Workshop Handout: CADD Essentials

1. Essential Databases & Tools

Bookmark these for daily use during the workshop.

Target (Protein) Resources

- **RCSB PDB (Protein Data Bank):** The primary repository for 3D protein structures.
- **UniProt:** For protein sequence and functional information (referenced in Module 1).
- **CASTp / COACH:** Servers for predicting active sites (binding pockets) on your protein.

Ligand (Drug) Resources

- **PubChem:** Massive database of chemical molecules and their activities.
- **ZINC Database:** Commercially available compounds for virtual screening.
- **SwissADME / pkCSM:** For predicting drug-likeness (ADMET) and toxicity (added in Module 1).

Software Toolkit

- **Visualization:** PyMOL, UCSF Chimera, Biovia Discovery Studio.
- **Docking:** AutoDock Vina, PyRx.
- **Simulation:** GROMACS (Linux-based MD engine).

---

2. Glossary of Key Terms

Structural Biology

- **Residue:** A single amino acid unit within a protein chain.
- **Active Site:** The specific "pocket" or region on a protein where a ligand binds to exert its effect.
- **Heteroatoms (HETATM):** Non-protein atoms in a PDB file, such as water, ions, or co-crystallized ligands, which often need to be removed before docking.

Docking & Screening

- **Ligand:** A small molecule (drug candidate) that binds to a biomolecule.
- **Pose:** A specific orientation and conformation of a ligand inside the binding site.

- **Binding Affinity (ΔG):** The strength of the interaction between the protein and ligand. Usually measured in **kcal/mol**. More negative values indicate stronger binding.
- **Grid Box:** The 3D user-defined space where the docking software searches for binding interactions.

Molecular Dynamics (MD)

- **Force Field:** A set of mathematical functions and parameters (like AMBER or CHARMM) used to calculate the potential energy of the system.
- **RMSD (Root Mean Square Deviation):** A metric used to measure the average distance between atoms of the protein over time. It indicates **structural stability**.
- **RMSF (Root Mean Square Fluctuation):** Measures how much each individual residue moves. It indicates **flexibility**.
- **Solvation:** The process of surrounding the protein-ligand complex with water molecules to mimic a biological environment.

---

3. The "Golden Workflow"

Keep this roadmap in mind for the next 10 days.

1. **Preparation:** Clean PDB (Protein) + Optimize Ligand (energy minimization).
2. **Grid Generation:** Define *where* the drug should bind.
3. **Docking:** Run the algorithm to find the best fit.
4. **Analysis:** Visual inspection (H-bonds, Hydrophobic interactions).
5. **Simulation:** Run MD (10–50 ns) to test if the drug *stays* bound under physiological conditions.