Technical Challenge – Computer-Aided Drug Discovery (CADD)

Objective: Demonstrate your ability to execute a small-scale computer-aided drug discovery workflow from ligand selection to docking, ranking, and ADMET evaluation. We are evaluating not only the quality of the docking and ranking results, but also your approach to workflow design, scalability, and reproducibility. Please include short notes on what could be improved, or bring these points up during the follow-up discussion.

Estimated effort: ~6–8 hours (spread over 1–2 days). If you encounter technical issues or realise that performing a task properly would take significant time, feel free to apply a simplified or approximate solution.

Assignment

1. Select one target: CDK2 (e.g., PDB 1H1Q) or Thrombin (e.g., PDB 1PPB).
2. Prepare a set of a few hundred purchasable molecules (e.g., from ChemDiv, Enamine, eMolecules, Mcule). Include a few analogs of the co-ligand (from the PDB structure) and some diverse compounds in both structural similarity and number of rotatable bonds.
3. Perform docking with both classical docking software (AutoDock Vina or Uni-Dock) and ML-based (DiffDock or Boltz). Use the same binding site as ligand in the original PDB structure. Document key parameters (box center, grid size, constraints, etc.).
4. Please provide a ranked list of candidates. Which other properties would you determine to identify the best candidates for further experiments? Please determine these properties for your top 5-10 candidates.
5. Create a summary table and include it in the github repository or send it to us
6. What options are available to get more accurate free energy binding scores
    - Make a short list of some ways and the steps to obtain them; include links to relevant packages
      Note: you don't need to actually perform free energy binding calculations
7. What would be the next experimental steps, and what types of results would you look for?
8. Provide a minimal reproducible setup (module, script or notebook + environment file) and a short README explaining how to run it and how you would manage scaling, logging, and results tracking for a large-scale CADD workflow.

Submission

9. High-level summary: A single slide (for reading) with the task description, the obtained results, challenges, and the next steps
10. Code: github repository with README documentation

Deadline: 7 days after receiving the task
Tools: open-source software preferred; GPU optional, AI tools are allowed (in fact encouraged), but list the AI tools you used and summarise how you work with them.

Your submission will be evaluated based on scientific reasoning, clarity of presentation, reproducibility, and understanding of workflow scalability.

If the instructions are unclear, please feel free to reach out to us.