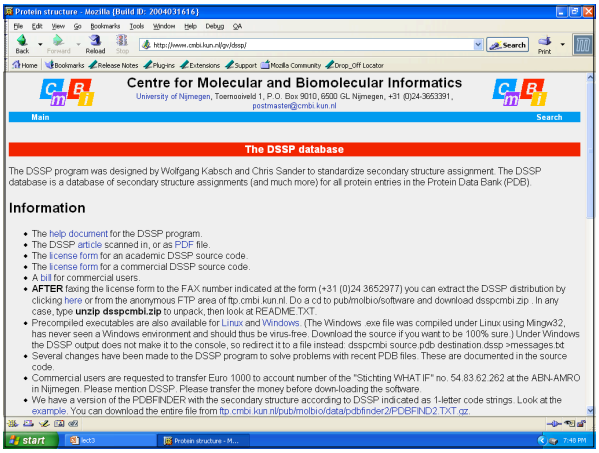
Protein Structure Prediction: Secondary Structure

Ingo Ruczinski

Department of Biostatistics, Johns Hopkins University



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: GHWIATRGQLIREAYEDYRHFSSECPFIP

Secondary structure: CEEEEECHHHHHHHHHHHHCCCHHCCCCCC

Notation: H: Helix E: Strand C: Coil

Conformational Preferences of Amino Acids

Amino acid	Preference α-helix	β-strand	Reverse turn
Glu	**1.59**	0.52	1.01
Ala	**1.41**	0.72	0.82
Leu	**1.34**	1.22	0.57
Met	**1.30**	1.14	0.52
Gln	**1.27**	0.98	0.84
Lys	**1.23**	0.69	1.07
Arg	**1.21**	0.84	0.90
His	**1.05**	0.80	0.81
Val	0.90	**1.87**	0.41
Ile	1.09	**1.67**	0.47
Tyr	0.74	**1.45**	0.76
Cys	0.66	**1.40**	0.54
Trp	1.02	**1.35**	0.65
Phe	1.16	**1.33**	0.59
Thr	0.76	**1.17**	0.90
Gly	0.43	0.58	**1.77**
Asn	0.76	0.48	**1.34**
Pro	0.34	0.31	**1.32**
Ser	0.57	0.96	**1.22**
Asp	0.99	0.39	**1.24**

Helical Preference.

Strand Preference.

Turn Preference.

Conformational Preferences of Amino Acids

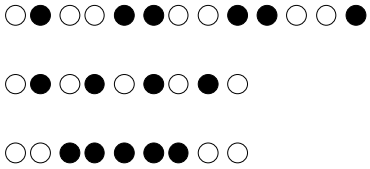
Amino acid	Preference α-helix	β-strand	Reverse turn
Glu	**1.59**	0.52	1.01
Ala	**1.41**	0.72	0.82
Leu	**1.34**	1.22	0.57
Met	**1.30**	1.14	0.52
Gln	**1.27**	0.98	0.84
Lys	**1.23**	0.69	1.07
Arg	**1.21**	0.84	0.90
His	**1.05**	0.80	0.81
Val	0.90	**1.87**	0.41
Ile	1.09	**1.67**	0.47
Tyr	0.74	**1.45**	0.76
Cys	0.66	**1.40**	0.54
Trp	1.02	**1.35**	0.65
Phe	1.16	**1.33**	0.59
Thr	0.76	**1.17**	0.90
Gly	0.43	0.58	**1.77**
Asn	0.76	0.48	**1.34**
Pro	0.34	0.31	**1.32**
Ser	0.57	0.96	**1.22**
Asp	0.99	0.39	**1.24**

Extended flexible side chains.

Bulky side chains, beta-branched.

Restricted conformations, side chain – main chain interactions.

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.

Rost et al (1994), JMB 235, pp 13-26.

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`

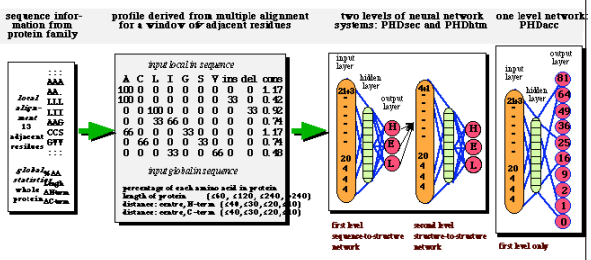
PSI-PRED

`http://bioinf.cs.ucl.ac.uk/psipred/`

JPRED

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

PHD



Consensus