Side-by-side workflow: Known vs. Unknown Ligands (Dengue NS3)

Step	Known Ligands (ChEMBL, labeled)	Unknown Ligands (ZINC, unlabeled)
Tut001	Query ChEMBL for Dengue NS3 protease bioactivity data (IC_{50}, pIC_{50}).	Not applicable yet — ChEMBL dataset is used first to build the training set.
Tut002	Apply Lipinski's Rule of Five filters to clean known ligands.	Apply same filters to ZINC compounds later for consistency.
Tut003	Remove PAINS, undesirable chemotypes, duplicates.	Apply identical cleaning to ZINC set to avoid false positives.
Tut004	Generate molecular fingerprints/descriptors for known ligands.	Generate same type of features for ZINC molecules.
Tut005	Cluster known ligands to inspect chemical diversity.	Optionally cluster ZINC predictions later to ensure diverse hit selection.
Tut006	Find Maximum Common Substructure (MCS) among Dengue NS3 actives.	Analyze MCS of top ZINC hits to understand scaffold similarity.
Tut007	Train & validate ML classifier (active/inactive) using known ligands.	Apply the trained model to ZINC ligands → predict their activity vs Dengue NS3.
Tut008	Prepare Dengue NS3 protease structure for docking (clean, protonate, remove water).	Dock top-predicted ZINC hits into the same NS3 structure.
Tut009	Define the binding site for Dengue NS3 from known ligand crystal structures.	Use same binding site definition for ZINC docking.
Tut010	Run docking on known ligands as benchmark.	Run docking on predicted ZINC hits for comparison.
Tut011	Visualize docking poses of known Dengue NS3 inhibitors.	Visualize docking poses of ZINC hits and compare with known actives.
Tut012	Evaluate docking scores & interactions (H-bonds, ionic, hydrophobic contacts).	Evaluate docking scores of ZINC compounds to shortlist the best ones.
Tut013	Create a final hit list from validated known ligands (benchmark set).	Select novel ZINC candidates for experimental validation.