GENETIC OPTIMIZATION FOR LIGAND DOCKING
G.O.L.D. TUTORIAL

Author: Vladyslav Kholodovych, Ph.D.
University of Medicine & Dentistry of New Jersey
Robert Wood Johnson Medical School
675 Hoes Lane
Piscataway, NJ 08854 U.S.A.
(732) 235-3229 phone
(732) 235-3475 FAX
kholodvl@umdnj.edu
http://www2.umdnj.edu/~kholodvl/

You did already all the dirty work by preparing files for docking in Dock and Autodock tutorials. To save time we will skip these steps in this tutorial and concentrate only on the GOLD related procedures.
We will use GOLD to compare the docking results from different docking programs, look for consensus and try to choose the best technique for our research project.

You have already two files extracted from the PDB structural complex 1EZQ for the Dock tutorial:
fxa_grd.mol2 – minimized structure of protein
rpr.mol2 – modified and minimized ligand structure

Create a new directory **gold** in your project directory and enter it.

mkdir gold
cd gold

Copy the original crystal structure complex, receptor and ligand files created in the Dock tutorial to the new gold directory
(assuming your Dock files are in projects/dock6 and you are in project/gold directory now)

cp ../dock6/pdb1ezq.ent ./ original PDB complex

cp ../dock6/fxa_grd.mol2 ./ minimized protein

cp ../dock6/rpr.mol2 ./ refined and minimized ligand

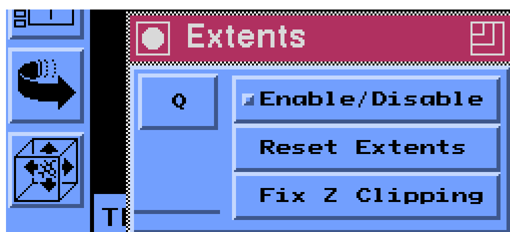
For GOLD docking you need to know coordinates of the center of binding pocket of the receptor. Because in this tutorial we are using a real crystal structure of a receptor–ligand complex we don't need to worry about that. Just take coordinates of any atom in the central part of RPR molecule (ligand) from the original PDB file pdb1ezq.ent

Open pdb1ezq.ent in Sybyl.

Biopolymer ->Read

Make sure **DO NOT CENTER** molecule.

Because we are working with real coordinates the molecule may be far away from the visible area in Sybyl. To bring molecule on the screen find the icon with cube and arrows at the bottom of left side menubar. Click on this icon and then on Reset Extents



Now you can see molecule on the screen. Zoom it in to a comfortable size.

Extract the ligand from complex as we did it in DOCK tutorial or copy it from Dock directory. We need it later for comparison of docking results with the original conformation from the crystal structure.

Build/Edit -> Extract…
in Atom Selection
click on Substructure and select RPR1
OK
OK

In Molecular Area
choose M2 <empty>
OK

Zap protein
Build/Edit
Zap (Delete) molecule
Select M1
OK

Because for the later comparison we need to preserve the atoms' coordinates. Only fix atom types and bonds as we did in the dock tutorial (and presumably at the beginning of this tutorial. Remember we just skipped this step at the beginning for simplicity and transferred the already prepared files from autodock) and do not minimize the ligand structure.

Name extracted molecule
Build/Edit-> Name Molecule
rpr_xray

and save as
rpr_xray.mol2

Change color of the ligand
View->Color-> By atom type

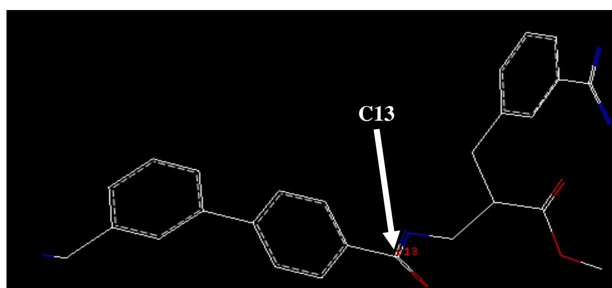
Notice any atom in the middle of the ligand, roughly equidistant from the edges. This atom will be a center of sphere in GOLD docking.
Find the name of this atom

View->Label->Atom Name

Click on this atom in the main window to select it
and click **OK** in Atom Expression

Write down this name and close Sybyl.

Find out coordinates of this atom
In the terminal window type
more rpr_xray.mol2



Partial listing of rpr_xray.mol2 file

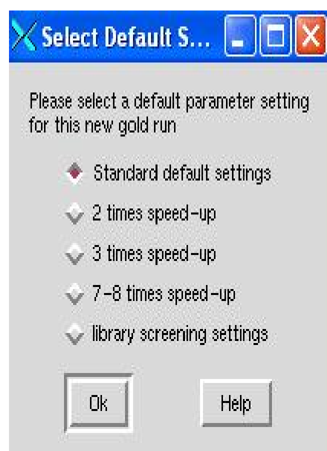
11	C11	7.2490	1.7840	19.7550	C.ar	1 RPR1	34.2300
12	C12	7.7700	3.0800	20.0090	C.ar	1 RPR1	33.1900
13	***C13***	***5.2730***	***4.5350***	***22.5310***	***C.2***	***1 RPR1***	***33.0400***
14	N14	5.4850	5.8740	22.2930	N.2	1 RPR1	28.9900
15	C15	4.9130	6.9410	23.1240	C.3	1 RPR1	25.5600

(**CAUTION! Atom name and coordinates may differ for your file**)

Alternatively, you can use coordinates from Autodock parameter file rpr.fxa.dpf for the small molecule center.

To start GOLD type
gold

From the dialog choose Standard default settings mode and press OK



The window on the right is the main configuration window of GOLD where you set up all docking jobs.

Press on ***Add/Delete Ligand*** button

In <u>Ligand Editor</u>
click on **ADD**

new window will appear

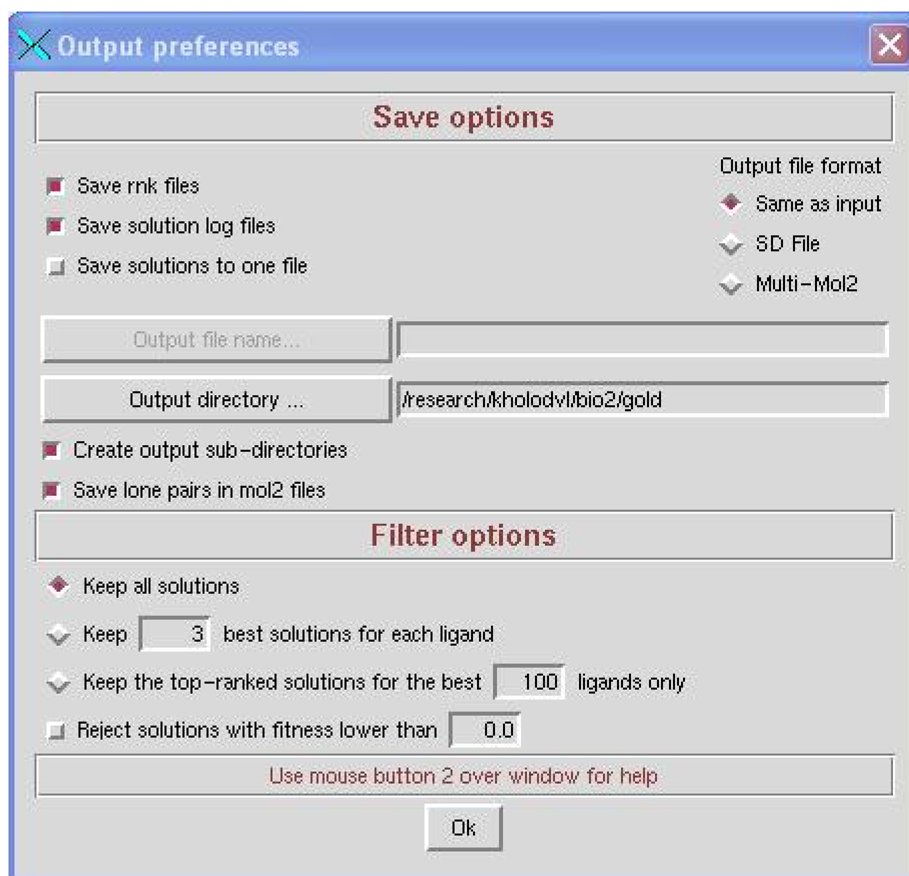
in <u>Add Ligand</u>
 click on **Filename** and choose rpr.mol2

With slider "***No Dockings***" you can choose how many separate docking runs you want to perform. Maximum is 50. For this tutorial we leave it at 10 (default value). Click on **ADD** and then **END**

GOLD 2.1
G.O.L.D. Genetic Optimisation for Ligand Docking
Run Default Save&Exit Submit&Exit Exit Configuration File gold.conf Help
Select editing panels: Input Fitness Function GA Parallel
Input Parameters and Files
Add/Delete Ligand Number of Ligands 1 Protein fxa.mol2
Set atom types: Ligand Protein
Allow early termination If top 3 solutions are within 1.5 Angstroms R.M.S.D.
Define active site from: Point Atom File Origin 6.547 4.046 21.307
Active site radius 10.0 Detect Cavity
Covalent Display ... Output ... Edit Parameters Parameter File DEFAULT
Fitness Function Settings
GoldScore ChemScore User Defined Score
Fitness Flags Edit Constraints Number of Constraints 0 Number of Ligand Bumps 0
Torsion Angle Distributions
Use Distributions Edit Distributions Distributions File DEFAULT
ChemScore Input Parameters
ChemScore File: chemscore.params
Read hydrophobic fitting points
Genetic Algorithm Parameters
Population Size 100 Selection Pressure 1.1 Number of Operations 100000
Number of Islands 5 Niche Size 2 Migrate 10 Mutate 95 Crossover 95
Click mouse button 2 over window for help.

New line will appear in the <u>Ligand Editor</u> indicating the ligand name and the number of dockings.
Press **OK** to return to the main window.

Click on ***Output…***



Check at
Save rnk files
Save solution log files
Output file format : Same as input
Output directory: your gold directory (project/gold)
Create output sub-directoires
Filetr options: Keep all solutions
Uncheck at
Save lone pairs in mol2 files

Click **OK** to return to the main window

Now add the protein
Click on **Protein** and add fxa_grd.mol2

Allow early termination if top 3 solutions are within 1.5 Angstrom RMSD

To define an active site in the section "Define active site from"
check at
Point and add coordinates that you found from rpr_xray.mol2 or an autodock configuration file (rpr.fxa.dpf).

Active site radius: 10
Check at
Detect cavity to allow GOLD automatically determine a free volume for docking within sphere with radius of 10 angstrom around defined Point.

For fitness function check Gold Score.
Leave all other parameters to their default values.

You are ready to submit the GOLD docking.
It is possible to run GOLD docking in the interactive mode where you will see all warning messages and output of running processes or in a hidden mode, the batch.

First method is a good choice if you are running GOLD for the first time and you do not know anything about your ligand and receptor interaction or you are not sure that binding pocket is defined correctly.
Run GOLD in the interactive mode by clicking on **RUN** button.
You will see several warnings about wrong types of atoms and bonds. This happens because GOLD and Sybyl have slightly different definitions for atoms. If you do not see any fatal error messages then let GOLD to take care about all necessary changes. Click on **Dismiss** button after checking the outputs.
After 2-3 minutes when you are sure that GOLD is working normally stop the interactive run and make a preliminary analysis. There is no reason to wait till GOLD finishes all dockings and then to find out that the whole procedure was set up wrong!
Press **Interrupt GA** and then **Dismiss** to return to the main window.

Open a new terminal console and go to your **gold** directory.
In this directory should be the protein, ligand and complex files and files produced by GOLD.
The most important file for a preliminary analysis is fit_pts.mol2. This file contains information about a binding pocket in which GOLD puts the ligand during the docking. To define it GOLD examines free space by filling all accessible volume with probe (dummy) atoms.
Open this file and a ligand extracted from the original pdb complex, rpr_xray.mol2 in Sybyl. If the binding cavity is set correctly, the majority of the ligand and fitting points should overlap.
If the ligand is covered by fitting points only partially (less than 10-15%) or it is far away from them, it means that a center of the binding pocket was set wrong and it should be redone. Alternatively, you can increase the radius of GOLD sphere to cover the ligand, but it is always better to set center more precisely and keep the docking cavity smaller.

You should see a similar picture on the screen.

(**<u>HINT</u>:** open the ligand molecule first. Position it in the center of the visible area, zoom if needed, and then open fit points).

Close Sybyl.

The drawback of the interactive mode besides that it is slow is that you can not log out from your session and have to wait until GOLD finishes a docking job.

Fortunately, you can run GOLD in a batch mode. In this case, you can even log out from the computer and GOLD will continue to run in a background. It is very useful if you have a long queue of ligands scheduled for docking so you can submit your job overnight instead of overload server resources with a computationally excessive job during the day time.

To run docking job in the batch click on **Submit&Exit** button.

You can save a configuration for a later submission if you click on Save&Exit. All settings will be stored in gold.conf file. Next time gold will read settings from this file so there is no need to run through the configuration procedure again.

While running, GOLD produces several files that can be found in the gold main directory.
The most important are:

gold.conf	settings for GOLD docking
gold.err	error output, contain log if GOLD can not start properly or is terminated. Always check this file for errors at the beginning after 2-3 minutes of the GOLD run and later after job finishes.
gold_protein.mol2	a protein molecule generated by GOLD and used for docking
gold_rpr.mol2	a ligand molecule generated by GOLD
fit_points.mol2	a detected cavity of the binding pocket in which ligands are docked
bestranking.lst	a listing of the best docking results for each ligand docked

A single GOLD run usually takes 1-2 minutes. There is no difference between number of molecules and number of dockings, i.e. 5 molecules set for 2 dockings each take roughly the same amount of time as one molecule docked 10 times.
So our docking should be done in 20-25 minutes.

ANALYSIS OF THE RESULT AND VISUALIZATION

We chose to save the docking output for each ligand in a separate directory. Usually this directory started with a ligand name by adding suffix "_m1". Accordingly docking results for RPR ligand are placed into a directory rpr_m1.

Enter this directory and check its content:
cd rpr_m1
ls -la

Two sets of files are generated by GOLD:
gold_soln_rpr_m1_X.mol2
ranked_rpr_m1_X.mol2

where X is a number from 1 to the total number of dockings, in our case from 1 to 10.

The gold solutions set is a set of conformations/orientations of a molecule GOLD placed in the binding pocket.
Ranked set is not real files. This is a set of symbolic links to the first set but sorted accordingly to the GOLD fitness function. The same information about ranking you can find in the rank file rpr_m1.rnk

To open all docking results in Sybyl you need to concatenate them in one file.

cat ranked_rpr_m1* > ranked_all.mol2

Analyze how close the docking results are to the original ligand in the crystal structure in Sybyl.

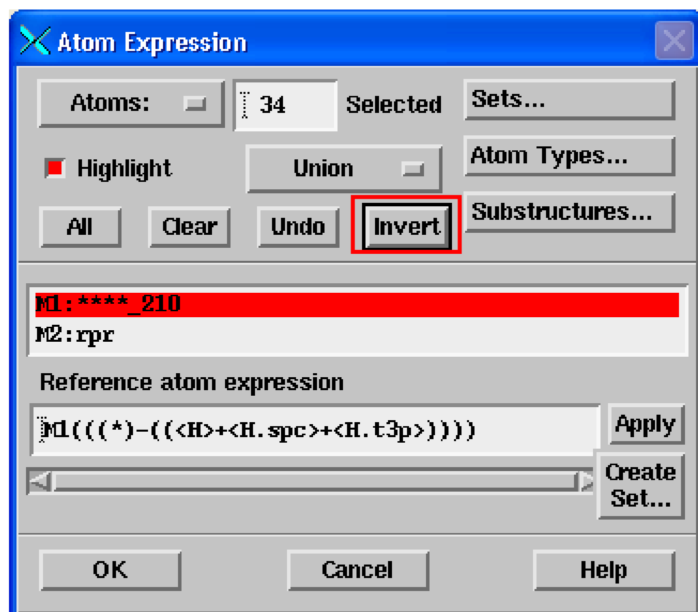
Open the original ligand extracted from PDB rpr_xray.mol2
and the best docking result based on the GOLD scoring function ranked_rpr_m1_1.mol2

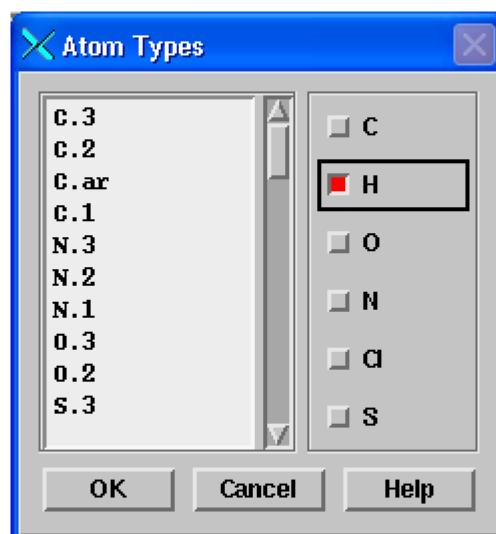
Notice how similar they are in terms of backbone orientation and a placement of functional groups (amidino, amide, carboxy).
Calculate RMSD between two molecules (based only on non hydrogen/heavy atoms).

In Sybyl

Analyze->Match…

IMPORTANT! You should always match docking resulting structures to a crystal structure. Make sure that the first molecule in the matching is your crystal structure, not *au contraire*.





in Atom Expression
select first molecule
M1: rpr_xray
click on **Atom Types…**
and **Check** at
hydrogen atoms in the right column
OK
Then click on **Invert,** you see that all non-hydrogen atoms are highlighted.
OK

Now select second molecule
M2: gold soln rpr m1
click on **Atom Types…**
and **Check** at
hydrogen atoms in the right column
OK, then **Invert**
OK

Go to Sybyl terminal window and check RMSD for these two ligands.

Repeat the same for the rest docking solutions and fill out a table.

Table 1. Two best dockings for GOLD, DOCK, and Autodock

File Name	Rank in Gold, DOCK or Autodock	RMSD with a crystal structure	Similarity/dissimilarity in the placement of the functional groups
Dock 1			
Dock 2			
Autodock 1			
Autodock 2			
Gold 1			
Gold 2			

Compare results with Dock and Autodock results.
Based on your analysis specify which docking method produced the best fitting. Why? Is it the top scorer of GOLD, Dock or Autodock solutions or any other conformer?

Which technique is more accurate?