

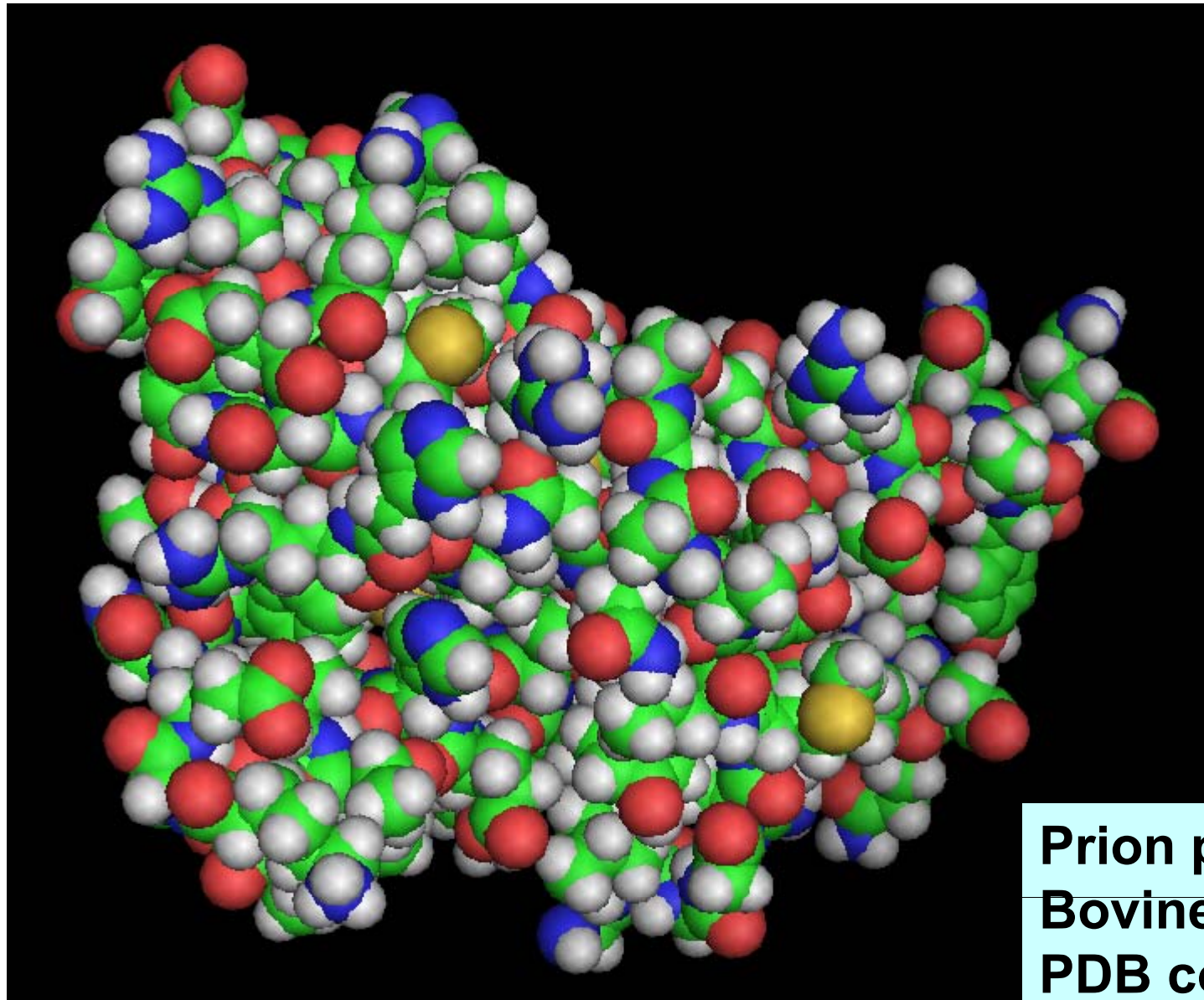
#12

Protein Structures

Topics:

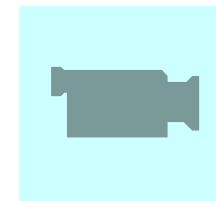
- Protein
 - Translation, Codon Table
 - Main Chain, Side Chain
 - Hydropathy index, hydrophilicity, hydrophobicity
- PDB Database
- Secondary Structure
 - alpha helix, parallel beta sheet, anti-parallel beta sheet
 - Ramachandran plot
 - DSSP code
 - Secondary Structure Prediction (Chou-Fasman method,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

Protein Structure Example

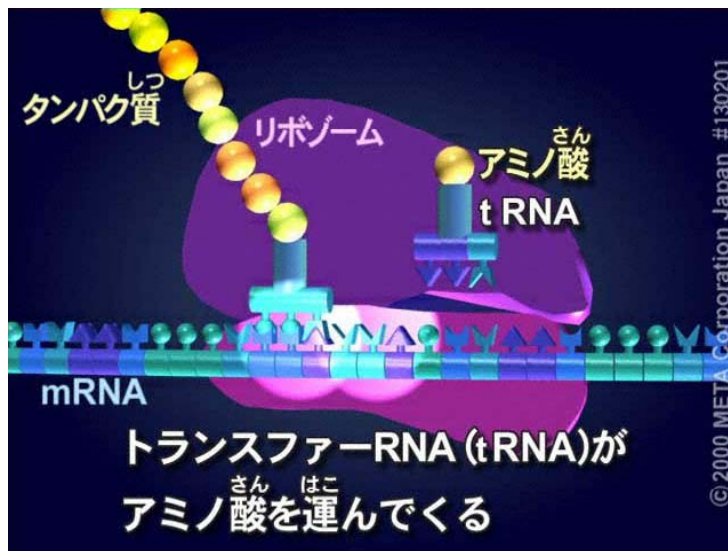


Prion protein
Bovine (cow)
PDB code = 1DX0

Translation of Protein



movie



Protein



Protein = poly peptide chain(s)

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

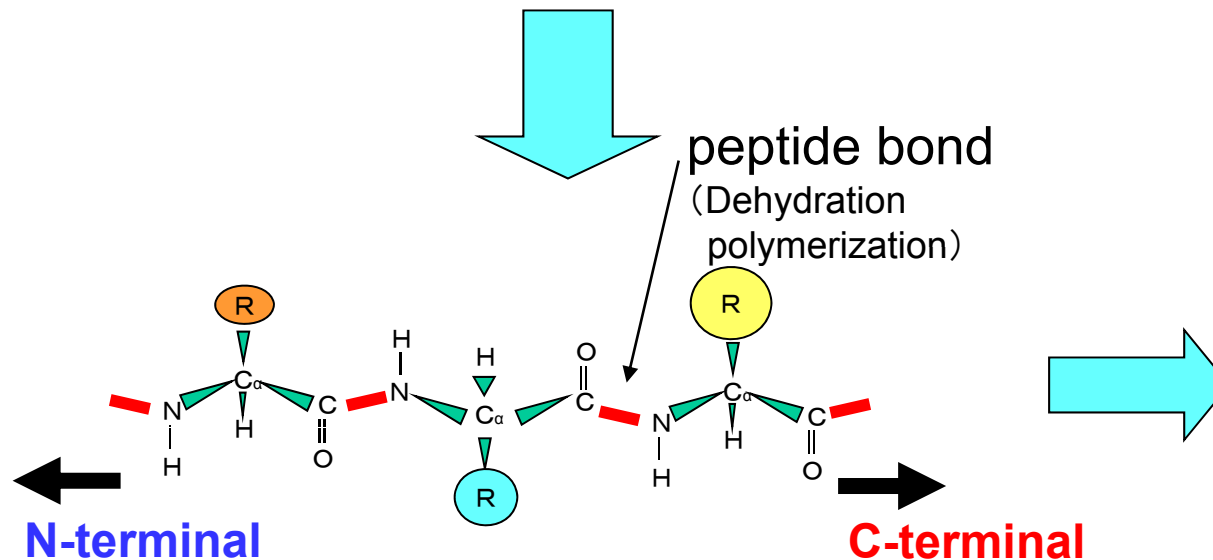
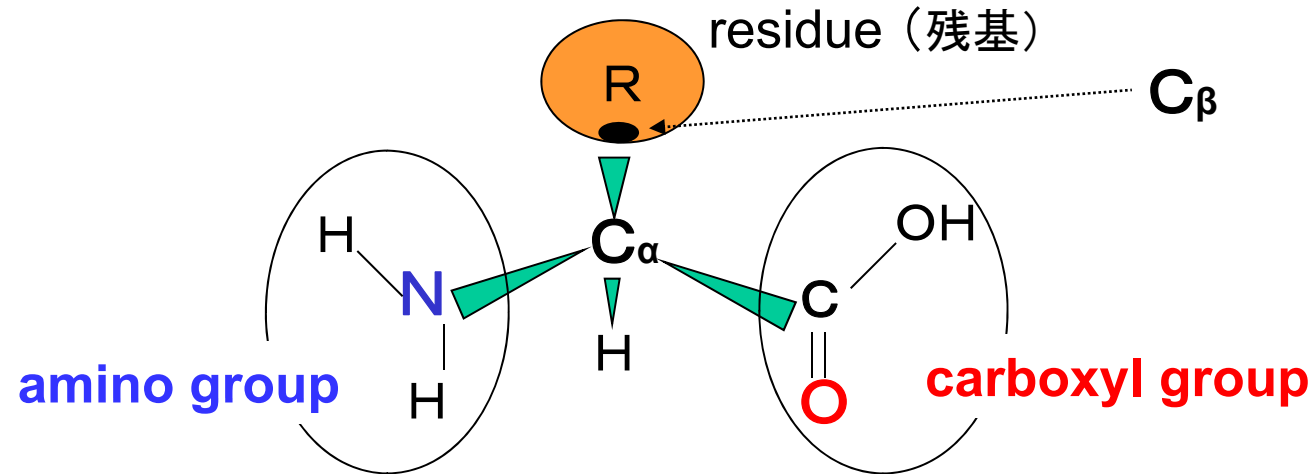
amino acid

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

(usually) 20 kinds

Protein Structure

• Amino acid as a building block



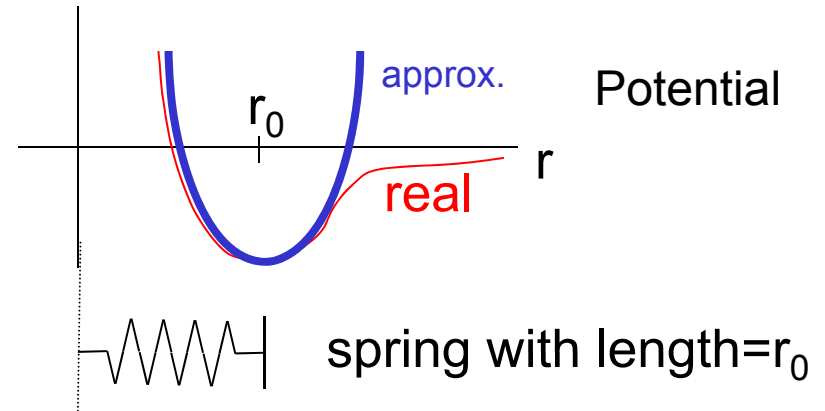
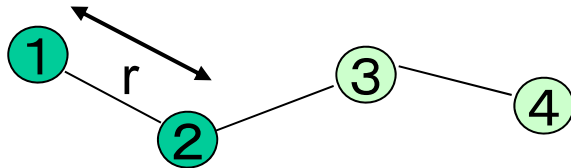
- S-S (disulfide bridge)
- Post-translational modification
acetylation,
phosphorylation,
sugar chain addition
- Dimerization, etc.

Main Chain

– bond length (結合長) :

approximation by Hooke's law

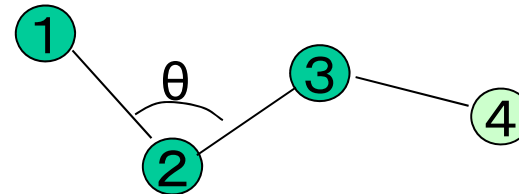
$$E_{\text{bond}} = \frac{1}{2} * k_b * (r - r_0)^2$$



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

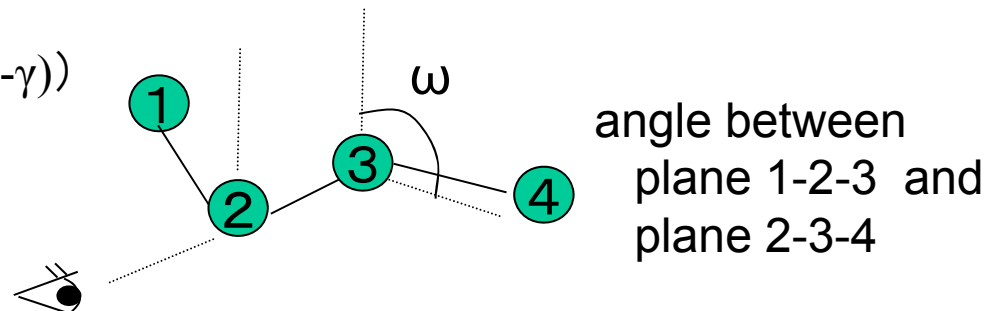
– bond angle (結合角) :

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



– dihedral angle (二面角) :

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

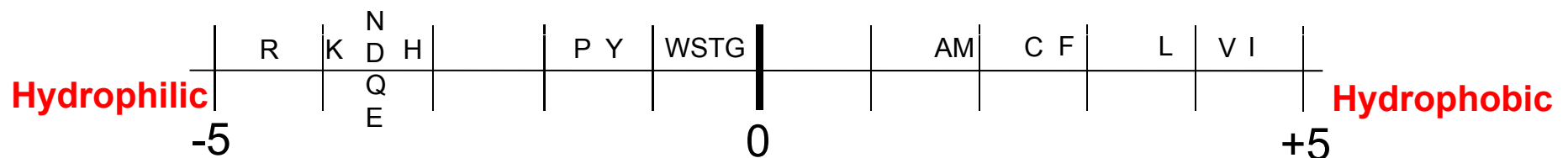


Side Chain

- 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)

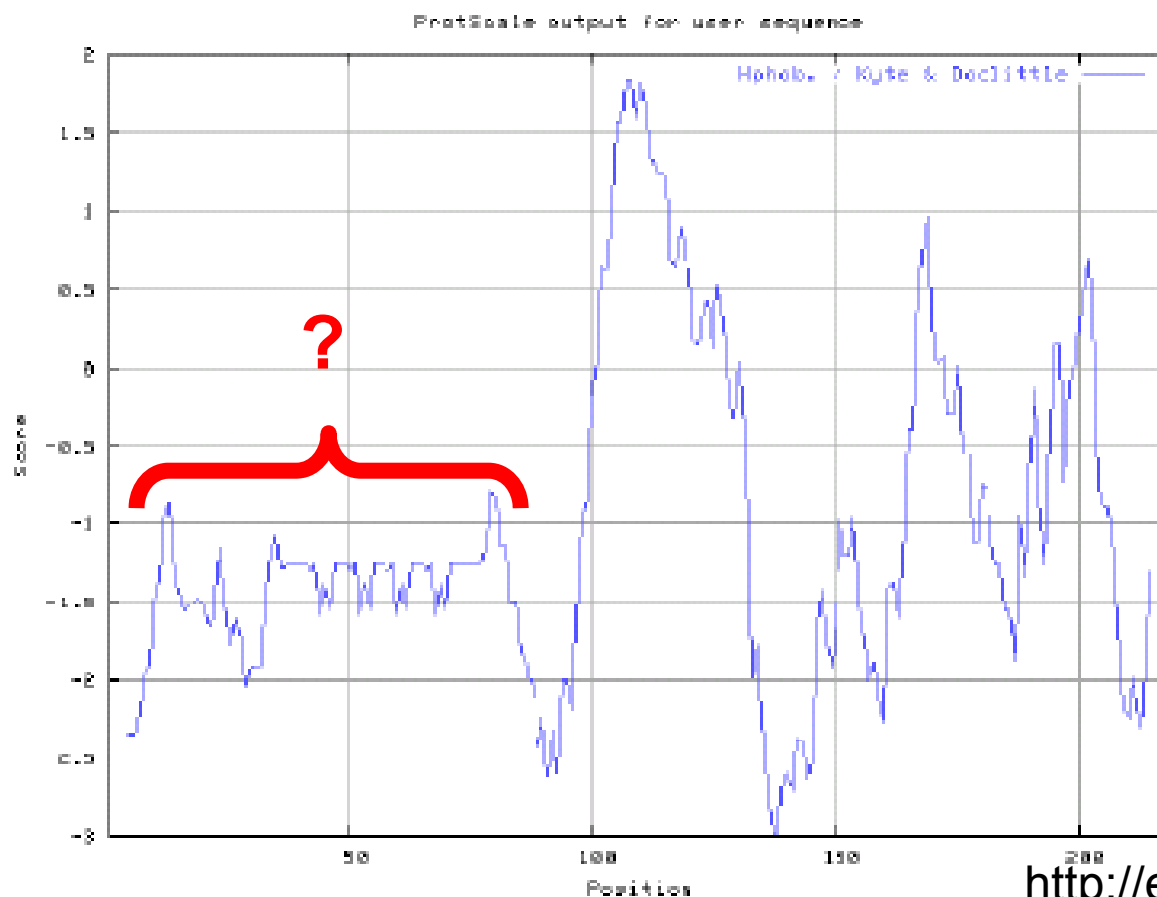
- hydropathy index** Kyte and Doolittle: *JMB*, **157**, 105-132 (1982)



Hydropathy Plot

>1DX0

GSKKRPKPGGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGGWGQPHGGGGWGQPHGG
 GWGQPHGGGGWGQPHGGGGWGQGGTHGQWNKPSKPKTNMKHVAGAAAAGAVVGGLGGYMLG
 SAMSRPLIHFGSDYEDRYRENMHRYPNQVYYRPVDQYSNQNNFVHDCVNITVKEHTVTT
 TTKGENFTETDIKMMERVVEQMCITQYQRESQAYYQRGA



**Prion protein
Bovine (cow)**

PDB code = 1DX0

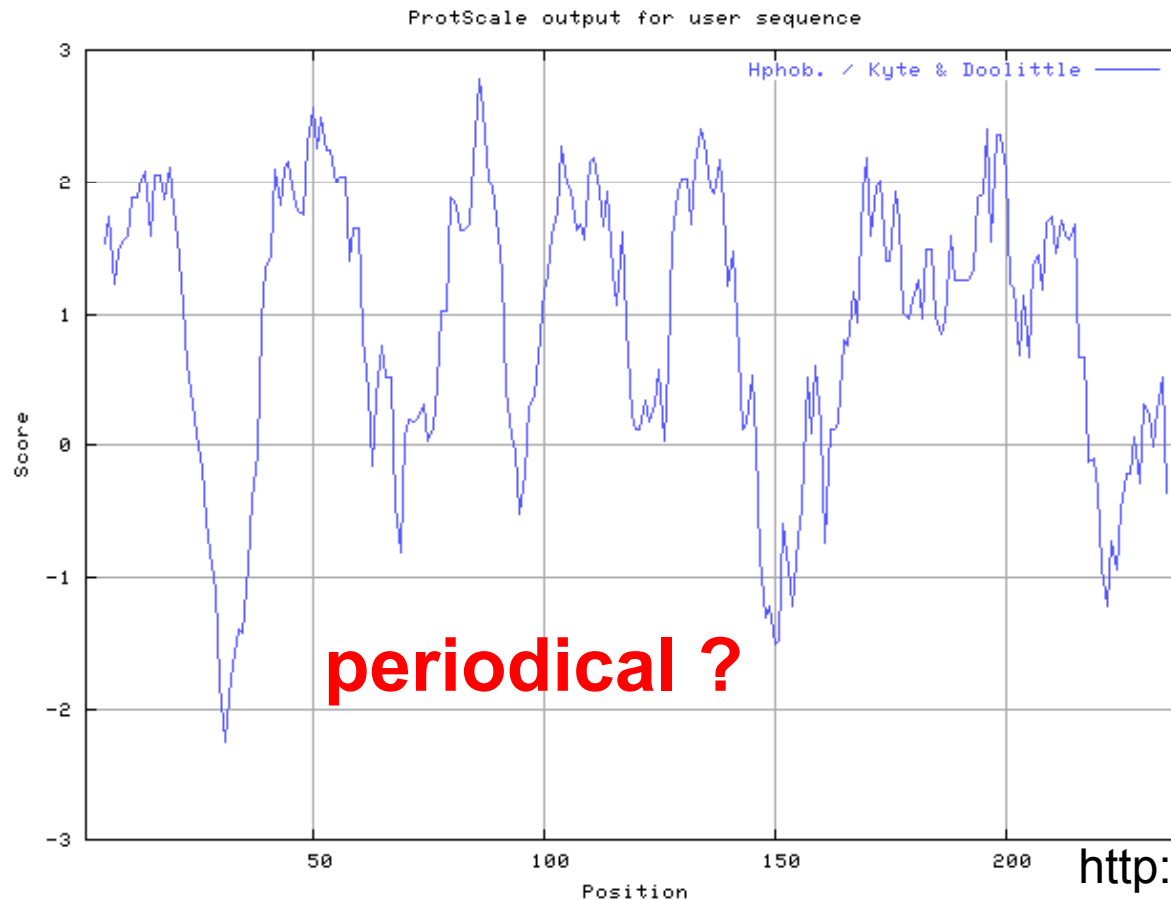
ProtScale

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

Hydropathy Plot

>1GU8

```
MVGLTTLFWLGAIGMLVGTLFAWAGRDAGSGERRYVTLVGISGIAAVAYVVMALGVGW
VPVAERTVFAPRYIDWILTTPLIVYFLGLLAGLDSREFGIVITLNTVVMLAGFAGAMVPGIERY
ALFGMGAVAFLGLVYYLVGPMTESASQRSSGIKSLYVRLRNLTIVLWAIYPFIWLLGPPGVAL
LTPTVDVALIVYLDLVTKVGFGFIALDAAATLRAEHGESLAGVDTDAPAVAD
```



**Sensory Rhodopsin II
(bacteria)**

PDB code = 1GU8

ProtScale

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

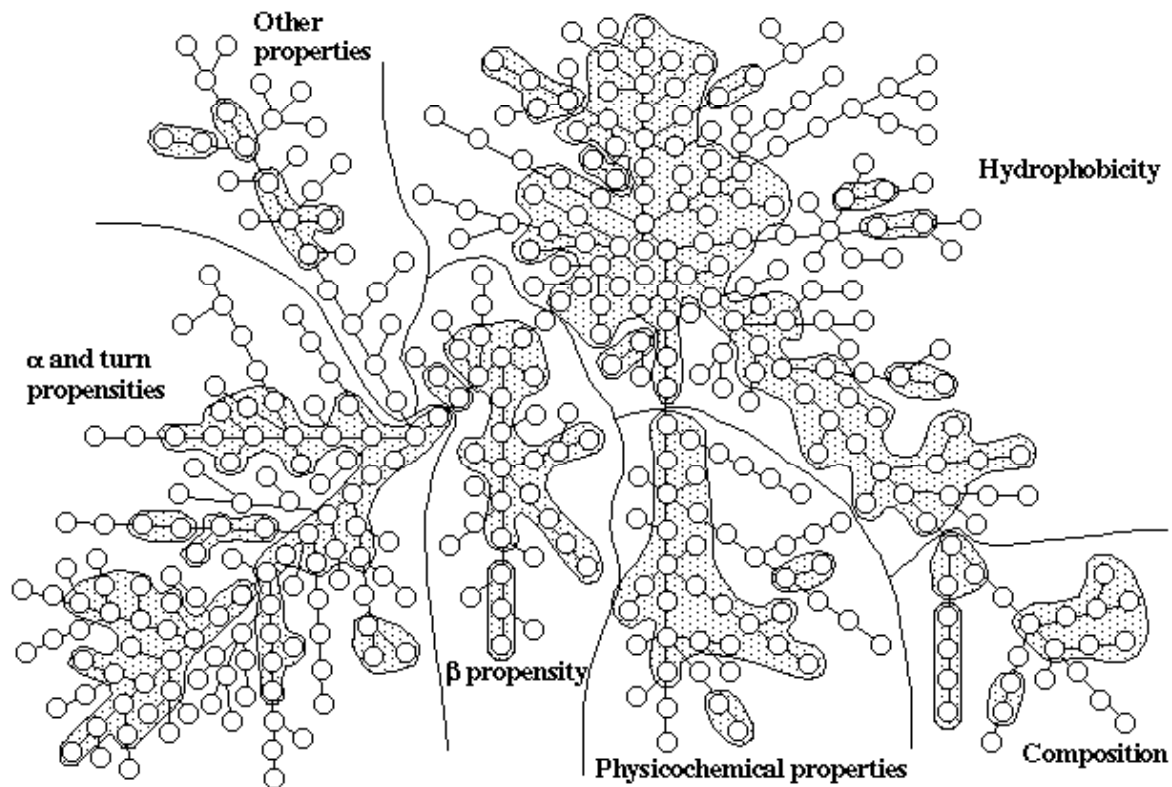
Other amino acid indices

AAindex database ver 9.1 (Aug. 2006)

Collection of 544 amino acid indices.

<http://www.genome.ad.jp/aaindex>

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



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

Getting PubMed abstract

The screenshot shows the PubMed website header with the NCBI logo and the text "A service of the National Library of Medicine and the National Institutes of Health". Below the header is a navigation bar with tabs for "All Databases", "PubMed", "Nucleotide", "Protein", "Genome", "Structure", "OMIM", "PMC", "Journals", and "Books". The search bar contains the text "PubMed" and "for 9053899[uid]". There are buttons for "Go", "Clear", and "Save Search". Below the search bar are buttons for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The "Display" section shows "AbstractPlus" selected, with "Show 20" and "Sort by" options. At the bottom of the search bar, it says "All: 1" and "Review: 0".

☐ 1: [Protein Eng.](#) 1996 Jan;9(1):27-36.

FREE FINAL TEXT
OXFORD JOURNALS

Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins.

[Tomii K](#), [Kanehisa M](#).

Institute for Chemical Research, Kyoto University, Japan.

An amino acid index is a set of 20 numerical values representing any of the different physicochemical and biochemical properties of amino acids. As a follow-up to the previous study, we have increased the size of the database, which currently contains 402 published indices, and re-performed the single-linkage cluster analysis. The results basically confirmed the previous findings. Another important feature of amino acids that can be represented numerically is the similarity between them. Thus, a similarity matrix, also called a mutation matrix, is a set of 20 x 20 numerical values used for sequence alignments and similarity searches. We have collected 42 published mutation matrices, performed hierarchical cluster analyses and identified several clusters. We then analyzed the nature of the data set and the method used for constructing the matrices. Further, we have tried to reproduce each mutation matrix by the combination of amino acid indices in order to understand which properties of amino acids are most important. There was a relationship between the PAM units of Dayhoff's mutation matrix and the amino acid indices.

Related Links

- ▶ Cluster analysis of amino acid indices for prediction of protein structure and function. [Protein Eng. 1996 Jan;9(1):27-36.]
- ▶ AAindex: Amino Acid Index Database. [Nucleic Acids Res. 1996 Jan;24(1):1-3.]
- ▶ AAindex: amino acid index database. [Nucleic Acids Res. 2000 Jan;28(1):1-3.]
- ▶ A method to estimate effects of amino acid substitution in blood coagulation factor IX from hemophilia B. [Medinfo. 1996;1:1-3.]

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>

- # SPring-8



-

- 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

RCSB PDB
PROTEIN DATA BANK

CONTACT US | HELP | PRINT PAGE

Home Search Structure Results
Queries

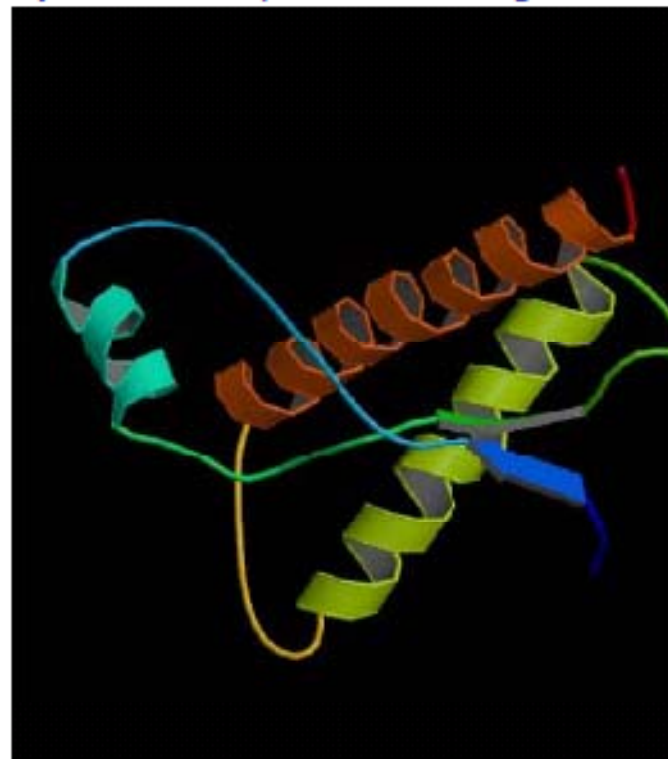
A MEMBER OF THE
An Information Portal to Biological Macromolecular
As of Tuesday May 29, 2007 there are 43755 Structures | PDB

☒ PDB ID or keyword ☐ Author

ALERT: Our data files are changing soon. Please visit <http://www.wwpdb.org> for more details.

Still Images for 1DX0

Asymmetric Unit / Assumed Biological Molecule



Render in VBT ProteinWorkshop

**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
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;

```

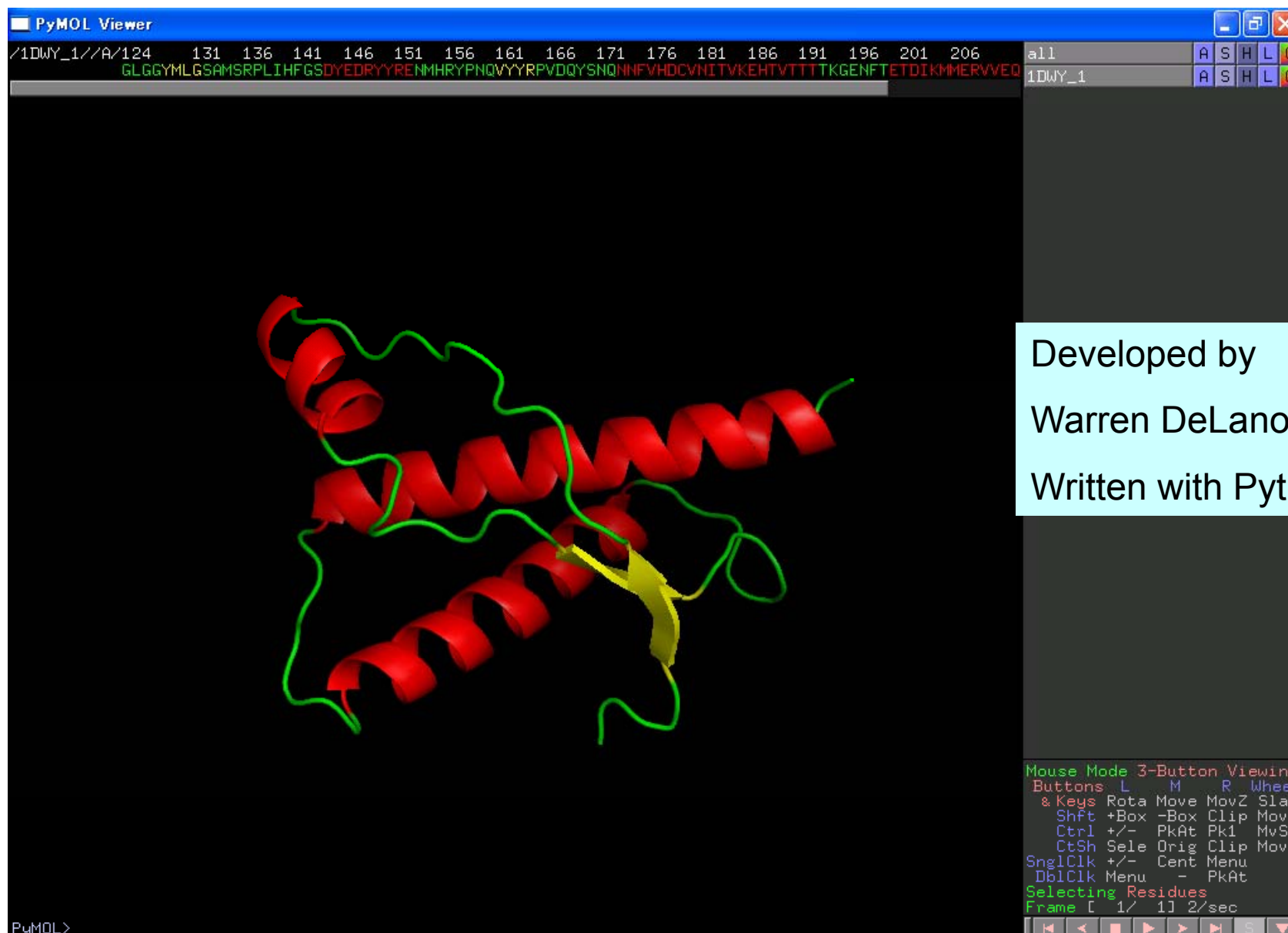
15-DEC-99 1DX0

first residue
(GLY 124)

X Y Z

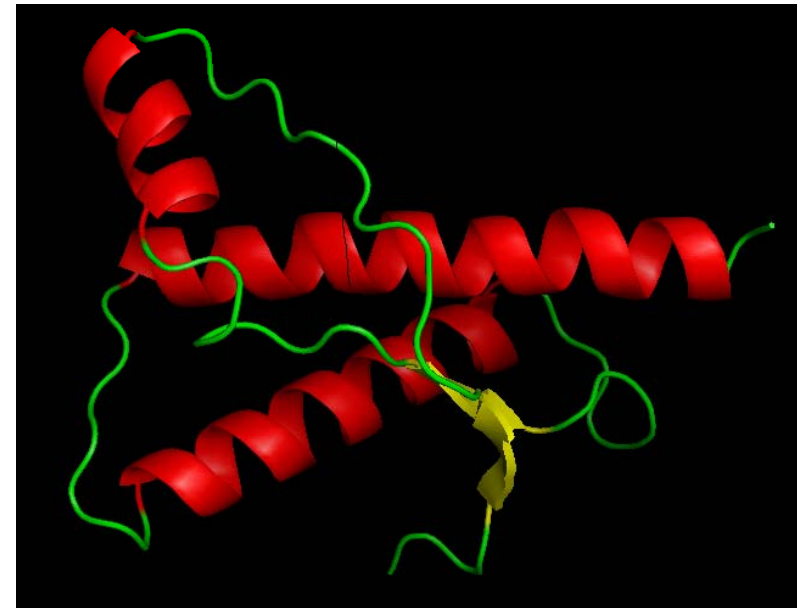
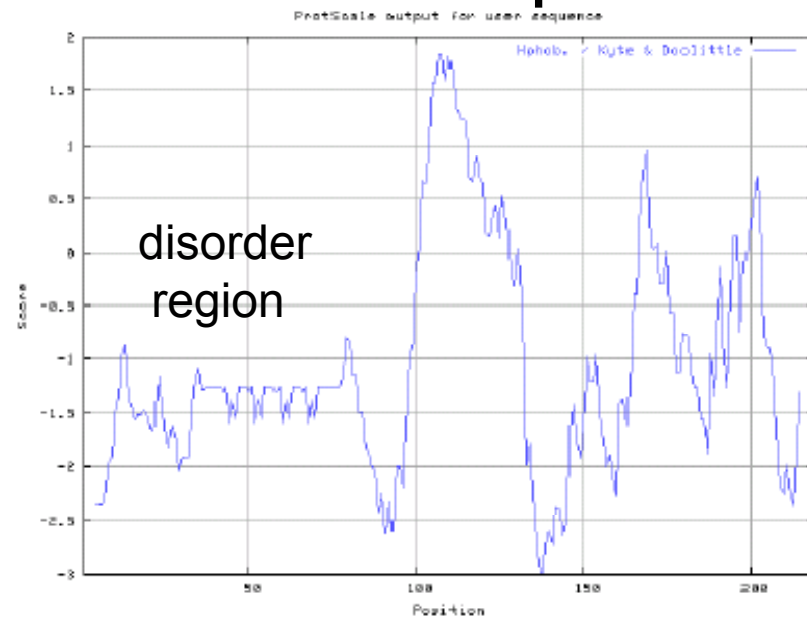
ATOM	1	N	GLY	A	124	11.500	-12.139	-5.556	1.00	0.00	N
ATOM	2	CA	GLY	A	124	12.523	-11.134	-5.837	1.00	0.00	C
ATOM	3	C	GLY	A	124	12.362	-9.873	-4.981	1.00	0.00	C
ATOM	4	O	GLY	A	124	12.934	-8.828	-5.305	1.00	0.00	O
ATOM	5	H	GLY	A	124	11.666	-12.848	-4.859	1.00	0.00	H
ATOM	6	1HA	GLY	A	124	12.474	-10.852	-6.890	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

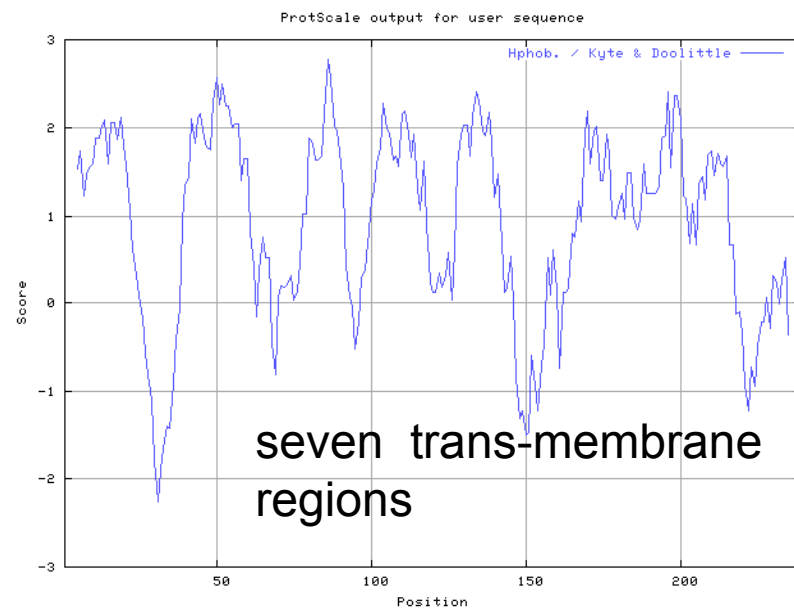
1DX0: Bovine Prion protein



1DX0



1GU8: Bacteria Sensory Rhodopsin II

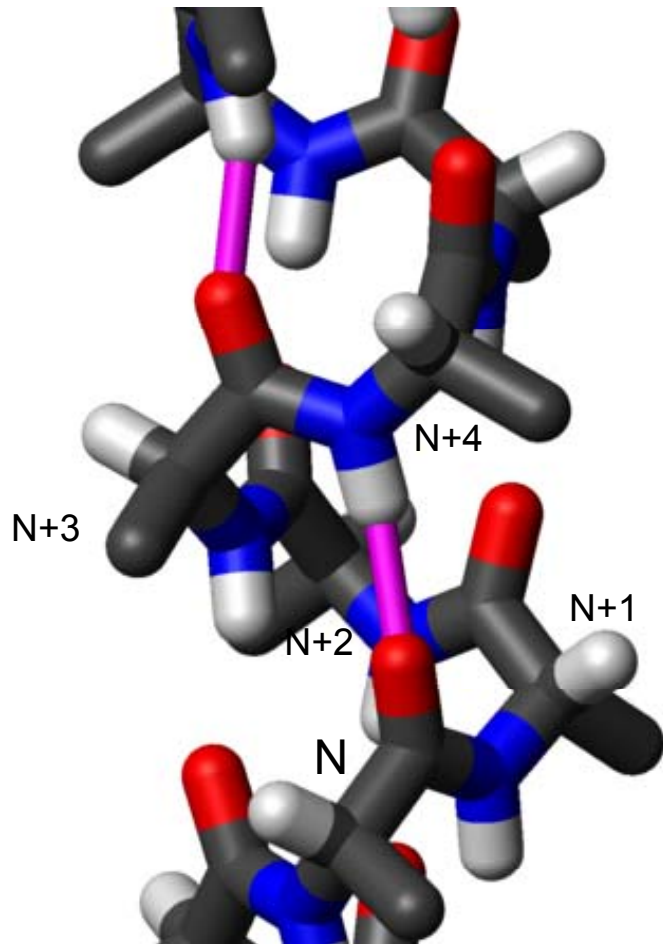


1GU8



Secondary Structure

1) alpha helix



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

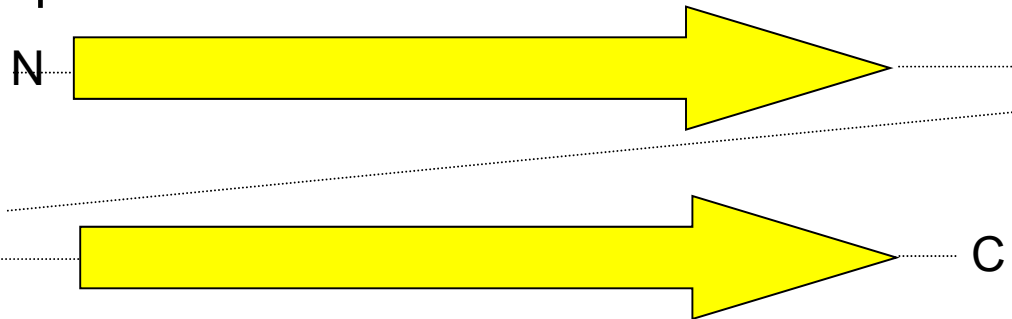


example) alpha helix regions (red)
in Cytochrome P450 2C5 (PDB: 1DT6)

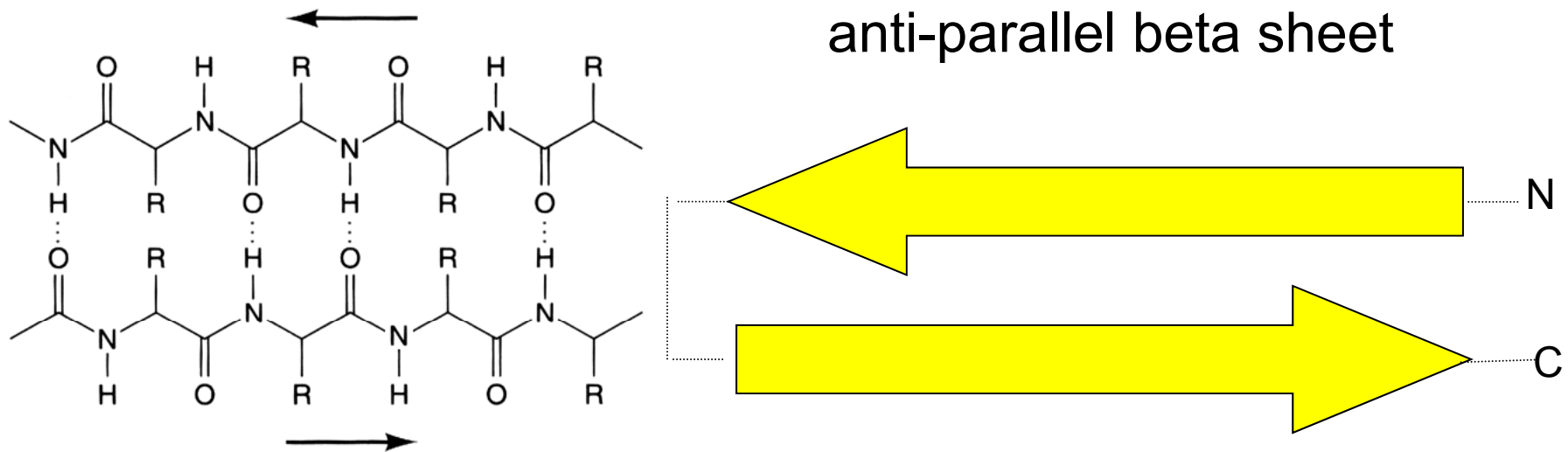
Secondary Structure

2) beta sheet

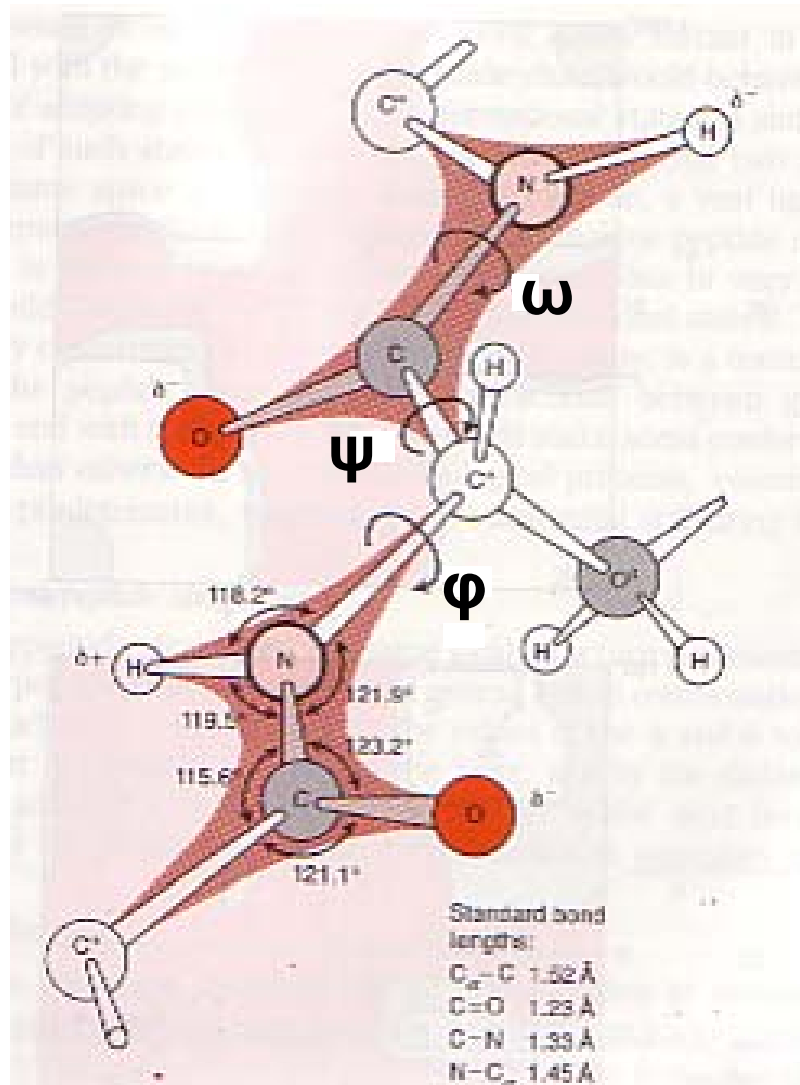
parallel beta sheet



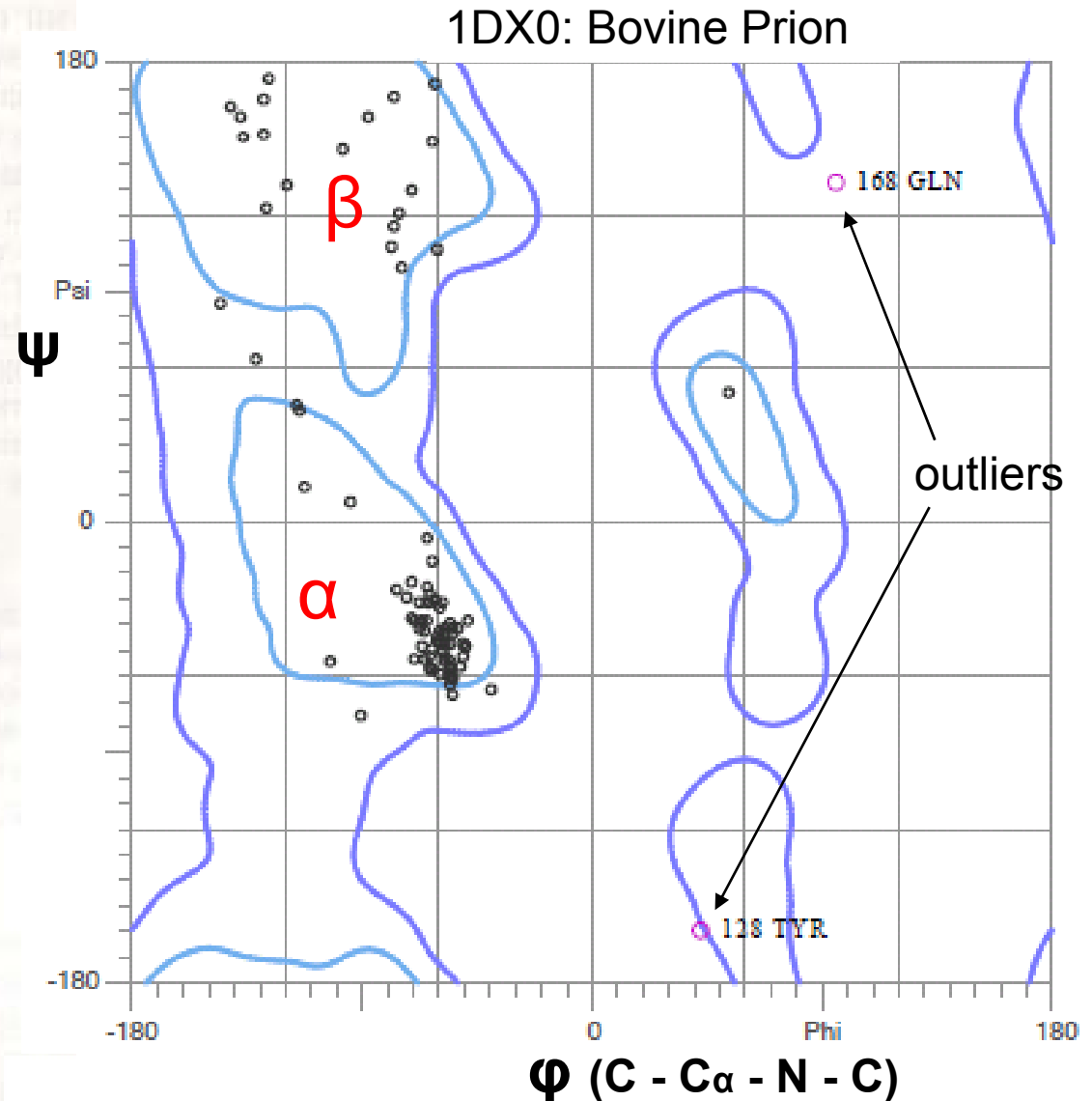
anti-parallel beta sheet



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 $n=3.6, r=2.3, d=1.5$ $(\varphi, \psi, \omega) = (-57^\circ, -47^\circ, +180^\circ)$

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 $(\varphi, \psi, \omega) = (-119^\circ, +113^\circ, +180^\circ)$

anti-parallel β sheet $(\varphi, \psi, \omega) = (-139^\circ, +135^\circ, -178^\circ)$

• Turn

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

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

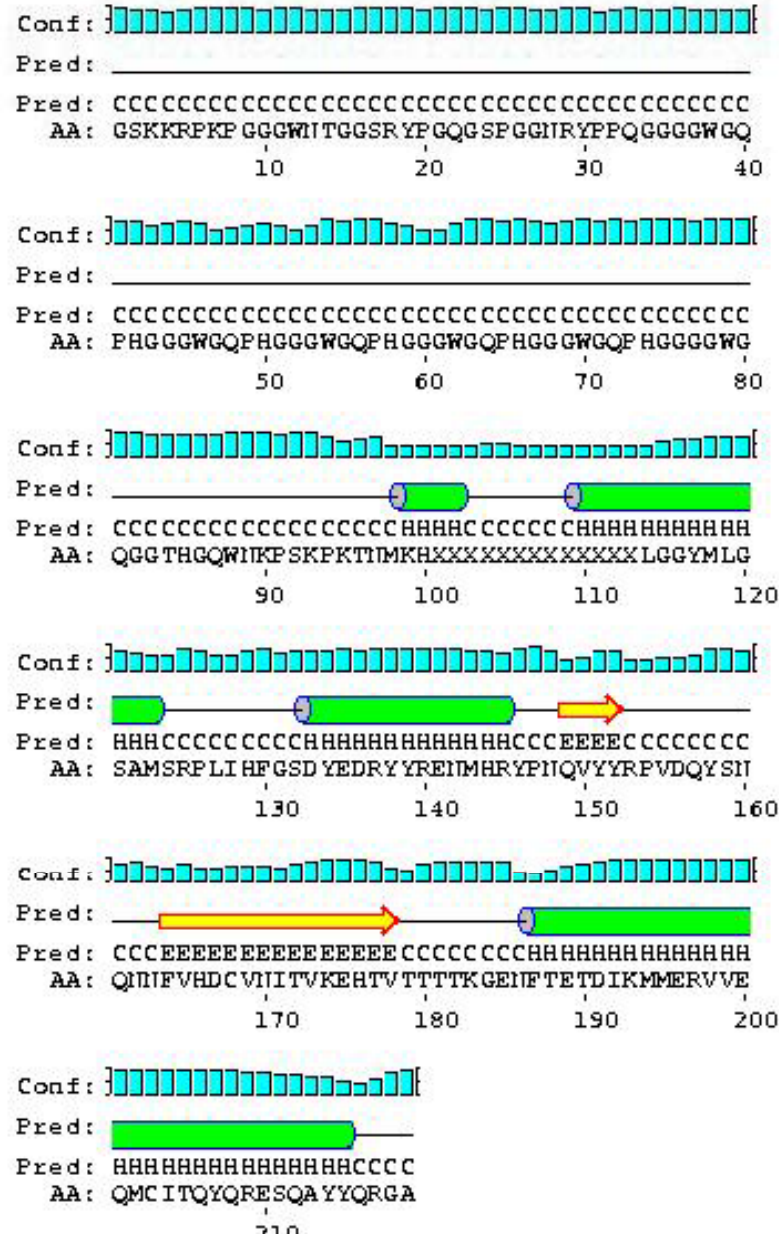
- Chou-Fasman method
- GOR method
- Qian-Sejnowski method (neural net)
- New Joint method (consensus method)

Each amino acid has “Secondary structure propensity” (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



[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.

