Introduction to Bioinformatics - 1MB438

Molecular Docking lab

December 2025

Philip Ullmann, Runze Zhang, Phong Lam
philip.ullmann@icm.uu.se, runze.zhang@icm.uu.se, phong.lam@icm.uu.se

1 Introduction

The goals of this practical session are to give you a first look at how molecular docking studies are carried out. This lab ties into material from **Lecture 6**. We will work with recently published data on a G protein-coupled receptor, the dopamine D2 receptor,[1] shown in **Figure 1**. Along the way, you will get familiar with ideas that come up all the time in molecular docking. The aim is to help you connect these ideas with the skills you have picked up in earlier labs and lectures.

Assessment of the practical session: Work in groups of 2 or 3 people. As you progress, take notes of answers to the questions highlighted in red in the designated black boxes, or on a paper sheet. At the end of the lab, you should approach one of the teaching assistants present, and ask them to check your answers. They will discuss incorrect answers with you. Once the assistants have approved your answers, they will mark your attendance and the practical session is considered completed.

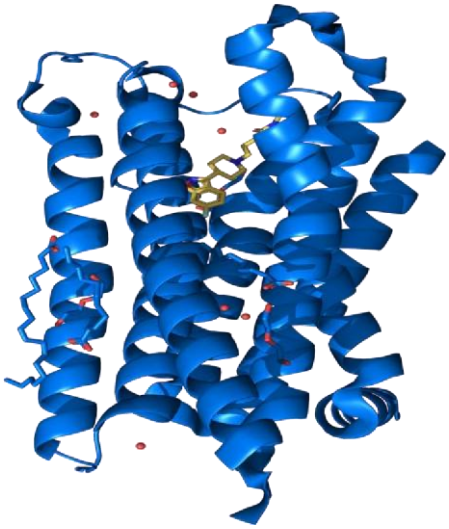


Figure 1: Dopamine D_2 receptor (blue) with its co-crystallized ligand, risperidone (yellow). PDB code: **6CM4**

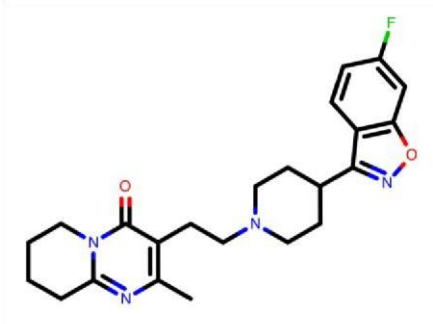


Figure 2: Chemical structure of risperidone, co-crystallized ligand in **6CM4** crystal structure.

2 Initial exploration of the binding pocket

Prior to any docking study, it is always good practice to have an initial glimpse at the binding pocket of the system you wish to investigate. In this section of the seminar, you will familiarize yourself with the binding pocket of the dopamine D_2 receptor.

Open PyMOL and load or fetch the crystal structure of the dopamine D2 receptor (**6CM4**). Zoom in on the ligand, risperidone. Show sidechains nearby to the ligand as lines or sticks.

You can either use the GUI or use these commands in the command palette:

- select ligand, resn 8NU
- select bdsite, (byres all within 4 of ligand) and polymer.protein
- show lines, bdsite
- zoom ligand
- label n. CA and bdsite, "%s %s" % (resn, resi)

Question 1: Try to identify three molecular interactions between the ligand and the residues of the protein receptor. Write down type of interaction, as well as the type of residue with its corresponding residue number. What, according to you, is the driving force for this particular ligand to bind the receptor in this way?

Think about the molecular interactions and the chemical structure of risperidone, as depicted in **Figure 2**.
(Hint: In a physiological environment, the protonation state of the molecule may change)

3 Redocking of risperidone

An initial hurdle to surmount in developing a model for docking studies is the redocking of the ligand in the crystal structure. To shift the focus from complex docking technicalities to analysis of structural bioinformatics data, we have already prepared all the following docking sessions for you. We sampled 10,000 orientational poses of several conformers of risperidone and ranked them accordingly using a scoring function.

Download the **risperidone-1.redocking.pse** file from Studium and open the file in a new PyMOL session. Scroll through the 300 different poses of risperidone and compare them with the crystal structure pose. For every docked pose, you should be able to find different energy contributions, as well as total energy, as estimated by the scoring function. Use **Figure 3** to support your answer for **Question 2**.

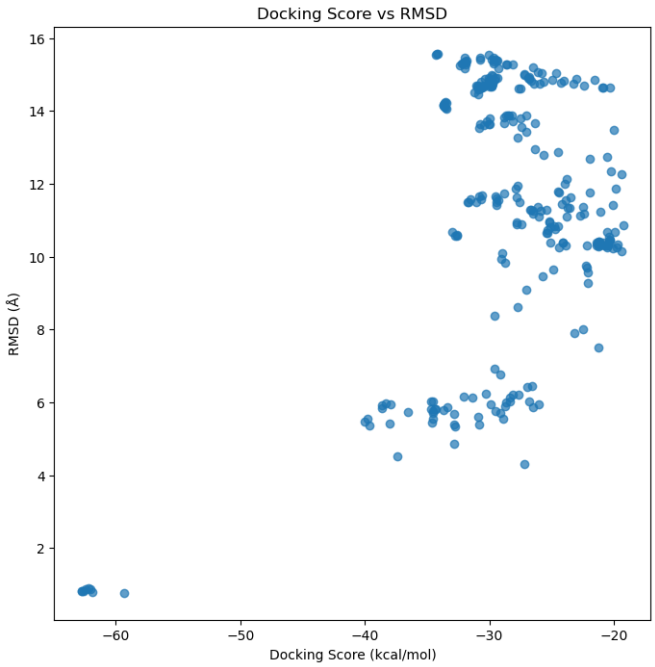


Figure 3: Scatter plot of the total energy (kcal/mol) of a risperidone pose versus the RMSD(A) of that pose with the˚crystal structure coordinates of risperidone.

Question 2: Do all docked poses geometrically align with the crystal structure binding pose, i.e. bioactive conformation[a]? The poses that align well, how is their total energy? Are molecules that are in a bioactive conformation always among the lowest in energy? Was the redocking of the ligand successful?

[a] Take into account that crystallization might affect the conformation of the ligand and/or protein

4 Decoys and ligand-enrichment

In this section, you will learn the concept of decoys and ligand-enrichment.

- **ligand**: a molecule with known affinity for the target receptor. Experimental data is usually available.
- **decoy**: a molecule that is expected not to bind, with similar properties[2] (logP, molecular weight, number of rotatable bonds, ...) of a ligand, but with a different topology (molecular connectivity). Experimental data on decoys are not always available, as they can be created virtually, but is recommended.
- **enrichment**: Figure 4 shows an empty ROC plot. The dashed curvy line is a model that does not enrich ligands or decoys and is therefore completely random. A curve to the left of the dashed line would enrich ligands, whereas a curve to the right would enrich decoys. We aim for the former when developing models to dock to.

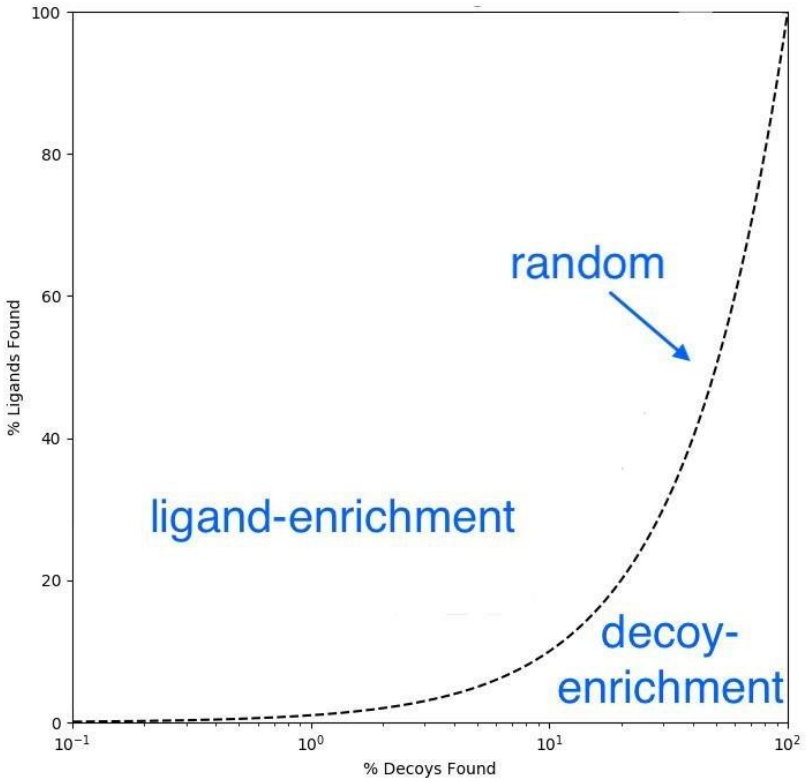


Figure 4: a generic ROC plot, demonstrating the regions of ligand and decoy enrichments. The **x-axis** is shown in **logarithmic form**, which helps draw attention to early enrichment. In practice, we usually expect true ligands to appear within roughly the first percentile of a docking campaign.

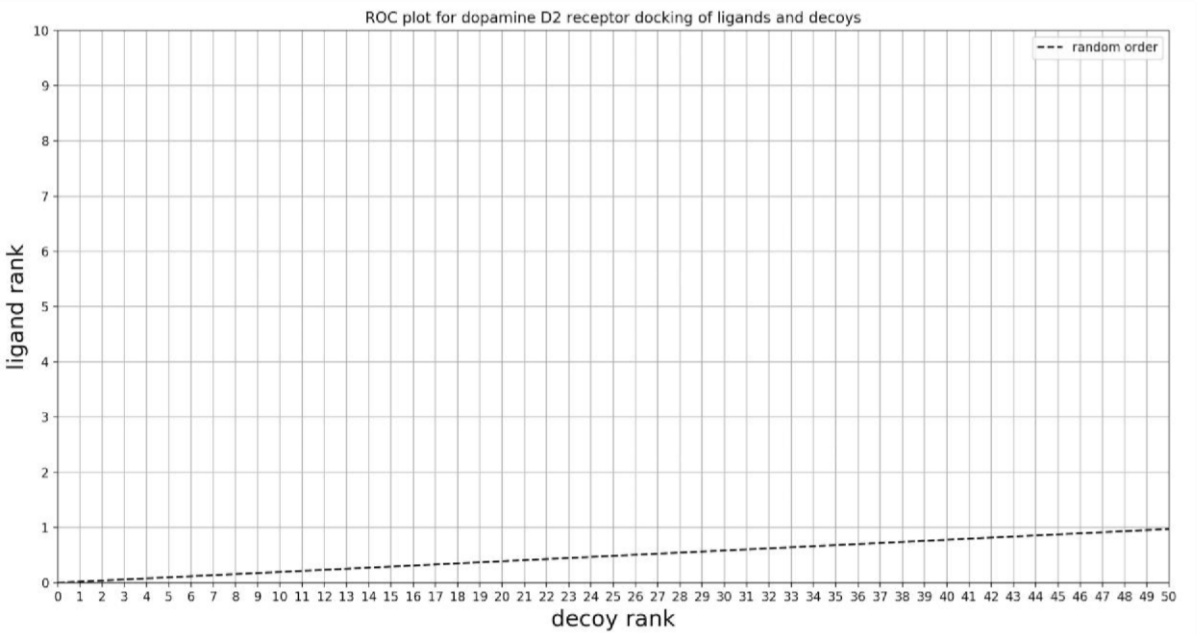
If the docking energies are reasonably good, the program should be able to recognize real binders in a set of molecules. We will use the ligands of the dopamine D_2 receptor, as it is a frequently studied GPCR and shown to be involved with several neurological disorders, such as schizophrenia,[3] bipolar disorder,[4] Parkinson's disease,[5] etc. All currently approved drugs were taken from the CHEMBL database and processed for a docking screen. The names of these compounds start with "CHEMBL". For each of these ligands, 50 decoys were generated. Decoys were selected from the ZINC database, and their name starts with "ZINC". Ligands and their corresponding decoys were docked, and results were extracted to support this section of the lab.

Download and open the **D2.example.structures.pdf** and **D2.drugsAndDecoys.properties.pdf** files from Studium to support your answer to **Question 3A**. Examine the chemical structures of the approved drugs from CHEMBL and their corresponding decoys. Use the distribution plots to help you answer question 3A.

Question 3A: What can you tell about the properties of the ligands, do they follow a specific rule? Is there any particular property in the ligands/decoys that supports your answer in Question 1 about molecular interactions? Do the decoys have similar properties to the approved drugs?

Open the **D2-1.decoy.docking.screen.pse** file in a new PyMOL session and scroll through the compounds. Have a look at the names of the compounds in **D2-1.decoy.dockingscreen.energies.txt**.

Question 3B: Based on the relative ranks and names of the first compounds, can you draw an estimate of the ROC curve on the grid below? Does this model enrich ligands, decoys, or is there no preference?



5 Docking screen of a virtual library

Once a reliable model of the protein has been established for molecular docking, we can advance towards a prospective screen of compounds. In this section of the lab, we have already docked a commercially available library containing 800.000+ fragments that can be found in the ZINC database.[6]

Open the **D2-1.fragment.dockingscreen.pse** from Studium in a new PyMOL session. Scroll through the top 100 compounds and analyze the interactions of the docked fragments with the binding site of the dopamine D2 receptor. You can also have a look at **D2-1.fragment.dockingscreen.energies.txt.**

Question 4: You work on a drug discovery project in a pharmaceutical company tackling Parkinson's disease. Pick three compounds from the screen you like and elaborate why. Think in terms of the molecular interactions and potential chemical expansion into a leadlike molecule. Keep your favorites in a session that you can show to the teaching assistant. See below on how to keep the favorite ones.

While iterating through the "states", you may see which state of the object you are at. For example, my favorite state is at 90/100.

PyMOL 3 PyMOL2

I can create a copy of it using the command below:

- create favorite_1, fragments, 90, 1

The first entry is the name of the object (favorite_1, favorite_2,etc). The second argument is the source for the object (we keep it as fragments for now). The third value is the source state (i.e. the state that you prefer). The last value is the destination state (we keep to 1 for now).

Version Information

The following Uppsala University staff contributed to writing this practical: Andreas Luttens, Jens Carlsson - 2018. Updated to PyMOL3 by Phong Lam and Philip Ullmann in 2025.

References

[1] Sheng Wang, Tao Che, Anat Levit, Brian K Shoichet, Daniel Wacker, and Bryan L Roth. Structure of the d2 dopamine receptor bound to the atypical antipsychotic drug risperidone. *Nature*, 555(7695):269, 2018.

[2] Michael M Mysinger, Michael Carchia, John J Irwin, and Brian K Shoichet. Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking. *Journal of medicinal chemistry*, 55(14):6582–6594, 2012.

[3] Stephen R Marder and Richard C Meibach. Risperidone in the treatment of schizophrenia. *The American Journal of Psychiatry*, 151(6):825, 1994.

[4] Jean A Frazier, Michele C Meyer, Joseph Biederman, Janet Wozniak, Timothy E Wilens, Thomas J Spencer, Grace S Kim, and Stephanie Shapiro. Risperidone treatment for juvenile bipolar disorder: a retrospective chart review. *Journal of the American Academy of Child & Adolescent Psychiatry*, 38(8):960–965, 1999.

[5] Angelo Antonini, Johannes Schwarz, Wolfgang H Oertel, Oliver Pogarell, and Klaus L Leenders. Long-term changes of striatal dopamine d2 receptors in patients with parkinson's disease: a study with positron emission tomography and [11c] raclopride. *Movement disorders: official journal of the Movement Disorder Society*, 12(1):33–38, 1997.

[6] Teague Sterling and John J Irwin. Zinc 15–ligand discovery for everyone. *Journal of chemical information and modeling*, 55(11):2324–2337, 2015.