

# ProtacMLA: PROTAC Mutation and Ligand analysis tool

By Tushar Gupta and Dr. Pradeep Pant

Department of Biotechnology

Bennett University

# ABOUT

- ProtacMLA is an online web-based tool that helps users extract, visualize, and analyse the Proteolysis targeting chimeras (PROTACs).
- This online freely accessible web tool helps to extract the PDB format file of PROTAC or the PROTAC complex and analyze the PROTAC region.
- This tool helps to:
  - Extract and analyse ligands acting as the E3 ligase binding ligand and the linker
  - Mutate the amino acid region that acts like the target protein-binding region of the PROTAC molecule
- The tool also enlists databases and tools that are useful in the PROTAC and protein study.

# Homepage

PDB  
Viewer  
section

The screenshot shows the ProtacMLA homepage with several sections:

- PDB Viewer section:** A pink box on the left labeled "PDB Viewer section" with a green arrow pointing to the top-left area of the interface.
- Mutation section:** A pink box on the left labeled "Mutation section" with a green arrow pointing to the "Peptide Mutation" and "PROTAC Ligand analysis" tools.
- Database and Tools section:** A pink box on the left labeled "Database and Tools section" with a green arrow pointing to the "Databases" and "Tools" sections at the bottom of the page.
- Ligand analysis section:** A pink box on the right labeled "Ligand analysis section" with a green arrow pointing to the "PROTAC Ligand analysis" tool.

**Peptide Mutation**  
Generate ready-to-use UCSF Chimera Python scripts with desired mutation and optional minimization.

**1) Structural Inputs**

PDB ID: e.g. 1ABC Local PDB structure file path: e.g. /home/user/model.pdb

Output structure location: e.g. /home/user/out/ Desired Output Format: PDB

**PROTAC Ligand analysis**  
Enter a PDB ID or upload a .pdb file. The tool finds non-polymer ligands and enriches them via the RCSB ChemComp API.

PDB ID: e.g. GLUT Load by ID

Or upload a .pdb file: Choose File No file chosen Analyze Upload

**Databases & Tools for PROTAC and Protein Research**

**Databases**

- PROTAC DB 3.0: Manually curated database for PROTAC molecules with degradation capacity, binding affinities and cellular activities.
- RCSB PDB: Archive of 3D structures of large biomolecules (Proteins, DNA/RNA, and their complexes).
- PROTCedia: Manually curated collection and comprehensive resource for PROTAC targeting chemists (PROTACs).
- PDBbind: Comprehensive collection of experimental Binding affinity data for biomolecules in PDB in RCSB PDB.
- AlphaFold DB: Collection of over 200 million predicted protein structures using AlphaFold.
- BindingDB: Binding affinity database focusing on protein complexes considered as drug targets.
- PubChem: Open chemistry database for small and large molecules.
- ZINC15: A free database of commercially available compounds for virtual screening.
- DrugBank: Comprehensive database of approved and experimental drugs.

**Tools**

- DeepPROTACs: A deep neural network model that predicts degradation capacity of PROTAC molecule based on structure of given target protein and E3 ligase.
- DHPROTACs: Transformers based model that learns and generates new PROTAC linkers based on given ligands.
- SwissADME: Website that allows to compute pharmacokinetic parameters and predict ADME parameters, pharmacodynamics, toxicity, drug-like nature and medicinal chemistry goodness of small molecule.
- PlayMolecule: A virtual environment for drug discovery where simulation, AI and data are integrated.
- ChemProp: A machine learning package for chemical property prediction using D-MPNN architecture.
- DeepChem: Open source ML toolkit that democratizes the use of Deep learning in drug discovery, material science, quantum chemistry and biology.
- ProteinPlus: A webserver for structure based molecular modeling, provides functionalities for structure quality assessment, hydrogen placement, search alternative conformations.

PBEE protein-peptide binding affinity predictor

# PDB structure viewer section

This section helps users visualize 3D structures along with their residues, chains, and involved ligands, allowing them to identify their target complex.

The viewer is designed in a way that it excludes common ions, water molecules, and uncommon ligands.

The diagram illustrates the PDB structure viewer section, showing the workflow from input to visualization and details.

**Input:** Users may enter the PDB ID or upload their PDB structure from the local system. This leads to the **Viewer Options** and **Structure Details** sections.

**Structure Visualization:** The 3D structure is loaded, which can be visualized here. The visualization shows a red ribbon model of a protein complex with green sticks representing ligands.

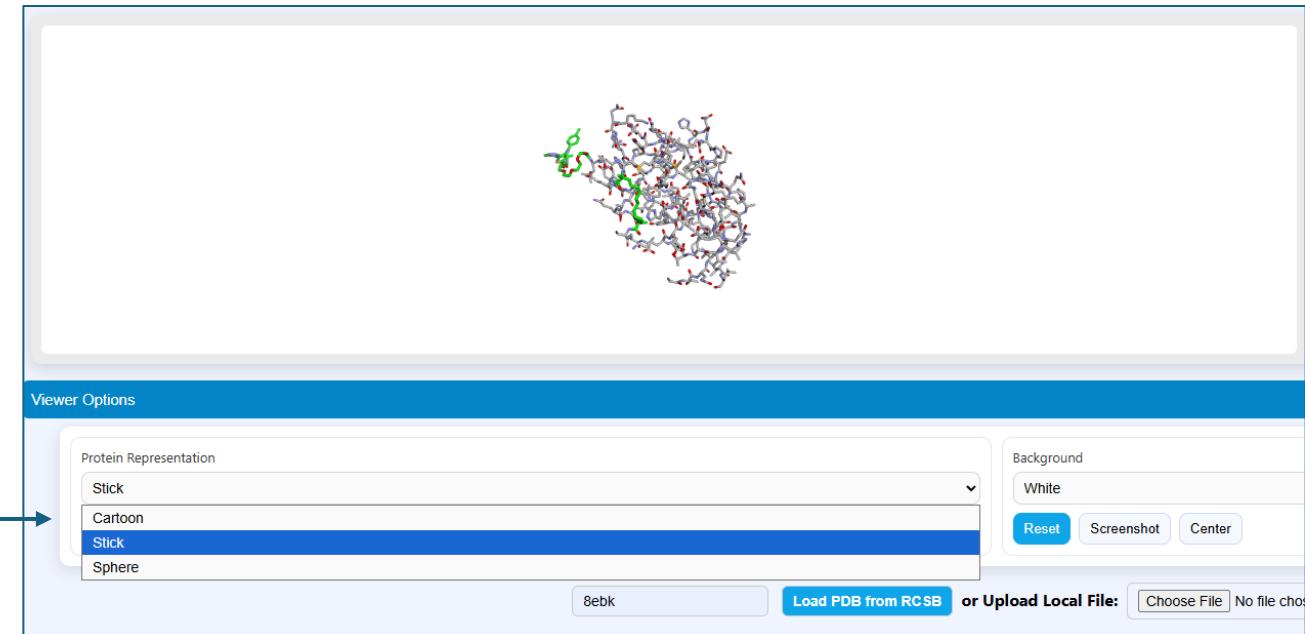
**Structure Details:** Details of the structure, including their residues, chain, and involved ligands, may be found here. The **Structure Details** panel displays residue lists for Chain A and Chain B, and a **Ligands** section.

**Viewer Options:** The viewer options are located at the bottom of both the input and visualization sections.

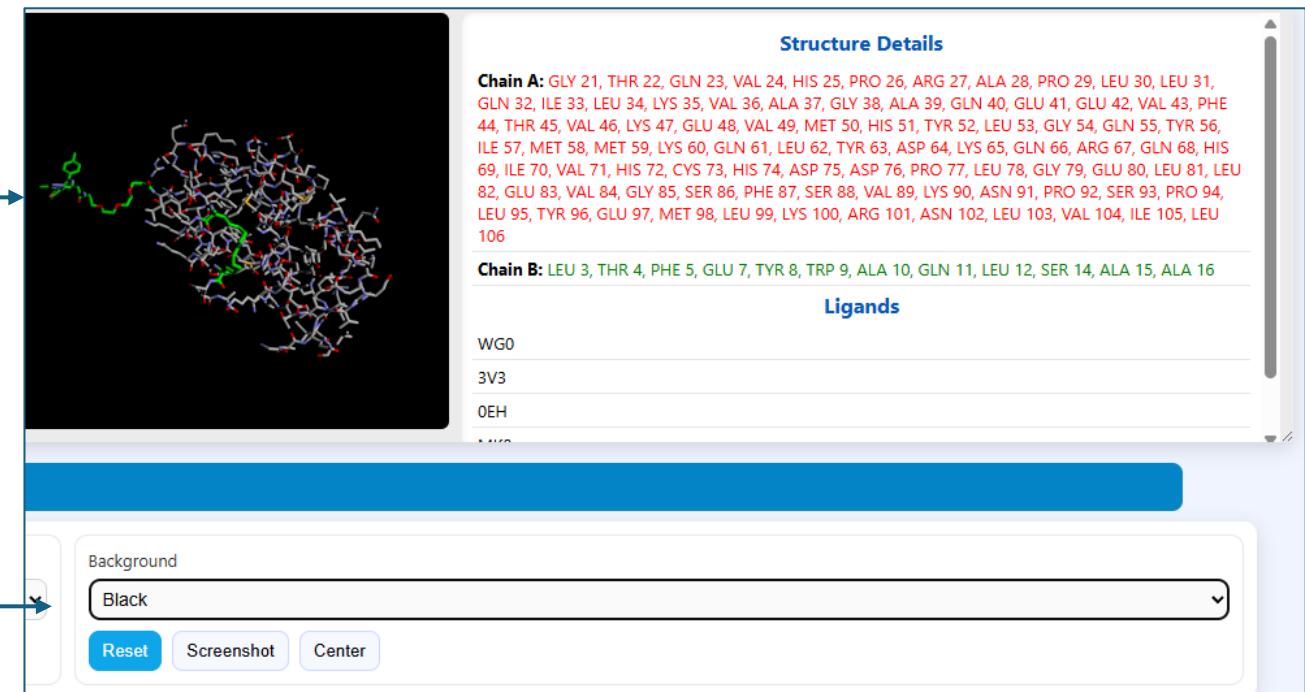
**Structure Details Panel:** The **Structure Details** panel includes fields for **Load a PDB to view details...**, **Ligands** (with a checkbox for **Show Ligands**), and a list of residues for Chain A and Chain B.

**Ligands Section:** The **Ligands** section lists ligand identifiers: WGO, 3V3, OEH, and 8ebk.

Using the ‘Viewer option’, the user may change the structure representation to cartoon/stick/sphere for easy visualization of the structure components.



Using the ‘Viewer option’, the user may also change the background to white/black/soft blue to increase the visibility of small components.



The options for Reset (reset to default settings), Screenshot (save structure image in PNG format), and Center (move the structure to the center of the view window) are also included.

# PROTAC mutation section

This section of the tool helps the user to mutate the amino acid sequence of the target protein binding site of the PROTAC molecule.

## 1) Structure Input:

It takes the input of the PDB ID or structure from the local system, the output file saving location, and the desired output file format (PDB/mmCIF/mol2)

## 2) Minimization (optional):

Takes the number of steps of Steepest Descent and Conjugate Gradient to run minimization of the mutated structure. This step is optional.

(Note: An Increased number of steps and structure complexity may increase the time of minimization.)

## 3) Mutation mode:

Allows users to select the type of mutation they want to perform on the structure. It includes,

Single or multiple mutation: Mutate one or more amino acids with any other amino acid. (Discussed in detail in the next section)

Class-wise mutation: Mutate one or more amino acids to all amino acids of any class. (Discussed in detail in the next section)

Enter the number of amino acids that you want to mutate and generate fields for input entry.

The user may also visualize the generated Python script and download it to directly run in Chimera 1.19+ for mutation and minimization, and save the final structures in the desired folder in the desired format.

The screenshot shows the 'PROTAC Mutation' interface. At the top, a header reads 'PROTAC Mutation' with a sub-instruction: 'Generate ready-to-use UCSF Chimera Python scripts with desired mutation and optional minimization.' Below this are three main sections: 1) Structural Inputs, which includes fields for 'PDB ID' (e.g. 1ABC), 'Local PDB structure file path' (e.g. \home\user\model.pdb), 'Output structure location' (e.g. \home\user\out\_folder\), and 'Desired Output Format' (set to PDB). 2) Minimization (optional), which has a checked checkbox for 'Select to perform energy minimization' and fields for 'Enter number of Steepest Descent Steps' (e.g. 2000) and 'Enter number of Conjugate Gradient Steps' (e.g. 200). 3) Mutation Mode, which offers two options: 'Single / Multiple' (selected) and 'Class-wise'. Below these are buttons for 'Generate Mutation Inputs', 'Preview Script', and 'Download Script'. A 'Generated Script Preview' section is at the bottom, currently empty.

## PROTAC Mutation

Generate ready-to-use UCSF Chimera Python scripts with desired mutation and optional minimization.

PDB ID or enter pdb structure location from local system

Enter preferred output file saving location

Enter the steps of steepest descent and conjugate gradient for minimization

Select the type of mutation to conduct on desired amino acids

Enter the number of amino acids to mutate

### 1) Structural Inputs

PDB ID

8ebk

Local PDB structure file path

e.g. \home\user\model.pdb

Output structure location

C:\Users\tusha\_t46syw9\Desktop\trial\

Desired Output Format

PDB

### 2) Minimization (optional)

Select to perform energy minimization



Enter number of Steepest Descent Steps

200

Enter number of Conjugate Gradient Steps

200

### 3) Mutation Mode

Single / Multiple

Class-wise

Enter number of mutations

1

Generate Mutation Inputs

B

(1)

3

(2)

Y (TYR - Tyrosine)

(3)

Preview Script

Download Script

### Generated Script Preview

```
from chimera import runCommand
runCommand("open 8ebk")
runCommand("swapaa tyr :3.B preserve true")
runCommand("minimize nsteps 200 cgsteps 200 nogui true")
runCommand("write B C:\\\\Users\\\\tusha_t46syw9\\\\Desktop\\\\trial\\\\tyr_3_min.pdb")
runCommand("close all")
```

Choose desired output format

Check the box if you want to run minimization on output structures

For the case of single/multiple mutation mode :

1. Enter the Chain ID of the desired amino acid to mutate
2. Enter the residue ID or residue number of the amino acid
3. Choose the amino acid you wish to mutate to  
Repeat the same for all the entries

## PROTAC Mutation

Generate ready-to-use UCSF Chimera Python scripts with desired mutation and optional minimization.

PDB ID or enter pdb structure location from local system

Enter preferred output file saving location

Enter the steps of steepest descent and conjugate gradient for minimization

Select the type of mutation to conduct on desired amino acids

Enter the number of amino acids to mutate

### 1) Structural Inputs

PDB ID

8ebk

Local PDB structure file path

e.g. \\home\\user\\model.pdb

Output structure location

C:\\Users\\tusha\_t46syw9\\Desktop\\trial\\

Desired Output Format

PDB

### 2) Minimization (optional)

Select to perform energy minimization



Enter number of Steepest Descent Steps

200

Enter number of Conjugate Gradient Steps

200

### 3) Mutation Mode

Single / Multiple

Class-wise

Enter number of classes

1

Generate Class Fields

Hydrophobic

(1)

A

(2)

3

(3)

Members: A (ALA - Alanine), V (VAL - Valine), I (ILE - Isoleucine), L (LEU - Leucine), G (GLY - Glycine), M (MET - Methionine), P (PRO - Proline)

Preview Script

Download Script

### Generated Script Preview

```
runCommand('u$wapaa val :3.A preserve true')
runCommand('u$minimize nsteps 200 cgsteps 200 nogui true')
runCommand('u$write @ C:\\\\Users\\\\tusha_t46syw9\\\\Desktop\\\\trial\\\\val_3_min.pdb')
runCommand('u$wapaa ile :3.A preserve true')
runCommand('u$minimize nsteps 200 cgsteps 200 nogui true')
runCommand('u$write @ C:\\\\Users\\\\tusha_t46syw9\\\\Desktop\\\\trial\\\\ile_3_min.pdb')
runCommand('u$wapaa leu :3.A preserve true')
runCommand('u$minimize nsteps 200 cgsteps 200 nogui true')
runCommand('u$write @ C:\\\\Users\\\\tusha_t46syw9\\\\Desktop\\\\trial\\\\leu_3_min.pdb')
runCommand('u$wapaa gly :3.A preserve true')
runCommand('u$minimize nsteps 200 cgsteps 200 nogui true')
runCommand('u$write @ C:\\\\Users\\\\tusha_t46syw9\\\\Desktop\\\\trial\\\\gly_3_min.pdb')
runCommand('u$wapaa met :3.A preserve true')
runCommand('u$minimize nsteps 200 cgsteps 200 nogui true')
runCommand('u$write @ C:\\\\Users\\\\tusha_t46syw9\\\\Desktop\\\\trial\\\\met_3_min.pdb')
runCommand('u$wapaa pro :3.A preserve true')
runCommand('u$minimize nsteps 200 cgsteps 200 nogui true')
runCommand('u$write @ C:\\\\Users\\\\tusha_t46syw9\\\\Desktop\\\\trial\\\\pro_3_min.pdb')
runCommand("close all")
```

Choose desired output format

Check the box if you want to run minimization on output structures

For the case of class-wise mutation mode :

1. Enter the preferred class of amino acid (all amino acids of each class are listed for reference)
2. Enter the chain ID for the amino acid
3. Enter the residue ID or residue number of the amino acid

Repeat the same for all the entries

# PROTAC ligand analysis section

This section of the tool helps users analyse features of ligands present in the structure playing role as E3 ligase binding ligand and linker.

This section also includes an exclusion list that enlists all the amino acids, small, uncommon ligands, and ions that may be changed by the user.

It takes the input pdb structure or is fetched using the PDB ID from RCSB PDB and displays ligands, including ones involved in E3 ligase binding ligand and as a linker for E3 ligase binding ligand and target protein binding region.

## PROTAC Ligand analysis

Enter a PDB ID or upload a .pdb file. The tool finds non-polymer ligands and enriches them via the RCSB ChemComp API.

**PDB ID**

e.g., 6LU7

We download the .pdb file from RCSB and analyze it in your browser.

**Or upload a .pdb file**

Choose File No file chosen

We read only in your browser; nothing is uploaded anywhere else.

**Load by ID**

**Analyze Upload**

**Exclusions (common non-ligands):**

We ignore standard amino acids, water, and typical ions/solvents by default. You can add/remove three-letter codes below (comma-separated).

ALA, ARG, ASN, ASP, CYS, GLN, GLU, GLY, HIS, ILE, LEU, LYS, MET, PH  
E, PRO, SER, THR, TRP, TYR, VAL, SEC, PYL, HOH, WAT, DOD, NA, CL, MG  
, ZN, CA, SO4, PO4, HEM, NAG, MAN, BMA, PEG, MPD, GOL

Tip: Remove entries like NAG if you \*do\* want carbohydrates included.

# PROTAC Ligand analysis

Enter a PDB ID or upload a .pdb file. The tool finds non-polymer ligands and enriches them via the RCSB ChemComp API.

Enter PDB ID or upload the PDB structure from the local system

PDB ID

8ebk

Load by ID

We download the .pdb file from RCSB and analyze it in your browser.

Or upload a .pdb file

Choose File No file chosen

Analyze Upload

We read only in your browser; nothing is uploaded anywhere else.

For downloading the list of ligands in CSV format

Found 4 ligand type(s) in 8EBK.

Download CSV

List of all the ligands in the PDB structure, excluding ones in the 'Exclusions' list

The tool finds non-polymer ligands and enriches them via the RCSB ChemComp API.

## Exclusions (common non-ligands):

We ignore standard amino acids, water, and typical ions/solvents by default. You can add/remove three-letter codes below (comma-separated).

ALA, ARG, ASN, ASP, CYS, GLN, GLU, GLY, HIS, ILE, LEU, LYS, MET, PHE, PRO, SER, THR, TRP, TYR, VAL, SEC, PYL, HOH, WAT, DOD, NA, CL, MG, ZN, CA, SO4, PO4, HEM, NAG, MAN, BMA, PEG, MPD, GOL

Tip: Remove entries like NAG if you \*do\* want carbohydrates included.

Ligand	Chemical Name	Formula	Weight (g/mol)	Instances (chain:resid)	Number of Atoms	Details
WG0	[(6S,10P)-4-(4-chlorophenyl)-2,3,9-trimethyl-6H-thieno[3,2-f][1,2,4]triazolo[4,3-a][1,4]diazepin-6-yl]acetic acid	C19 H17 Cl N4 O2 S	400.882	B:1	26	<button>View</button>
3V3	1-amino-3,6,9,12-tetraoxapentadecan-15-oic acid	C11 H23 N O6	265.303	B:2	17	<button>View</button>
0EH	(2R)-2-amino-2-methylnonanoic acid	C10 H21 N O2	187.279	B:6	12	<button>View</button>
MK8	2-methyl-L-norleucine	C7 H15 N O2	145.199	B:13	9	<button>View</button>

Click on 'View' for more information

## Downloaded ligand table with features

Ligand	Name	Formula	Weight_g_mol	Instances_chain:resid	TotalAtoms
WG0	[(6S,10P)-4-(4-chlorophenyl)-2,3,9-trimethyl-6H-thieno[3,2-f][1,2,4]triazolo[4,3-a][1,4]diazepin-6-yl]acetic acid	C19 H17 Cl N4 O2 S	400.882	B:1	26
3V3	1-amino-3,6,9,12-tetraoxapentadecan-15-oic acid	C11 H23 N O6	265.303	B:2	17
0EH	(2R)-2-amino-2-methylnonanoic acid	C10 H21 N O2	187.279	B:6	12
MK8	2-methyl-L-norleucine	C7 H15 N O2	145.199	B:13	9

After clicking the ‘View’ button, the window expands to list features of the specific ligand, including IUPAC name, chemical formula, molecular weight, and 2D structure.

3D structure of the ligand is displayed using PubChem, which may be extended further to see other features of the ligand by clicking the link ‘Explore more on PubChem’.

The screenshot shows a ligand table on the left and a detailed view for ligand WG0 on the right.

**Ligand Table:**

Ligand	Chemical Name	Formula	Weight (g/mol)	Instances (chain:resid)	Number of Atoms	Details
WG0	[(6S,10P)-4-(4-chlorophenyl)-2,3,9-trimethyl-6H-thieno[3,2-f][1,2,4]triazolo[4,3-a][1,4]diazepin-6-yl]acetic acid	C19 H17 Cl N4 O2 S	400.882	B:1	26	<a href="#">View</a>
3V3	1-amino-3,6,9,12-tetraoxapentadecan-15-oic acid	C11 H23 N O6	265.303	B:2	17	<a href="#">View</a>
0EH	(2R)-2-amino-2-methylnonanoic acid	C10 H21 N O2	187.279	B:6	12	<a href="#">View</a>
MK8	2-methyl-L-norleucine	C7 H15 N O2	145.199	B:13	9	<a href="#">View</a>

**Detailed View for Ligand WG0:**

- WG0**
- Name:** [(6S,10P)-4-(4-chlorophenyl)-2,3,9-trimethyl-6H-thieno[3,2-f][1,2,4]triazolo[4,3-a][1,4]diazepin-6-yl]acetic acid
- Formula:** C19 H17 Cl N4 O2 S
- Weight:** 400.882 g/mol
- 2D Structure:** A 2D chemical structure diagram showing the complex heterocyclic system of the ligand.
- 3D Structure:** A 3D ball-and-stick model of the ligand.
- PubChem Link:** [Explore more on PubChem](#)

# Databases and Tools for PROTAC and Protein Research

This section of the tool is designed to help users access various available resources for computational drug discovery and bioinformatics using PROTAC.

The included databases and tools range from several PROTAC-specific resources to general ones.

## Databases



PROTAC-DB 3.0

RCSB PDB

PROTACpedia

PDBbind+

AlphaFold DB

BindingDB

PubChem

ZINC15

DrugBank

## Tools



DeepPROTACs

DiffPROTACs

SwissADME

PlayMolecule

ChemProp

DeepChem

ProteinPlus

# DATABASES

## PROTAC-DB 3.0

- Purpose:** A manually curated database specifically for PROteolysis-TArgeting Chimeras (PROTACs). It's a specialized resource for researchers working on targeted protein degradation.
- Key Features:** It provides detailed information on PROTAC molecules, including their chemical structures, and crucial biological data like degradation capacity (DC50 and Dmax), binding affinities (IC50, Ki, Kd), and cellular activities. The updated version, 3.0, has expanded its entries and now includes pharmacokinetic data, which is essential for assessing the druggability of these molecules. The database also categorizes the components of a PROTAC: the warhead (the part that binds to the target protein), the E3 ligase ligand, and the linker.

## PDBbind+

- Purpose:** A specialized database that bridges structural and energetic information. It's a comprehensive collection of experimentally determined binding affinity data for protein-ligand complexes that have a corresponding 3D structure in the RCSB PDB.
- Key Features:** It is an invaluable resource for developing and validating computational methods in drug discovery, such as docking and scoring functions. The database provides a "refined set" of high-quality data, which is a standard benchmark for testing the performance of molecular modeling algorithms.

## PubChem

- Purpose:** An open chemistry database maintained by the National Institutes of Health (NIH). It's a foundational resource for a vast array of chemical information.
- Key Features:** It includes a massive collection of information on chemical structures, physical properties, biological activities, and more. Data is contributed by a wide range of sources, including government agencies and vendors. It's a go-to resource for chemical searching, whether by structure, name, or other identifiers.

## RCSB PDB (Protein Data Bank)

- Purpose:** The central global archive for 3D structures of large biomolecules. It serves as a fundamental resource for structural biologists, providing a detailed view of the atomic coordinates of proteins, DNA, RNA, and their complexes.
- Key Features:** PDB entries are derived from experimental methods like X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy. The database is a go-to source for understanding molecular function, interactions, and for structure-based drug design. It includes extensive annotations, visualizations, and tools for searching and analyzing structures.

## AlphaFold DB

- Purpose:** A collection of over 200 million predicted protein structures generated by Google DeepMind's AlphaFold AI system. It dramatically expands the number of available protein structures beyond those determined experimentally.
- Key Features:** For many organisms, it provides a comprehensive predicted proteome. Each structure comes with a per-residue confidence score (pLDDT), allowing users to assess the reliability of the prediction for different regions of the protein. This database is accelerating research by providing structural models for proteins that have not yet been experimentally characterized.

## ZINC15

- Purpose:** A free and comprehensive database of commercially available chemical compounds. It is optimized for virtual screening.
- Key Features:** ZINC15 contains over 230 million purchasable compounds, all in "ready-to-dock," 3D formats. This is a crucial feature for computational chemists who need to quickly prepare large libraries of molecules for docking simulations. It also provides various filters based on molecular properties like molecular weight and LogP, making it easy to create focused screening libraries.

## PROTACpedia

- Purpose:** A collaborative, high-quality, and freely accessible resource for PROTACs. Unlike commercial databases, it's designed to be a community-driven platform.
- Key Features:** It provides a manually curated collection of data on PROTAC molecules. The collaborative nature of the platform means that registered users can contribute new data, helping to expand the resource. This makes it a dynamic and up-to-date source of information for the PROTAC research community.

## BindingDB

- Purpose:** A publicly accessible database focused on the binding affinities of small molecules to proteins, particularly those considered to be drug targets.
- Key Features:** It contains millions of data points, including Ki, Kd, IC50, and EC50 values. This data is critical for medicinal chemists and computational modelers for developing Structure-Activity Relationships (SAR), training machine learning models, and validating docking methods. It also provides links to related information in other databases like RCSB PDB and PubChem.

## DrugBank

- Purpose:** A unique bioinformatics and cheminformatics resource that combines detailed drug information with comprehensive drug target data.
- Key Features:** It includes data on approved and experimental drugs, providing their chemical, pharmacological, and pharmaceutical details. It also offers a wealth of information on the drugs' protein targets, including sequences and pathways. This database is a powerful tool for linking drugs to their mechanisms of action and for drug repurposing studies.

# TOOLS

## DeepPROTACs

- **Purpose:** A deep neural network model designed to predict the degradation capacity of a PROTAC molecule.
- **Key Features:** It takes the 3D structures of the target protein and the E3 ligase as input, along with the PROTAC molecule's structure. The model uses a combination of Graph Convolutional Networks (GCNs) and other neural network architectures to predict whether a given PROTAC will effectively degrade its target. This tool helps in the early-stage rational design of PROTACs, as it can filter out potentially ineffective molecules before synthesis.

## DiffPROTACs

- **Purpose:** A generative AI model based on diffusion and transformer architectures for designing new PROTAC linkers.
- **Key Features:** This tool addresses a major challenge in PROTAC design: the linker. It can learn the properties of existing linkers and generate novel ones that connect a given target ligand and E3 ligase ligand. It's a powerful approach for exploring the vast chemical space of potential linkers to find ones that optimize the properties of the PROTAC molecule.

## PlayMolecule

- **Purpose:** A virtual environment for drug discovery that integrates simulations, AI, and data. It provides a comprehensive platform for various computational workflows.
- **Key Features:** It offers a suite of applications for tasks like protein and small molecule preparation, virtual screening, binding mode analysis, and relative binding affinity predictions. It can be accessed via a graphical user interface, a Python API, or the command line, providing flexibility for different user needs. It's designed to streamline complex drug discovery processes.

## ChemProp

- **Purpose:** An open-source machine learning package for chemical property prediction, specifically using Directed Message Passing Neural Networks (D-MPNNs).
- **Key Features:** This is a coding-focused tool for researchers who want to build their own predictive models. It simplifies the process of training and using powerful graph neural networks on molecular data. It can predict a wide range of properties for single molecules, reactions, and multi-molecule systems, making it highly versatile for cheminformatics tasks.

## SwissADME

- **Purpose:** A web-based tool for computing physicochemical descriptors and predicting Absorption, Distribution, Metabolism, and Excretion (ADME) parameters for small molecules.
- **Key Features:** It helps researchers quickly assess the "druglike" nature of their compounds. It predicts properties such as water solubility, gastrointestinal absorption, blood-brain barrier penetration, and adherence to rules like Lipinski's Rule of Five. This is an essential tool for early-stage drug discovery to filter out molecules with poor pharmacokinetic properties.

## DeepChem

- **Purpose:** An open-source machine learning toolchain that aims to democratize the use of deep learning in drug discovery, materials science, quantum chemistry, and biology.
- **Key Features:** It provides high-level APIs for building and deploying deep learning models on molecular datasets. It includes a wide variety of models and data featurizers, making it easy to get started with complex machine learning tasks. It's a foundational library for researchers who want to apply AI to scientific problems.

## ProteinPlus

- **Purpose:** A web server for structure-based molecular modeling, focusing on supporting life scientists who work with protein structures.
- **Key Features:** It provides a suite of functionalities for preparing and analyzing protein structures, especially in the context of protein-ligand interactions. Key tools include structure quality assessment, hydrogen placement, binding site prediction (DoGSiteScorer), and the generation of aligned protein structure ensembles. It also features tools for 2D ligand interaction diagrams and finding alternative conformations.