

Tutorial : Docking small molecules in a receptor using GOLD

Context

Docking of small molecules can be performed on-line with several tools. For a recent review see [Sulimov VB et al. Curr Top Med Chem, 2021, 21:507-546](#)

Some of these tools are available through the public European Galaxy instance dedicated to cheminformatics: <https://cheminformatics.usegalaxy.eu>. There, you can use the [AutoDock VINA software](#) for instance.

At the Loria, we currently use the [GOLD](#) software from Cambridge Crystallographic Data Centre (CCDC). CCDC provides a free license to academics, in particular to CNRS and their co-workers. On our local platform, we have interfaced GOLD with a Galaxy instance that has been duplicated on a secure hosting provider for remote access <https://galaxy.gabek.fr/>.

GOLD enables flexible docking (AutoDock VINA also, but we did not conduct any comparison study between these two software). For each molecule, a set of conformers is produced, by default 50. Moreover, during the docking, certain movements of the receptor side chains are allowed when computing minimal energy per conformer. More information in the [CCDC GOLD Userguide](#).

Our case-study will concern Covid nsp13 protein (apo form PDB:7NIO) in line with the article presented in the seminar ([Islam MK et al., 2023](#)) and with the article referenced therein ([White MA et al., 2020, J Phys Chem Lett. 11:9144-9151](#))

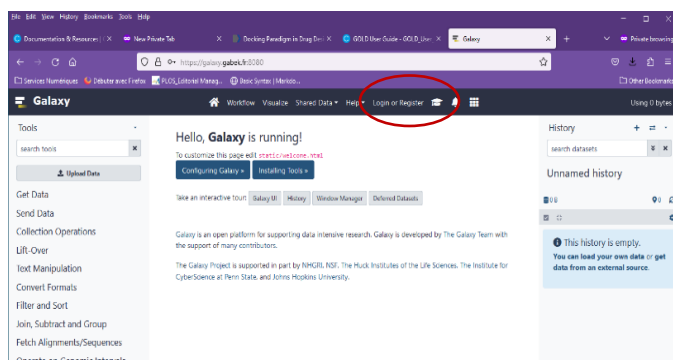
The top-65 ligands tested in the paper have been decomposed in 4 lists of 20 ligands. Each list contains the same 5 reference ligands and 15 additional ones. Each group of students will test GOLD with one of these lists and PDB:7NIO.

All necessary files for this tutorial will be provided as a compressed file : DataForGOLDTutorial-29nov2023.zip (903K), that you should download and unzip in a dedicated directory.

I. Connexion to Galaxy through the Gabek secure hosting provider.

Go to <https://galaxy.gabek.fr:8080/>

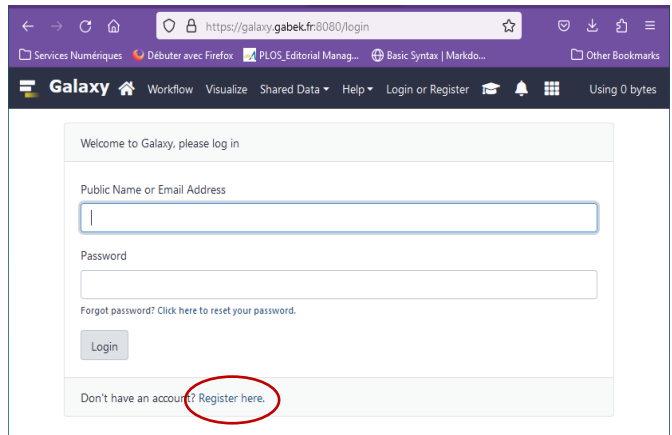
The site is asking you to sign in. Usernames and password are available from the text file 'List_Accounts_Passwords_MasterClassBrazil2023.txt'. Please wait until we distribute one username to each participant. Then use it and copy-paste the password.



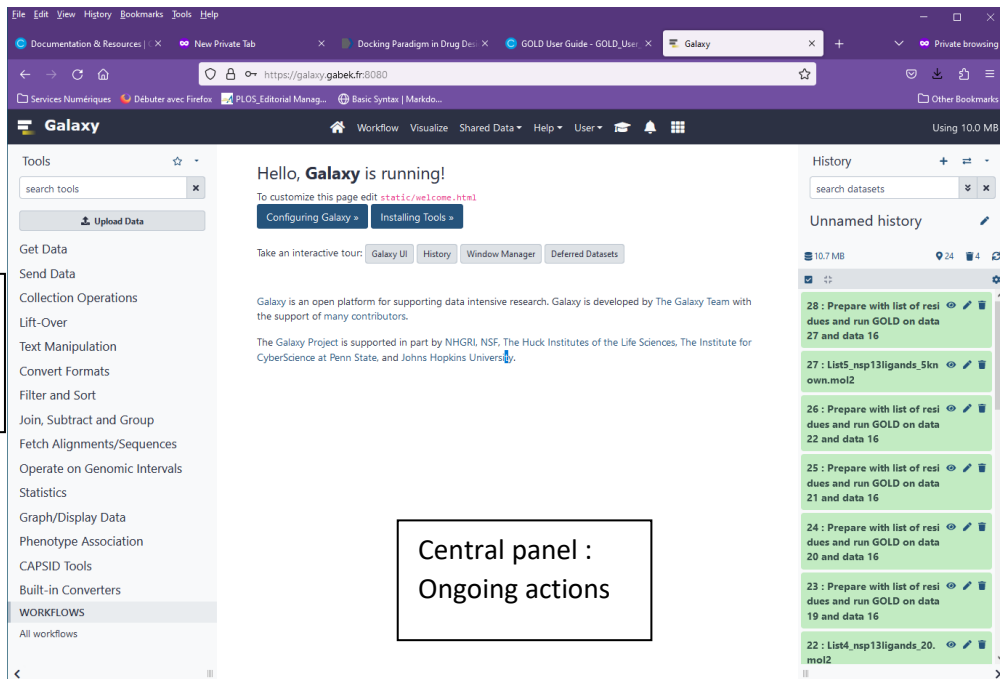
Then the Galaxy interface is displayed like here. IMPORTANT : Before starting to work, you need to **register locally** on Galaxy in order to keep track of your session history. This is important as the jobs will not be available immediately and you may need to logout and login again later. In that case, your local account will allow you to find all the jobs you have launched in an earlier session.

Click on Login or Register. This windows open and the first time you need to register and enter an e-mail address (which will not be used by this Galaxy instance to send you anymail as this server is not configured to send any mail to the users), a username and a password that you will use to log in as a registered user. **Note your username and password to register again later.**

Be patient, the gabek server is quite slow !



When you are logged in, explore the Galaxy interface:



Left panel :
The list of tools.

Central panel :
Ongoing actions

Right panel :
Job history and access to the results

II. Preparing and launching the GOLD jobs

1. Uploading input data on Galaxy

Left panel : **Get Data**

Upload from your computer (directory where you have the uncompressed zip file)

- nsp13_7nio_forGOLD.pdb
- one of the four lists of 20 ligands: 15 particular ligands + 5 common ligands. These lists will first be assigned among participants, don't take all the same one, e.g. List1_nsp13ligands.mol2)

Attention when filling the **upload form** -> specify the **file type (pdb or mol2)**

2. Launching GOLD with a list of residues

Left panel : **CAPSID Tools > Prepare with list of residues and run GOLD**

The corresponding form is displayed in the central panel with default filenames (the last files uploaded in your session).

- Check that you have **nsp13_7nio_forGOLD.pdb** in the protein field and **your list of ligands** in the ligands file.
- Keep **Nb Conformation = 50** (default value)
- Enter a list of residues by copying-pasting the list of residues from the file named **"list_residues_nsp13.txt"** provided in the compressed archive. Open the file with a text editor and copy-paste only the list. These residues are part of the binding pocket and are known to interact with the known ligands of nsp13 (see Fig.3 in the paper by White MA et al.). GOLD will use them as guides during the docking process.

GLU261 ASN265 GLY287 LYS288 SER289 HIS290 LYS320 LYS323 TYR324 ARG442 GLN404 GLU375 ASP374 ARG567

- Keep **Number of ligand per batch = 8** (default value) The program works with each batch in parallel to speed up the execution.
- Click on the **Execute** button

The job is created with an incremented number in the right panel. It will stay with a salmon color until it is finished. The execution can take time as it is not performed on Gabek itself but transferred to a cluster at the LORIA behind our firewall by an accredited program.

====**==**==**==**==**==**

We'll make a break in the tutorial at that stage and switch to the Pharmit tutorial (virtual screening with pharmacophore). We'll come back to this tutorial this afternoon.

====**==**==**==**==**==**

III. Analysis of GOLD results

- In the right panel, click on the terminated job. Note its **size**. The **"i"** button at the bottom provides information about the run (important when a bug has happened)
- Click on the **"eye"** icon on the right of the job name to download the compressed file containing the results of the job.
- Check also the function of the **"pencil"** icon for example to rename a job or add annotations.
- Fetch the results compressed file in your Download directory:

Typical filename: Galaxy3-[Prepare_with_list_of_residues_and_run_GOLD_on_data_1_and_data_2]

- Unzip to a new directory (option from your zip application)

By default the new directory will have the same name as the results compressed file. Keep this name

- Open the new directory: it contains a compressed file again
- Unzip this file but this time you can remain in the same directory (option "unzip here" from your zip application)
- One text file (bestranking.lst) and one directory (solutions) are produced.

- Rename the file and the directory by adding a prefix to recall the job number and the list of molecules tested. Example of prefix: Job3-List1_ . This will lead to the text file renamed as Job3-List1_bestranking.lst and the directory renamed as Job3-List1_solutions.

3. Analysis of best-ranked molecules

- Open the text file (for example Job3-List1_bestranking.lst) with a text editor (for example NotePad++)
- One row per molecule, 9 fields per row. The first one is the GOLD score, the next 6 are descriptors of the run that we will not explain due to the limited time (refer to GOLD user guide if you are interested), the last two fields contain the filename of the solution corresponding to this row and the pubchem ID of the molecule.

In the provided spreadsheet with the contents of the lists of molecules (Results_nsp13ligands.ods or .xlsx), add a new column to the list you have used and copy-paste the score (field_1), and the filename of the solution (field_8).

- Rank by decreasing GOLD score and check the rank of the 5 known ligands.
- Interpret and conclude.

To compare your results with the other groups, report also the ranking of your ligands in the [shared Spreadsheet on GoogleDrive](#).

IV. Visualization of best solutions with PyMOL

1. Opening a solution file with PyMOL

Open PyMOL. To visualize the docked molecules you need to load two files in PyMOL : the structure you have used for the docking (nsp13_7nio_forGOLD.pdb) and one solution available from your GOLD 'solutions' directory (take for example the best ranked solution, usually NADH).

Open these two files in PyMOL and take some time to explore the various functionalities of this software.

The PyMOL window is divided in a top and a bottom panel.

- The top panel displays the commands that are executed by PyMOL during your session as well as shortcut buttons for frequently used commands (on the right)
- The bottom panel contains the main “viewing window”: you can click and drag with the left mouse button to rotate the molecule. Dragging with the right mouse button will allow you to zoom in and out. Finally, dragging with the middle mouse button will translate the structure in the X-Y plane of your monitor.
- On the right a smaller panel contains a table that displays all the structures and selections you have considered in your session (“molecular objects”, one row per object) and allows you to modify them according to 5 menus:
 - A for Action (for example to rename the object selected in a row)
 - S for Show (to modify the display of the object selected in a row)
 - H for Hide (to hide everything or part of the object selected in a row)
 - L for Label (to activate or modify the label of the object selected in a row)
 - C for Color (to color in a given color the object selected in a row)

- At the bottom is a summary of all actions you can do by combining the three buttons of the mouse and its wheel with the Shift, Ctrl and CtrlShift tabs. It also contains additional buttons for making molecular movies.

For more information about PyMOL see the [document prepared for a bootcamp by the Fitskee lab](#) (Mississippi State university). They also provide a very good [tutorial for Pymol](#)

2. Visualization and comparison of docked ligands

From the ranked list of ligands open the best docked poses. Color them with different colors and see how they superpose in the nsp13 binding pocket.

Visualize also a ligand with the minimal score -> can you propose a reason why this ligand has a poor score with GOLD ?

Highlight the binding residues of the protein used for the docking. For this you can create a PyMol object by copy-pasting this command in PyMol command textfield.

```
create bindingsite, (resi 288,289,290,320,374,375,404,442,443,540,567)
```

- Use the 'S' menu to display residues as 'stick' : this shows the complete structure of the residues
- Use the 'C' menu to color them as you wish
- Use the 'L' menu to label each residue and recognize it from the list

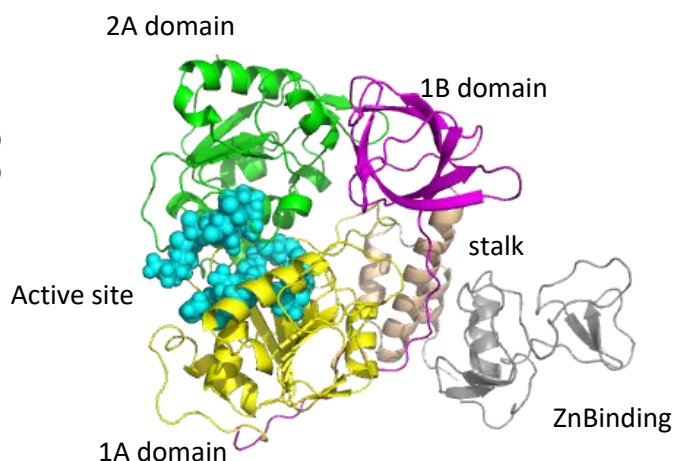
Can you identify the most important ones : involved for all the ligands ?

```
GLU261 ASN265 GLY287 LYS288 SER289 HIS290 LYS320 LYS323 TYR324  
ARG442 GLN404 GLU375 ASP374 ARG567
```

3. Optional : commands to color the nsp13 protein according to its domains

Nsp13 is an helicase, critical for viral replication, and very much conserved across coronavirus strains. It contains 5 domains from N-term to C-term : Zinc Binding, stalk, 1B domain, 1A domain, 2A domain.

```
create ZnBinding, (resi 1:98)  
create stalk, (resi 99:149)  
create 1BDomain, (resi 150:249)  
create CoreDomain1A, (resi 250:440)  
create CoreDomain2A, (resi 441:603)  
show cartoon, ZnBinding  
color grey, ZnBinding  
show cartoon, stalk  
color wheat, stalk  
show cartoon, 1BDomain  
color magenta, 1BDomain  
show cartoon, CoreDomain1A  
color yellow, CoreDomain1A  
show cartoon, CoreDomain2A  
color green, CoreDomain2A
```



CONCLUSION

The ranking of the ligands is not strictly reproducible. This is due to the GOLD algorithm. Nevertheless, well-ranked ligands remain well ranked even if the rank is not exactly the same.

Visualisation with PyMol allows to see which residues are in contact with the ligand. This can be automated by various programs. In particular PLIP : Protein Ligand Interaction Profiler available at Dresden University (<https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index>).
