PyMOL practical exercise

A few important notes before you start…

- Reinitialize your PyMOL workspace (`File -> Reinitialize`) at the beginning of each new exercise.
- Whenever you find written "Load […]" use the `Plugin -> PDB Loader Service` option from the external interface to load the PDB file .
- Every time that "enter the command" is mentioned, it refers to typing the suggested line of text in the command line area of either the external or the internal GUI of PyMOL.
- For most purposes, starting from the cartoon representation of molecule is the most practical thing to do. Unless stated otherwise, proceed to hide everything/show cartoons for each loaded PDB file.
- At the end of each section, before reinitializing PyMOL for the next exercise, you can save a snapshot ("scene") of the current PyMOL status selecting the command (external GUI):

  `File -> Save Session As…`

  then save the file where you can retrieve it later (e.g. on your network workspace). You can recall the session at a later stage for further processing.

Part I – Proteins (secondary structure, co-factors, & domains)

1. Myoglobin.

Load the myoglobin structures with (PDB ID: 1BZR) and without (PDB ID: 1BZP) carbon monoxide bound.

a. What secondary structure elements are present in the structure?
b. Try to hand-draw a schematic figure showing the arrangement of the secondary structure elements and how they are connected.
c. Where is the haeme group located? (hint: `S -> organic`)
d. How does the heme group bind? (hint: show sequence, locate the haeme group – under the name HEM – then select it either by clicking on it in the sequence or with the command "`select resn HEM`"; use the expand selection/show stick action sequence that you used with the calcium ion in the calmodulin structure).
e. What effect does CO binding have on the iron co-ordination? (CO is represented as "`CMO`"; issue the command "`select resn CMO`", turn on the spheres visualization for the new `sele` object, then type the command "`set sphere_scale, 0.6`")

2. Haemoglobin

Load deoxy haemoglobin (PDB ID: 2DN1) and oxy-haemoglobin (PDB ID: 2DN2), switch to the cartoon representation.

a. How many molecules are present in the deoxy haemoglobin structure? (hint: use the "color by chain" action to discern between different subunits)
b. How many subunits do you normally find in haemoglobin (oxy-haemoglobin)?
c. align the two structures entering the command:

```
align 2DN1, 2DN2
```

   How do the oxy- and deoxy- forms of haemoglobin compare with each other? (hint: define two new objects corresponding to two aligned chains, hide everything except for the two aligned objects)
d. What kind of movements can be seen, what kind of distances are involved? (hint: select the haeme groups then expand the selection - `A -> modify -> expand -> by 8 A` - and show haeme-coordinating residues as sticks).

e. Load myoglobin (PDB ID: 3RGK), align it on the deoxy-hemoglobin with the command:

```
align 2DN2, 3RGK
```

   How does haemoglobin compare with myoglobin?

f. The disease sickle cell anemia ($\alpha_2,\beta^s{}_2$) is caused by a single DNA nucleotide mutation resulting in a substitution from Glutamate to Valine at position 6. Locate this residue on the deoxy-haemoglobin sequence (hint: `Display -> sequence` from the external GUI menu). Is it close to the haeme group? How might this amino acid substitution result in the disease?

3. Triose phosphate isomerase (TPI)

Load the TPI (PDB ID: 1NEY). You may wish to “create an object” of just one of the monomers, rather than work with the dimer.

a. What are the characteristic secondary structure elements (hint: colour by secondary structure)?
b. Assign the molecule a secondary structure based colour scheme, the select the first option in the “`colour by element`” object sub-menu (only affects non-alpha carbons). Select the beta sheet-portion of the molecule by entering the command:

```
select ss s
```

   Examine the hydrogen bonding characteristics of the parallel strands using the measurement wizard (hint: show sticks to visualize interacting residues between different strands).
c. What class of proteins does TPI belong to?
d. Where are the hydrophobic side-chains located? (hint: use the hydrophobic.pml script)
e. Where are the hydrophilic side-chains located? (hint: use the hydrophilic.pml script)
f. Locate the substrate DHAP (hint: `show -> organic -> sticks`)
   i. Which residues are involved in binding?
   ii. How is the phosphate stabilized?
g. Reinitialize the PyMOL workspace, reload the structure of the open TPI conformation (PDB ID: 1NEY), then load the closed form of TPI (PDB ID: 2V2D). Align the two structures with the command

```
align 2V2D, 1NEY
```

   Which part(s) moved?

4. Domains and movement: Adenylate kinase

Adenylate kinase is a phosphoryl transfer enzyme that catalyses the inter-conversion of adenine nucleotides e.g 2 ADP <–> ATP + AMP. The enzyme can be found in an open form (PDB ID: 2AK3) and a closed form (PDB ID: 1ANK). Load the two structures in PyMOL, represent them with cartoons, colour them using one of the two different secondary-structure based colouring schemes (`C -> by ss -> …`).

a. Which secondary structural elements do you find in the enzyme, and which fold do they represent?
b. What substrates can be found in the open form? Closed form? (hint: show “`organics -> sticks`”)
c. Where do the substrates/products bind?
d. How is the binding relative to the strands? (hint: is it at the N-terminal end, or C-terminal end of the strands?)
e. How is the binding co-ordinated? (hint: select the substrates by entering the command “`select organic`”, the expand the selection and show the co-ordinating residues with sticks)

Domains are “sub-sections” of a protein that maybe associated with a particular function. At times can be somewhat subjective, domains do have distinct catalytic activities associated with them.

f. Are there any domains present? If so, where might they be?

Extract as a new object one single chain for each PDB file (right click on one of the chains, `chain -> extract object`, generating `obj01` for 2AK3 – the open form - and `obj02` for 1ANK – the closed form); align the two objects by entering the command:

```
align obj01, obj02
```

g. Which parts move in response to substrate binding? How large are the movements?

Part II – Interactions (local, subunit and environmental)

5. Helix-Helix interactions: Tropomyosin

Tropomyosin (PDB ID: 1IC2) is an actin-binding protein which is important in many processes including muscle contraction. The protein contains a classic helix-helix interaction, namely “knobs in holes”, where the side-chains of one helix fit into holes created by the other helix.

Load the tropomyosin PDB file: the crystal structure contains 4 molecules, so you will need to use the `extract object` feature (right click on a chain `-> chain -> extract object`) to simplify things (hint: hide everything, switch to the cartoon structure display, then `color by chain` to be able to locate the different chains). Create an object for two of the interacting helices. Represent two interacting helices as cartoons, sticks, surface, mesh, *etc etc* to examine how the two structures interact.

a- Which residues are responsible for the helix-helix interactions?
b- Are the interactions hydrophilic or hydrophobic? (hint: you can take advantage of the hydrophobic.pml and hydrophilic.pml scripts).

6. Sub-unit interactions: Short Chain Dehydrogenases (SDRs).

The short-chain dehydrogenases are a large family of enzymes that typically utilize NADH, or NADPH, as a co-substrate to catalyse a variety of different substrates. Despite only *ca*. 25-30% sequence identity, the enzymes share a common 3D-structure. The substrates range from steroid hormones, to secondary metabolites (e.g antibiotics), and vary greatly in size. The enzymes contain a nucleotide binding-domain, and a substrate-binding domain, which is where most of the structural variation can be found due to the diversity of substrates.

a. Examine the clavulanic acid dehydrogenase (PDB ID: 2JAH).
   I. Which secondary structure elements are present?
   II. How are they arranged?
   III. What type of fold is present?
   IV. Can you identify the two domains?

b. Rename the 2JAH molecule as “D”, then load (and rename) the following molecules belonging to the SDR family using their PDB IDs: 2HSD (rename as “A), 1GEG (rename as “C”), 1DHR (rename as “E”), 1FJH (rename as “F”) ”) , 1EDO (rename as “G”). Align the molecules issuing the commands:

```
cealign A, C
cealign A, D
cealign A, E
cealign A, F
cealign A, G
```

   Take a look at the differences among the aligned structures:
   I. Which parts of the molecule are most similar/conserved?

II. Which parts of the molecule are most different?

c. Look at a typical tetrameric SDR, 2JAH ("D"). How do the:
   I. A-C monomers interact?
   II. A-D monomers interact?
   III. A-B monomers interact?
   IV. Which interactions do you think is the strongest? The weakest?

d. Two of the SDRs in the alignment are incapable of forming tetramers, but instead form dimers. Using the structure alignment in (b) and your knowledge of the subunit interactions from (c):
   I. Identify which two SDRs can't form tetramers
   II. Why are they unable to form tetramers?

7. Environment interactions. Membrane proteins.

The mitochondrial cytochrome *bc1* complex (PDB ID: 1BYG) and the Ferrichrome-Iron receptor (PDB ID: 1BY3) are two examples of proteins found in cell membranes. The membrane proteins contain domains that are located inside, and outside the cell, with a transmembrane region linking the two. Load in PyMOL the structures of the two proteins.

a. Which secondary structures elements do you find in the two membrane proteins?
b. Locate the trans-membrane regions (hint: take advante of the hydrophobic.pml script).
   a. What characteristics do the trans-membrane regions have?
   b. What is the length of the trans-membrane region?
c. Identify the pore formed through the membrane.

And if you have time…….. turn the page.

Part III - using your PyMOL skills to examine a DNA binding protein

8. Leucine zipper's

The GCN4-bZIP/DNA complex (PDB ID: 2DGC) can be used as a model to dissect the leucine zipper motif/DNA interaction. Load the structure of complex in PyMOL and generate its symmetric pair by selecting the selectin the command (object menu for the 2DGC object):

```
A -> generate -> symmetry mates -> within 4 A
```

Delete all the newly created objects (`A -> delete object`), showing up as "`2DGC_#######`", **<u>EXCEPT</u>** the only one which symmetrically pairs with the original 2DGC structure.

Display the molecule as cartoons, sticks, surface, mesh, or whatever you want and answer the following questions:

a. Helix interactions:
   - i- Which residues are responsible for the helix-helix interactions?
   - ii- Are the interactions hydrophilic, hydrophobic?
   - iii- Where are the leucines located in the leucine zipper? (hint: PyMOL command line: `select resn leu`)
   - iv- How are the charge distributions along the helix?
     - a. Hint- first use the macro charged.pml
     - b. Actions menu "*generate*" -> "*vacuum electrostatic*" -> "*protein contact potential*"

b. DNA:
   - i- Use the measure function to define the hydrogen bonds between A-T and G-C
     Top menu "*wizard*" -> "*measurement*"
   - ii- Measure the distance between the backbone phosphates in the major groove, and in the minor groove. What are the differences in distance? does the protein bind in the major or minor groove?
   - iii- In which groove is the methyl group of thymine exposed, and can it participate in discriminatory recognition?

c. DNA:
   - i- Which residues interact with the DNA?
   - ii- Which interactions are sequence specific?
   - iii- Which are non sequence specific?

2009 – 2011 – Alasdair K. Mackenzie and Gabriele Cordara

University of Oslo