PDBx/mmCIF: Foundation for FAIR Data Access in the Protein Data Bank

Brinda Vallat

RCSB Protein Data Bank and Institute for Quantitative Biomedicine

brinda.vallat@rcsb.org

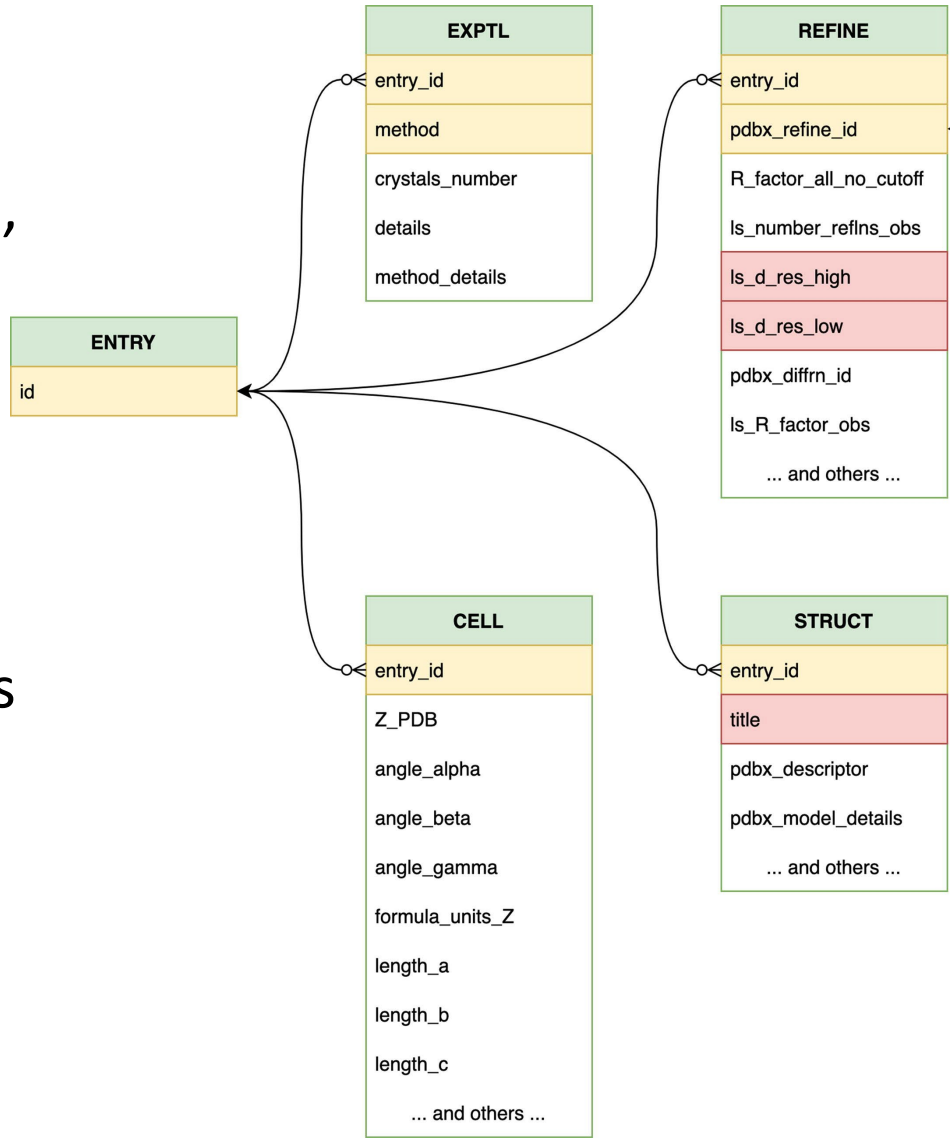
RCSB.org

RCSB PDB
PROTEIN DATA BANK

PDBx/mmCIF Data Standard

RUTGERS–NEW BRUNSWICK
Artificial Intelligence
and Data Science Collaboratory

- Data standard: Technical specifications describing the logical organization of data and metadata
  - Pre-requisite for automated data collection, processing, and dissemination
- PDBx/mmCIF is the data standard for representing macromolecular structures archived in the PDB
- Origin of PDBx/mmCIF
  - CIF: By IUCr for small-molecule diffraction studies
  - mmCIF: Macromolecular X-ray diffraction investigations
  - PDBx/mmCIF: Other experimental methods in the PDB (*e.g.*, NMR, 3DEM)
- Based on a dictionary definition language
  - primary data types (*e.g.*, integers, real numbers, text)
  - controlled vocabularies and boundary conditions
  - linking of data items together to express relationships



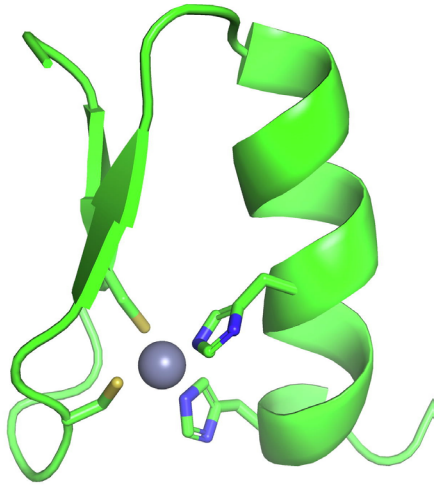
RCSB.org

mmcif.wwpdb.org

PDBx/mmCIF: Features and Advantages

RUTGERS–NEW BRUNSWICK
Artificial Intelligence
and Data Science Collaboratory

```
loop_
_struct_conn.id
_struct_conn.conn_type_id
_struct_conn.ptnr1_label_asym_id
_struct_conn.ptnr1_label_comp_id
_struct_conn.ptnr1_label_seq_id
_struct_conn.ptnr1_label_atom_id
_struct_conn.ptnr1_symmetry
_struct_conn.ptnr2_label_asym_id
_struct_conn.ptnr2_label_comp_id
_struct_conn.ptnr2_label_seq_id
_struct_conn.ptnr2_label_atom_id
_struct_conn.ptnr1_auth_asym_id
_struct_conn.ptnr1_auth_comp_id
_struct_conn.ptnr1_auth_seq_id
_struct_conn.ptnr2_auth_asym_id
_struct_conn.ptnr2_auth_comp_id
_struct_conn.ptnr2_auth_seq_id
_struct_conn.ptnr2_symmetry
_struct_conn.pdbx_dist_value
_struct_conn.pdbx_value_order
metalc1 metalc D ZN . ZN 1_555 C HIS 25 NE2 C ZN 201 C HIS 25 1_555 2.138
metalc2 metalc D ZN . ZN 1_555 C CYS 7  SG  C ZN 201 C CYS 7  1_555 2.232
metalc3 metalc D ZN . ZN 1_555 C CYS 12 SG  C ZN 201 C CYS 12 1_555 2.440
metalc4 metalc D ZN . ZN 1_555 C HIS 29 NE2 C ZN 201 C HIS 29 1_555 1.876
```



```
_struct.entry_id                    '4HHB'
_struct.title
; THE CRYSTAL STRUCTURE OF HUMAN
DEOXYHAEMOGLOBIN AT 1.74 ANGSTROMS RESOLUTION
;
_struct.pdbx_descriptor             'HEMOGLOBIN (DEOXY)'
_struct.pdbx_model_details          ?
_struct.pdbx_CASP_flag              ?
_struct.pdbx_model_type_details     ?
```

```
_exptl.entry_id          4HHB
_exptl.method            'X-RAY DIFFRACTION'
_exptl.crystals_number   ?
```

RCSB.org

Standardization

Format, Syntax, Data Standards

Community Support

Extensibility

Support for evolving science
ModelCIF, IHMCIF

PDBx/mmCIF

Automation

Software amenable Tools for data management

PDB

mmcif.wwpdb.org

FAIR Data

Findable, Accessible, Interoperable, Reusable

Westbrook JD et al. *J. Mol. Biol.* 2022, 434: 167599
Westbrook J et al. *International tables for crystallography G*. 2005, 195-198