

Protein Structure Prediction: Secondary Structure

Ingo Ruczinski

Department of Biostatistics, Johns Hopkins University

Protein Folding vs Structure Prediction

- Protein folding is concerned with the process of the protein taking its three dimensional shape.
- Protein structure prediction is solely concerned with the 3D structure of the protein.

Levinthal's Paradox

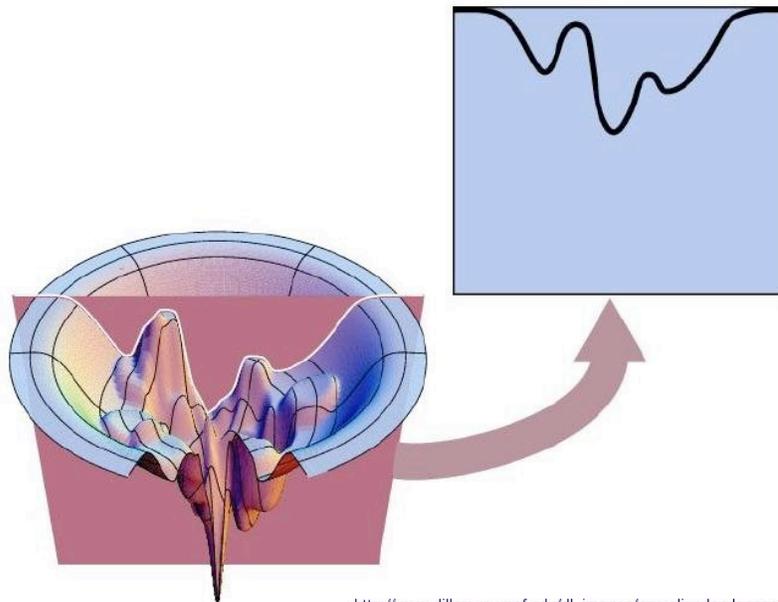
Consider this greatly simplified view of protein folding for a protein of 100 amino acids:

- If each amino acid can adopt only 3 possible conformations, the total number of conformations is $3^{100} = 5 \times 10^{47}$.
- Assuming it would take 10^{-13} seconds to change each conformation, the time required to test all conformations would be 5×10^{34} seconds, or 10^{27} years (the age of the universe is thought to be about 10^{10} years).
- Proteins usually fold in seconds or fractions thereof.

The Central Dogma

“The three-dimensional structure of a protein is determined by its sequence and its environment without the obligatory role of extrinsic factors”.

Energy Landscape



http://www.dillgroup.ucsf.edu/dl_images/one-slice-landscape.jpg

Flavors of Structure Prediction

- Homology modeling.
- Fold recognition (also called 'inverse folding', 'threading', or 'sequence-structure threading').
- Ab initio (also called 'de novo' or 'new folds methods').

Predicting protein interaction (for example docking) also has to do with structure prediction, but is not considered in this lecture.

Protein structure - Mozilla (Build ID: 2004031616)

File Edit View Go Bookmarks Tools Window Help Debug QA

Back Forward Reload Stop <http://www.cmbi.kun.nl/gv/dssp/> Search Print

Home Bookmarks Release Notes Plug-ins Extensions Support Mozilla Community Drop_Off Locator

CMBI Centre for Molecular and Biomolecular Informatics
University of Nijmegen, Toernooiveld 1, P.O. Box 9010, 6500 GL Nijmegen, +31 (0)24-3653391,
postmaster@cmbi.kun.nl

Main Search

The DSSP database

The DSSP program was designed by Wolfgang Kabsch and Chris Sander to standardize secondary structure assignment. The DSSP database is a database of secondary structure assignments (and much more) for all protein entries in the Protein Data Bank (PDB).

Information

- The [help document](#) for the DSSP program.
- The [DSSP article](#) scanned in, or as [PDF file](#)
- The [license form](#) for an academic DSSP source code.
- The [license form](#) for a commercial DSSP source code.
- A [bill](#) for commercial users.
- **AFTER** faxing the license form to the FAX number indicated at the form (+31 (0)24 3652977) you can extract the DSSP distribution by clicking [here](#) or from the anonymous FTP area of <ftp.cmbi.kun.nl>. Do a cd to pub/molbio/software and download dsspcmbi.zip . In any case, type **unzip dsspcmbi.zip** to unpack, then look at README.TXT.
- Precompiled executables are also available for [Linux](#) and [Windows](#). (The Windows .exe file was compiled under Linux using Mingw32, has never seen a Windows environment and should thus be virus-free. Download the source if you want to be 100% sure.) Under Windows the DSSP output does not make it to the console, so redirect it to a file instead: `dsspcmbi source.pdb destination.dssp >messages.txt`
- Several changes have been made to the DSSP program to solve problems with recent PDB files. These are documented in the source code.
- Commercial users are requested to transfer Euro 1000 to account number of the "Stichting WHAT IF" no. 54.83.62.262 at the ABN-AMRO in Nijmegen. Please mention DSSP. Please transfer the money before down-loading the software.
- We have a version of the PDBFINDER with the secondary structure according to DSSP indicated as 1-letter code strings. Look at the [example](#). You can download the entire file from <ftp.cmbi.kun.nl/pub/molbio/data/pdbfinder2/PDBFIND2.TXT.gz>.

start lect3 Protein structure - M... 7:48 PM

Secondary Structure Assignment

Eight states from DSSP:

- H: α -helix
- G: 3_{10} helix
- I: π -helix
- E: β -strand
- B: bridge
- T: β -turn
- S: bend
- C: coil

CASP standard:

H = (H, G, I), E = (E, B), C = (C, T, S).

Secondary Structure Prediction

Given the sequence of amino acids of a protein, what is its secondary structure?

Primary structure: GHWIATRQLIREAYEDYRHFSSECPFIP

Secondary structure: CEEEECHHHHHHHHHHCCHHCCCCC

Notation: H: Helix E: Strand C: Coil

Secondary Structure Prediction



Helix



Edge strand



Buried strand

By eye!

A Little Bit of History...

The early methods for secondary structure prediction suffered from lack of data, and were usually performed on single sequences.

1974: Chou and Fasman.

Propensities of formation based upon frequency of occurrence, rule based.

1974: Lim.

Theory based on chemical side-chain properties, very complex rules.

1978: Garnier, Osguthorpe, Robson.

Sliding window, consensus approach.

The prediction accuracy for all of those methods were roughly 50-55%.

Measures for Prediction Accuracy

The standard measure for prediction accuracy is (still) the Q3 measure. It is simply the proportion (in percent) of all amino acids that have correct matches for the three states C, E, H.

In recent years, the segment overlap measure (SOV) has been used more extensively. It aims for measuring how well secondary structure elements have been predicted rather than individual residues.

Automated Methods

The availability of large families of homologous sequences together with advances in computing techniques has pushed the prediction accuracy well above 70%. Most methods are available as web servers. They include:

PHD

<http://www.embl-heidelberg.de/predictprotein/predictprotein.html>

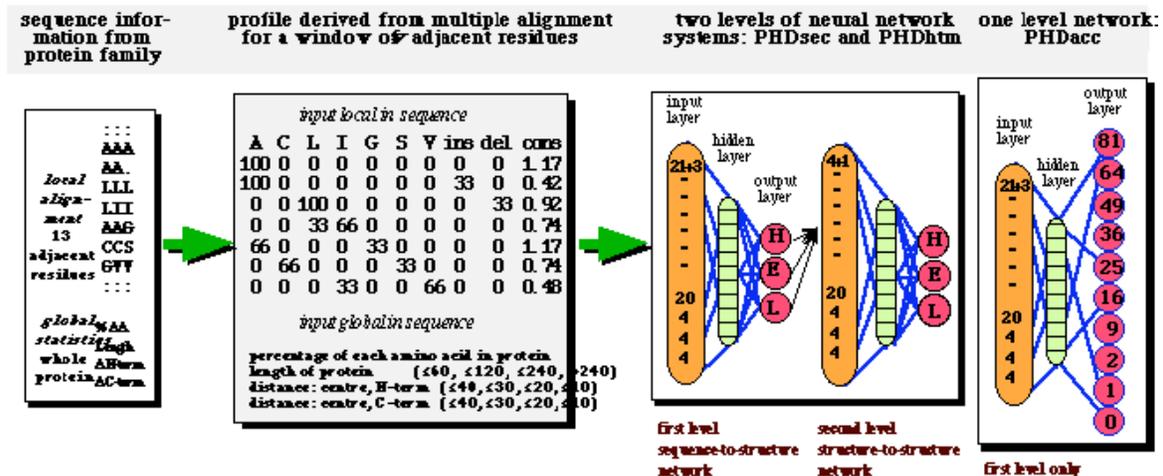
PSI-PRED

<http://bioinf.cs.ucl.ac.uk/psipred/>

JPRED

<http://www.compbio.dundee.ac.uk/~www-jpred/>

PHD



Other References

- Zvelebil, M.J.J.M., Barton, G.J., Taylor, W.R. & Sternberg, M.J.E. (1987). Prediction of Protein Secondary Structure and Active Sites Using the Alignment of Homologous Sequences *Journal of Molecular Biology*, 195, 957-961. ([ZPRED](#))
- Geourjon, C. & Deleage, G. (1994), SOPM : a self optimised prediction method for protein secondary structure prediction. *Protein Engineering*, 7, 157-16. ([SOPMA](#))
- Solovyev V.V. & Salamov A.A. (1994) Predicting alpha-helix and beta-strand segments of globular proteins. (1994) *Computer Applications in the Biosciences*, 10, 661-669. ([SSP](#))
- Wako, H. & Blundell, T. L. (1994), Use of amino-acid environment-dependent substitution tables and conformational propensities in structure prediction from aligned sequences of homologous proteins. 2. Secondary Structures, *Journal of Molecular Biology*, 238, 693-708.
- Salamov A.A. & Solovyev V.V. (1995), Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiply sequence alignments. *Journal of Molecular Biology*, 247, 1 ([NNSSP](#))
- Mehta, P., Heringa, J. & Argos, P. (1995), A simple and fast approach to prediction of protein secondary structure from multiple aligned sequences with accuracy above 70 %. *Protein Science*, 4, 2517-2525. ([SSPRED](#))
- King, R.D. & Sternberg, M.J.E. (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci*, 5, 2298-2310. ([DSC](#)).

Consensus

