# Algorithms in Structural Bioinformatics
# Structural Classification and Alignment Practical

TME-1

---

**Exercice - 1 Experimental structure determination**

---

**1-** Go to the Protein Data Bank (PDB) website (`http://www.pdb.org`) and look for the entries with PDB codes 1N0S and 1T0V.

**a-** What do they have in common? Determine the experimental methods used for solving the two structures.
**b-** Fill in the table below with the characteristics of each structure.

| | 1N0S | 1T0V |
|---|---|---|
| Resolution | | |
| R-free | | |
| Number of conformers | | |
| Refinement method | | |

**2-** Download the coordinate PDB files in text format.

**a-** Localise the information of the table above in these files. In which sections or subsections is the information given?
**b-** How many chains do the PDB files contain? Which delimiter is used to separate them? Are there any compound present in addition to the protein? Any water molecules? What are the maximum values for b-factors and to which regions do they correspond? To answer these questions, you can refer to the format of the coordinate section which presents the atomic coordinates for standard amino acids and nucleotides (ATOM records) and for non-polymer chemical coordinates (HETATM records):

| COLUMNS | DATA TYPE | FIELD | DEFINITION |
|---|---|---|---|
| 1 - 6 | Record name | "ATOM " | |
| 7 - 11 | Integer | serial | Atom serial number. |
| 13 - 16 | Atom | name | Atom name. |
| 17 | Character | altLoc | Alternate location indicator. |
| 18 - 20 | Residue name | resName | Residue name. |
| 22 | Character | chainID | Chain identifier. |
| 23 - 26 | Integer | resSeq | Residue sequence number. |
| 27 | AChar | iCode | Code for insertion of residues. |
| 31 - 38 | Real(8.3) | x | Orthogonal coordinates for X in Angstroms. |
| 39 - 46 | Real(8.3) | y | Orthogonal coordinates for Y in Angstroms. |
| 47 - 54 | Real(8.3) | z | Orthogonal coordinates for Z in Angstroms. |
| 55 - 60 | Real(6.2) | occupancy | Occupancy. |
| 61 - 66 | Real(6.2) | tempFactor | Temperature factor. |
| 77 - 78 | LString(2) | element | Element symbol, right-justified. |
| 79 - 80 | LString(2) | charge | Charge on the atom. |

---

**Exercice - 2 Protein domain assignment**

---

**1-** Search for the sequence of the human protein VASP in UniprotKB/SwissProt database (`www.uniprot.org`). How many amino acid residues does the sequence contain?

**2-** How many domains are there? Where are they localised? Use the links to Interpro, Pfam...etc Can you find an indication of a similarity to domains WH1 and HOMER? What are the functions of these domains? Is there a link in Pfam to the protein RanBP1? What can you deduce from this?

**3-** Are there structures of this protein available in the PDB? Which one would you choose and why?

---

**Exercice - 3 Structure classification**

---

**1-** Go to the PDB website (`http://www.pdb.org`) and look for the entries with PDB codes 3PDZ, 1Z86, 1RGW. Download the corresponding PDB files and the sequences in FASTA format.

**2-** From the information available on the PDB website, determine to which CATH classification these proteins belong.

**3-** Go to Pfam database website (`http://pfam.sanger.ac.uk/`). Perform a sequence search with 3PDZ sequence as the query. Determine the function of the domain corresponding to 3PDZ, and to the associated structures.

---

**Exercice - 4 Structure alignment**

---

The *Pymol* program will be used to visualise the aligned structures.

**1-** Use the program *ClustalW* (for example from the Mobyle portal `http://mobyle.pasteur.fr/`) to perform multiple sequence alignment of the sequences downloaded previously (paste the sequences in FASTA format one after the other). Analyse the obtained sequence alignment, focusing on the conservation degrees of the residues in each position. Are these sequences highly conserved?

**2-** Use the program *Stretcher* (for example from the Mobyle portal `http://mobyle.pasteur.fr/`) to perform global alignment of 3PDZ and 1Z86 sequences. Use the program *SSM* (`http://www.ebi.ac.uk/msd-srv/ssm/`) to perform structural alignment of 3PDZ and 1Z86 structures.

**a-** Enable the option "coordinate file" in the fields "Query" and "Target" as a source and provide the PDB files previously downloaded. What are the values of the z-score and root mean square deviation? What do they indicate?
**b-** Tick the "x" box and download the aligned sequences. After having clicked on the column "##", download the structures after superimposition, and visualise them with *Pymol*. You can look at the sequences aligned according to the structural alignment at the bottom of the page. Compare the sequences aligned by sequence alignment and by structural alignment.

**3-** Download the coordinate file of protein 1FCF as well as its sequence in FASTA format.

**a-** Use the program *LALIGN* (`http://embnet.vital-it.ch/software/LALIGN_form.html`) to perform local and global sequence alignment between 1FCF and 3PDZ. Can you comment on the results ?
**b-** Use *SSM* to perform structural alignment between 3PDZ and 1FCF. What are the values of the alignment scores? Visualise the superimposed structures with *Pymol*. What conclusions can you make ? Propose an interpretation of these results in parallel to those obtained from the sequence alignment.