POCK-DB 2026: a standardized protein–ligand binding pocket database integrating ligand-based and geometry-based definitions

Léa Dufay[1], Anne-Claude Camproux[1]

[1]Team: In Silico Pharmacological Profiling, Université Paris Cité, CNRS 8251, INSERM U1133, 35 rue Hélène Brion, 75205, Paris, France

We present Pock-DB 2026, a large-scale database of experimentally observed protein–ligand binding pockets extracted from the Protein Data Bank [1] (PDB; release 02/2026). The aim of this resource is to provide a standardized and reusable representation of binding pockets for chemoinformatics and structure-based drug discovery. To build the database, we developed an automated workflow that combines two complementary estimations of each pocket: a ligand-based pocket estimated by proximity to the bound ligand [2] and a geometry-based pocket detected independently from protein cavity geometry [3].

Each entry in Pock-DB 2026 integrates protein and ligand annotations from RCSB PDB and PDBe [4], local structural environment information, and pocket-level geometric and physicochemical descriptors [2]. The pipeline includes biological assembly processing, structure cleaning, exclusion of non-relevant ligand contexts, and overlap-based validation between ligand-based and geometry-based pocket estimations. This yields a coherent dataset of experimentally resolved binding pockets suitable for large-scale analysis, while preserving both complementary pocket descriptions for each retained complex.

Pock-DB 2026 is intended as a general resource for binding pocket characterization, pocket comparison, descriptor analysis, virtual screening, and machine-learning workflows. By making experimentally resolved binding pockets and their two pocket estimations available in a homogeneous format through both downloadable files and an interactive web interface, the database facilitates the exploration of binding-site diversity and supports data-driven chemoinformatics applications.

Bibliography :

[1] Berman H.M.; et al. Nucl. Acids Res., 2000, 28, 235-242.
[2] Hussein H.A.; et al. Nucl. Acids Res., 2015, 43, 436-442.
[3] Le Guilloux V.; et al. Bioinformatics, 2009, 10, 168.
[4] Consortium. Nucl. Acids Res., 2022, 50, D534-D542