

## MS12-P02 | ENHANCING SMALL MOLECULE INFORMATION IN THE PROTEIN DATA BANK IN EUROPE

Pravda, Lukáš (GBR); PDBe (PDBe, Hinxton, GBR)

The Protein Data Bank (PDB) provides a wealth of data on biomacromolecules and their complexes with small molecules/ligands. Information about each unique small molecule is provided through the PDB chemical component dictionary (CCD), including chemical composition, connectivity, model and idealized coordinates, and descriptors such as InChI or SMILES. Entries in the CCD and PDB, however, lack information on the role of bound molecules, e.g. as a cofactor, drug or inhibitor. Information on structural similarity, e.g. common scaffolds or fragments, and data on common interactions formed between small molecules and macromolecules are also difficult to access.

As part of the data enrichment process in the Protein Data Bank in Europe [1] we have extended small molecule information to include scaffolds, fragments, physicochemical properties, and 2D coordinates for collision-free depictions. We also provide cofactor annotation within a biological context and mapping of unique small molecules in the CCD to other common databases using UniChem. Interaction information between macromolecules and small molecules is already available through the new aggregated views for proteins [2] which are part of the Protein Data Bank in Europe Knowledge Base (PDBe-KB; [pdbe-kb.org](http://pdbe-kb.org)) a new, community-driven resource of functional annotations for structure data.

Enrichment of small molecule data provides the basis for a new service presenting aggregated views for small molecules/ligands in the PDB, putting them in their biological context. The service is planned to be introduced at the end of 2019.

[1] PDBe; <https://pdbe.org>

[2] <https://pdbe-kb.org/proteins>