Practical course 2
Classification and protein structure alignment

1. **Search for protein domains and their structure**

Search for the sequence of the human VASP protein in the UniprotKB/Swissprot database (www.uniprot.org). What are the different domains of this protein and what are their limits ? Use the links to the Interpro or Pfam domain database. Do you find information about the similarity with the WH1 and HOMER domains ? What is the biological function of these domains ? Do you find, in Pfam, a link between VASP and Ran_BP1 ?

Describe the protein classification of the EVH1 (WH1) domain according to SCOP (link from Pfam for instance).

From Uniprot, give the PDB codes of the structures of the EVH1 (WH1) domain. Which structure should you use and why (you can use the links to the PDB database to answer this question) ?

2. **Classification database**

A] Find on the protein databank website (http://www.rcsb.org) the proteins with the following PDB code: 3PDZ, 1Z86 and 1RGW. Download these files and the sequences in FASTA format (menu "Download Files" on the top right of the web page).

B] What is the CATH classification of these proteins (click on the "Annotations" at the top of the web page)?

C] Go to the Pfam database (http://pfam.xfam.org). Perform a search in Pfam by using the sequence of 3PDZ ("Sequence Search" tool). What is the biological function of the 3PDZ domain and of the related structures?

3. **Structure superimposition**

Pymol will be used to visualize the superimposed structures. Pymol will not be used to superimpose the structures.

A] Use *ClustalOmega* (http://www.ebi.ac.uk/Tools/msa/clustalo) to perform a multiple sequence alignment of the sequences downloaded in section 2A (paste all the sequences in FASTA format). Analyse the sequence alignment, and focus on the conserved amino acids at the different sequence positions. Are these sequences highly conserved?

B] Perform a **global** sequence alignment between 3PDZ and 1Z86 with *LALIGN* (http://www.ch.embnet.org/software/LALIGN_form.html); the sequences must be provided in plain text format and not in FASTA.

Use *PDBeFold* (http://www.ebi.ac.uk/msd-srv/ssm/ssmstart.html) to align the structures 3PDZ and 1Z86. Choose in the "Query" and "Target" fields the option "coordinate file" as a source and give the PDB file downloaded in section 2A. What are the values of the *z-score* and of the *rmsd*, and what do they mean?

Click on the "1" in the "##" column and then download the superimposed structures (click on the 2 "download" tabs). Use *Pymol* to visualize these superimposed structures. The aligned sequences according to the structure superimposition are available lower on the web page.

Compare the aligned sequences obtained by sequence alignment (*LALIGN*) and by structure superimposition.

D] Download the PDB file of 1FCF and its sequence in FASTA format. Use *LALIGN* to perform a local and a global sequence alignment between 1FCF and 3PDZ. Comment the result. Use *PDBeFold* to superimpose the structures of 1FCF and 3PDZ. What are the alignment scores? Visualize the superimposed structures in *Pymol*. What are your conclusions? Compare these results with those obtained with the sequence alignment.