

## MOL217 Lecture 1

### CATH & SCOP Hands-on Session

#### **Part I – Exploring SCOP**

Go to the root page: <http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.html> and click on the “?” button for the manual. You can read the overview of SCOP’s organisation here.

Return to the root page and click on the “Alpha and beta proteins (a/b)” link under “Classes”.

- a. How many “Folds” are there?
- b. What do they describe?

Click on TIM beta/alpha barrel, which is the first on the list. You will then see another list that is referred to as “Superfamilies”.

- c. What does the term “Superfamilies” refer to?

You may explore the rest of the hierarchy laid out in SCOP, namely “Family”, “Protein” and “Species”. Please note that some of the links may be dead on this site, particularly the coloured thumbnails referring to visualisations of the proteins.

In the search box at the bottom, enter “1N55” and note down the “Lineage” and the 4 PDB IDs within that list. Look at these “PDB domain entries” on the PDB website (<http://www.rcsb.org/pdb/home/home.do>) by entering the PDB IDs you have just noted in the search box. You can use JMOL to visualize those structures by clicking the “View in JMOL” button found under the image of the structure on the right hand side of the structure’s page. For each of the structures, check that you are viewing only one chain by scrolling down to the “Displaying options” on the JMOL page and select “Displaying Asymmetric Unit”. You will be able to see how similar these proteins look, especially with regard to their fold.

#### **Part II – Exploring CATH**

Click on the link to access the CATH database: <http://www.cathdb.info/>

In the “Search CATH by keywords or ID” field, type in “triose phosphate isomerase” and press the return key.

View all entries under CATH Superfamilies when the Match results in CATH are retrieved.

- d. What is the superfamily identified for this enzyme?

Click on the link for the superfamily. You can see that this superfamily is characterized by attributes such as the GO Diversity, EC Diversity, Species Diversity, Structural Diversity, Domain Organization, Enzyme Function, Functional Families, etc. For a description of these, click here:

[http://www.cathdb.info/wiki/doku?id=tutorials:workshop#the\\_hup\\_superfamily](http://www.cathdb.info/wiki/doku?id=tutorials:workshop#the_hup_superfamily)

On the navigation menu on the top left-hand corner, click on “Classification/Domains”

- e. What is the “Classification Lineage” of this superfamily of proteins? Click on “All Domains”.
- f. Can you find “1n55A00”? What does “A00” refer to? What do “S35”, “S60”, “S95” and “S100” refer to? For more information, check this:  
[http://www.cathdb.info/wiki/doku?id=faq#faqcath\\_data](http://www.cathdb.info/wiki/doku?id=faq#faqcath_data) and  
[http://www.cathdb.info/wiki/doku?id=blog:2008:08:22faq\\_questions\\_about\\_the\\_definition\\_of\\_cath\\_code](http://www.cathdb.info/wiki/doku?id=blog:2008:08:22faq_questions_about_the_definition_of_cath_code)
- g. How many proteins are there with the exact same classification as “1n55A00”?

**If you have the time...****Structure search** - [http://www.cathdb.info/search/by\\_structure](http://www.cathdb.info/search/by_structure)

Download and save the PDB file from

<http://www.rcsb.org/pdb/download/downloadFile.do?fileFormat=pdb&compression=NO&structureId=1KKO> and submit in the search field and click "Next". On the next page, submit only "chain A" in the search, and provide a job title and your email address so that the tool can inform you when the search is complete. Note that this search takes quite some time. Can you find the CATH classification for this structure?

**SSAP** - pairwise structural alignment program: <http://www.cathdb.info/cgi-bin/SsapServer.pl>. Type "1N55" and "1E15" in the two fields of the query and examine the output.