X4BI040 - Bioinformatique structurale niveau III
X4BI040 - Structural Bioinformatics level III
(séance 1/session 1) - Friday 25th of January 2019

Part 1 - Deriving data from a subset of PDB (1 session)

You want to get from PDB, protein chains filtered at a cut-off value of 20% for sequence identity, a resolution cut-off of 1.6 angstroms, and a R-factor cutoff of 0.25. This can only be obtained from specialised servers like the PISCES (http://dunbrack.fccc.edu/Guoli/pisces_download.php).

1. Download the latest pre-compiled subset from this server that corresponds to these criteria. How many protein chains does this dataset contains ?
2. Run DSSP on one of these files. Examine the output of the files carefully. Check the documentation of DSSP online to identify the nature of the data provided by DSSP. In particular, look for the secondary structure assignment, the phi and psi dihedral angles and the relative solvent accessibility.
3. Run DSSP on the whole subset.
4. Write a small script to extract from each of the file generated by DSSP, the amino acid sequence, the dihedral angles and the relative solvent accessibilities which you will save in three separate files (.seq, .ang and .acc). Care should be taken to take into account chain breaks.
5. **Submit these sets of three files via Madoc + a 1 page report on your work + the script that you used, all packed into a single .tar.gz file.**

Keep this data for the next sessions.