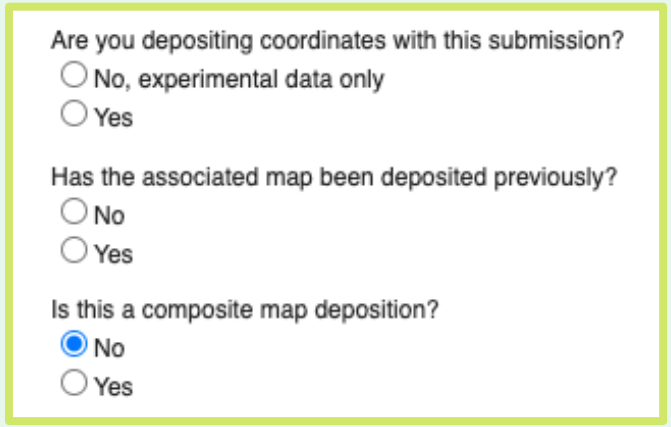
Must have associated map!

EMBL-EBI

EMDB Deposition – Brain Page

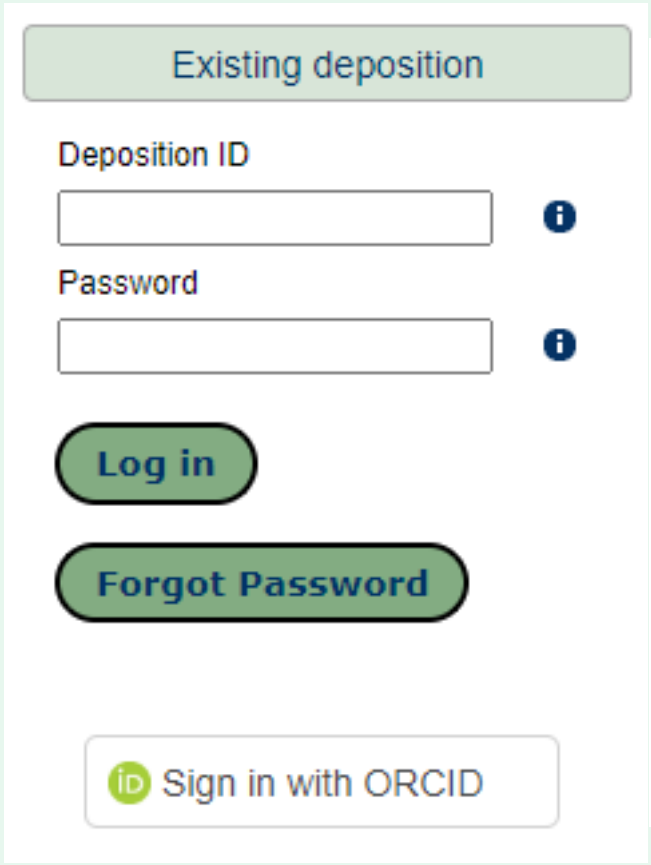






EMBL-EBI

EMDB Deposition - ORCID Login

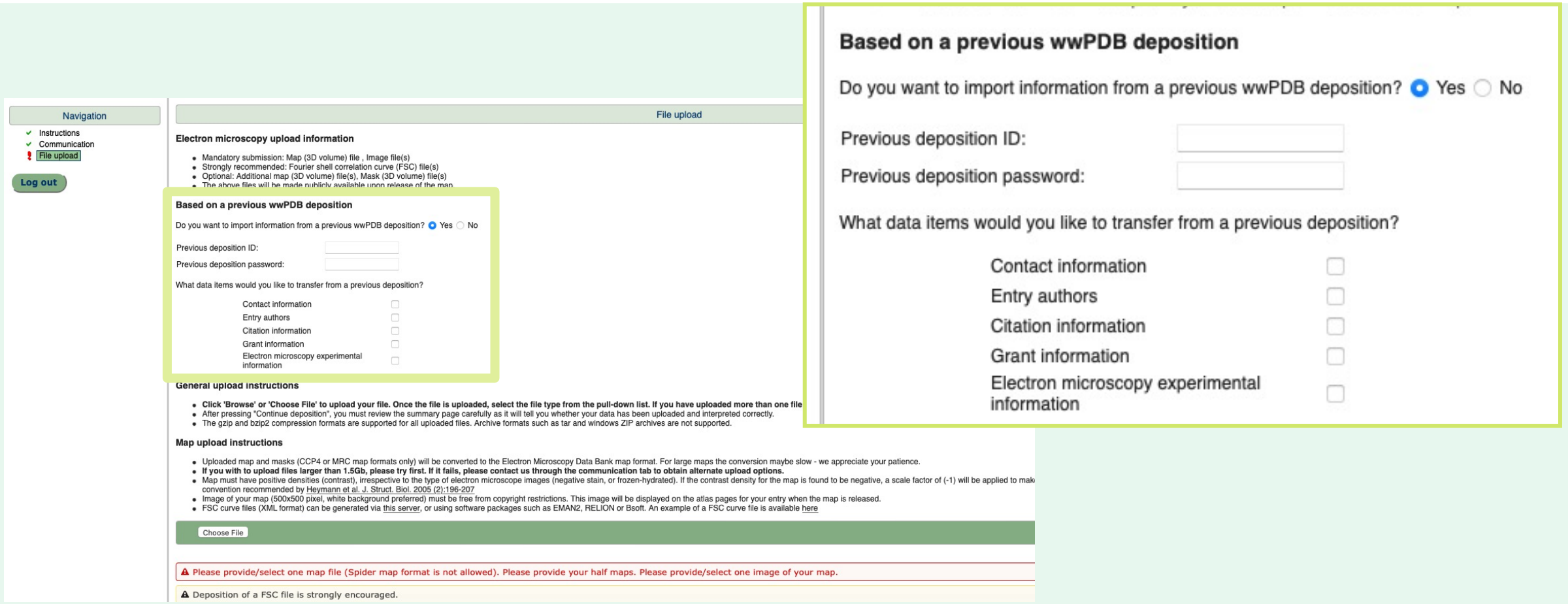


Deposition list

Depositions available to 0000-0012-3456-789X

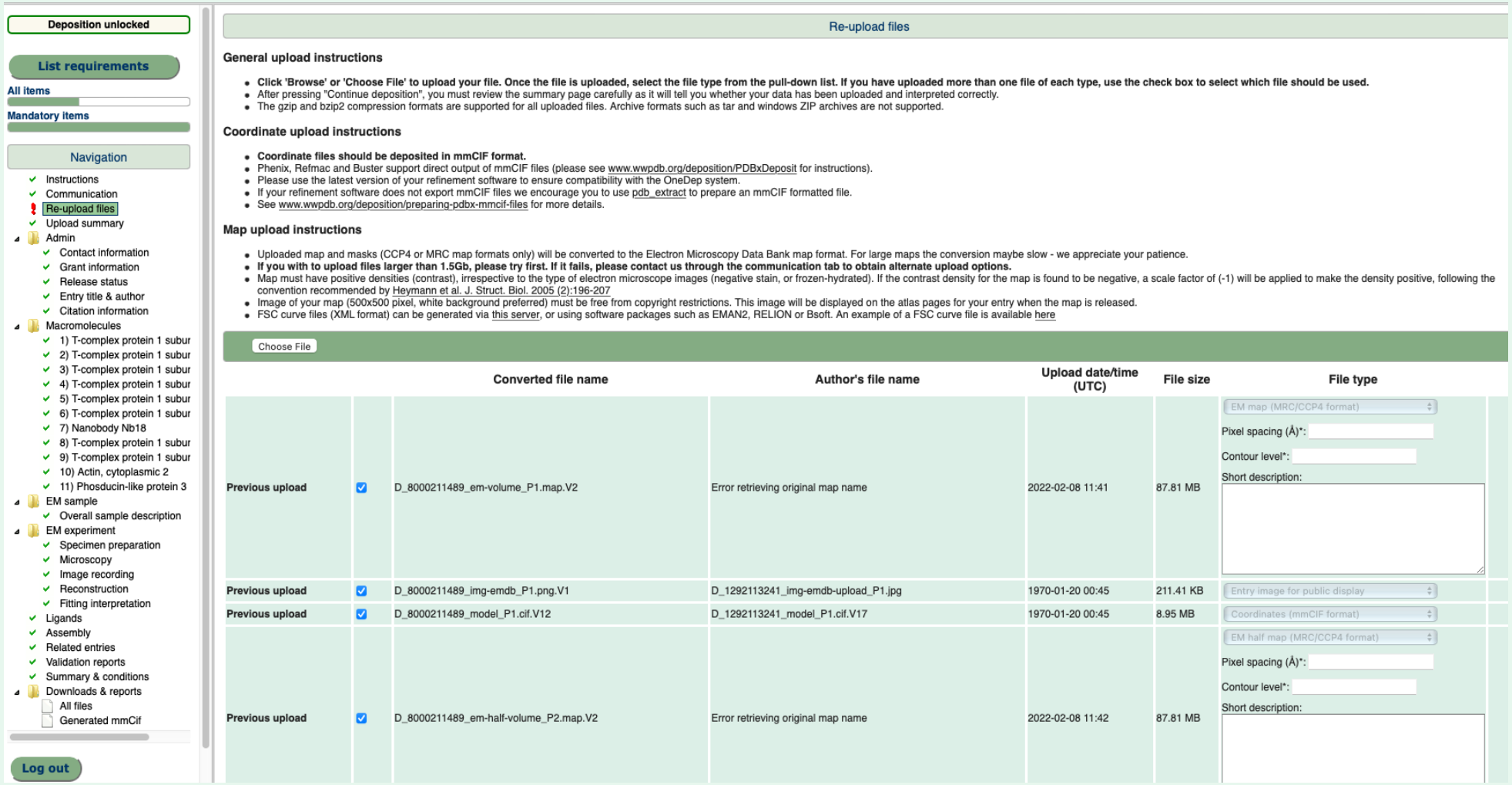
Deposition ID	Entry ID	Entry Title	Created	Site	Status	Last login
D_1292111636	7axh	Crystal structure of the hPXR-LBD in complex with alpha-zearalanol	17/03/2023	RCSB	PROC	17/03/2023
D_1292111623	7ax8	Crystal structure of the hPXR-LBD in apo form (P43212 SG)	09/01/2023	PDBe	AUTH	16/03/2023
D_1292100980	6qvt	CMP-Sialic acid bound structure the human ST6Gal1	14/07/2022	PDBj	HPUB	22/09/2022
D_1292100979	6qvs	Unliganded structure of the human wild type ST6Gal1...	04/03/2019	PDBe	REL	15/05/2020
D_1200009063	6fwu	Crystal structure of human wild type B4GalT1 in apo-closed dimeric form	07/03/2018	PDBe	REL	03/02/2019
D_1290050811	4adp	HCV-J6 NS5B POLYMERASE V405I MUTANT	02/01/2012	RCSB	REL	-
D_1290044331	2xi3	HCV-H77 NS5B Polymerase Complexed With GTP	25/06/2010	PDBj	REL	-
D_1290044262	2xhu	HCV-J4 NS5B Polymerase Orthorhombic Crystal Form	21/06/2010	RCSB	REL	-

EMDB Deposition – File Upload Page



EMBL-EBI

EMDB Deposition – Deposition Interface



EMBL-EBI

EMDB Deposition – Tips & Tricks

- mmCIF is a metadata file!
- The provided sequence should be the full sequence in the sample, including any unmodeled regions.
- Half-maps must be unmasked, unfiltered and unsharpened raw maps.
- Defocus values must be in nm and positive values refer to defocus.
- Maps and models should all be present in the same coordinate space and overlay correctly.
- When depositing volume EM data make use of EMPIAR

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

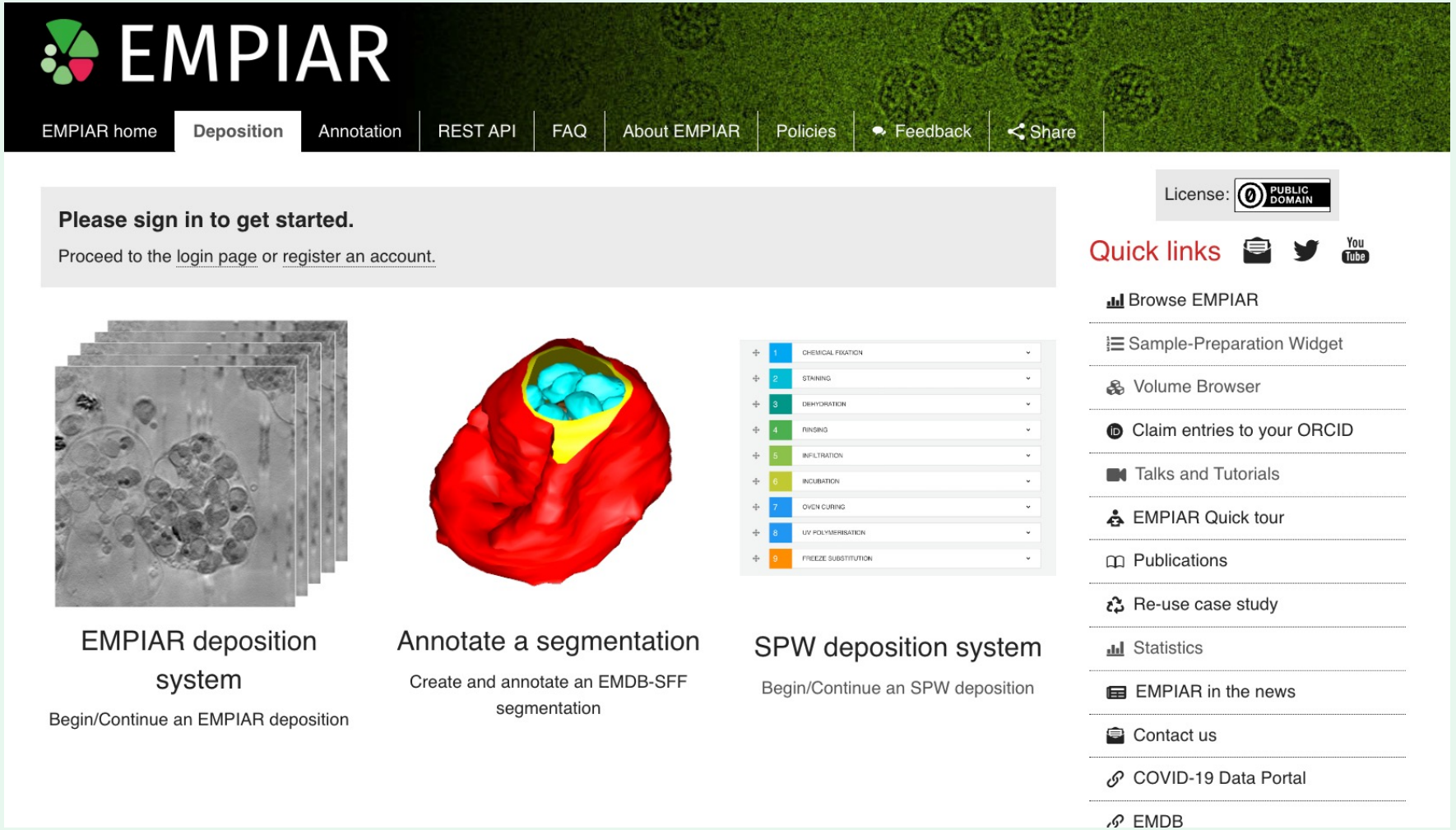
EMPIAR Deposition Policies

- 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

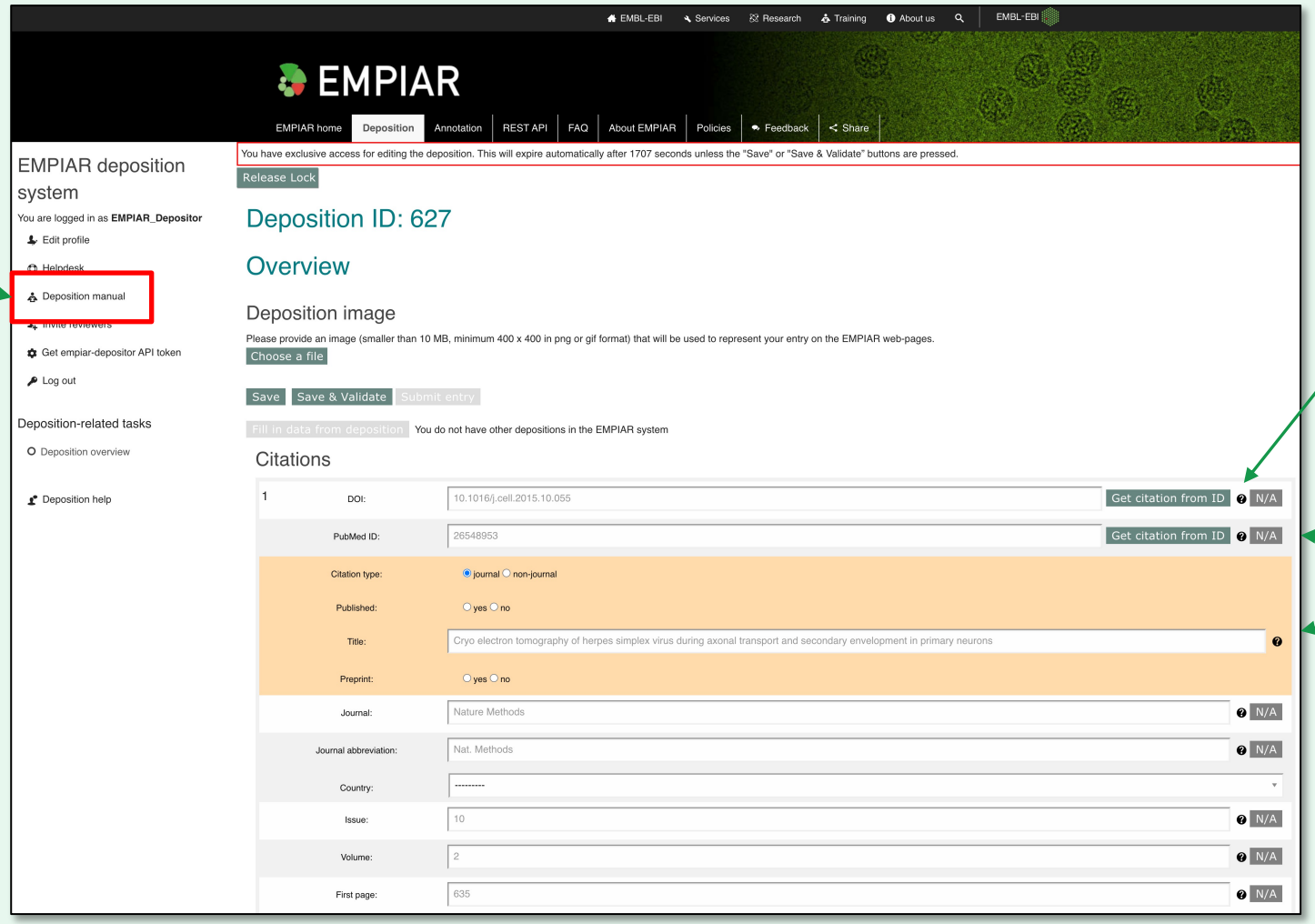


https://www.ebi.ac.uk/empiar/deposition/choose-action/

EMBL-EBI

EMPIAR Deposition Overview Page

Deposition manual available if you get stuck



Citation can be pulled from DOI or PubMed ID

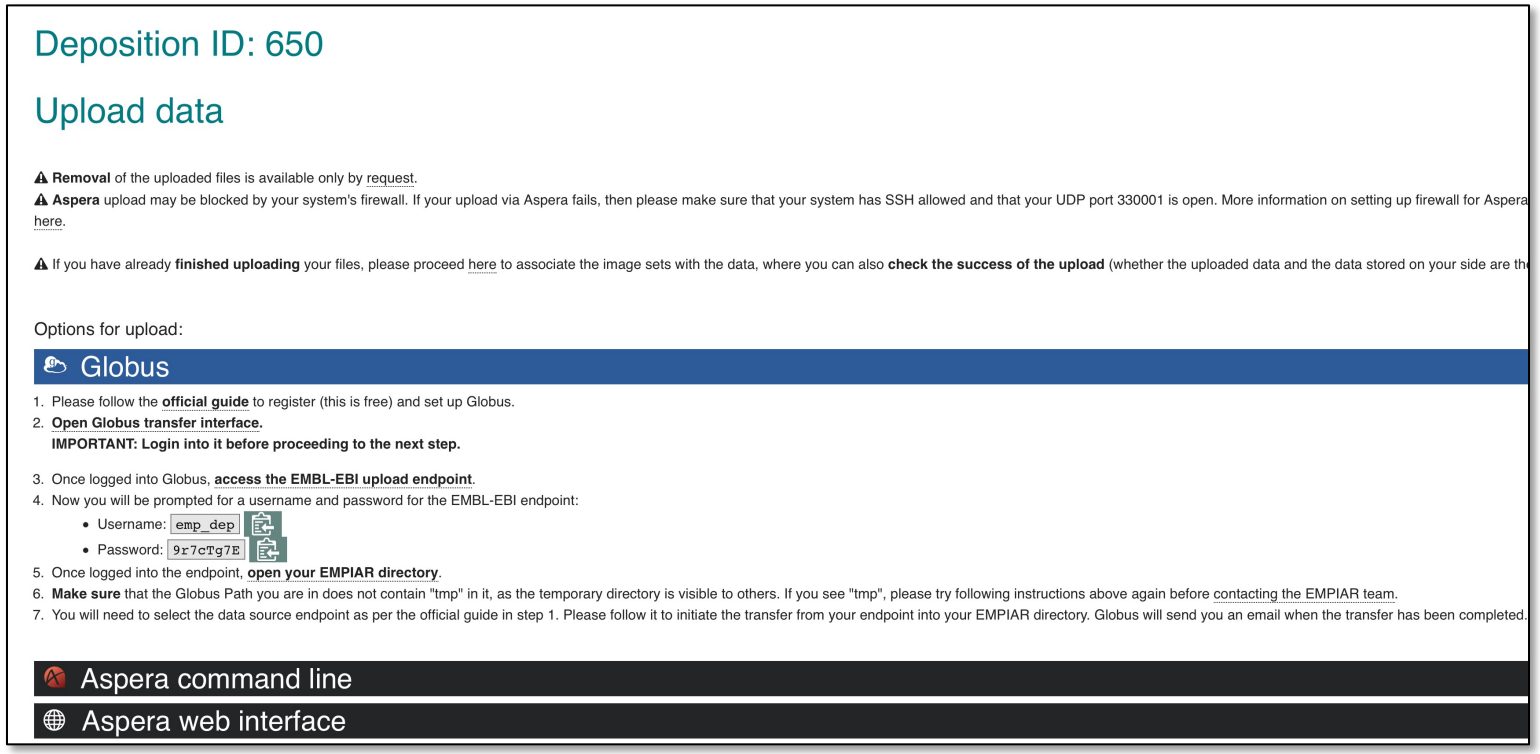
Grey = Optional
Orange = Mandatory
(N/A is clickable for optional fields)

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

EMBL-EBI

EMPIAR Deposition - File Upload

- Aspera
- Globus
- FTP (Not recommended)



EMBL-EBI

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)

Questions!

EMDB Website

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

EMDB Chart Builder

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

Feedback

EMBL-EBI