# *Section II*

## **The Structure of Macromolecules:** *Protein-Ligand Interactions. Molecular Docking (FRED, FlexX) - tutorial*

*Andrei Leitão, Dan C. Fara and Tudor I. Oprea*

Objective

To learn how to use FRED and FlexX software for structure-based virtual screening (docking).

### 1. Background - Nuclear receptors, hormones, cancer and inhibitors: the biological context of diseases

Nuclear receptor is a family of proteins that, upon activation, stimulates transcription of DNA in the nucleus. It can be done in two main ways – see Figure 1:

- directly (named genomic pathway): the hormone binds to the receptor (left part of the Figure 1) and this complex goes to the nucleus, binds to DNA and promotes the transcription;
- indirectly (also known as rapid response): the compound binds to a receptor at the membrane (right part), then second messengers are released and also influence the transcription, among other effects.

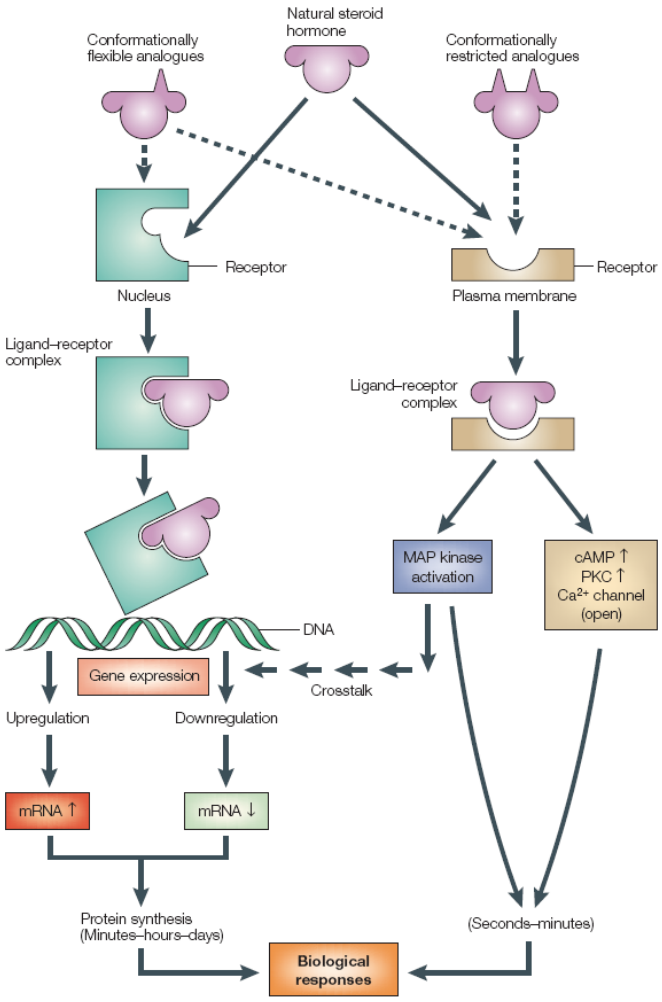


**Figure 1**. Mechanisms of activation of nuclear receptor and promotion of transcription into the nucleus (taken from reference 1)

Estradiol is a hormone that regulates many physiological functions by binding to some proteins called nuclear receptors. More specifically, it interacts with estrogen receptors subtypes alpha (ERα or ERa, $K_i$ = 0.2 nM)[2] and beta (ERβ or ERb, $K_i$ = 0.5 nM) and, as recently described, with GPR30 (a transmembrane receptor)[3].

It is well-known the importance of ERa to the development of breast cancer. The growth of the cancer is related to the stimulus of the ERa by the estradiol or other agonists. It is estimated that 40,460 women and 450 men died of breast cancer in the USA in 2007.[4] Among men, it is estimated that the prostate cancer had been the cause of death for 27,050 individuals in 2007 in the USA. This cancer is related to the androgen receptor (AR)[5], in which testosterone and dihydrotestosterone bind and stimulate the receptor (the second hormone is much more effective) – see Figure 2.

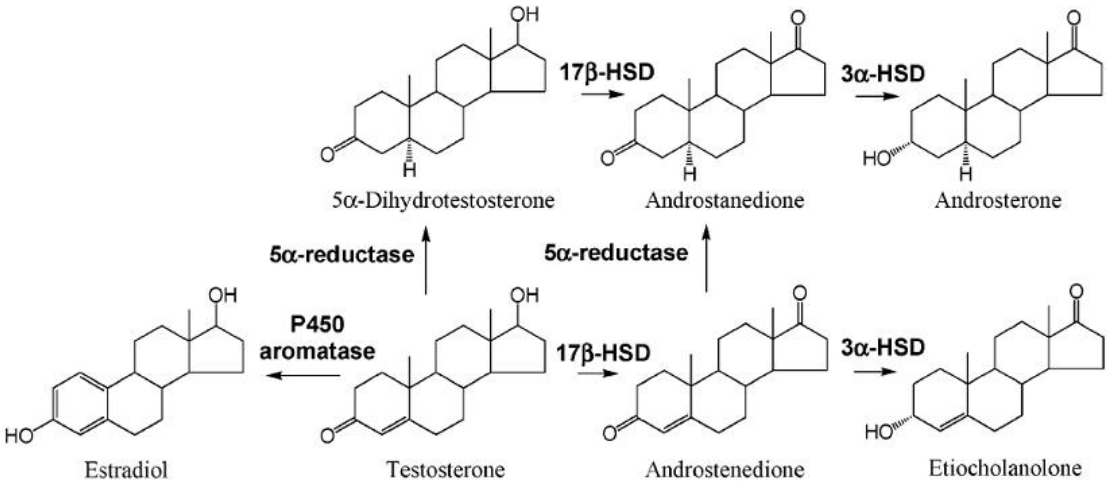



**Figure 2**. Part of the biochemical cascade of metabolism of testosterone (taken from reference 6)

Estrogen receptor is composed by two subunities (dimmer) in which the hormone estradiol binds and activates the co-complex, with the binding of a polypeptide sequence – see Figure 3. This polypeptide sequence that binds to the active form of the receptor is important for the next step of the DNA transcription. Estradiol binds to the ERa in a pocket composed by hydrophobic residues, which interact in the center of the scaffold (not shown in Figure 3 to help visualization), and also forms hydrogen bonds with glutamate 353 (Glu353), arginine 394 (Arg394) and histidine 524 (His524), besides one water molecule (red sphere in Figure 3). Inhibitors, such as 4-hydroxytamoxifen and raloxifene, bind into this cavity, but also interact with aspartate 351 and displace the polypeptide sequence as shown in Figure 3 for estradiol-ERa.

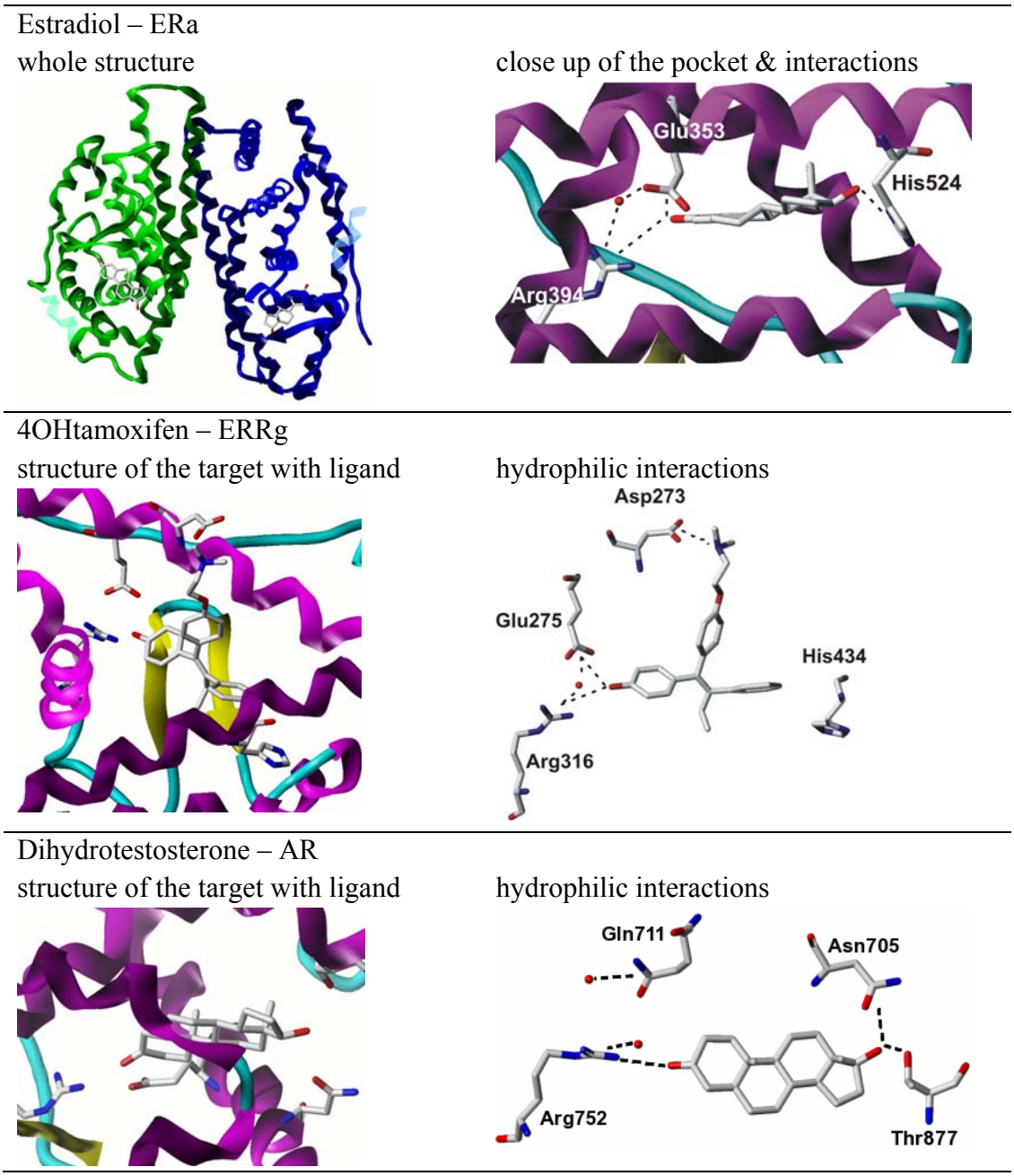



**Figure 3**. Interaction patterns of nuclear receptors and ligands

Dashed lines represent proposed hydrogen bond interactions with the amino acids. Ligands and important amino acids in the pocket are shown as sticks, colored by atom type. Cartoon representations of the protein were colored based on the structure: helices (pink), loops (light green), β-sheets (yellow). Polypeptide sequence is shown as lines.

Other compounds can also interact with these targets, and they can be agonists, partial agonists, inverse agonists and antagonists. These molecules modulate the outcome of the nuclear receptor response.

New selective inhibitors are needed to interfere with this unbalanced biological system and, consequently, to stop the cancer growth. Finding active compounds is not a difficult task, but the problem arises when it is necessary to identify selective compounds[7], because some of these nuclear receptors have similar structure, i.e. when the compound has high affinity to one receptor, it can also bind to others (the lack of selectivity can be related to the some of the side effects of these compounds). For example, the drug 4-

hydroxytamoxifen (also written as 4-OHtamoxifen, OH-tamoxifen) is used to treat breast cancer, but administered as a pro-drug, tamoxifen, binds to three other different nuclear receptors: ERa and ERb ($K_i$ around 0.5 nM) and also estrogen-related receptor gamma (ERRg) with high affinity ($EC_{50}$ = 15 nM). The physiological function of ERRg is not well established, but it is surprisingly that this receptor does not bind estradiol with high affinity, although it is very similar to ERa, with the same pattern of binding to 4-hydroxytamoxifen – see Figure 3.

The newly potentially active and selective molecules are studied based on the way they interact with the biological target[8]. For example, estradiol establishes hydrogen bonds with histidine, glutamate and arginine, but there are also lypophilic interactions with many residues in the middle of the core. 4-hydroxytamoxifen also establishes hydrogen bonds to histidine and glutamate for both ERa and ERRg, but instead of interacting with histidine at the bottom of the pocket, it forms hydrogen bond with aspartate at the top – see Figure 3. Dihydrotestosterone is not an excellent binder to ERa ($K_i$ = 69.9 nM) because it lacks a hydroxy group in one end – see Figure 2. In contrast, it is the hormone that binds with high affinity to the androgen receptor ($K_i$ = 0.3 nM), even better than testosterone ($K_i$ = 1.4 nM). Non-steroidal compounds also bind to this receptor, like metribolone (also known as R1881), which is a potent agonist molecule.[9]

Those studies are primarily done by using many different kinds of computer simulations that comprise, for example, the representation of the target-ligand, the three-dimensional simulation of the receptor and its interactions with the ligands (Figure 3) and also the study and selection of new compounds based on similarities with known binders (this field of work is known as cheminformatics). All these topics will be further discussed.

## 2. Molecular docking using FRED and FlexX - tutorial

Docking is a tool commonly used to discover new chemicals based on the interaction between the ligand and the biological target in the three-dimensional (3D) space. It comes from the knowledge that there are somewhat complementary interactions between the ligand and the biological target. Better ligands have lower (negative) interaction energy and it can be captured by what is known as scoring function. However, the scoring functions give only a rough idea about the ligand – target system.

For instance, two molecules are similar to each other (A and B) but A interacts with the protein (C) way better than B – see Figure 4. A good docking program can capture this information accurately. To accomplish that, the program must change the position of the ligand to fit it into the pocket of the protein – see Figure 4, part 1 – and to detect the correct ligand-protein interactions, while it discards bad solutions – see Figure 4, part 2.

1. Given a database and a receptor site the docking program will change the 3D coordinates of the ligand to match the target
2. After that, the program starts to check how well the ligand fit the target (hydrogen bonds are shown as dashed lines)
3. the best result for each ligand is ranked (it can be also a set of results for each ligand depending on the procedure)

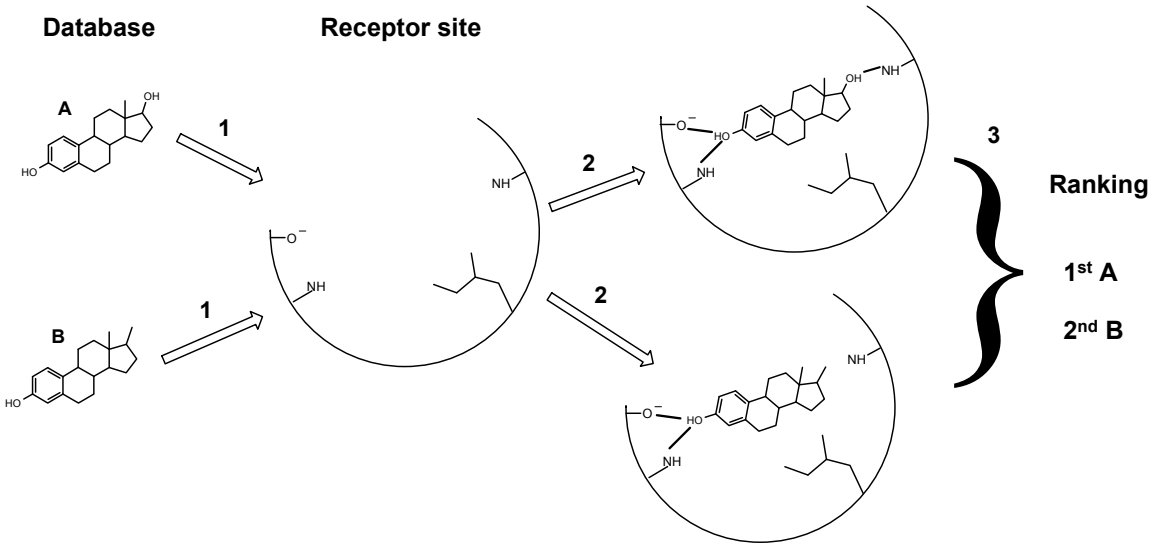



**Figure 4**. Interaction profiles for molecules A and B and the docking protocol

The main idea of ligand-protein interaction comes from the lock-and-key model, where both ligand and protein are rigid bodies, as shown in Figure 4. In reality, the ligand and protein are not rigid, one or even both of them being flexible during the process of recognition. There are many different docking programs that try to solve this apparent simple question (in fact, none of them can be used to all systems).

Here you will use two different programs, FRED and FlexX, to dock small molecules from a 3D database (199 compounds) into three biological targets: estrogen receptor alpha (ERa), estrogen-related receptor gamma (ERRg) and androgen receptor (AR). The outcome of each ligand-receptor interaction will be automatically analyzed by a scoring function which, in turn, ranks the compounds.

## 2.1 FRED

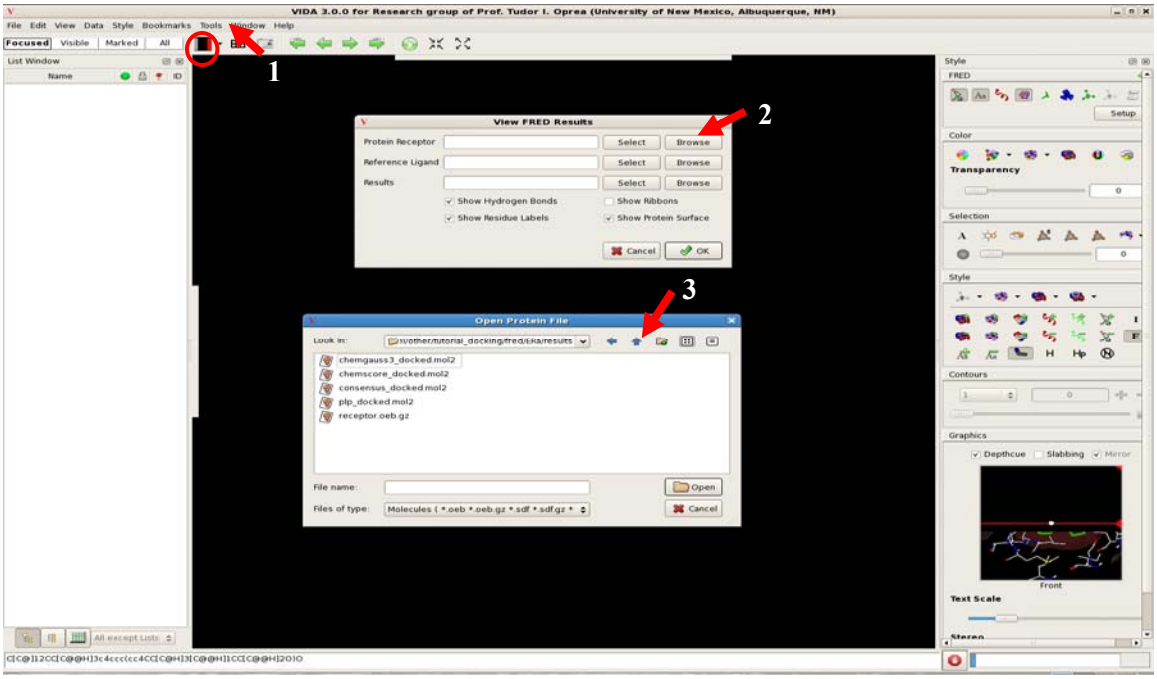
This program uses the lock-and-key idea presented above. A bunch of conformations from each ligand interacts with the receptor and the best solution is stored to be ranked at the end of the process. Programs like this one give results much faster than others that deal with the flexibility of the compound.

- Connect to hatch.health.unm.edu, your account
- Type **cp –r /home/app/tutorial-docking ~**
- Go to the tutorial directory (type **cd ~/tutorial_docking** on the shell) and see what is in there (use the list command: **ls**). The following three directories are listed: (i) *compounds*, (ii) *FlexX*, and (iii) *fred*. The first one contains the compound database, with the 3D structures generated and a set of conformations for each one. Open *fred* directory (**cd fred**) and note that each one of the three directories (ERa, ERRg and AR) contains the files needed to run the program. Go to ERa (type **cd ERa**) and list the files again (**ls**): 3ERT.oeb is the receptor already prepared; OHtamoxifen.mol2 is the ligand; setup.txt has all the information necessary to run the program; and the directory result is empty (if you want, you can check it before running FRED). Now let us run the program: **fred -param setup.txt** It takes time to dock the entire database (around 15 minutes) because this kind of calculation is much more complex (see Figure 4 with the procedure).

- Go to the directory result (**cd results**) to check it (**ls**). Many files were created, but you will analyze the consensus_docked.mol2 (with the conformations of the compounds docked into the site) and consensus_scores.txt (the ranking of the compounds).
- Open the file to check the results: **gedit consensus_scores.txt &**. Try to find the position of the compounds that were added to the set (ohtamoxifen is the best ranked, estradiol - the 69$^{th}$, while dihydrotestoterone is the worst of them, as expected). The top 5 compounds will be analyzed more deeply, but it can also be done to other compounds.
- Close the gedit window.
- Open a visualization program named VIDA to see how good the compounds fit the pocket of the receptor: type **vida** on the shell. An interface like the one shown in Figure 5a will open. The program has the following layout (if the layout is different, Click on the button showed by the red circle, Figure 5a):
  - the top part contains the main menu and many options to change the display (Figure 5a. number 1)
  - the left part harbors the ligand list and other information (two white windows)
  - the main windows (middle black one) is where you visualize the compounds
  - the right window has some visualization buttons
- Click on 'Tools' and 'FRED view' on the menu (Figure 5a, number 1). The box shows the options to fill in (number 2):
  - Protein: 3ERT.oeb (click in 'browse' and go one directory up, number 3)
  - Reference ligand: OHtamoxifen.mol2 (click in 'browse')
  - Docking results: consensus_docked.mol2 (click in 'browse' and click on 'result'). The result will appear on the main window, loading the list of compounds docked (Figure 5b). Now it is time to see how good the results are.
- Hide the hydrogens by clicking on the H symbol, highlighting the icon (skip this step if hydrogens are not shown) – see Figure 5b, number 1. Look at the main window with the structures. You are seeing two structures (green is the tamoxifen from the X-ray, the template, and gray the docked one). They are inside the ERa pocket, shown by the surface, and there are some hydrogen bonds with amino acids, highlighted by the green dashed lines (number 2).
- Hide the standard compound by clicking on the green balls in the left panel (number 3). Now, only the docked compound (gray carbons) is shown. As can be seen only a weaker hydrogen bond interaction is shown, because the dashed yellow line is finer than the standard compound-ERa interactions. These hydrogen bonds with arginine and glutamate are important for those compounds, because it helps the stabilization of the co-complex. To rotate the co-complex, drag the mouse with the left button pressed. There is no hydrogen bond with histidine because of the lack of complementarities. Another hydrogen bond, with aspartate 351, is not shown on the standard X-ray structure (green) because of the distance between the amino group and the side chain of the amino acid. In other X-rays this hydrogen bond is better defined for antagonist compounds.

You can also check other structures by clicking on the codes or in the green ball column – see Figure 5b, number 3. Remember to hide the latest compound opened before looking at the next one in order to facilitate the visualization (do not do that if you want to compare the structures).

(a)




(b)

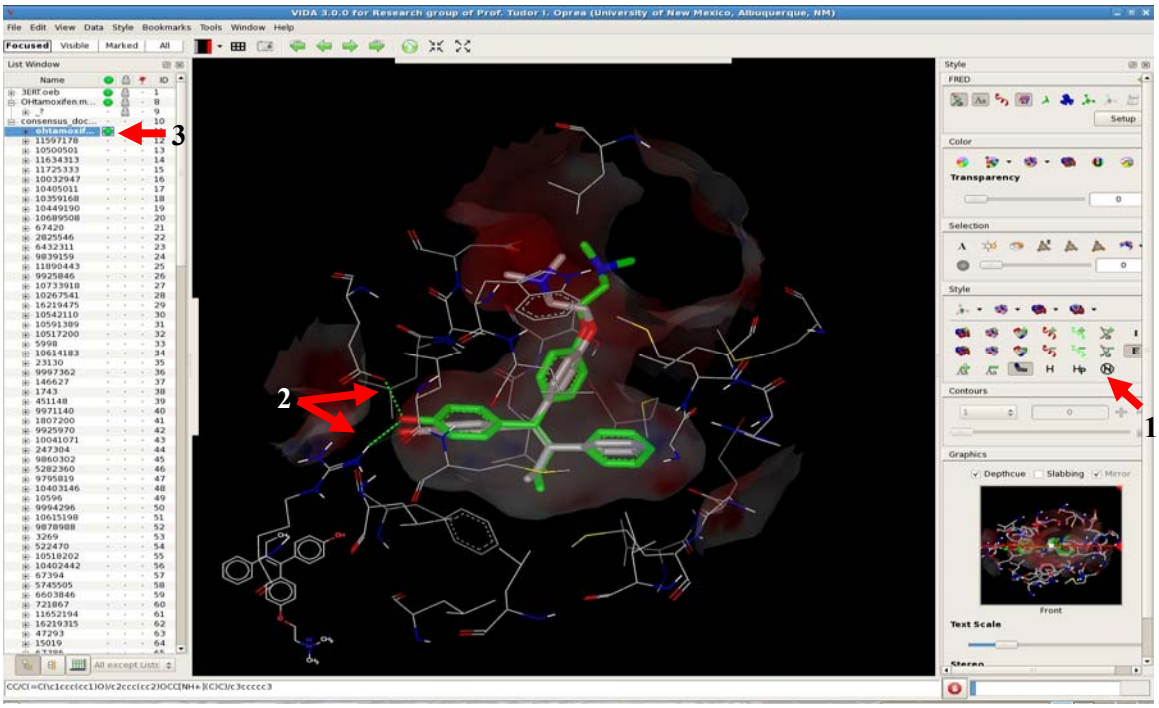



**Figure 5**. VIDA interface and docking results for ERa

Figure 5b: Note that when you select a compound the structure of the ligand appears at the bottom left corner

- Go to the ERRg (**cd ../../ERRg**) and also AR directories and repeat the previous given steps.
- Open the results.

Table 1 shows the top 5 compounds for those three results.

Table 1. Ranking of the best 5 compounds for each target

| Name | Consensus Score AR | Name | Consensus Score ERa | Name | Consensus Score ERRg |
|---|---|---|---|---|---|
| 10314457 | 7 | ohtamoxifen | 0 | 10405011 | 5 |
| 10041071 | 7 | 11597178 | 6 | 10689508 | 14 |
| 11722940 | 7 | 10500501 | 7 | 11597178 | 15 |
| 10470789 | 16 | 11634313 | 10 | ohtamoxifen | 18 |
| 10402442 | 21 | 11725333 | 16 | 10449190 | 22 |

Try to compare the best five ranked compounds with the standard ones in the study (Figure 6). Note that they interact similarly with the target because they share some similar features with the scaffold.

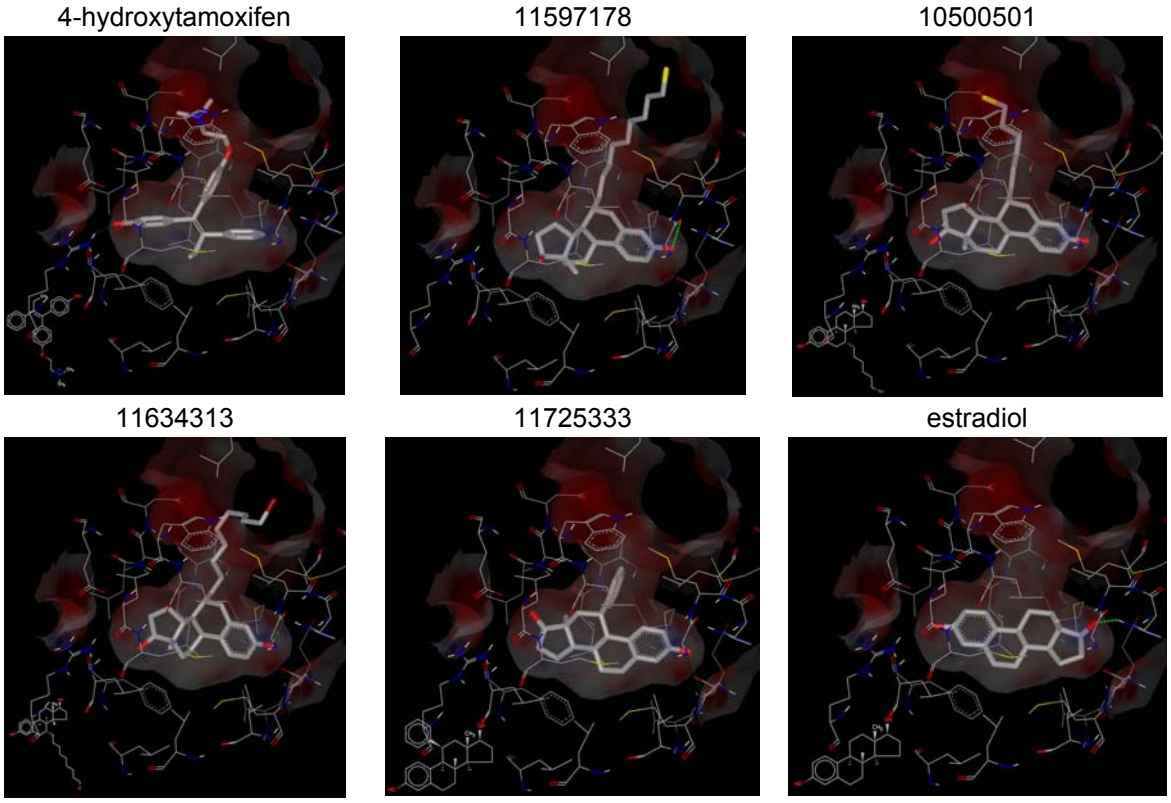



**Figure 6**. Comparison of the top 5 docked compounds and estradiol in ERa

- Parameters to use in VIDA after running the docking program for ERRg target:
  - Protein: 3ERT.oeb (click in 'browse' and go one directory up, Figure 5a, number 3)
  - Reference ligand: ohtamoxifen.mol2 (click in 'browse')
  - Docking results: consensus_docked.mol2 (click in 'browse' and click on 'result')

Note that the OHtamoxifen appears on the top 10 list. Find estradiol and dihydrotestosterone in consensus_scores.txt and observe that these compounds had bad results (at the end of the list).

- Parameters to use in VIDA after running the docking program for AR target:
  - Protein: *.oeb (click in 'browse' and go one directory up, Figure 2a, number 3)
  - Reference ligand: dihydrotestosterone.mol2 (click in 'browse')
  - Docking results: consensus_docked.mol2 (click in 'browse' and click on 'result')

Check if the results are good or not. OHtamoxifen was at the end of the list. It is expected, once that it not AR binder. Now compare estradiol and dihydrotestosterone. Are they ranked well according to what you already now about these ligands? Try to figure out what is happening with estradiol and dihydrotestosterone in relation to AR binding.

### 2.2 FlexX

This program is slower than FRED because the conformation of each compound changes on-the-fly (i.e. the conformation is changed while the program is running the calculation). In fact, FlexX uses another concept, based on cutting the compound in pieces, docking the anchor points and assembling these parts to re-generate the compound. This is known as dynamic change of the conformation in part, once that only the ligand can change its conformation during the interaction with the rigid receptor.

The procedure is also straightforward. Being connected to hatch.health.unm.edu, your account, go to the FlexX directory: **cd tutorial_docking/FlexX**. Again, three directories with the targets' files are prepared for running the study. The procedure for ERa target is given below and is the same for the other two targets. Go to the **ERa** directory. The following files are available: (i) 3ERT.rdf, which is the receptor definition file, (ii) 3ERT.pdb, which is the original receptor, (iii) 3ERT_surf.sdf, which is the surface created to the pocket, and (iv) 3ERT_poc65.pdb, which is the pocket generated by considering all amino acids at 6.5 Å distance from the ligand ohtamoxifen.mol2. The script.bat file has all the information to run the calculation, which is based on the definition file (config.dat). The compound database is located on the directory /tutorial_docking/compounds.

- Type the command **flexx**
- Run the program by typing: **script script**
- Wait until the end of calculation and close the program by typing **quit**.
- Open VIDA to analyze the results (type **vida** in the shell)
- Go to: *Tools* > *FRED View*
- Click in *browse* to find the Protein receptor file: 3ERT.pdb; Reference Ligand: ohtamoxifen.mol2; Results: *.*

Use the left button of the mouse to rotate the complex and the middle one to zoom in and out. Note that the docking result is much better when FlexX was used for hydroxytamoxifen, but in the same time the calculation took longer because the conformation is interactively changed to fit the target when docking into the ERa site (remember that FRED uses a set of conformations previously generated to dock).

Follow the same steps for ERRg. Go to the **ERRg** directory (**cd tutorial_docking/FlexX/ERRg/**) and list the files (**ls**). The receptor files are as follow: (i) 2GPU.pdb, the original file with the coordinates of the receptor, (ii) 2GPU_poc65.pdb that defines the pocket size of the receptor, (iii) 2GPU.rdf, the receptor file, and (iv) 2GPU_surf.sdf, which is the surface of the receptor. The ligand file, ohtamoxifen.mol2, is the reference compound.

Finally, do the same to AR and analyze all results, comparing with the FRED ones.

---

[1] Norman, A. W.; Mizwicki, M. T.; Norman, D. P. G.; Steroid-Hormone Rapid Actions, Membrane Receptors and a Conformational Ensemble Model, *Nature Reviews Drug Discovery* **2004**, *3*, 27-41

[2] $K_i$ and $EC_{50}$ data extracted from WOMBAT2006.2, Sunset Molecular Discovery LLC.

[3] Bologa, C. G.; Revankar, C. M.; Young, S. M.; Edwards, B. S.; Arterburn, J. B.; Kiselyov, A. S.; Parker, M. A.; Tkachenko, S. E.; Savchuck, N. P.; Sklar, L. A.; Oprea, T. I.; Prossnitz, E. R.; Virtual and Biomolecular Screening converge on a Selective Agonist for GPR30, *Nature Chemical Biology* **2006**, *2*, 207-212

[4] http://www.cancer.org/downloads/STT/CAFF2007PWSecured.pdf

[5] Taplin, M.-E.; Drug Insight: Role of the Androgen Receptor in the Development and Progression of Prostate Cancer, *Nature Clinical Practice Oncology* **2007**, *4*, 236-244

[6] Gao, W.; Bohl, C. E.; Dalton, J. T.; Chemistry and Structural Biology of Androgen Receptor, *Chemical Reviews* **2005**, *105*, 3352-3370

[7] Blizzard, T. A.; Gude, C.; Chan, W.; Birzin, E. T.; Mojena, M.; Tudela, C.; Chen, F.; Knecht, K.; Su, Q.; Kraker, B.; Holmes, M. A.; Rohrera, S. P.; Hammond, M. L.; Bridged androstenediol analogs as ER-b selective SERMs, *Bioorganic & Medicinal Chemistry Letters* **2007**, *17*, 2944–2948

[8] Menezes, I. R. A; Leitão, A.; Montanari, C. A.; Three-Dimensional Models of Non-Steroidal Ligands: A Comparative Molecular Field Analysis, *Steroids* **2006**, *71*, 417-428

[9] Ariazi, E. A.; et al. Exemestane's 17-hydroxylated Metabolite Exerts Biological **E**ffects as an Androgen, *Mol. Cancer Ther.* **2008** *in press*