Prediction of protein disorder

Exercise 1

DISPROT database and analyzing calcineurin A

1. Find calcineurin A (PP2BA_HUMAN) in DISPROT

http://www.disprot.org/DP00092

You can start from UNIPROT, it has a link to DISPROT or search DISPROT directly by keyword (sequence search does not work).

1. What kind of information can you find on this page?
2. Which regions are marked as disordered in DISPROT?
3. Which regions are marked as ordered DISPROT?
4. By what experimental techniques?
5. What can ambiguous region mean?

Exercise 2

Disorder prediction methods

1. Collect prediction outputs for calcineurin A using various methods

The input can be:

- the amino acid sequence in FASTA format
- amino acid sequence in raw format (without header)
- UNIPROT ID or accession number

Please note, some methods are sensitive to line breaks. Minimum and maximum length of sequence

Some disorder prediction methods:

- IUPred http://iupred.enzim.hu
- Globplot http://globplot.embl.de/
- PONDR-FIT http://www.disprot.org/pondr-fit.php
- PredictProtein http://ppopen.informatik.tu-muenchen.de/ (MD, UCON, Norsp, Profbval)
- DISOPRED3 http://bioinf.cs.ucl.ac.uk/psipred/ (choose Disopred2 and 3 option!!!)
- or any other method you like.

1. Do the predictions agree with the experimental characterization of disorder?
2. Do the predictions agree with each other?
3. Which method predicts the most disorder?
4. Note the differences in the running time of the methods.

Exercise 3

MobiDB database

1. Find calcineurin in MOBIDB http://mobidb.bio.unipd.it/entries/Q08209

(or using the link from the Disprot database)

You can find the output of several other methods there.

1. Which regions are predicted as disordered by the majority of methods?
2. How are ambiguous regions are predicted?

Exercise 4

Analyze DISPROT DP00039 : a highly disordered protein

1. Predict protein disorder for DISPROT DP00039
2. Count number of positively charged amino acids
3. Count number of negatively charged amino acids
4. Calculate net charge (or use the protparam server)
5. Check low complexity segments (you can take these from PFAM through uniprot)
6. Check PFAM domains

Is there a contradiction between PFAM domain assignments and predicted disorder?

Exercise 5

Finding suitable targets for structure determination.

One of the main applications of disorder prediction methods is to find regions that are suitable for structure determination

1. Which region of this protein would you try to crystallize?

```
>mystery protein
MMQDLRLILIIVGAIAIIALLVHGFWTSRKERSSMFRDRPLKRMKSKRDDDSYDEDVEDD
EGVGEVRVHRVNHAPANAQEHEAARPSPQHQYQPPYASAQPRQPVQQPPEAQVPPQHAPH
PAQPVQQPAYQPQPEQPLQQPVSPQVAPAPQPVHSAPQPAQQAFQPAEPVAAPQPEPVAE
PAPVMDKPKRKEAVIIMNVAAHHGSELNGELLLNSIQQAGFIFGDMNIYHRHLSPDGSGP
ALFSLANMVKPGTFDPEMKDFTTPGVTIFMQVPSYGDELQNFKLMLQSAQHIADEVGGVV
LDDQRRMMTPQKLREYQDIIREVKDANA
```

Try various predictions methods, check domains, signal sequences, low complexity regions, …

2. Check, which region has a structure?

Blast the sequence against Uniprot or PDB, or do a Blast against the PDB.

Exercise 6

Disordered binding regions for human p53

1. Predict disordered binding regions for p53 Use ANCHOR http://anchor.enzim.hu and Disopred DISOPRED results can be found here : http://bioinf.cs.ucl.ac.uk/psipred/result/a17c05d4-d847-11e6-9c28-00163e110593

2. The following structures all involve the C-terminal region of p53. What type of secondary structures does this region adopt in the complex ?

   check PDBs:

   1. 1ma3
   2. 1h26
   3. 1jsp
   4. 1dt7

2. How well predictions agree with known binding region?
3. Check the predicted secondary structure for this protein (using PSIPRed that comes with DISOPRED) What is the predicted secondary structure for the C-terminal region?

Exercise 7

Filtering motif hits

Dynein light chain protein binds to disordered segments that have a TQT binding motifs. One of its known interaction partner is FA83D (Q9H4H8) with the region `VGTQTS`. We found the same sequence in the protein `ASNSD1`.

1. Do you think it can be a valid binding partner?

Hint: Predict disordered binding regions (e.g. with ANCHOR)

You can add the `VGTQTS` motif to the search too in the motif window

Is the matching region predicted to have a disordered binding region?