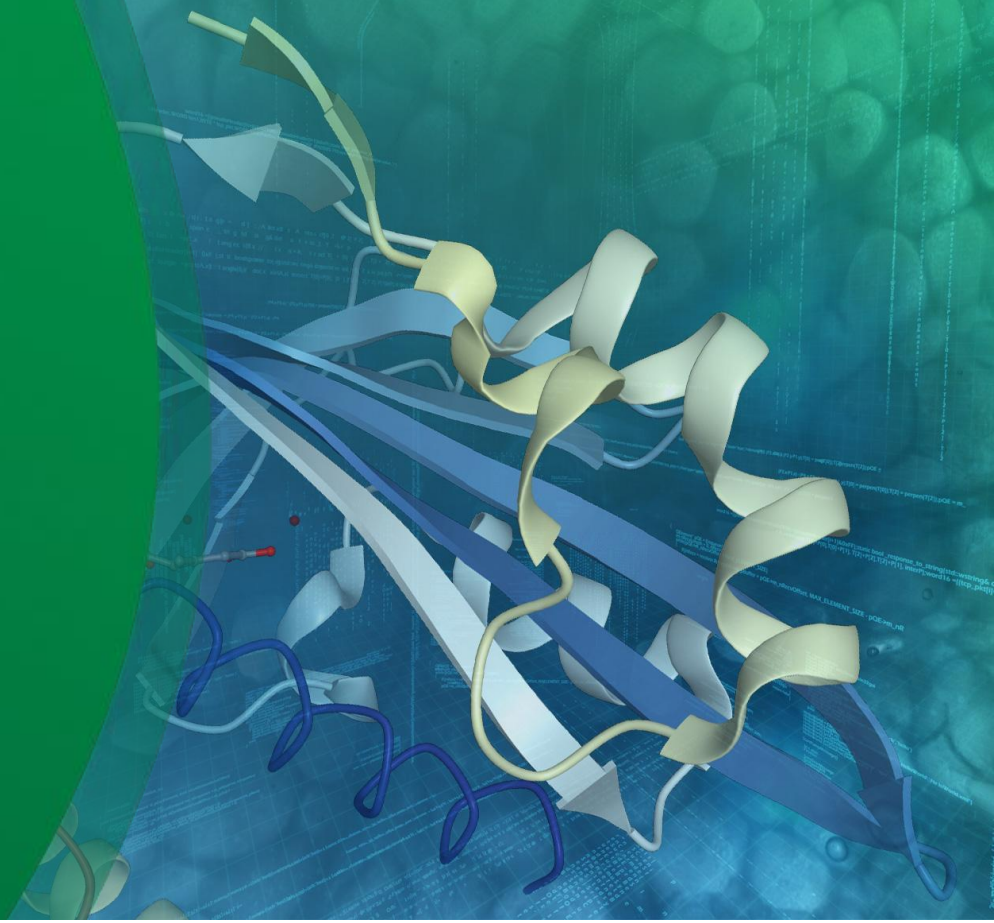
Introduction to the Protein Data Bank



Dr. Deborah Harrus

PDBe Archive Project Leader

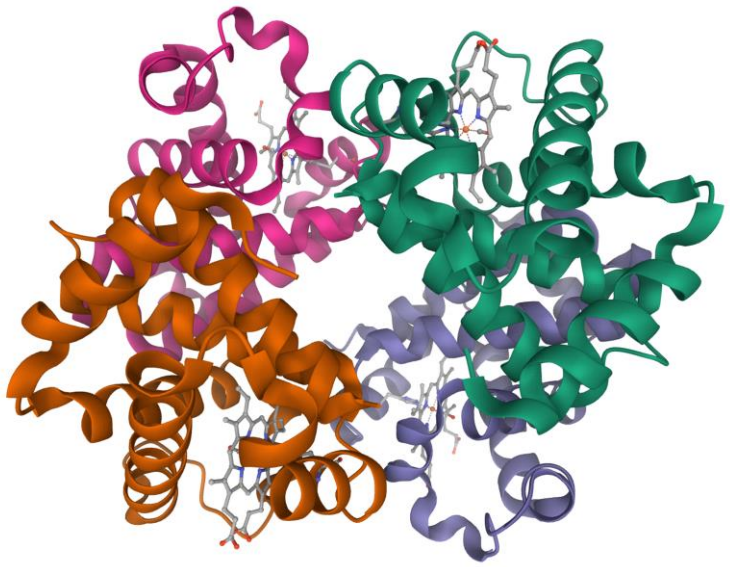
PDBe
Protein Data Bank in Europe

EMBL-EBI

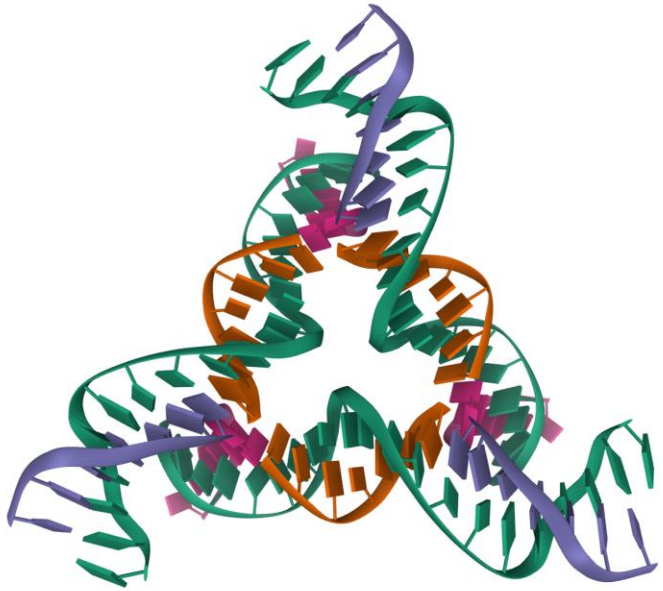
CCDC Virtual Workshop

5th of November 2024

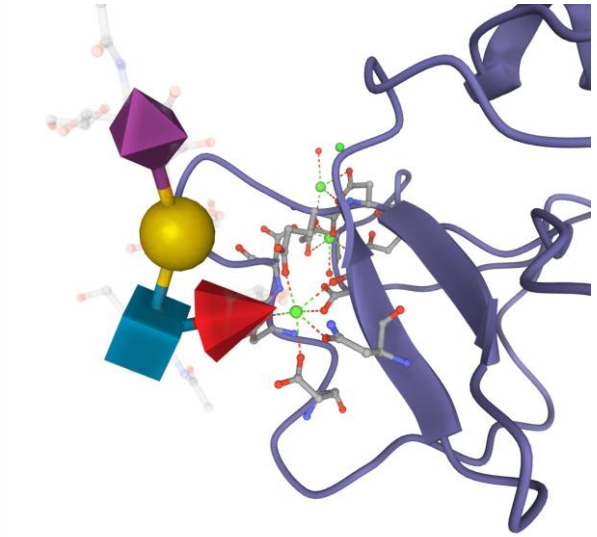
The Protein Data Bank (PDB) is an archive of **experimentally determined** 3-dimensional structures of biological macromolecules



Proteins



DNA/RNA



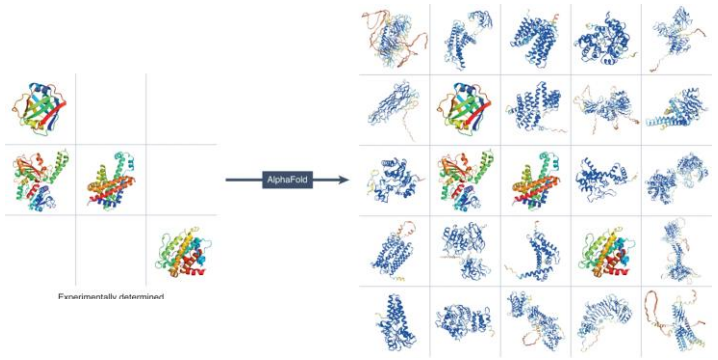
Bound ligands

EMBL-EBI

Our mission…

Establish a **high-quality**, **up-to-date** and **integrated infrastructure** and innovative **data analysis tools** for **3D macromolecular structures** to support basic and applied **research** and **education** across the sciences

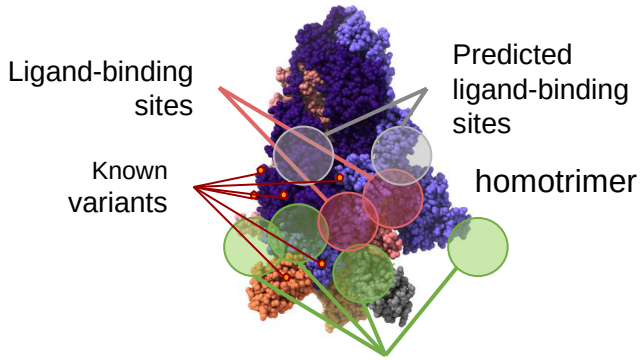
Macromolecular Structure Data



WORLDWIDE PDB
PROTEIN DATA BANK

AlphaFold
Protein Structure Database

Information



Macromolecular interaction interfaces

PDBe-KB
Protein Data Bank in Europe - Knowledge Base

Knowledge

Mutation *X* disrupts function of enzyme *Y*, which causes disease *Z*

EMBL-EBI

How chemical information is stored in PDB

CCD (chemical component dictionary) definitions (reference dictionary)

Ideal coordinates (*Experimental coordinates*)

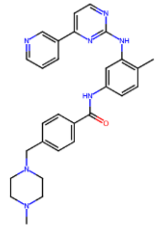
Name, synonyms, formula, weight

Atom-level info: unique ids, element, charge, R vs S

Bond-level info: aromaticity, stereochemistry, bond order

Descriptors: SMILES, InChI, InChIKey

```
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.alt_atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.charge
_chem_comp_atom.pdbx_align
_chem_comp_atom.pdbx_aromatic_flag
_chem_comp_atom.pdbx_leaving_atom_flag
_chem_comp_atom.pdbx_stereo_config
_chem_comp_atom.model_Cartn_x
_chem_comp_atom.model_Cartn_y
_chem_comp_atom.model_Cartn_z
_chem_comp_atom.pdbx_model_Cartn_x_ideal
_chem_comp_atom.pdbx_model_Cartn_y_ideal
_chem_comp_atom.pdbx_model_Cartn_z_ideal
_chem_comp_atom.pdbx_component_atom_id
_chem_comp_atom.pdbx_component_comp_id
_chem_comp_atom.pdbx_ordinal
STI C1 C1 C 0 1 Y N N N N N 16.356 100.406 50.614 2.740 -0.022 8.889 C1 STI 1
STI C6 C6 C 0 1 Y N N N N N 15.218 100.012 51.414 1.914 0.207 7.802 C6 STI 2
```



CCD ID: STI
Mol. Name: Imatinib

https://ftp.ebi.ac.uk/pub/databases/msd/pdbechem_v2/ccd/S/STI/STI.cif

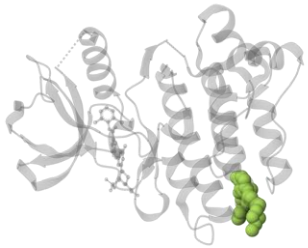
PDB entries (experimental 3D structures)

Experimental coordinates

Number of instances observed

Subject of investigation

```
#
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num
ATOM   1    N  N   . ASP A 1 5   ? -22.963 -26.563 -47.248 1.000 77.262  ? 252 ASP A N   1
ATOM   2    C  CA  . ASP A 1 5   ? -22.572 -26.099 -45.878 1.000 75.723  ? 252 ASP A CA  1
.
.
.
HETATM 5887 C  C1  . STI D 2 .   ? -17.422 -2.386  -7.775  1.000 44.873  ? 601 STI A C1  1 ?
HETATM 5888 C  C6  . STI D 2 .   ? -16.730 -3.540  -7.470  1.000 45.283  ? 601 STI A C6  1 ?
```



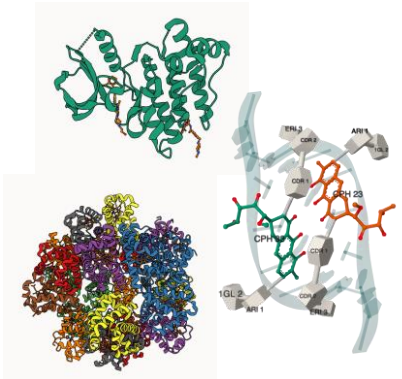
PDB ID: 7n9g
Protein Name: Tyrosine-protein kinase ABL1

https://www.ebi.ac.uk/pdbe/entry-files/7n9g_updated.cif

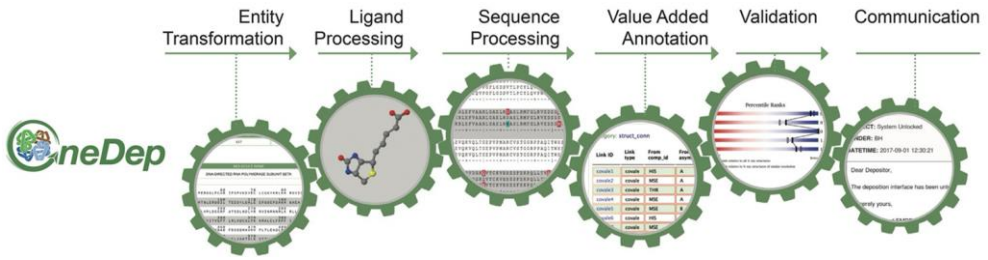
https://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Index/index.html

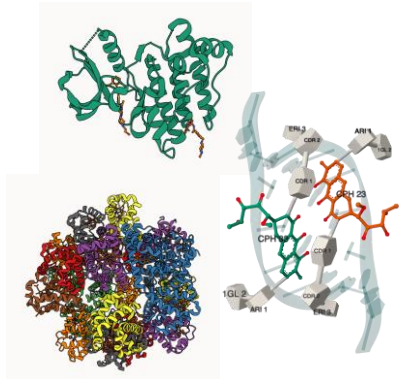
EMBL-EBI

Small molecules annotation

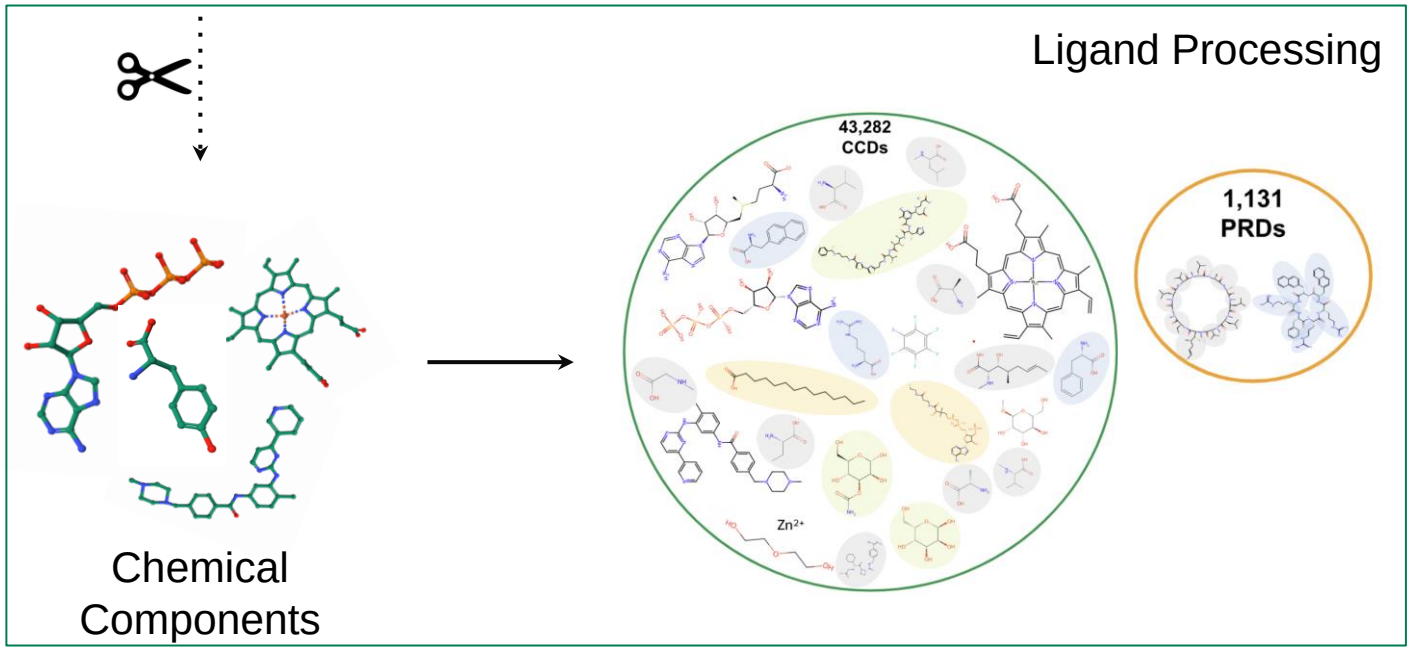


Deposited Structures

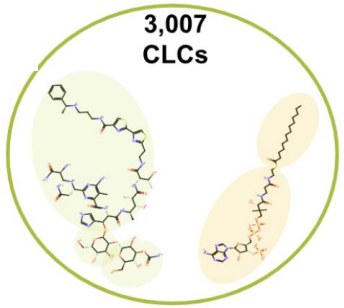




Curated Structures



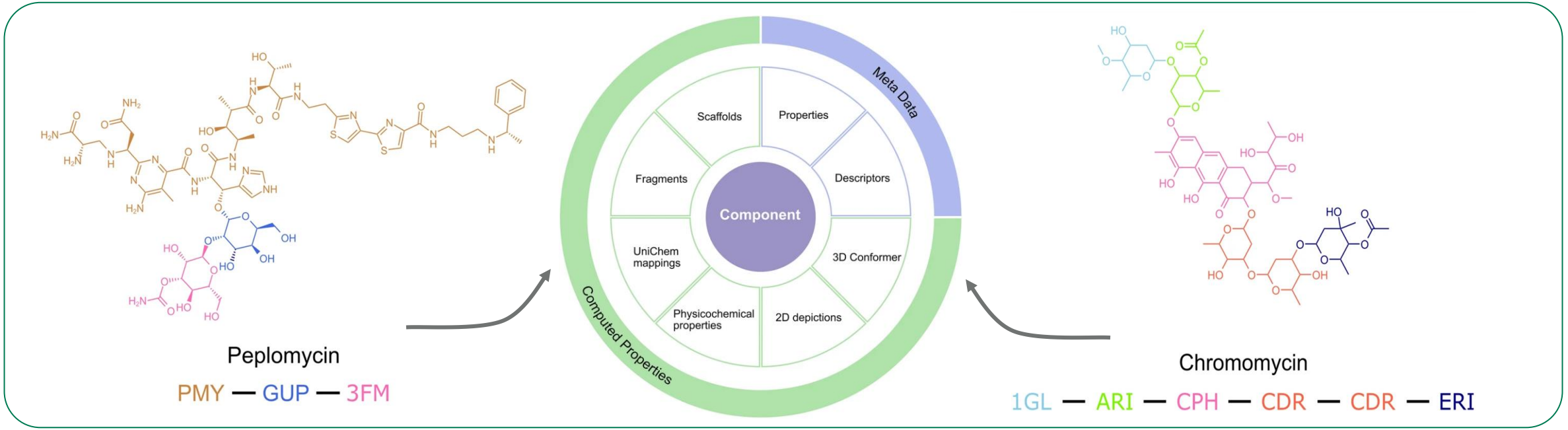
PBDe CCDUtils



EMBL-EBI

CCDUtils: Structural analysis of small molecules in the PDB

Toolbox for chemoinformatic visualization, property calculations, and CLC identification.



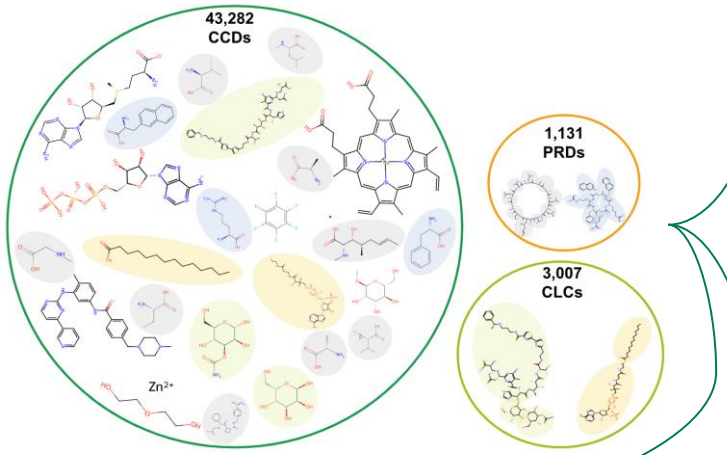
Kunnakkattu et al. (2023)

doi.org/10.1186/s13321-023-00786-w

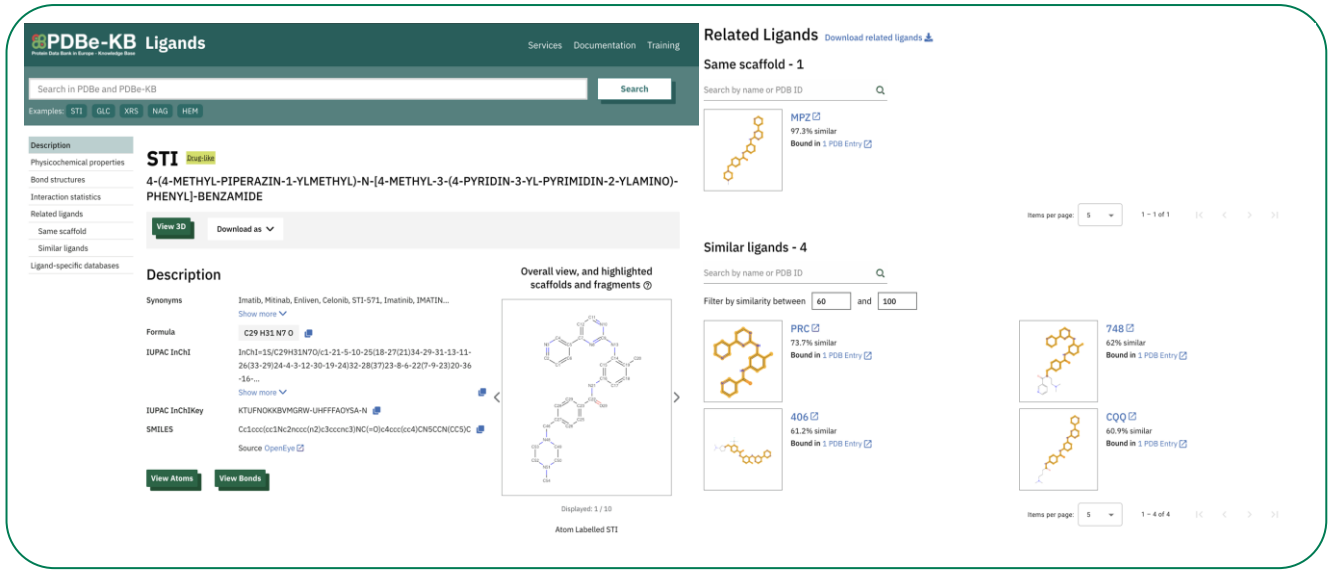
https://github.com/PDBeurope/ccdutils

EMBL-EBI

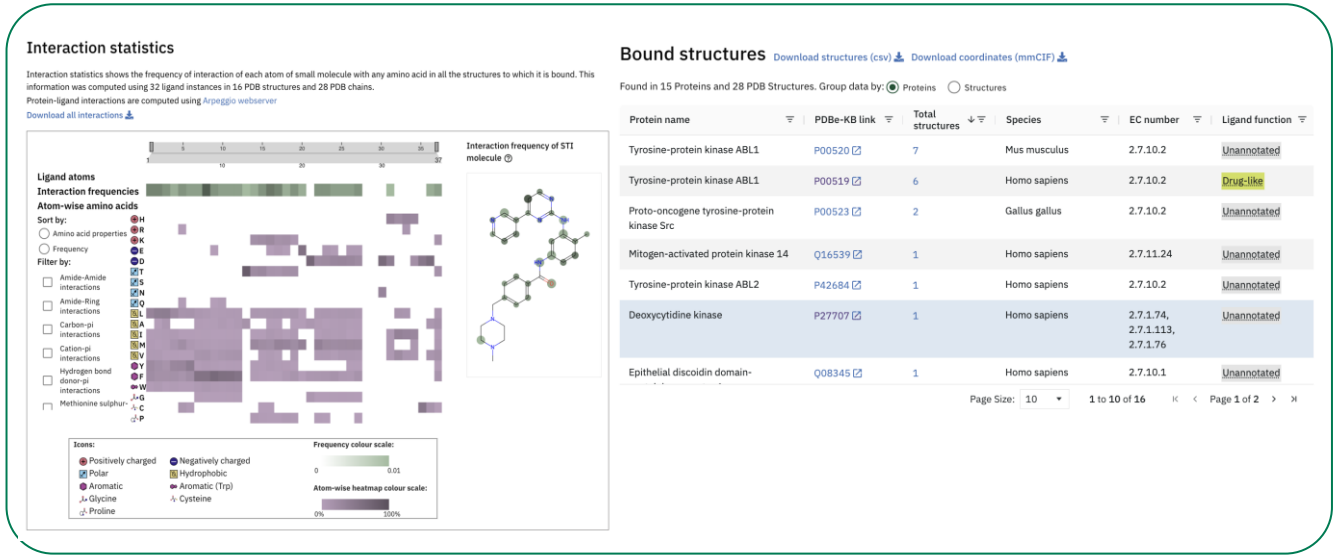
PDBe Ligand Pages



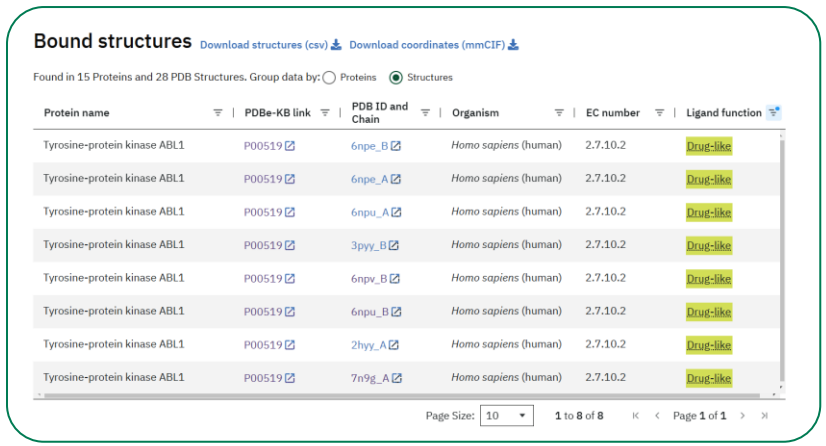
[PBDe CCDUtils]



[PBDe Arpeggio]



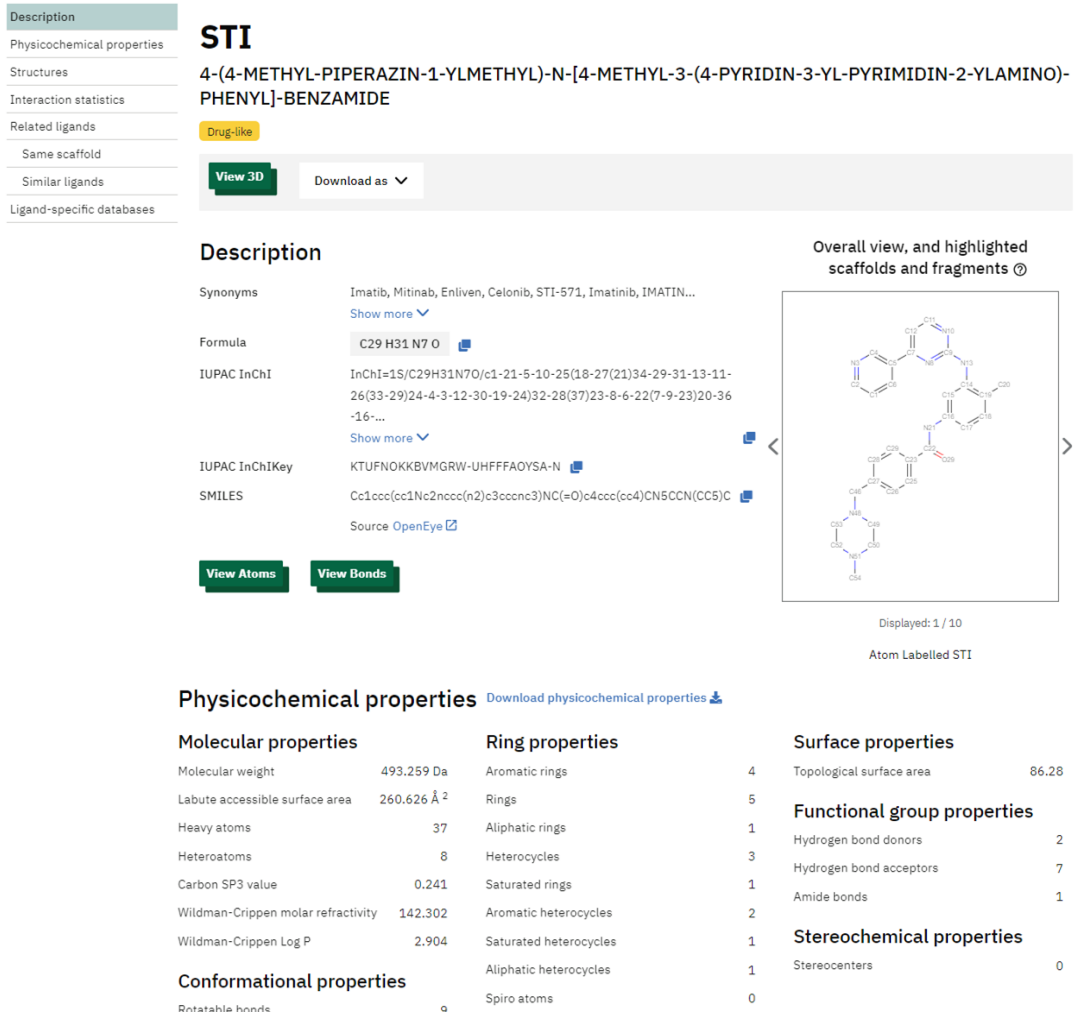
[PBDe RelLig]
(Relevant Ligands in proteins)



EMBL-EBI

PDBe Ligand Pages

Keyed on CCD/PRD/CLC IDs



Releasing soon (this November)

These webpages will replace [pdbechem service]

Following data will be accessible

- Basic physicochemical properties
- Cross-references to other resources
- Functional annotations
- Descriptors, synonyms
- All the structure to which it binds
- Similar ligands
- Ligands with same scaffold
- Stereoisomers
- All the interacting proteins, enzymes
- 2D and 3D views of ligand/protein

https://wwwdev.ebi.ac.uk/pdbe/connect/ligands/STI

EMBL-EBI

Acknowledgements

PDBe
Protein Data Bank in Europe

PDBe-KB
Protein Data Bank in Europe - Knowledge Base

CCDC

ChEMBL

PDBe
@PDBeurope
@PDBeurope
proteindatabank
proteindatabank
pdbeurope

The Protein Databank in Europe team

EMBL

UKRI UK Research and Innovation

wellcome

Funded by the European Union

Google DeepMind

iNEXT DISCOVERY

EMBL-EBI