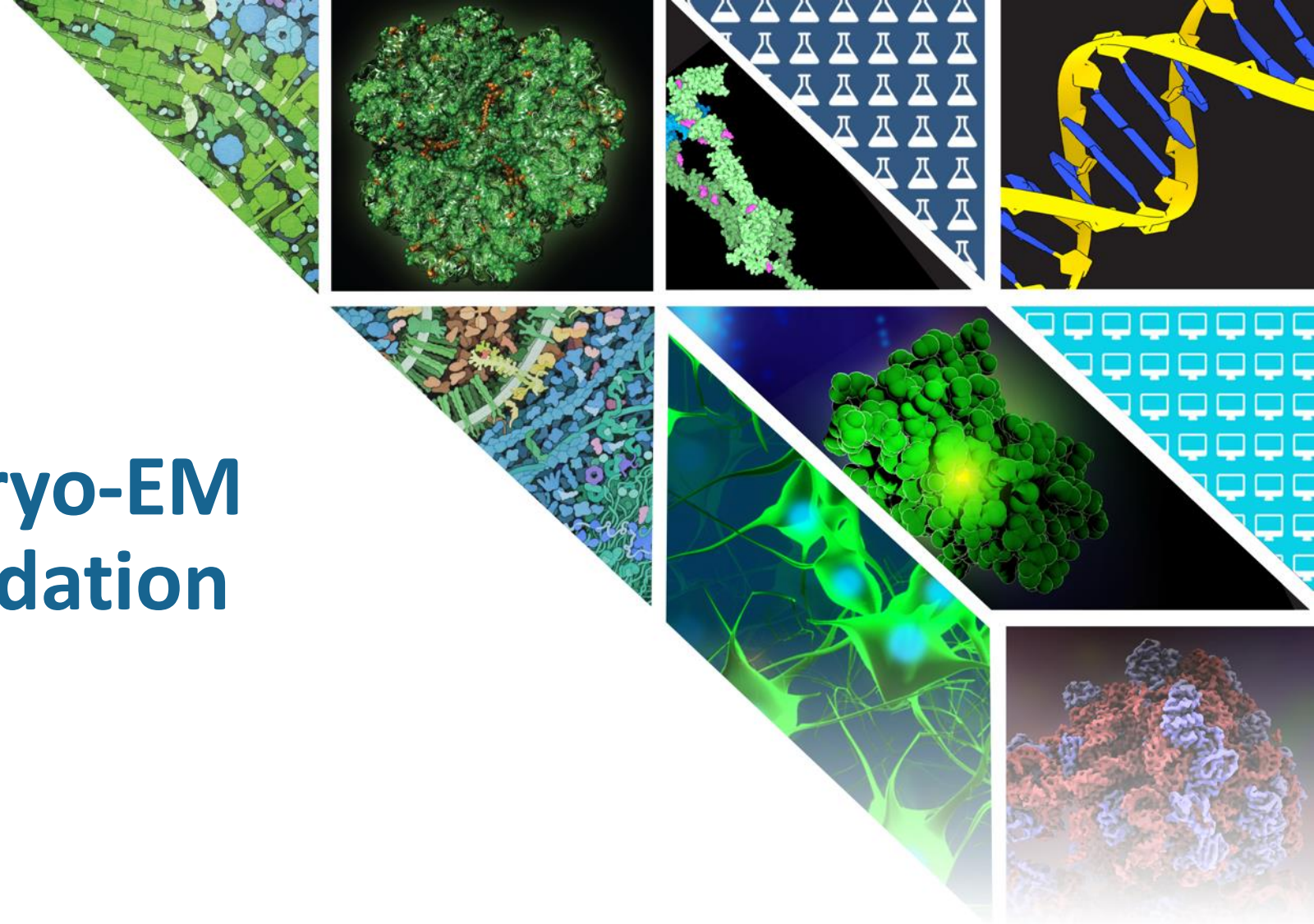
RCSB PDB
PROTEIN DATA BANK

rcsb.org



Best Practices for Cryo-EM Deposition and Validation

Justin W. Flatt, Ph.D.
Biocurator, RCSB Protein Data Bank
Rutgers, The State University of New Jersey

NYSBC Single Particle Short Course March 20, 2026

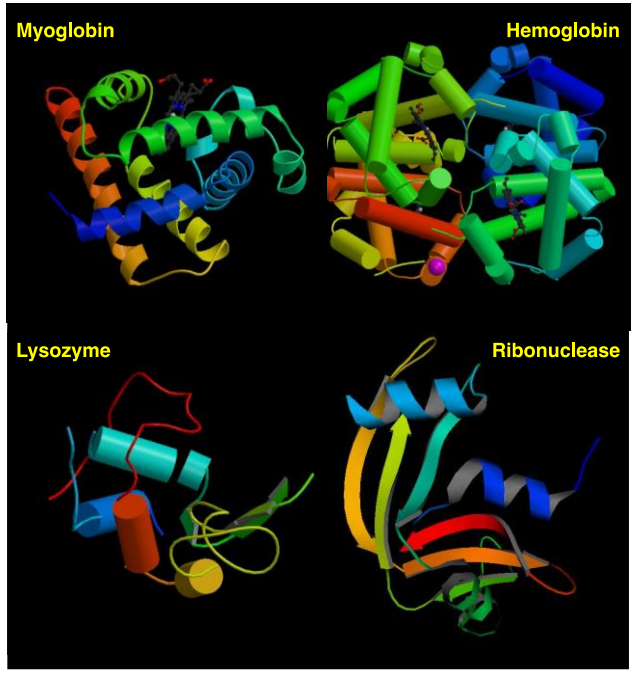
Outline

I. Frozen in Motion: 3DEM Then, Now, and Beyond – wwPDB milestones and evolution
II. Meeting the Data Deluge Challenge – Extended PDB IDs, PDBx/mmCIF, OneDep
III. Quality at Scale: Validation of 3DEM structures – Insights, Best practices, What's Next

I. Frozen in Motion: 3DEM Then, Now, and Beyond – wwPDB milestones and evolution

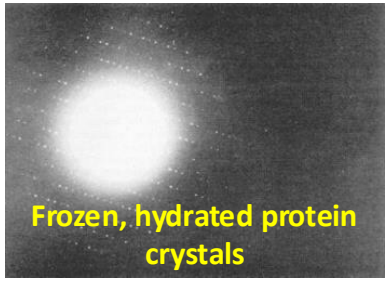
Protein Data Bank (Established 1971)

- A single global archive for structural biology data
- Started with just 7 protein structures
- Now over 250,000 structures freely available!
- Jointly managed by wwPDB partnership started in 2003
- Members: RCSB PDB (US), PDBe (Europe), PDBj (Japan), and PDBc (China); plus EMDB (3DEM) and BMRB (NMR)



X-ray structures led to the launch of archive

Cryo-EM not yet born when the PDB began

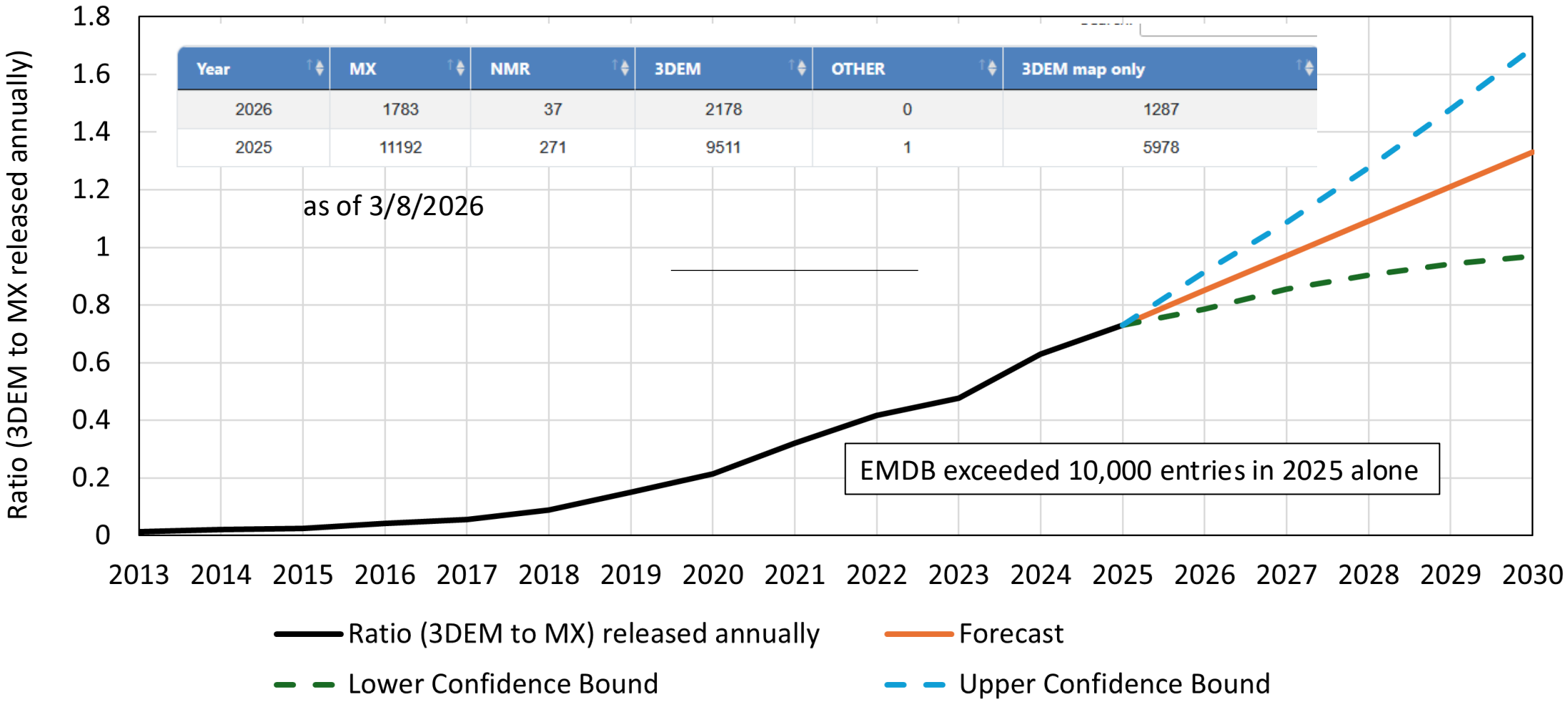


Protein Data Bank (1971) *Nature New Biology 233,* 223
Worldwide Protein Data Bank (2019) *Nucleic Acids Research 47*, D520–D528

Annual PDB Structure Releases for 3DEM ~~Will Soon~~ Eclipses Macromolecular Crystallography

Year	MX	NMR	3DEM	OTHER	3DEM map only
2026	1783	37	2178	0	1287
2025	11192	271	9511	1	5978

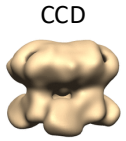
as of 3/8/2026

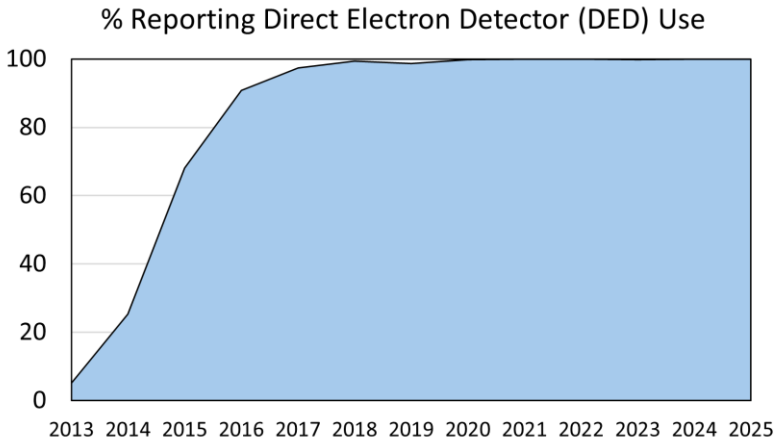


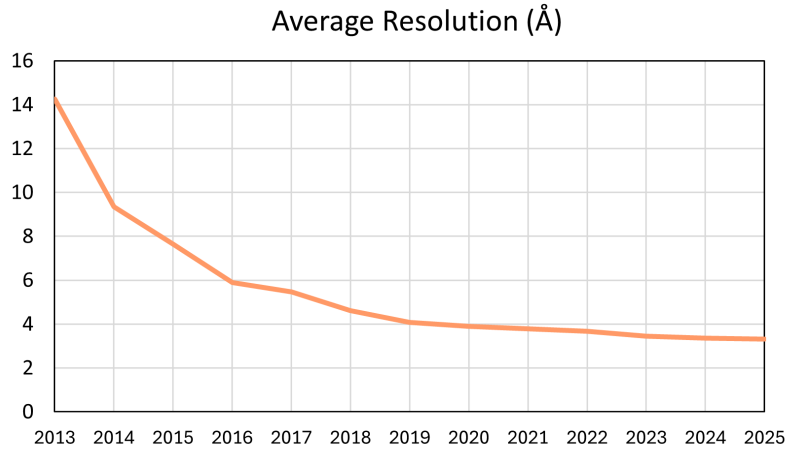
The Cryo-EM Resolution Revolution as a Data Revolution in Structural Biology

Facilitated by

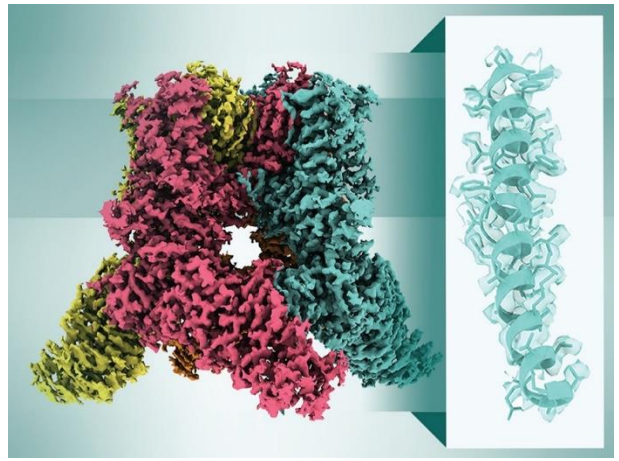
- Direct electron detectors
- Modern electron optics
- Improved data collection software
- Sophisticated image processing with improved algorithms
- FAIR and sustainable data management







300 keV • Direct Detector • RELION (1.9 Å)



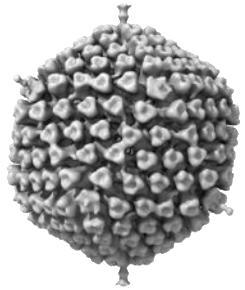
EMD-23853, pdb_00007mij

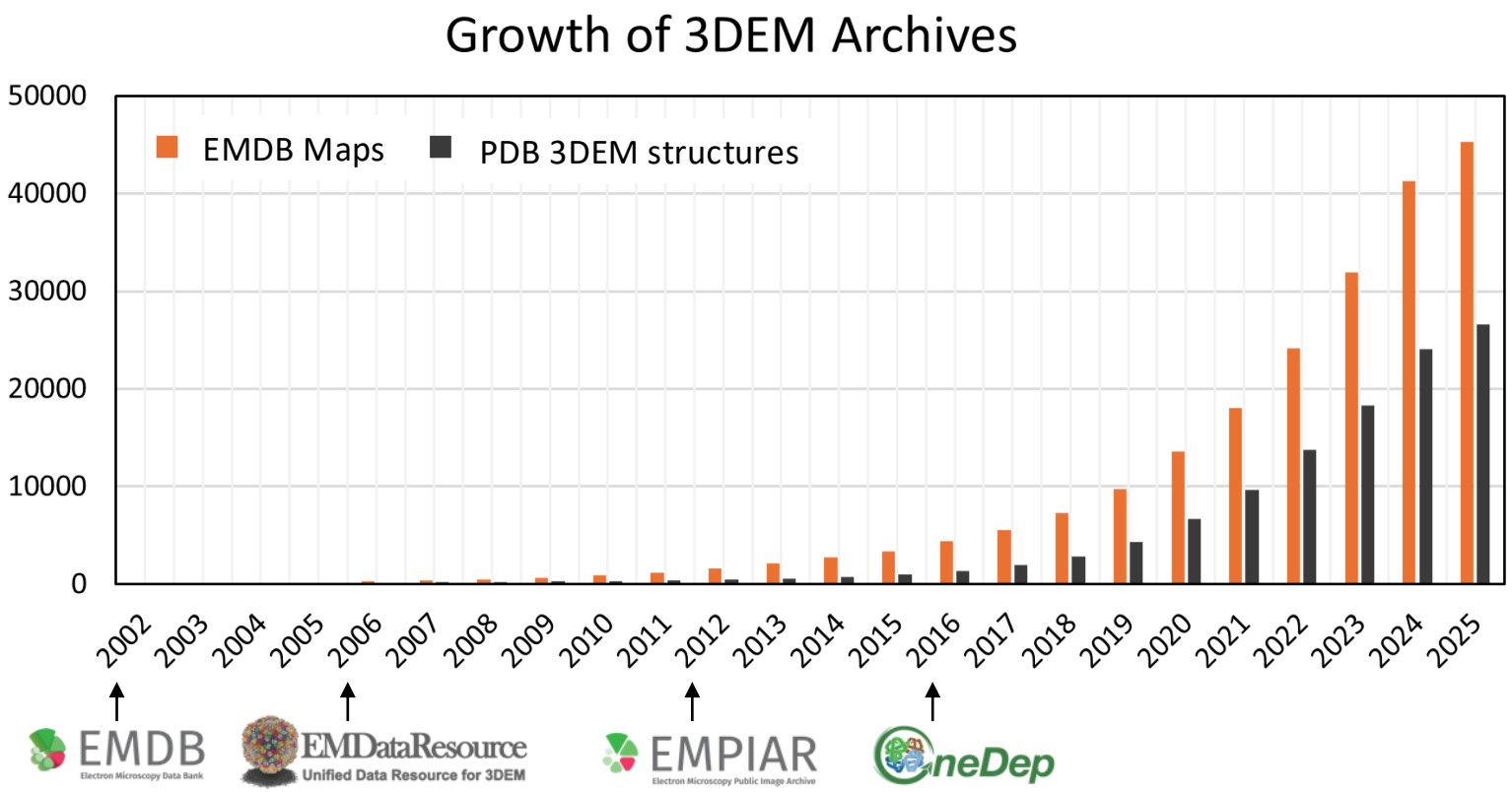
Community-wide deposition and validation standards critical for data quality, consistency, and trust

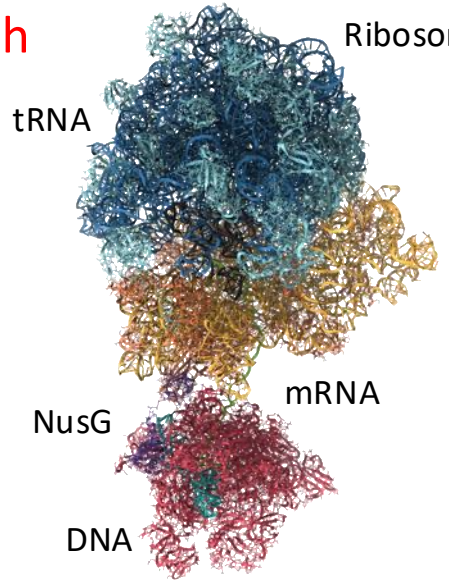
Frozen In Motion: Then, Now, and Beyond

Ongoing Challenge: Flat funding with fixed biocuration resources despite continued growth

Human Adenovirus 2
EMD-1016 (25 Å)
Released 2002

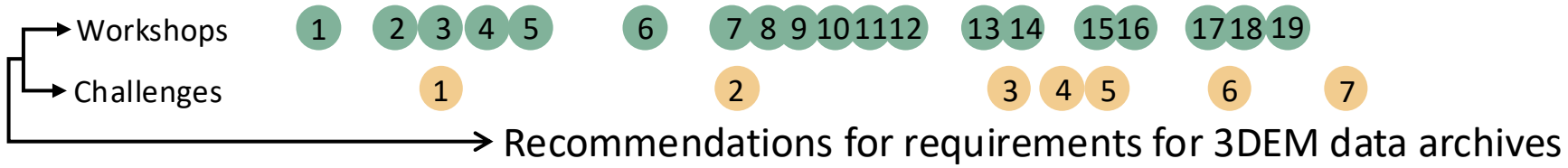






NusG coupled expressome
EMD-11418, pdb_00006ztj (3.4 Å)
Released 2020

Archives and Access



Data Standards

AI-Fueled Future Needs Experts Now More than Ever

Cryo-EM Deposition Workshop
2004
Rutgers, USA

3DEM Validation Task
Force 2010
Rutgers, USA

Nat Methods. 2024 Jun 25. doi: 10.1038/s41592-024-02321-7. Online ahead of print.

Outcomes of the EMDataResource cryo-EM Ligand Modeling Challenge

IUCrJ. 2024 Mar 1;11(Pt 2):140-151. doi: 10.1107/S2052252524001246.

Community recommendations on cryoEM archiving and validation

Cryo-EM Joint Map and Model Challenges
Workshop 2017
SLAC, Stanford, USA

Depositors as Part of the Solution

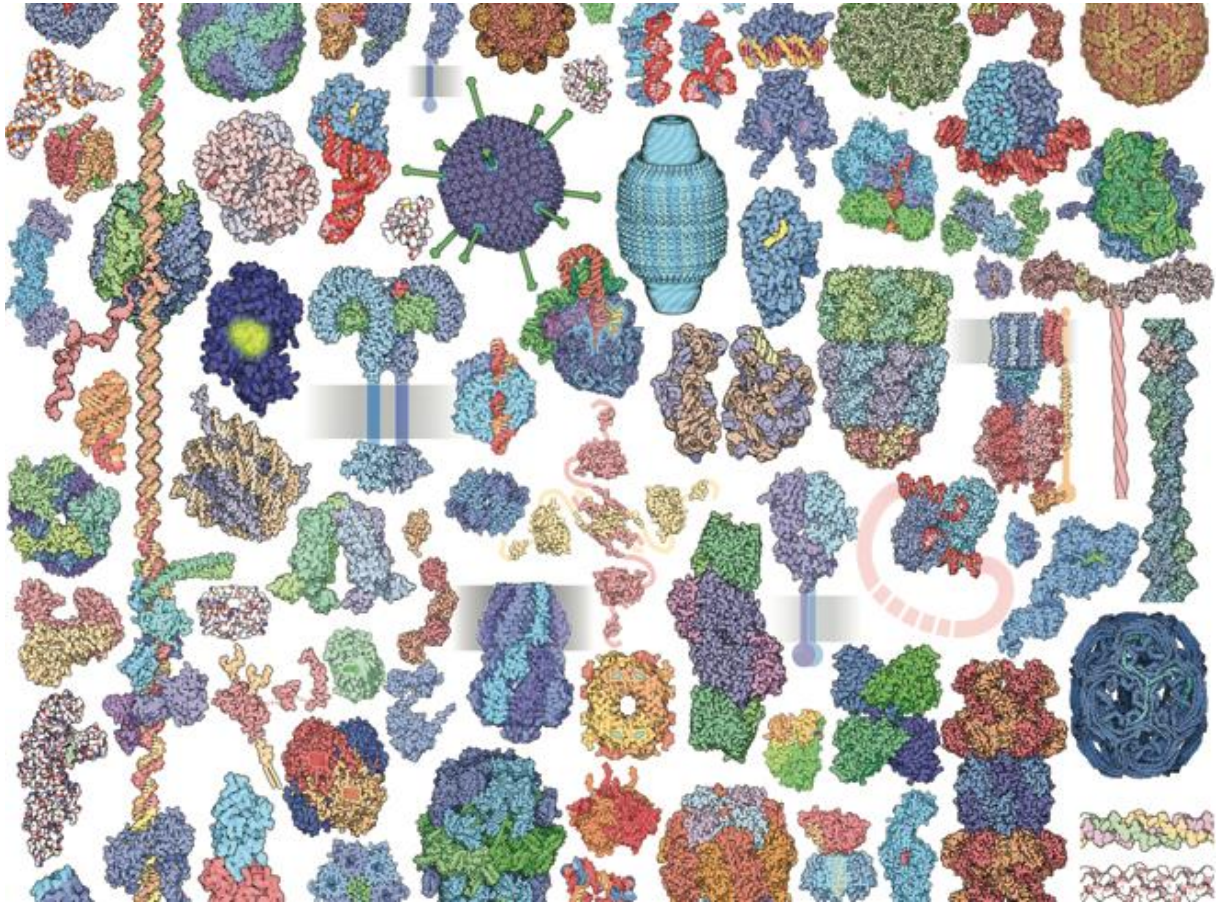
- Follow community standards
- Provide complete, accurate data
- Engage early and stay informed
- Enable high-quality annotations

wwPDB: Data Stewardship by Domain Experts

wwPDB as the Other Part of the Solution

- Steward the global archive
- Centralize deposition & validation (OneDep)
- Curate and annotate expertly
- Guide depositors with tools and feedback

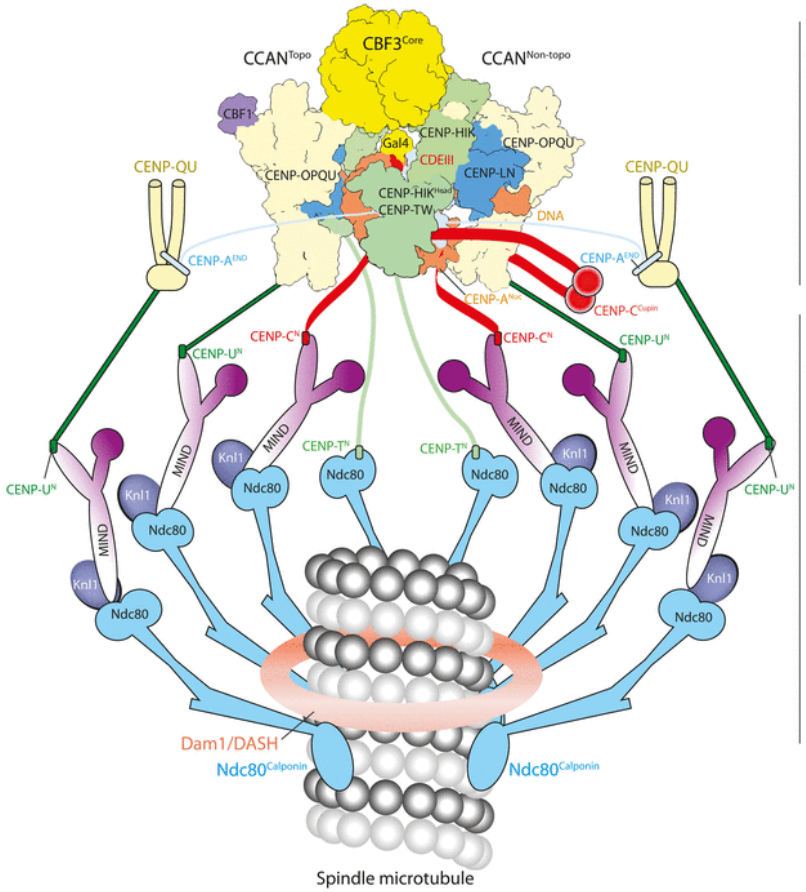
We must work together to keep pace with rapid growth and evolving standards!

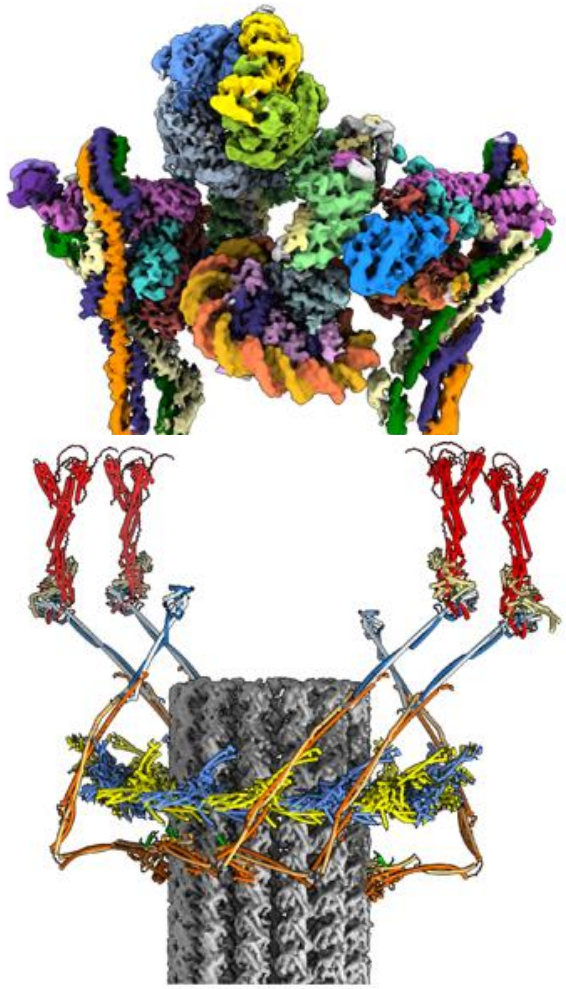


WORLDWIDE PDB
PROTEIN DATA BANK

A New Era of Unprecedented Structural Complexity

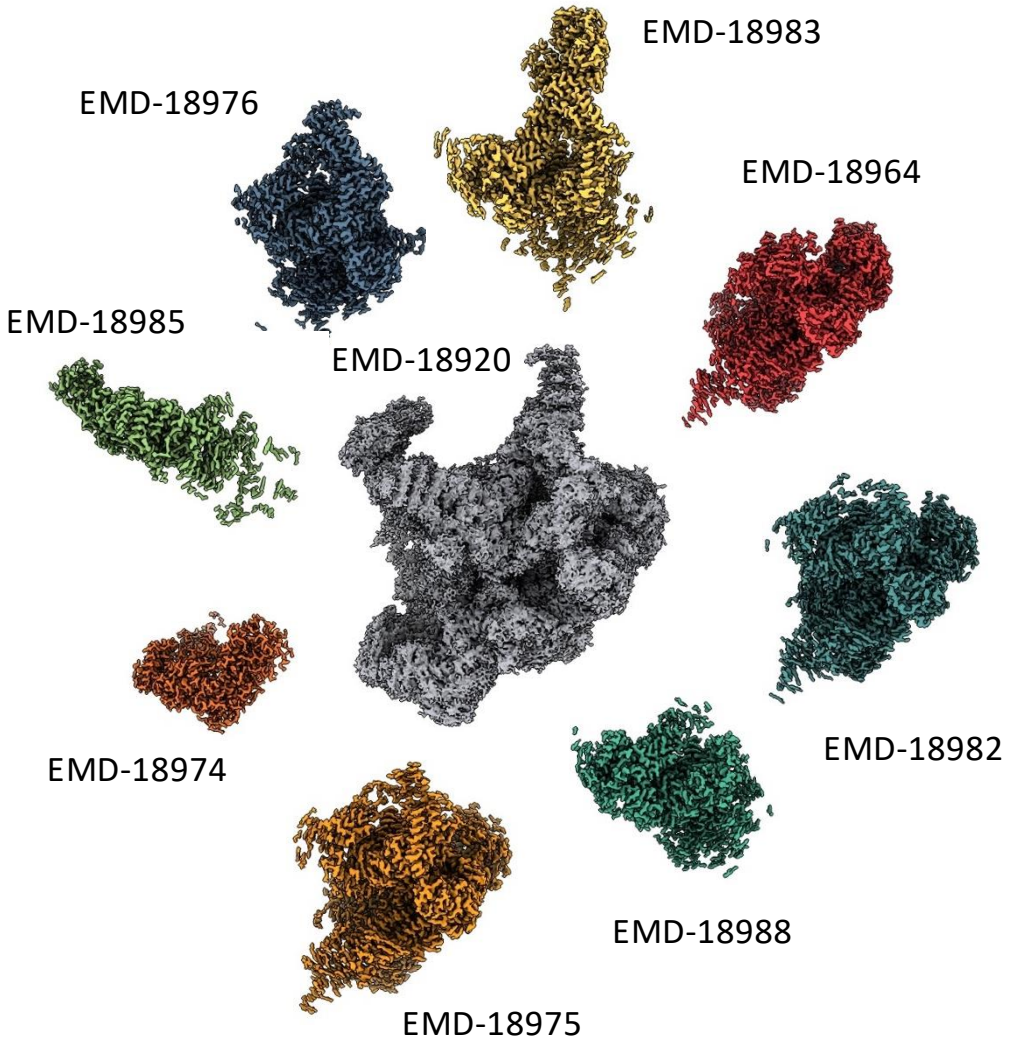
- Previously intractable systems can often be studied at resolutions suitable for atomic model building
- AI/ML predictions can help (but be careful!)
- Experimental structures are essential for understanding biology



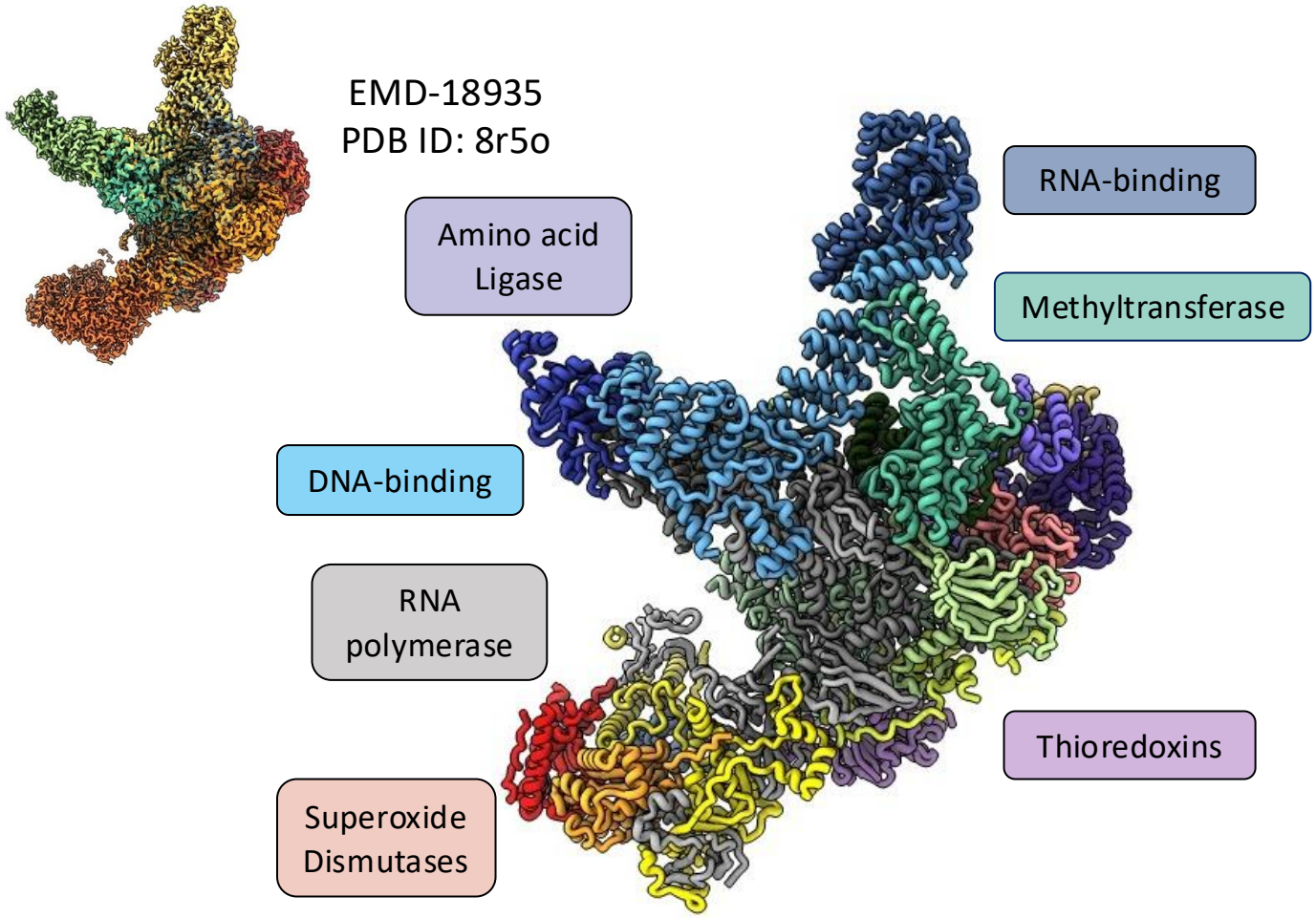


Dendooven T. et al. Science Advances 9(30), eadg7480 (2023). doi:10.1126/sciadv.adg7480
Muir K. W. et al. Science 382(6675), 1184–1190 (2023). doi:10.1126/science.adj8736

A New Era of wwPDB Data Management



Supporting Consensus and Focused Maps

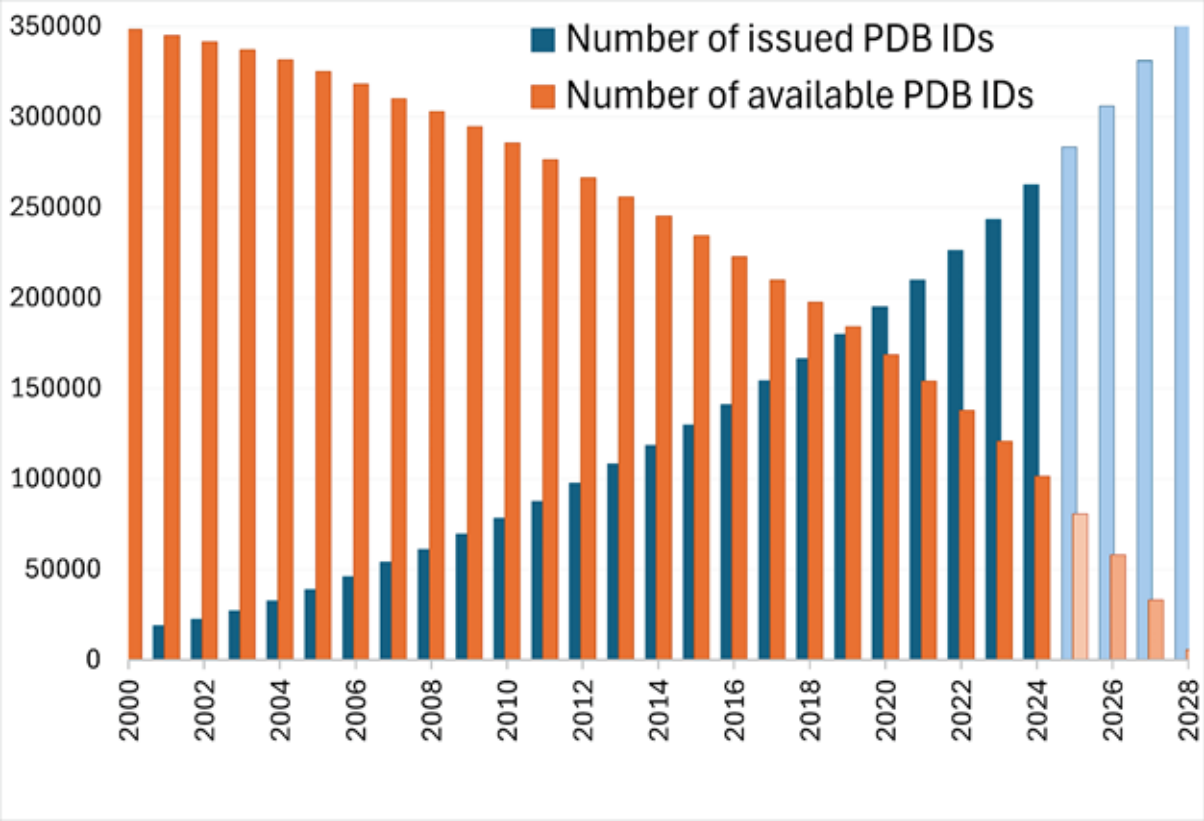


Composite Structure

Vergara-Cruces Á. et al. Cell 187(5), 1145–1159.e21 (2024). doi:10.1016/j.cell.2024.01.036

II. Meeting the Data Deluge Challenge – Extended PDB IDs, PDBx/mmCIF, OneDep

Transitioning to Extended PDB IDs and PDBx/mmCIF Format to Support Growth



Challenges:

- Since 1971, over 240,000 entries have used 4-character PDB IDs
- Software frequently built around legacy PDB file format cannot support extended IDs
- Users, journals, and software developers need to adopt to new PDB IDs and PDBx/mmCIF format now

New PDB ID has 12-character length with “pdb_” prefix: pdb_[a-z0-9]{8}

Example: pdb_00006vxx

Available Transition Resources

Dedicated webpage describing the project

Documentation, FAQs, software tools, etc.
https://www.w wpdb.org/documentation/new-format-for-pdb-ids

Extended IDs in the public files – ready for testing

For the current entries, the extended PDB IDs are generated by adding prefix “pdb_0000” to their 4-character IDs: **4HHB / pdb_00004hhb**

All current archival files have been updated with the extended PDB ids, users can start testing /adopting the extended IDs now:

```
_database_2.database_id                 PDB
_database_2.database_code               4HHB
_database_2.pdbx_database_accession     pdb_00004hhb
```

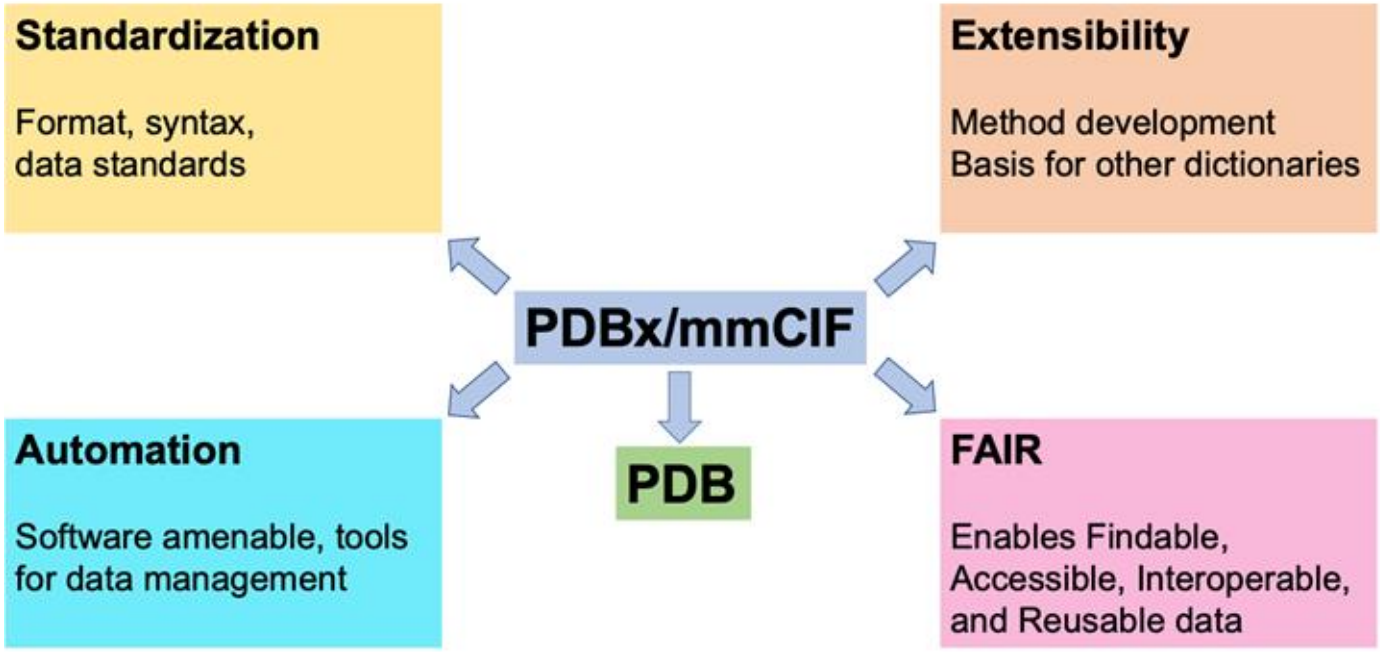
Example files at GitHub – ready for testing

https://github.com/wwPDB/extended-wwPDB-identifier-examples

PDBx/mmCIF Data Format

PDBx/mmCIF: The master archival format of the PDB since 2014

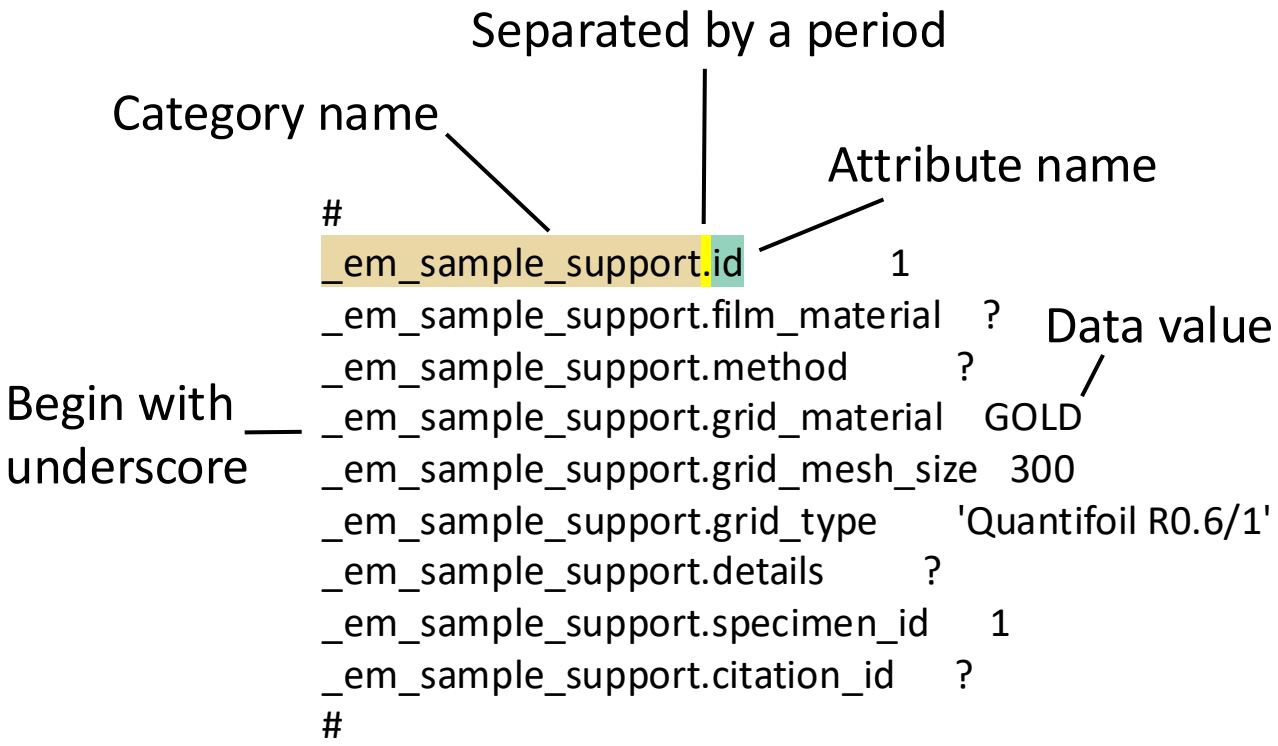
- Community-developed
- Supports structures of any size and complexity
- Defines data types, boundary conditions, enumerations



Entries with extended PDB IDs will be available only in PDBx/mmCIF format

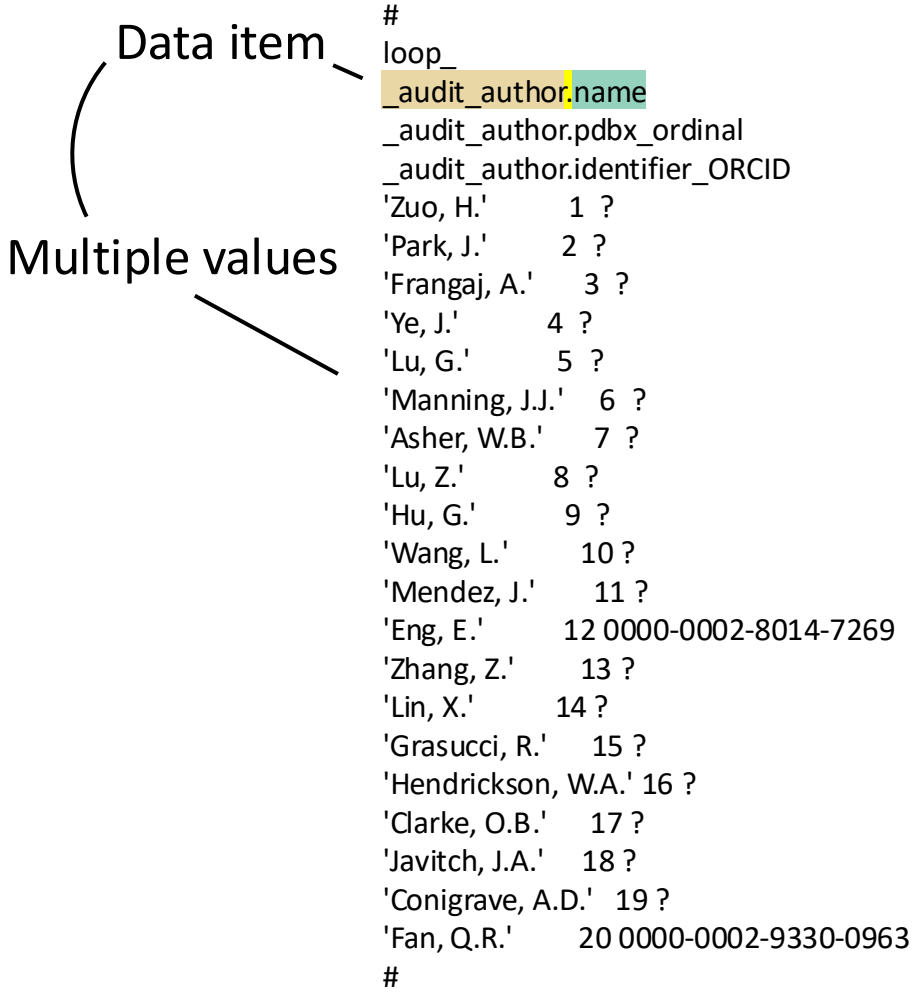
How to read a PDBx/mmCIF File

Key-value style:



Category + attribute = Data item

Tabular style (multiple values for single data item):



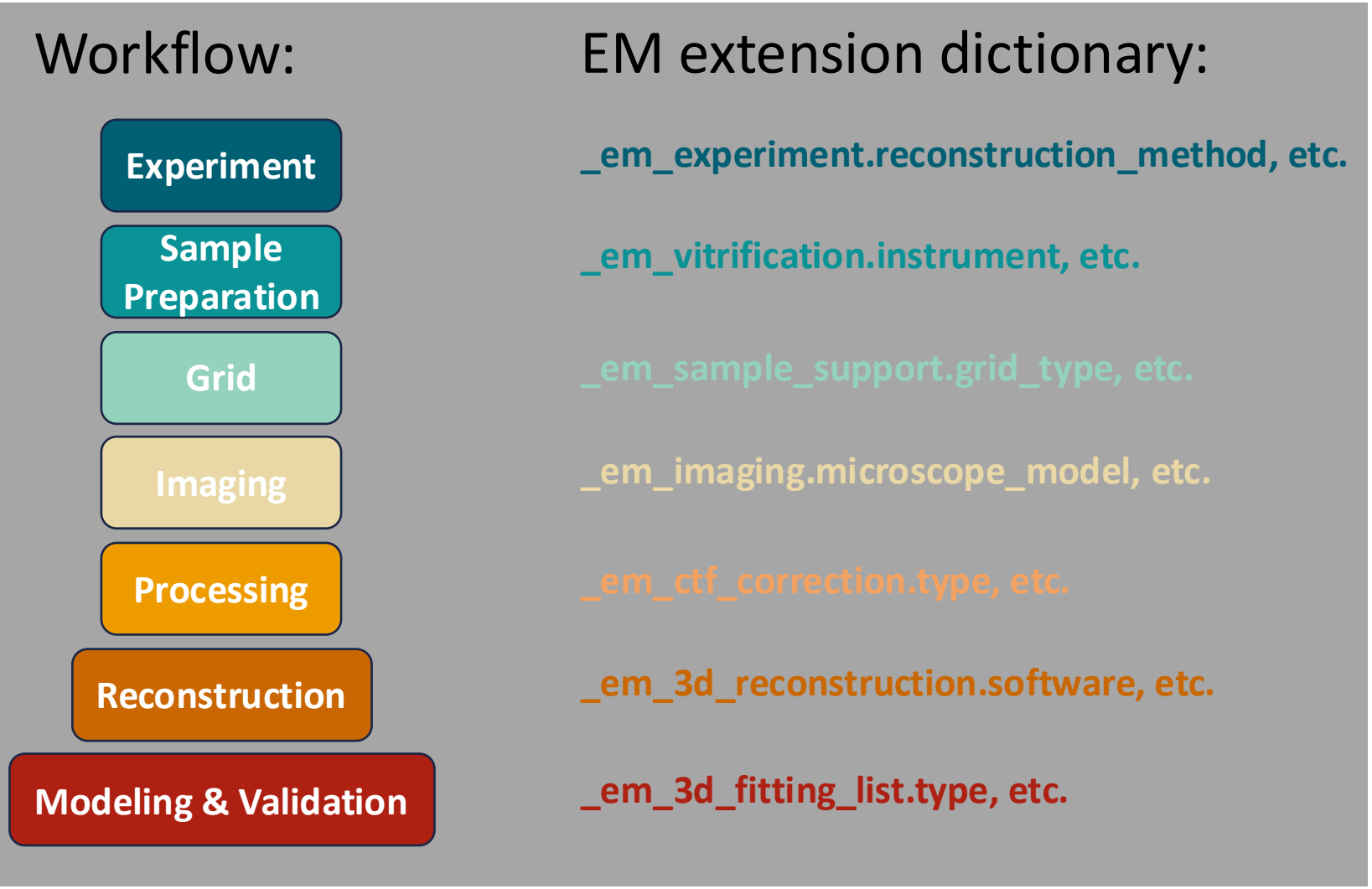
PDBx/mmCIF Data Resources

Documentation, FAQs, user guide available at https://mmcif.wwpdb.org/docs/user-guide/guide.html

Major structural biology software tools support PDBx/mmCIF

- Software libraries, parsers, format converters, file editors
- Visualization tools: Chimera, Mol*, PyMol, Jmol, Coot, CCP4mg
- Refinement programs: Phenix, Refmac, Buster

PDBx/mmCIF Powers Deeper Insight and Wider Impact for 3DEM

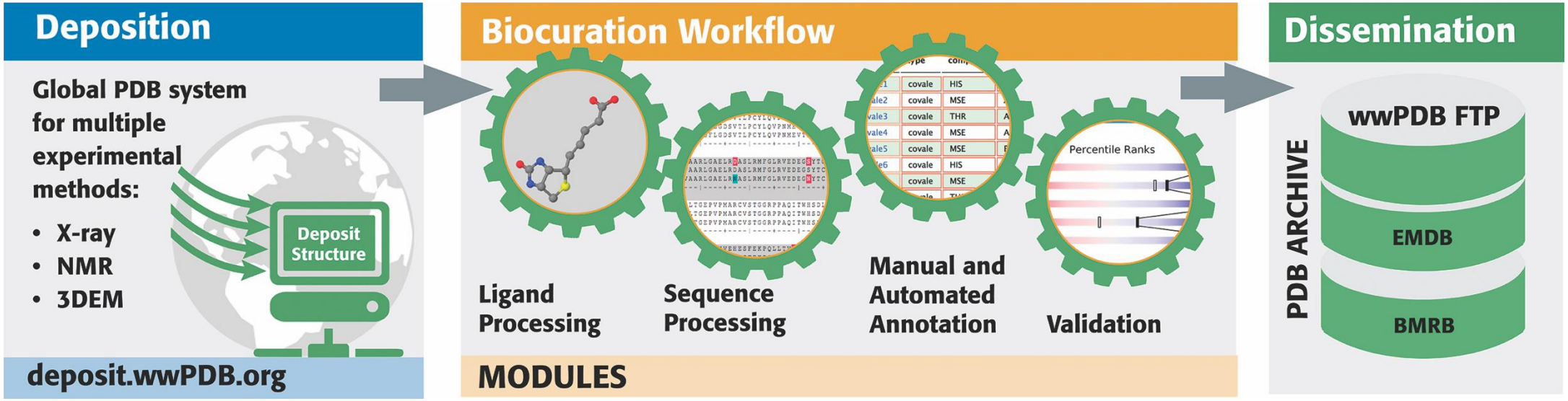


Your data, fully realized:

- No limits on atom count or polymer chains
- Continuously updated data dictionary
- wwPDB validation reports now available in PDBx/mmCIF format

Example:
_em_vitrification.instrument
'CRYOSOL VITROJET'

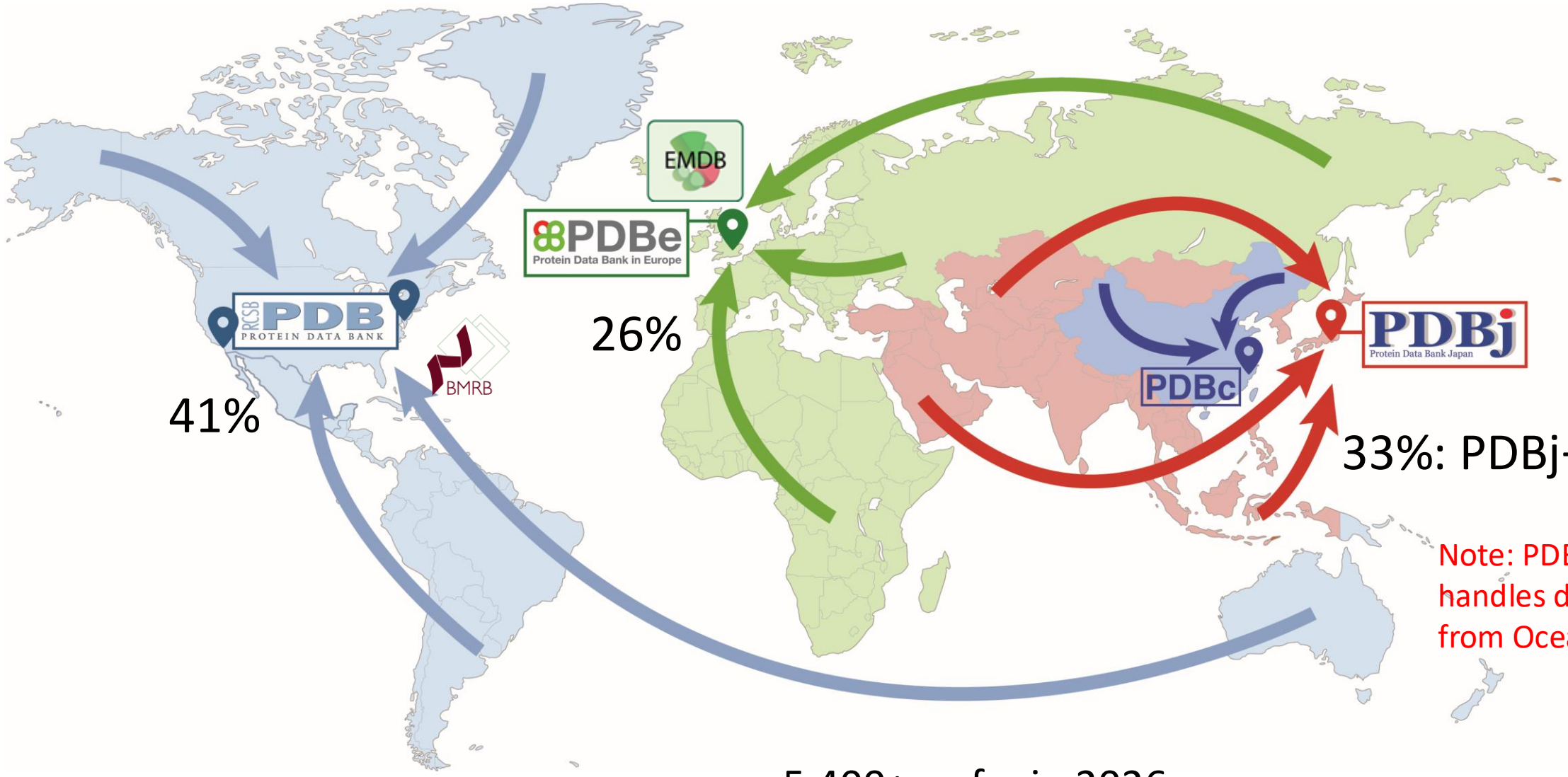
OneDep: wwPDB's Unified Deposition System



- Supports all major experimental techniques
- Enhanced with extended metadata, more enumerations, and stronger data checks
- Prevents incomplete depositions

- Distributes responsibilities among wwPDB data centers
- Provides equal service to depositors worldwide
- Simplifies data exchange and archival updates
- Allows reliable remediation with version control

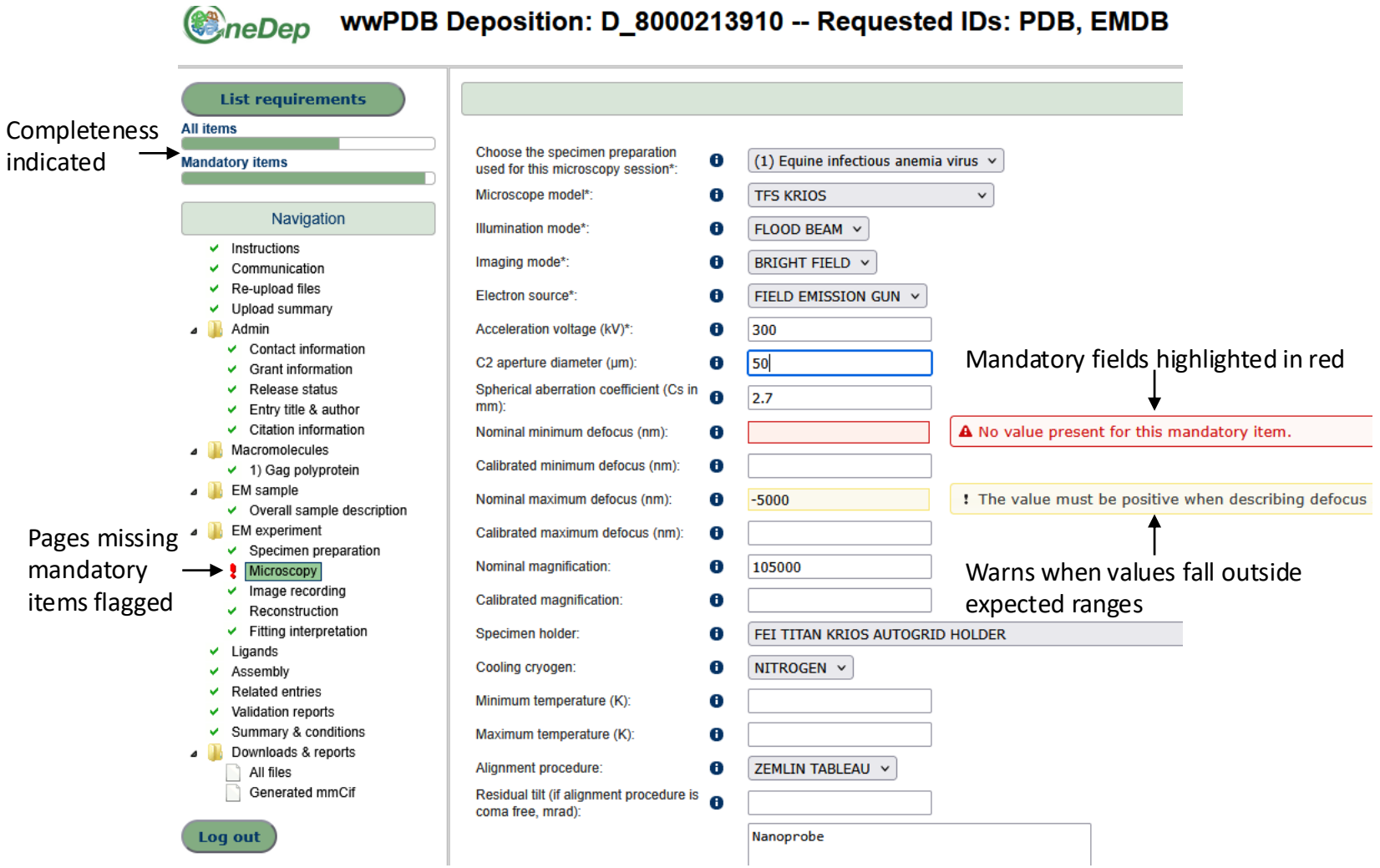
OneDep: A Global Tool for a Global Community



5,400+ so far in 2026

Depositions Become “Easier” Inside OneDep

- Direct data harvesting from uploaded PDBx/mmCIF files
- Dep UI collects metadata missing from uploaded files
- Issues PDB IDs for 3DEM models and EMDB IDs for 3DEM maps



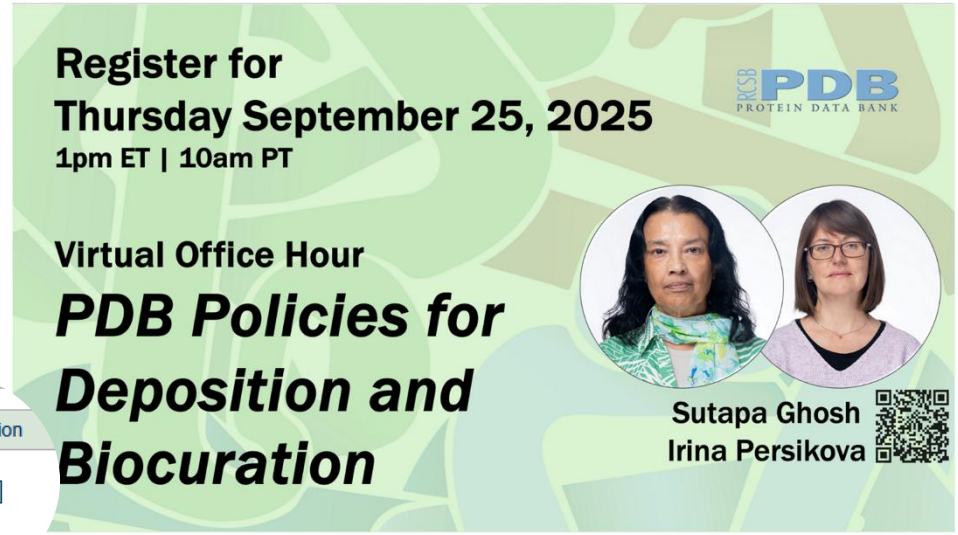
Structure 25: 536-545 doi: 10.1016/j.str.2017.01.004

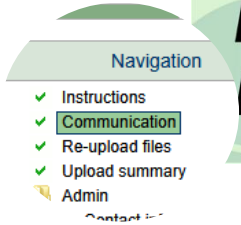
Tips for a Smoother Deposition Process

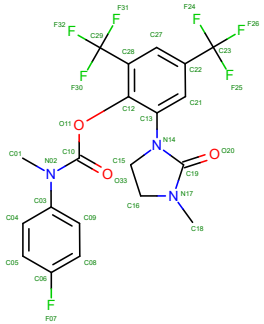
- Register for training events
- Prepare files and validate before deposition
- Ask questions anytime for guidance and support

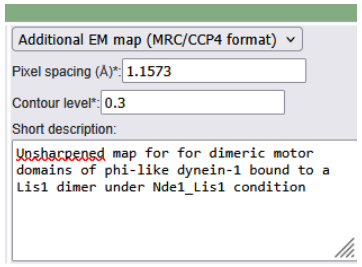
  General: deposit-help@mail.wwpdb.org

  Specific: Log in to your session and select the Communication Page
- Provide accurate info for sequences and ligands
- Carefully review the validation report and address errors
- Share as you want others to share with you





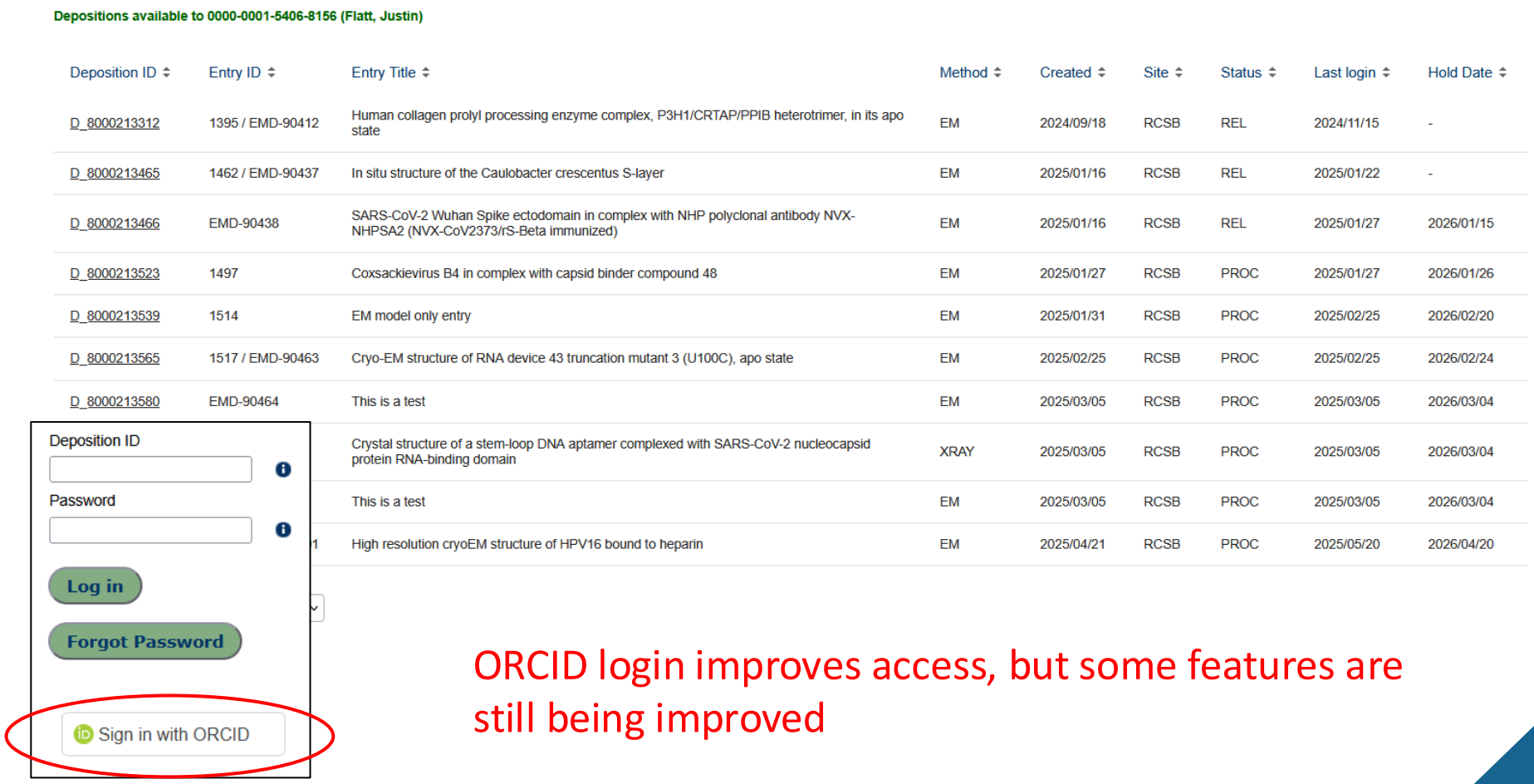




File	Description	Details
emd_47378_additional_1.map.gz	Unsharpened map for for dimeric motor domains of phi-like dynein-1 bound to a Lis1 dimer under Nde1_Lis1 condition	⊞

Simplify OneDep Login and Entry Tracking with ORCID

- Authenticate as a contact author without sharing passwords
- ORCID login provides a summary table of your entries
- Traditional deposition ID/password still available

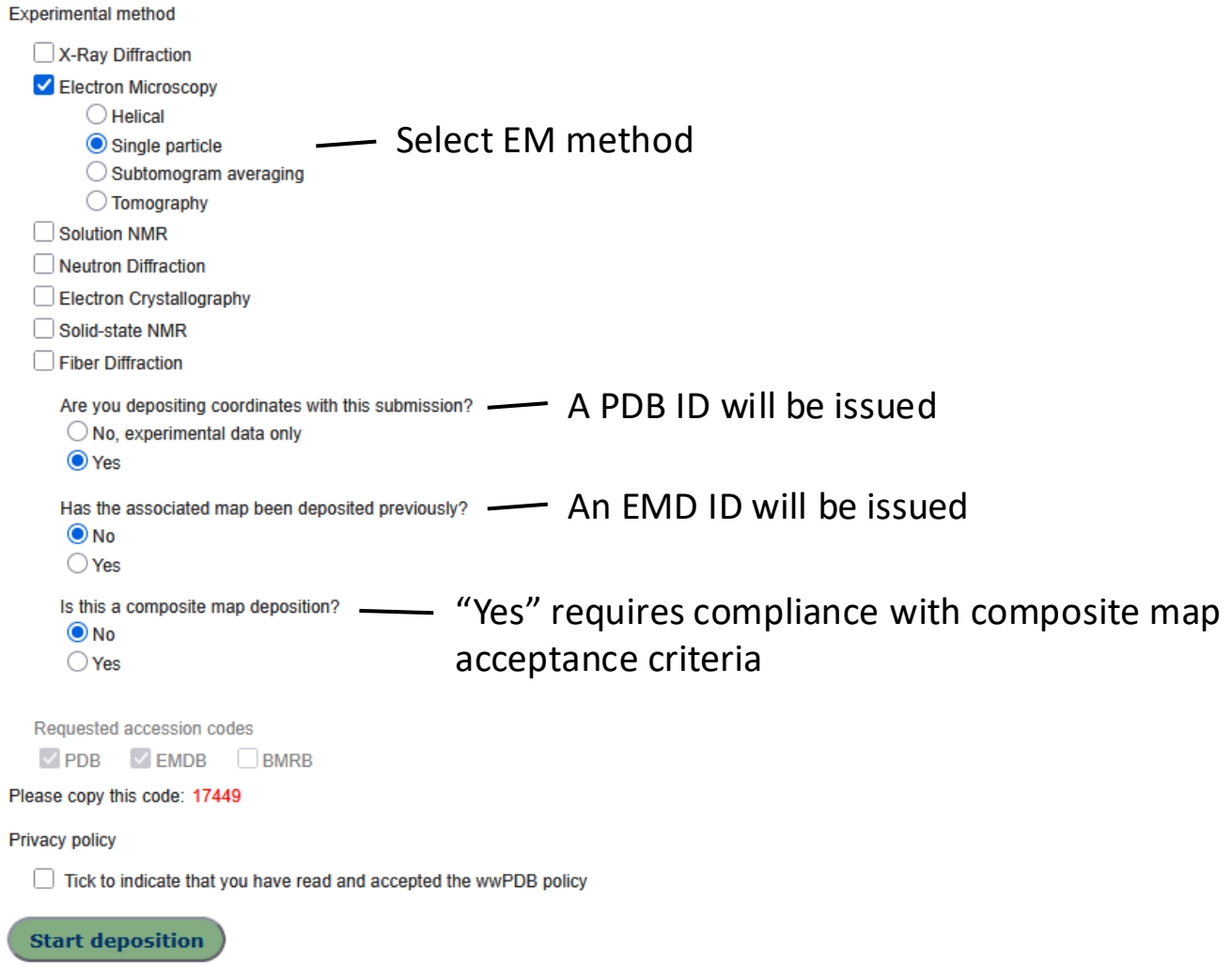


ORCID login improves access, but some features are still being improved

Define Your Deposition Type Carefully at the Start As Requirements and IDs Cannot Be Changed

- Define the deposition type at the start
- OneDep uses these selections to determine entry requirements and validation
- Once you click “Start deposition”, ID settings cannot be changed

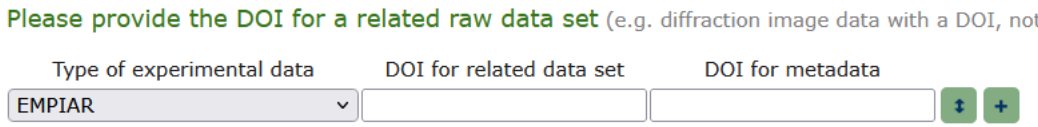
Errors in defining an entry require withdrawing it and starting a new deposition from scratch



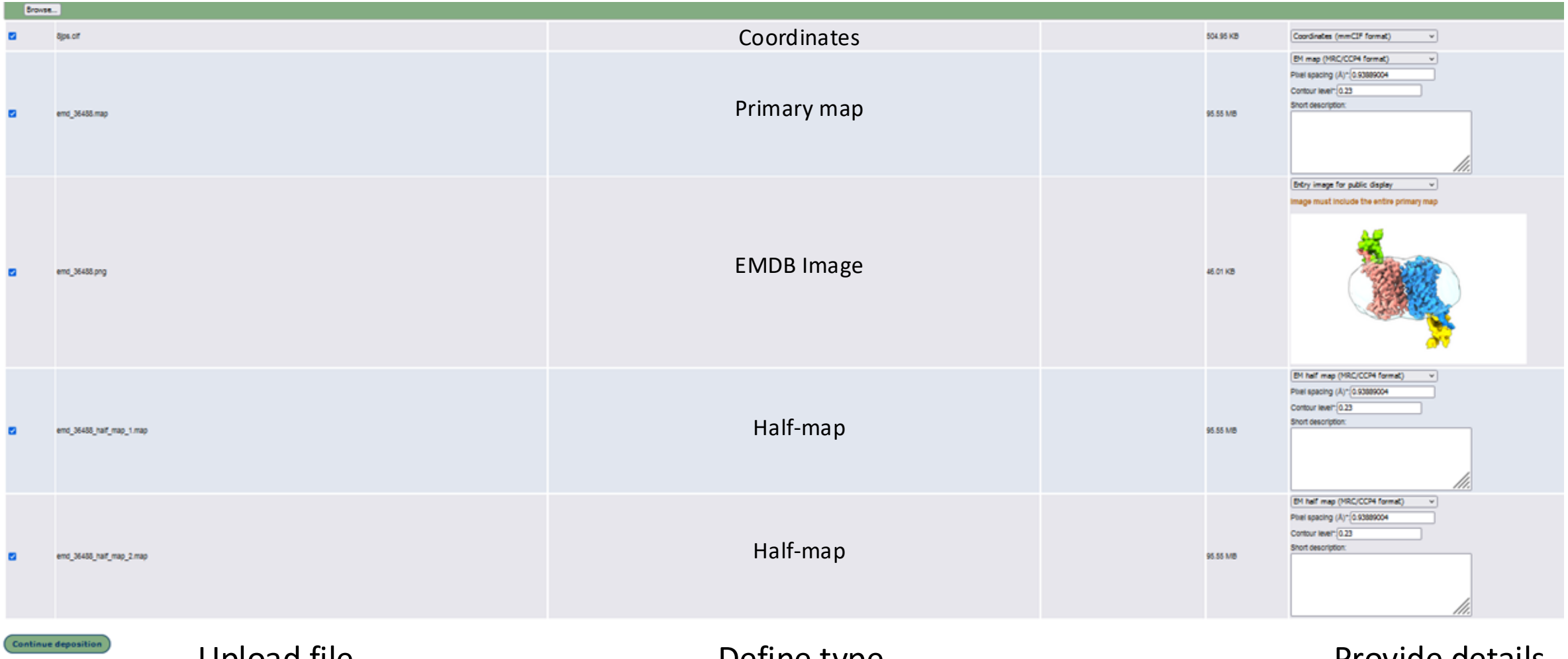
Once Entry Type is Defined, OneDep Sets File Requirements

File	When Required	In Validation Report?
Coordinates (PDB/mmCIF)	Map+Model, Model-only	Yes
Primary map (MRC/CCP4)	Map+Model, Map-only	Yes
Half-maps (MRC/CCP4)	SPA/Helical/STA	Yes
EMDB public image	Required for all 3DEM entries	No
FSC curve (XML)	Optional (recommended)	Yes
Mask (MRC/CCP4)	Optional (recommended)	Yes
Additional maps (MRC/CCP4)	Optional (upload any relevant supporting maps)	No
Metadata (mmCIF)	Optional (Map-only; Recommended for Batch Submissions)	Some

Depositing raw cryo-EM image data to EMPIAR is best practice as it preserves and shares raw data for you study



Strive To Go Beyond The Minimum When Sharing your Data



Upload file

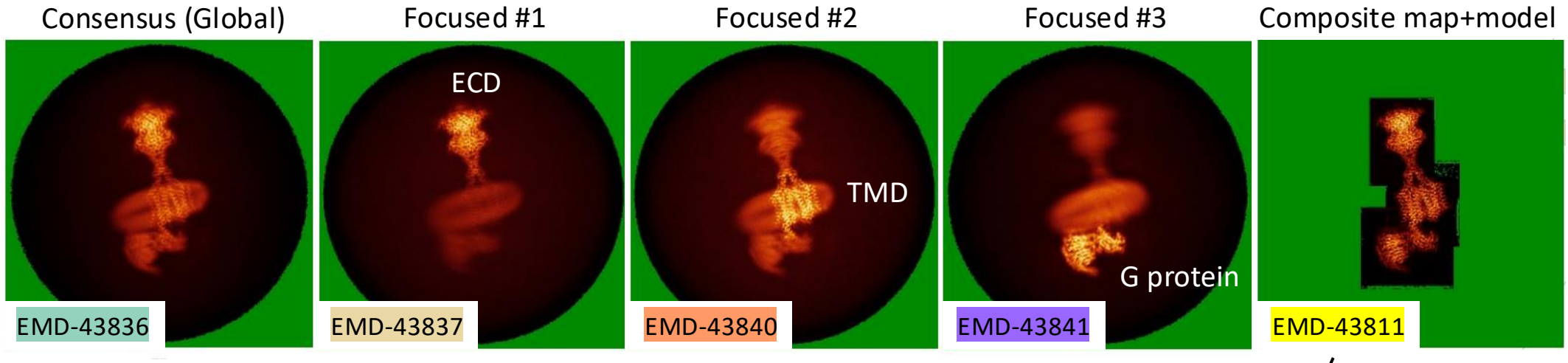
Define type

Provide details

Include all helpful files and provide sufficient descriptions to maximize validation, reuse, and future discovery

Composite Map Entries Must Meet Additional Acceptance Criteria

All supporting maps must be deposited as separate entries

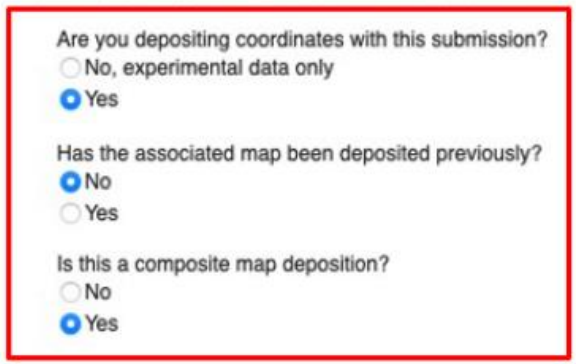


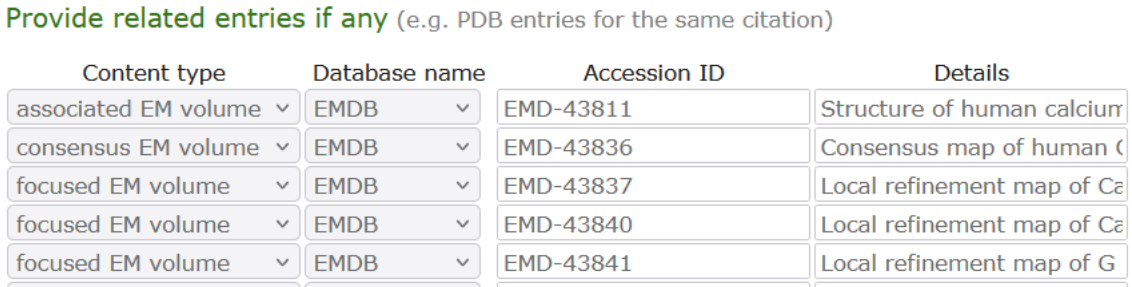
Each supporting map (e.g., consensus or focused EM volume) must be clearly defined and linked to the composite model entry

```
loop_
_pdbx_database_related.db_name
_pdbx_database_related.details
_pdbx_database_related.db_id
_pdbx_database_related.content_type
EMDB 'Structure of human calcium-sensing receptor in complex with chimeric Gq (miniGisq) protein'  EMD-43811
'associated EM volume'
EMDB 'Consensus map of human CaSR-miniGisq complex in nanodiscs'  EMD-43836
'consensus EM volume'
EMDB 'Local refinement map of CaSR extracellular domain in nanodisc-reconstituted human CaSR-miniGisq complex' EMD-43837
'focused EM volume'
EMDB 'Local refinement map of CaSR transmembrane domain in nanodisc-reconstituted human CaSR-miniGisq complex' EMD-43840
'focused EM volume'
EMDB 'Local refinement map of G protein in nanodisc-reconstituted human CaSR-miniGisq complex'  EMD-43841
'focused EM volume'
```

Composite Map Deposition Workflow

1. Review the EMDB Composite Map Deposition Guide: https://www.ebi.ac.uk/emdb/documentation/deposition/composite_map
2. Create supporting depositions for:
   - The consensus volume
   - Each focused map
3. Create a deposition for the composite map
4. Link all related entries in the depUI or request annotator assistance during processing





Associating Multiple Atomic Models With One EM Map

- Avoid depositing the same map multiple times
- Start by submitting one model as a Map+Model deposition
- Submit additional models as EM-model-only
- Link all models to the original Map+Model entry

Note: EM-model-only validation - coordinates assessed, map fit not yet included

Recommended:


Resource

A Multi-model Approach to Assessing Local and Global Cryo-EM Map Quality

Mark A. Herzik Jr. [1], James S. Fraser [2], Gabriel C. Lander [1 3]

Show more

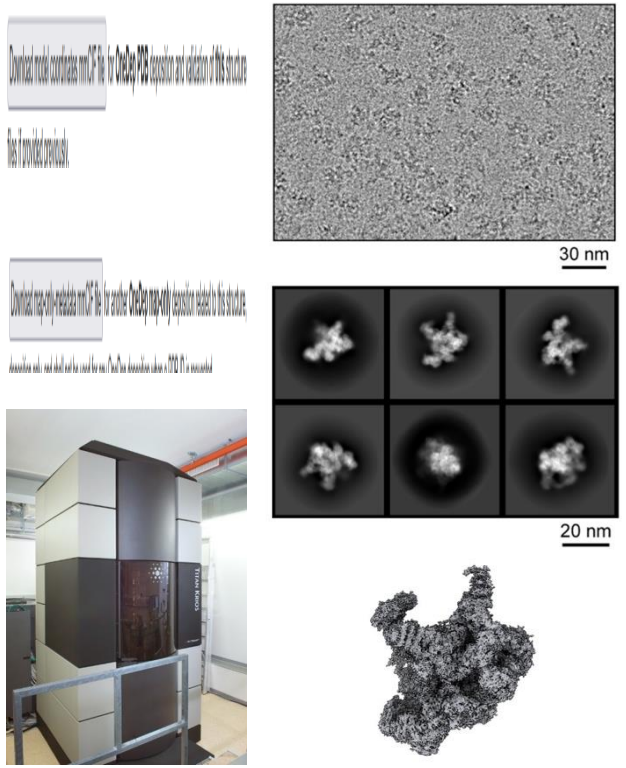
+ Add to Mendeley Share Cite

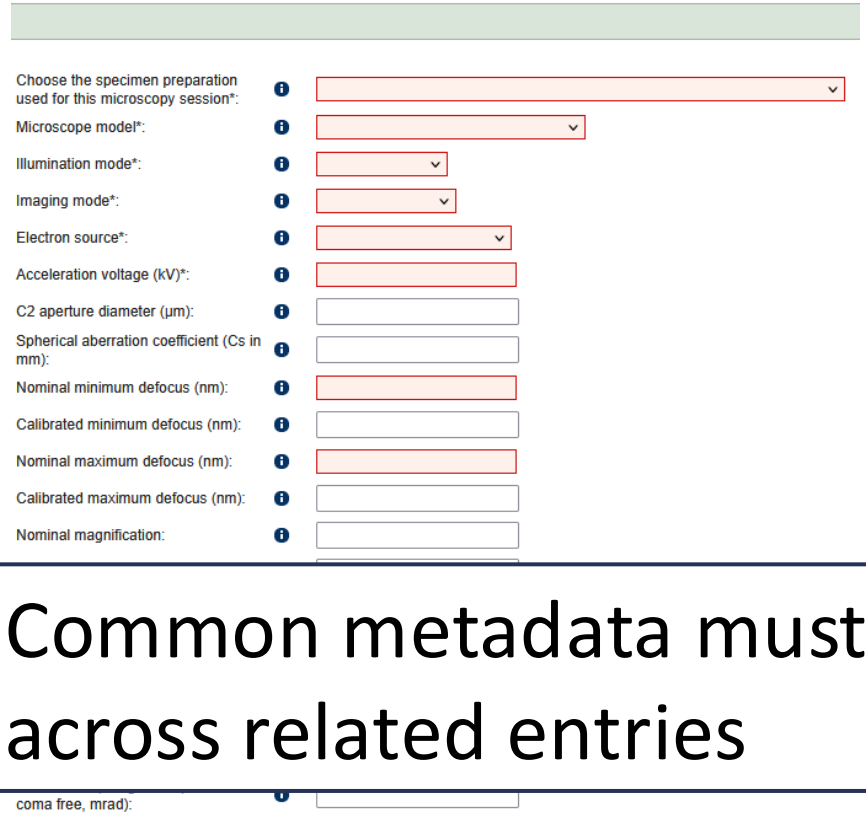
https://doi.org/10.1016/j.str.2018.10.003 ↗ Get rights and content ↗

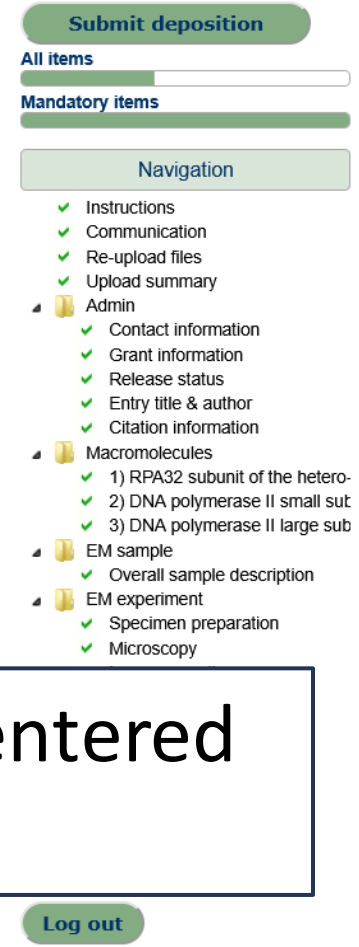
Watch for evolving best practices;
EMDB is available for guidance

Typing the Same Metadata Across Large Related Submissions Is a Major Burden

Experimental Workflow







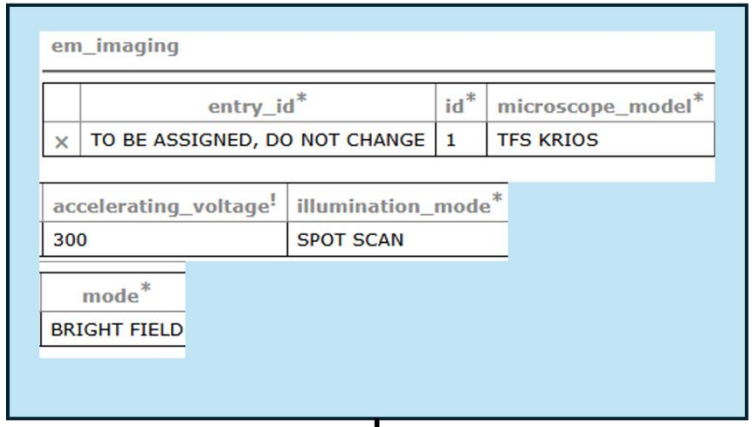
Common metadata must be entered across related entries

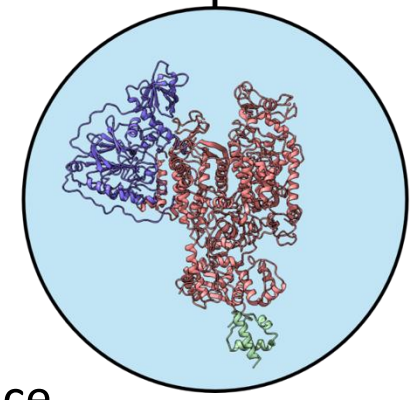
Upload Structure

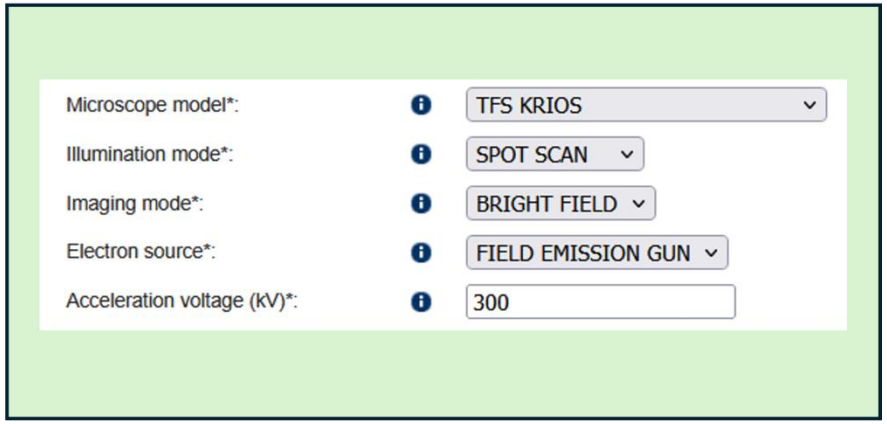
Input Metadata

Submit

Prepare Reusable Metadata for Multiple Structure Submissions Using pdb_extract







Composite (Map+Model)

Consensus (Map-only)

Focused (Map-only)

Reduce effort, minimize errors

Enter metadata only once
Generate deposition-ready mmCIF files

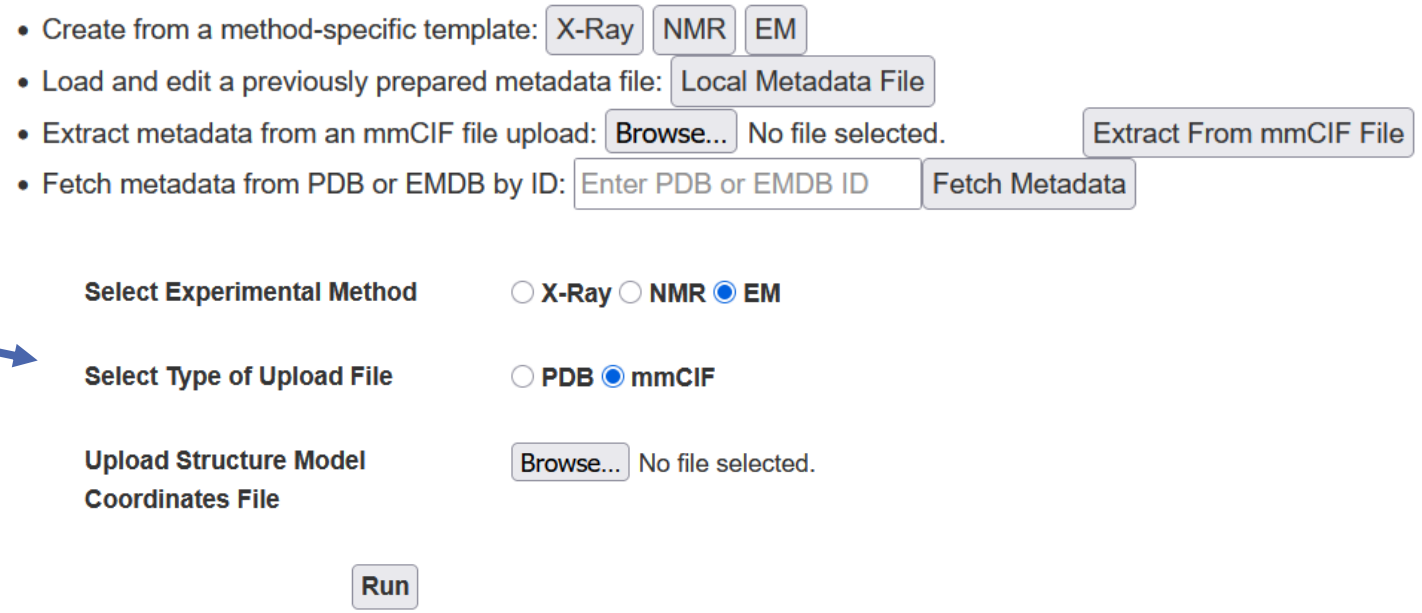
Upload to OneDep
Data-entry pages auto-populated

pdb_extract Makes Metadata Preparation Simple

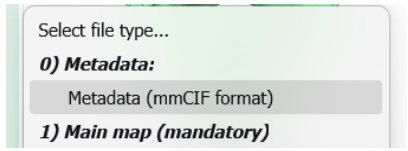
New and improved version

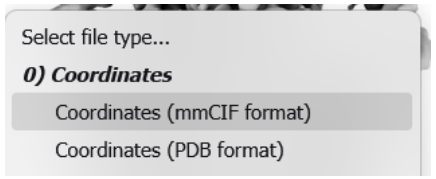
For now use: https://pdb-extract-dev.wwpdb.org/ instead of https://pdb-extract.wwpdb.org/

1. Prepare reuseable EM metadata template
2. Make files deposition-ready
3. Upload to OneDep



map-only or coordinates





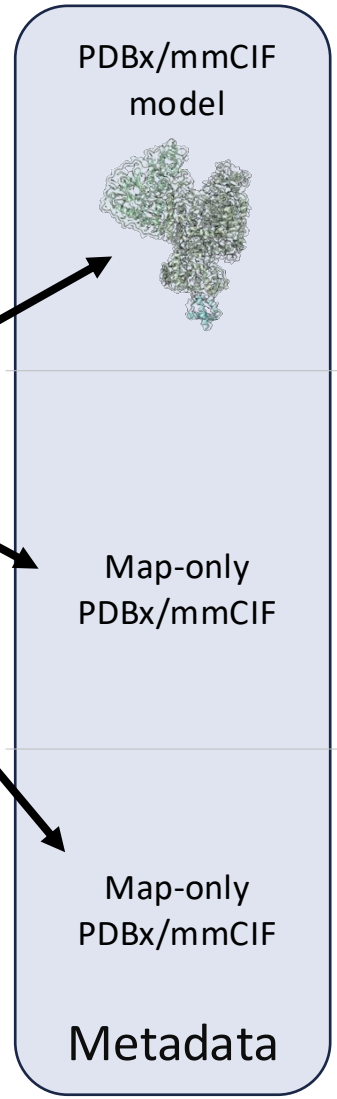
Process large batches (e.g., 50+) quickly using standalone command-line version

pdb_extract Reduces the Burden of Batch Submissions to OneDep

Simple composite case:

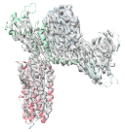
Download model coordinates mmCIF file for **OneDep PDB** deposition and validation of **this** structure

Download map-only-metadata mmCIF file for another **OneDep map-only** deposition related to this structure



Composite Map+Model

Files:
Primary map
Coordinates



Focused map-only

Files:
Primary map
Half-maps
Mask
FSC

Consensus map-only

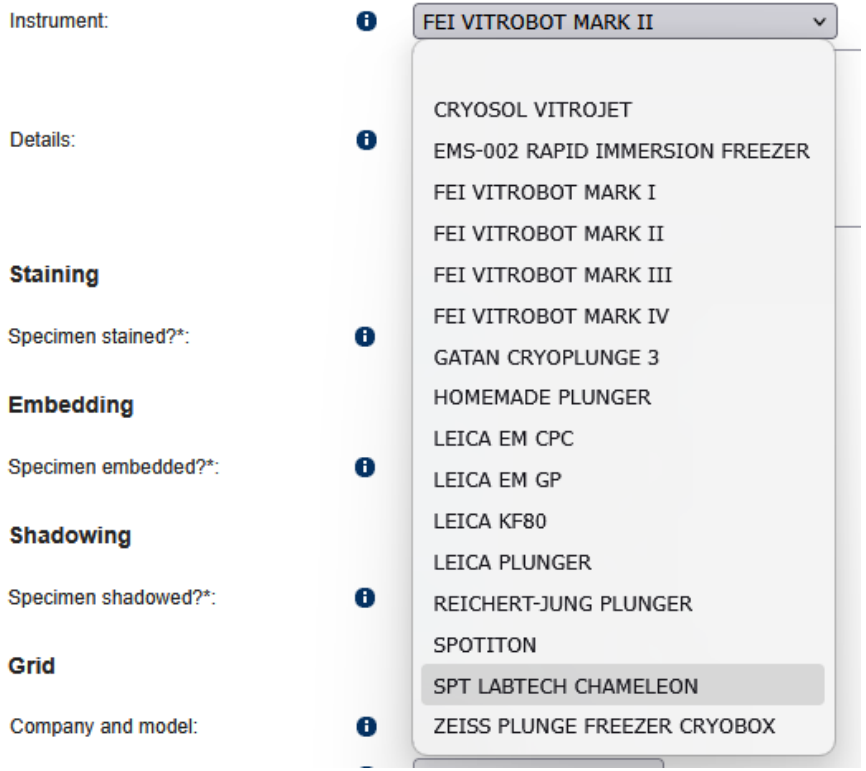
Files:
Primary map
Half-maps
Mask
FSC

For help please email
deposit@deposit.rcsb.org,
chenghua.shao@rcsb.org, or
justin.flatt@rcsb.org

Working Together For Accurate Records

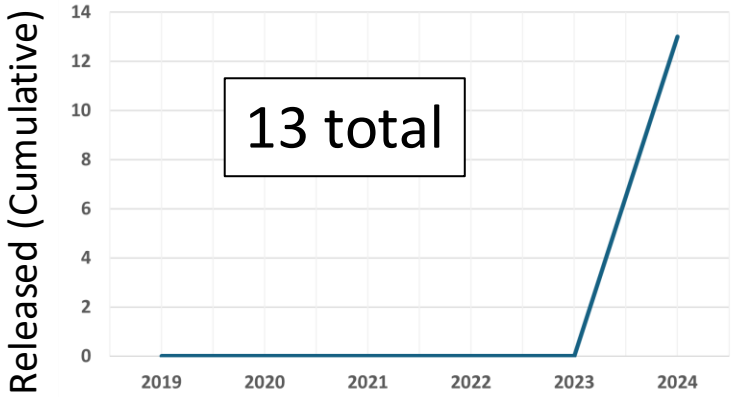
Please report any needed changes upon noticing!

SPT Labtech introduced the Chameleon in 2019



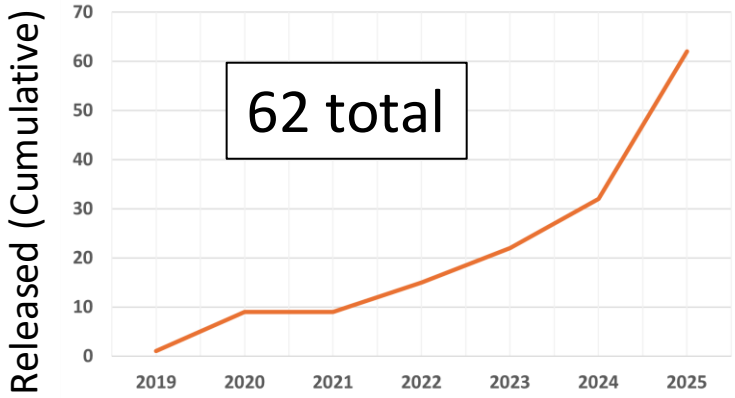
Request was not received to add the device to the pull-down until 2024

Entries with Chameleon Vitrification



vs

Spotiton with Chameleon in Title or Details

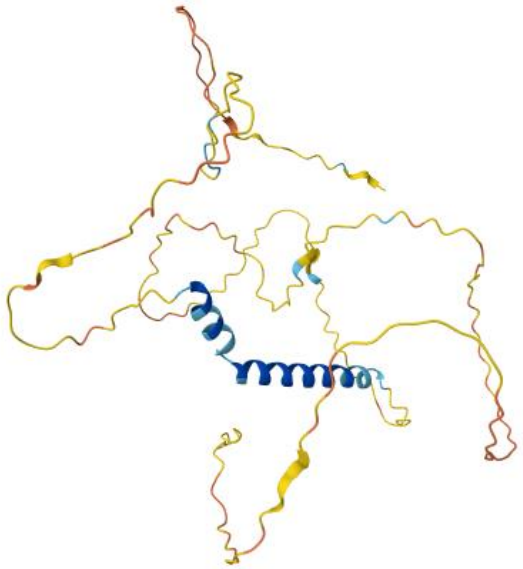


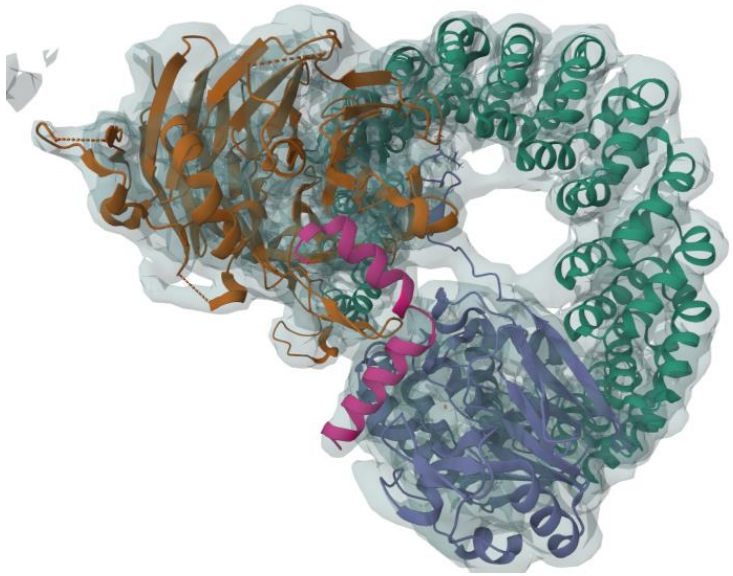
As a result, existing entries need to be updated to specify chameleon

Complete Workflow Descriptions Enable Learning and Broader Adoption of More Robust Methods

Starting Model(s)

Initial Refinement Model(s)		
Type	**Source**	**Accession Code**
experimental model	PDB	3DW8
in silico model	AlphaFold	Q96E09



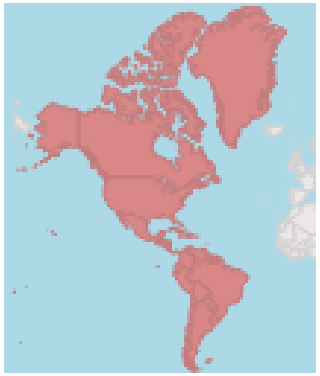


PDB ID: 8so0/EMD-40644

EM Software		
Task	**Software Package**	**Version**
PARTICLE SELECTION	RELION	4.0
PARTICLE SELECTION	Topaz	0.2.5
IMAGE ACQUISITION	EPU	3.0
CTF CORRECTION	CTFFIND	4.1.14
CTF CORRECTION	RELION	4.0
MODEL FITTING	UCSF ChimeraX	1.4
INITIAL EULER ASSIGNMENT	RELION	4.0
FINAL EULER ASSIGNMENT	RELION	4.0
RECONSTRUCTION	RELION	4.0
MODEL REFINEMENT	PHENIX	1.20.1-4487
MODEL REFINEMENT	Coot	0.9

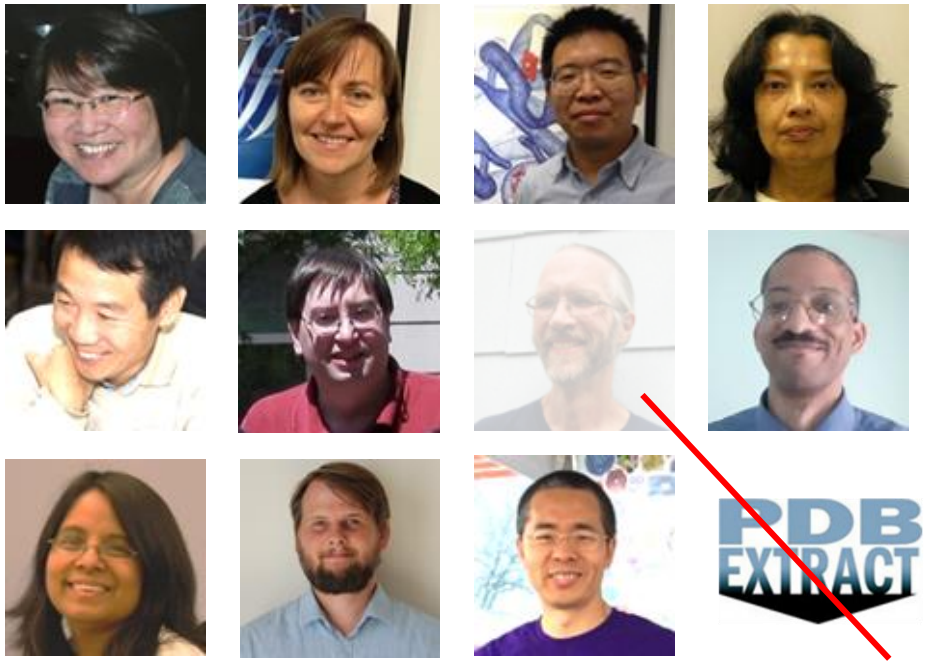
Map-Model Fitting and Refinement	
Id	1 (3DW8)
Refinement Space	REAL
Refinement Protocol	FLEXIBLE FIT
Refinement Target	Cross-correlation
Overall B Value	
Fitting Procedure	
Details	Iterating between manual refinement in Coot and automated real-space refinement in Phenix

RCSB PDB Support for 3DEM Structure Deposition



RCSB team handles all depositions submitted from Americas

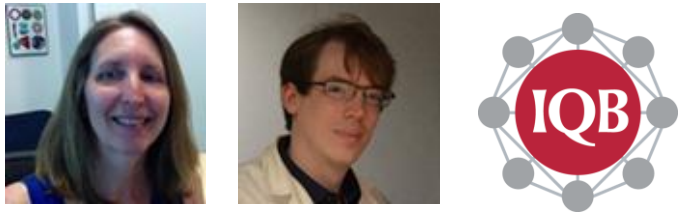
Team



RCSB PDB has processed over 17,600 EM entries. Always happy to help the 3DEM community!

Brian retired (Feb 2026); position open for EM expert!

Advisors



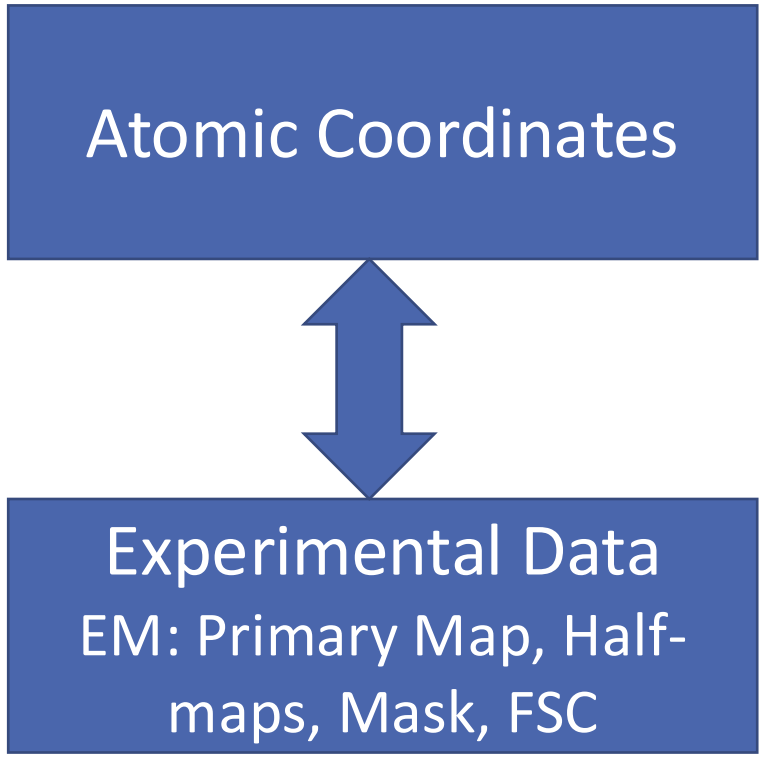
EMDataResource
Unified Data Resource for 3DEM

III. Quality at Scale: Validation of 3DEM structures – Insights, Best practices, What's Next

Evolving Validation Standards for 3DEM

What Gets Validated:

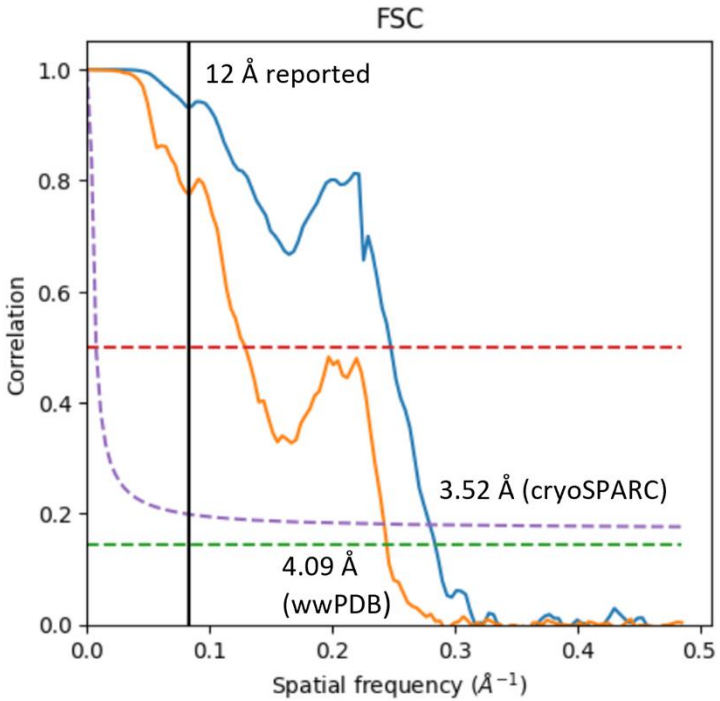
- Molecular geometry quality against established chemical references
- Experimental data quality
- Goodness-of-fit between atomic coordinates and experimental data

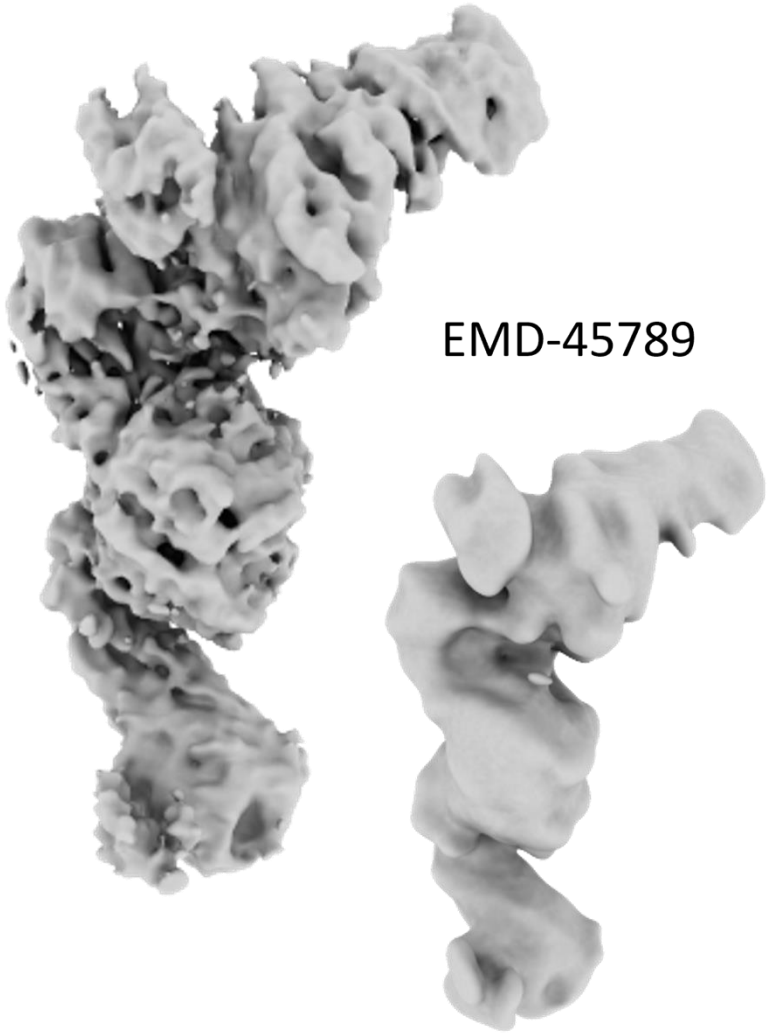


- High-quality PDB data essential for biomedical research
- Structure validation standards and tools developed by community experts
- 3DEM Validation Task Force publishes recommendations and consensus OneDep implementation

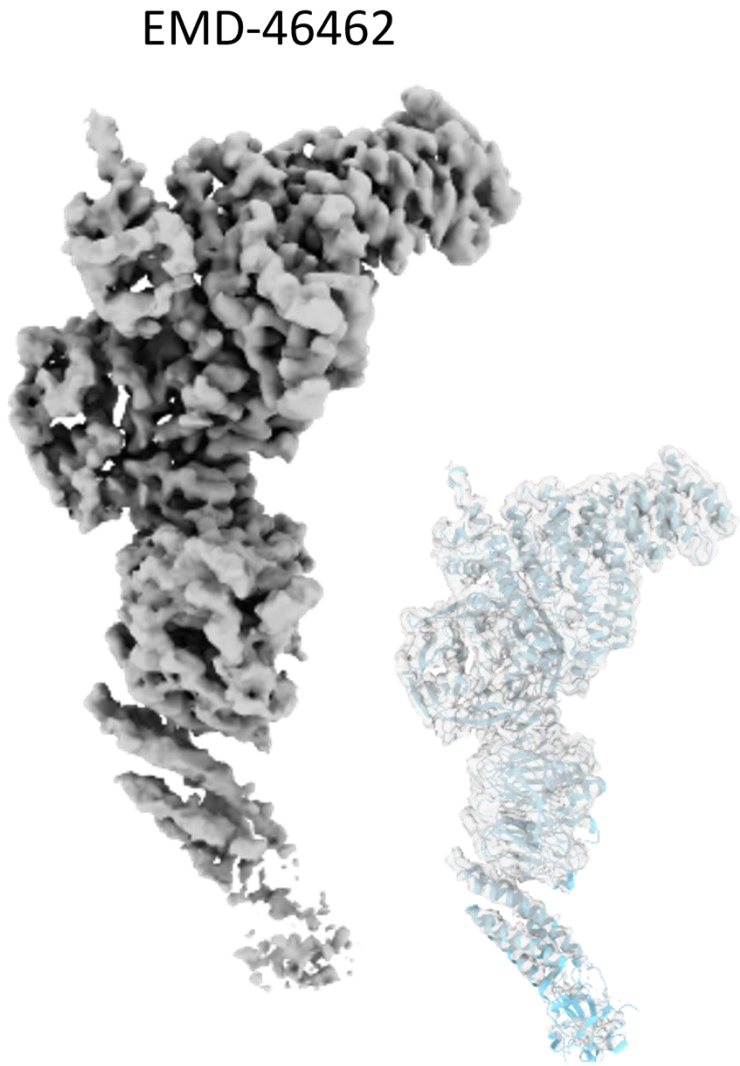
Struct Dyn. 2020 Jan 24;7(1):014701. doi: 10.1063/1.5138589.

Never Take the FSC at Face Value



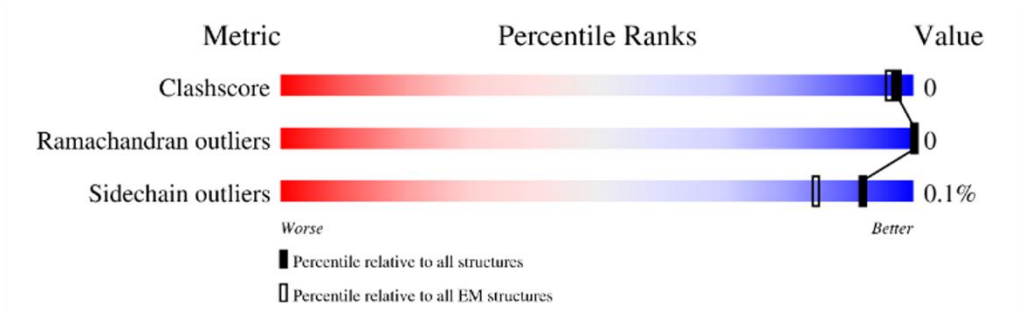


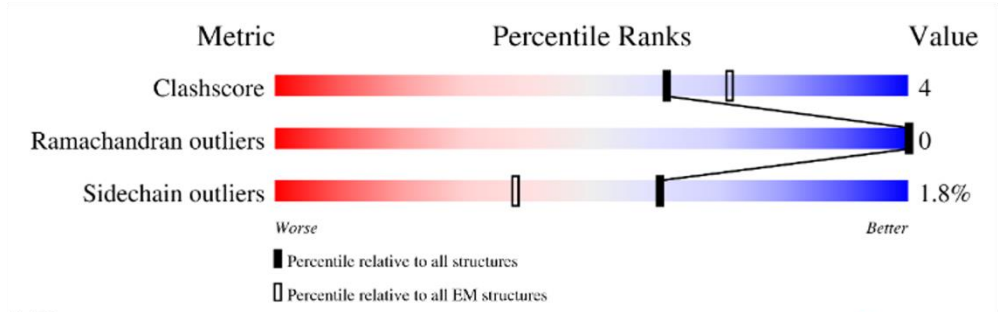
vs.



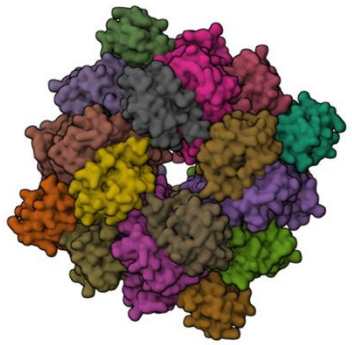
Always check local density and compare to similar-resolution maps

Looking Good is not Always Being Right





VS.

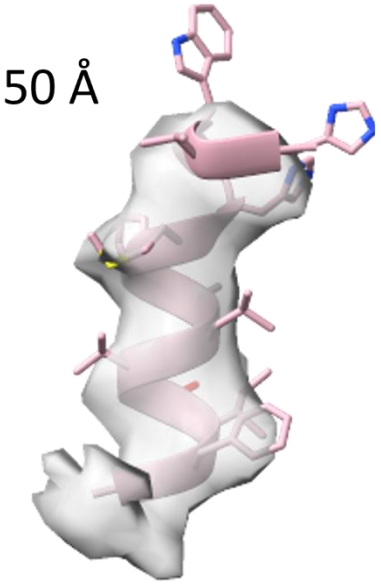


Subtomogram averaging

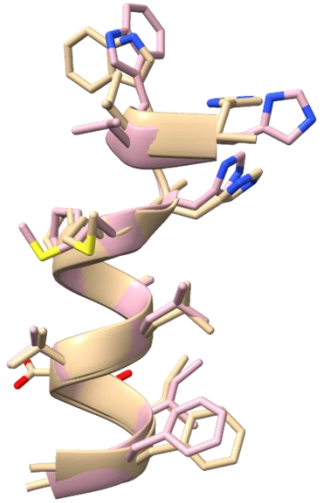
9fwv/EMD-50836

Chain C: 383-HVWHMPALVEIFGG-395

3.50 Å

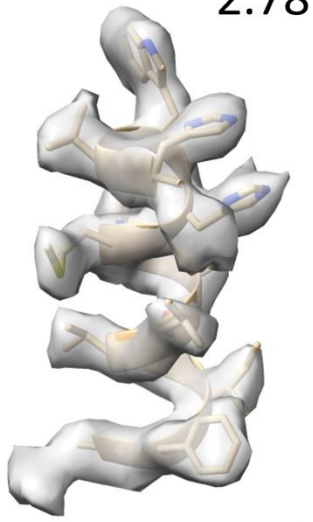


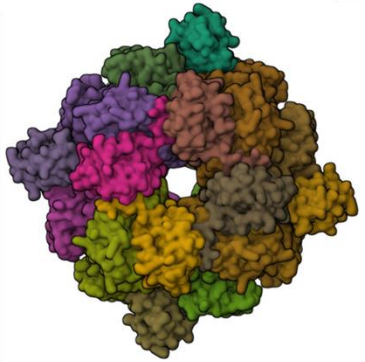
Fit to map



Superimposed on 6hbc

2.78 Å





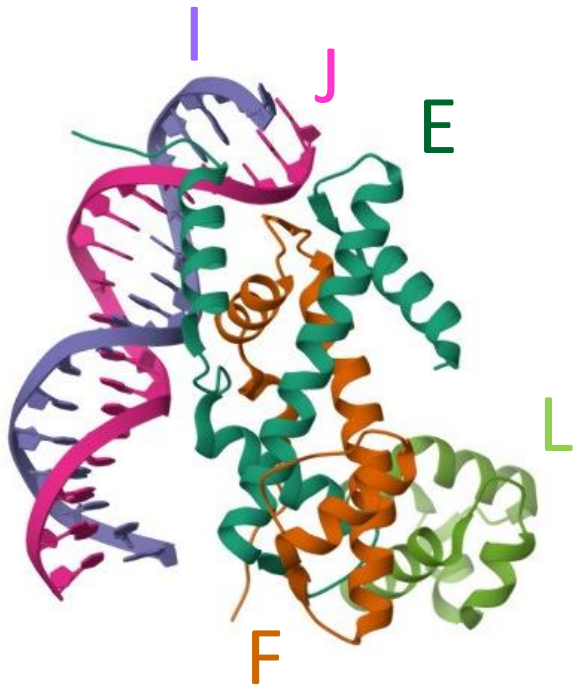
Single Particle

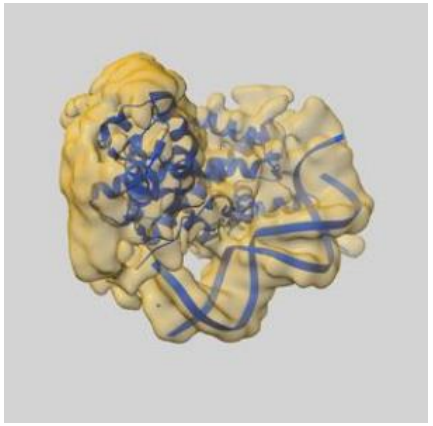
6hbc/EMD-0180

Chain C: 383-HVWHMPALVEIFGG-395

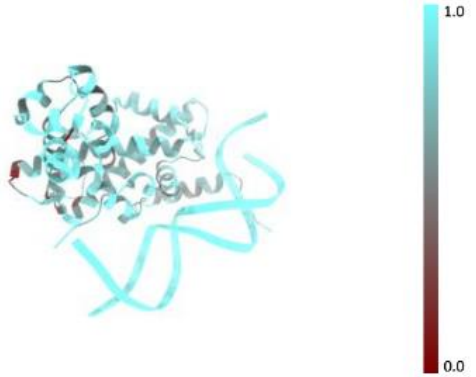
Outliers can be refined away, creating deceptively perfect scores for imperfect structures

Looking “Too Good” is Easier to Spot

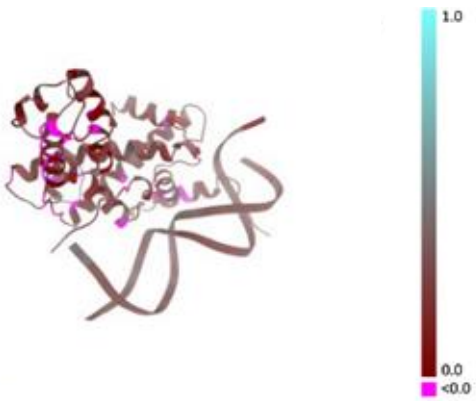




Atom inclusion



Q-score



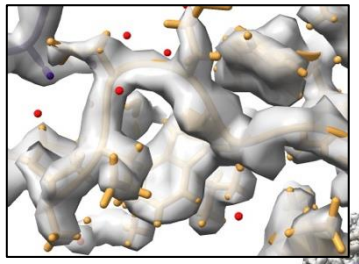
Chain	Atom inclusion	Q-score
All	0.8450	0.2190
E	0.7770	0.1630
F	0.7830	0.2130
I	1.0000	0.3320
J	0.9890	0.3210
L	0.8030	0.1510

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	2.89	-	-
Author-provided FSC curve	2.89	3.23	2.94
Unmasked-calculated*	3.49	4.29	3.57

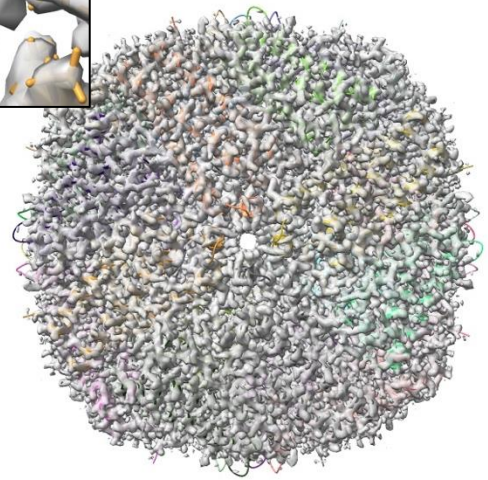
Pintilie, G., Zhang, K., Su, Z. *et al.* Measurement of atom resolvability in cryo-EM maps with *Q*-scores. *Nat Methods* **17**, 328–334 (2020). https://doi.org/10.1038/s41592-020-0731-1

Good Compared to What?

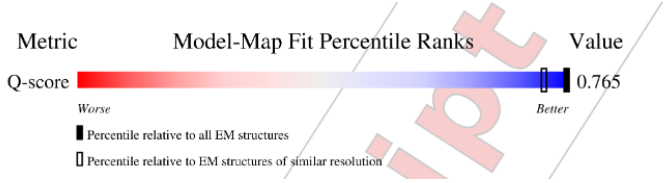
Well-fitted



2.0 Å Reported
2.26 Å Calculated

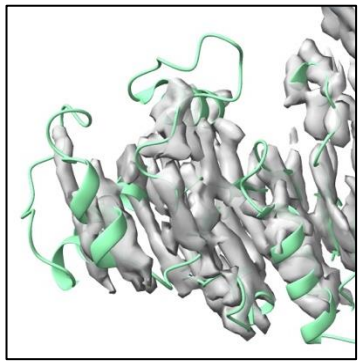


PDB ID: 6wx6
EMD-21951

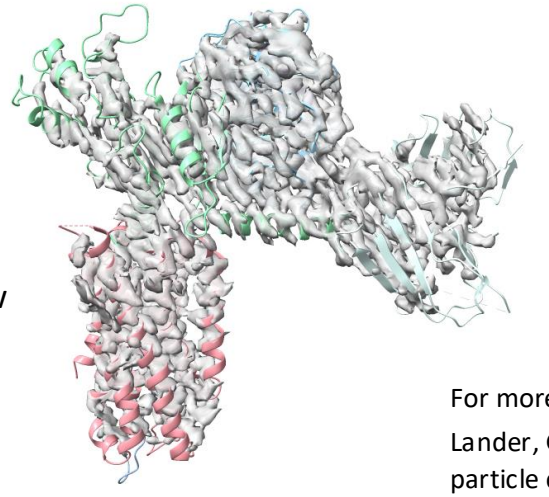


Compared to 25,397 EM structures
1,659 structures at similar resolution (1.50-2.50)

Preferred Orientation

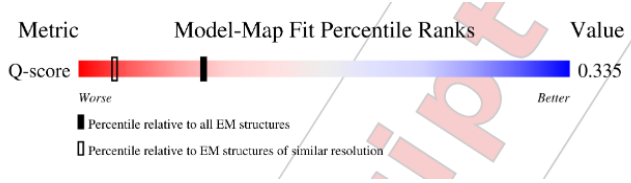


Rotated around X axis 45°



PDB ID: 7f8w
EMD-31494

3.1 Å Reported
No half-maps, no estimate



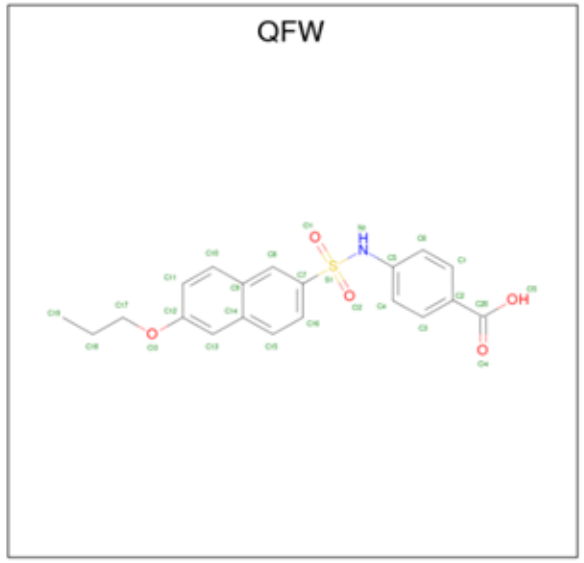
Compared to 25,397 EM structures
14,724 structures at similar resolution (2.60-3.60)

For more on this case see:

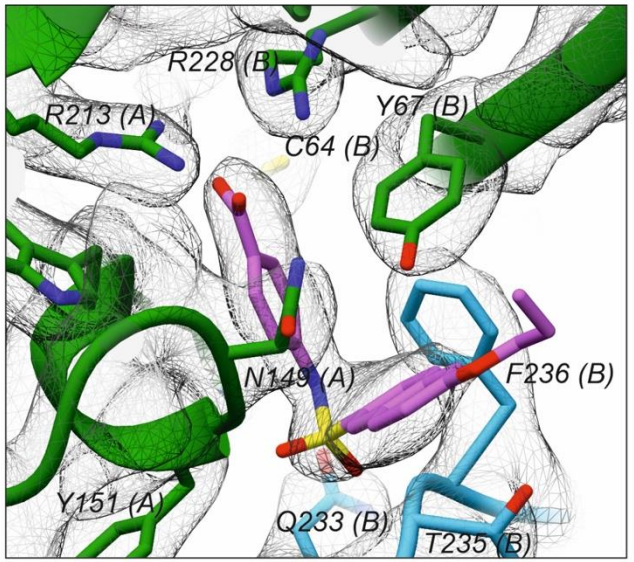
Lander, G. C. (2024). Single particle cryo-EM map and model validation: It's not crystal clear. Current Opinion in Structural Biology, 89, 102918. https://doi.org/10.1016/j.sbi.2024.102918

Modeling Ligands Are Double the Trouble

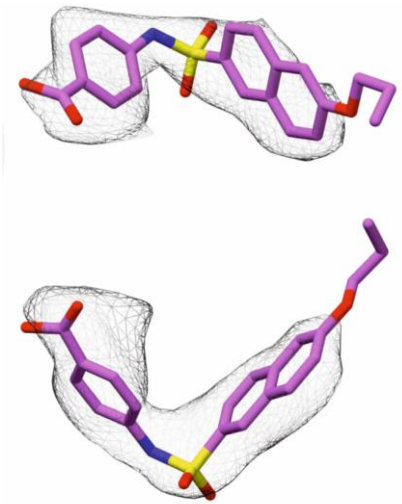
Ligand description



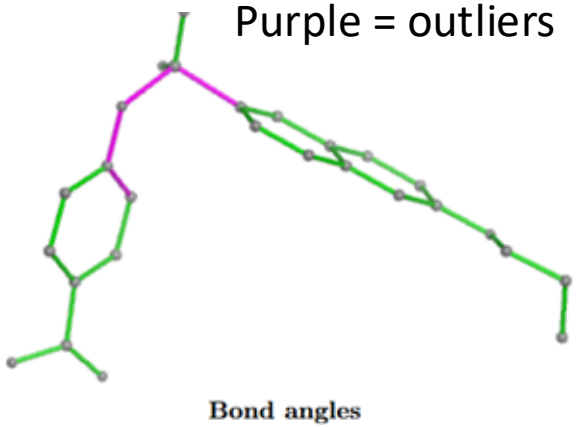
PDB 6zck/EMD-11165



Flatt, J.W., *et al.* (2020) *Commun Biol* https://doi.org/10.1038/s42003-021-01779-x

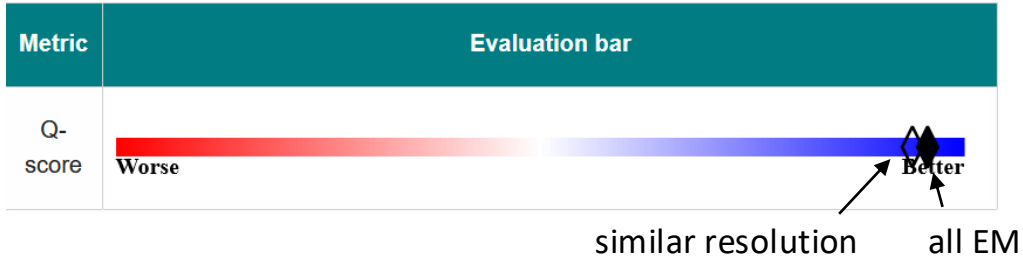


Avg. Q-score: 0.624
Ligand Q-score 0.544



Geometry validated against the Cambridge Structural Database (CSD); environmental clashes measured; per-ligand fit to experimental data assessed using Q-score

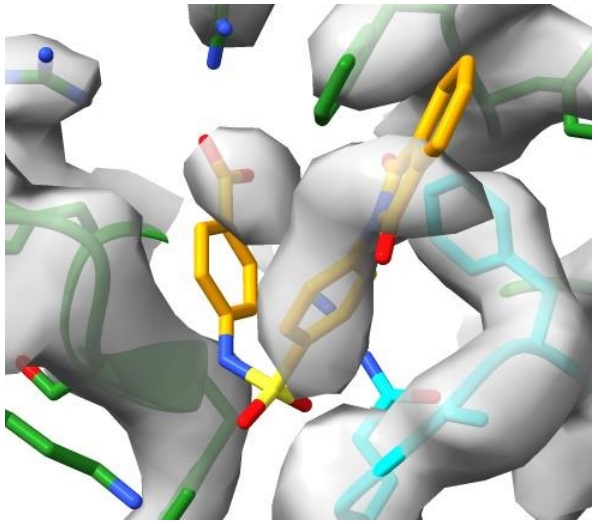
Always verify that map fit, stereochemistry, and interactions are chemically sensible.



Let Q-scores be a Guide to Your Ligands

PDB ID: 6gzv/EMD-0103

4.0 Å
Q Struct: 0.454
Q Ligand: 0.2

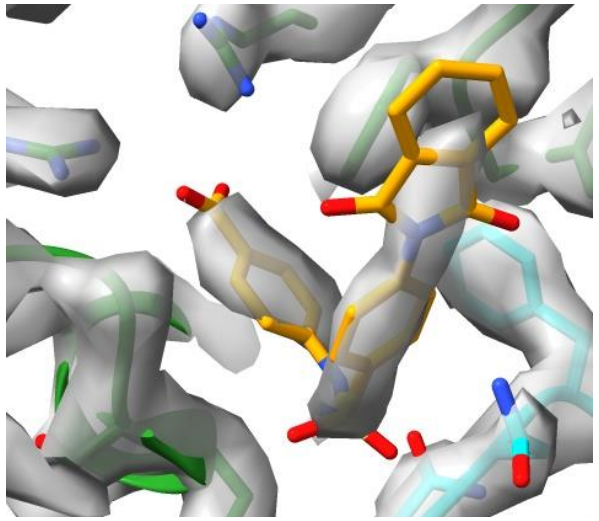


vs.

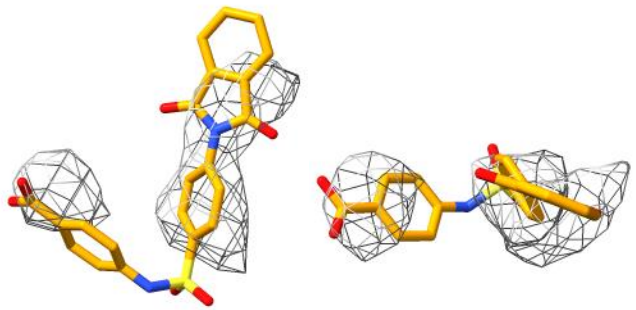
PDB ID: 6zcl/EMD-11166

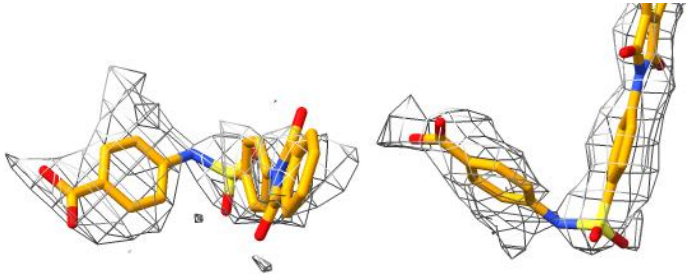
Reprocessed and remodeled

2.8 Å
Q Struct: 0.62
Q Ligand: 0.52



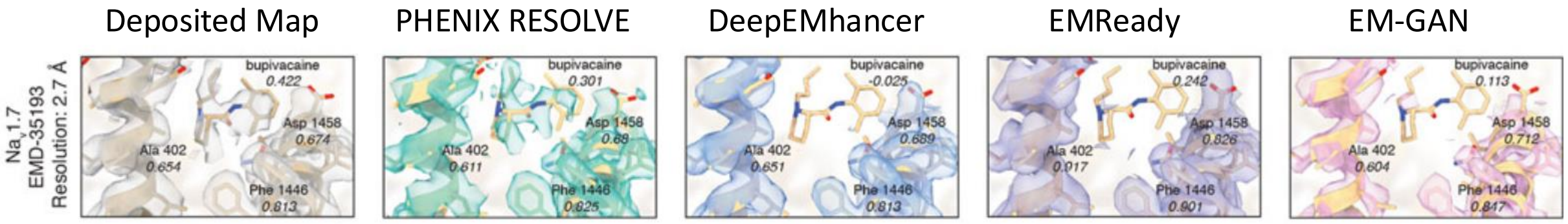
Fit 1





Fit 2

Caution: ML Tools Can Distort Ligand Density

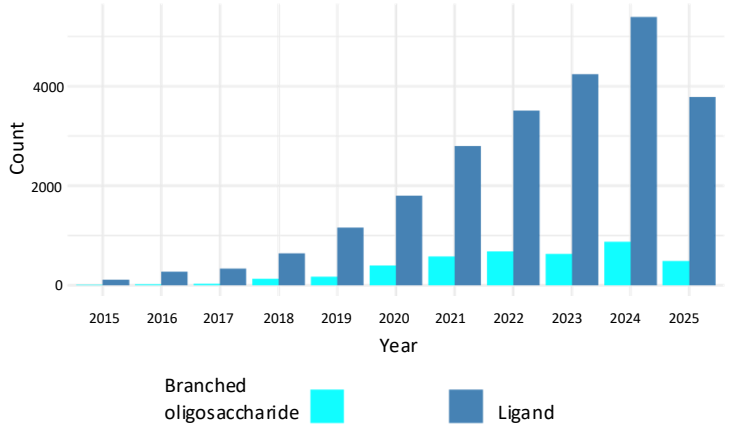


Berkeley, R.F., Cook, B.D., Herzik Jr, M.A. (2024). Machine learning approaches to cryoEM density modification differentially affect biomacromolecule and ligand density quality. Front. Mol. Biosci, Sec. Structural Biology. Vol 11, https://doi.org/10.3389/fmolb.2024.1404885.

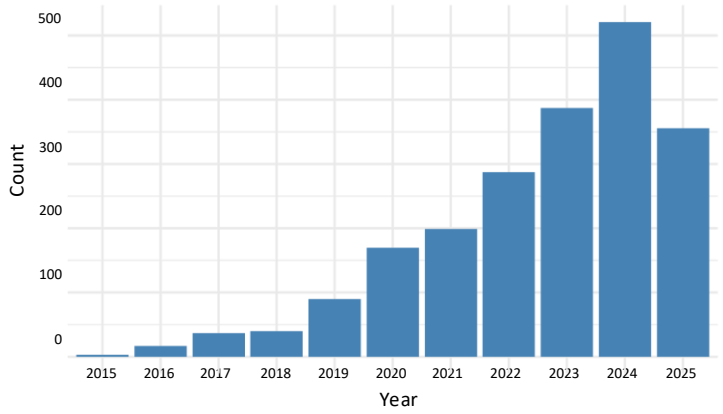
Extreme cases show ligands erased from the map

Bird’s Eye View of Ligand Quality Using Q-scores

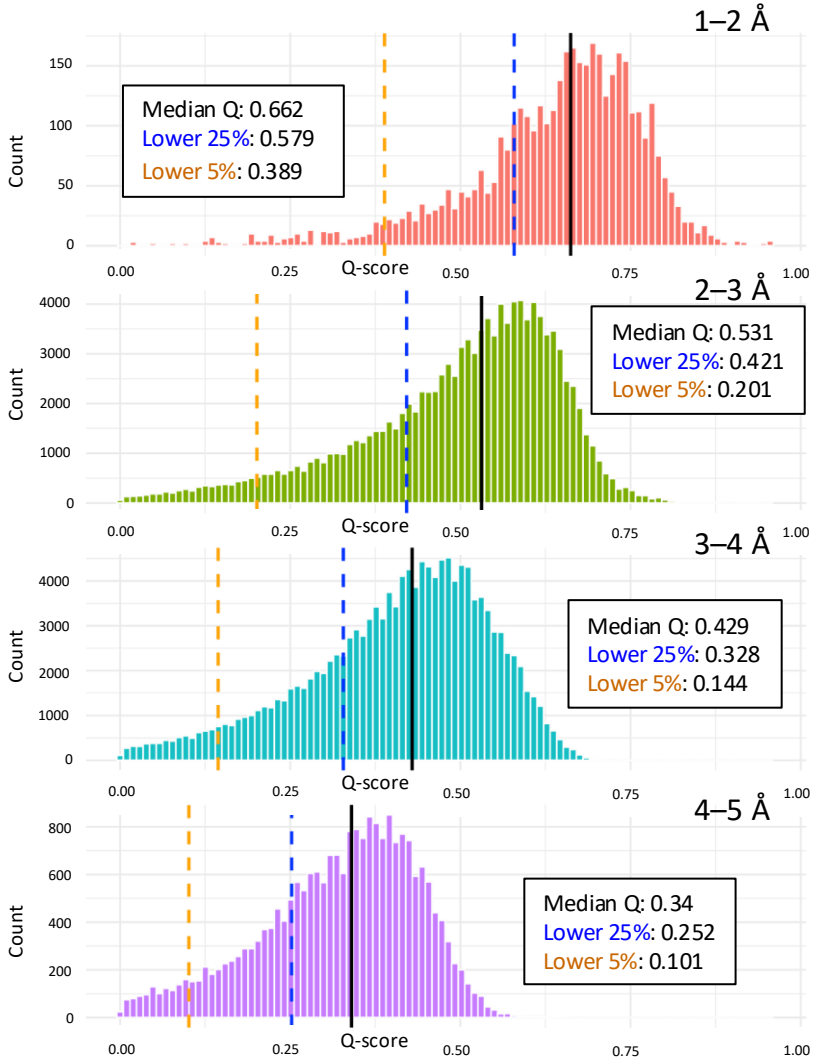
Cryo-EM Depositions with Ligands or Branched Oligosaccharides

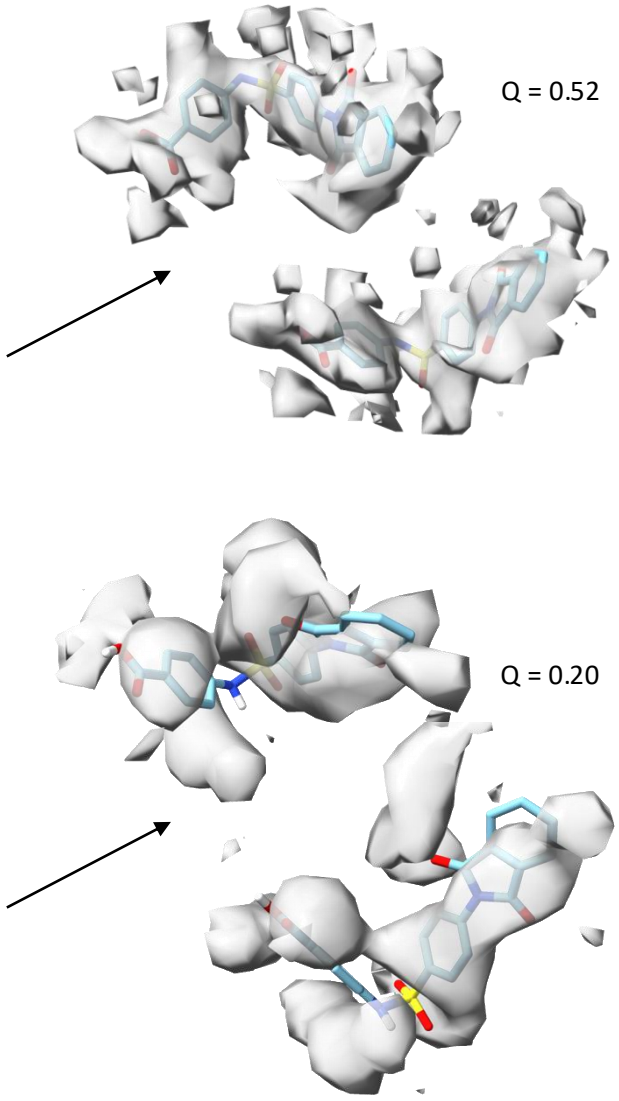


New Ligands Entering the CCD



Ligand Q-score Distribution Across Resolution Bins





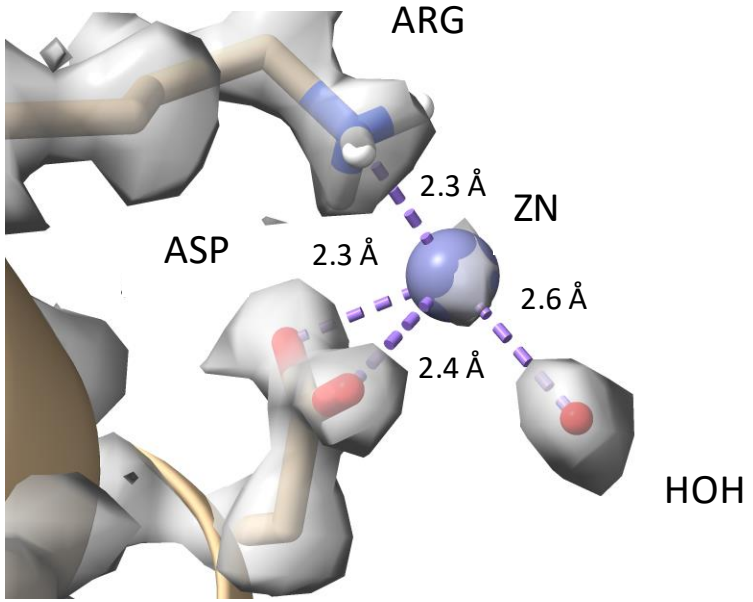
Locking in Your Best Possible Model

- Validate early, validate often
- Leverage all available validation tools
- Model within the map's resolution limits
- Visually inspect local fit across the structure
- Review official validation report and any correspondence carefully
- Note: Coordinates can be updated post-release, but the primary map cannot

https://validate.wwpdb.org

Welcome to the wwPDB validation system!

This server performs the same validation as you would observe during the deposition process.



Version Number	Version Date	Version Type/Reason	Version Change
1.0	2023-10-11	Initial release	
1.1	2023-12-13		Data collection, Database references
2.0	2024-08-21	Coordinate replacement / Model orientation/position	Advisory, Atomic model, Data collection, Database references, Derived calculations, Experimental preparation, Refinement description, Source and taxonomy, Structure summary

The most valuable resource is the EM community

RCSB PDB PROTEIN DATA BANK

RCSB.ORG • info@rcsb.org

Core Operations Funding

US National Science Foundation (DBI-2321666), National Institute of General Medical Sciences, National Institute of Allergy and Infectious Disease, and National Cancer Institute (NIH R01GM157729), and the US Department of Energy (DE-SC0019749)

Management

RUTGERS–NEW BRUNSWICK
Artificial Intelligence and Data Science Collaboratory

UC San Diego

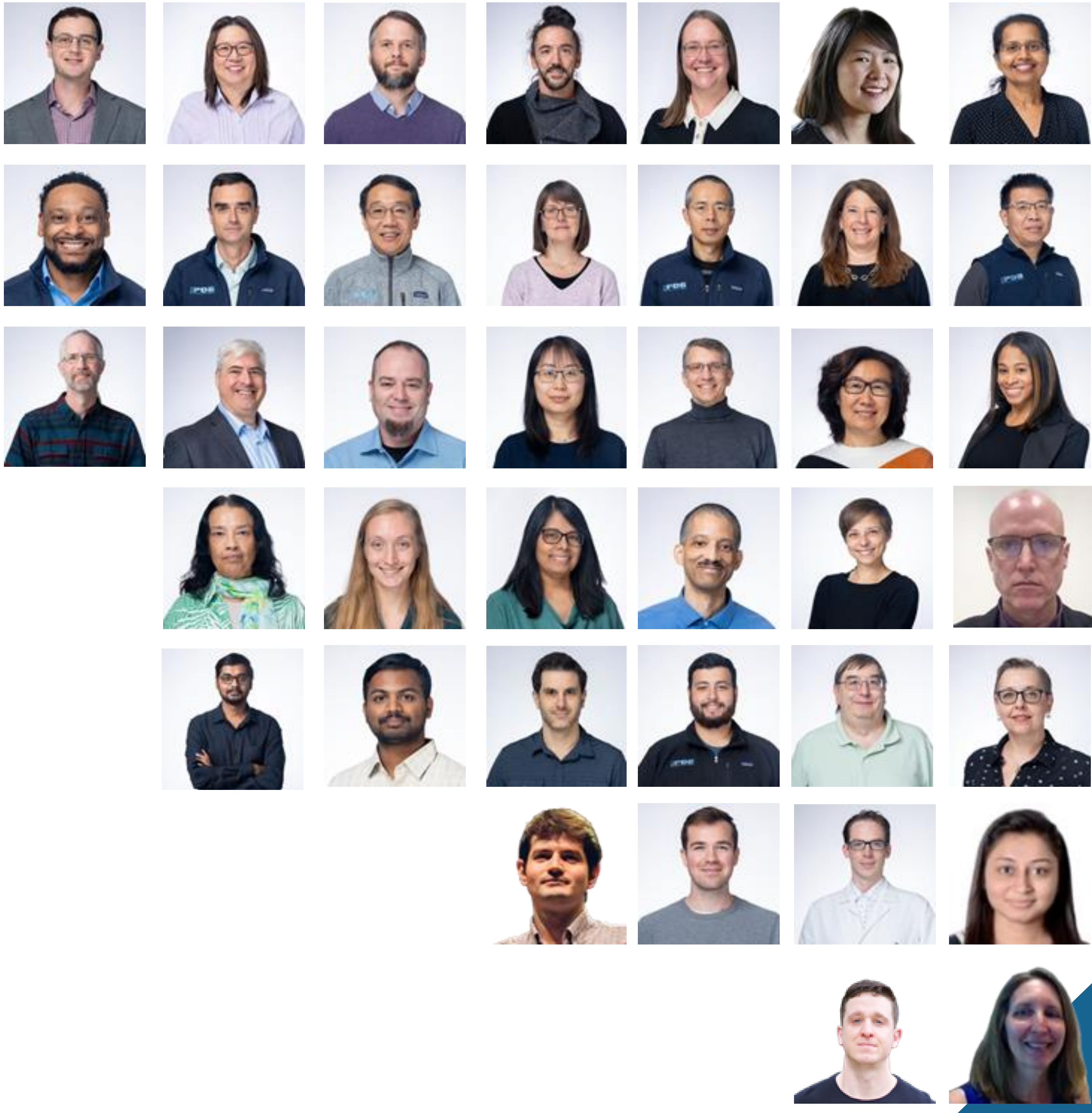
SDSC SAN DIEGO SUPERCOMPUTER CENTER

UCSF University of California San Francisco

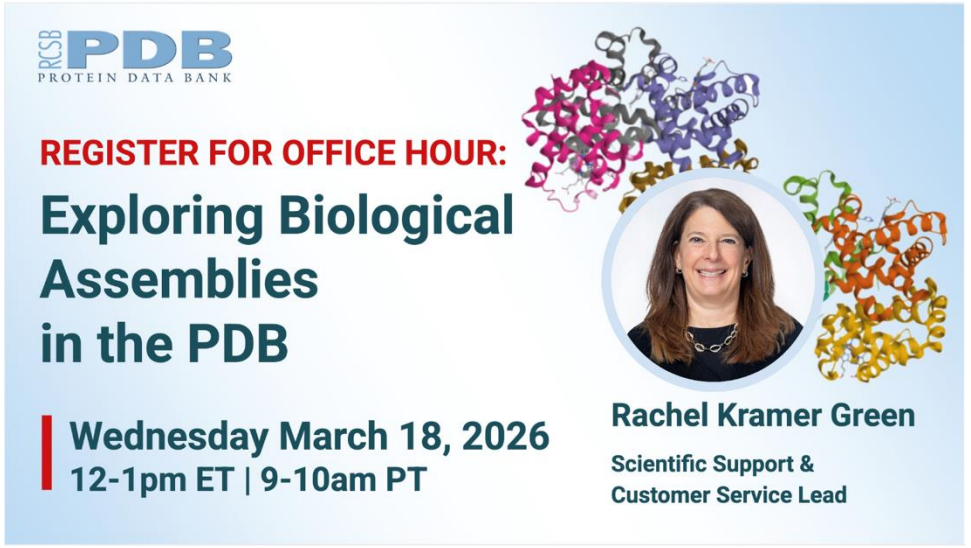
Follow us

WORLDWIDE PDB PROTEIN DATA BANK

Member of the Worldwide Protein Data Bank (wwPDB; **wwpdb.org**)



Stay Connected with the RCSB PDB Team



Recordings posted at PDB-101.rcsb.org