TUTORIAL: DOCKING WITH MOE AND GOLD

LAPATINIB STUDY

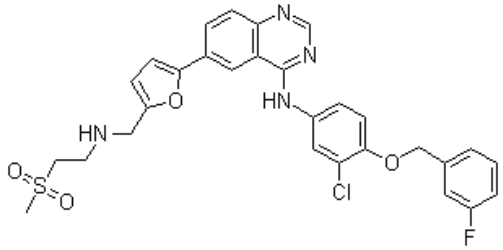
Author: Vladyslav Kholodovych, Ph.D.
University of Medicine + Dentistry of New Jersey
Robert Wood Johnson Medical School
675 Hoes Lane
Piscataway, NJ 08854 U.S.A.
(732) 235-3229 phone
(732) 235-3475 FAX
kholodvl@umdnj.edu
http://www2.umdnj.edu/~kholodvl/

INTRODUCTION

In this tutorial you will study the docking of a tyrosine kinase inhibitor that is a potent dual inhibitor of epidermal growth factor receptor (EGFR, ErbB-1) and ErbB-2 and approved by FDA as a drug for treatment solid tumors such as breast cancer.

History and Overview
(Excerpt from article on Wikipedia http://en.wikipedia.org/wiki/Lapatinib as on 11/20/2008)

N-[3-Chloro-4-[(3-fluorophenyl)methoxy]phenyl]-6-[5-[(2-methylsulfonylethylamino)methyl]-2-furyl]quinazolin-4-amine



Lapatinib (INN) or lapatinib ditosylate (USAN) (Tykerb/Tyverb, GSK) is an orally active chemotherapeutic drug treatment for solid tumours such as breast cancer. During development it was known as small molecule GW572016. Patients who meet specific indication criteria may be prescribed lapatinib as part of combination therapy for breast cancer. Pharmacologically, lapatinib is a dual tyrosine kinase inhibitor that interrupts cancer-causing cellular signals.

Status
On March 13, 2007, the FDA approved lapatinib in combination therapy for breast cancer patients already using capecitabine (Xeloda®, Roche).

Mode of action
Lapatanib inhibits the tyrosine kinase activity associated with two oncogenes, EGFR (epidermal growth factor receptor) and HER2/neu (Human EGFR type 2). Over expression of HER2/neu can be responsible for certain types of high-risk breast cancers in women. Lapatanib inhibits receptor signal processes by binding to the ATP-binding pocket of the EGFR/HER2 protein kinase domain, preventing self-phosphorylation and subsequent activation of the signal mechanism

Clinical application
Lapatinib is used as a treatment for women's breast cancer in patients who have HER2-positive advanced breast cancer that has progressed after previous treatment with other chemotherapeutic agents, such as anthracycline, taxane-derived drugs, or trastuzumab (Herceptin,Genentech).

A 2006 GSK-supported randomized clinical trial on female breast cancer previously being treated with those agents (anthracycline,a taxane and trastuzumab) demonstrated that administrating lapatinib in combination with capecitabine delayed the time of further cancer growth compared to regime that use capecitabine alone. The study also reported that risk of disease progression was reduced by 51%, and that the combination therapy was not associated with increases in toxic side effects.

The results from studies like these leave lapatinib with its somewhat complex and rather specific indication — use only in combination with capecitabine for HER2-positive breast cancer afflicted women who have completely responded to previous chemotherapy with anthracycline, taxanes and trastuzumab.

A number of studies are underway attempting to evaluate the efficacy of lapatinib as a first-line therapy for HER2-positive cancer. As of 2007 they have only progressed to Phase II trials.

PDB depository has two x-ray structures with lapatinib: a molecular complex of EGFR domain (PDB ID: 1xkk) and more recent entry with lapatinib co-crystalized with ERbB4 kinase (PDB ID: 3bbt). Unfortunately in the last complex the molecule of lapatinib was partially truncated during crystallization. Therefore we will work with an older EGFR domain complex.

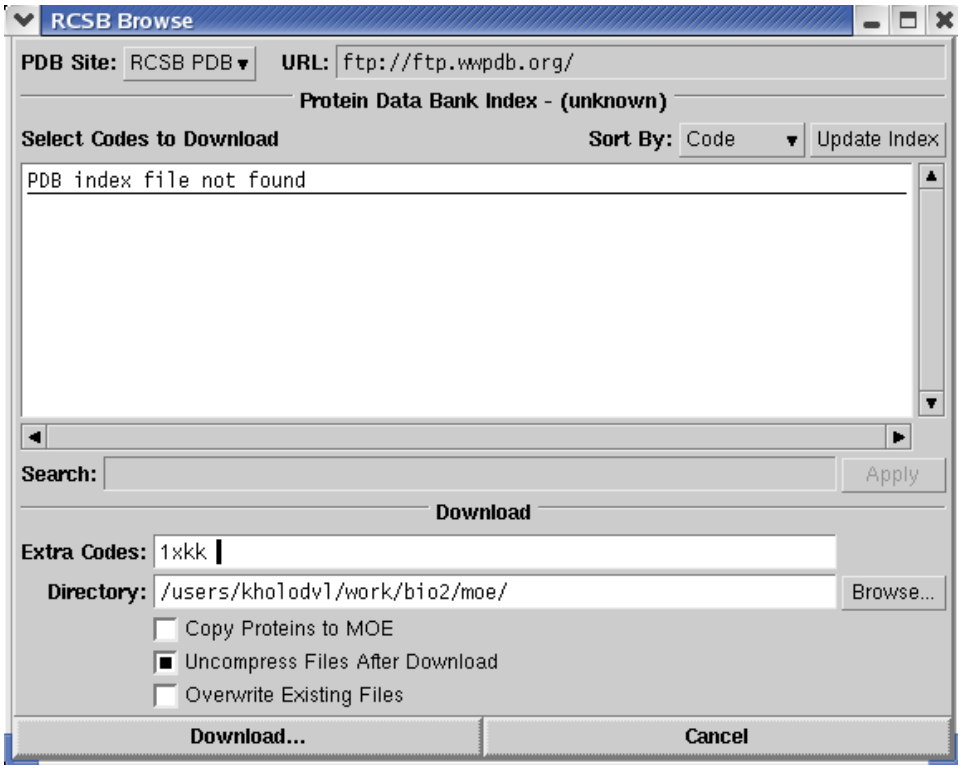
Primary citation:
Wood, E.R., Truesdale, A.T., McDonald, O.B., Yuan, D., Hassell, A., Dickerson, S.H., Ellis, B., Pennisi, C., Horne, E., Lackey, K., Alligood, K.J., Rusnak, D.W., Gilmer, T.M., Shewchuk, L. (2004) A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. Cancer Res. **64:** 6652-6659

PART I. DOCKING WITH MOE

Create a new directory **moe** in your project directory and enter it.
mkdir moe
cd moe

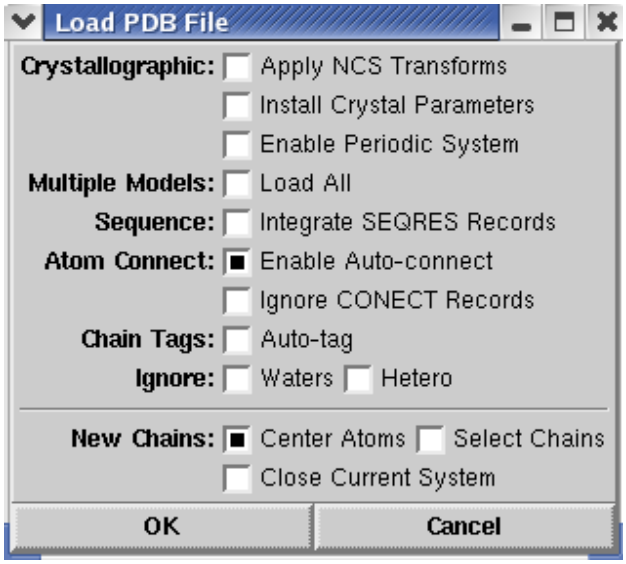
Start MOE program by typing
moe

Download molecular complex of EGFR domain with lapatinib (PDB ID: 1xkk)
In MOE main window
File->RCSB Download
Extra codes: **1xkk**
Select box **Uncompress Files After Downloading**
Click on **Download...** button and in the next window on **Start.**
When download is complete exit the dialog windows by clicking **Close** and **Cancel**.



Open file in MOE
File -> Open
Select pdb1xkk.ent OK

Check at
Enable Auto-connect and **Center Atoms** boxes and uncheck at **Ignore CONECT Records**



For easier visual inspection mark chains by color
Render->Color->Chain

A protein (fortunately it's only one chain!) turns one color and the ligand/ligands the other.
You may notice that the chain 2 consists of 3 fragments. Let's check what they are.

Open Sequence editor
Window->Sequence Editor
As we know already the protein has only one chain – the amino acid sequence is in Chain 1
Chain 2 has three residues: PO4 PO4 FMM
Chain 3 is a bunch of water molecules

It turns out that the chain 2 has a ligand (FMM) and two anions of phosphoric acid. We do not need phosphate anions for our exercise as well as water. Let's delete them.

Click on chain 3 with the right mouse button and select **Delete selected chain**, then **OK**.
Now let's get rid off anions.

Click on PO4 with the mouse left button, you may need to use Shift or Ctrl to highlight both residues.
Then **Edit->Delete Selected residues**
OK

Now only the protein (chain 1) and lapatinib (chain 2) remain on the screen.

INTERACTION ANALYSIS – SURFACES, POCKET, INTERACTIONS

Let's take a look how ligand binds to a protein and what part of the ligand contributes the most to interactions with the receptor.

In main window
Compute->Surfaces and Maps
Surface: Gaussian Contact
Name: LP Interactions
Atoms: Receptor Atoms
Now go to Sequence Editor and click on Chain 2 – ligand atoms

Return to Surfaces and Maps window and choose for **Near** field **Selected Chains**
Click **Apply**.

We created a surface of the receptor binding pocket that is color-coded according to the interactions with the ligand.

Let's hide atoms of the protein and examine only ligand binding site.
In main window **Render->Hide-> All**

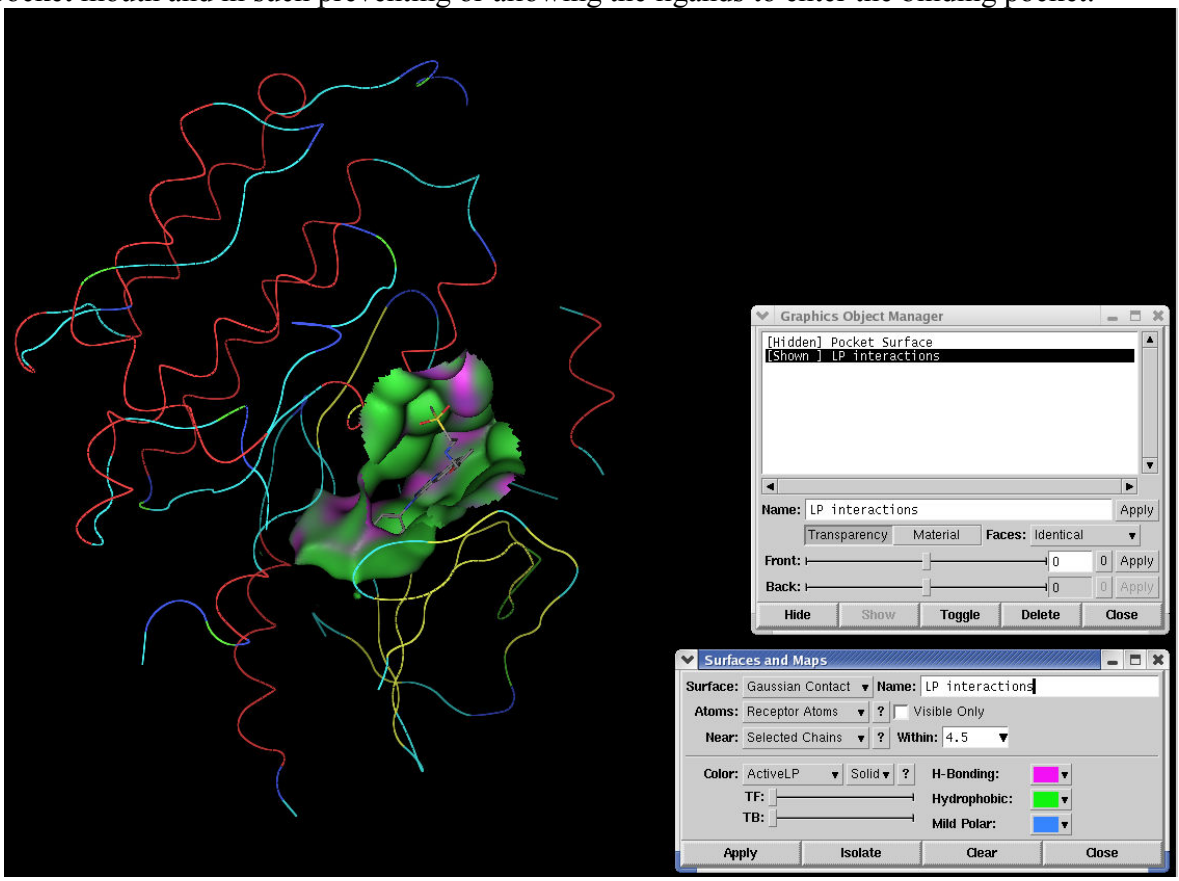
Now visualize the ligand atoms.
In sequence editor click on FMM with the mouse right button
In the menu select
Atoms-> Show

It will bring the ligand atoms on the screen.

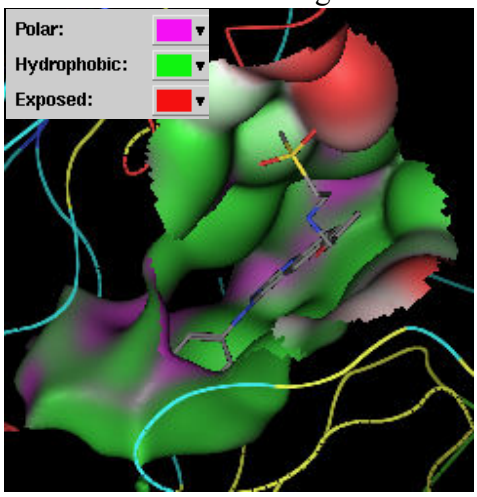
Now change the view of the ligand.
Click again on the FMM
Render->Stick
In the side bar of the main window click on **Color->Element** to display atom types of the ligand.

Display the protein in a schematic view.
Click with the mouse right button on the Chain 1
Backbone->Line

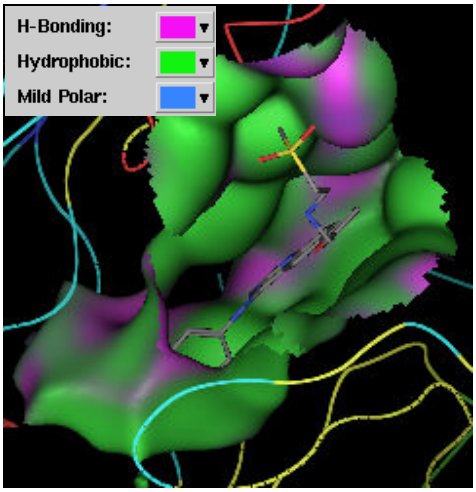
You can see that the binding pocket is in the middle of the protein between several alpha helices highlighted in red.It has an opening on the protein surface through which ligands enter the pocket. On the side of the opening (at the bottom on the picture below) there is a beta-sheet region (highlighted in yellow) which is very common for many receptors and sometimes serves as a lid for closing and opening the pocket mouth and in such preventing or allowing the ligands to enter the binding pocket.



Let's take a closer look on the ligand and the pocket



A



B

Molecule of Lapatanib in the ATP-binding pocket of the EGFR/HER2 protein kinase domain.
A. Surface of the pocket is color-coded by the nature of the pocket B. Surface is color-coded according to the character of the ligand protein interactions.

As you see the pocket is a profound groove which main function is binding the ATP for further phosphorylation. The ligand sits deep inside the pocket with its substantial part buried in the protein. Pocket surface is mostly hydrophobic (green color on both panels A and B) and it matches well with the hydrophobic rings of the lapatinib. The polar hydrophilic tail of the lapatinib (sulfur and oxygen atoms) on the other side of the molecule cannot enter the pocket and remain on its external surface exposed to interactions with a polar solvent (water molecules).

Notice the orientation and character of interactions between the ligand and the protein. You will need to use this knowledge later when you are going to analyze docking results.

Before proceed to the next step save the protein and the ligand in separate files for further work.

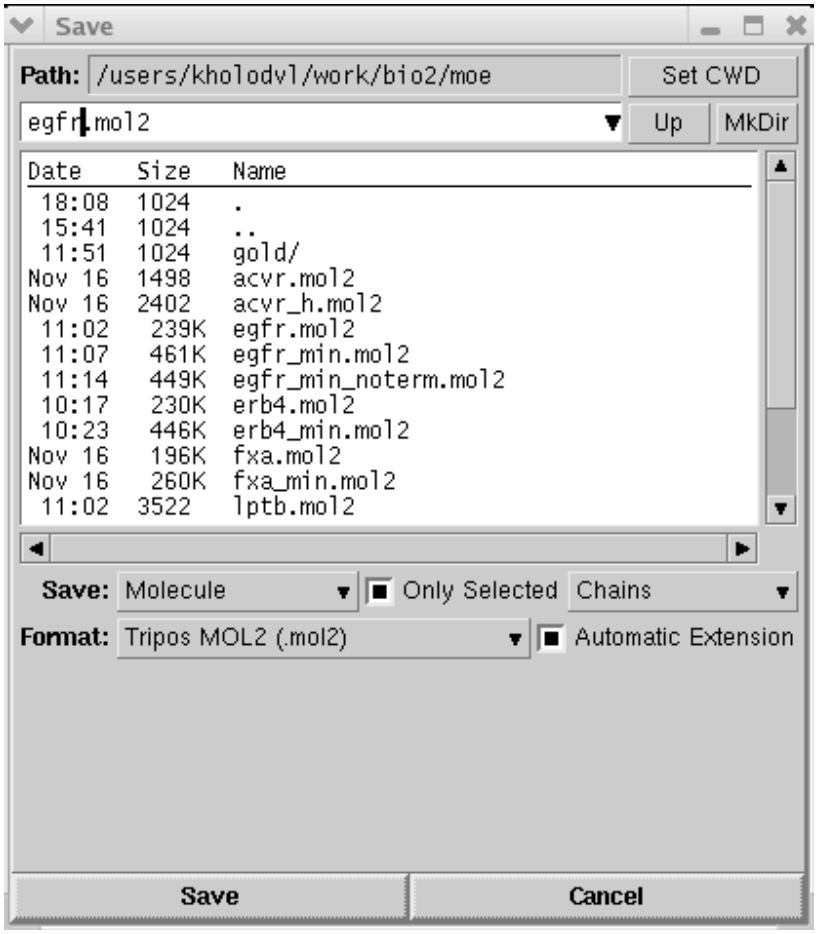
Click on chain 1 and then in Sequence Editor (SE) window
File-> Save
In the new window:

Save : Molecule

check at **Only selected** and make sure it refers to **Chains**

Format: **Tripos Mol2**

Name molecule egfr.mol2
and click on **Save**.



Now select Chain 2 in SE:
File-> Save
In the new window:
Save : Molecule
check at **Only selected** and make sure it points to **Chains**
Format: **Tripos Mol2**
Name molecule lptb.mol2
and click on **Save**.

Now close everything – in a side bar click **Close**

Well done, you dig the ore already now let's make some gold out of it! Okay, okay, we agreed to start with MOE - let's make some MOE then.

PROTEIN PREPARATION

Open the protein file that you have just saved
Open ->File
egfr.mol2
You may notice that some fragments of the protein sequence are missing. It's not unusual that very flexible regions like external loops are not crystallized well and such they are omitted from the final x-ray complex.
Before proceed any further we need to make sure that those omissions do not affect the structural unity of the ligand binding site.

In SE
Display -> Residue UID

Now you see actual sequence numbers of residues. Check consistency of those numbers.
Here is one gap between Pro 733 and Val 738. Holding the Ctrl key click on both residues with the mouse left button to highlight them.
Another gap is between Glu 749 and Lys 754 - holding the Ctrl highlight them.
One more is between Met 987 and Ser 995 - highlight them as well.

Now highlight the ligand - do not forget hold the CTRL key.
Make all selected residues and the ligand glow on the main screen

In SE
Selection->Atoms-> of Selected Residues

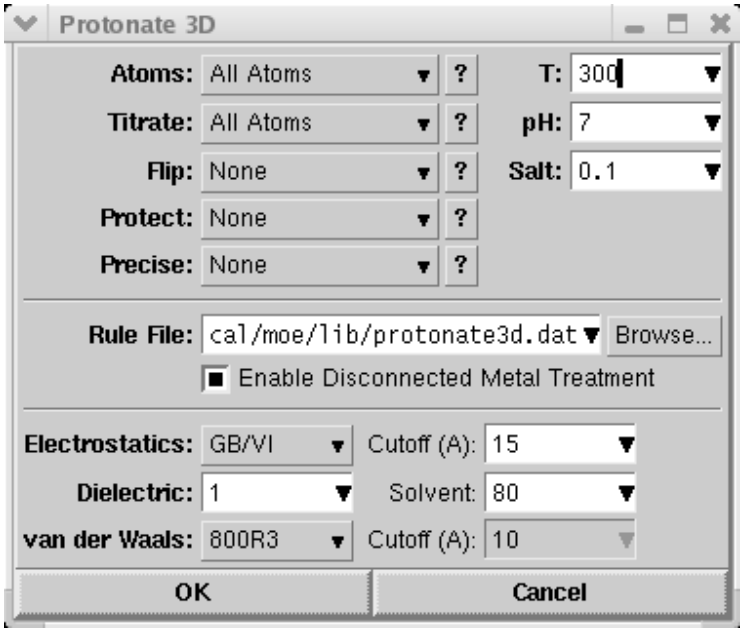
Examine the structure in the main window. In the main window click on **Label** in the side bar and select **Residue**

As we thought the selected residues are most likely the anchor points of the flexible external loops that have not been crystallized. Fortunately for us they are far away from the ligand binding site. Thus we do not need to fix missing regions for the docking.

Now delete ligand - chain 2 and focus on protein preparation for now.

Add hydrogens to protein and minimize their positions.

In Main Window
Compute -> Protonate 3d
Change **Flip** field to None
Keep all other settings at their default values
Click **OK**

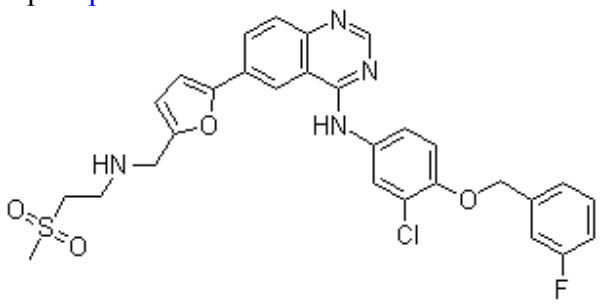


After minimization save the molecule as **Tripos Mol2**
Name: egfr_min.mol2

Close the protein file.

LIGAND PREPARATION

Open lptb.mol2 file



Structure of Lapitinib

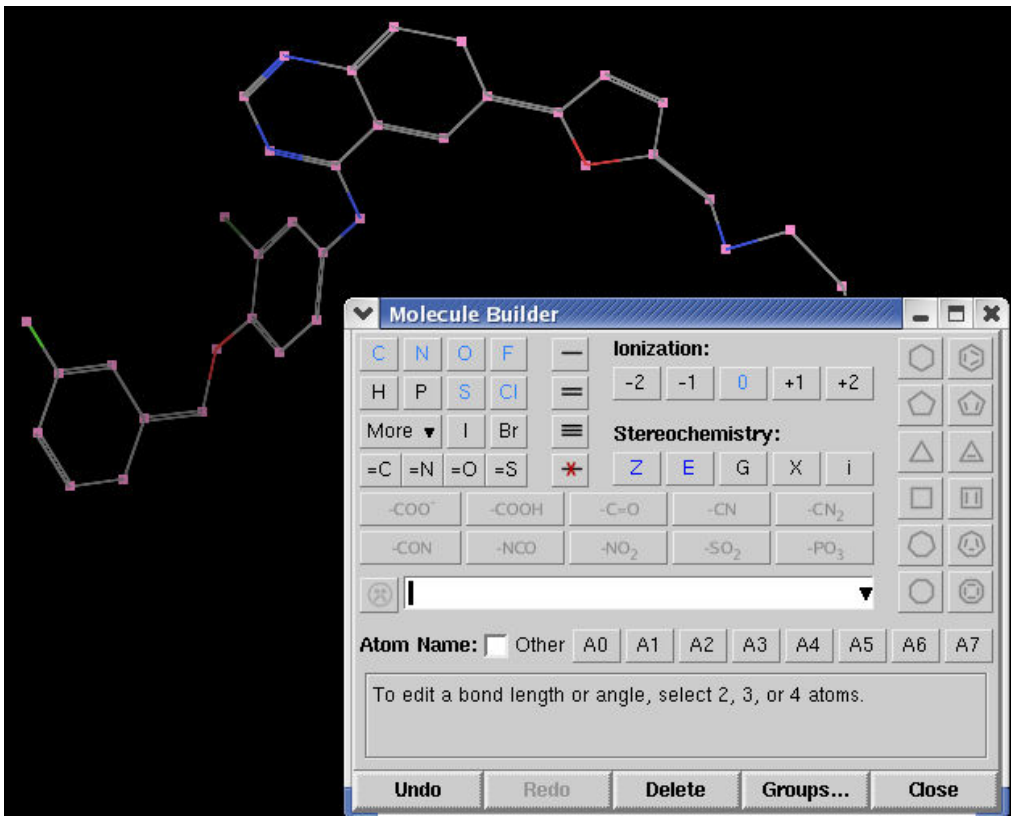
Compare the molecule on the screen with the structure from the scheme above. You may need to fix some bonds, i.e. replace single bonds with double bonds and vice versa.
If both structures look identical proceed to the next step of adding hydrogens.

Adding hydrogens

In Main Window
Edit->Hydrogens->Add hydrogens

You may see that some hydrogens have been added but surprisingly there are no hydrogens on rings. Obviously something is wrong. It happens because a pdb file does not have hydrogens and when we save molecule as mol2 some carbon atoms are saved in ionized (charged) form.
We need to fix it.

With the mouse select all atoms of the ligand on the screen.
Click on **Builder** in the side bar of the main window and then click on zero for atoms ionization.
Close Molecular Builder.



Examine the structure in the main window.
What a mess - single bonds and double bonds are mixed up now!
Using **Builder** fix the bonds according to the scheme of lapatinib from the previous page.

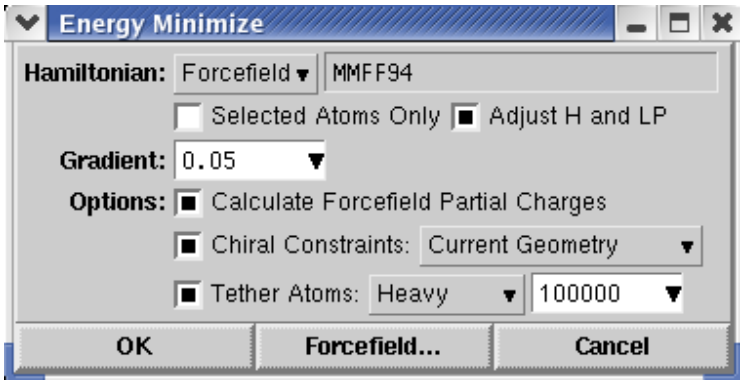
Alternatively you may open lptb.mol2 file in Sybyl, it does recognize all atom types and charges correctly, then add hydrogen atoms in Sybyl- do not minimize the structure! - and save the molecule as lptb_syb.mol2. Then open lptb_syb.mol2 in MOE.

When you finish fixing bonds add hydrogen atoms.
Edit-> Hydrogens -> Add hydrogens

Optimize only hydrogen positions.

Compute->Energy minimize...
Click on the **Forcefield...** button
Select **MMFF94** force field
Click on **Fix charges** then **Close**

In Energy Minimize window
Uncheck at **Selected atoms only** and check at all other boxes
Constrain chirality to **Current Geometry**
and **Tether Heavy** atoms to their current positions with max strength **100000**
Click **OK**



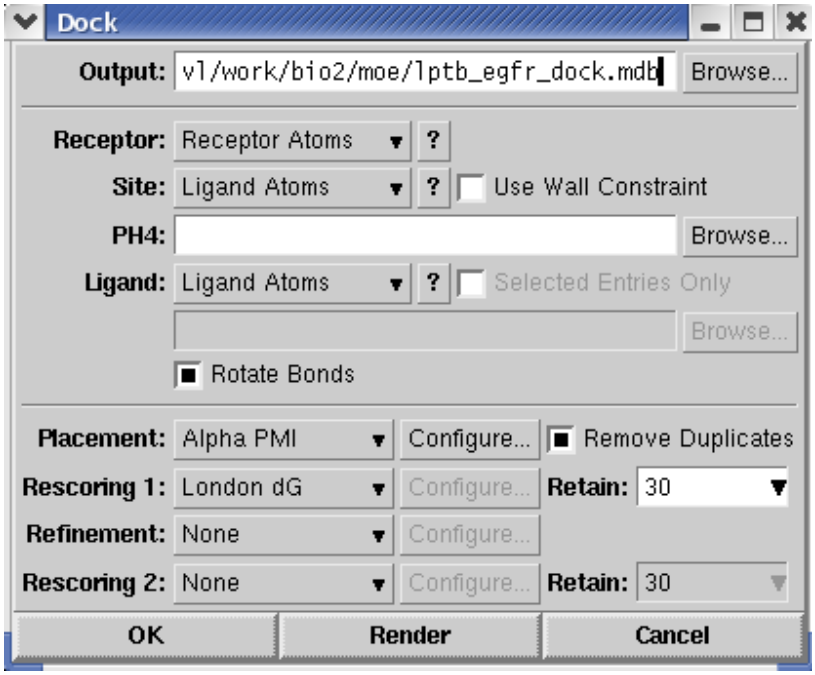
Minimization is done instantly!

Save molecule as **Tripos mol2**
Name: lptb_h.mol2

DOCKING

You should have already the ligand open in the main window. If you closed it after minimization open it again lptb_h.mol2
Open the receptor file egfr_min.mol2

Compute->Simulations->Dock...



Output: lptb_egfr_dock.mdb
Receptor: Receptor Atoms
Site: Ligand Atoms

Click on the question mark next to **Ligand Atoms** to check whether the program correctly determines the ligand. All ligand atoms blink several times in the main screen.

As you see not only lapatinib atoms are highlighted. A small part of the protein - remember we found several missing loop regions in the sequence before- is detected by program as a small peptide and such as a possible ligand.

Thus we have to specify the ligand manually.

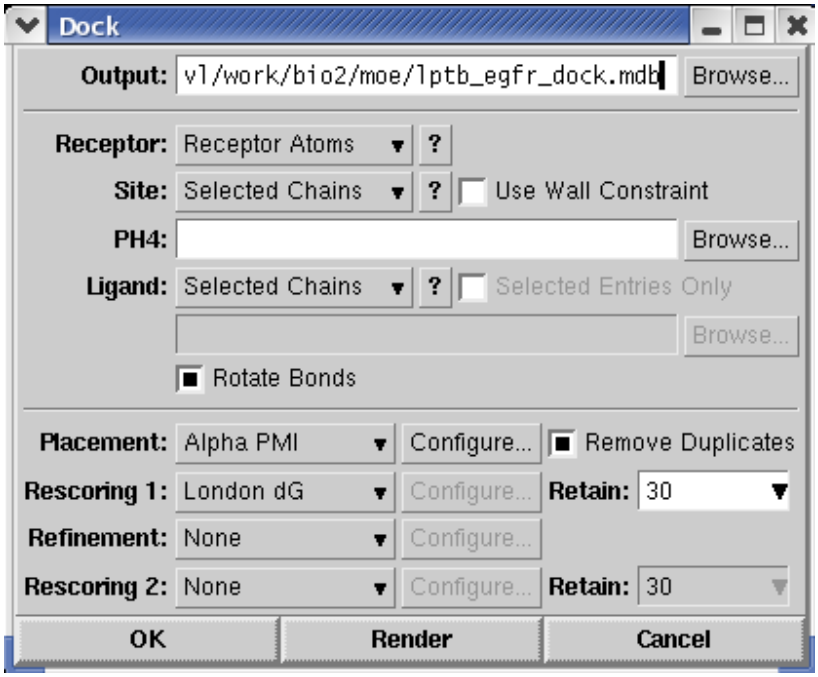
In SE click on chain 1 - FMM.

Go back to Dock window and for the **Site** put **Selected chains**
Click on the question mark to verify the ligand.
Now only atoms of lapatinib are highlighted.

Put **Selected Chains** for **Ligand** as well.
Check at **Rotate Bonds** for flexible docking

Placement: **Alpha PMI**
Rescoring 1: **London dG**
Refinement: None
Rescoring 2: None

Click **OK**



Wait till the table is filled up with docking results in about 10 min.

While you are waiting download a script file fragment_superpose.svl from
www2.umdnj.edu/~kholodvl/files/fragment_superpose.svl
and place it in your project directory.

We will need it later for analysis of the docking results.

Dock writes the top scoring poses to a database containing some or all of the following fields:

Field(s)	Description
`mol`	An output pose.
`mseq`	The molecule sequence number. If a single ligand is docked then the `mseq` field will be 1 for each pose. If multiple ligands are being docked, and the source database does not itself contain an `mseq` field (i.e. not a conformation database), the `mseq` field for the output results will be set to the entry number in the source. If the molecule source contains groups of conformations, the original `mseq` number of the group will be emitted. When examining the output database, it is safe to assume that if two consecutive entries have a different `mseq` number, they belong to a different group of bound

	poses. The `mseq` field is not guaranteed to be group-consecutive or unique.
S	The final score, which is the score of the last stage that was not set to *none*.
E_conf	The energy of the conformer.
E_place	Score from the placement stage.
E_score **E_score1** **E_score2**	Score from the rescoring stage(s).
E_ASE **E_ASE1** **E_ASE2**	The ASE score(s) from the rescoring stage(s). The ASE score is proportional to the sum of Gaussians $R_1R_2\exp(-0.5d^2)$ over all ligand atom - receptor atom pairs and ligand atom - alpha sphere pairs. R_1 and R_2 are the radii of the atoms in Å, or is `-1.85` for alpha spheres. `d` is the distance between the pair in Å
E_refscore	Score from the refinement stage.
RMSD	The root mean square deviation of the pose, in Å, from the original ligand, if there was one in MOE.

More details about docking parameters can be found in MOE manual here
file:///usr/local/moe/html/apps/docking.htm (copy the link into your web browser on Linux boxes)

For our docking run RMSD has not been calculated.
Let's do it manually, that is why you have downloaded a svl script.

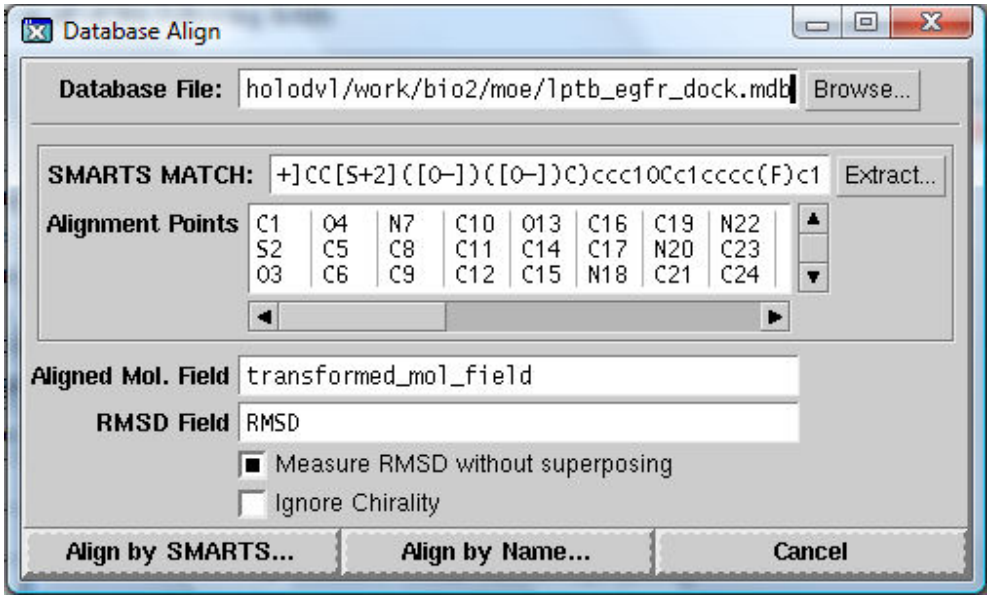
ANALYSIS OF DOCKING

First, close the protein file
in SE delete chain 2.

Now run a svl script
In main window
File->Open
select fragment_superpose.svl and click **Ok**

In the next window agree to use the suggested database (it is your docking result database and it is already open) . Click **OK**

The following window appears



Molecular alignment for RMSD calculations may be performed two ways: semi-automatic based on a SMILES/SMARTS match of the common parts of molecules or manually by selecting specific atoms from the alignment points list.

In the main window you may select with a mouse the common atoms of the ligand for matching. If no selection is made the program takes all visible atoms for calculation.
We want to calculate RMSD based on all atoms thus select everything or just leave the ligand unselected.
Click on **Extract** button.
The program writes the SMILE string for selected (all) atoms in the SMARTS MATCH field.

Important!. By default program aligns molecules from a database in such a way to maximize their overlapping with the reference molecule on the screen. We want to compare ligand poses/placements after docking and calculate RMSD between those poses without changing their positions.

Check at the box **Measure RMSD without superposing**
Uncheck at **Ignore Chirality**

Click on **Align by SMARTS**… button.

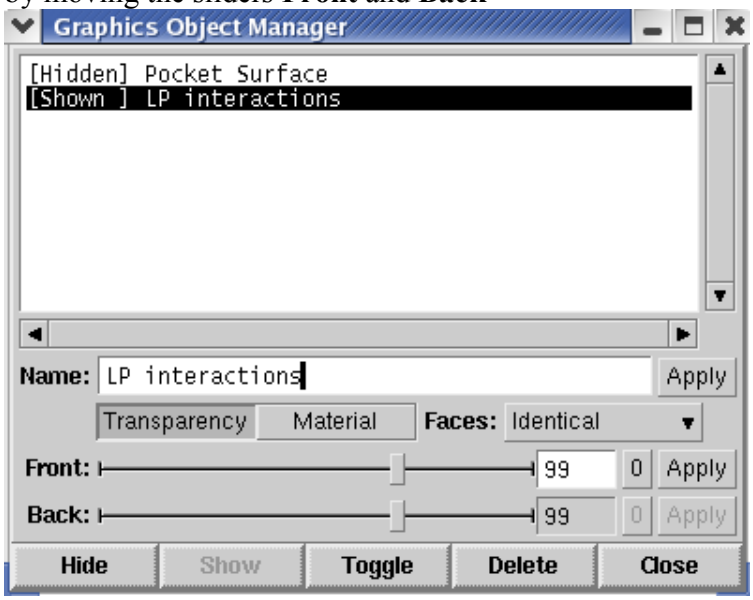
New fields named **RMSD**, **transformed_mol_field** and **comment** will appear in the database.

Sort docking results based on RMSD.
In Database Viewer
Click on **RMSD** header to highlight it
Then **Compute->Sort**
Select RMSD in Priority field
Click **OK**

Analyze how the order of molecules in the database changed and notice what the score (S field) and Energy for the lowest RMSD conformer are. Is the top molecule (sorted by RMSD) the best scorer in S and E fields?

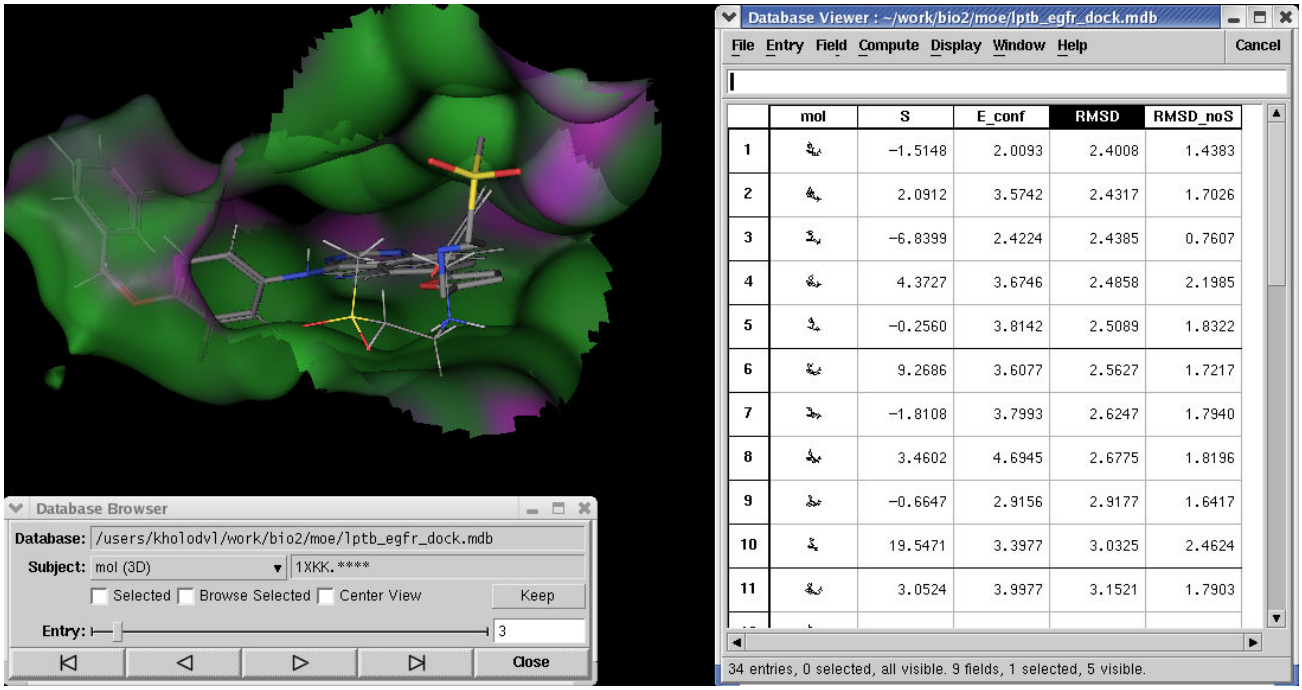
VISUALIZE DOCKING RESULTS

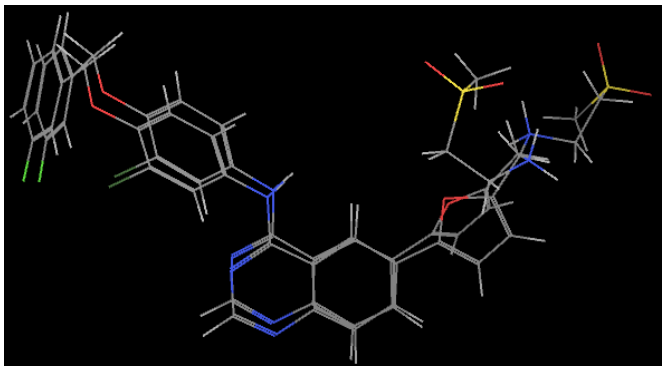
You may want to generate a surface of the pocket as we did it at the beginning of this tutorial for a complex analysis or just compare results with a ligand from a crystal structure. You may adjust a transparency of the pocket surface form
Window->Graphic Objects..
by moving the sliders **Front** and **Back**



In Database Viewer
File->Browser

Move the slider or click on the buttons to switch between conformers from the database.





Notice how ligand poses produced by docking are placed in the binding pocket. What you may say about accuracy of the docking results and the docking scoring function. Is the lowest RMSD conformer aligned well with the reference molecule? What is the “troubled” region? Is there any better aligned conformer from a dataset if a sulfur contained tail of lapatinib is truncated?
Calculate RMSD_noS for the truncated form as it shown in the right panel above – choose the highlighted atoms for calculation.

Write down your observation and fill up the table at the end of this tutorial.

PART II. DOCKING WITH GOLD

Okay NOW it's time for GOLD.
You did already all the dirty work by preparing files for docking.

We will use GOLD for comparison with other docking programs, look for consensus docking and try to choose the best technique for the future research projects.

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 to the new gold directory

pdb1kxx.ent original PDB complex

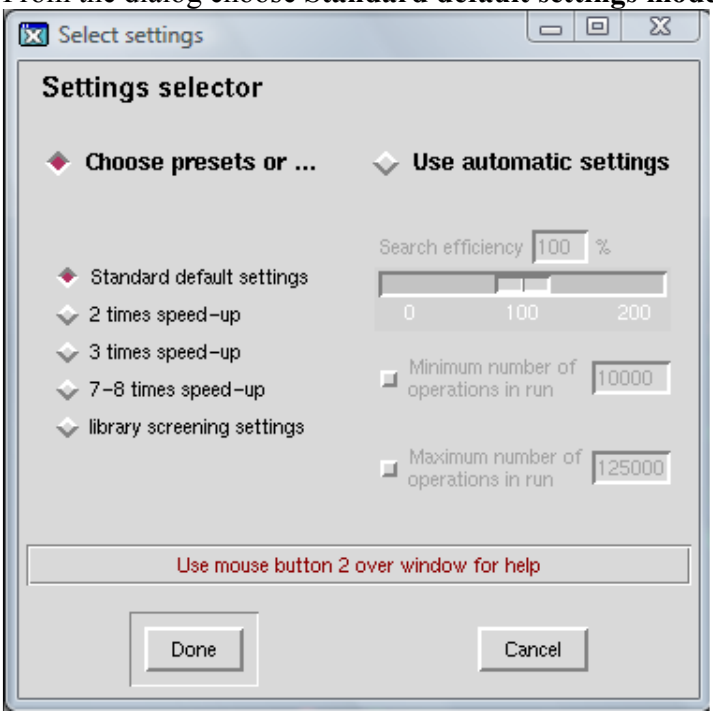
egfr_min.mol2 minimized structure of protein

lptb_h.mol2 refined and minimized ligand

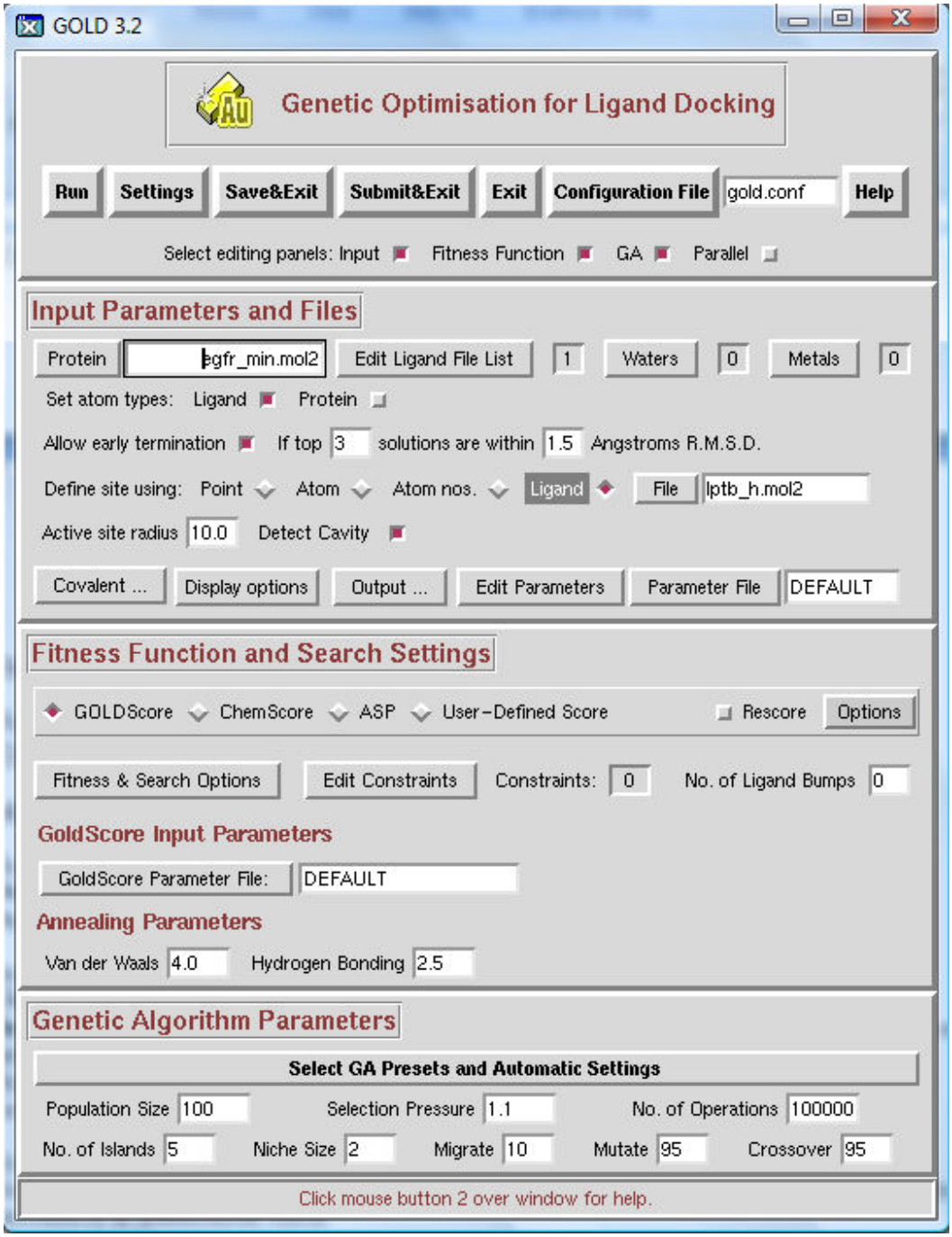
For successful docking you need to know the place where to dock the ligand, i.e. coordinates of the binding pocket or the groove on the surface of the receptor. In this tutorial we are using a real crystal structure of the receptor–ligand complex so we don't need to worry about finding the docking place. We will use a ligand from a crystal structure to define coordinates.

To start GOLD type
gold

From the dialog choose **Standard default settings mode** and press OK



This is the main window of GOLD where you set up all docking jobs. Keep it handy so you know what field/button this tutorial refers to.

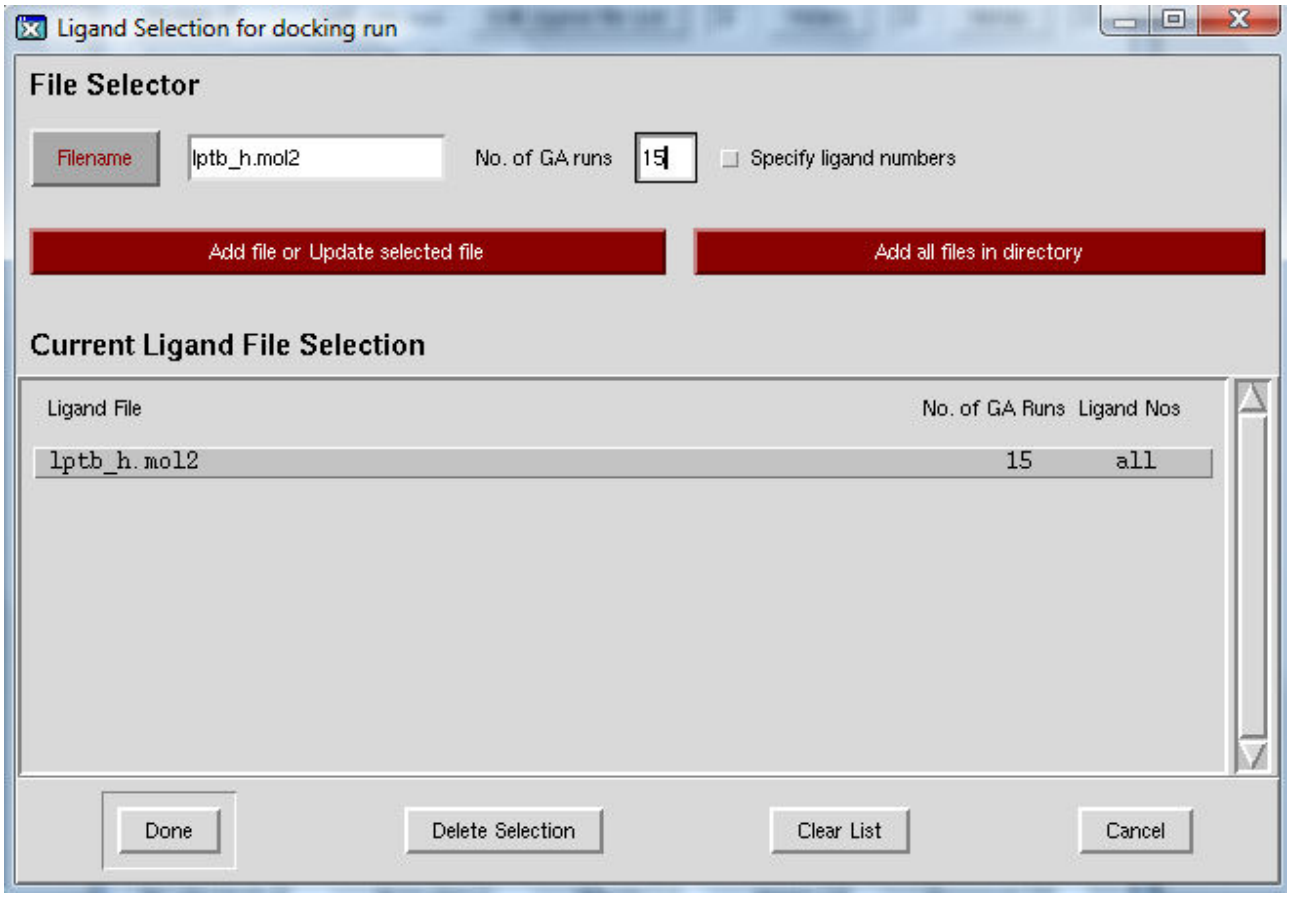


Press on ***Edit Ligand File List*** button

new window will appear

in Ligand Selection for docking run window
click on **Filename** and choose lptb_h.mol2

Put **15** in the box **No. of GA runs** and click on **Add file or Update selected file** button.



New line will appear in the Ligand Editor indicating the ligand name and the number of dockings.
Press **Done** to return to the main window.

Now add the protein
Click on **Protein** and select egfr_min.mol2

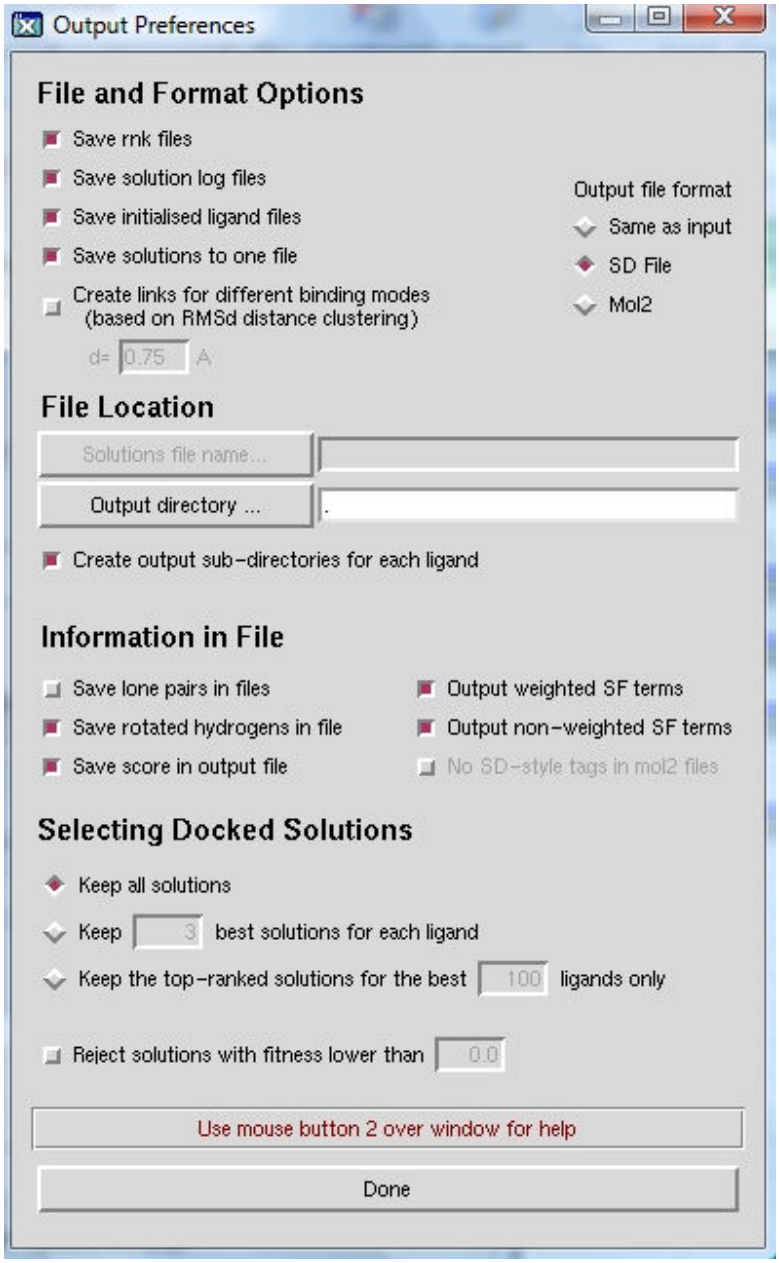
Allow early termination if top 3 solutions are within 1.5 Angstroms R.M.S.D.

In the section **Define site using**
check at
Ligand and select the ligand file lptb_h.mol2.

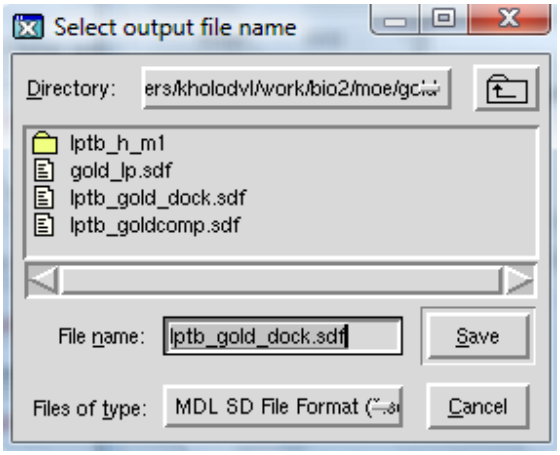
Active site radius: 10
Check at **Detect cavity** to allow GOLD automatically determine a free volume for docking within a sphere with radius of 10 angstrom around the Ligand.
For fitness function check **Gold Score**.

Click on ***Output…***

Check at – see the picture for details
Save rnk files
Save solution log files
Save solutions to one file
Output file format : **SD file**
Output directory: type in your project/gold directory path. If you started gold from this directory keep the path to the current directory (in this case it should be "dot" sign in the field)
Create output sub-directoires for each ligand
Save rotated hydrogens in file
Save score in output file
Output weighted SF terms
Output non-weighted SF terms
Keep all solutions
Uncheck at
Save lone pairs in files



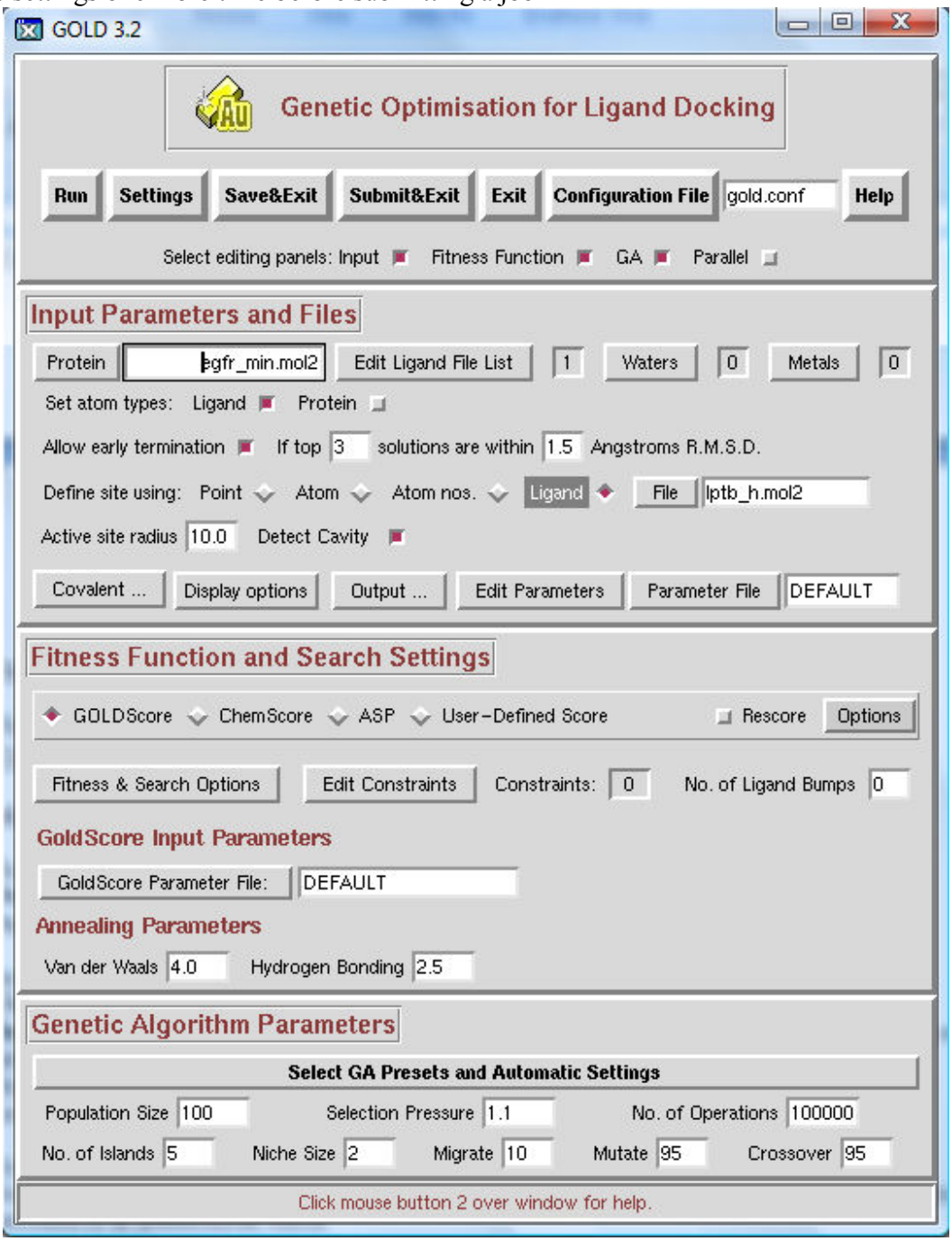
Click on **Solutions file name** and enter the name lptb_gold_dock.sdf
Format **MDL SD File**



Click **Done** to return to the main window
Leave all other parameters to their default values.

You are ready to submit the GOLD docking

Check the settings one more time before submitting a job



It is possible to run GOLD docking in either in an interactive mode where you see all warning messages and output notes of running processes or in the hidden batch mode.

First method is a good choice if you run GOLD for the first time and do not know much about the ligand and receptor interaction or if you are not sure that the binding site is defined correctly.
For the first time let's run GOLD in an interactive mode. Click on **RUN** button.
Several warnings about wrong types of atoms and bonds appear in new windows. Do not panic, as long as they are not about a fatal error and /or crash the docking will continue running. Those messages are

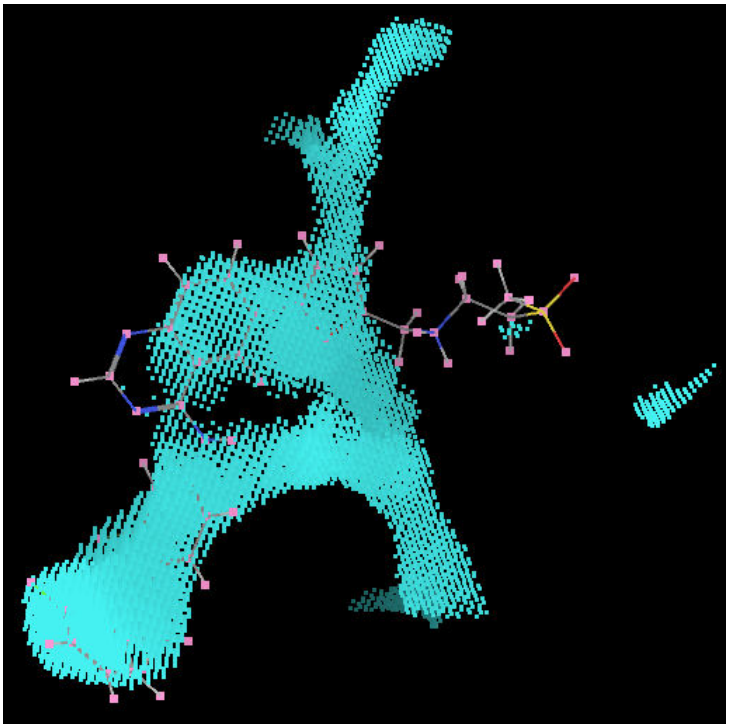
just notification that GOLD has its own parameters for atoms and bonds. If there is no fatal errors let GOLD takes care of all necessary adjustments. Click on **Dismiss** button after examining the output. After 2-3 minutes in the run when you are sure that GOLD is working normally interrupt the interactive run and make preliminary analysis. There is no reason to wait till GOLD finishes all dockings and then to find out that the job was set up incorrectly!
Press **Interrupt GA** and then **Dismiss** button to return to the main window.

Open a new terminal window on Linux box and enter your **gold** directory.
In this directory you should see protein, ligand, complex and other files produced by GOLD run.
The most important file for our preliminary analysis is fit_pts.mol2. This file contains information about the binding site where GOLD is trying to place a ligand during docking. GOLD examines a space within a sphere of the predefined radius and fills the accessible volume with probe (dummy) atoms. Open this file and the ligand lptb_h.mol2 in MOE. If the binding site is set correctly, the ligand and fitting points should overlap.

HINT: Because the fitting point file may contain thousands of points it may significantly slow down the program response time especially when rotating or moving objects on the screen. Open the ligand molecule first, position it in the center of the visible area on the screen, zoom if needed. Then open fitting points file.

If the ligand covered by fitting points less than 10-15% or if it is far away from them then the binding site was set incorrectly. You may want to increase the sphere radius to include more accessible space for docking but it is always better to set the docking site more precisely.

It seems that in our example everything matches perfectly, see the picture below.



Close MOE.

An interactive mode has its own drawback. Besides that it is slower than the batch mode it does not allow you to log out from the computer without losing the submitted jobs. Thus you have to wait till GOLD finishes all docking tasks.

Fortunately, GOLD can be run in a batch mode. In this case you can close the active session or even log out from the system completely and GOLD will continue to run in the background. It is very useful if you have a large set of ligands (high throughput libraries) that you want to test. You may submit the docking job overnight instead of overload the server with a computationally excessive job during the day time.

DOCKING

It is easy to submit the GOLD job in the batch – simply click on **Submit&Exit** button.
The current setup for GOLD run will be automatically saved in the configuration file by default named gold.conf. You may use it later for re-submission of another GOLD job if you plan to use the same protein and the same binding site. Next time when you open GOLD it will read all settings from this file so there will be no need to run through the configuration procedure again.

During the run GOLD produces several files that can be found in the gold main directory.
The most important of them are followed:

gold.conf	settings for GOLD docking
gold.err	error outputs. If GOLD failed to start or has been terminated during the run it writes the error messages to this file. Always check this file at the beginning after about 2-3 minutes of the run and after the run is completed.
gold_protein.mol2	molecule of protein generated by GOLD and used in docking
gold_xxx.mol2 or gold_xxx.sdf	molecule of ligand generated by GOLD before the run
fit_points.mol2	the accessible volume of the binding site where actual dockings are performed
bestranking.lst	list of the fitness scores of the top-ranked poses for each ligand

A single GOLD run usually takes approximately 45-60 sec on Linux boxes in V-12 cluster. There is no difference between the number of molecules and number of dockings, i.e. a set of 5 molecules set for 2 dockings each lasts roughly the same amount of time as 1 molecule set for 10 dockings.
Thus allow about 12-15 minutes for your docking to be completed.

ANALYSIS OF DOCKING

We chose to save the docking output for each ligand in a separate directory. Usually the name of this directory started with a ligand name and ended with the suffix "_m1". Accordingly docking results for lptb_h ligand are placed into directory lptb_h_m1.
If there is no such directory then the results should be in the GOLD main directory.

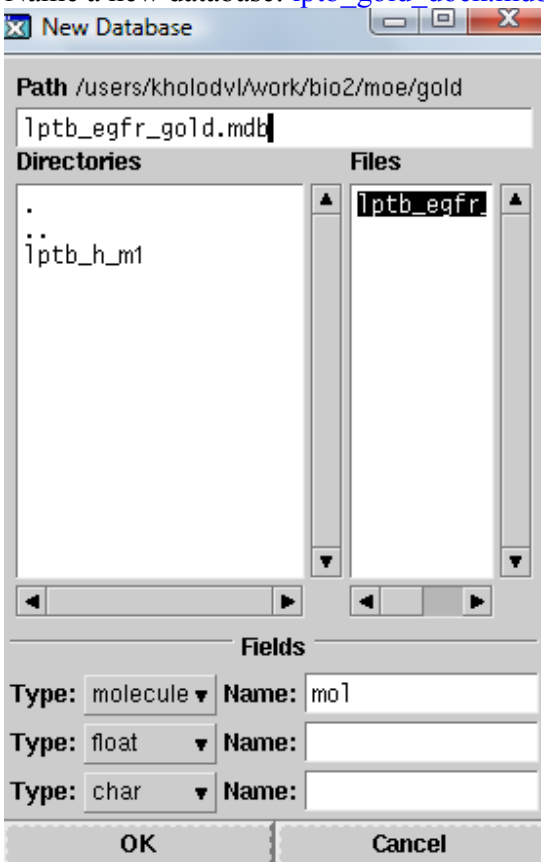
Enter the appropriate directory and check its content.
If you select an output to be saved in mol2 format then you see two sets of files generated by GOLD:
gold_soln_ligandname_m1_X.mol2
ranked_ligandname_m1_X.mol2
where X is a number from 1 to the total number of dockings, in our case from 1 to 15.
The "gold solutions" is a set of conformations/orientations of ligand poses as GOLD placed them in the binding pocket.

Another files, the “ranked”set are not real structural files. They are just symbolic links to the first set but sorted accordingly by the GOLD fitness function. Same ranking information is stored in the rank file in the ligand output directory ligandname_m1.rnk
Be careful when moving or copying docking files from one place to another. By copying only the rank set you may lose all the docking results. Always copy gold_soln files.

But wait a minute, we did not set our output to be mol2 format, we wisely selected that GOLD results to be written in a single SDF file! That’s right, while it is perfectly fine for our exercise to have a sdf file as an output however sometimes for a real project it is preferable to work with mol2 files. Mol2 Tripos format keeps a unique name for each atom in the molecule while SDF/MOL format refers to the position of the atom on the list. Thus from MOL2 format you always can find out what atom C23 or H35 is just simply refer to it by its name while in SDF you have to count the lines in the file to find the atom. There are some advantages and drawbacks specific for each file format but that is out of scope of this tutorial.

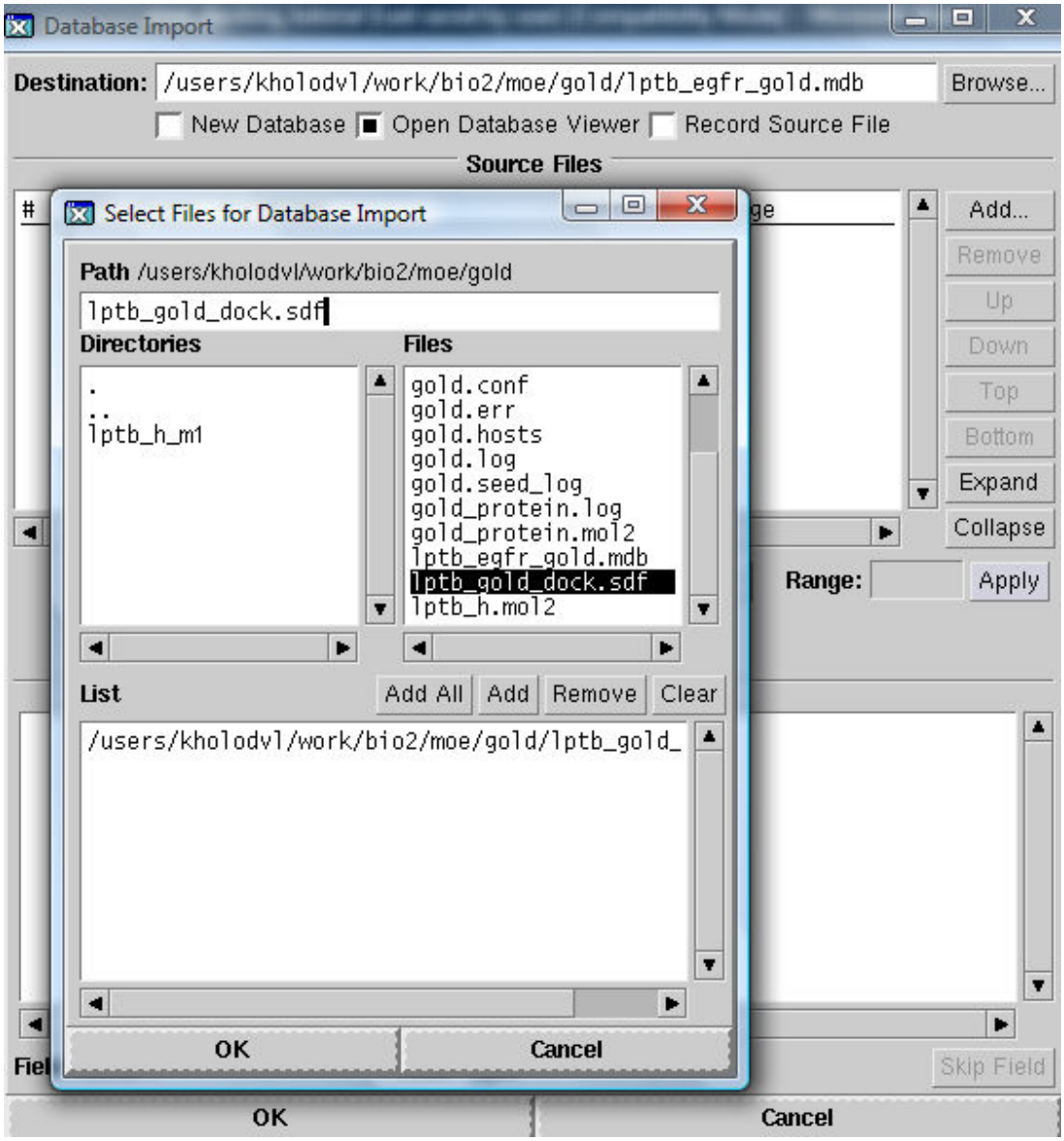
Let’s examine the results from a SDF file now.

Start MOE
Create a new database file where we import GOLD docking results
In MOE
File->New->Database
Name a new database: lptb_gold_dock.mdb



Select for the first field **Type: molecule** and **Name** it **mol**
Click **OK**

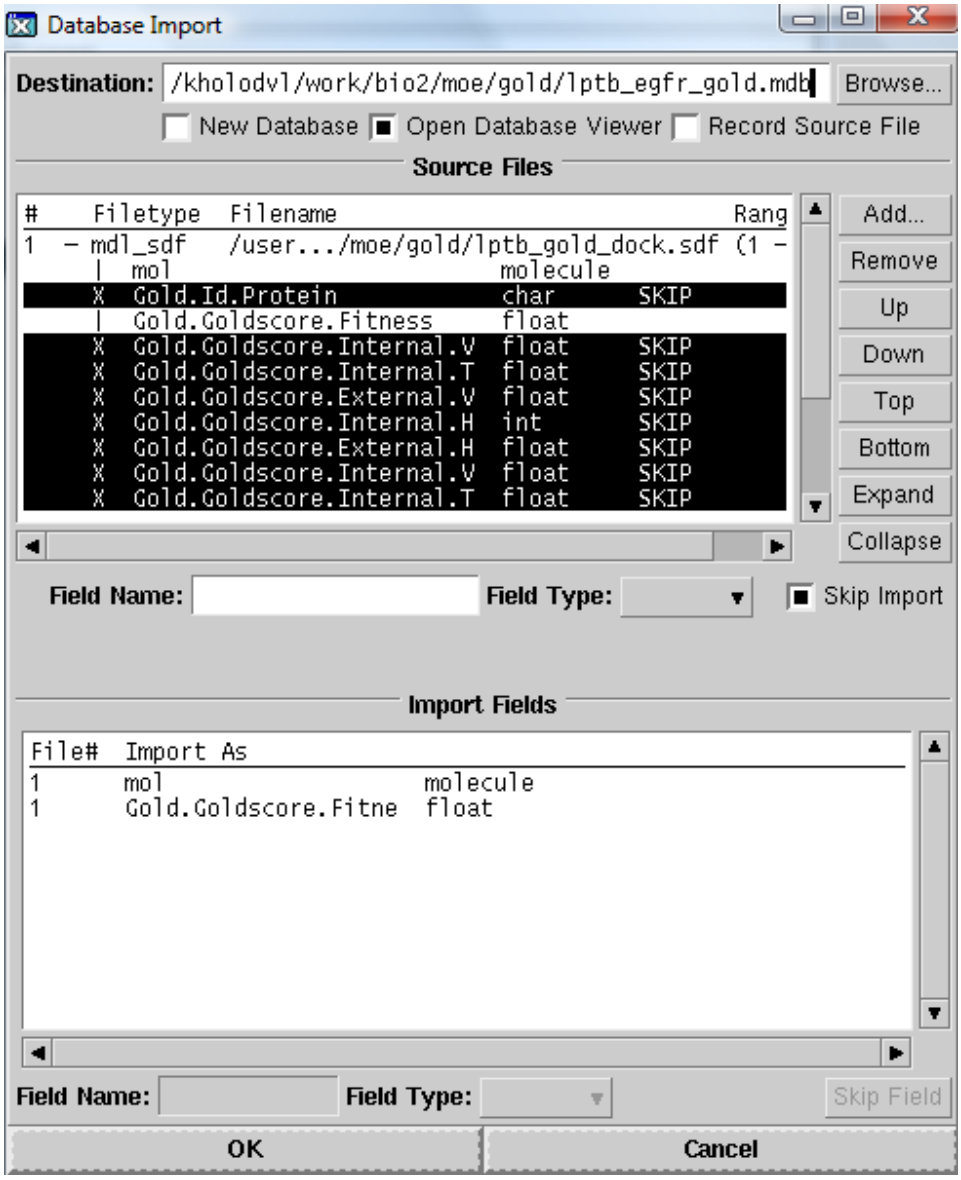
A new empty database will open in the **Database Viewer**.
Import GOLD docking results into MOE database
In Database Viewer
File->Import
in **Database Import** click on **Add**
in a new dialogue window **Select Files for Database Import**
select GOLD output SDF file
lptb_gold_dock.sdf
Click **OK**



It brings you back to a **Database Import** window

In **Import Fields** section you may notice many fields that GOLD added to the resulting file structural file. We need only two of them – **mol** -the molecular structures of ligand poses and GOLD **Fitness**

Select with a mouse all fields in the **Import Fields** section
Click on the first line and holding the SHIFT key click on the last line in the list.
Now pressing the CTRL key click on **mol** and **Gold.Goldscore.Fitness** (lines 1 and 3)
It makes those lines unselected from the list.
Click on **Skip Filed** button at the bottom
Now there are only two fields in the **Import Field** window
Click **OK**



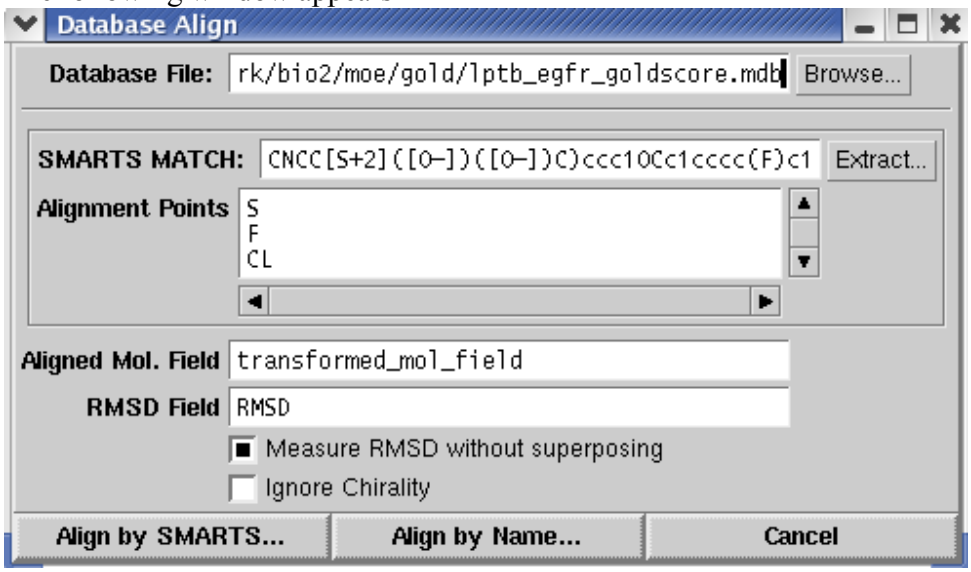
In the **Database Viewer** new entries are added to the database. Number of entries may vary, it depends on how many output poses within 1.5 Angstroms RMSD a GOLD run produced.

Calculate RMSD for GOLD results
Open reference molecule lptb_h.mol2
Run the script that you downloaded for MOE part of this tutorial
www2.umdnj.edu/~kholodvl/files/fragment_superpose.svl

In main window
File->Open
select fragment_superpose.svl and click **OK**

In the next window agree to use the suggested database (it is your docking result database and it is already open) . Click **OK**

The following window appears



Calculate RMSD based on all atoms thus select everything or just leave the ligand unselected.
Click on **Extract** button.
The program writes the SMILE string for selected (all) atoms in the SMARTS MATCH field.

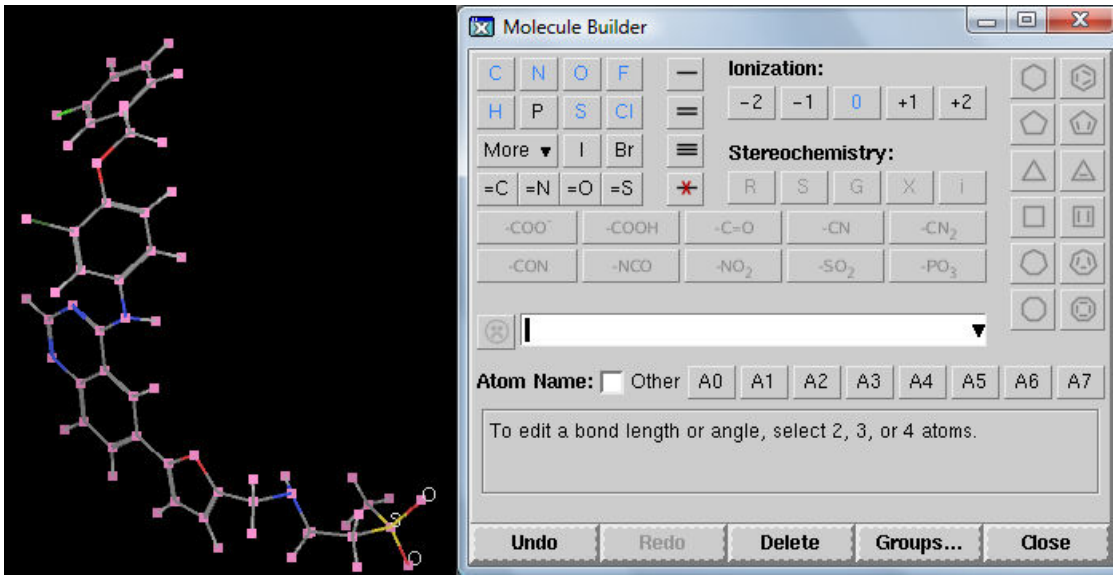
Important!. By default program aligns molecules from a database in such a way to maximize their overlapping with the reference molecule on the screen. We want to compare ligand poses/placements after docking and calculate RMSD between those poses without changing their positions.

Check at the box **Measure RMSD without superposing**
Uncheck at **Ignore Chirality**

Click on **Align by SMARTS**… button.

New fields named **RMSD**, **transformed_mol_field** and **comment** will appear in the database.
Unfortunately as it states in the **comment** field that no matches were found for the ligand and RMSD field remains empty.
It happened because GOLD reassigned atom types and charges to the ligand and now some atoms have a different ionization state. It is a similar problem that we saw before with PDB-MOL2 file conversion in the MOE part of this tutorial. Let's fix the ligand.

With a mouse select all atoms of the ligand on a screen.
Click on **Builder** in the side bar of the main window and then click on zero for atoms ionization.

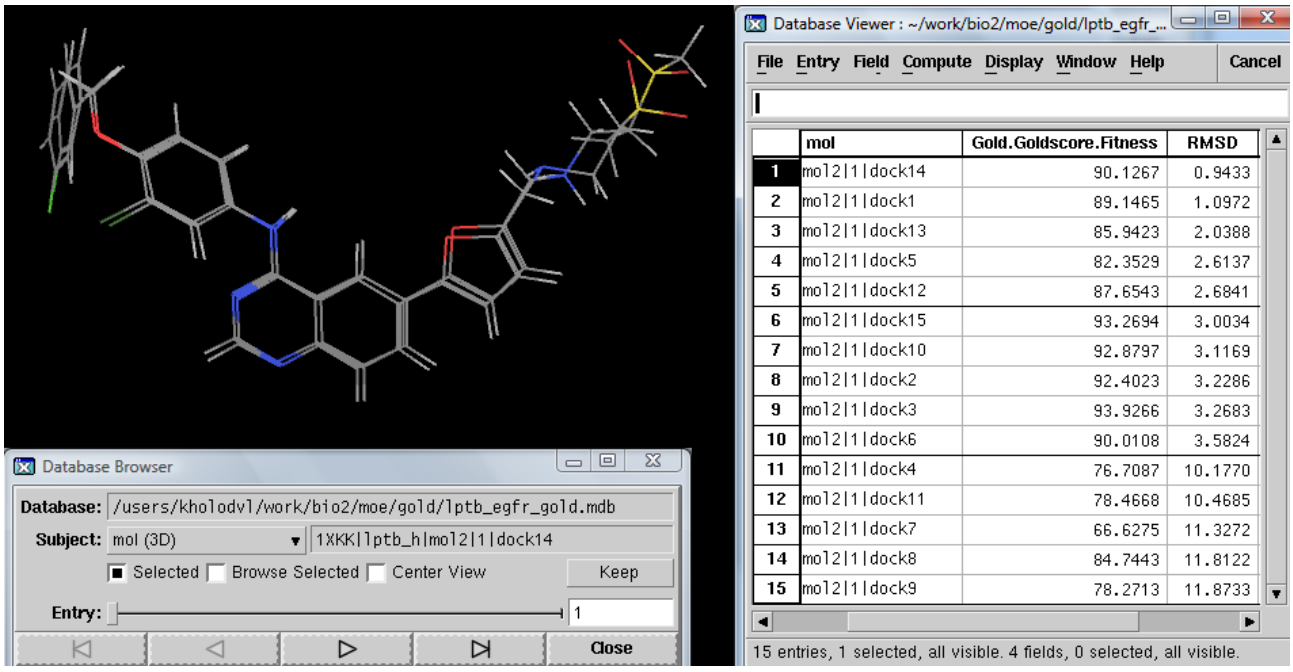


Close **Molecule Builder**

Repeat the step of RMSD calculation. It should be fine now.

VISUALIZE AND ANALYZE DOCKING RESULTS

Refer to the corresponding section in the MOE part of this tutorial.



Notice how accurate/not accurate GOLD docking is.
Compare GOLD Fitness and RMSD ranks. Is the top scorer in Fitness also the conformer with the lowest RMSD score? If not explain why.

Summary and Homework

Fill up the table for the best Score/Fitness and RMSD conformers from MOE and GOLD docking.

Table 1.

File Name	Rank in MOE or GOLD based on Fitness/Scoring	RMSD with the reference molecule	Similarity of poses with the ligand from X-ray complex (rings and functional groups placement, overall orientation)
MOE 1			
MOE 2			
GOLD 1			
GOLD 2			

Compare the results obtained in MOE and GOLD.
Based on your observations decide what docking result is the best and explain why. Is it the top scorer in MOE or in GOLD or another conformer?

Which method/program is more accurate? Are both docking programs consistent in placement and orientation of the ligand in the binding pocket?
Is there a consensus in ligand orientation between both program?
How is it possible to increase the accuracy of the docking?