Lab Session 6: Prediction of protein structure

Date:

Concept Courtesy: ITQB Lisbon
OS: Windows/LINUX

Objective: Derive the structure of proteins using comparative modelling methods.

Note: 3D models obtained by comparative modelling are approximations and are as better is the homology between their sequences and known (experimental) structures.

Step1: Select a sequence to model.

Model 1: Major Cold shock protein from *Staphylococcus aureus*

The Major Cold Shock protein (CspA) is present in many bacteria and is induced by cold shock. This protein binds to single stranded nucleic acids and it is believed to be involved in the stability of mRNA upon these conditions (as a RNA "chaperone"). Our objective here is to structurally characterise the CspA from *Staphylococcus aureus*, a human pathogen that has its protein sequence determined, but for which no experimental structural information exists.

The sequence can be found in:
http://www.uniprot.org/uniprot/Q6G9F9
Staphylococcus aureus (strain MSSA476) [TaxID:282459]
MKQGTVKWFN AEKGFGFIEV EGENDVFVHF SAINQDGYKS LEEGQAVEFE
VVEGDRGPQA ANVVKL

Model 2: Antifreeze protein from *Pachycara brachycephalum*

Antifreeze proteins are found in organisms that live in cold environments. They limit the growth of ice crystals that otherwise would be very harmful for cells and organisms. The sequence of this 86 residues long protein can be found in http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-e+[uniprot-id:(Q1AMQ4_PACBR)]|[uniprot-acc:(Q1AMQ4_PACBR)]+-noSession.

MKSVVLTGLLFVLLCVDHMSSATKSVVASQLIPINTALTPAMMKAKEVSPKGIPAEEMSKIVGMQVN
RAVNLDETLMPDMVKTYQK

Model 3: Dihydrofolate reductase from *Vibrio cholerae*

Dehydrofolate reductase is an enzyme that catalyses the conversion of folic acid (folate) into tetrahydrofolic acid, a fundamental molecule in the synthesis of DNA precursors. The homology with proteins in the PDB database is much lower than the one found for CpsA from *S. aureus*, presenting a more challenging problem for comparative modelling procedures. The sequence of this protein is 157 residues long and can be found at http://www.uniprot.org/uniprot/Q7BN39 :

MKLSLMAAIS KNGVIGNGPD IPWSAKGEQL LFKAITYNQW LLVGRKTFES
MGALPNRKYA VVTRSSFTSS DENVLVFPSI DEALNHLKTI TDHVIVSGGG
EIYKSLIDKV DTLHISTIDI EPEGDVYFPE IPSSFRPVFS QDFVSNINYS
YQIWQKG

See if you can find homologues

Your first task is to determine whether your sequence has homologues in the PDB data bank. If these homologues exist, then the structure for your sequence can be determined.

You can go to the EBI (European Bioinformatics Institute) and submit this sequence to Blast:
http://www.ebi.ac.uk/Tools/blast2/index.html

- Place the protein sequence there.
- Change the database from the default UniProt Knowledgebase to Structure -> Protein Structure Sequences.
- Check the alignments and the percentage of identity between your structure and the found hits at PDB.

Model a structure by comparative modelling

Use the comparative modelling servers in automatic mode. This means that they may use different templates than the ones you have found in this first step. This first step is therefore a proof of concept.

Model the structure of your protein. Use the following web based servers to derive the structure.

1) Swiss Model
   http://swissmodel.expasy.org/
   Use the Automated Mode.

2)ModWeb
   https://modbase.compbio.ucsf.edu/scgi//modweb.cgi
   This one is a front end for MODELLER, a program that does comparative modelling by satisfaction of spatial restrains (which is the one we use in the lab – not this front end). This front end does everything totally automatically but uses only one structure to model your sequence (it can actually return models based on different targets), it does not mix targets. For that you need to really use MODELLER, like a pro...
   You need to place your name, email (it does not work interactively, like Swiss Model). You also need to place the MODELLER Key. Use also a Dataset name to identify your job in the Queue. Place your sequence and select in "Other options" "Very Fast".

The MODELLER license key is MODELIRANJE. You may use it if prompted

Exercises:

- Modelled structures will (may) be different from different servers
- Load it in PyMol/SPDBV as well as one of the hits you have found.
- Align the proteins using the align command, Calculate RMSD
- Load one representative structure in VAST, and see whether you can fetch other model pdb structures

  VAST: https://structure.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml.

- Generate VAST score comparison values for structural alignment.