No more ~~PDB~~ format files! Welcome to PDBx/mmCIF ecosystem

Kumaran Baskaran, Jonathan R. Wedell, Hongyang Yao, Dimitri Maziuk, Hamid R. Eghbalnia, Michael M. Gryk, and Jeffrey C. Hoch

Department of Molecular Biology and Biophysics, UConn Health, 263 Farmington Avenue, Farmington, CT 06030, USA

wwPDB data deposition

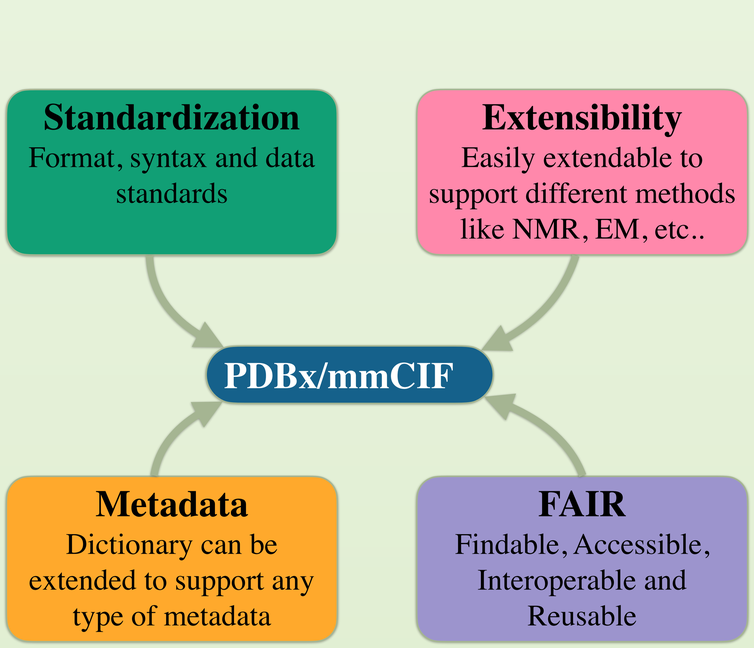
Model File
PDBx/mmCIF
oordinate data and meta data

Experimental File
NMR-STAR/NEF
Chemical shifts, restraints, peak list and other experimal data

wwPDB accept data coming from structure determination studies whereas BMRB accepts all kids of NMR data including raw experimental data.

PDBx/mmCIF Format

PDBx(Protein Data Bank exchange)/mmCIF(macromolecular Crystallographic Information File) is the master archival format of the PDB since 2014



Limitations of legacy PDB format

- Fixed size columns
- Column informations hard coded
- Limited scope for meta data
- Can't be used represent large molecules like viruses
- No support to define biological assemblies
- No way define ligand interaction
- No way to define ensemble properties
- No way to represent conformers

Scan here for the PDF version of this poster

Metadata

A

```
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
. . .
#
_cell.entry_id              1ZAA
_cell.length_a              45.400
_cell.length_b              56.200
_cell.length_c              130.800
_cell.angle_alpha           90.00
_cell.angle_beta            90.00
_cell.angle_gamma           90.00
_cell.Z_PDB                 8
#
_atom_sites.entry_id                    1ZAA
_atom_sites.fract_transf_matrix[1][1]   0.022026
_atom_sites.fract_transf_matrix[1][2]   0.000000
_atom_sites.fract_transf_matrix[1][3]   0.000000
. . .
```

B

```
LINK         ZN    ZN C 201                NE2 HIS C  25     1555   1555  2.14
LINK         ZN    ZN C 201                SG  CYS C   7     1555   1555  2.23
LINK         ZN    ZN C 201                SG  CYS C  12     1555   1555  2.44
LINK         ZN    ZN C 201                NE2 HIS C  29     1555   1555  1.88
. . .

CRYST1   45.400   56.200  130.800  90.00  90.00  90.00 C 2 2 21       8

SCALE1      0.022026  0.000000  0.000000        0.00000
SCALE2      0.000000  0.017794  0.000000        0.00000
SCALE3      0.000000  0.000000  0.007645        0.00000
```

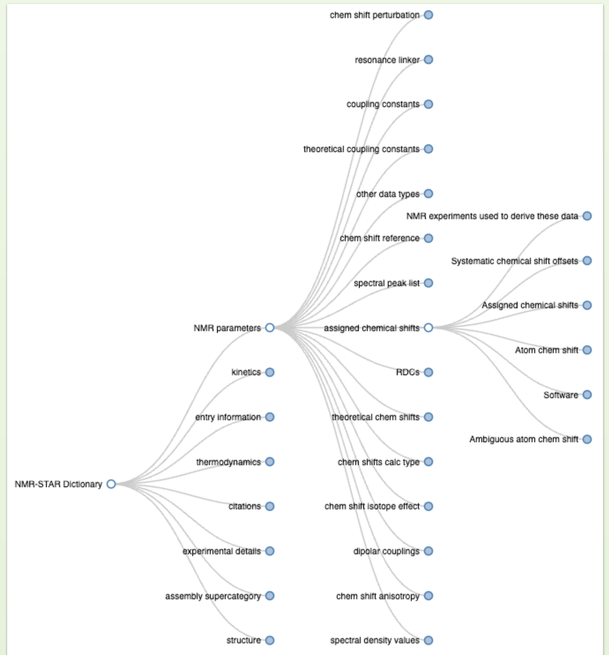
(A) Partial PDBx/mmCIF file for PDB ID 1ZAA. N.B.: Every data value has a key and multiple rows of data may be described in a table. The yellow highlighting describes the category and attributes. For the _struct_conn category, green depicts the residue numbers and cyan the component type. (B) Equivalent metadata records in legacy PDB format. Similar color coding depicts the mapping between category keys and record names as in (A), with LINK records highlighting the residue number and cyan the chemical component type. Inset figure, one of the zinc finger domains in 1ZAA depicting the sidechains that interact with the bound zinc ion codified in (A) and (B).

PDBx/mmCIF Resource

- Dictionary resource https://mmcif.wwpdb.org/
- Virtual crash course https://pdb101.rcsb.org/train/training-events/mmcif
- Create your mmCIF https://pdb-extract.wwpdb.org/
- Software resource https://mmcif.wwpdb.org/docs/software-resources.html

NMR-STAR data model

NMR-STAR is the archival format for NMR data at wwPDB and BMRB.



NMR-STAR Resource

- Dictionary resource https://bmrb.io/dictionary/
- Parser https://pypi.org/project/pynmrstar/
- Data visualization https://pypi.org/project/pybmrb/

Why PDBx/mmCIF and NMR-STAR?

Integrative and hybrid methods have become more common and model files are required to carry the necessary meta data from different experiments in addition to the coordinate information. The complexity of the information like multiple conformations, ligand interactions, screening studies require proper data modeling with can only be possible using PDBx/mmCIF and NMR-STAR.

Join Us!

If you are a software developer and your software is capable of generating model data and NMR data then join us as an NMR mmCIF working group member and contribute to the data modeling and software development.

Contact: help@bmrb.io

1. Westbrook, *et al*., (2022) PDBx/mmCIF Ecosystem: Foundational Semantic Tools for Structural Biology *JMB* 434: 167599 doi: 10.1016/j.jmb.2022.167599

2. Ulrich, *et al*. NMR-STAR: comprehensive ontology for representing, archiving and exchanging data from nuclear magnetic resonance spectroscopic experiments. *J Biomol NMR* **73**, 5–9 (2019). doi: 10.1007/s10858-018-0220-3