**Excercise - Analysis of structures of protein-ligand complexes**

In this excercise you will examine the geometrical features of six x-ray structures in order to understand the interactions between ligands and receptors and to get an overall understanding of structure complexes. We will the ligand binding domains of nuclear receptors for this purpose since they bind their ligands with high affinity and specificity, and several interesting structures have been solved.

The first half (step 1-4) will cover three complexes of receptors with their cognate (natural) ligand. In the second half, the effect of a mutation associated with disease will be looked at (step), and the structural mechanism of drugs acting on receptors will be shown (steps).

**1. Starting ICM**

The molecular modeling program ICM will be used for analysis of complex structures. (This free version is available after registration at www.molsoft.com.) Make sure the computer is Linux booted, log on, open a unix terminal window (right click on background, select "open terminal") and execute

```
mozilla www.molbio.gu.se/~brive/tmp &
```

Right-click on the link and save the file on the disk (Save link target as...). Then execute this in the command window:

```
icmbrowserpro nuclearReceptors.icb
```

Hint: Pressing the tab key will complete words if possible, so the above line can probably be entered using

```
icm<tab> nucl<tab>
```

You should now see a window containing, amongst other things, a ribbon structure of the androgen receptor. Maximize the window by clicking the button in the upper right corner of the ICM frame.

**2. Getting along with ICM**

There are four parts of the window, that will be refered to as the workspace, the html window, the graphics window and the command window. (See figure below) In the next step (3) you will be guided through their use, but below is a short description of each for reference. Read through it quickly before doing step 3. You can try things out if you wish, and restart ICM before doing step 3 to make sure things are ok again.

- workspace

This is the list of the six PDB structure complexes that we'll use, the names being 1i37, 1a28, 1ert, ... Clicking on the object name will display or undisplay the structure in a predefined way. Clicking the triangle next to the object's name brings up a sub-listing of protein, ligand and water molecules. You can turn on or off the individual parts there.

(The objects have been modified from the original PDB file: Only one protein molecule and its associated ligands and solvent have been kept. The original file can be retrieved and checked by typing, e.g., ' nice "1i37" ' (without the single quotes) in the command window.)

Further details on the display features will be shown during the excercise.

- html window

This window contains buttons that brings shortcuts to the visualization. Their names are self-explanatory. The pocket representations that can be switched on or off have been generated by mapping the solvent-accessible surface of protein and crystal waters (tightly bound, can be considered part of the protein) within 8 Ångström from the ligand, and shows the shape of the binding pocket. This window is specifically made for this excercise, so if you run ICM some other time it will not be available.

The tools in the different tabs are avaiable all the time, so if you for example need to reset the view, click the "pockets" tab (if it's not already active) and use the "reset view" link.

- graphics window

The graphical objects are shown here. The most important controls are these:
(LM=left mouse button, MM=middle mouse button, RM=right mouse button, LR=left/right, UD=up/down)

| | |
|---|---|
| LM | rotate in the viewing plane |
| MM | move in the viewing plane |
| LM near top edge (LR) | rotate around the axis perpendicular to the screen |
| LM near left edge (UD) | zoom |
| | |
| LM near top right edge (UD) | move the back clipping plane |
| LM near middle right edge (UD) | move through the clipping plane |
| LM near lower right edge (UD) | move the front clipping plane |

One of the buttons on the right can be used to rock the molecule which might be useful. Move the mouse without clicking over the buttons to identify the button. Clicking the correct button will start the rocking, pressing and holding the mouse button will bring up a list of different rocking modes. Stop the rocking by clicking the button again, or press the escape key on the keyboard when the pointer is in the graphics window.

If you spin and twist the molecule so much that you get lost, click the "reset view" link in the html window.

You can make the graphics window cover the whole screen by moving the mouse to the graphics window and press <control> F . Go back again with the same two-key combination.

- command window

The actual commands that get executed when buttons are clicked will be shown here, don't worry about it. You may close it if you wish. The adventureous and brave folks can type stuff here, but go through the whole excercise first and don't blame anyone else if things go wrong... Those that manage to make some sense of it are recommended to type "help" followed by a command word.

**3. Guided tour of the androgen receptor (1i37)**

If you clicked around a lot before coming here, quit ICM without saving and start anew . On the pull-down menu above the workspace, select "ligand + ribbon".

You should see the androgen receptor (1i37) structure model as a ribbon, with binding pocket side-chains and the ligand shown as sticks. Make sure no other object is shown by checking the workspace for high-lighted names.

First rotate the model using the left mouse button in the graphics window to see what the structure looks like. How many helices and beta strands are there? The overall structure is often described as a "three-layered helical sandwich". The nuclear receptor family is structurally well conserved and have about the same number of secondary structures arranged in roughly the same way. The structures in this excercise are all of steroid hormone nuclear receptors, which is a sub-family of nuclear receptors that all bind steroid ligands (see figure) that makes them even closer in structural similarity.

Speaking of ligands, now zoom in on the ligand by clicking the "zoom in on ligand" link in the html window. If you wish you can adjust the clipping planes using the middle mouse button in the right edge of the graphics window, see step 2 above. You can return to the predefined view by clicking the "zoom" link again.

Look at the ligand and see which interactions exist. Two important factors for ligand recognition are shape and electrostatic complementarity. Let's begin with the electrostatic properties. Remember that hydrocarbon functions are hydrophobic, and that polar and charged functions are hydrophilic. What properties does the ligand have? Hint: Think of the molecule as having three regions along its long axis, which are hydrophobic and hydrophilic? Are there hydrogen bond donors/acceptors?

Next, look at the residues that line the binding pocket. The stick representation of binding site residues is needed for this, make sure it is displayed by selecting "ligand+ribbon" on the pulldown menu above the workspace. Are the properties of the ligand matched by the protein? If you need to be reminded about the properties of residues, click the "color residues by residue polarity" button in the html window. Hydrophobic residues will be colored green, negatively/positively charged red/blue, and polar yellow.

List the four residues that make specific polar interactions with the ligand. Hydrogen bonds are not displayed, but can you see which pairs of heteroatoms (nitrogens and oxygens) that are in good position to form hydrogen bonds? Remember that hydrogen bonds are around 3-5 Å ngström between the heavy atoms. For comparison, the covalent bonds between heavy atoms is roughly 1.5  Ångström. What is the function (and properties) of the other residues? You can turn on residue labels by clicking the link in the html window.

Next, we'll have a look at the shape complementarity of the protein and ligand. The shape of the pocket is easier to see if the surface of the protein around the ligand is displayed. These surfaces are available as precalculated objects that can be turned on or off using the links in the html window. Click the 1i37 pocket "ON", click the zoom to ligand link and see how the shape of the ligand and pocket matches. Remember that the pocket was calculated from all atoms except the ligand, so a perfect fit cannot be expected.

Now look at the other natural steroids progesterone, estradiol, and 9-fluorocorticosterone on the paper hardcopy. Take a few minutes to predict which ones would also match the pocket shape just by looking at their 2D structure. Which one would show the worst match? Write it down. Also image that you are to modify the receptor in a way that it gets specific for another ligand. How would you re-design the protein?

Note that some ligands are extremely similar, and yet their receptors can recognise them with very high specificity!

Now that you have ideas how other ligands would match the shape, check the answer by displaying them using the workspace panel. List the components of a particular pdb entry by clicking the triangle, and click on the ligand's name (three letter code, not "a", "water", "acbm" or "apo4"). Were you right?

There is only a small shape difference between the ligands progesterone (1a28 - astr) and testosterone (1i37 - dht), and both may seem to match reasonably. Display the pocket of 1a28 along with 1i37 to see where the shape differs.

**4. Analysis of the other pdb entries with cognate ligands.**

Repeat the steps above with the progesterone (PR, 1a28) and estrogen receptor (ER, 1gwr). Figure out how they have evolved to specifically recognize their ligands.

**5. The effect of a mutation**

Bring up only the structure of the AR receptor, the bound testosterone and the side-chains of the binding pocket. Add residue labels to the pocket by clicking "labels on" in the html window. Rotate the view such that you can identify residues T877 and L701. Now suppose that these residues were mutated to alanine and histidine, respectively (T877A, L701H). This double mutant has been observed for a prostate cancer patient that suffered from the so called androgen independent syndrome, which means that the tumour can grow in the absence of androgens. It was found that it can be activated by progesterone and corticosteroids. Look at the 2D structures of the ligands and the 3D models on screen, and try to explain why the double mutation can lead to activation by these two non-androgen hormones.

When you know the answer, you can discuss it with the teacher or check what the authours of these two papers thought:
Matias PM et al. (2002) J. Med. Chem. 45(7):1439 (the report of the structure)
Krishnan AV et al. (2002) Endocrinology 143(5):1889

**6. Binding of the coactivator**
The early steps of transcription activation by nuclear receptors is that a conformational change occurs upon binding of ligand, which triggers the binding of coactivator proteins to the ligand binding domain. For the coactivator to bind, it is required that the C-terminal helix (helix 12) adopts the conformation that we saw for the hormone receptors previously. This creates a hydrophobic patch where the coactivators hydrophobic residues will bind to. We will briefly look at two structures to understand how this occurs.

Click the "coact" tab in the html window. Then click "display the open NR conformation" to bring up the structure model of the retinoic acid X receptor alpha (RXRalpha) that was crystallized in the absence of a ligand. Note that most of the fold is the same, but that the red C-terminal helix is not folded against the rest of the protein as was the case for the ligand-bound hormone receptors that we've looked at previously.

Next, switch to the ligand bound structure of the thyroid hormone receptor (TR) with the pdb code 1bsx, by clicking on the link "display thyroid hormone receptor. Notice the new component in the upper left corner - the magenta-colored helix of the coactivator. Zoom in on it by clicking on the "center on coactivator peptide" link. Notice how there are multiple interactions between hydrophobic residues of the coactivator (cyan colour) and the TR (green color). The receptor side chains of a lysine and a glutamic acid are also shown. Those that are interested are welcome to suggest the function of these residues for binding of the coativator.)

**7. How two drugs prevent the activation of their receptors**
The importance of the hydrophobic binding surface to form in order to activate transcription is exploited by some antagonists (drug ligands). Look at the ligand structures on the hardcopy - the compounds raloxifene and tamoxifene are used for breast cancer prevention and treatment. They act by binding to the estrogen receptor and preventing coactivator binding. Now let's see how this happens.

Go to the drugs tab in the html window and click "Display the estrogen receptor". This is the same structure that we looked at previously. Click "zoom in on ligand binding site" and follow the transition so you know the orientation of the structure, which is slightly different than the one for hormone receptors. The C-terminal helix that is important for coactivator binding is left of the ligand. Now add tamoxifene from the 3ert structure model by clicking "display tamoxifen". Tamoxifen carbons are coloured green to distinguish tamoxifen from estradiol. The antagonist structure is added, superimposed on the 1gwr ER structure. Notice how tamoxifen has a long "arm", sticking out in the direction of the C-terminal helix, that causes steric clashes with three residues near helix 12. When tamoxifen binds, it also prevents the positioning of helix 12 for coactivator binding. Click "display raloxifene" and notice how it acts in the same manner as tamoxifen.

So where does helix 12 go when tamoxifen binds? Click "display tamoxifen-bound ER" to find out. The ribbon of 3ert (the tamoxifen structure) is colored green to distiguish it from the estradiol-bound structure, so you need to trace the ribbon to find out where the C-terminal helix 12 is. Do you recognise the new positioning of helix 12? For clarity, you may want to remove 1gwr from the display using the workspace window. Think a while before you continue reading.

You have seen this conformation before - helix 12 now binds to the coactivator binding site! Therefore, coactivators cannot bind and the receptor cannot stimulate gene transcription. Actually, there is a conserved sequence motif in coactivators LXXLL that is used for binding to the ligand binding domain, and a similar sequence motif exists in helix 12 of several nuclear receptors. If you want to find out more, check the literature references for the relevant pdb entry by looking up the pdb code at www.rcsb.org.

When you are done, quit from ICM and delete the icb file from the UNIX command window:

```
rm nuclearReceptors.icb
```