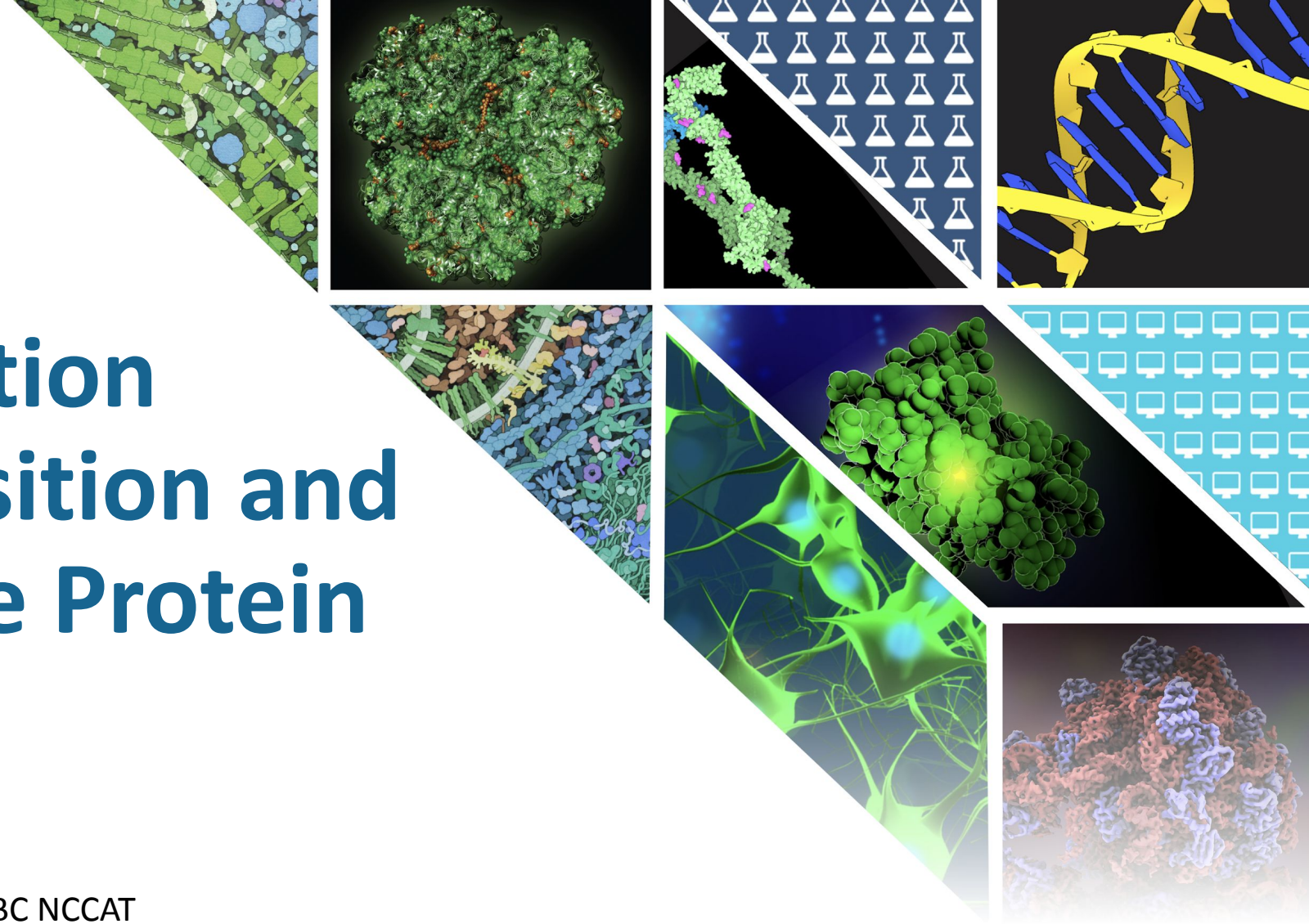
RCSB PDB
PROTEIN DATA BANK

rcsb.org



Electron Diffraction Structure Deposition and Validation in the Protein Data Bank

Nov 3rd, 2025

MicroED Symposium and Short Course at NYSBC NCCAT

Chenghua Shao, Ph.D.

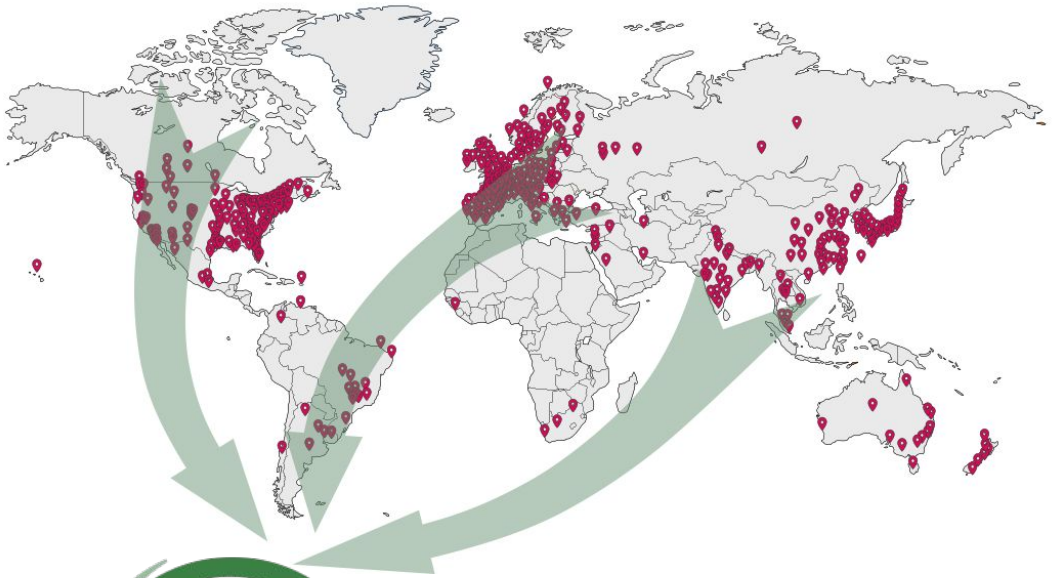
RCSB Protein Data Bank, Rutgers, The State University of New Jersey

Outline

- Electron Crystallography and 3D ED/MicroED structures in the PDB
- Current PDB 3D ED/MicroED structure deposition and validation
- Improved 3D ED/MicroED data model for future deposition and validation

Protein Data Bank (PDB) Archive

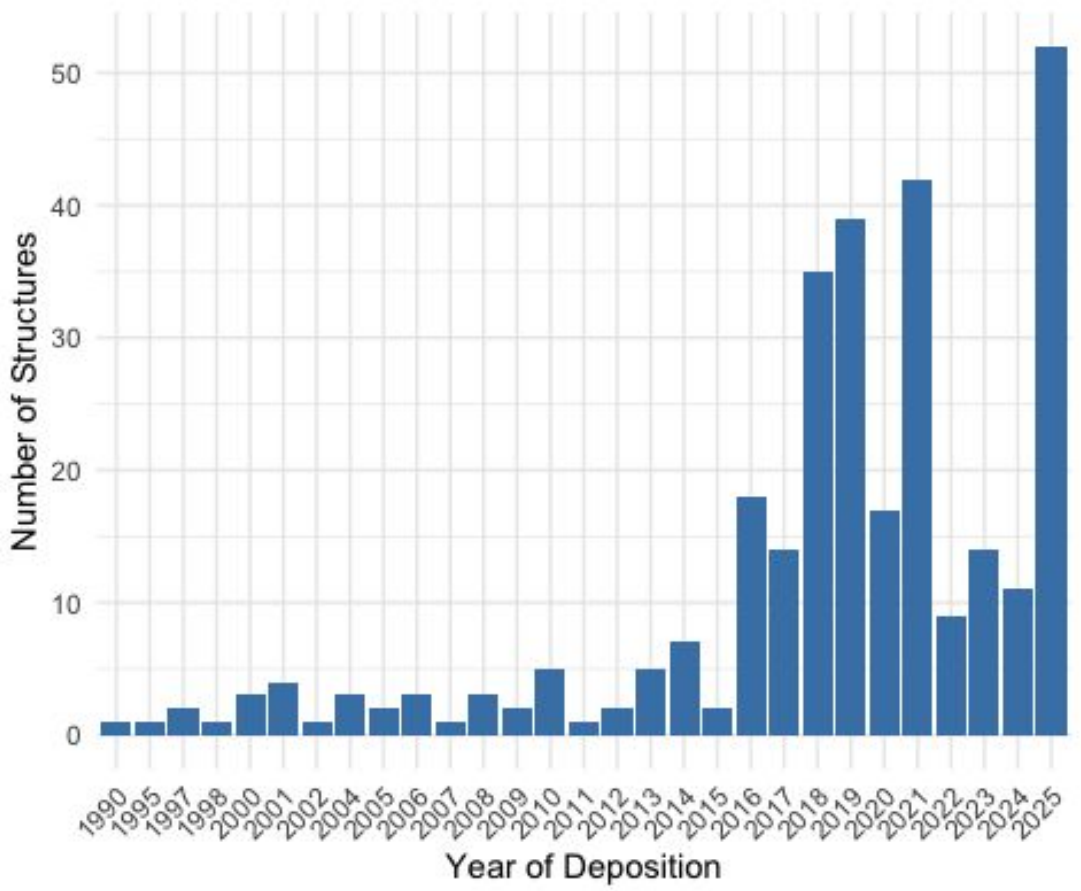
- 1st open access digital data resource in all of Biology established in 1971
- Single global archive for protein and nucleic acid experimental structures with over 240,000 structures; 2025 yearly deposition approaching 20,000
- Managed jointly by Worldwide PDB (wwPDB) regional partners
    - RCSB PDB (US)
    - Protein Data Bank in Europe (PDBe)
    - PDB Japan (PDBj)
    - Associate Member: PDB China (PDBc)
    - Plus EMDB and BMRB
- All PDB data are validated, deposited, and biocurated using OneDep



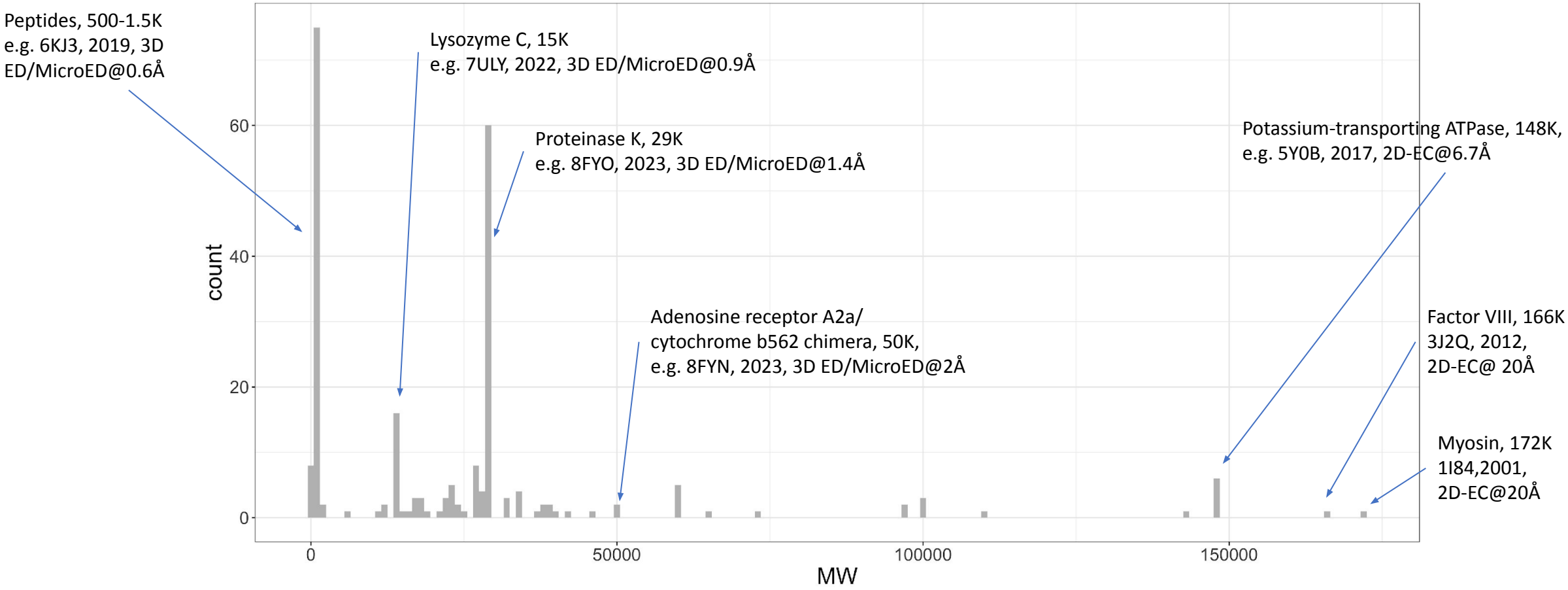
OneDep

Electron Diffraction Structures in the PDB

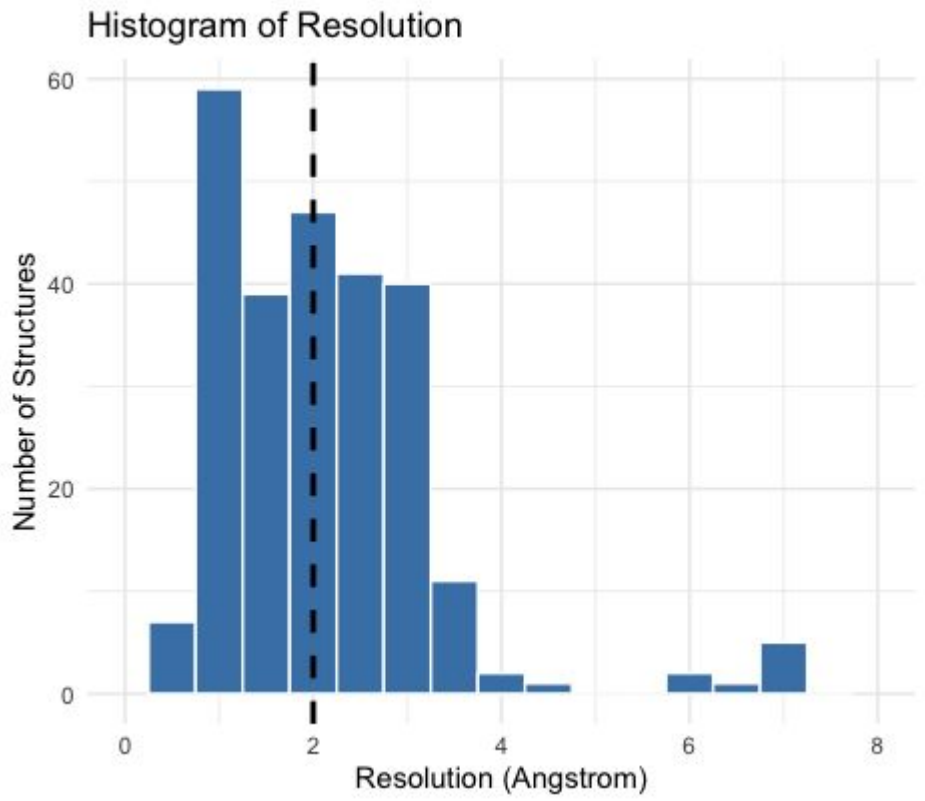
- 2D EC (2-Dimensional Electron Crystallography) was used for protein structures before 3D ED/MicroED
- 1990: 1st PDB 2D EC structure of bacteriorhodopsin @ 3.5 Å (PDB: 1BRD)
- 2013: 1st 3D ED/MicroED structure of lysozyme @ 2.9 Å (PDB: 3J4G)
- 2016-2025: 3D ED/MicroED dominates, with a record of 52 in 2025 by Oct
- 300 deposited into PDB: ~20% from 2D EC; ~80% 3D ED/MicroED

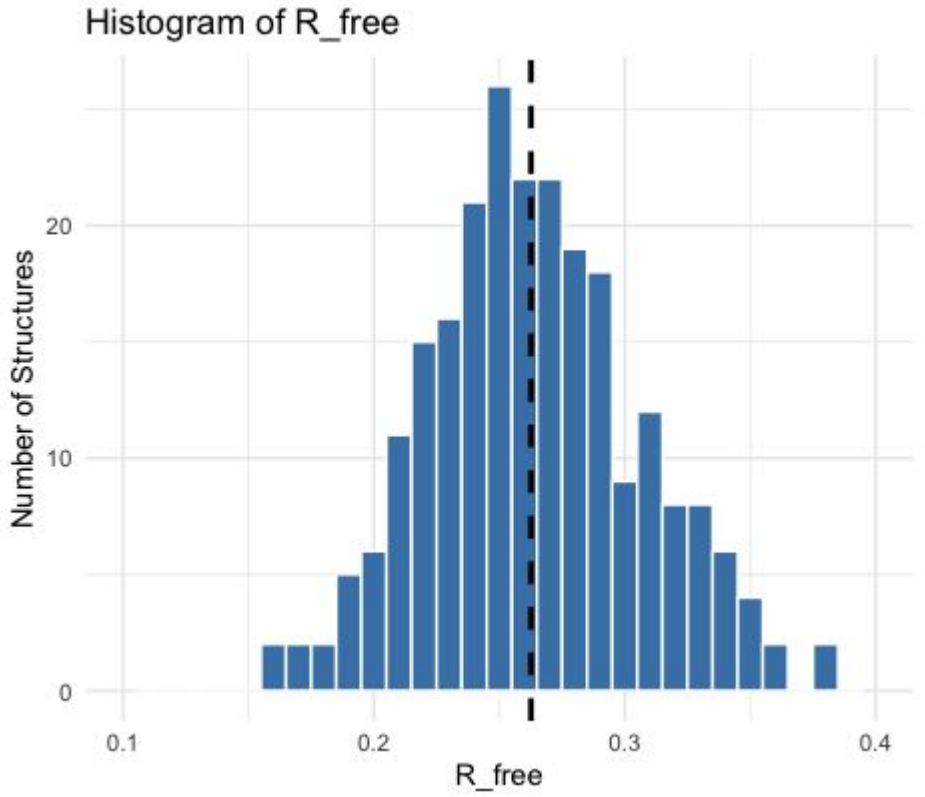


Molecules Studied by Electron Diffraction



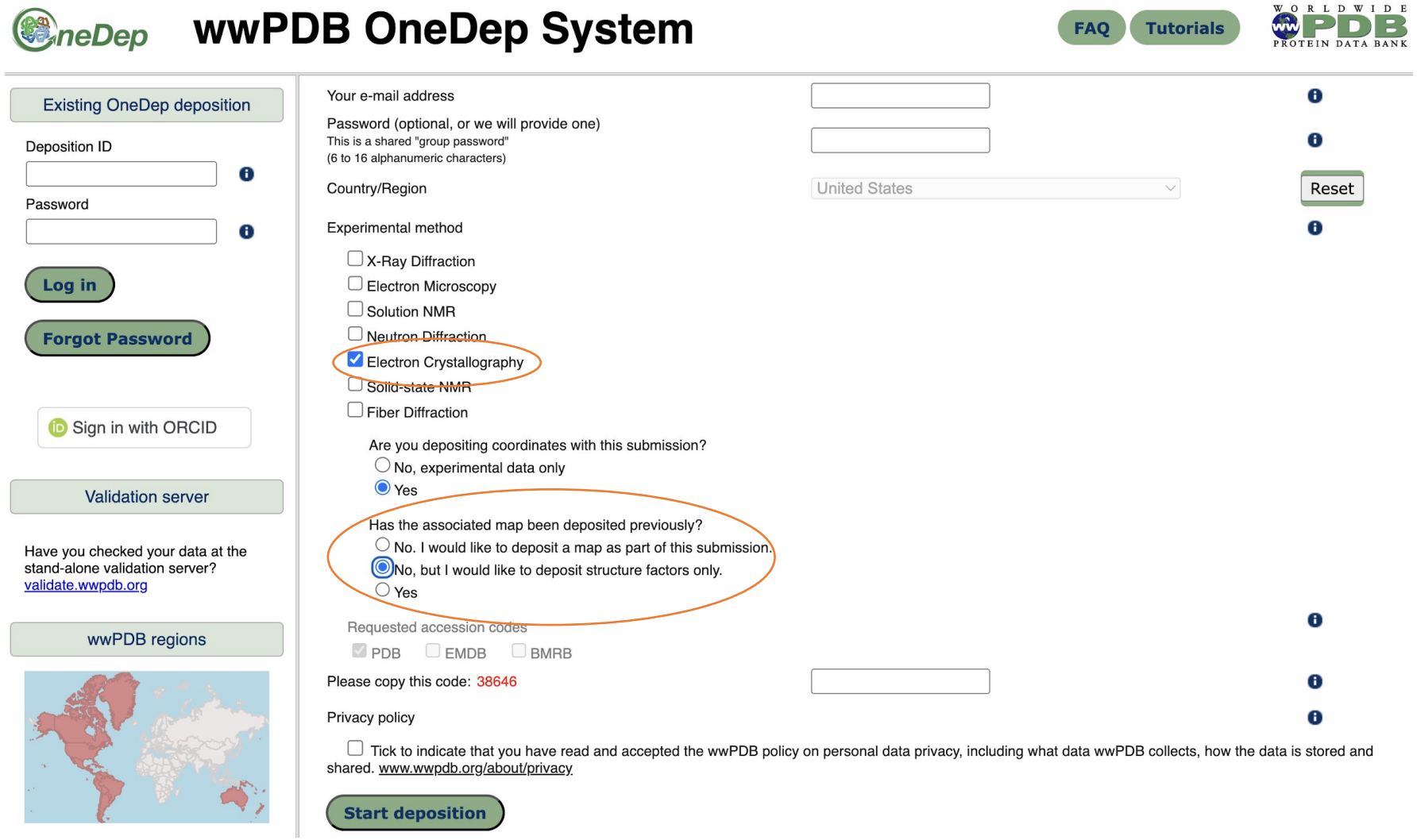
Electron Diffraction Data and Structure Quality



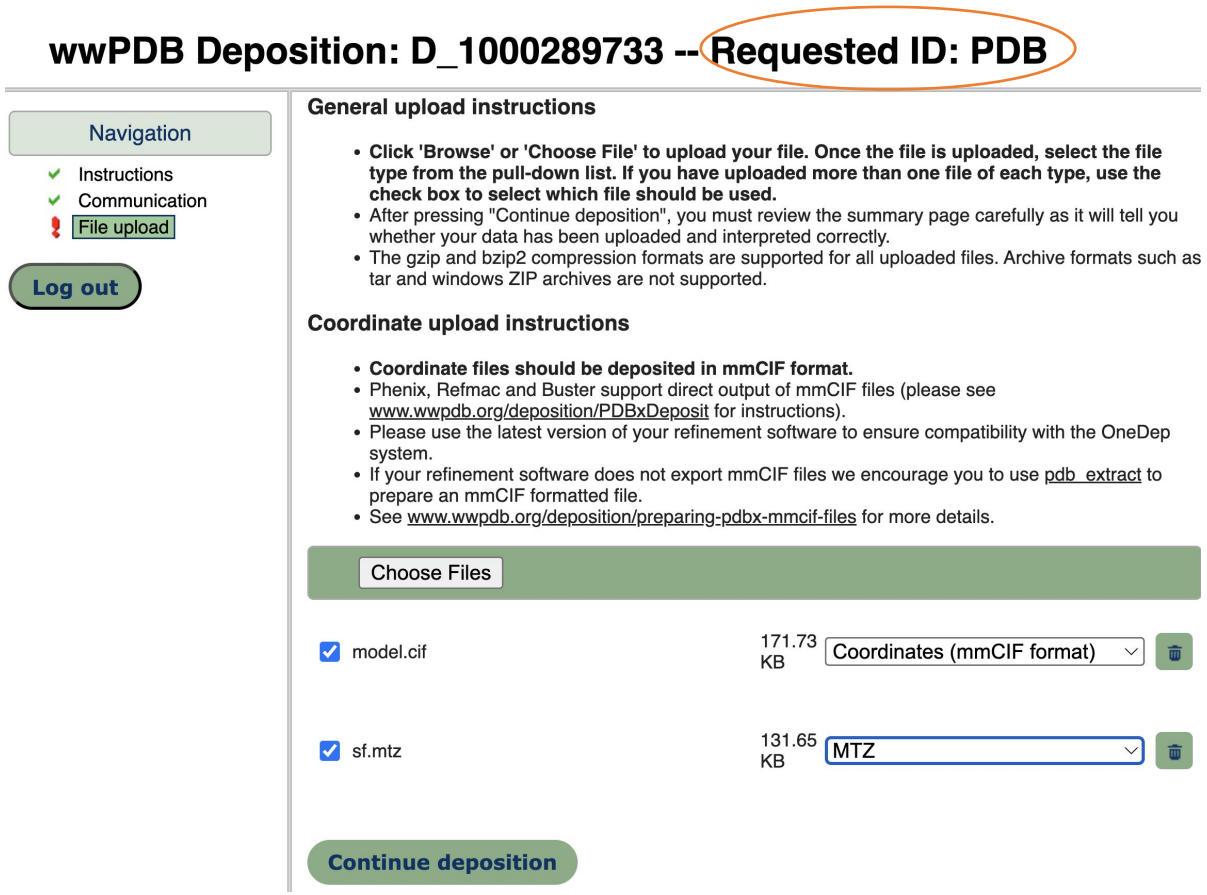


Current 3D ED/MicroED Data Deposition and Validation

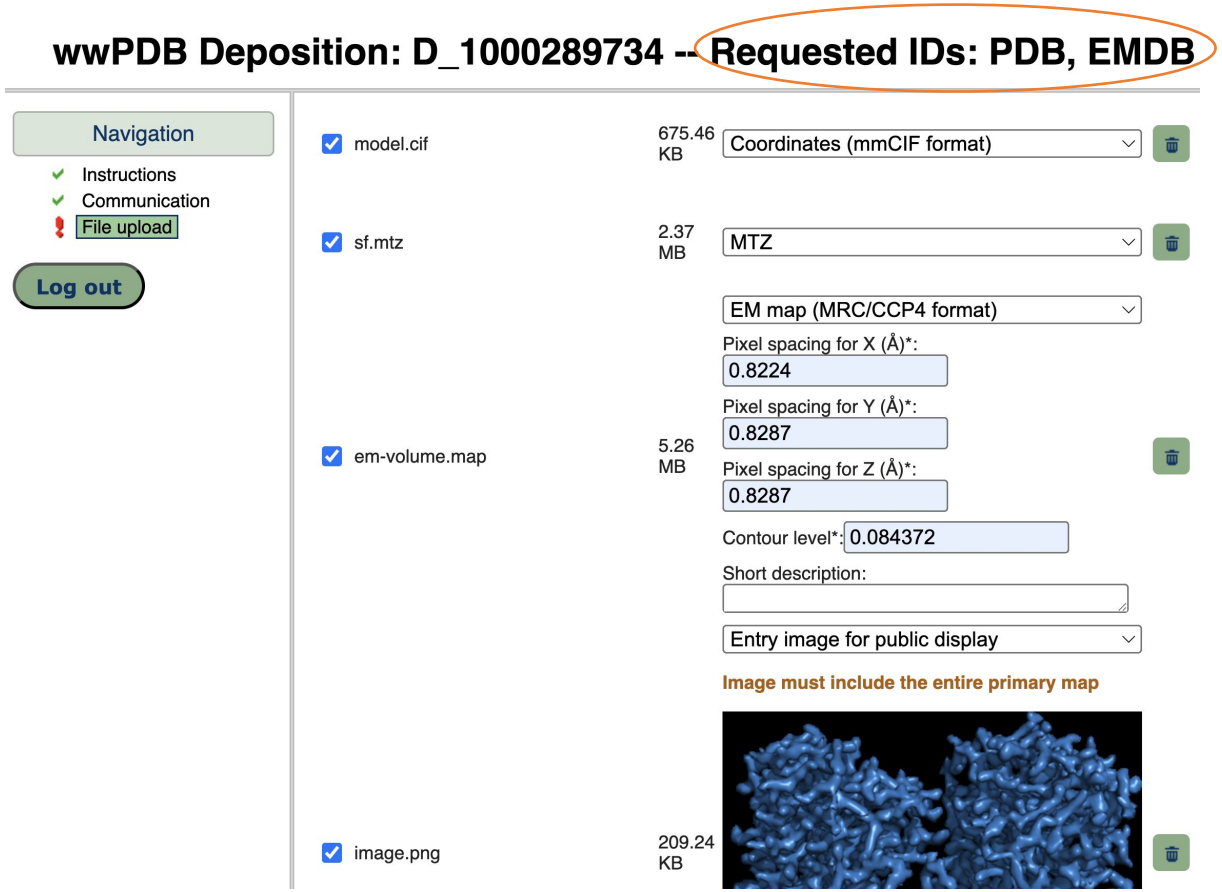
Start a 3D ED/MicroED Deposition



Upload Files

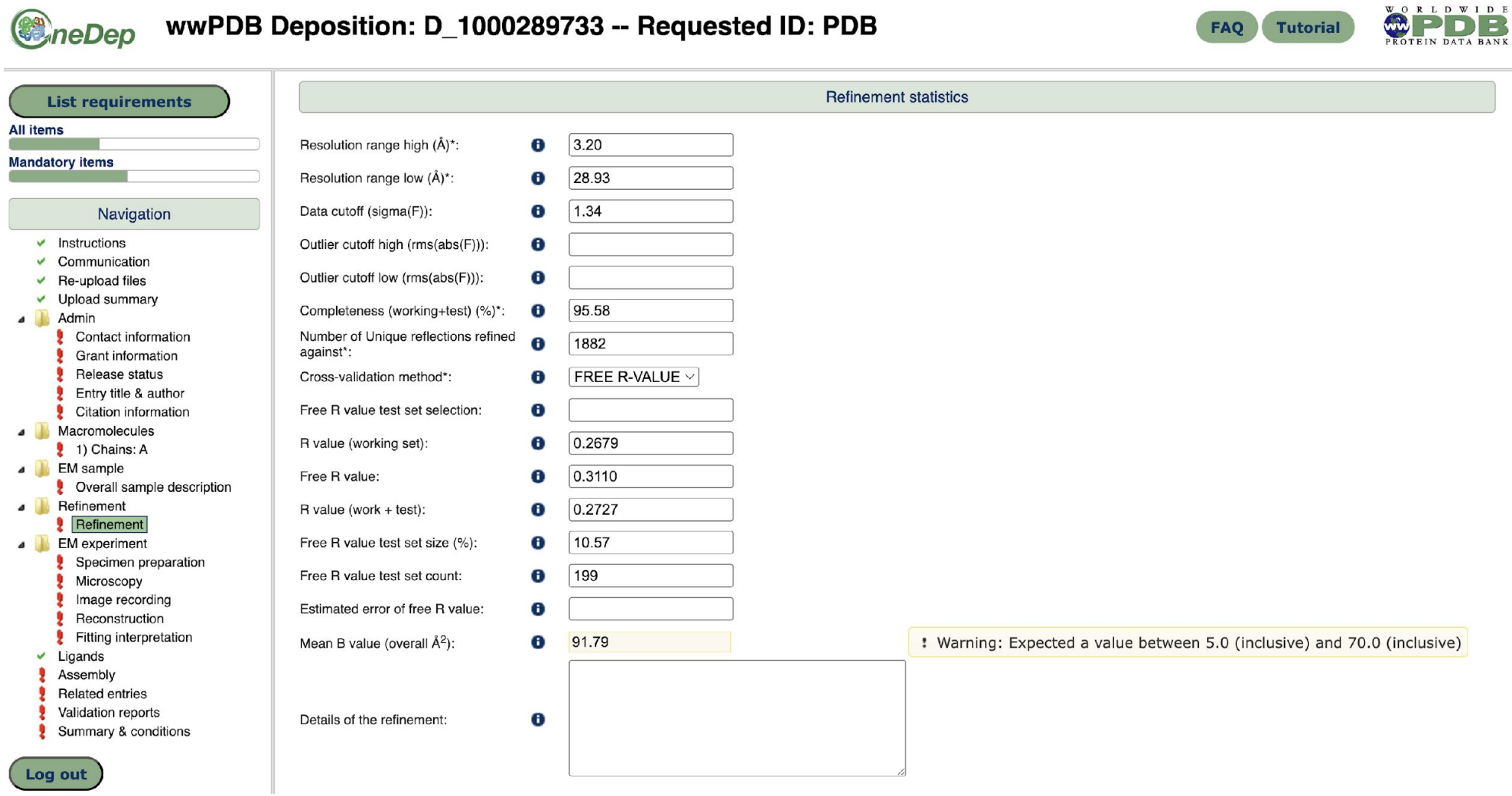


Model and SF



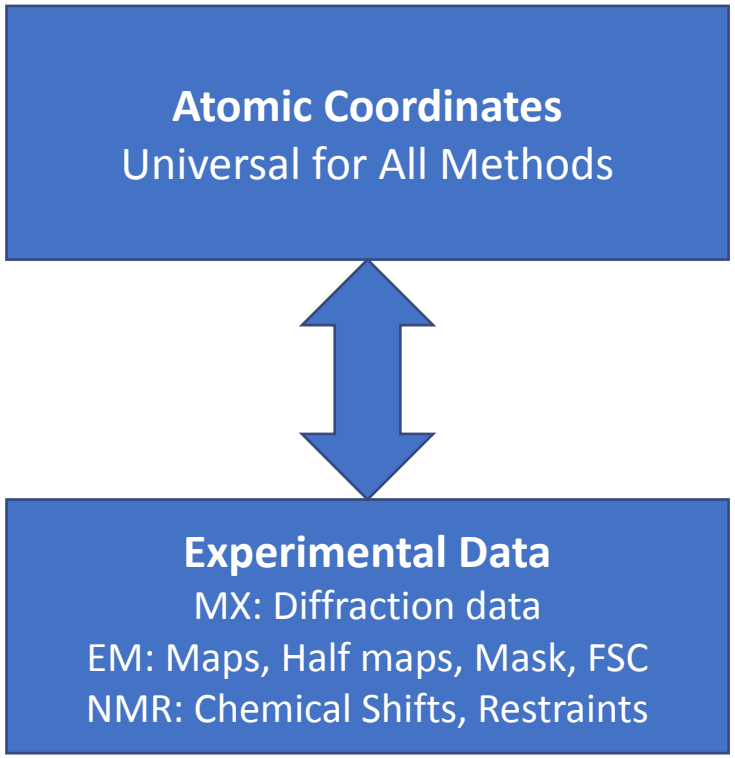
Model, SF, and Map

Provide Metadata at the Deposition Interface



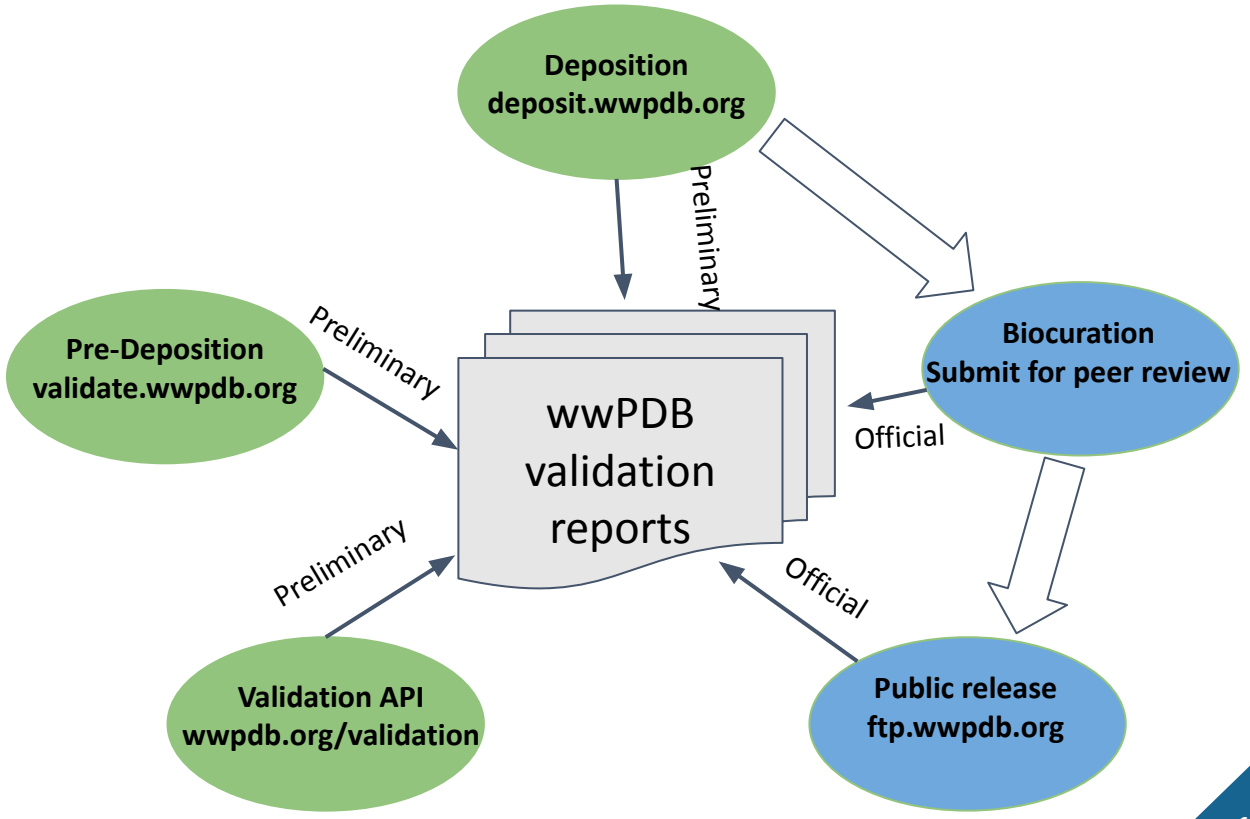
wwPDB Validation Scope

- Molecular geometry agreement with established chemical references (bond lengths, bond angles, *etc*.)
- Experimental data quality
- Goodness-of-fit between atomic coordinates and method-specific experimental data
- Global *vs.* local structure validation
- Validation for distinct molecular components (polymers, ligands, etc.)



wwPDB Validation Reports Tailored to Different Audiences

- Data Authors/Depositors: Can generate and access watermarked reports pre-/post-deposition
  - Deposition site
  - Standalone validation server
  - Application Programming Interface (API)
- Journals: Supporting peer review
  - Authors provide reports to journals
  - Journals provide reports to referees
  - Required by many journals
- Data Consumers
  - Access reports on all wwPDB partner sites
  - CIF/XML/PDF formatted reports available for download and analysis



Chemical Geometry Analysis

Polymers are analysed for the following geometry issues

- Bond Lengths
- Bond Angles
- Atom Clashes
- Ramachandran Outliers
- Sidechain Conformers
- Chirality Issues
- RNA backbone quality

Molprobity + PDB software

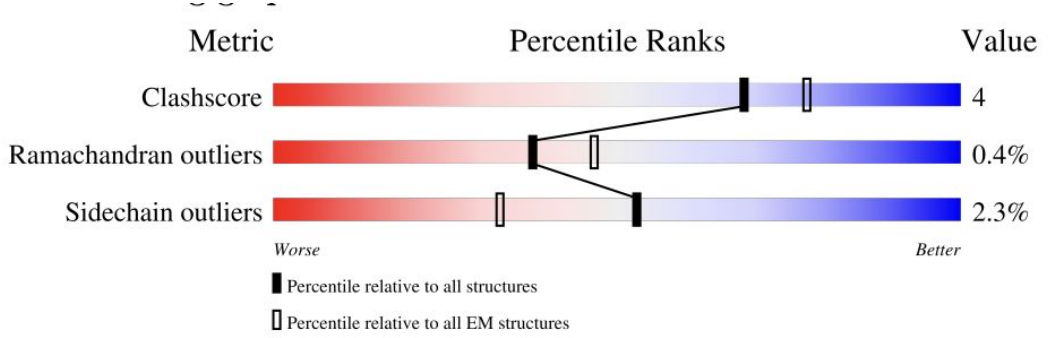
- Model quality
  - Standard Geometry
  - Too-close Contacts
  - Torsion Angles
  - Polymer Linkage Issues
- Residue-property plots

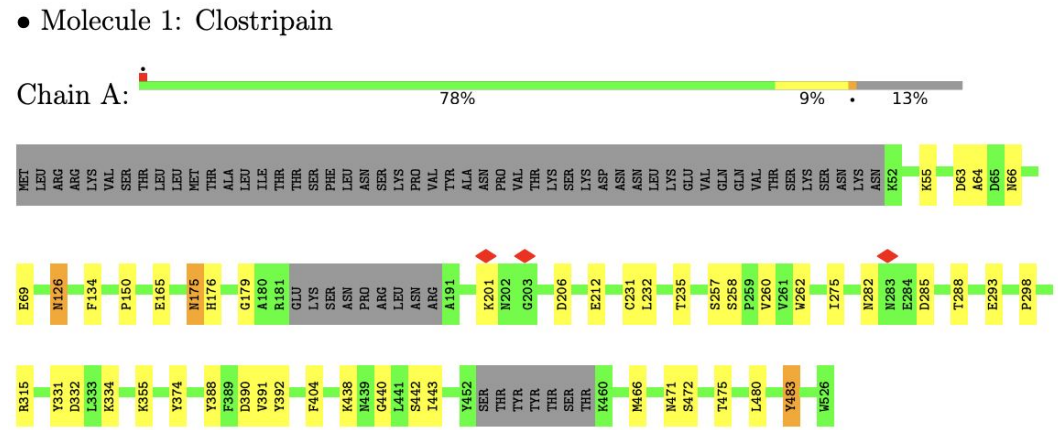
Reference for Protein: Engh, R. A. and Huber, R., Accurate bond and angle parameters for X-ray protein structure refinement, Acta Cryst. A47:392-400, 1991; Engh, R. A. and Huber, R., Structure quality and target parameters, International Tables for Crystallography (2006). Vol. F, ch. 18.3, pp. 382-392

Reference for nucleic acid: Parkinson, G.N., *et. al.*, New parameters for the refinement of nucleic acid containing structures. Acta Cryst., D52:57–64, 1996

Polymer Chemical Geometry: Overall Structure and Individual Residues

- Overall review of Clashscore, Ramchandran and Rotamer outliers
- In sequence graphical display, green, yellow, orange and red color coding indicates the fraction of residues with 0, 1, 2, ≥3 chemical geometry outliers, respectively
- Gray for unobserved residues
- Red diamond indicates poor fit to map





PDB: 9CIP

Ligand Chemical Geometry

- PDB validation focuses on Ligand Of Interest (LOI) designated by authors or potential LOI with MW > 250 Da
- Agreement with known chemistry in Cambridge Structural Database (CSD) of small molecule crystal structures
  - Bond Lengths: RMSZ, # |Z|>2 Bond Angles: RMSZ, # |Z|>2
  - Analyses of Chirality, Torsions, Rings
- 2D graphical depiction for geometrical metrics
  - Green: within normal range
  - Magenta: statistical outlier
  - Gray: not applicable, or insufficient chemical reference data to assess

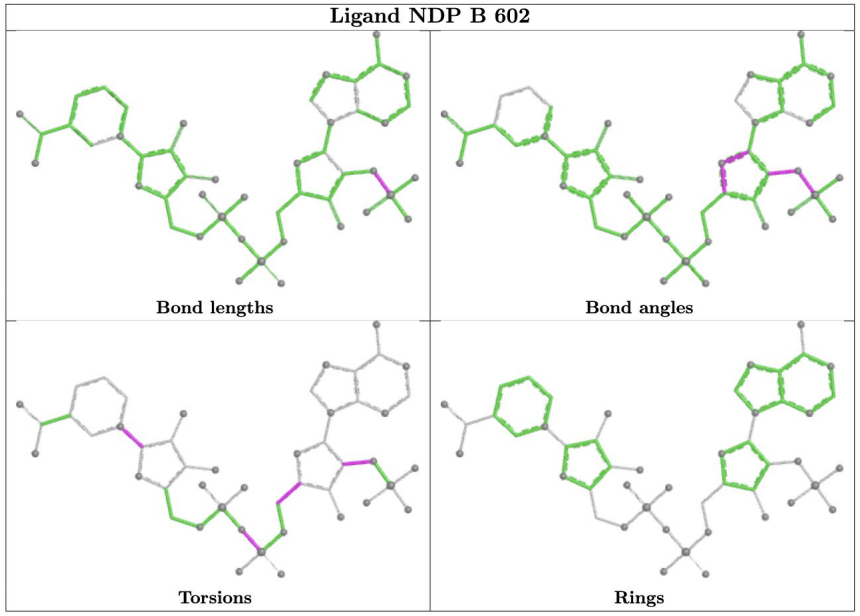
Summary

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	NDP	B	602	-	47,52,52	0.71	1 (2%)	61,80,80	0.95	2 (3%)

List of component outliers

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	602	NDP	P2B-O2B	2.03	1.63	1.59

Graphical depiction of outliers



PDB: 9BDJ

Overall Experimental Data Assessment

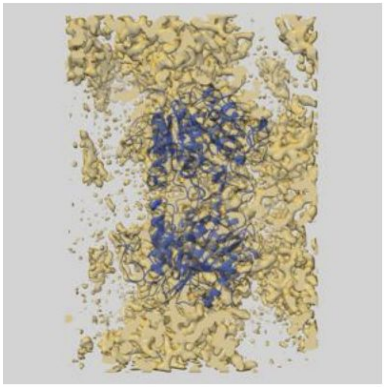
- Symmetry and Unit Cell
- Resolution limit (Å)
- Diffraction data summary: completeness (%), Rmerge, I/σ
- R/Rfree etc

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α, β, γ	31.24Å 57.87Å 59.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.93 – 3.20 28.93 – 3.20	Depositor EDS
% Data completeness (in resolution range)	95.6 (28.93-3.20) 95.4 (28.93-3.20)	Depositor EDS
R_{merge}	0.69	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.99 (at 3.18Å)	Xtriage
Refinement program	PHENIX 1.21.1_5286	Depositor
R, R_{free}	0.268 , 0.311 0.286 , 0.343	Depositor DCC
R_{free} test set	1765 reflections (10.58%)	wwPDB-VP
Wilson B-factor (Å^2)	85.0	Xtriage
Anisotropy	0.779	Xtriage

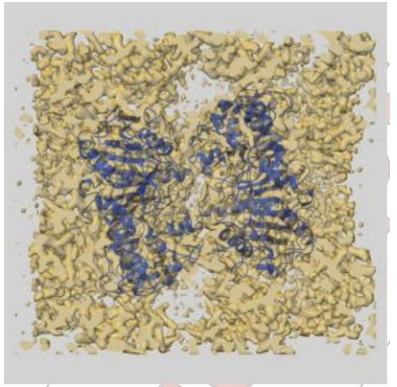
PDB: 9DKZ

Map *vs*. Atomic Model: Visualization

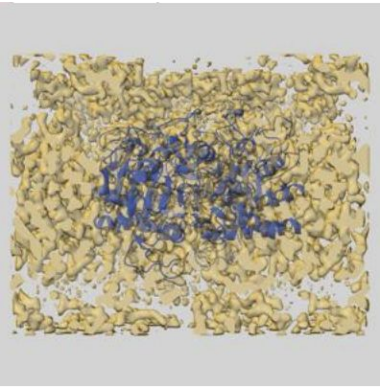
- Projection views of the Map (yellow, at author-selected contour)
- Ribbon representation of the Atomic Coordinates (blue)
- Regions with poor fitting to the map indicate insufficient experimental support



X-axis

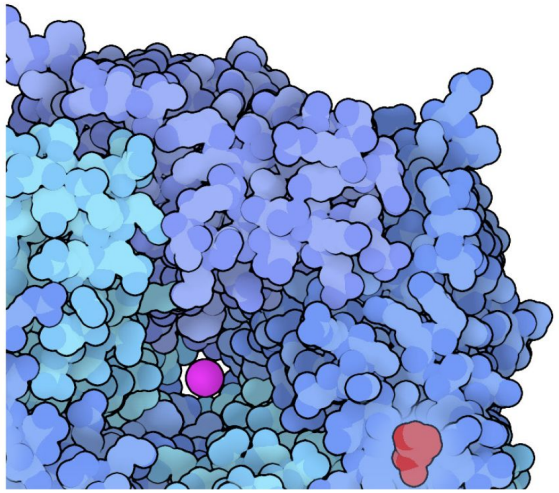


Y-axis



Z-axis

PDB: 9CIP



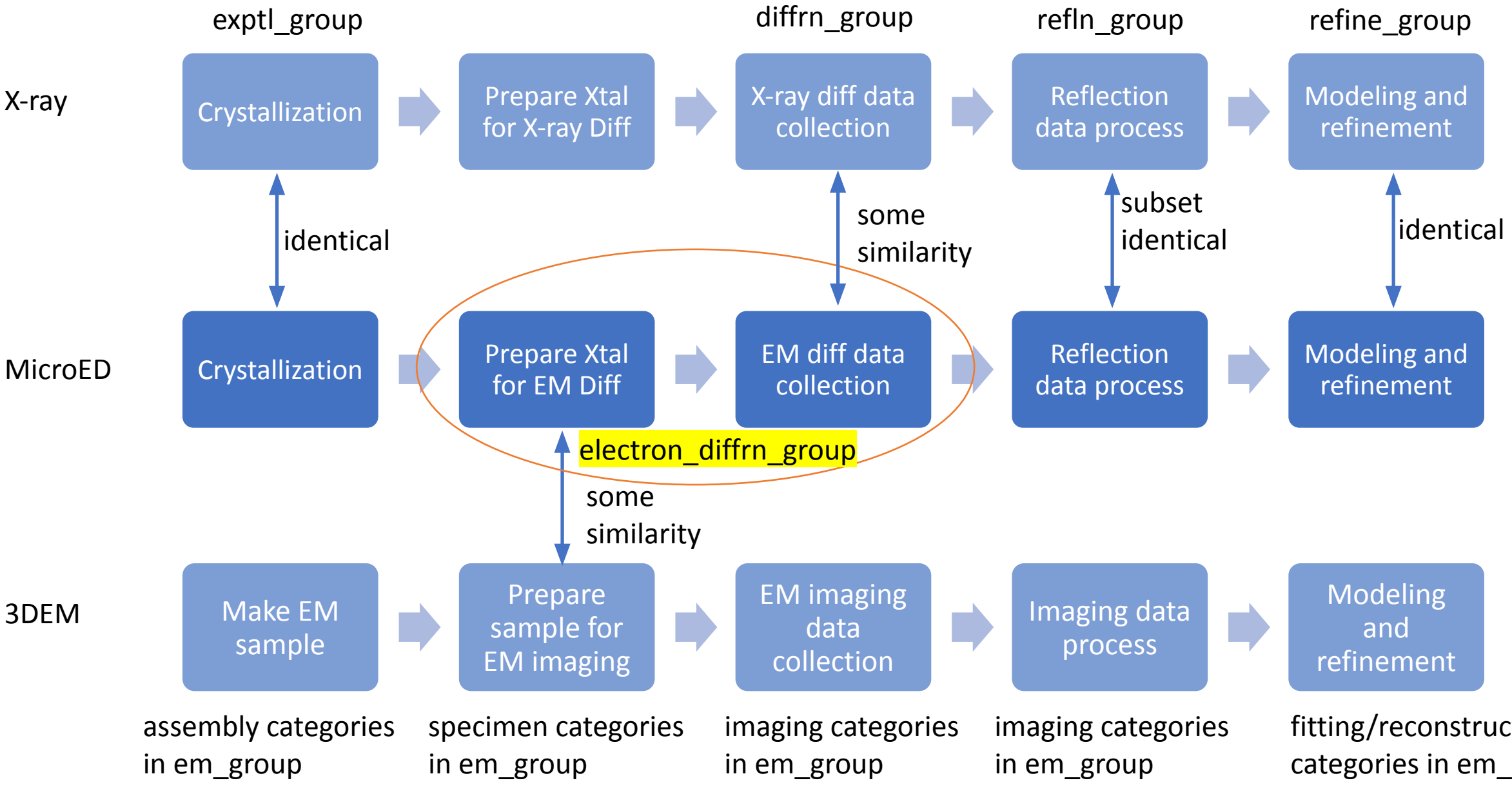
Improving 3D ED/MicroED Data Model for Future Deposition and Validation

Community Engagement and Feedback

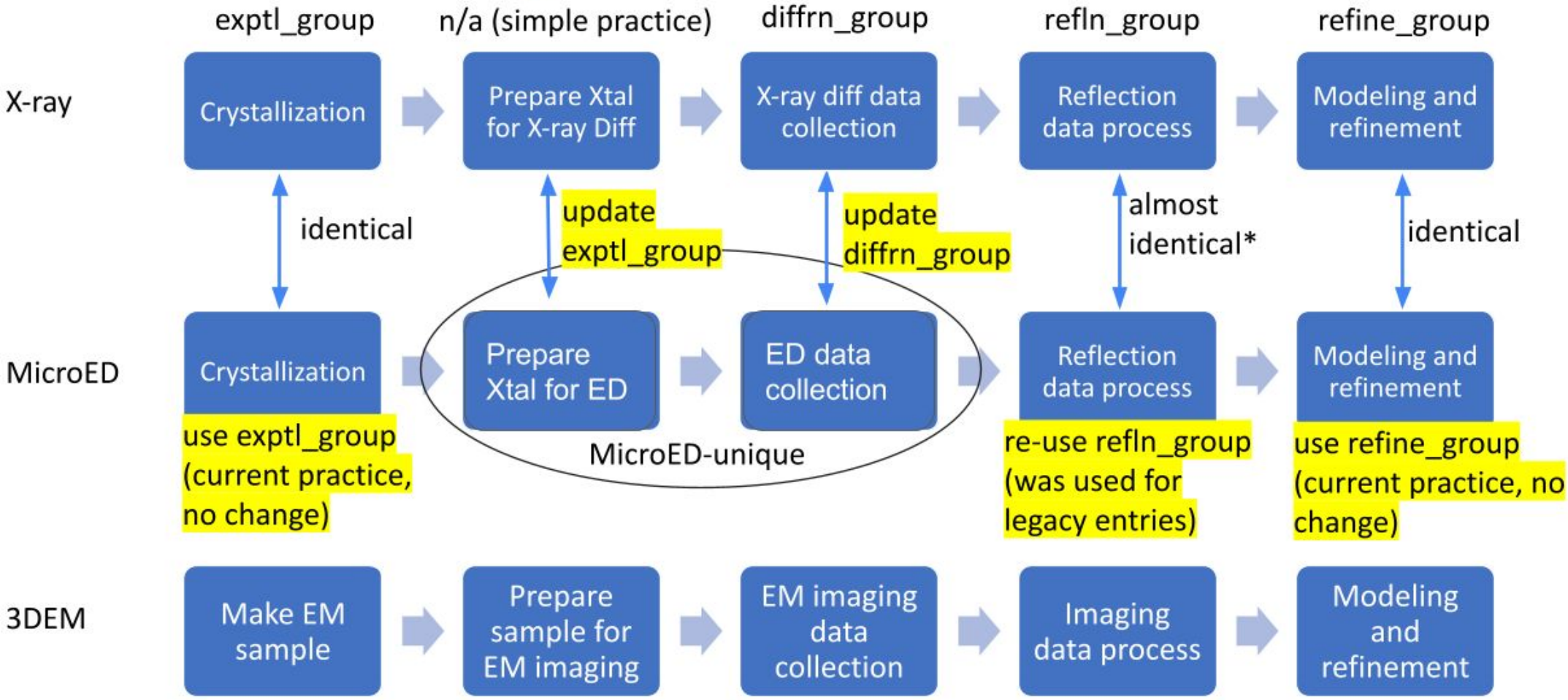
- Stakeholders were consulted since 2023:
  - 18 people including 7 from US, 7 from Europe, and 4 from Japan
  - Including researchers, institute scientists, and software team leads
- Feedback also collected from the depositors, and at the meetings such as 2024 3D ED/MicroED Symposium
- Data models and updates have been sent to the stakeholders for review

Feedbacks and Recommendations

- 3D ED vs MicroED?
- 3D ED/MicroED map to be handled as derived data, “automatically generate map from uploaded MTZ”.
- Structure factors should be collected and used for validation
- Do not ask for data that do not apply to diffraction, e.g. defocus, staining.
- Differentiate between 3D ED/MicroED and 2DEC, e.g. continuous rotation procedure description and microcrystal info should be collected
- Diffraction data process and software info resembles that of X-ray, “we recommend creating a new template derived from the deposition system for XRD” (XRD for X-Ray Diffraction)


exptl_group
diffrn_group
refln_group
refine_group
X-ray
Crystallization
Prepare Xtal for X-ray Diff
X-ray diff data collection
Reflection data process
Modeling and refinement
identical
some similarity
subset identical
identical
MicroED
Crystallization
Prepare Xtal for EM Diff
EM diff data collection
Reflection data process
Modeling and refinement
electron_diffrn_group
some similarity
3DEM
Make EM sample
Prepare sample for EM imaging
EM imaging data collection
Imaging data process
Modeling and refinement
assembly categories in em_group
specimen categories in em_group
imaging categories in em_group
imaging categories in em_group
fitting/reconstruction categories in em_group

MicroED vs X-ray and 3DEM mmCIF Data Model



mmCIF Data Model Summary

- Existing mmCIF data groups to be reused
  - exptl_group
    - Add **pdbx_exptl_subtype** category to differentiate between sub-methods such as 2DEC and MicroED
    - Add **pdbx_exptl_crystal_process** category to record microcrystal production, which can be used for both 3D ED/MicroED and BioXFEL
  - diffrn_group
    - Update **diffrn_measurement** category with IUCr-approved Core cif ED extension
    - Add **pdbx_diffrn_ed** category to record 3D ED/MicroED-specific data in the new
  - refln_group
  - refine_group
  - xfel_group
    - re-used for emerging serial ED. Despite the name, the xfel group was designed to handle serial crystallography applicable to both X-ray and 3D ED/MicroED
- Details and examples available for review at *https://github.com/wwPDB/microed-extension*

Dictionary Definition Example for diffrn_measurement

Data items in the DIFFRN_MEASUREMENT category record details about the device used to orient and/or position the crystal during data measurement and the manner in which the diffraction data were measured.

- **_diffrn_measurement.pdbx_angle_start** : The starting angle for crystal rotation.
- **_diffrn_measurement.pdbx_angle_end** : The ending angle for crystal rotation.
- **_diffrn_measurement.pdbx_rotation_rate** : The speed of crystal rotation.
- **_diffrn_measurement.pdbx_exposure_time_per_image** : The exposure time per image for the crystal and detector/camera.
- **_diffrn_measurement.rotation_mode** : Originated from IUCr Electron Diffraction CIF extension. Code describing the technique used to sample as completely as possible the accessible range of reciprocal space. Rotation mode indicates the detector records diffracted intensities continuously while the sample is rotated. Stepwise model indicates the detector records diffracted intensities while the sample is stationary. Enumeration:
  a. rotation
  b. stepwise

……

Example of diffrn_measurement

```
#
_diffrn_measurement.diffrn_id                          1
_diffrn_measurement.details                        1
_diffrn_measurement.method                         ?
_diffrn_measurement.method_precession                  N
_diffrn_measurement.pdbx_angle_start               0
_diffrn_measurement.pdbx_angle_end                 20
_diffrn_measurement.pdbx_exposure_time_per_image   1
_diffrn_measurement.pdbx_rotation_rate                 0.0476
_diffrn_measurement.rotation_mode                  rotation
_diffrn_measurement.sample_tracking                ?
_diffrn_measurement.sample_tracking_method             ?
#
```

Example of pdbx_exptl_crystal_process

```
#
_pdbx_exptl_crystal_process.id                          1
_pdbx_exptl_crystal_process.crystal_id                  1
_pdbx_exptl_crystal_process.microcrystal_method         "FIB milling"
_pdbx_exptl_crystal_process.microcrystal_instrument     "Helios Hydra 5 CX dual-beam plasma FIB/SEM (Thermo Fisher
Scientific)"
_pdbx_exptl_crystal_process.microcrystal_min_dim        0.3
_pdbx_exptl_crystal_process.microcrystal_max_dim        5
_pdbx_exptl_crystal_process.microcrystal_description    lamella
_pdbx_exptl_crystal_process.details                     "stepwise protocol to an optimal thickness of
approximately 300 nm using a 30 kV Argon plasma ion beam"
#
```

Example of pdbx_diffrn_ed

```
#
_pdbx_diffrn_ed.id                          1
_pdbx_diffrn_ed.diffrn_id                   1
_pdbx_diffrn_ed.beam_diameter_sample_plane   3.5
_pdbx_diffrn_ed.camera_length               1402
_pdbx_diffrn_ed.c2_aperture_diameter        50
_pdbx_diffrn_ed.details                     "The microscope was aligned for low flux density using spot size 11,
and gun lens setting 8 for a less bright but more coherent illumination. The energy filter was tuned to pass
electrons with energy losses less than 10 eV, with the zero-loss peak centered in defocused diffraction"
_pdbx_diffrn_ed.fluence_accumulated         0.84
_pdbx_diffrn_ed.fluence_rate                0.002
_pdbx_diffrn_ed.electron_source             "FIELD EMISSION GUN"
_pdbx_diffrn_ed.energyfilter_name           "TFS Selectris"
_pdbx_diffrn_ed.energyfilter_upper          ?
_pdbx_diffrn_ed.energyfilter_lower          ?
_pdbx_diffrn_ed.recording_mode              "Electron Counting"
_pdbx_diffrn_ed.sa_aperture_diameter        150
#
```

Future 3D ED/MicroED Deposition and Validation

- Data Model to be implemented in 2026
- Deposition interface will resemble X-ray's, under MX framework
  - SF file is sufficient. No map.
  - Shared page templates with X-ray, Neutron diffraction
  - Dedicated page for 3D ED/MicroED
- Validation process will resemble X-ray's, under MX framework
- Supports from other programs, e.g.
  - Working with the Phenix team
  - pdb_extract to incorporate 3D ED/MicroED deposition preparation and log parsing

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

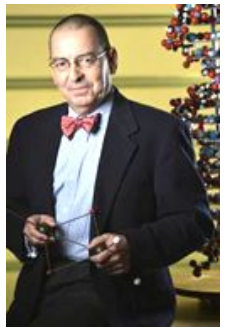
UC San Diego
SDSC SAN DIEGO SUPERCOMPUTER CENTER

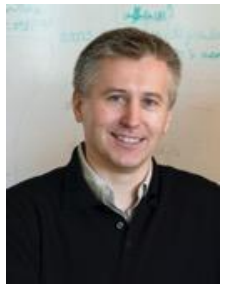
UCSF
University of California San Francisco

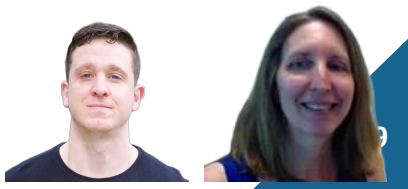
Follow us

WORLDWIDE PDB
PROTEIN DATA BANK

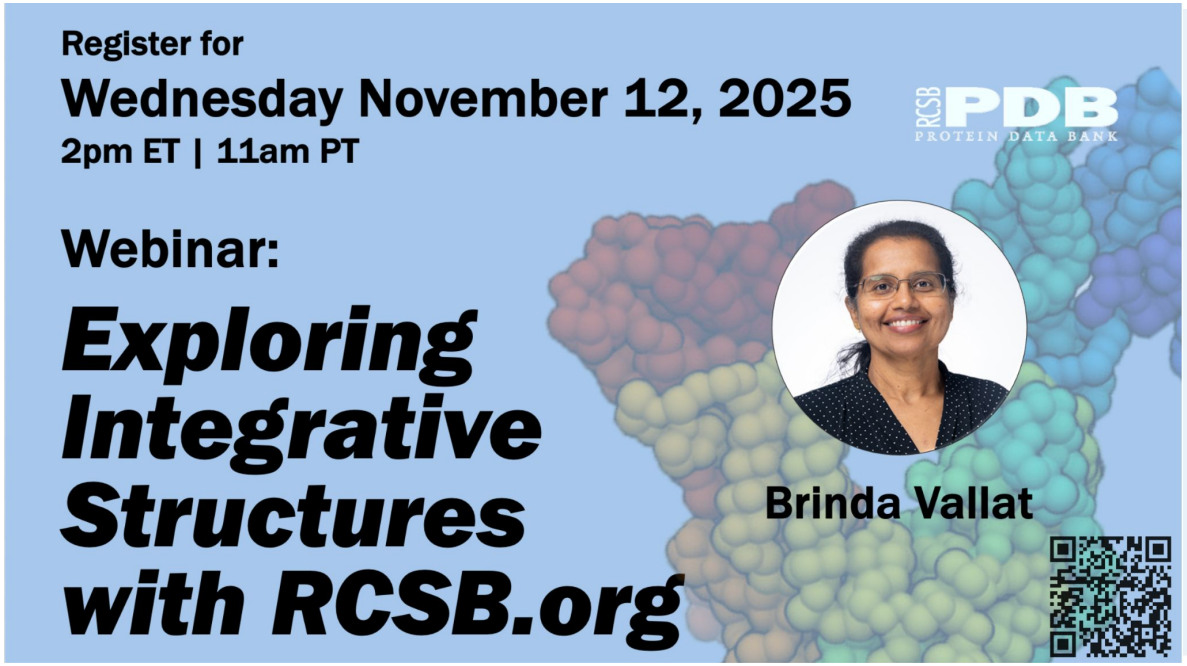
Member of the Worldwide Protein Data Bank (wwPDB; **wwpdb.org**)

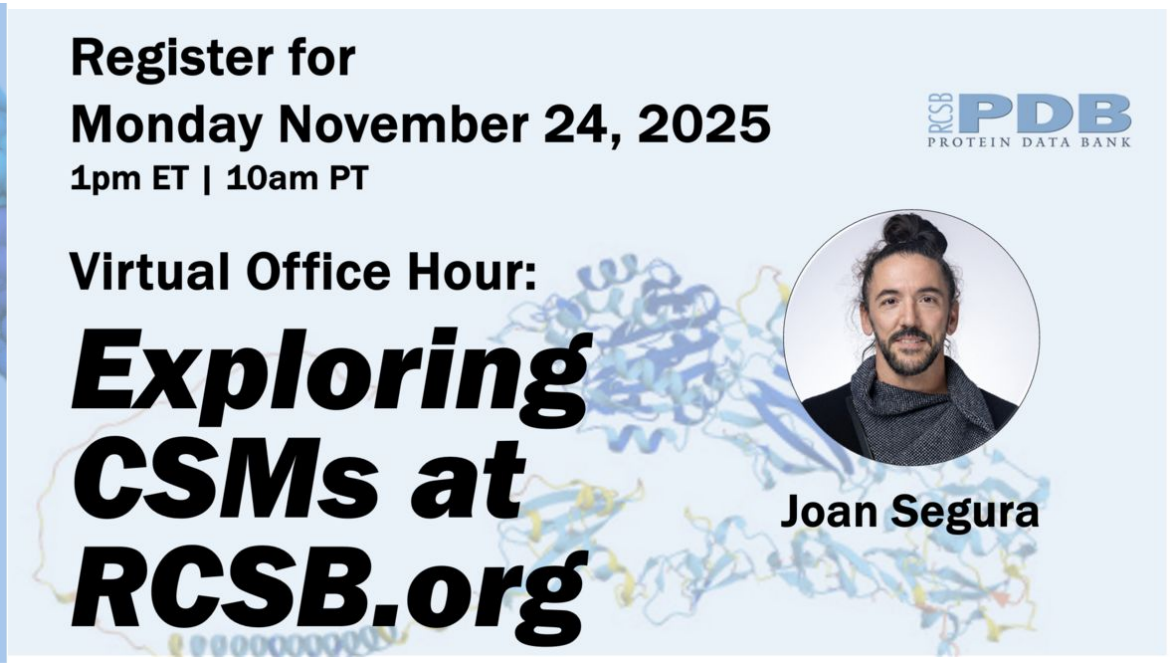






Register for Upcoming Events





Recordings will be posted at PDB-101.rcsb.org

Announcement: Ligand Expo To Be Retired in 2025

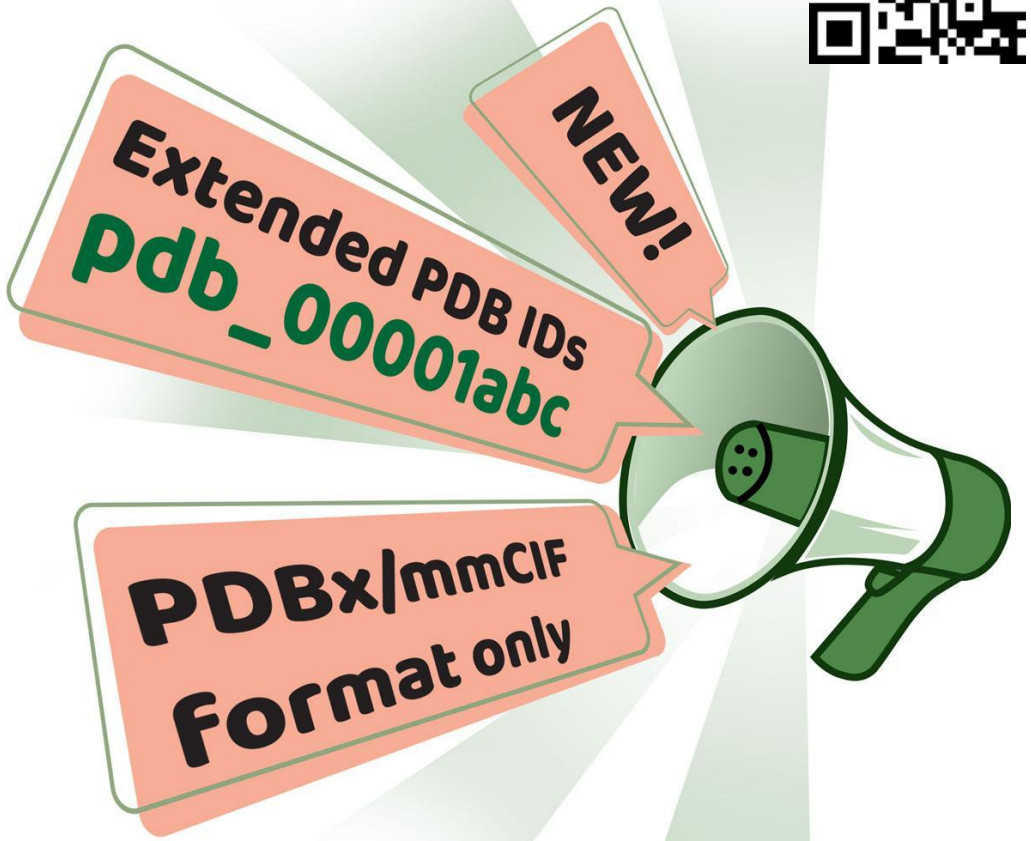
Users should transition to RCSB PDB and wwPDB services as soon as possible

Transition to Extended PDB IDs and PDBx/mmCIF File Format

By 2028
4-character PDB IDs (*e.g.* **1abc**) will be fully allocated.
After that, all new entries will be assigned only extended PDB IDs.

Extended PDB ID Format
8 alphanumeric characters with prefix *pdb_*
e.g. **pdb_00001abc**

Entries with extended PDB IDs will be available only in PDBx/mmCIF format.
All PDB users, including software developers, and journal editors must transition to this format.
Start using extended IDs and PDBx/mmCIF data files today



WORLDWIDE PDB
PROTEIN DATA BANK

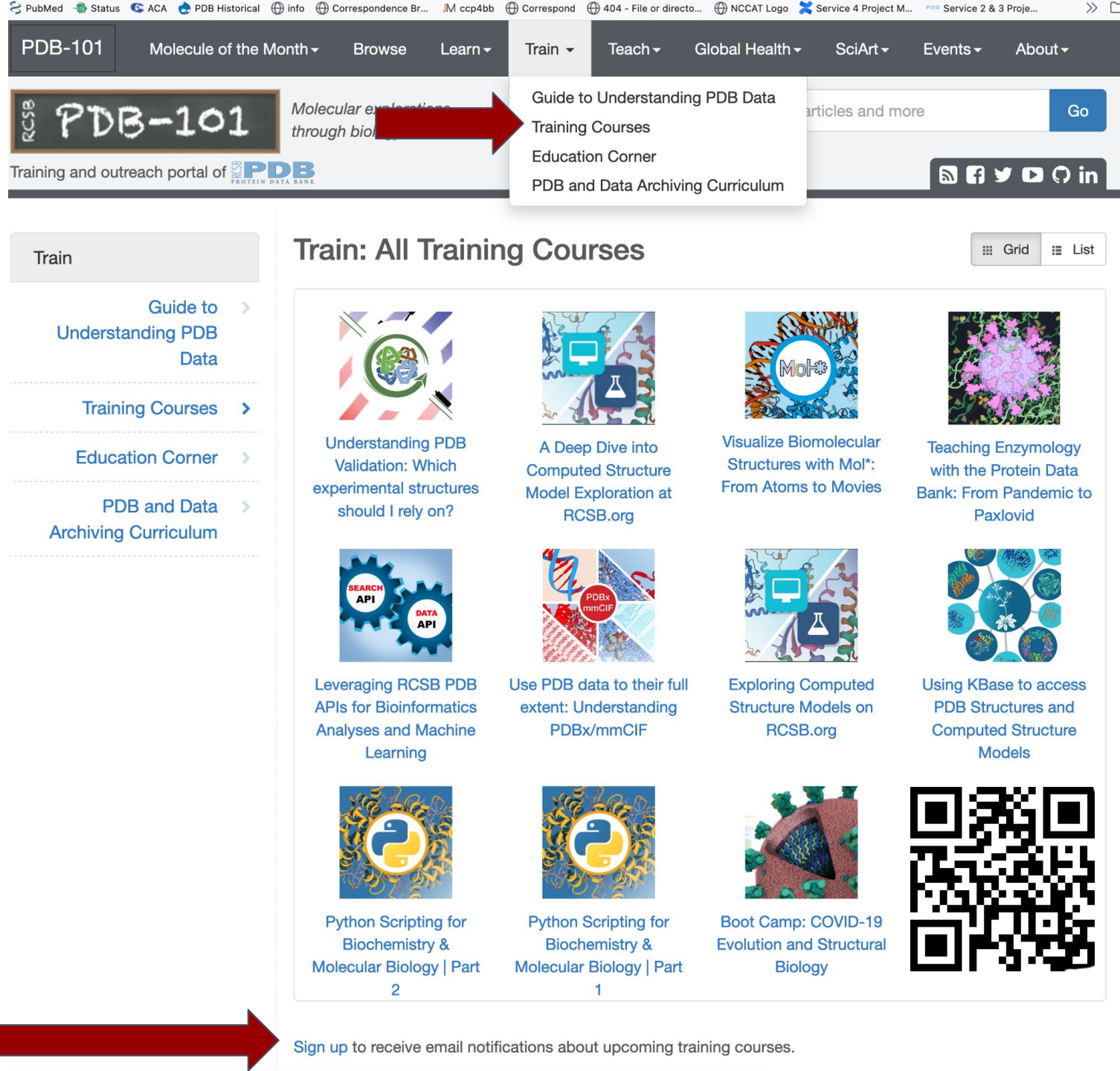
wwpdb.org/
documentation/new-format-for-pdb-ids

Course Materials

Course recordings and presentations may be published at PDB-101.rcsb.org later this summer.

They will not be emailed separately.

Sign up to receive emails about upcoming training events



Questions?