Bioinformatics

Exercices– Protein Structure Prediction

1. The CspA protein (major cold shock protein) is produced by some bacteria in response to a cold shock. This protein binds to single-strand nucleic acids and it is believed to stabilized them under extreme conditions. Experimente structures are available for several members of this family, and we want to build a structural model for the CspA proteinA from *Staphylococcus aureus*, whose sequence, but not the structure, is known.
   a) Search the *S.aureus* CspA sequence against the sequences of proteins of known structures. To do so, retrieve the sequence from Uniprot (www.uniprot.org) and using the "Blast" tab, search against the PDB database. Select the hit with the best E-value and longest sequence length.
   b) Inspect the Uniprot entry of the selected template to find a PDB code for the corresponding structure, with a resolution better than 2.0 Å .
   c) In order to build the structural model for *S.aureus* CspA, we will use the automated modeling server http://swissmodel.expasy.org , with the option "Automated Mode". In "Advanced Options" you will insert the PDB code and chain of the selected template. You should also input a valid email address. After submission, the job should take a few minutes to complete, after which the web page is updated with the final results of the modeling procedure.
   d) Download the PDB file for the model ("download model as pdb"), and open it in PyMOL together with the PDB for your selected template. Compare the two structures, looking for the differences in sequence and structure.

2. We shall now consider a much more difficult case, the generation of an homology model for the protein dehydrofolate reductase from *Vibrio Cholerae*, where again the sequence, but not the structure, is known. In this case you will not find templates more than about 35% identical to our target. Repeat the steps in **1**. and observe the much greater differences between final model and template.