

Introduction to Structural Bioinformatics

Sheh-Yi Sheu

Department of Life Science

National Yang-Ming University

Outline

◆ Introduction

- Database — Protein Data Bank
- Visualization Tools
- Protein Structure
 - Primary Structure
 - Secondary Structure
 - Tertiary Structure
 - Quaternary Structure

- ◆ Sequence Analysis
- ◆ Look at structure and interpret function
- ◆ Sequence — structure relationships
- ◆ **Homology modelling**
- ◆ **Fold families / classification**
- ◆ **Threading and structure prediction**
- ◆ **Protein modelling and drug discovery**
- ◆ **Overview of available software packages**

◆ **Modeling Protein structure and Homology**

- **What is Homology modeling?**
- **Building-Model by Homology**
- **Applications**
- **ProMod: Swiss-Model server**

◆ **Computer-Aided Drug Design**

- **Structure-based drug design**
- **Procedures**
- **drug-receptor interaction and rational drug design**
- **Application**

◆ **Overview of available software packages**

--- **Insight II, CHARMM**

Introduction

The relationship between **protein structure** and its biological function

**Chromosome
(Cell and Nucleus)**



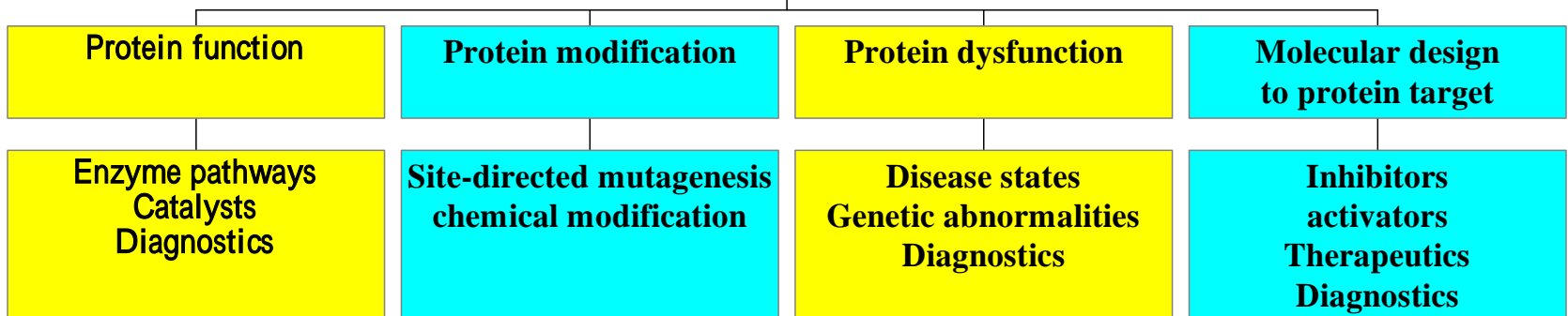
Gene sequence



Protein sequence



Protein 3D structure



Amino acid Gene sequence

Computation

Structure prediction

database searching/generation
homology building
ab initio calculations

energy minimization
molecular dynamics
van der Waals surfaces
Electrostatic potential

Computer graphics
3D visualization

Experiment

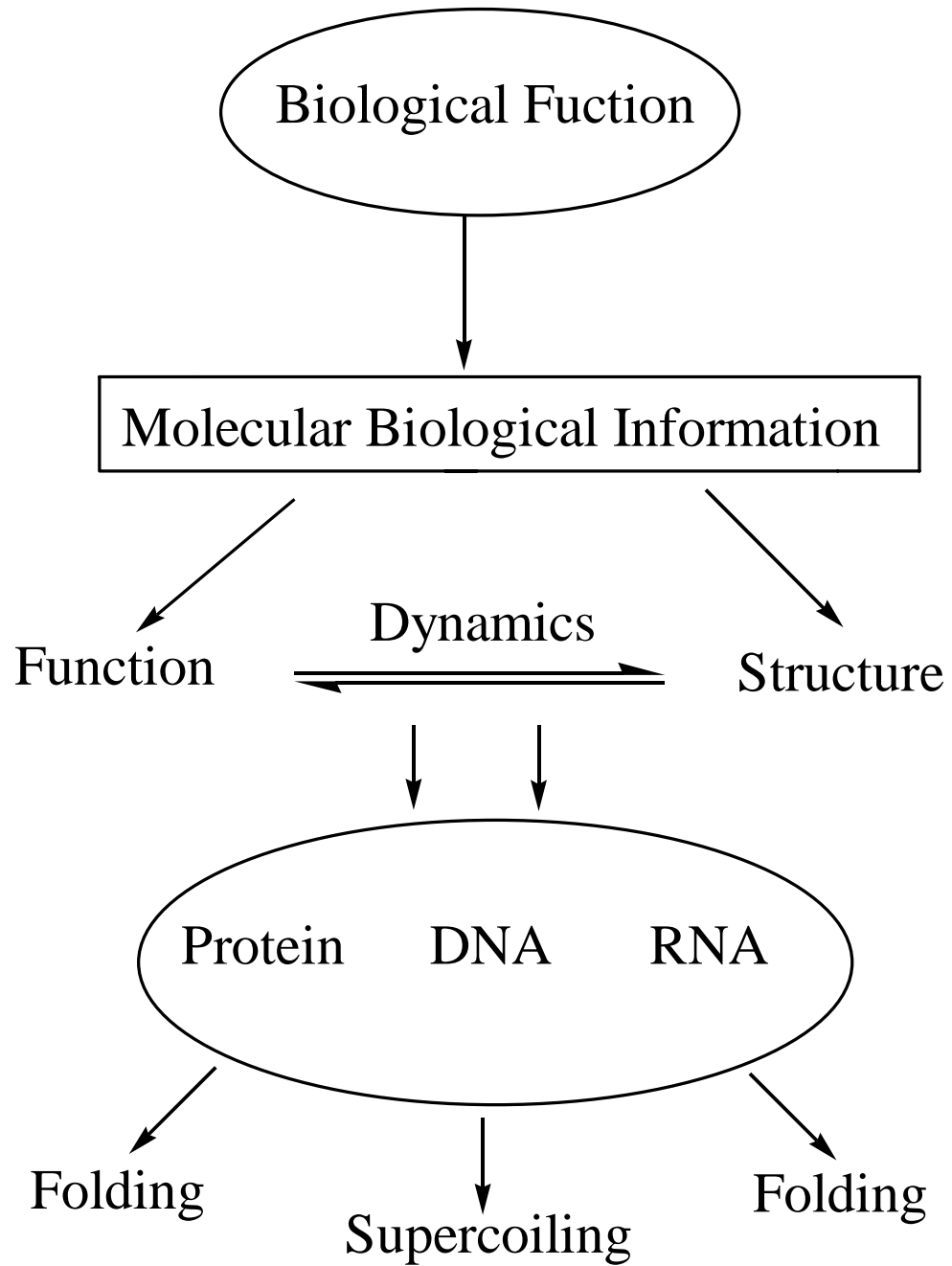
Spectroscopy
NMR, UV, IR,
Fluorescence,
Diffraction, X-ray, neutron

Object

- **enzyme -inhibitor docking**
- **active analog/lead compound**
- **structure modification/stabilization**
- **site-directed mutagenesis**
- **quantitative structure-activity relationship**

APPLICATIONS

- **Enzyme inhibitor/regulator design**
- **Enzymatic pathway regulation**
- **Reduction of toxic/side effect**
- **Enhancement of chemical production**
- **Structural stabilization**
- **Specificity/reactivity modulation of enzymes**



Sequence



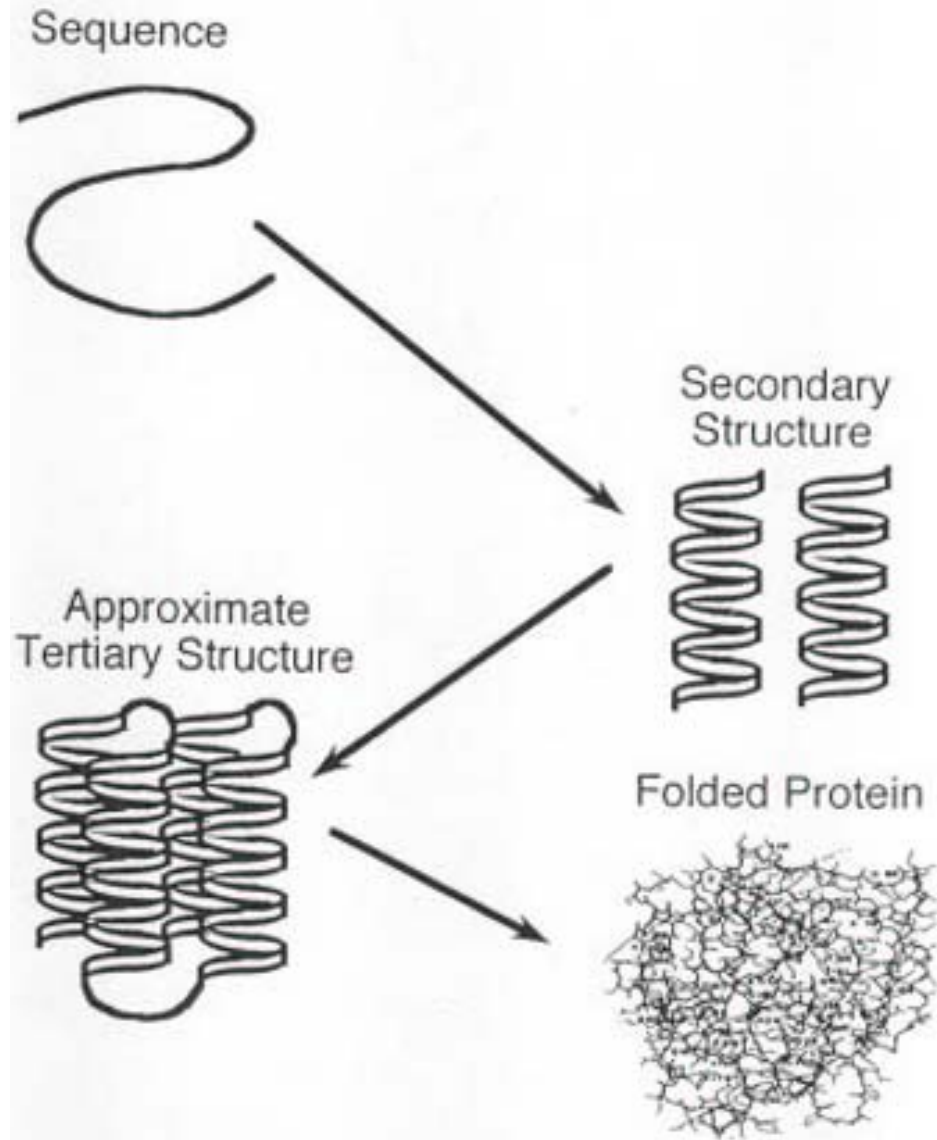
Structure



?



Function



A hierarchical condensation model for protein folding. Sequence determines secondary structure, and secondary structure elements assemble to form an approximate tertiary structure. Energy refinement yields a detailed three-dimensional structure.

Protein folding problem

Why does a certain protein sequence lead to specific protein fold?

- 巨分子結構形成的原理
- 分子與分子間的作用力
- 分子摺疊的理論基礎
- 分子突變後結構的影響
- 分子結構預測
- 分子功能與結構間之關係
- 蛋白質工程
- 分子動力學之模擬
- 酵素反應之模擬
- 酵素與受質之交互作用
- 理性藥物設計

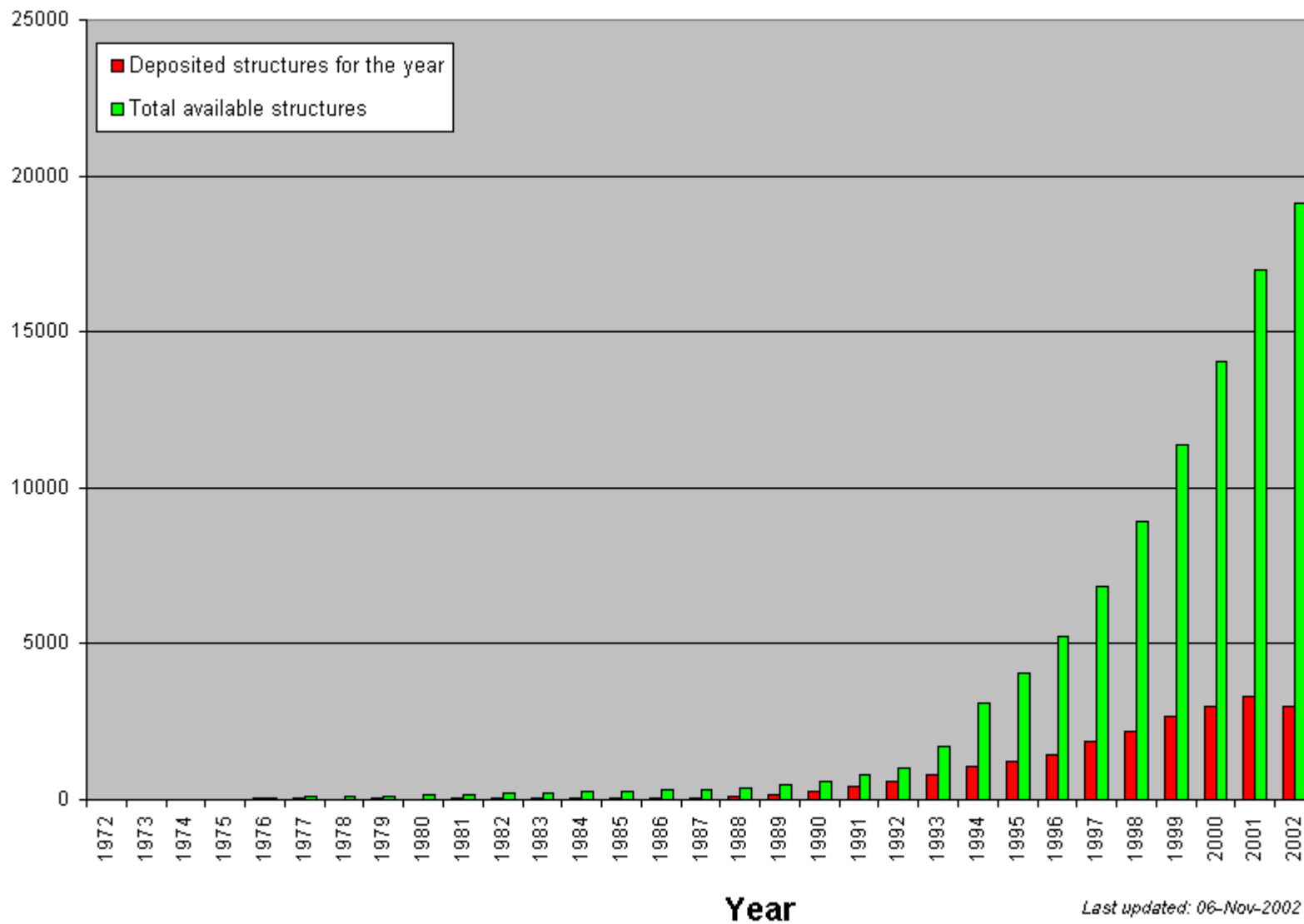
Protein Data Bank

PDB -----Protein Data Bank

the single worldwide repository for the processing and distribution of 3D biological macromolecular structure data.

<http://www.rcsb.org/pdb/>

PDB Content Growth



Visualization Tools and Molecular Modeling software systems

Visualization Tools

- Rasmol & RasTop
- Chime
- Weblab viewer
- Molscript
- Kinemage
- VMD
- InsightII
- Quanta
-

Molecular Modeling software systems

AMBER – Molecular Mechanics and Dynamics

CAMSEQ – Molecular Mechanics and Molecular Display

CHEMLAB – Molecular Mechanics, Quantum Mechanics,
Molecular Display

CHEM-X – Molecular Mechanics, Dynamics and Display

DISCOVER – Molecular Mechanics and Display

INSIGHT – Molecular Display

FRODO – Molecular Display (especially macromolecular
crystallography)

GRAMPS – General Graphical Display System

HYDRA – Molecular Mechanics, Molecular Dynamics,
Molecular Display

MACROMODEL – Molecular Mechanics and Molecular Display

MIDAS – Molecular Display

MMS – Molecular Display

SYBYL – Molecular Mechanics and Molecular Display

MENDYL – Macromolecular Mechanics and Molecular Display

 Introduction to Molecular Modeling

----- A Tutorial for RasMol

[http://www.usm.maine.edu/~rhodes/RasTut/
text/RasTut.html](http://www.usm.maine.edu/~rhodes/RasTut/text/RasTut.html)

Introduction of Protein Structure

Principles of Protein Structure, Comparative Protein Modelling and Visualisation

-----Nicolas Guex and Manuel C. Peitsch

Part I: Introduction to Protein structure

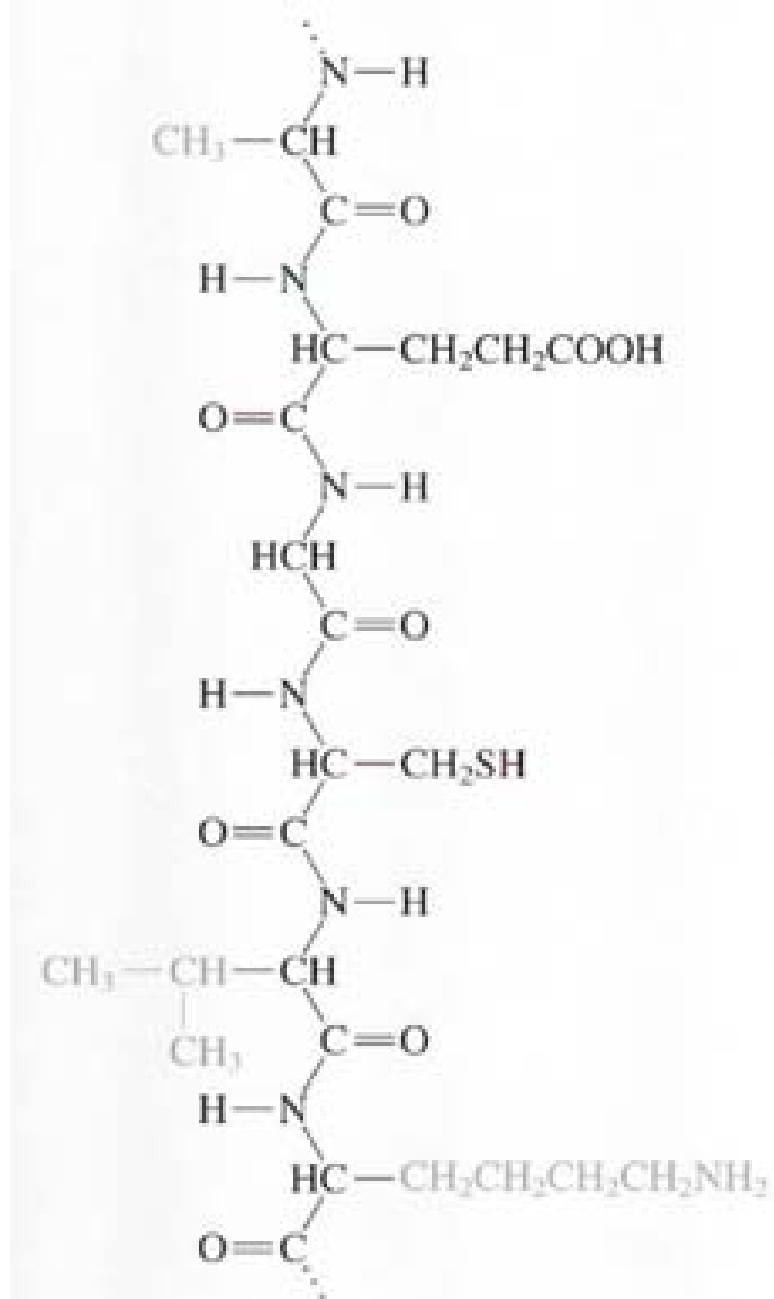
Part II: Protein modelling

Part III: Model quality

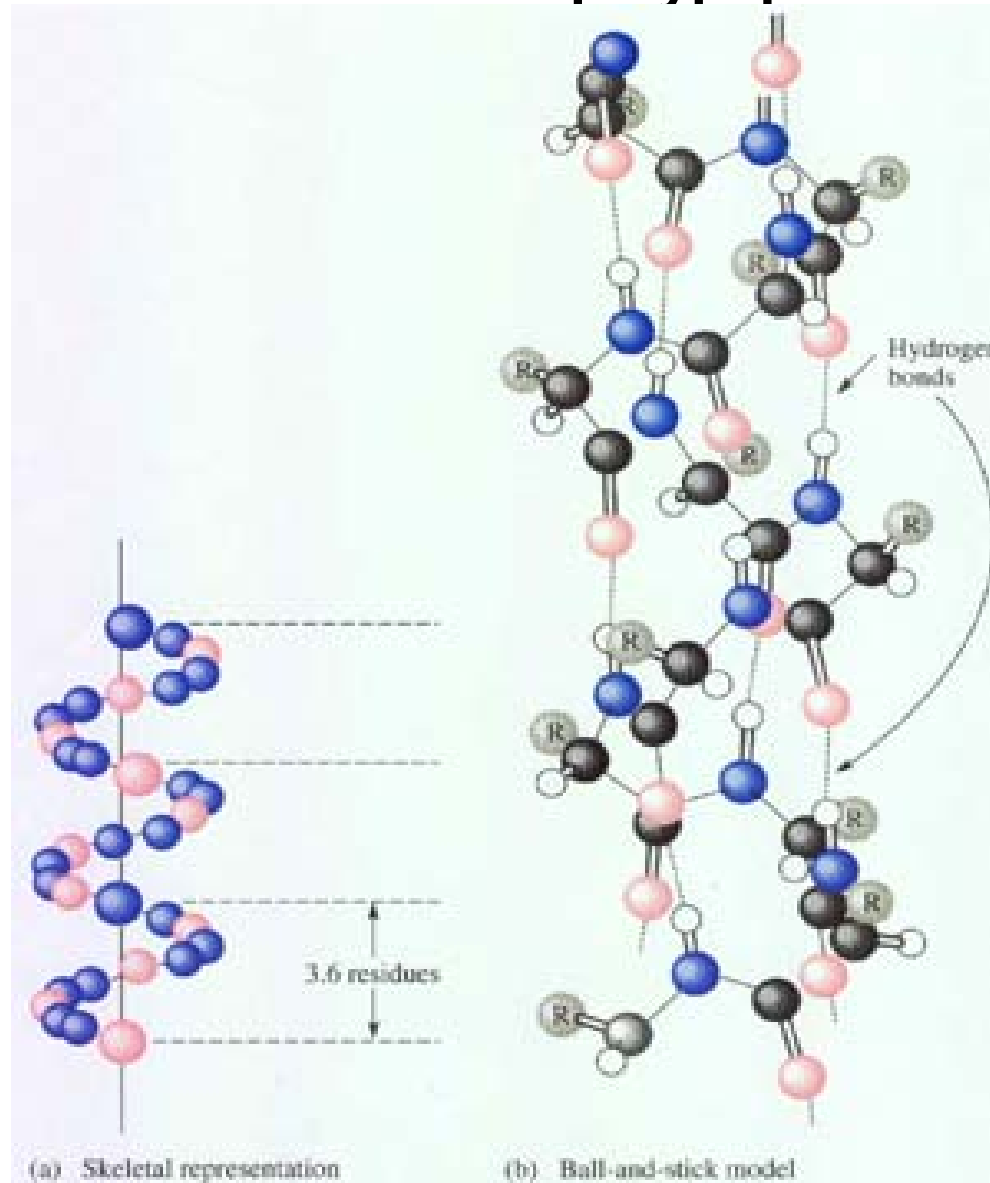
Part IV: The SWISS-MODEL modelling environment

<http://www.expasy.ch/swissmod/course/course-index.htm>

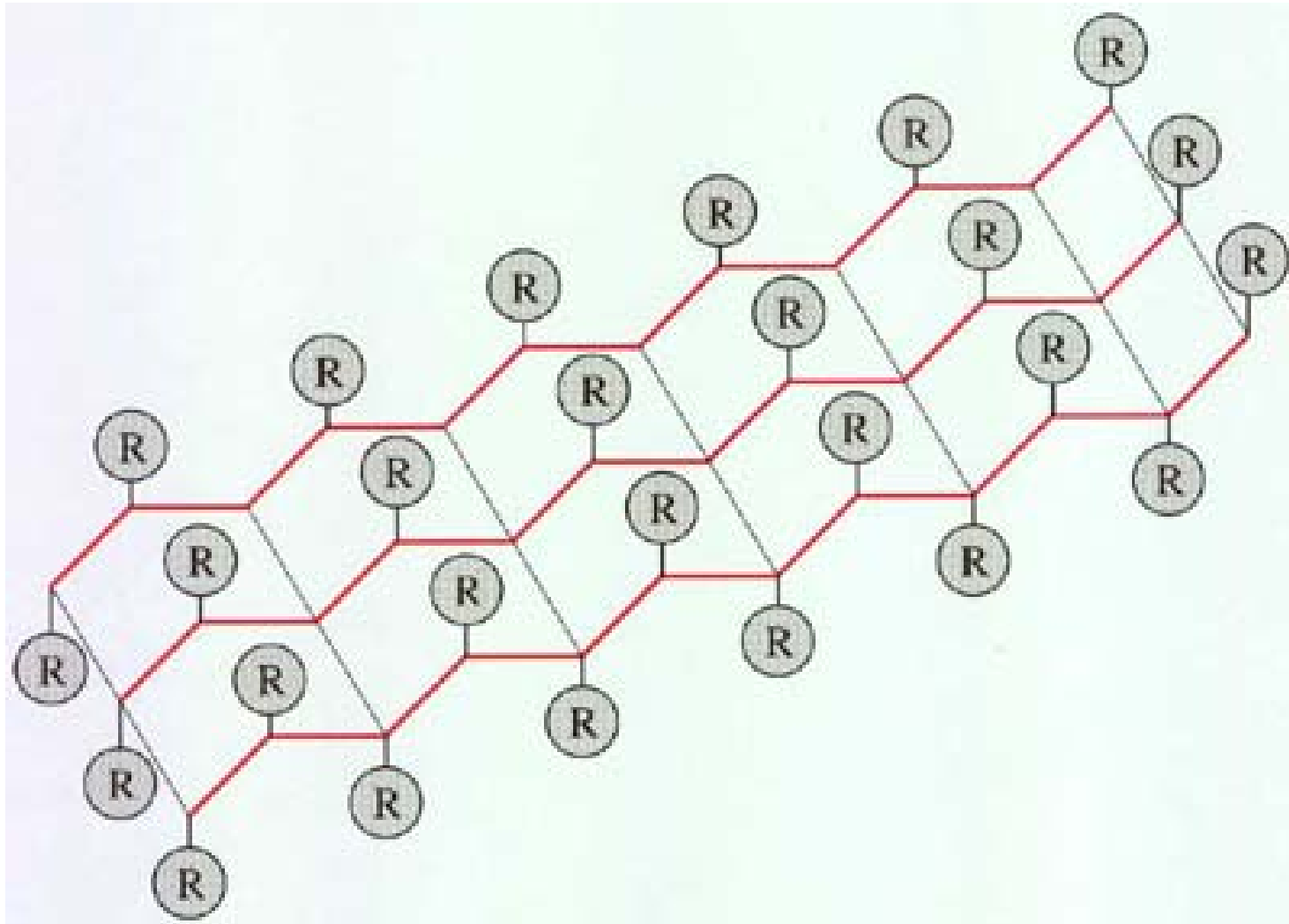
Structural formula of a protein molecule.



Two representations of the alpha-helical conformation of a polypeptide chain

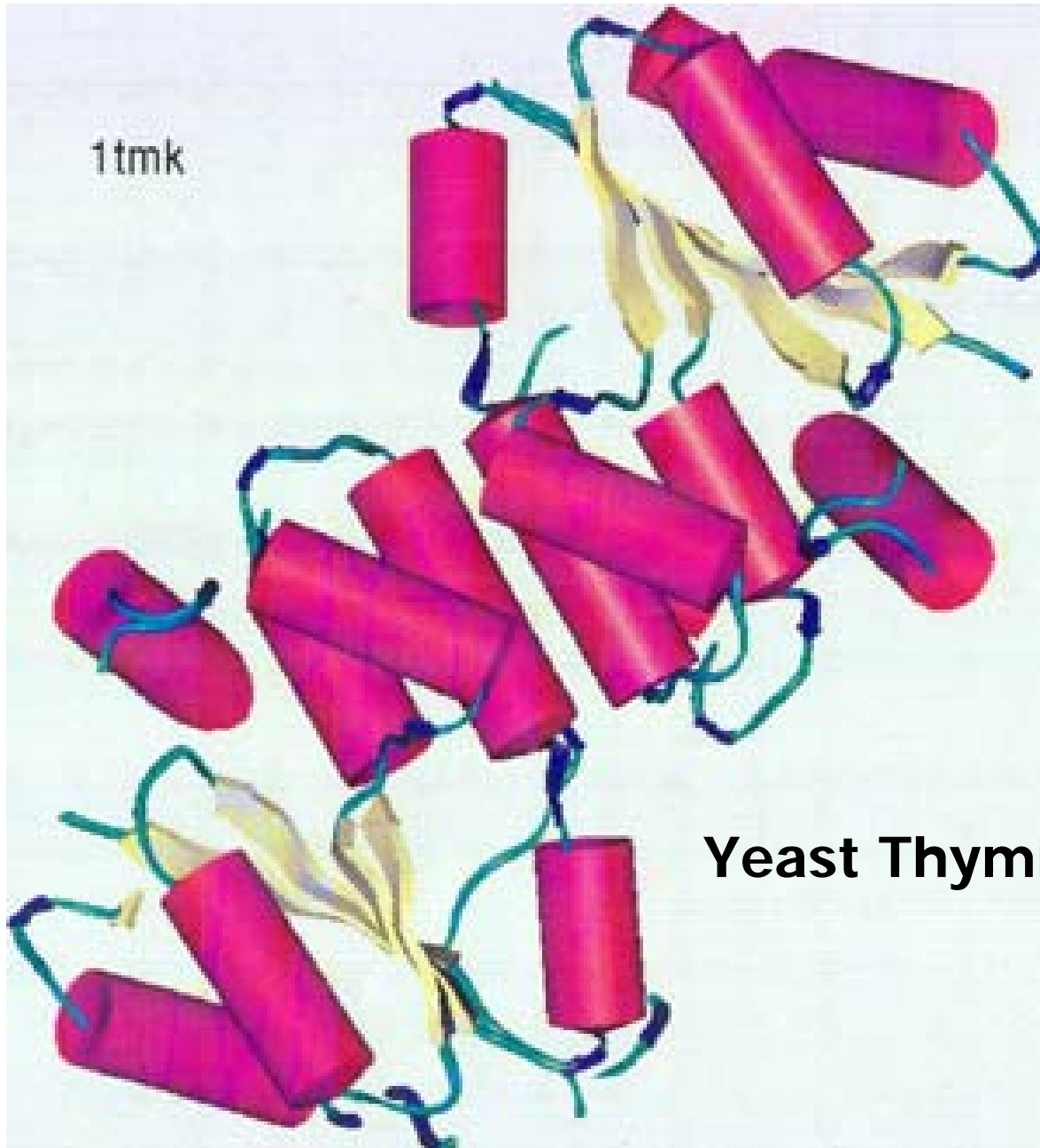


The pleated sheet conformation of polypeptide chains.



The tertiary structure of proteins is maintained by four different types of interactions.






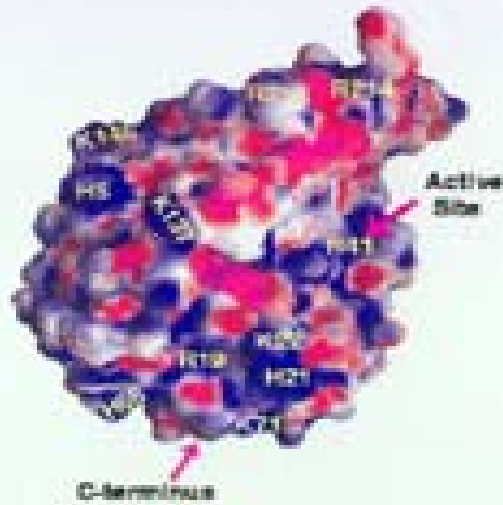
Yeast Thymidylate kinase

mouse Trypsase

human Trypsase

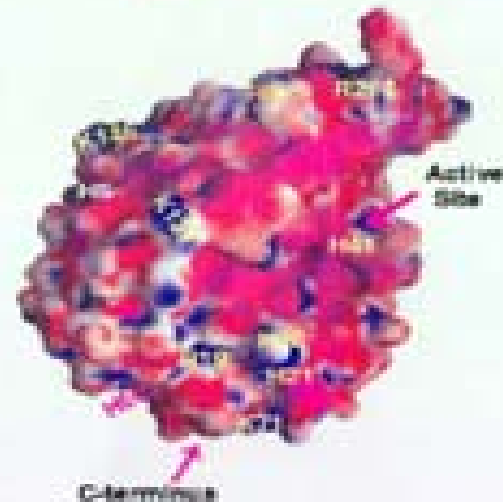
Surface Potential (kV) 

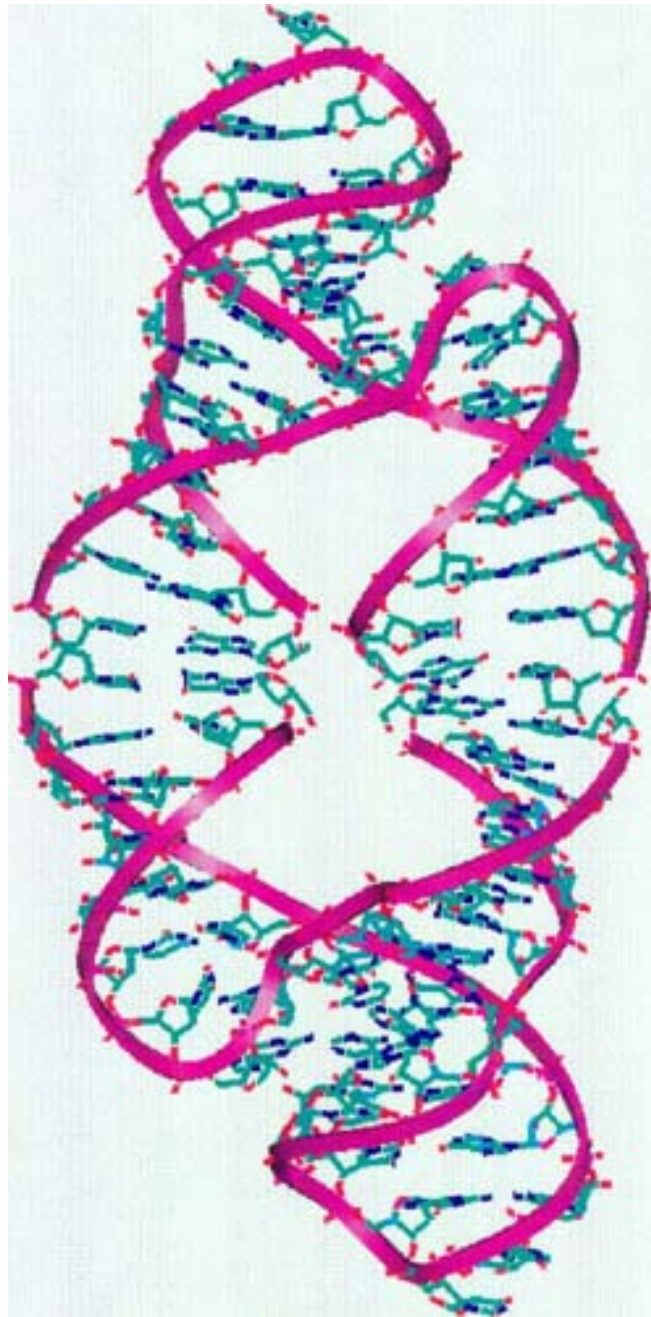
Surface Potential (kV) 



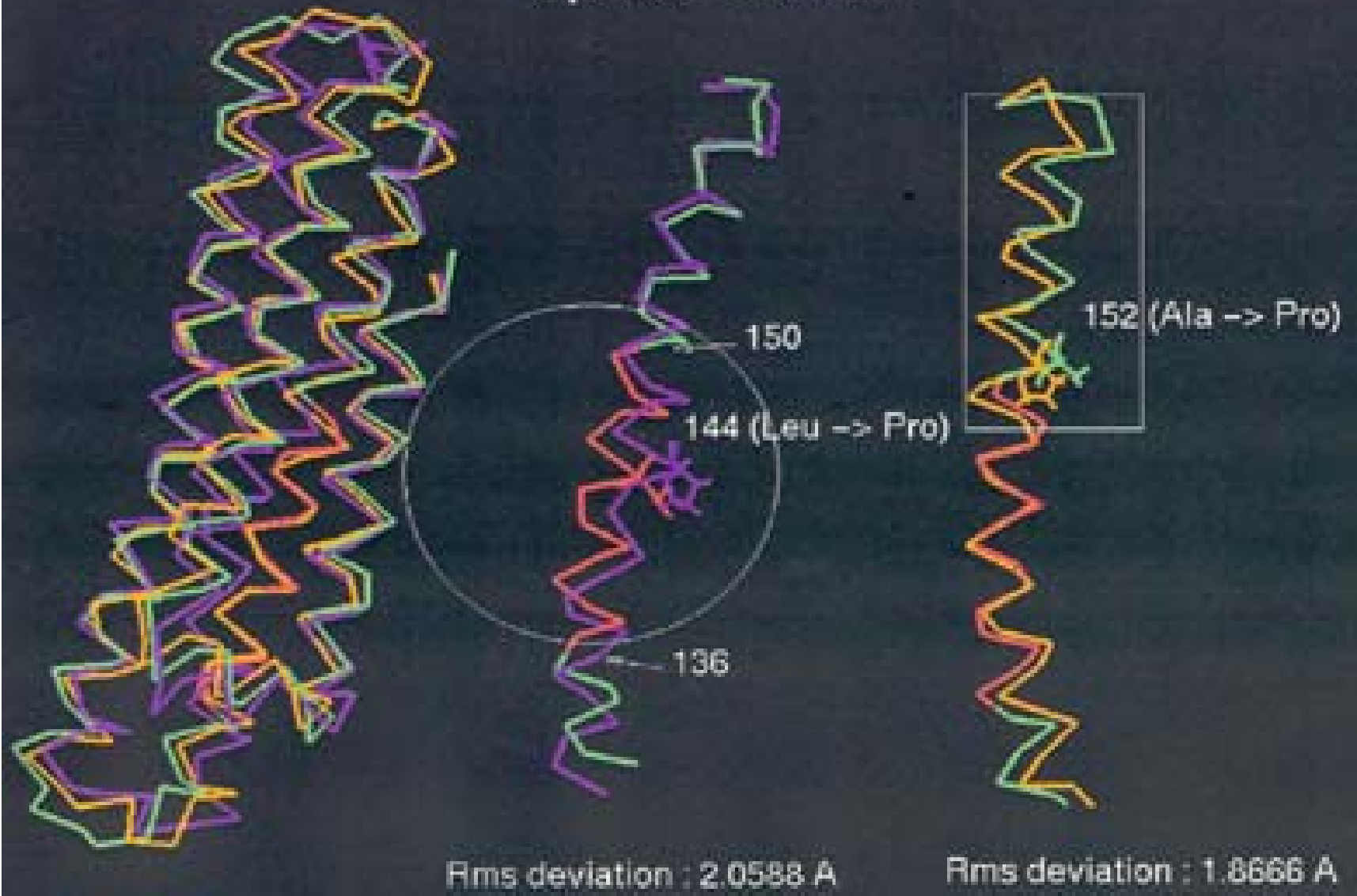
Surface Potential (kV) 

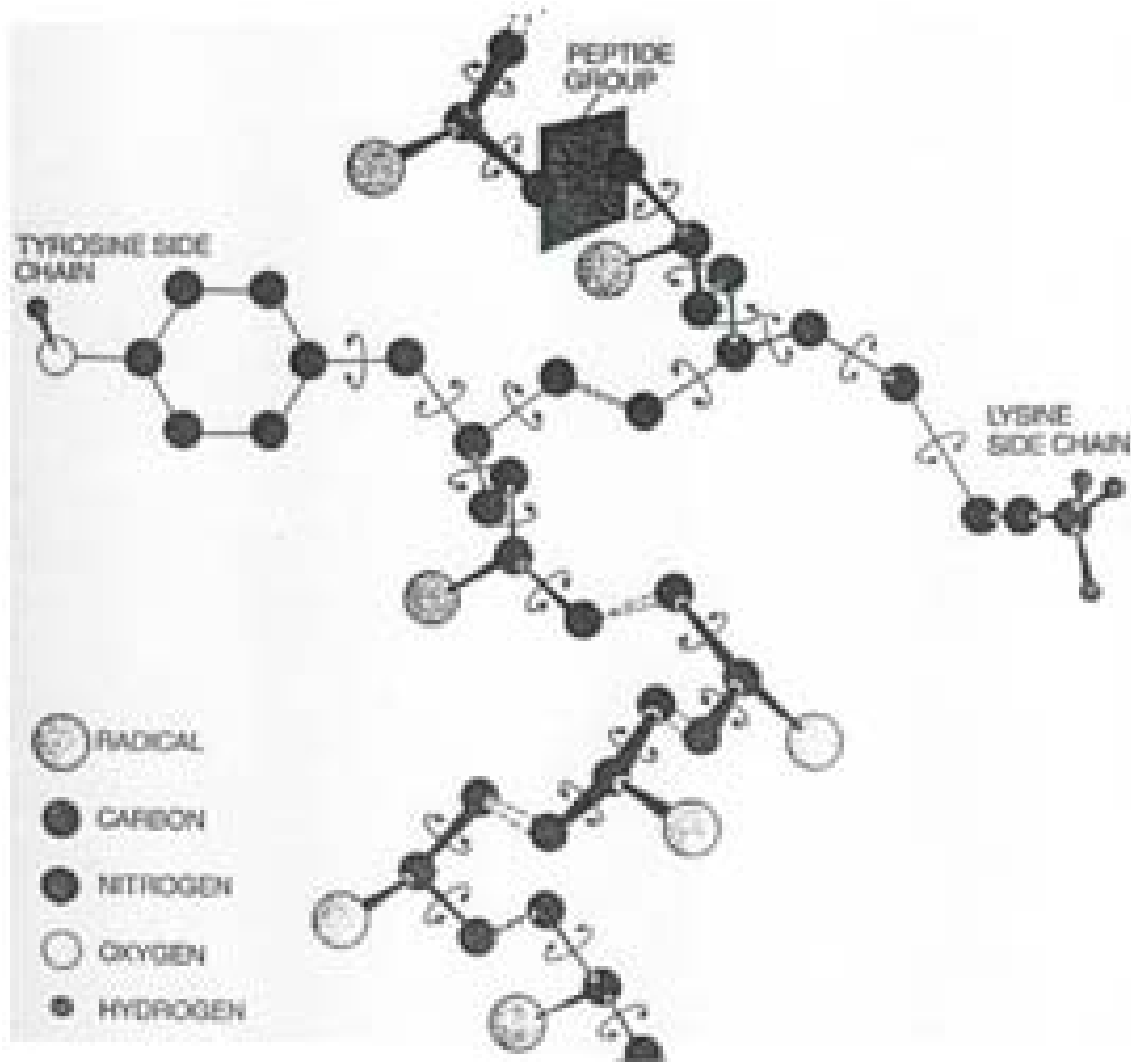
Surface Potential (kV) 





Apo E3 -> Mutant



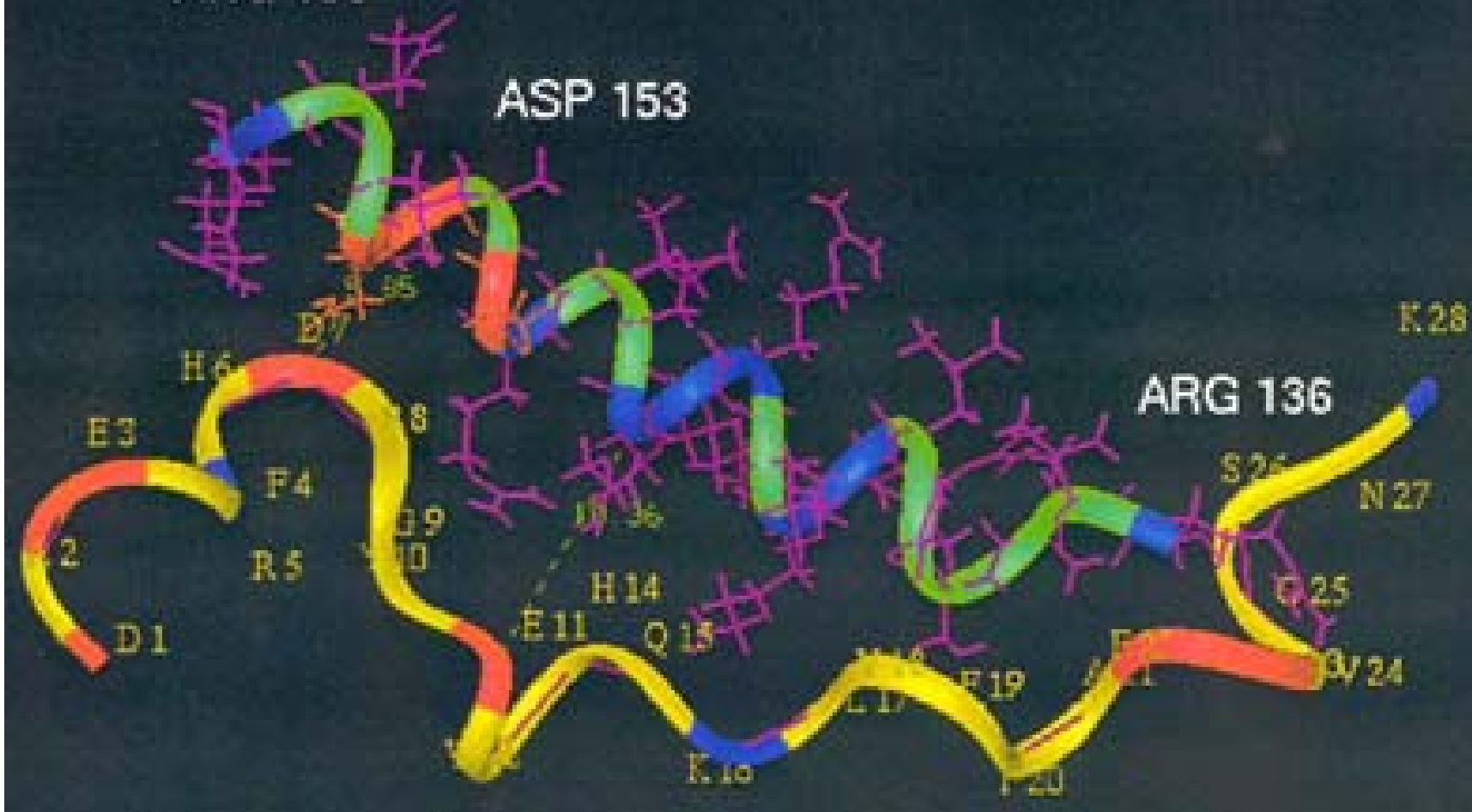


Sites of Flexibility in a polypeptide chain enable the chain to fold into the conformation characteristic of the protein; the sites also facilitate the fluctuations the protein atoms make with respect to their average positions.

ARG 158

ASP 153

ARG 136



Molecular Modeling and Computer simulation

Goals for Macromolecular modeling

- **How a molecule is represented on a computer**
- **What a potential function is**
- **How energy minimization and molecular dynamics work**
- **How modeling methodology enters into the refinement of structure by X-ray or NMR**
- **Why we cannot predict the 3D structure of a protein from its amino acid sequence**

Why to study modeling ?

- **Structures, pictures, sets of atomic coordinates models**
- **The more real data that goes into it, the better the model**
- **Predictive models are generally better than descriptive models**

* Swiss-Model -----ExPASy Molecular Biology Server
<http://tw.expasy.org/>

* The Molecular Modelling Toolkit 2.0
----- An open source program library for
molecular simulation applications
<http://dirac.cnrs-orleans.fr/programs/MMTK/>

What is computer Simulation ?

- **Computational Science uses computers to model, understand and predict properties. Computational modeling offers a tool, complementary with experiment, capable of providing valuable insight into complex reaction mechanisms.**

Understanding the dynamics of biomolecular processes

Macromolecular complexes involved in cellular signaling processing and transmission of signals by conformational changes

Protein 3D Structure Prediction Methods

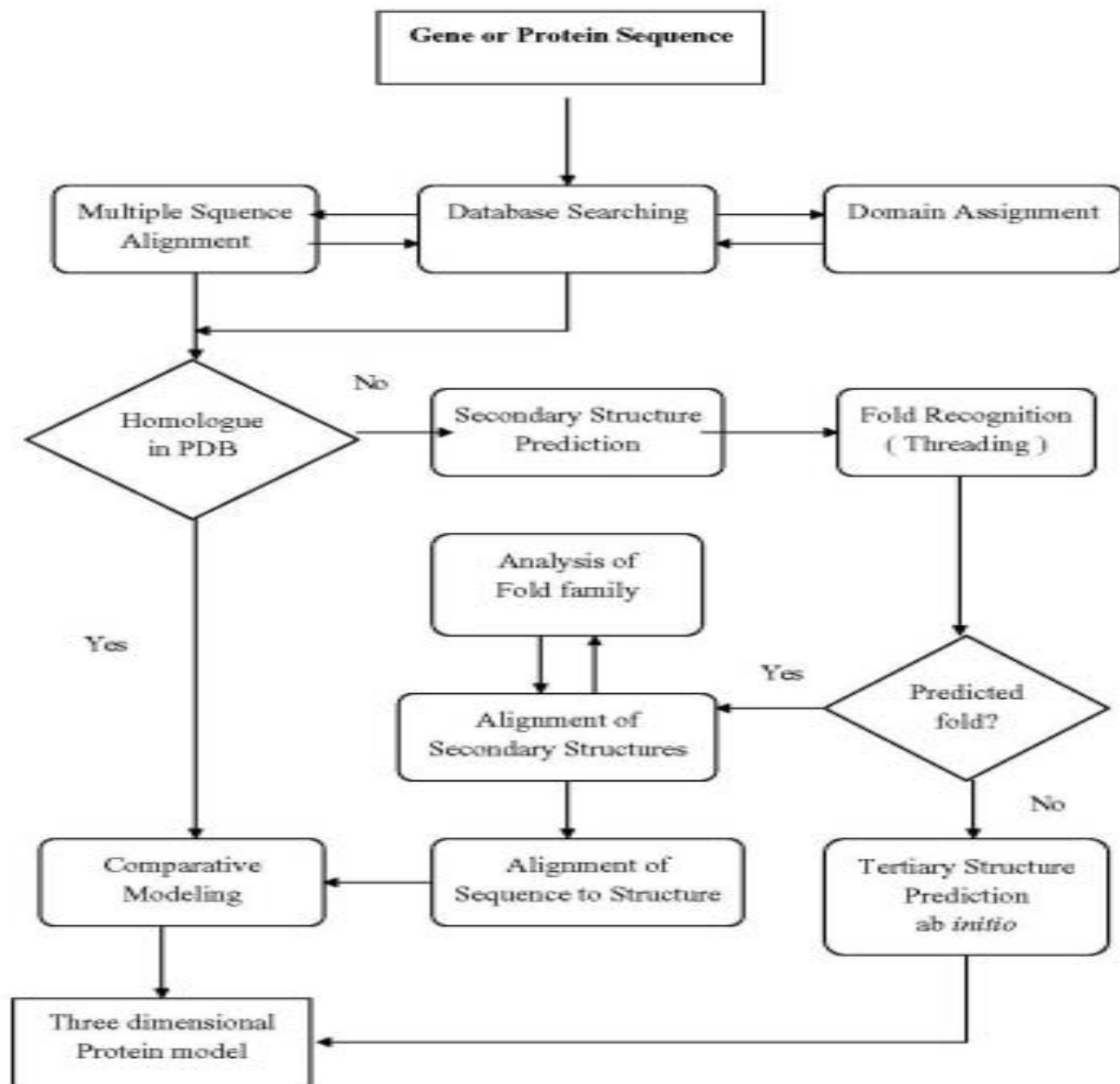


Figure 1: The principle of protein structure prediction.

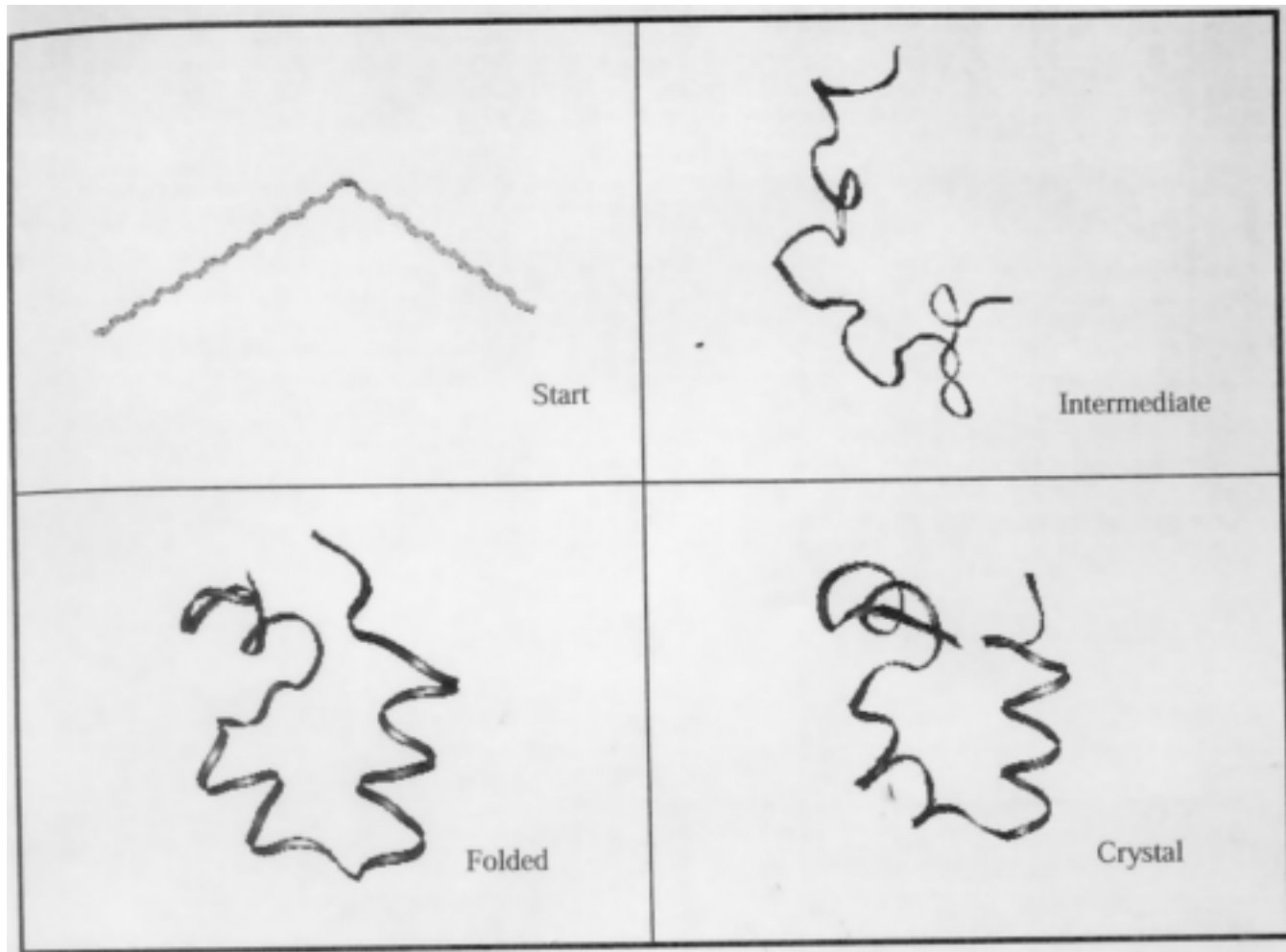
合理的蛋白質分子模型

- 可靠的或可信度高的
- 能夠和 x-光晶體或是多維度NMR光譜實驗分析結果比較
- 能夠符合實驗結果及蛋白質分子結構特徵
- 能夠提供分子結構與生物功能之間關係的詳細資料

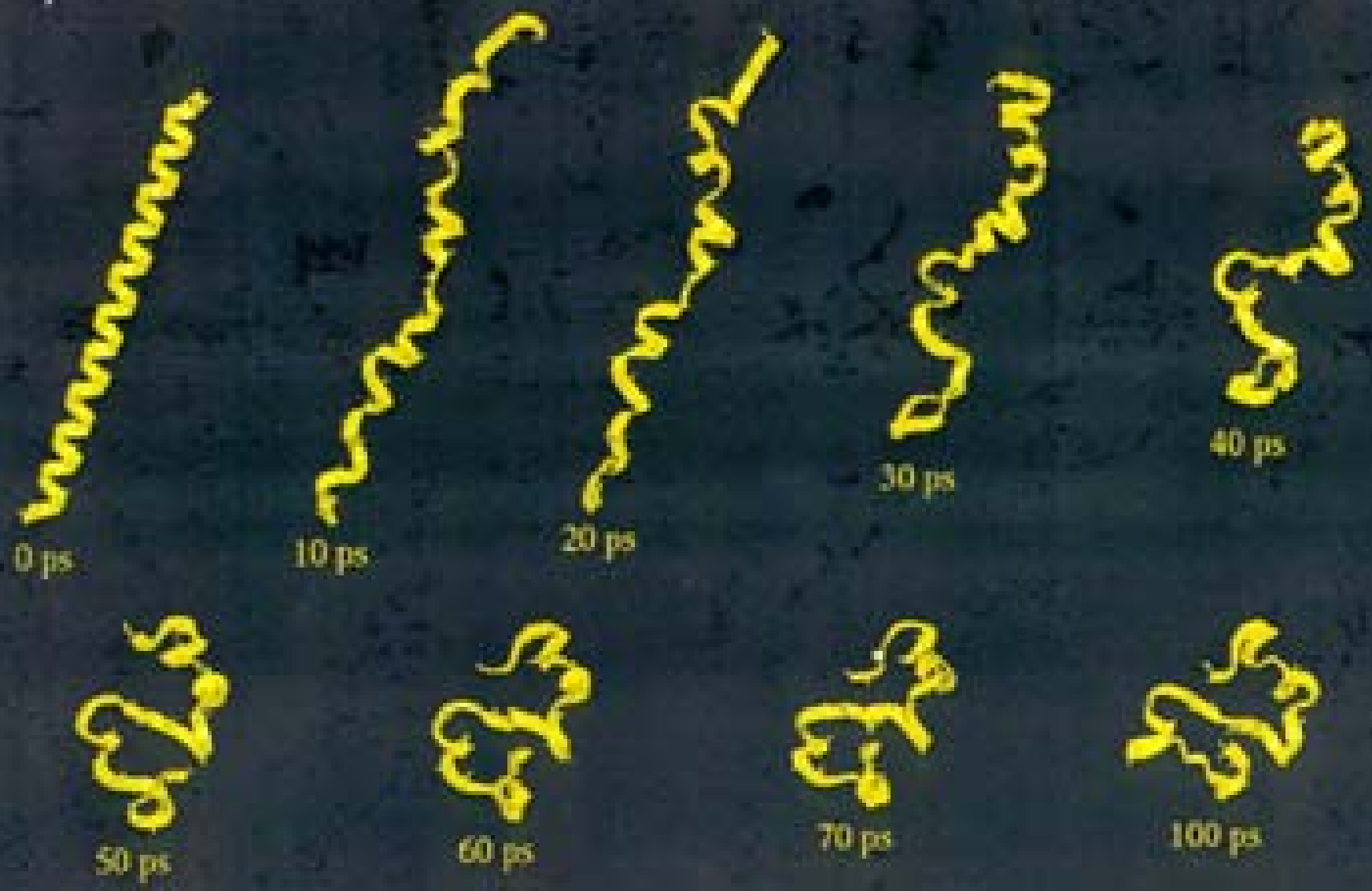
Experimental Data

- **Disulphide bonds, which provide tight restraints on the location of cysteines in space**
- **Spectroscopic data, which can give you an idea as to the secondary structure content of your protein**
- **Site directed mutagenesis studies, which can give insights as to residues involved in active or binding sites**
- **Knowledge of proteolytic cleavage sites, post-translational modifications, such as phosphorylation or glycosylation can suggest residues that must be accessible**
- **Etc.**

How does protein fold?



A β (1-40)



Approach methods

- Homology modeling-----
 - Based on statistical (data base) force field
 - Use of sequence homology with peptides of known three-dimensional structure
- Ab initio methods
 - Based on physically representative force field
 - Use of empirical energy functions ab initio to derive the tertiary structure of minimum potential energy
- Fold recognition (threading) ----- Combinatorial approach
 - Prediction of secondary structure units by the assembly of these units into a compact structure

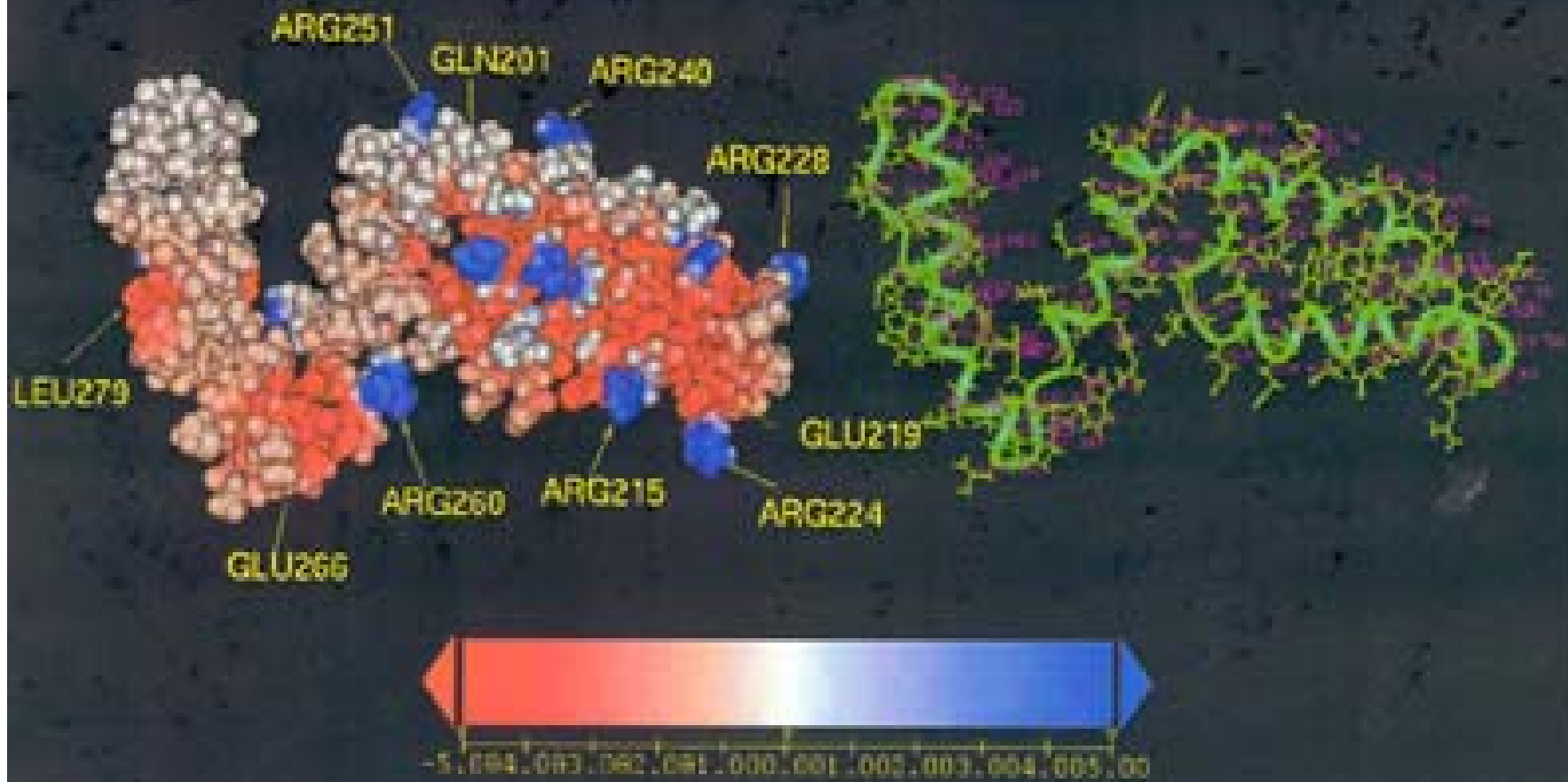
Combinatorial approach

(Cohen, Sternberg et al. 1979-1987)

Three stages are:

- (1). Predict the regular secondary structures, possible with up to 80% accuracy**
- (2). Pack the α helices and β strands into an approximate native fold**
- (3). Use energy calculations to refine the fold into the native structure**

ApoE200-299_Delphi



Ab initio methods

Liu and Beveridge, Proteins, **46**, 2002, p.128

- **Small molecules**
- **Folding too slow to do by full MD therefore have to simplify**

Procedures:

- **Start with random structure (sampling problem)**
- **Randomly move to new conformations**
- **Use MD force field to score results**
- **Keep lowest fraction of structures, repeat**
- **Uses concept of folding funnels**

Secondary structure prediction methods

1. Chou and Fasman method

----- based on the empirical statistically method

(Chou, P. Y. and Fasman, G. D., Ann. Rev. Biochem., 1978, 47, 251)

Ala, Arg, Gln, Glu, Met, Leu and Lys → **helices**

Cys, Ile, Phe, Thr, Trp, Tyr and Val → **sheets**

2. Garnier, Osguthorpe and Robson (GOR)

----- statistically based method

(In "Prediction of Protein Structure and the Principles of Protein Conformation", ed. By Fasman, G.; plenum: new York, 1989, Chapter 10, 417-465)

- 3. Profile 3D, Eisenberg and Lim et al. ----- hydrophobic, hydrophilic and electrostatic properties of side chains.**
- 4. JAMSEK and ALB programs ----- Statistical and stereochemical rules**
- 5. Sander ----- evolutionary information
(neural networks)**
- 6. PHD ----- neural network-based method,
70% accuracy**
- 7. DSSP program ----- based on the known atomic coordinates**
- 8. EMBL in Heidelberg
(<http://www.embl-heidelberg.de>)**

Modeling Protein Structure and Homology

What is Homology modeling?

The protein sequences of unknown structure are matched against the protein sequences of known structures via sequence searching.

- Automatic Sequence alignment methods:
 1. Needleman and Wunsch alignment algorithm
(Needleman, S. B. and Wunsch, C. D., Y. Mol. Biol.,
48, 443(1970))
 2. Biosym's new alignment procedure
 3. Scoring matrices

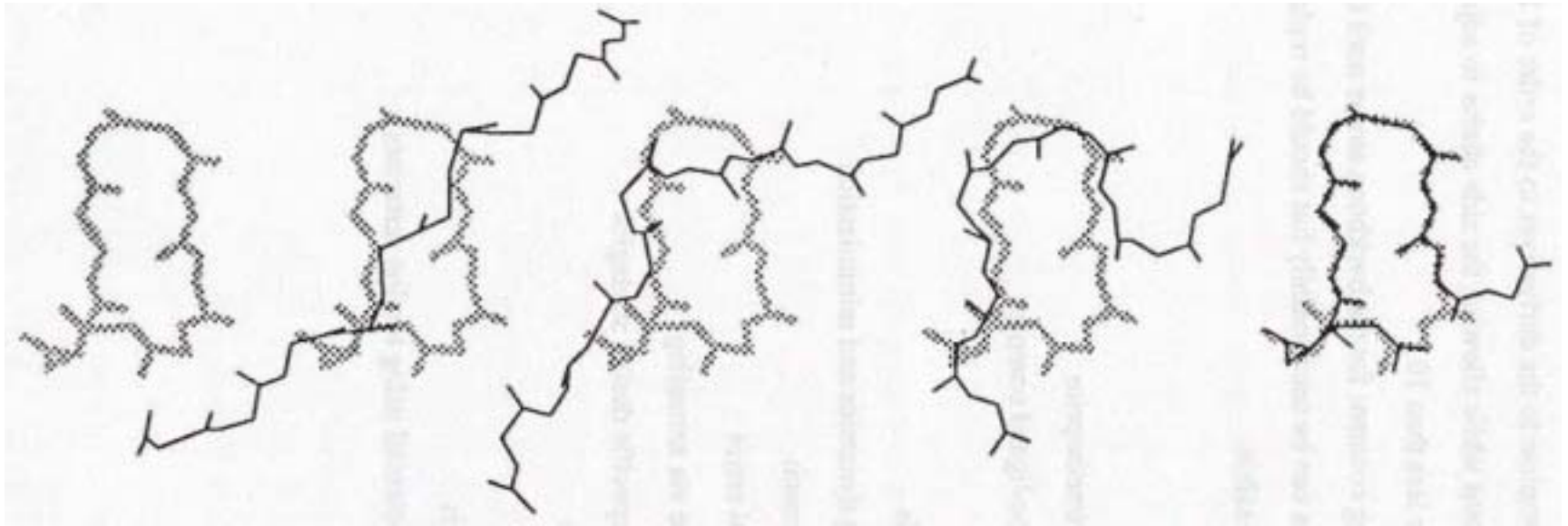
Sequence Alignment

- Sequence based similarity alignment
- Structure based similarity alignment

Theory and Methodology:

1. Determine Structurally Conserved Regions (SCR's)
2. Sequence alignment ---- between the unknown protein and the reference protein within the SCR's
3. Assigning Coordinates within the SCR's
4. Building loop or variable regions
5. Refinement of the structure using Molecular Mechanics: energy minimization and molecular dynamics

Template forcing:



$$F = V + k \left[\sum_i^N (x_i - x_i^0)^2 / N \right]^{1/2}$$

penalty function

x_i^0 : template atom coordinate (analog)

x_i : forced molecule

What is sequence searching?

The sequence of the unknown structure is compared with sequences of known structures stored in a sequence database. Results are scored according to identical and close matches. Certain residues may be weighted more heavily than others (CYS and PRO, for example).

SSKCSRLUTACVYHK

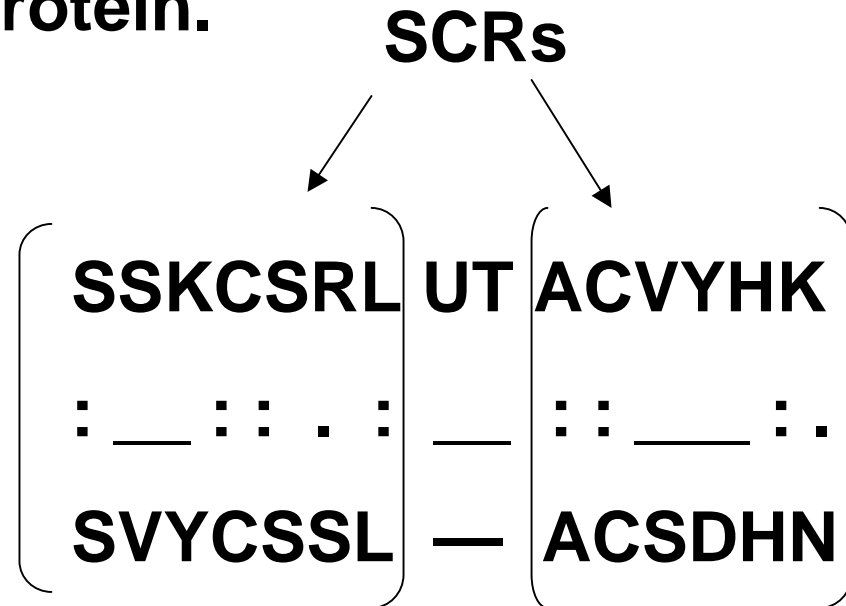
: _ : : . : _ : : _ : .

SVYCSSL — ACSDHN

What are structurally conserved regions?

Structurally conserved regions (SCRs) are those sequences of residues in the protein of unknown structure which are highly homologous with those in a known structure.

It is assumed that proteins with high sequence homology also share high structural homology. SCRs are used to create the bulk of the model of the unknown protein.



Sequence-directed Protein Folding

- **Searching structure databases (PDB or Custom)**
- **Search for similar sequences**
- **Predict Secondary Structure**
- **Align Sequences**
 - § *by Amino Acid sequence*
 - § *by Structures*
- **Superpose Structurally Conserved Regions**
- **Create Templates**
 - § *Single Structure*
 - § *Average of Multiple Structures*
- **Mutate and place side chains**

Tertiary structure prediction method ----

Profile-3D

What is Profiles 3-D ?

A method to fold protein sequences into a known 3D structure.

Algorithm:

This method is developed by Dr. David Eisenberg at UCLA and measures the compatibility of an amino acid sequence with a 3D structure by reducing the structure to a 1D representation (as 3D profile) that can be aligned with the unknown-structural sequence.

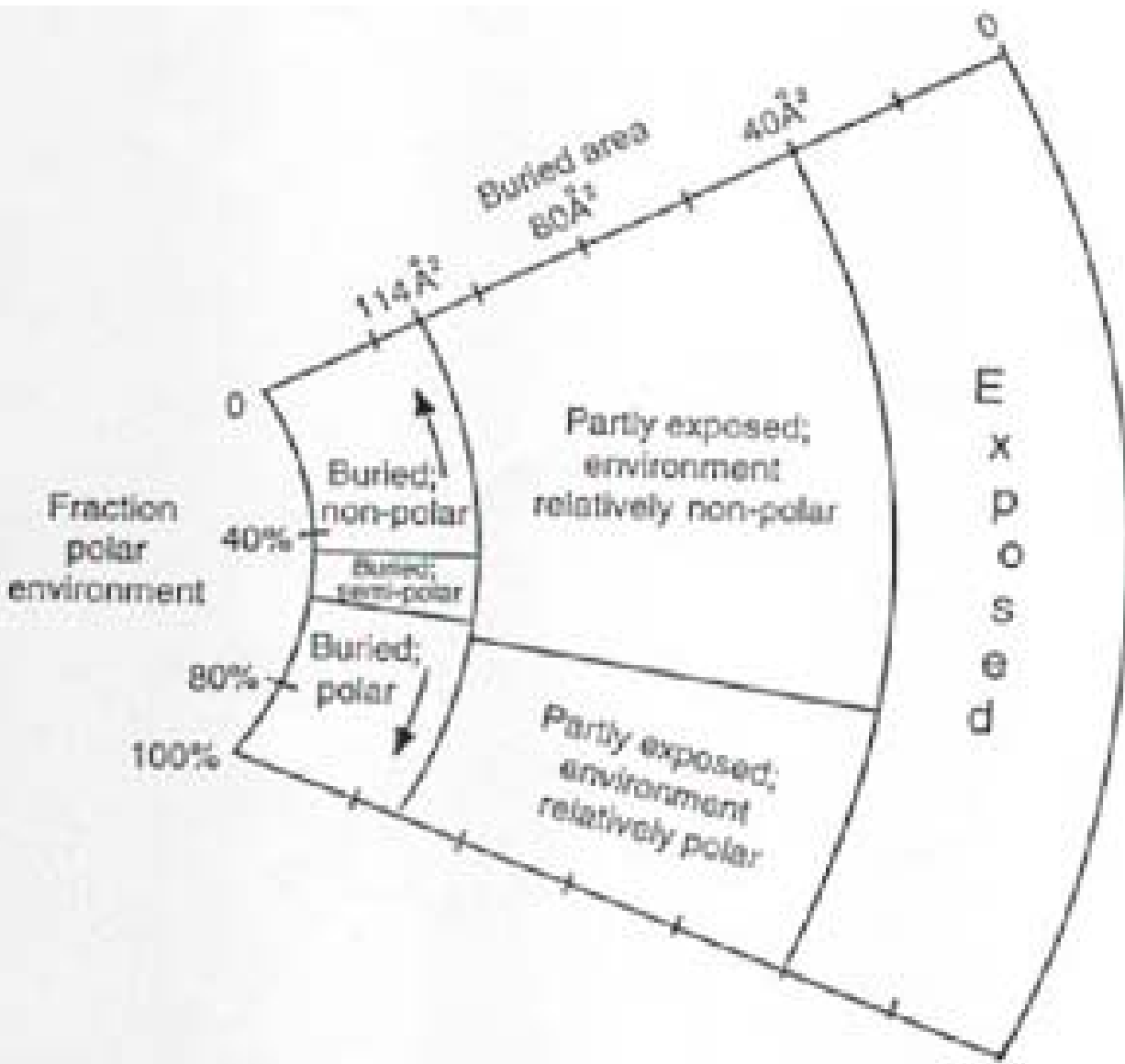
Principle is based on :

- i. The area of the residue buried in the protein and inaccessible to solvent**
- ii. The fraction of side-chain area that is covered by polar atoms**
- iii. The local secondary structure**

**(J. U. Bowie, R. Luthy, and D. Eisenberg,
Science, 253, 164, (1991))**

Procedures :

- 1. Searching a Database of 3D Profiles**
- 2. Assessment of Hyperthetical protein structures**
- 3. Creating 3D Profiles**



Bowie, Luethy and Eisenberg [30] characterise the environments of residues in proteins in three categories: the degree of their exposure to solvent, the polarity of the atoms with which they are in contact (six classes are shown here.) Secondary structure: helix, sheet and other. This gives a total of $3 \times 6 = 18$ classes. The statistical preference of certain amino acids for certain classes can be applied to methods for identifying folding patterns and detection of errors in structure.

From “Computer Modeling in Molecular Biology”, 7^{ed.}
By J. M. Goodfellow.

W = TRP F = PHE Y = TYR L = LEU

I = ILE V = VAL M = MET A = ALA

S = SER Q = GLN N = ASN E = GLU

D = ASP H = HIS K = LYS R = ARG

B, A=B1 buried; hydrophobic environment

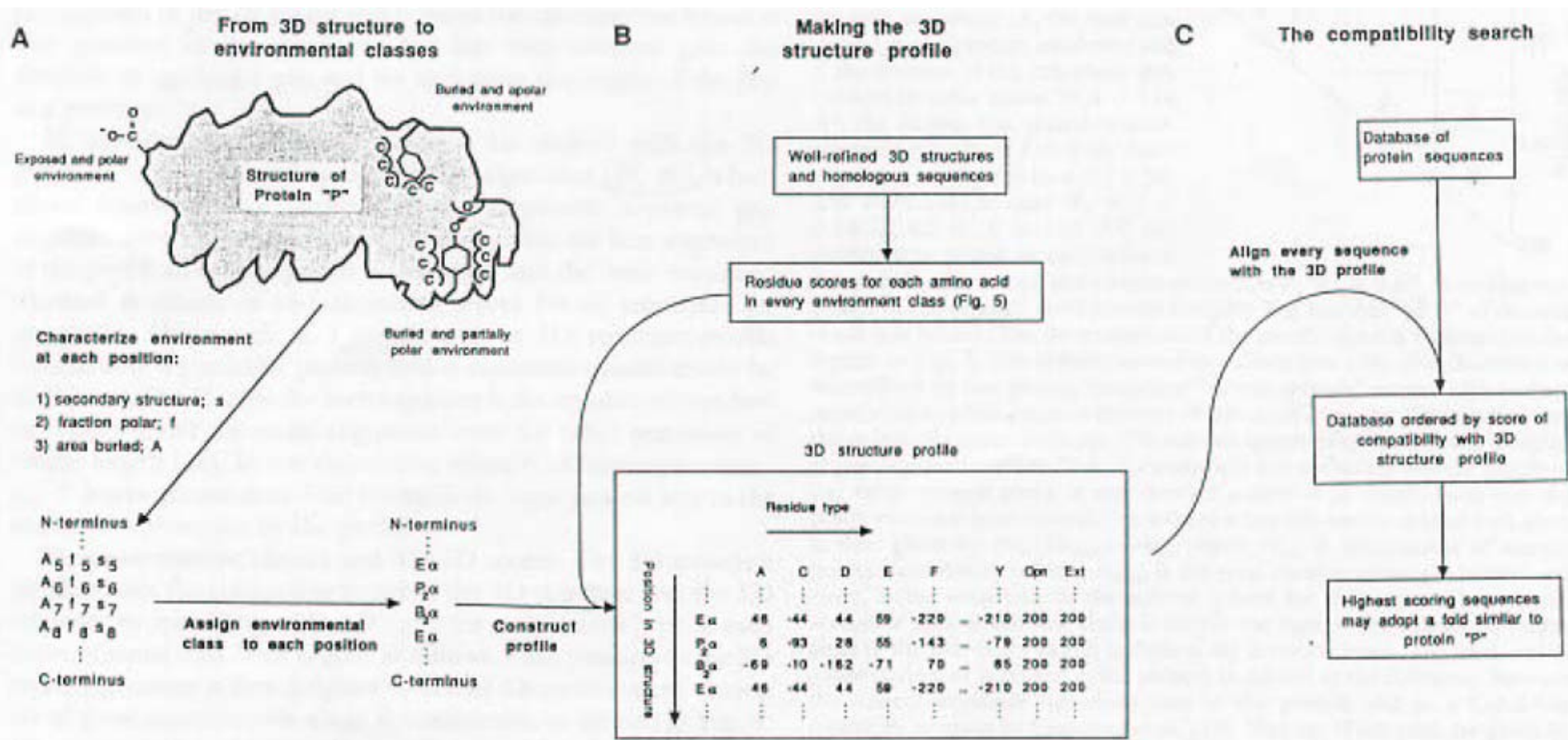
b, B=B2 buried; moderately polar environment

3, C=B3 buried; polar environment

P, D=P1 partially buried; moderately polar environment

P, E=P2 partially buried; polar environment

E, F=E exposed to solvent

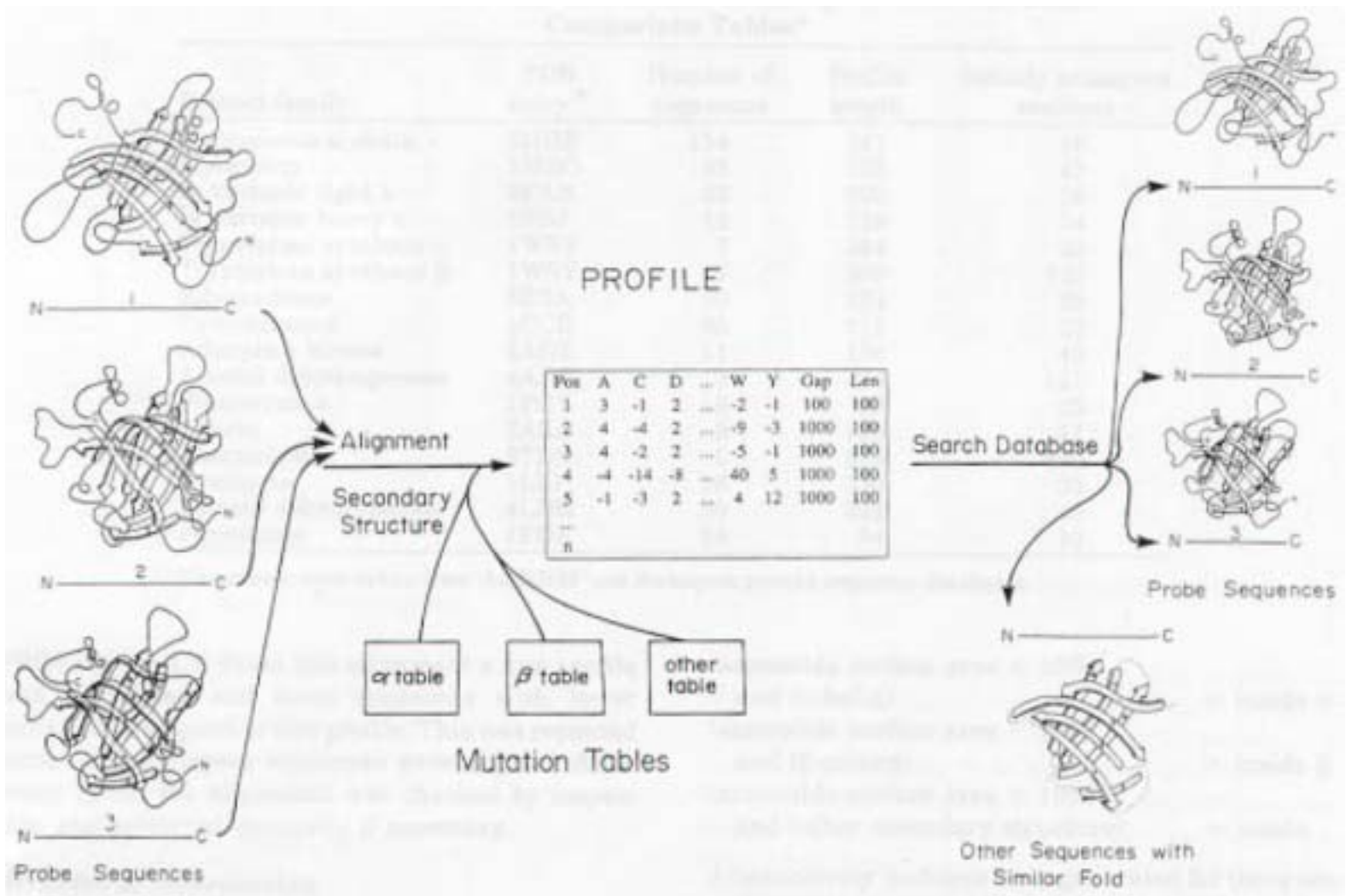


“A Method to identify Protein Sequences that fold into a known 3-D Structure”, Science, 253, 164 (1991), D. Eisenberg et al

Fig – Schematic description of the construction of a 3D structure profile (A and B) and of a 3D compatibility search of the sequence database (C). The 3D structure profile shown at the bottom of (B) is a protein of the profile for sperm whale myoglobin, giving scores for only four positions of the structure (corresponding to residues 5, 6, 7, and 8) and for only 6 of the 20 amino acids.

The profile is aligned to every sequence in database with Profile-Search and the scores of the alignments are normalized.

$$Z - \text{score} = \frac{\text{normalized score} - \text{mean score}}{\text{standard deviation}}$$



Sequence: *Neuraminidase*

Normalization:

Curve fit using 43 length pools

0 of 43 pools were rejected

Normalization equation:

Predicted_Score = 13.95 * (1.0 - exp(-0.0027 * Profile_Length - 0.6271)

Correlation for curve fit: 0.7342

Z score calculation:

Average and standard deviation calculated using
847 scores.

5 of 852 scores were rejected

Z_Score = (Score/Predicted_Score - 1.0705) / 0.2310

Summary:

/usr/biosym/220/data/profiles_3d/pdb1e1l.prf	:	15.89
/usr/biosym/220/data/profiles_3d/pdb1e1n.prf	:	12.61
/usr/biosym/220/data/profiles_3d/pdb3sgb.prf	:	5.57
/usr/biosym/220/data/profiles_3d/pdb3fgf.prf	:	4.71
/usr/biosym/220/data/profiles_3d/pdb2pab.prf	:	4.38
/usr/biosym/220/data/profiles_3d/pdb1rbp.prf	:	3.83
/usr/biosym/220/data/profiles_3d/pdb3rhe.prf	:	3.79
/usr/biosym/220/data/profiles_3d/pdb5ega.prf	:	3.78
/usr/biosym/220/data/profiles_3d/pdb3cna.prf	:	3.50
/usr/biosym/220/data/profiles_3d/pdb1ton.prf	:	3.48
/usr/biosym/220/data/profiles_3d/pdb1ifb.prf	:	3.41
/usr/biosym/220/data/profiles_3d/pdb1phv.prf	:	3.40
/usr/biosym/220/data/profiles_3d/pdb2phv.prf	:	3.17
/usr/biosym/220/data/profiles_3d/pdb1gct.prf	:	3.09
/usr/biosym/220/data/profiles_3d/pdb8gch.prf	:	3.05
/usr/biosym/220/data/profiles_3d/pdb1ppd.prf	:	3.02
/usr/biosym/220/data/profiles_3d/pdb3gct.prf	:	3.00
/usr/biosym/220/data/profiles_3d/pdb3ifb.prf	:	2.93
/usr/biosym/220/data/profiles_3d/pdb3ega.prf	:	2.92
/usr/biosym/220/data/profiles_3d/pdb3gct.prf	:	2.87

Position in fold	Environment class	Amino acid type														Gap penalty	
		A	C	D	E	F	G	...	R	S	T	V	W	Y	Opn	Ext	
1	E	12	-46	22	3	-190	113	...	-32	32	12	-91	-214	-94	2	0.02	
2	B ₂	-66	-5	-128	-135	105	-166	...	-80	-117	-76	60	102	112	2	0.02	
3	E α	46	-44	44	59	-220	68	...	-34	15	-17	-110	-135	-210	200	200	
4	P ₂ α	6	-93	28	56	-143	-50	...	50	-18	-5	-48	-114	-79	200	200	
5	E α	46	-44	44	59	-220	68	...	-34	15	-17	-110	-135	-210	200	200	
6	P ₂ α	6	-93	28	56	-143	-50	...	50	-18	-5	-48	-114	-79	200	200	
7	B ₂ α	-69	-10	-162	-71	90	-149	...	6	-147	-150	68	50	85	200	200	
8	E α	46	-44	44	59	-220	68	...	-34	15	-17	-110	-135	-210	200	200	
9	P ₂ α	6	-93	28	56	-143	-50	...	50	-18	-5	-48	-114	-79	200	200	
10	B ₁ α	-66	-73	-197	-174	132	-253	...	-167	-273	-129	66	100	18	200	200	
.	
.	
.	

Environment class	W	F	Y	L	I	V	M	A	G	P	C	T	S	Q	N	E	D	H	K	R
D_{1a}	1.00	1.32	0.18	1.27	1.17	0.88	1.28	-0.88	-0.53	-1.18	-0.73	-1.28	-0.73	-1.08	-1.83	-1.74	-1.87	-0.24	-1.32	-1.67
$D_{1\beta}$	1.17	0.85	0.07	1.13	1.47	1.09	0.55	-0.79	-2.02	-0.94	-0.22	-1.12	-0.91	-1.67	-1.42	-1.85	-2.58	-1.91	-0.59	-1.18
D_1	1.05	1.45	0.17	1.10	1.11	1.02	0.98	-0.91	-1.80	0.26	-1.23	-1.23	-0.91	-1.17	-0.42	-0.22	-1.78	-1.12	-0.58	-2.18
D_{2a}	0.50	0.90	0.85	1.01	0.63	0.88	1.12	-0.88	-1.48	-0.21	-0.10	-1.52	-1.47	-0.23	-0.61	-0.71	-1.82	0.23	-0.78	0.06
$D_{2\beta}$	0.01	1.18	1.08	0.78	1.21	1.08	0.64	-1.25	-2.28	-0.49	-0.87	-0.27	-1.77	-1.22	-0.27	-1.07	-1.41	-0.77	-1.14	-0.25
D_2	1.02	1.25	1.12	0.84	0.81	0.80	0.98	-0.88	-1.88	0.19	-0.08	-0.78	-1.17	-0.78	-0.68	-1.26	-1.28	0.48	-2.34	-0.80
D_{3a}	0.82	-0.03	0.58	0.15	0.24	-0.02	0.88	-0.57	-1.84	-0.68	-1.58	-0.57	-0.98	0.22	-0.58	0.58	-0.50	0.73	0.43	0.96
$D_{3\beta}$	0.75	0.81	1.20	0.18	0.84	0.88	-0.27	-0.92	-1.80	-0.34	-0.84	-0.44	-0.74	0.21	-0.24	-0.14	-0.88	0.82	-0.23	0.13
D_3	1.07	0.70	1.12	0.28	-0.17	-0.02	0.23	-0.98	-0.94	-0.12	-1.28	-0.52	-0.54	0.08	0.24	-0.38	-1.08	1.01	0.18	0.88
F_{1a}	-1.28	-0.82	-0.58	-0.52	-0.24	0.18	-0.22	0.73	-0.43	-0.25	0.95	0.21	0.24	-0.14	-0.54	-0.17	-0.25	-0.52	-0.21	-0.28
$F_{1\beta}$	0.28	-0.49	0.17	-1.02	0.28	0.48	-0.27	0.84	-0.82	-0.25	1.48	0.93	0.25	-0.27	-1.22	-0.73	-1.07	-0.42	-1.21	-0.77
F_1	-1.28	-1.22	-1.21	-0.82	-0.22	-0.21	-1.18	0.48	-0.24	0.88	1.28	0.28	0.48	-0.82	-0.12	-0.21	0.28	-1.12	-0.74	-1.25
F_{2a}	-1.14	-1.42	-0.79	-0.25	-0.54	-0.48	-0.43	0.08	-0.20	-0.28	-0.92	-0.02	-0.18	0.22	-0.02	0.28	0.28	0.28	0.81	0.20
$F_{2\beta}$	-0.78	-0.54	-0.84	-1.20	-0.22	0.12	-0.72	-0.22	-0.28	-1.28	-0.27	0.24	0.28	-0.28	-0.18	0.22	0.18	-0.87	0.28	0.12
F_2	-0.82	-0.88	-0.51	-0.70	-1.09	-0.88	-0.88	-0.12	-0.40	0.44	-0.22	0.08	0.28	0.27	0.20	0.27	0.49	0.12	0.44	0.20
G_a	-1.25	-2.20	-2.10	-1.58	-0.78	-1.10	-0.72	0.48	0.88	0.04	-0.44	-0.17	0.12	0.28	0.28	0.28	0.44	-0.18	0.12	-0.24
G_{β}	0.84	-0.20	0.28	-1.08	-1.47	-1.74	-0.68	0.28	1.48	-0.98	-0.24	0.14	0.82	-0.18	-0.58	-0.12	-0.78	-0.82	-0.22	-0.49
G	-0.14	-1.20	-0.84	-1.12	-1.81	-0.81	-1.67	0.12	1.12	0.20	-0.48	0.12	0.22	-0.22	0.41	0.02	0.22	-0.22	-0.14	-0.20

Fig. 5 – The 3D-1D scoring table. The scores for pairing a residue i with an environment j is given by the information value (61),

$$3D - 1D \text{ score } ij = \ln\left(\frac{P(i:j)}{P_i}\right)$$

When $P(i:j)$ is the probability of finding residue i in environment j and P_i is the overall probability of finding residue i in any environment. These probabilities were determined from a database of 16 known protein structures and sets of homologous sequences aligned to the sequence of known structure as described in Luthy et al. (28). For each position in the aligned set of sequences, we determined the environment category of the position from the known structure and counted the number of each residue type found at the position within the set of aligned sequences. A residue type was counted only once per position. For example, if there were ten aspartates and one glycine found at a position in a set of aligned sequences, then both the Asp and Gly counters were both incremented by only one. The total number of residue replacements in our database was 8273. If the number of residues i in an environment j was found to be zero, the number was increased to one so that $P(i:j)$ was never zero. Boundaries for the environment categories (shown in Fig. 3) were adjusted iteratively to maximize the total 3D-1D score summed over all residues in our database:

$$\text{Total } 3D - 1D \text{ score} = \sum_{ij} N_{ij} \ln\left(\frac{P(i:j)}{P_i}\right)$$

Where N_{ij} is the number of residues i environment j . in this case, if N_{ij} was zero, the number was not increased to one. Instead, that term in the sum was treated as zero

Figure 2: An outline of the state of the computer modeling in structure prediction

Input sequence	Output prediction			
	High resolution model	Rough 3D model	Fold 2D structure	2D structure prediction
Sequences same as protein of known structure	← Energy minimization and Molecular dynamics →			
Sequences differs from proteins of known structure only at point mutations	← Homology modeling →			
Sequences differs from proteins of known structure only in loop regions	← Homology modeling →			
Sequence's similarity is >40%	← Homology modeling →			
Sequences distantly related to several proteins of known structure	← Threading →			
Sequence related to many sequences but none of known structure or Sequence unrelated to any other known sequence	← Predicted interactions between 2D structural units →			
	← 2D structure prediction or <i>ab initio</i> →			

Fig – An overview of the state of the art in structure prediction. This figure is organized into different categories of potential input information and different categories of potential output information. It describes the dependence of the quality and extent of detail of possible inferences, on the information available when undertaking a modeling project. The terms 1-D, 2-D and 3-D structure under “Output Prediction” refer to the information predicted: **1-D** means that only the structural state (e.g. α -helix, β -strand, turn, coil) of a residue is predicted. **2-D** means that a list of a interactions to each residue is predicted (e.g. contact map). **3-D** means that the geometry of interactions of each residue is predicted (e.g. full threedimensional model). **Note that** the prediction of 1-D information may involve contributions from 2-D information (e.g. Secondary Structure Prediction (1-D) involves $i-i-n \dots i-l+n$ terms).

From “Computer Modeling in Molecular Biology”, ed. by J. M. Goodfellow