Sequence Analysis Spring 2000 Dr. Richard Friedman
(212)305-6901 (76901) friedman@cuccfa.ccc.columbia.edu 130BB

Lesson 12 Protein secondary structure prediction

In this class we are going to learn various methods of predicting protein secondary structure, including secondary structural class, membrane-spanning segments and topology, coiled-coils, globular regions, and signal peptides.

Reading:

1. GCG Program manual:

   A. Peptidestructure
   B. Plotstructure
   C. Pepplot
   D. Seg
   E. Coilscan

2. The following web-sites:

http://dodo.cpmc.columbia.edu/predictprotein/predictprotein.html
http://www.cbs.dtu.dk/services/SignalP/

Summary of Commands:

Note: In this document different fonts have different meanings:

Times is used to explain commands.

`Courier is used to indicate commands and command options.`

`Courier italics are used to indicate command parameters, for example, filenames.`

`Courier bold is used to indicate commands that are not displayed.`

`Courier bold italics are used to indicate computer-generated output.`

Helvetica is used to indicate menu items.

`tk`	Sets GCG so that graphics are output to the screen.
`lw`	Sets GCG so that graphics are output to a postscript file.

`lz`	Sets GCG so that graphics are output to cuccfa's laser printer.
`peptidestructure`	Predicts a protein's secondary structure by the Chou-Fasman method. Generates a *.p2s file that can be plotted by `plotstructure`.
`plotstructure`	Plots the information in a *.p2s file created by peptidestructure.
`pepplot`	Predicts protein secondary structure information, including Chou-Fasman hydrophobicity. It plots this information and can also output additional information in text files where specified.
http://cubic.bioc.columbia.edu/predictprotein/	PredictProtein (PHD) server web site. This server can predict protein secondary structure, transmembrane helices, and transmembrane topology.
return PHD predictions additionally in column format	Will return predictions in column format so that it can be pasted into a spreadsheet.
http://jura.ebi.ac.uk:8888/	JPRED site for secondary structure prediction.
`seg -win=45 -low=3.4 -high=3.75`	Predicts which regions of proteins will be globular.
`coilscan`	Predicts regions capable of forming coiled-coils.
http://www.cbs.dtu.dk/services/SignalP/	Predicts which regions will be signal peptides (better than the GCG program `spscan`).

Lab

A. Staphylococcal Nuclease

1. Obtain a copy of the sequence of staphylococcal nuclease, a protein whose SwissProt identifier is NUC_STAAU .
2. Convert the file to Fasta format and download a copy to your Mac or PC (If you use Netscape on Cuccfa via an X-Windows interface, the downloading step is not necessary).
3. Use `peptidestructure` to generate a *.p2s file that contains predictions as to the protein's secondary structure.
4. Use `tk` to set GCG's graphics output to your terminal screen.
5. Use `plotstructure` to display a graph displaying your protein's secondary structure on your Mac or PC screen.
6. Use `lw` to set GCG's graphics output to a postscript file.
7. Use `plotstructure` to create a postscript file that describes the protein's secondary structure. (Alternatively, if you have GCG-Figure on your Mac, use `plotstructure -fig` to create a Figure file). Download this file to your diskette.
   Use `pepplot` to create a postscript file (or `peptideplot -fig` to create a figure file) displaying the protein's secondary structure. Download this file to your diskette.
8. Use Netscape to access the URL of the PredictProtein (PHD) server

   http://dodo.cpmc.columbia.edu/predictprotein/predictprotein.html

9. Use the PredictProtein server to predict the secondary structure of the protein. Use the following option.

   return PHD predictions additionally in column format

10. Use Netscape to access the URL of the SignalP server.

    http://www.cbs.dtu.dk/services/SignalP/

11. Use Gapped Blast to identify the protein whose structure has been determined that is most closely homologous to staphylococcal nuclease. Download the sequence of the protein.

B. Bacteriorhodopsin

1. Obtain a copy of the sequence of bacteriorhodopsin precursor from Halobacterium halobium, Swiss-Prot identifier, BACR_HALHA.
2. Use Netscape to access the URL of the PredictProtein (PHD) server

   http://dodo.cpmc.columbia.edu/predictprotein/predictprotein.html

3. Use the PredictProtein server to predict which residues reside inside the membrane, and the membrane protein topology.

C. Human Wiskott-Aldrich syndrome protein

1. Obtain a copy of the sequence of human Wiskott-Aldrich syndrome protein (accession number u12707).
2. Use Seg with the following parameters to identify its globular and non-globular domains.

   ```
   seg -win=45 -low=3.4 -high=3.75
   ```

D. Human FOS-JUN protein.

1. Obtain a file containing the sequence of the human Fos-Jun protein (SwissProt identifier: FOS_HUMAN).
2. Use `coilscan` to obtain the regions of this protein capable of forming a coiled-coil monomer-monomer interface.

Web sites:

The web equivalent of Coilscan is available at:

http://ulrec3.unil.ch/software/COILS_form.html

Otherwise refer to web sites as given above.

Seqlab Users:

Repeat exercises using `peptidestructure`, `plotstructure`, and `coilscan` using `Seqlab`.

Bibliography:

1. References cited in the program documentation.
2. Biochemistry, L. Stryer, 5th Ed., W.H. Freeman, 1995.
3. Introduction to Protein Structure. C. Branden and J. Tooze, Garland, 1991.