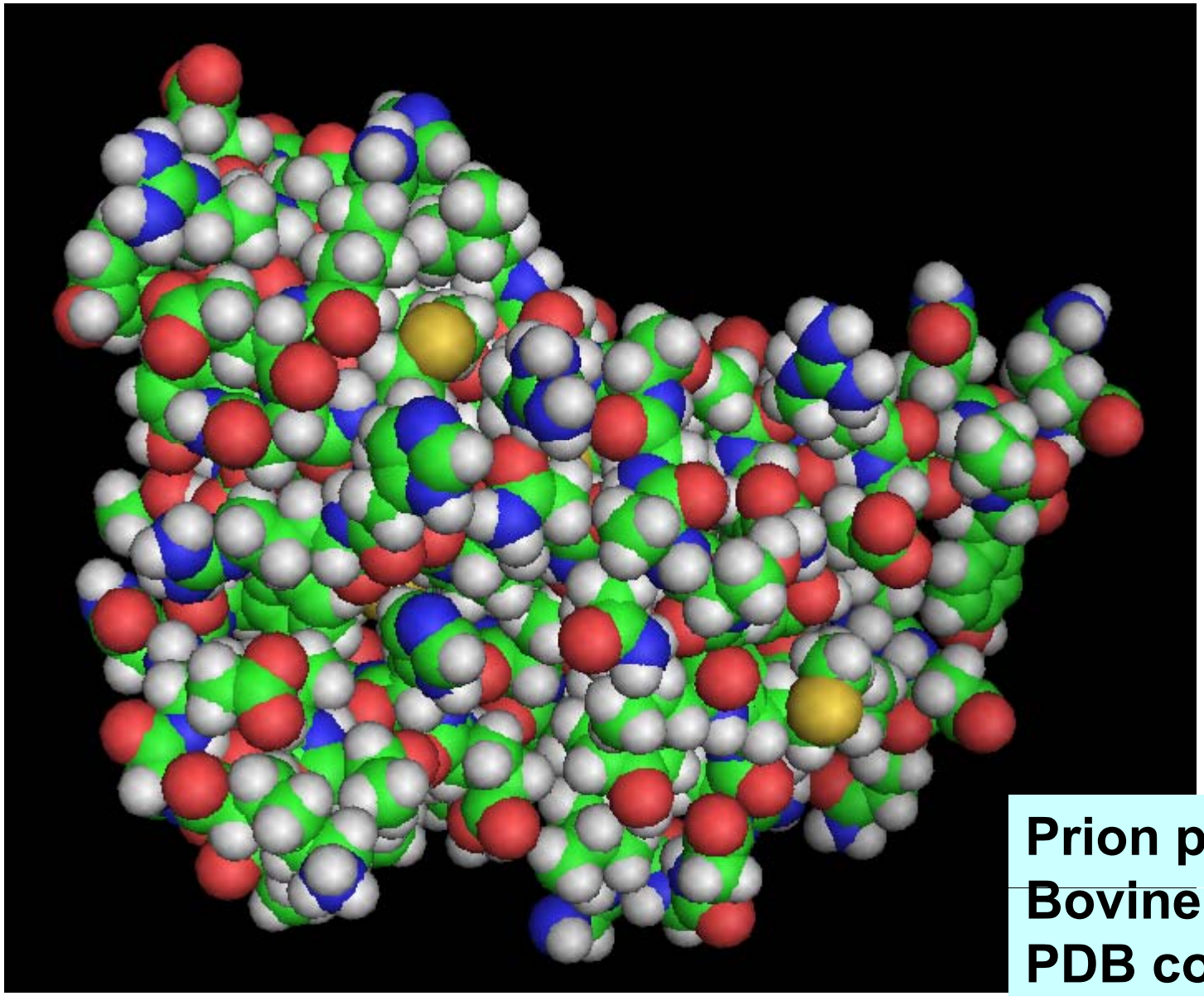
#12
Protein Structures

Topics:

- Protein
  - Translation, Codon Table
  - Main Chain, Side Chain
  - Hydropathy index, hydrophilicity, hydrophobicity
- PDB Database
- Secondary Structure
  - alpha helix, parallel beta sheet, anti-parallel beta sheet
  - Ramachandran plot
  - DSSP code
  - Secondary Structure Prediction (Chou-Fasman method, etc.)

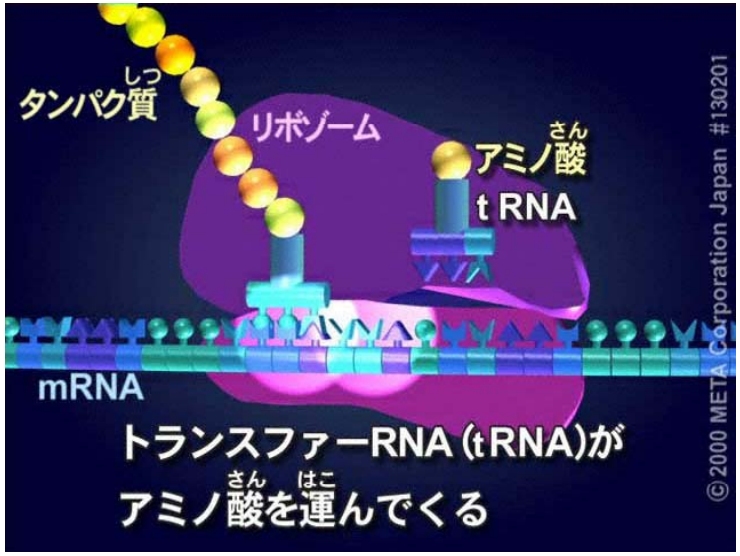
Protein Structure Example

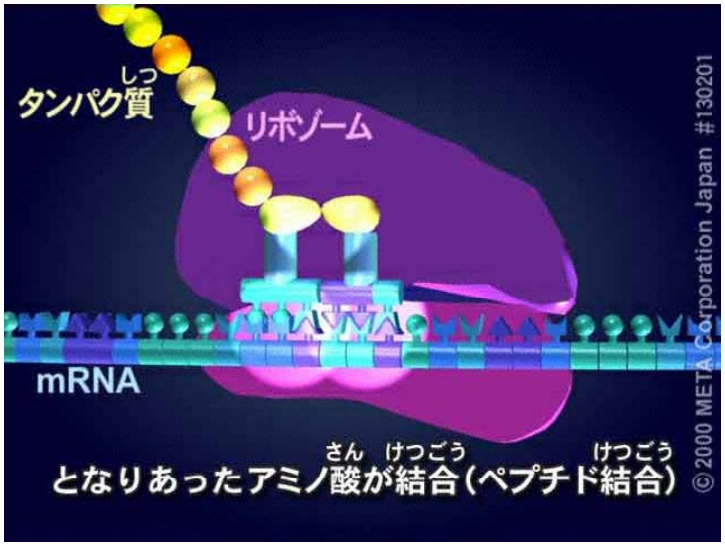


Prion protein
Bovine (cow)
PDB code = 1DX0

Translation of Protein

movie





From : IPA Educational Material Archive, http://www2.edu.ipa.go.jp/gz/

Protein

Protein = poly peptide chain(s)

poly peptide = (long) peptide
peptide = chain of amino acid

amino acid

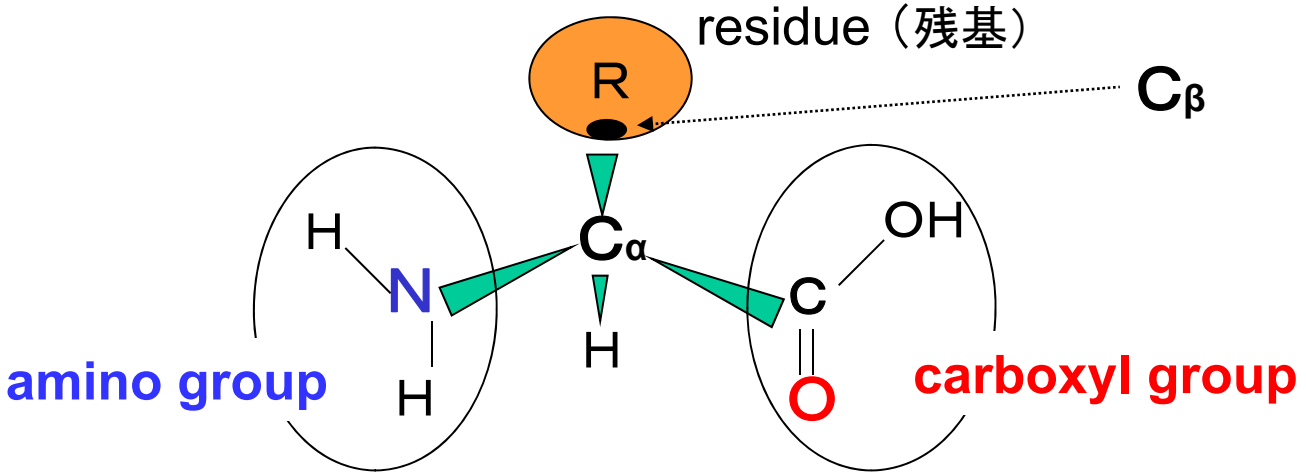
G,	A,	V,	L,	I,
S,	T,	C,	M,	D,
N,	E,	Q,	R,	K,
H,	F,	Y,	W,	P

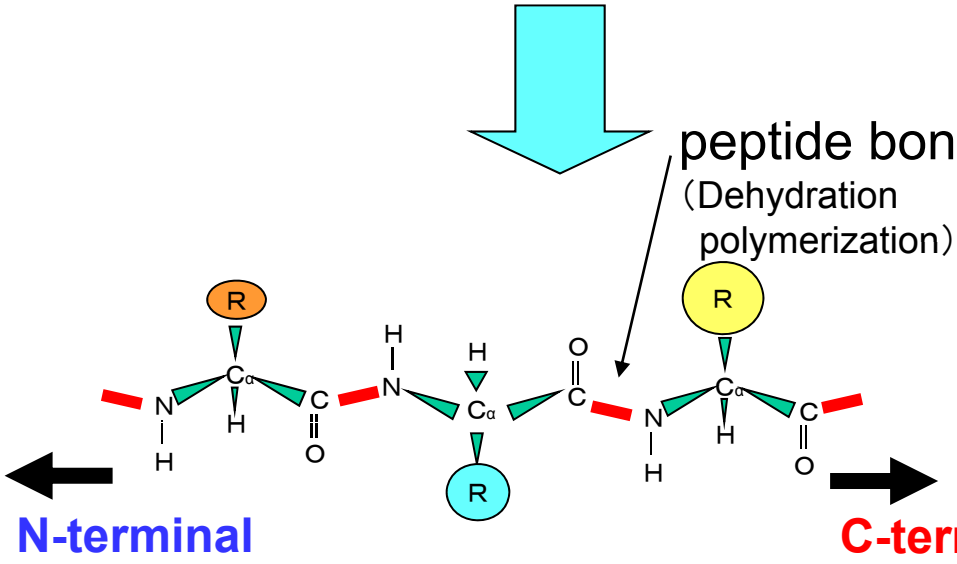
(usually) 20 kinds

From : IPA Educational Material Archive, http://www2.edu.ipa.go.jp/gz/

Protein Structure

・Amino acid as a building block





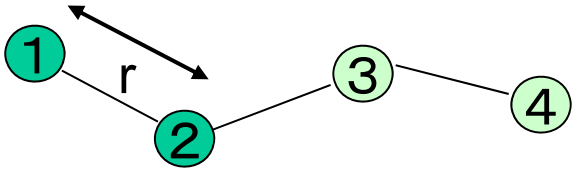
・S-S (disulfide bridge)
・Post-translational modification
acetylation,
phosphorylation,
sugar chain addition
・Dimerization, etc.

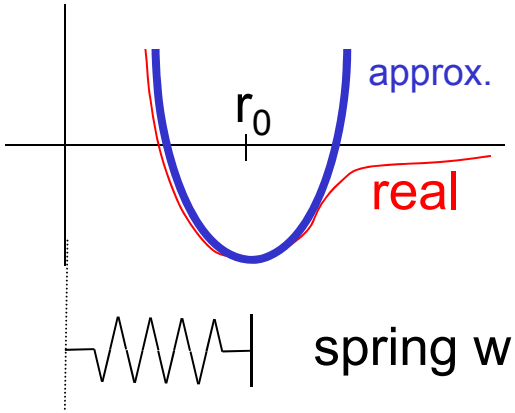
Main Chain

– **bond length (結合長)**:

approximation by Hooke's law

$E_{bond} = 1/2 * k_b * (r - r_0)^2$

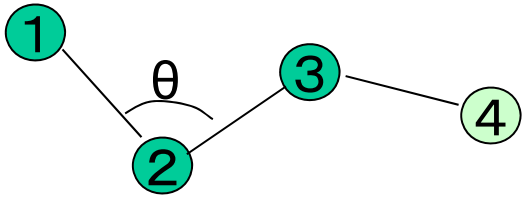




cf) Morse function $= D\{1-\exp(-k(r-r_0))\}^2$

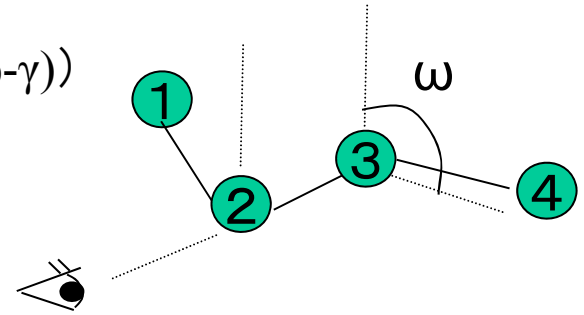
– **bond angle (結合角)**:

$E_{angle} = 1/2 * k_a * (\theta - \theta_0)^2$



– **dihedral angle (二面角)**:

$E_{torsion} = 1/2 * k_t * (1 + \cos(n\omega - \gamma))$



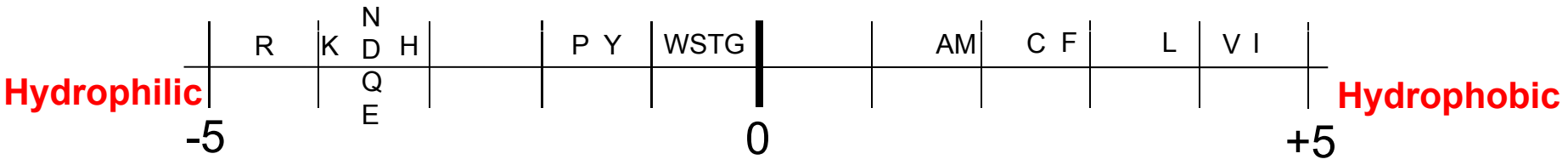
angle between
plane 1-2-3 and
plane 2-3-4

Side Chain

・20 “standard” amino acids are usually used in bioinformatics
（cf. unusual amino acids
selenocysteine = UGA codon, pyrrolysine = UAG codon）

Charged	Arg(R) , Lys(K), His (H) : Positive（basic） Asp(D), Glu (E) : Negative（acidic）
Polar	Ser(S), Thr(T), Tyr(Y), Cys(C), Asn(N), Gln(Q), Trp(W)
Hydrophobic	Gly(G), Ala(A), Phe(F), Pro(P), Met(M) Val(V), Ile (I), Leu (L)

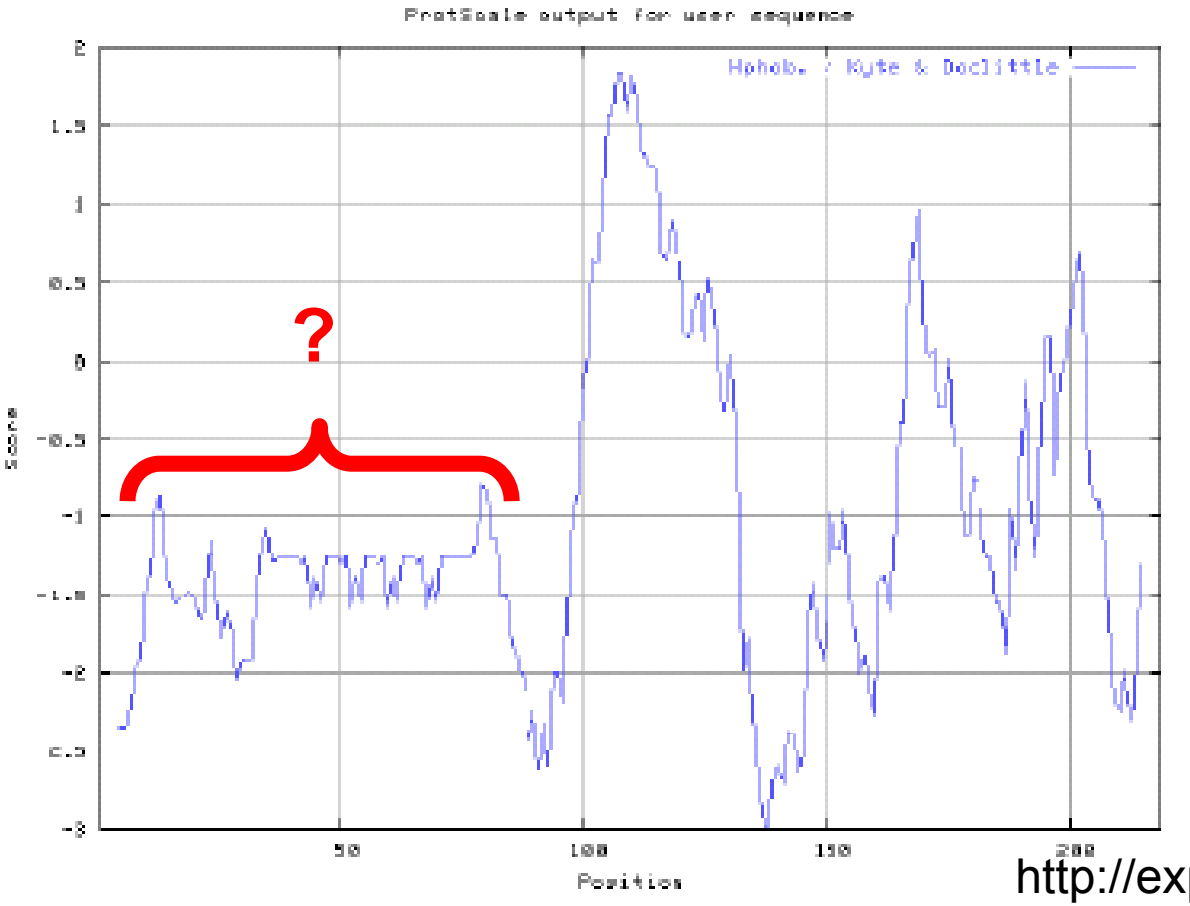
・**hydropathy index** Kyte and Doolittle: *JMB,* **157**, 105-132 (1982)



Hydropathy Plot

>1DX0

GSKKRPKPGGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGWGQPHGGGWGQPHGG
GWGQPHGGGWGQPHGGGGWGQGGTHGQWNKPSKPKTNMKHVAGAAAAGAVVGGLGGYMLG
SAMSRPLIHFGSDYEDRYYRENMHRYPNQVYYRPVDQYSNQNNFVHDCVNITVKEHTVTT
TTKGENFTETDIKMMERVVEQMCITQYQRESQAYYQRGA



Prion protein
Bovine (cow)

PDB code = 1DX0

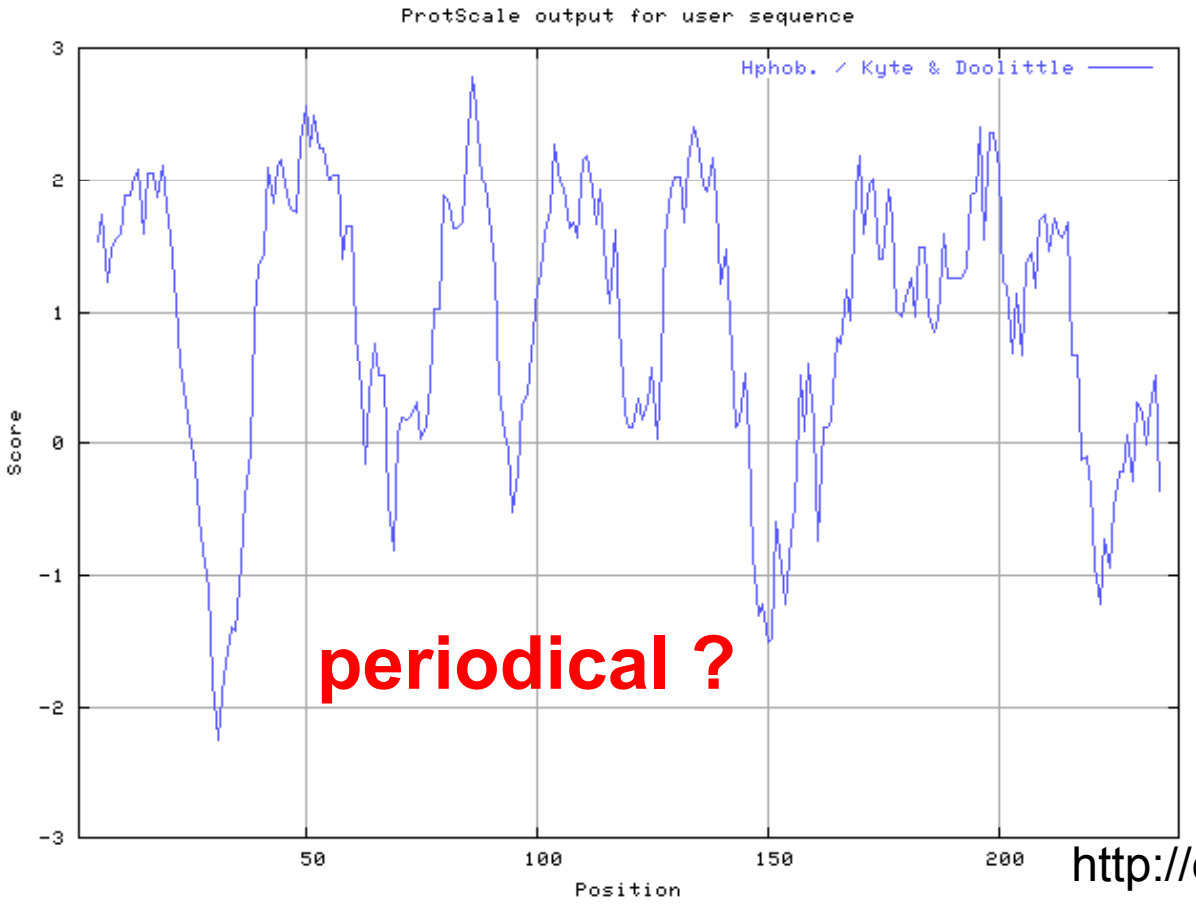
ProtScale

http://expasy.org/cgi-bin/protscale.pl

Hydropathy Plot

>1GU8

MVGLTTLFWLGAIGMLVGTLAFAWAGRDAGSGERRYYVTLVGISGIAAVAYVVMALGVGW
VPVAERTVFAPRYIDWILTTPLIVYFLGLLAGLDSREFGIVITLNTVVMLAGFAGAMVPGIERY
ALFGMGAVAFLGLVYYLVGPMTESASQRSSGIKSLYVRLRNLTVILWAIYPFIWLLGPPGVAL
LTPTVDVALIVYLDLVTKVGFGFIALDAAATLRAEHGESLAGVDTDAPAVAD



Sensory Rhodopsin II (bacteria)

PDB code = 1GU8

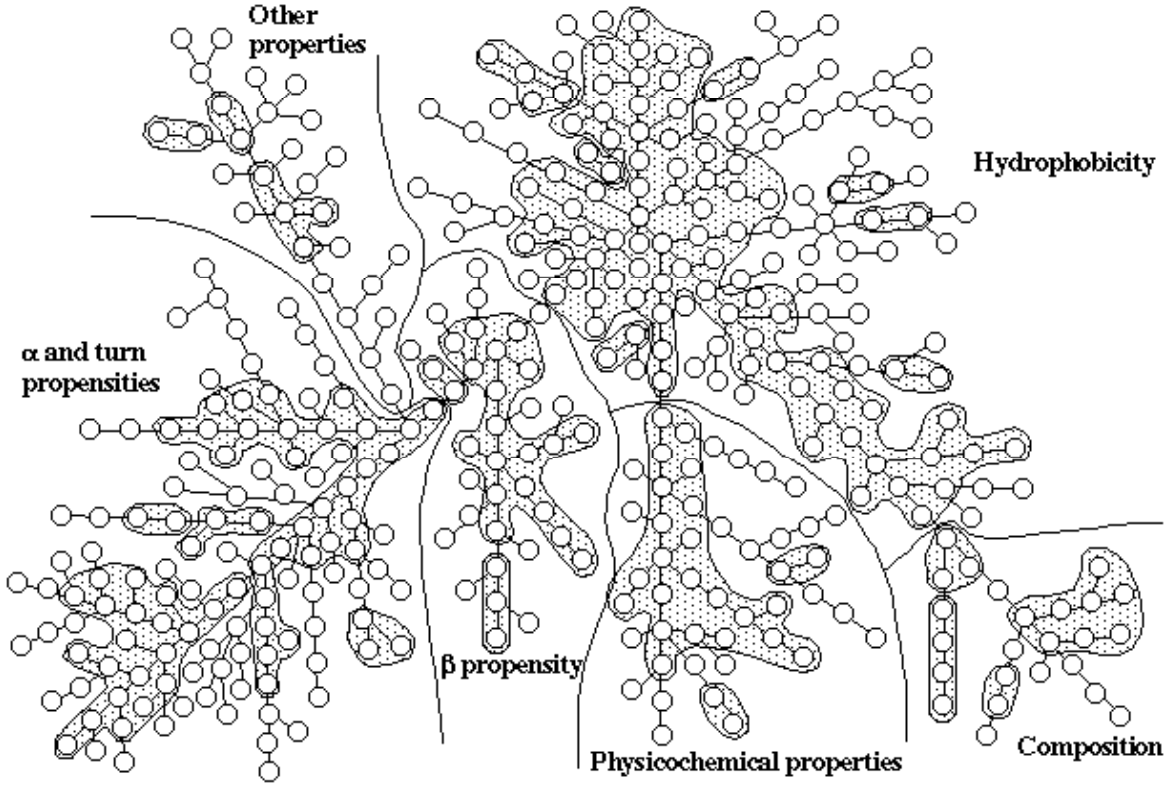
ProtScale

http://expasy.org/cgi-bin/protscale.pl

Other amino acid indices

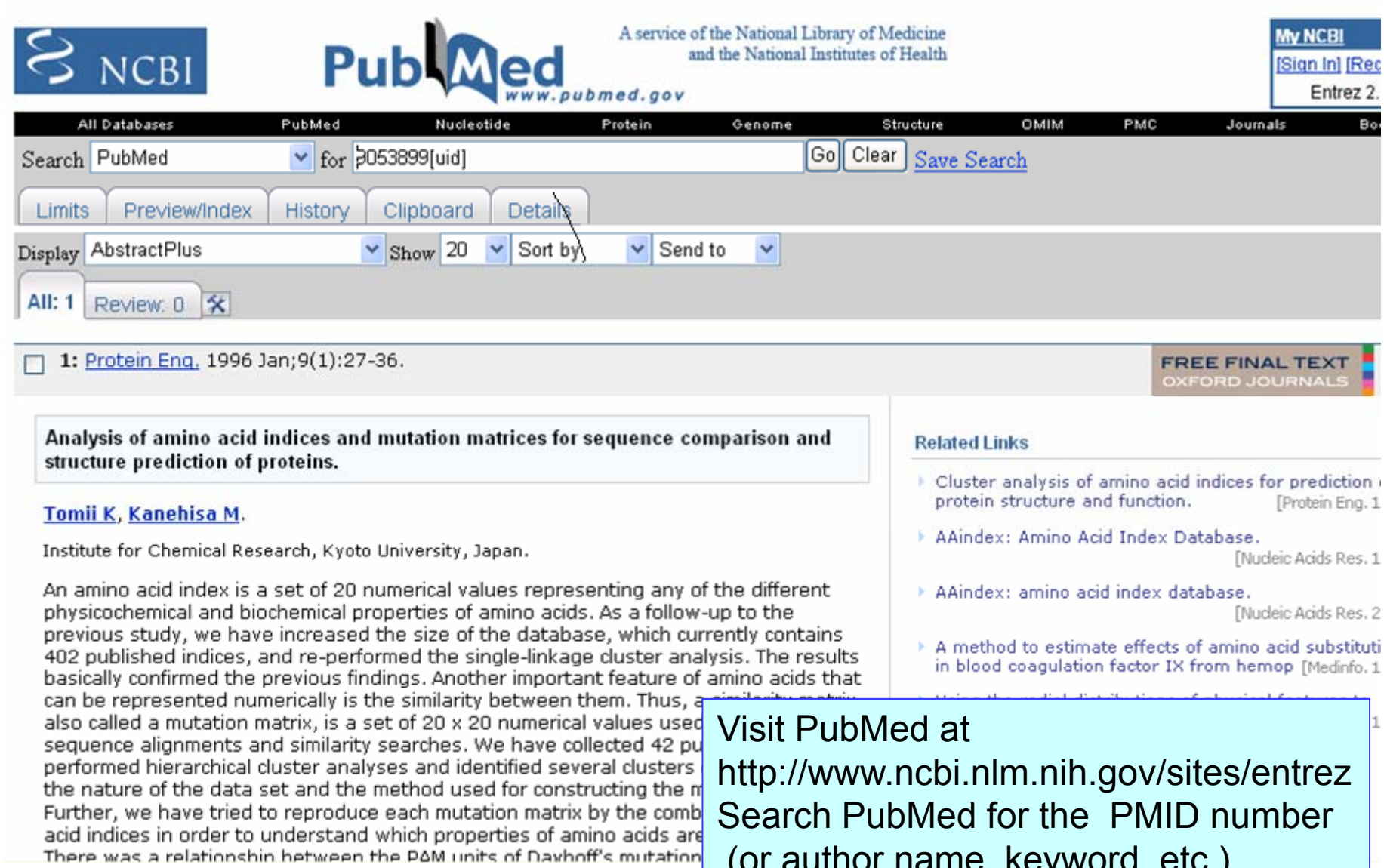
AAindex database ver 9.1 (Aug. 2006)
Collection of 544 amino acid indices.
http://www.genome.ad.jp/aaindex

ex.) Population (Dayhoff), Secondary Structure Propencity(Chou-Fasman)
Hydrophobicity (Kyte-Doolittle), Volume, Surface area, etc.



Tomii, *et.al* (1996)
Protein Eng. 9, 27-36
[PMID:9053899]

Getting PubMed abstract



Visit PubMed at
http://www.ncbi.nlm.nih.gov/sites/entrez
Search PubMed for the PMID number
(or author name, keyword, etc.)

http://www.ncbi.nlm.nih.gov/pubmed/9053899

Protein Structure Determination

- X-ray diffraction（X線構造解析）
  - Protein crystalization is needed.
  - Electron density map is obtained from X-ray diffraction data（intencity, phase）through Fourier transformation calculation
  - Then, residue assignment to the density map is requred. Several support software are available: X-PLOR, X-AUTOFIT

SPring-8

- NMR spectroscopy（核磁気共鳴）
  - Protein should be water-soluble, and smaller than 10^5 Dalton (mass).
  - NOE (nuclear Overhauser effect) cross peaks between close protons are measured.
  - 3D structure is obtained by “Distance geometry” calculation.

- cryo-electron microscope（極低温電子顕微鏡）

Protein Data Bank

- 3-D coordinates database for Protein, DNA, and Complex

Registration 58,236 entry (as of 2009.06.16)

PDB Current Holdings Breakdown

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	53562	1212	2507	17	57298
NMR	7371	914	157	7	8449
ELECTRON MICROSCOPY	201	17	77	0	295
HYBRID	21	1	1	1	24
other	125	4	4	13	146
Total	61280	2148	2746	38	66212

http://www.rcsb.org

- history
  - 1971 Started at Brookhaven National Laboratory
  - 1998 Moved to RCSB (Research Collaboration for Structural Bioinformatics)
  - 2006 wwPDB (The Worldwide Protein Data Bank) by US, Europe, and Japan.

Protein Structure Example

Prion protein
Bovine (cow)

PDB code = 1DX0

1. visit RCSB PDB site at http://www.rcsb.org/pdb/
2. make keyword search for “Prion Bovine”
3. choose 1DX0, for example

http://www.rcsb.org/pdb/explore/explore.do?pdbId=1DX0

PDB format (text file)

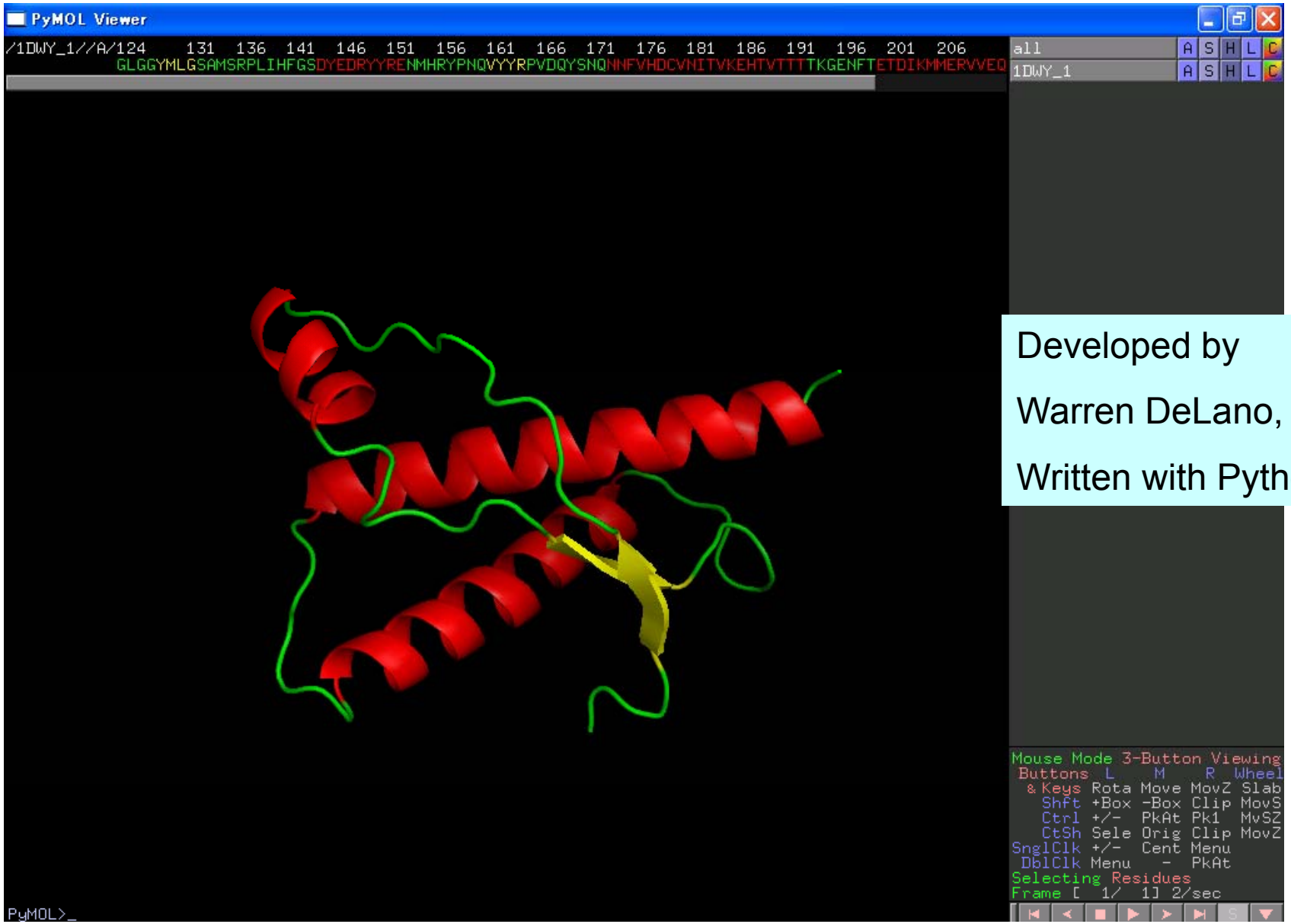
```
HEADER    PRION PROTEIN                           15-DEC-99   1DX0
TITLE     BOVINE PRION PROTEIN RESIDUES 23-230
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: PRION PROTEIN;
COMPND   3 CHAIN: A;
COMPND   4 SYNONYM: PRP, MAJOR PRION PROTEIN;
COMPND   5 FRAGMENT: RESIDUES 23-230;
COMPND   6 ENGINEERED: YES;
COMPND   7 MUTATION: YES;
SOURCE    MOL_ID: 1;
  :
```

X Y Z

first residue (GLY 124)

```
ATOM      1  N   GLY A 124      11.500 -12.139  -5.556  1.00  0.00           N
ATOM      2  CA  GLY A 124      12.523 -11.134  -5.837  1.00  0.00           C
ATOM      3  C   GLY A 124      12.362  -9.873  -4.981  1.00  0.00           C
ATOM      4  O   GLY A 124      12.934  -8.828  -5.305  1.00  0.00           O
ATOM      5  H   GLY A 124      11.666 -12.848  -4.859  1.00  0.00           H
ATOM      6 1HA  GLY A 124      12.474 -10.852  -6.890  1.00  0.00           H
ATOM      7 2HA  GLY A 124      13.507 -11.563  -5.647  1.00  0.00           H
ATOM      8  N   LEU A 125      11.588  -9.939  -3.891  1.00  0.00           N
ATOM      9  CA  LEU A 125      11.436  -8.860  -2.922  1.00  0.00           C
ATOM     10  C   LEU A 125      12.790  -8.554  -2.286  1.00  0.00           C
ATOM     11  O   LEU A 125      13.152  -7.402  -2.083  1.00  0.00           O
ATOM     12  CB  LEU A 125      10.479  -9.293  -1.794  1.00  0.00           C
ATOM     13  CG  LEU A 125       8.986  -9.494  -2.095  1.00  0.00           C
  :
```

Structure Viewer: PyMOL

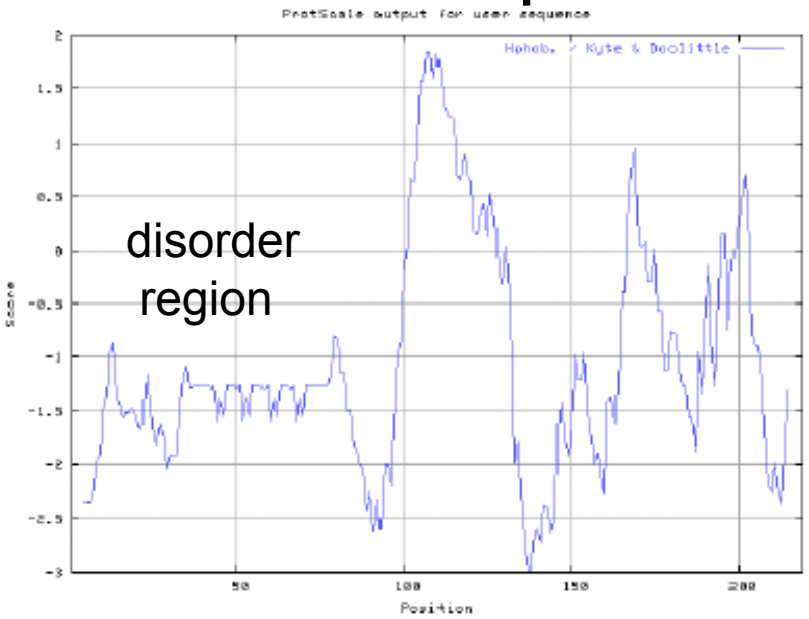


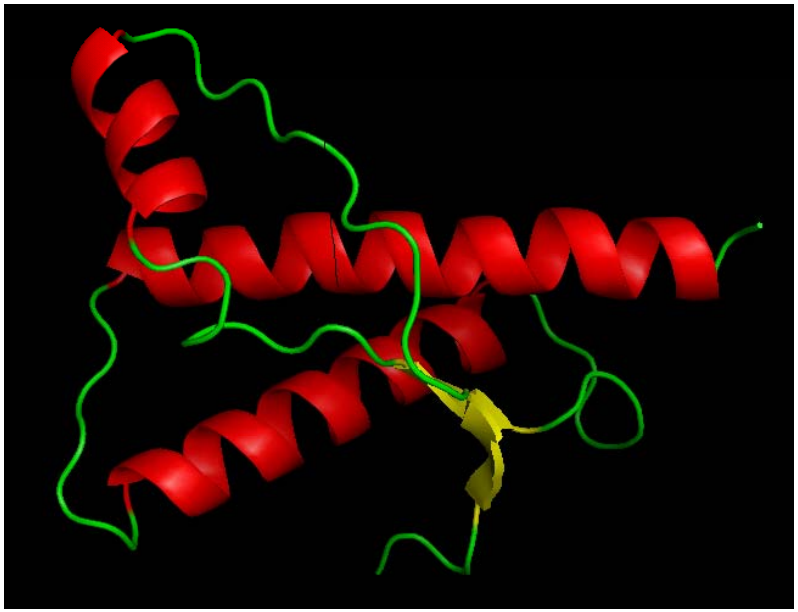
Developed by

Warren DeLano,

Written with Python

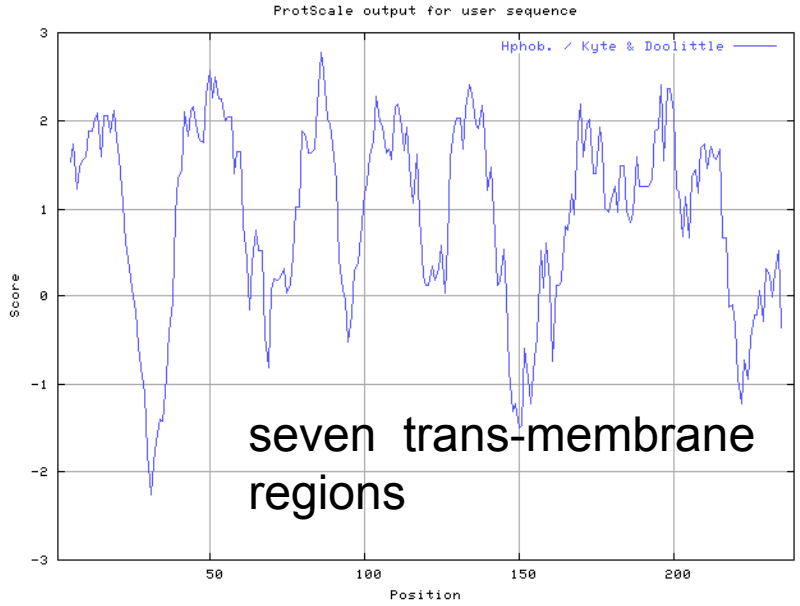
1DX0: Bovine Prion protein





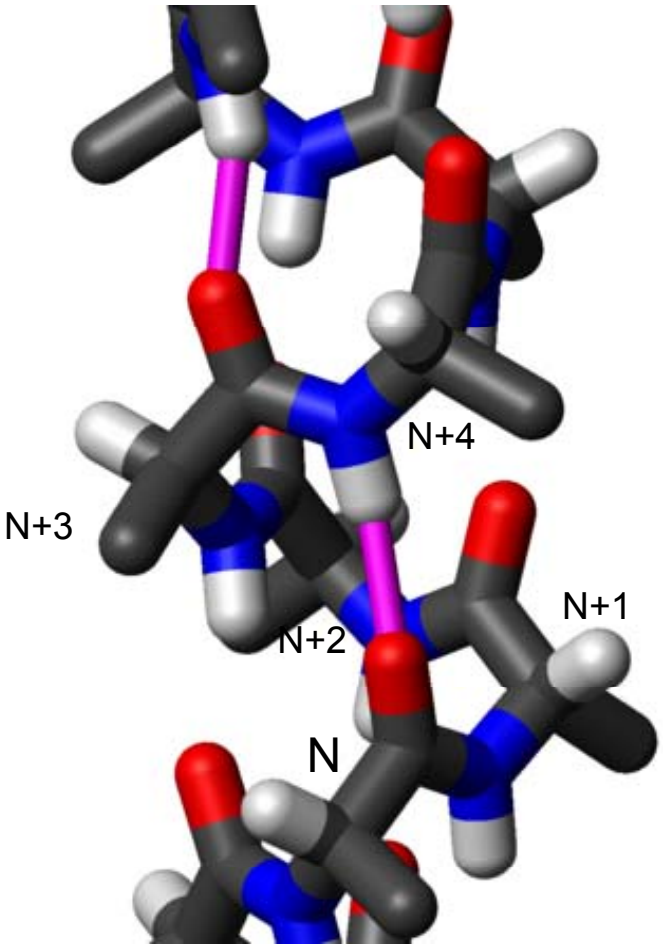
1DX0

1GU8: Bacteria Sensory Rhodopsin II



1GU8

Secondary Structure
1) alpha helix

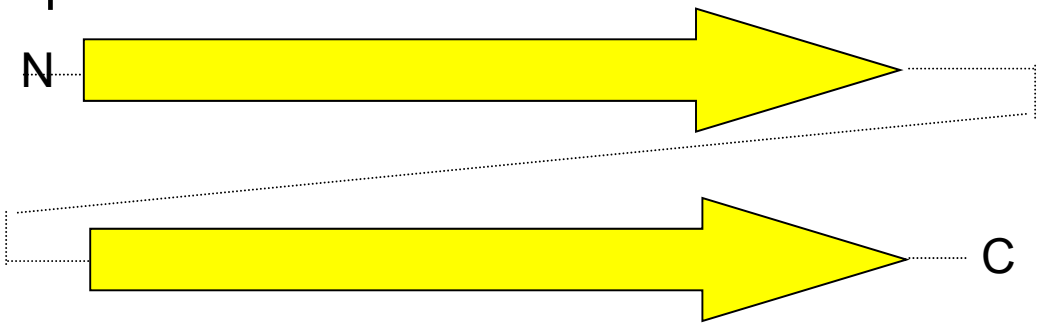


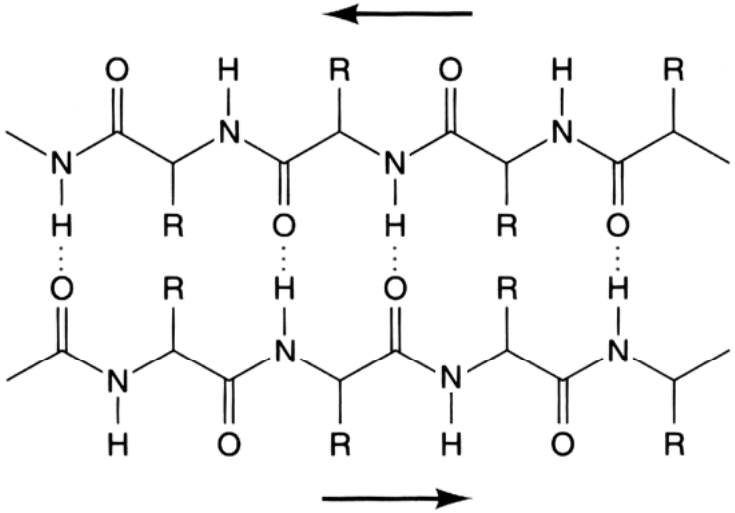
oxygen atom at N -th residue
is **hydrogen-bonded** with
nytrogen atom at N+4 -th residue

example) alpha helix regions (red)
in Cytochrome P450 2C5 (PDB: 1DT6)

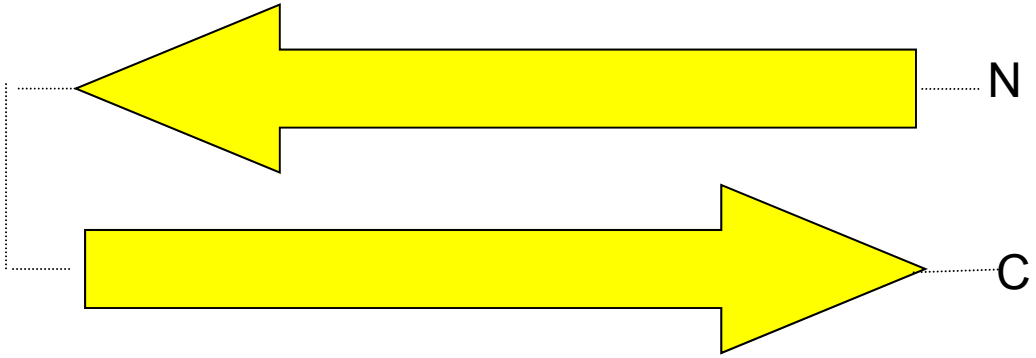
Secondary Structure
2) beta sheet

parallel beta sheet

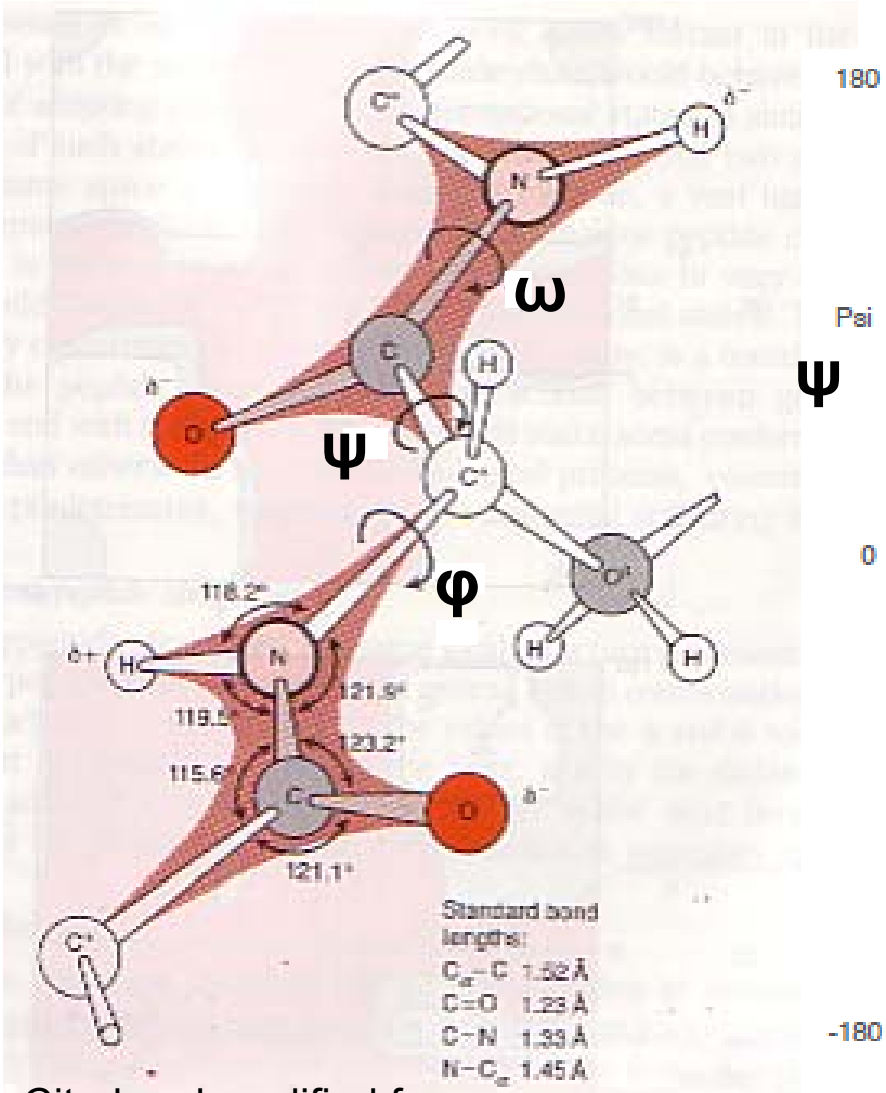




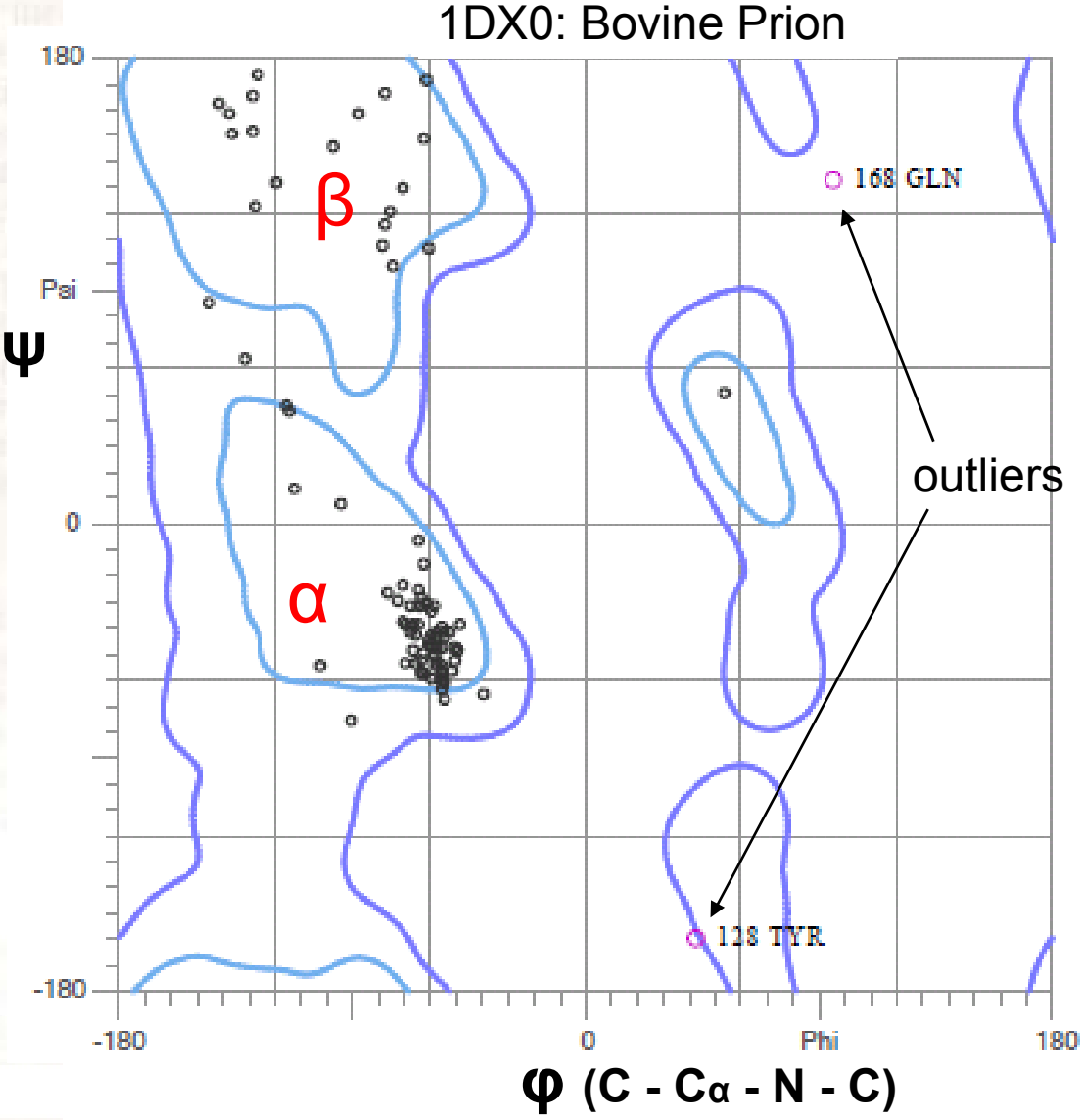
anti-parallel beta sheet



Ramachandran Plot



Cited and modified from
Darby and Creighton: Protein Structure,
Oxford University Press (1993)



Secondary Structures (advanced)

·α helix

3.6 residue / 1 round, spiral radius = 2.3 Å,
pitch (in Z-axis) = 1.5 Å / residue, pitch (in Z-axis) = 5.4 Å / round.
Dipole Moment (N-terminal charged as "+"、C-terminal as "－")
thus two helices prefer to align in "anti-parallel" manner.
Not stable in its dihedral angles, only stable with hydrogen bonding.
Proline residue is a strong helix breaker.

α helix n=3.6, r=2.3, d=1.5 (φ, ψ, ω) = (-57°, -47°, +180°)
3_{10} helix n=3, r=1.9, d=2.0
π helix n=4.3, r=2.8, d=1.1

·β sheet

almost doubly extended structure of α helix
side chains are placed orthogonal to the sheet, toward two sides in turn.
parallel β sheet (φ, ψ, ω) = (-119°, +113°, +180°)
anti-parallel βsheet (φ, ψ, ω) = (-139°, +135°, -178°)

·Turn

β turn（sharp turn with 4 residues）、 γ turn（sharp turn with 3residues）

DSSP code

- **H = alpha helix**
- B = residue in isolated beta-bridge
- **E = extended strand, participates in beta ladder**
- G = 3-helix (3/10 helix)
- I = 5 helix (pi helix)
- T = hydrogen bonded turn
- S = bend

Secondary Structure of a protein can be represented as a ***character string*** with DSSP code

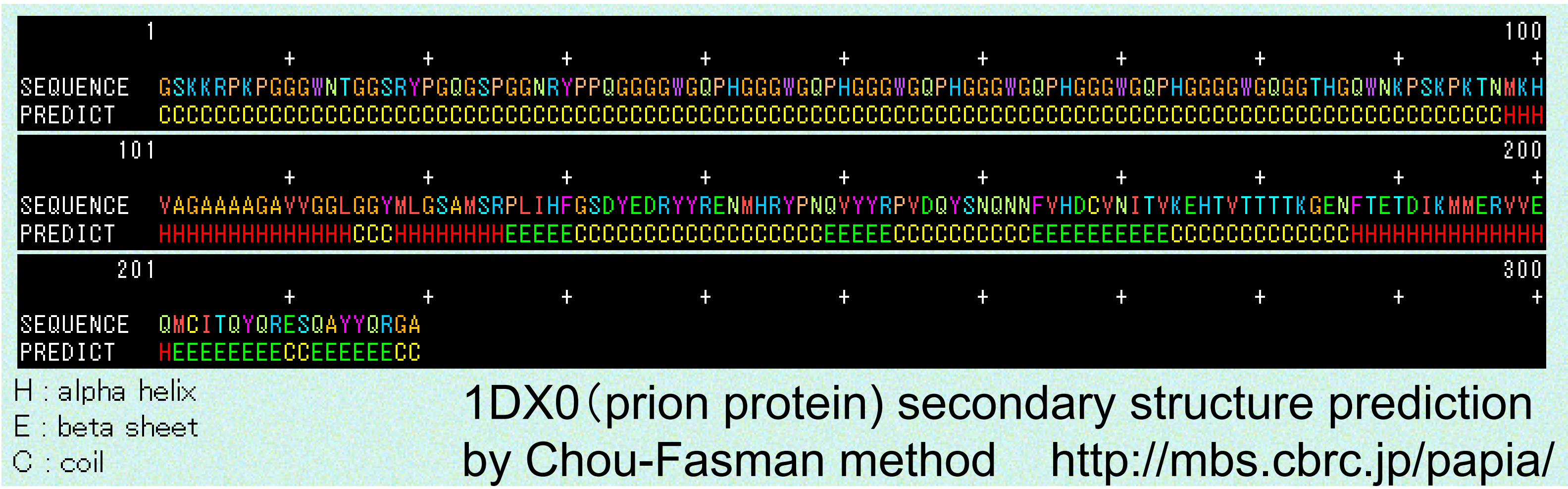
"….HHHHH….EEEEEEE….HHHH…."

Secondary Structure Prediction

- Chou-Fasman method
- GOR method
- Qian-Sejnowski method (neural net)
- New Joint method (consensus method)

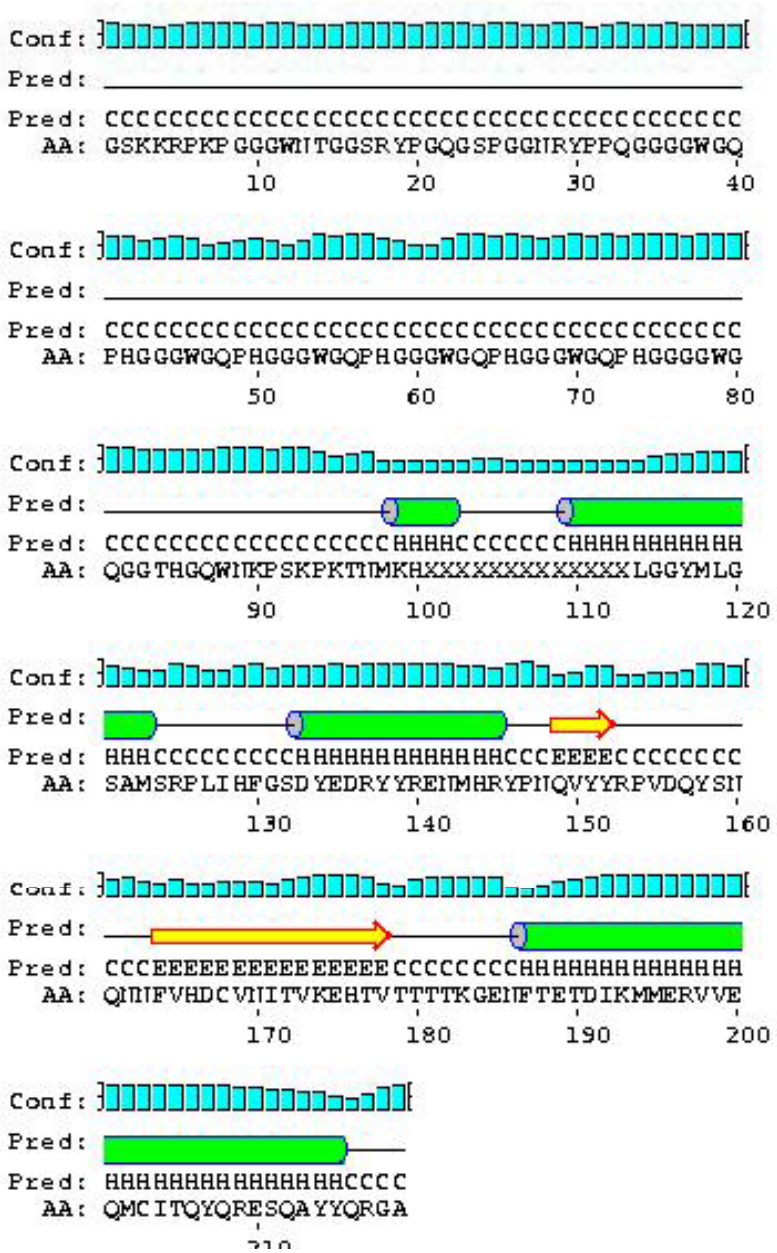
Each amino acid has “Secondary structure propencity” (as amino acid index).

In Chou-Fasman method, core regions (a block of “H” or “E”) are extracted from the sequence at first, and then expand “H”or “E” regions toward neighbors.



1DX0 (prion protein) secondary structure prediction by Chou-Fasman method http://mbs.cbrc.jp/papia/

Secondary Structure Prediction



Bioinformatics Group — UCL

Jones home>
McGuffin home>
Bryson home>

The PSIPRED Protein Structure Prediction Server

David T. Jones, Liam J. McGuffin & Kevin Bryson

Description

PSIPRED Server Help Page

PSIPRED Server History

PSIPRED FAIR USAGE POLICY

The PSIPRED protein structure prediction server allows you to submit a protein sequence, perform a prediction of your choice and receive the results of the prediction via e-mail. You may select one of three prediction methods to apply to your sequence: PSIPRED - a highly accurate method for protein secondary structure prediction, MEMSAT - our widely used transmembrane topology prediction method and GenTHREADER - a sequence profile based fold recognition method. More...

CLICK HERE TO ACCESS THE SERVER

For queries regarding PSIPRED: psipred@cs.ucl.ac.uk

1DX0（prion protein) Secondary Structure Prediction on UC London **PSIPRED** server.