Bioinformatics, computer lab XI

University of Potsdam WS 2019/2020, S. Hartmann

Deciding on a method for protein structure prediction

In this computer exercise, you will compare two proteins sequences with unknown structure against the sequences for which structures are known. Based on the BLAST results, you will then determine which of the three approaches for protein structure prediction is the best choice in each case.

1. Prunasin hydrolase from black cherry leaves and seeds (*Prunus serotina*) is involved a multi-step defense mechanism against herbivores. The sequence of prunasin hydrolase has been determined, but the structure of this protein has not yet been resolved.

✎ Download the prunasin hydrolase protein sequence with the Accession number AAL39079 from NCBI's GenBank.

✎ Using the online BLAST tool, compare the prunasin hydrolase protein sequence with sequences for which a structure is available in PDB.

- what is the E-value of the best BLAST hit?
- how do you statistically interpret this result?
- inspect the alignment of prunasin hydrolase and the best BLAST hit: how similar are the two sequences? Does the BLAST alignment cover the entire length of one or both sequences?
- what is the function of the protein resulting in the best BLAST hit?
- how do you biologically interpret this result? In other words, do you think the best BLAST hit represents a homolog?
- which approach for predicting the structure of prunasin hydrolase do you suggest? Justify your answer.

2. Q8N0U7 from human is a protein with unknown function. It's NCBI Accession number in GenBank is Q8N0U7. Homologous proteins have been found in many other animals, but their function remains unknown.

✎ Download the Q8N0U7 protein sequence from NCBI's GenBank.

✎ Using the online BLAST tool, compare the Q8N0U7 protein sequence from human with sequences for which a structure is available in PDB.

- what is the E-value of the best BLAST hit?
- how do you statistically interpret this result?
- inspect the alignment of Q8N0U7 and the best BLAST hit: how similar are the two sequences? Does the BLAST alignment cover the entire length of one or both sequences?
- what is the function of the protein resulting in the best BLAST hit?

- how do you biologically interpret this result? In other words, do you think the best BLAST hit represents a homolog?
- which approach for predicting the structure of Q8N0U7 do you suggest? Justify your answer.

Optional: exploring online resources

The prediction of protein tertiary structure is computationally very intensive and generally takes longer than we have available during our computer lab. Please take this into consideration and use the services wisely when exploring the following resources.

Homology modeling can be done at the Swiss server `http://swissmodel.expasy.org/`. Note that registration is required for using this and similar services.

Threading. Different servers are available for structure prediction using threading, these include

- Phyre at `http://www.sbg.bio.ic.ac.uk/~phyre/`
- HHpred at `http://toolkit.tuebingen.mpg.de/hhpred`
- GenThreader at `http://bioinf.cs.ucl.ac.uk/psipred/`