Wintersemester 2014/2015, AST

Übung: Protein Domains

Assignment due date: 30.11.2014
Exercise on 17.11.2014 and 24.11.2014

We will use the protein `1CXL`, a cyclodextrin glycosyltransferase, as an example in this exercise. You can also try `1QCF`, also from the lecture slides. In the homework assignment everybody will get their own protein to work on.

PDB web tools

Go to the PDB website at `www.pdb.org` and type the PDB identifier `1CXL` into the search box. You will be presented with a summary of the experimental information for this PDB record. If you have never seen a PDB webpage before it's hard to not feel overwhelmed by information.

At the top are tabs labelled "Summary", "3D View", "Sequence", etc.

A first overview of the protein can be found on the summary page in the box labelled "Molecular Description". The picture inside shows, in a linear depiction, our protein `1CXL` and how it relates to data found in other databases. At the top of the picture is data from UniprotKB, a large protein sequence database (green bar). Then comes a list of SCOP domains (orange bar), and below information from the PDB entry (blue bar). Move the mouse over the text to get more information.

Now, click on the small plus symbol just below the picture to get a more detailed view. We can see that the domain assignments for `1CXL` (and `1QCF` too) are slightly different.

On the Annotations tab, you will also be shown a list of domain assignments from SCOP, CATH, Pfam, as well as Gene Ontology (GO) information which are standardised terms for the biological processes and functions the protein is involved in.

PDB website: 3D view of domains

Click on the "3D View" tab (next to the "Summary" tab). You should see a 3D "ribbon" depiction of the protein. Keeping the left mouse button pressed allows you to rotate the 3D view, and you can zoom with the mouse wheel.

Below the 3D display there is a list of domains, assigned by three different methods. By clicking on a domain, the part of the structure belonging to this domain is highlighted in the 3D view.

PDB website: sequence view of domains

Click on the "Sequence" tab. The will be a picture showing your protein sequence and its secondary structure. In the drop-down box "Add Annotations", choose "SCOP", "CATH", "Domain Parser", "Protein Domain Parser", "PFAM", and "InterPro".

Sequence-based domain assignment: Pfam and Interpro

Unfortunately, not all information from external databases is complete on the PDB website, and currently this seems to be the case for Pfam and InterPro data. But we can still visit the Pfam / InterPro websites and do the analysis ourselves.

Another good reason to visit these sites manually is if you want to predict domains of a protein, but you only know its sequence.

The website for Pfam is `http://pfam.sanger.ac.uk/`, and InterPro can be found at `https://www.ebi.ac.uk/interpro/`. On the Pfam website, click on "Sequence Search" to be able to input a protein sequence.

We will need the FASTA sequence of our protein to submit to Pfam / InterPro. To get it, click on the Link labelled "Display Files" on the top right corner of the PDB webpage for your protein. Then you can copy and paste the sequence into the Pfam / InterPro submission form.

Your protein

Matrikelnr. endet in	PDB ID	Protein
...2507	1TAU	DNA Polymerase
...9224	1U04	Argonaute
...2020	1A36	Topoisomerase
...4818	1IGT	Immunoglobulin
...2394	1TTT	EF-Tu
...8480	3DAG	Hydrogenase
...3749	1DGS	DNA Ligase
...2760	2SRC	Protein kinase
...9714	2ADM	Methyltransferase
...0256	1NSF	SNARE protein
...8434	1DKF	Vitamin D receptor
...9692	1AX8	Leptin
...5015	1F88	Rhodopsin
...2578	3FXI	Toll-like receptor
...8842	3OS1	Integrase
...6409	3I0G	ABO blood type glycosyltransferases
...5410	4ESV	DNA helicase

...7204	**3VKH**	Dynein
...6934	**1B7T**	Myosin
...4978	**4IAR**	Serotonin receptor

Your assignment

Please submit a brief report with your answers to ast_uebung[at]zbh[dot]uni-hamburg[dot]de (please include your full name). Attach one file containing the answers to all assignments (only .txt or .pdf).

1. Skim the references given on the PDB webpage for the domain assignments of SCOP, CATH, Domain Parser, Protein Domain Parser, Pfam, and Interpro. You may also look for additional information on the Internet, e.g. Wikipedia. Write a short summary about how each method works, especially if it uses sequence or structural information.

2. If you had access to a wet lab, how else could you try and assign protein domains? Why could the lab-based results differ from the results derived from protein sequence and structure?

3. Find your protein for this assignment from the table in the section "Your protein". Navigate to the PDB webpage (`www.pdb.org`) for this protein. In a few sentences, summarise this protein's biological significance.

4. Compare the domain assignments for SCOP, Domain Parser, Protein Domain Parser, Interpro, and Pfam if they exist. Don't rely on the PDB webpage for the Pfam / InterPro results, go to their respective webpages. Also look and compare the structural domain assignments in the 3D view (under the tab "3D View"). Do you see any differences between the assignments? If yes, what might be the cause of the differences?