Exercise 1.

CATH

We have the structure 1GCQ.

What types of domains does it have? Check for example on the PDB page. Check the structure with chimera. Can you see separate domains?

Go to the CATH homepage (http://www.cathdb.info/). Explore the web site, what kind of information can you find here? Type the PDB code into the search box on right corner on the top.

How many domains the structure have? What type of superfamilies they have? Which class do they belong to?

If you follow the individual domain links you will get the detailed data about the domains. If you click on the sequence tab, you can find the domain boundaries as well.

In Chimera, color the domains according to the domain boundaries.

Exercise 2.

PFAM

Using the Pfam annotation, what is the domain structure of the following proteins?

RGS1_HUMAN
RK_HUMAN
ARBK1_HUMAN
SNX13_HUMAN

Do they have any common domains?
What is the function of these proteins? Do they have the same function?
What type of function additional domain provide?

In what species the common domain can be found?
(Under Species)

Which positions are important in this domain?
(Check out HMM logo)

How many different domain architecture was observed with this domain?
(Domain organisation)

Exercise 3.

Interpro

Search the sequence at the Intepro database (http://www.ebi.ac.uk/interpro/)

```
>squirrel_seq
MALPARLVPLCCLALLALPAQSCGPGRGPVGRRRYVRKQLVPLLYKQ
FVPSVPERTLGASGPAEGRVARGSERFRDLVPNYNPDIIFKDEENSG
ADRLMTERCKERVNALAIAVMNMWPGVRLRVTEGWDEDGHHAQDSLH
YEGRALDITTSDRDRNKYGLLARLAVEAGFDWVYYESRNHVHVSVKA
GTVGGGCFRETEAAQLWGDARGLRELHRAWVLAADAAGRVVPTPVLL
FLDRDLQRRASFVAVETERPPRKLLLTPWHLVFAARGPAPAPGDFAP
VFARRLRAGDSVLAPGGDALRPARVARVAREEAVGVFAPLTAHGTLL
VNDVLASCYAVLESHQWAHRAFAPLRLLHALGALLPGGAVQPTGMHW
YSRFLYRLAEELLG
```

What kind of domain information can you find? What sequence family information?
What does it say about the function of this protein?

Exercise 4.

Search for protein Q3JCG5 at Interpro.

What sequence family this protein belongs to?
Is the family annotation in agreement with the domain annotations?

What is that function and structure of the domains you can find in the sequence?

Is there any conserved motif there?

How conserved is the motif?

Find the annotated motif in Uniprot. Note the sequence.
Then find the Pfam familiy annotation at the Interpro page. Click on the Pfam link on the left.
Load the alignment into Jalview, find the motif, and check how conserved is the motifs/

Advanced way: generate the alignments yourself
Help: Carry out a sequence search (using your favourite method (BLAST (NCBI) or BLAST (UNIPROT) , save the similar sequences, align them and visualize the alignment using jalview, then check if the motif is really present in all of the sequences.

Exercise 5.

Search for specific domain architecture

You can search for domain architecture (specific combination of domains) at the Interpro database.
http://www.ebi.ac.uk/interpro/search/domain-organisation (IDA)

Are there proteins which have both the VIT and the 14-3-3 domain? Which proteins are these? What other domains are associated with these two domains?

Exercise 6. Advanced

Generate an Hmm profile on your own!!!

Let's take the structure 2KX7. Search for this protein in the PDB database. Which uniprot entry does belong to this structure? Save the sequence of the PDB structure (sequence display, then download.)

Let's try to generate a PFAM family for the region corresponding to this structure!

We are going to use the HHMER program.
http://www.ebi.ac.uk/Tools/hmmer/search/phmmer

Go to the search option, and upload the sequence that belongs to the 2kx7 structure.
Select the sequence database reference proteome rp75.
Start the search.

Once you get the result, check the conservation and coverage along the sequence.
Is the equally good everywhere?

Examine the low scoring hits. Would it be better to change the cutoff value?
For this check:

The taxonomy distribution of the hits
The domain architecture of the hits.

Can you find outliers? Make note of the outliers.

Go the Download tag, and save the multiples sequence alignment in the FASTA format.

Open the alignment in Jalview.

- color it
- filter at 80% redundancy
(Edit-> Remove redundancy) → 80
- remove empty columns
(Edit-> Remove empty columns)
- Remove the less conserved regions at both termini
- Remove outlier sequences

Save the alignment in a different name in fasta format.
File → Save As

Go back to the HMMER page.
At the search options, select hmmsearch .
Upload the alignment .
Start the search on rp75 database.

What changed compare to the previous search? What is the score value below and above the cutoff?

Now using the profile you just generated, search in the Uniprot database. How many hits did you get?

Save the profile.

Congratulations, you have successfully generated a sequence family profile!