PRAC-G PRIVATE LIMITED

AI IN DRUG DISCOVERY

A Beginner's Hands-on Workshop

Understanding AI, Molecular Prediction and Lead Prioritisation

Duration: 3 Days | 2 Hours/Day | Total: 6 Hours

Format: Lecture + Demonstration + Guided Hands-on Practical

1. Workshop Overview

This beginner-level workshop introduces students to the role of Artificial Intelligence (AI) and Machine Learning (ML) in modern drug discovery. The workshop follows a practical, stepwise approach in which students move from understanding a biological target and chemical data to AI-based prediction, molecular docking, ADMET/drug-likeness assessment, and final lead prioritisation.

The workshop follows the practical structure of the reference Bioinformatics curriculum: learning objectives, topics covered, tools/materials, hands-on activities, and a final deliverable.

2. Aim

To provide beginners with a clear conceptual and practical understanding of how AI/ML can support drug discovery and how computational predictions can be integrated to prioritise potential drug candidates.

3. Learning Outcomes

- Explain the basic concepts of AI, Machine Learning and their relevance to drug discovery.
- Describe the major stages of the drug-discovery pipeline and identify where AI can be applied.
- Explore basic biological and chemical databases used in drug discovery.
- Understand how molecular structures are represented as numerical features for ML.
- Understand and apply a simple ML-based molecular activity prediction workflow.
- Understand the purpose of protein structure analysis and molecular docking.
- Interpret basic drug-likeness and ADMET predictions.
- Integrate AI prediction, docking and ADMET information to prioritise candidate molecules.
- Recognise that computational predictions support, but do not replace, experimental validation.

4. Workshop Learning Journey

DAY 1 - UNDERSTAND: Drug Discovery + AI + Biological/Chemical Data

DAY 2 - PREDICT: Machine Learning + Molecular Activity Prediction

DAY 3 - PRIORITISE: Docking + ADMET + AI-Assisted Lead Selection

Overall workflow: DATA → AI → PREDICTION → VALIDATION → LEAD PRIORITISATION

DAY 1 — Introduction to AI in Drug Discovery

Theme: How is AI changing Drug Discovery?

Time	Session	Format
6:00-6:20	Introduction to Drug Discovery	Lecture
6:20-6:45	Where AI/ML Fits into Drug Discovery	Lecture + Discussion
6:45-7:05	Drug & Biological Databases	Demo
7:05-7:30	Understanding Molecular Data	Lecture + Demo
7:30-7:50	Hands-on: Explore a Target & Its Known Drugs	Practical
7:55-8:00	Break / Q&A	-

Module 1 - Drug Discovery Fundamentals

Drug discovery pipeline:

Disease → Target identification → Hit identification → Lead optimisation → Preclinical testing → Clinical trials

AI application areas:

- Target identification
- Drug-target prediction
- Virtual screening
- Molecular property prediction
- ADMET prediction
- Drug design

Module 2 - What are AI and ML?

- Artificial Intelligence: the broader concept of machines performing intelligent tasks.
- Machine Learning: algorithms learn patterns from data.
- Deep Learning: neural networks learn complex patterns from large or complex datasets.

Simple biological example: Molecular data → ML model → Activity prediction

Module 3 - Exploring Drug Discovery Databases

Recommended beginner resources (for example):

- PubChem - chemical information
- ChEMBL - bioactivity data
- RCSB PDB - protein structures
- UniProt - protein information

Hands-on activity: Students select one target and retrieve its protein information, known inhibitors/ligands, chemical structures and an available protein structure.

Day 1 takeaway: Students should be able to answer - "What is my biological target, and what molecules are already known to interact with it?"

DAY 2 - Machine Learning for Drug Discovery

Theme: Can AI predict which molecules might work?

Time	Session	Format
6:00-6:25	Machine Learning Basics	Lecture
6:25-6:50	How Molecules Become AI Data	Lecture + Demo
6:50-7:10	Building a Simple Drug-Activity Model	Demo
7:10-7:20	Break / Q&A	-
7:20-7:50	Hands-on: Predict Molecular Activity	Practical
7:50-8:00	Results & Interpretation	Discussion

Module 4 - Machine Learning Basics

- Supervised learning
- Training data, test data, features and labels
- Prediction and model performance
- Overfitting conceptual understanding

Beginner-friendly models: Random Forest (primary), Logistic Regression and SVM (brief introduction).

Module 5 - How Does AI Understand a Molecule?

Molecule → SMILES → Molecular descriptors/fingerprints → Numerical features → Machine Learning → Prediction

Basic descriptors: molecular weight, LogP, H-bond donors, H-bond acceptors and TPSA.

Recommended tool: RDKit.

Module 6 - Hands-on AI Prediction

Students work with a small, instructor-prepared dataset containing Compound ID, SMILES and activity.

1. Load and inspect the dataset.
2. Examine basic molecular properties.
3. Train or apply a simple ML model.
4. Predict activity for unseen compounds.
5. Rank the predicted candidates.

Day 2 deliverable: Top 5 AI-predicted compounds with prediction scores and a short rationale.

Key scientific message: AI prediction helps prioritise experiments; it does not prove that a compound is a drug.

DAY 3 - From AI Prediction to Drug Candidate

Theme: Can we combine AI with structure-based drug discovery?

Time	Session	Format
6:00-6:20	Protein Structure & AI	Lecture + Demo
6:20-6:40	Protein-Ligand Interaction	Lecture
6:40-7:10	Molecular Docking Demonstration	Demo
7:10-7:20	Break / Q&A	-
7:20-7:40	ADMET & Drug-Likeness	Lecture + Demo
7:40-7:55	Integrating AI + Docking + ADMET	Practical
7:55-2:10	Final Conclusion	Discussion

Module 7 - AI & Protein Structure

Protein sequence → Protein structure → Binding pocket → Drug binding

- What is AlphaFold?
- Why protein structure matters
- Predicted vs experimental structures
- Structural confidence
- Using structure to support drug discovery

Module 8 - Molecular Docking

Core question: "Can this molecule fit into the target protein's binding site?"

Protein + Ligand → Docking → Binding pose → Docking score → Interaction analysis

Recommended tools for demonstration: AutoDock Vina and PyMOL (or an equivalent molecular viewer).

For beginners, docking should be a guided demonstration using one target and a small number of compounds rather than a lengthy installation/troubleshooting exercise.

Module 9 - ADMET & Drug-Likeness

Drug-likeness: Molecular weight, LogP, H-bond donors/acceptors and Lipinski's Rule of Five.

ADMET: Absorption, Distribution, Metabolism, Excretion and Toxicity.

Key concept: Good binding does not automatically mean a good drug candidate.

Module 10 - Lead Prioritisation

Students integrate three evidence streams:

- AI prediction
- Docking / interaction analysis
- ADMET and drug-likeness

The objective is not simply to select the molecule with the best docking score, but to identify the candidate with the strongest overall computational evidence.

5. Final Hands-on Activity AI-Assisted Lead Selection

Students receive five candidate molecules and compare AI prediction, docking and ADMET/drug-likeness information.

Compound	AI Prediction	Docking	ADMET	Final Decision
A	High	Good	Good	Priority
B	High	Poor	Good	Consider
C	Medium	Excellent	Good	Priority
D	Low	Good	Poor	Reject
E	Medium	Moderate	Good	Consider

Final Deliverable

AI-Assisted Lead Selection Table + 3-5 sentence scientific rationale for the final selected candidate(s).

6. Recommended Tools & Resources

Category	Tools / Resources
Biological / Chemical Databases	PubChem, ChEMBL, UniProt, RCSB PDB
Molecular Representation	RDKit
Machine Learning	Python, Google Colab/Jupyter, scikit-learn
Protein Structure	AlphaFold DB, RCSB PDB
Molecular Docking	AutoDock Vina
Structure Visualisation	PyMOL or equivalent
ADMET / Drug-likeness	Beginner-friendly web-based prediction platform