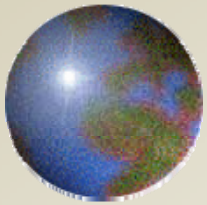
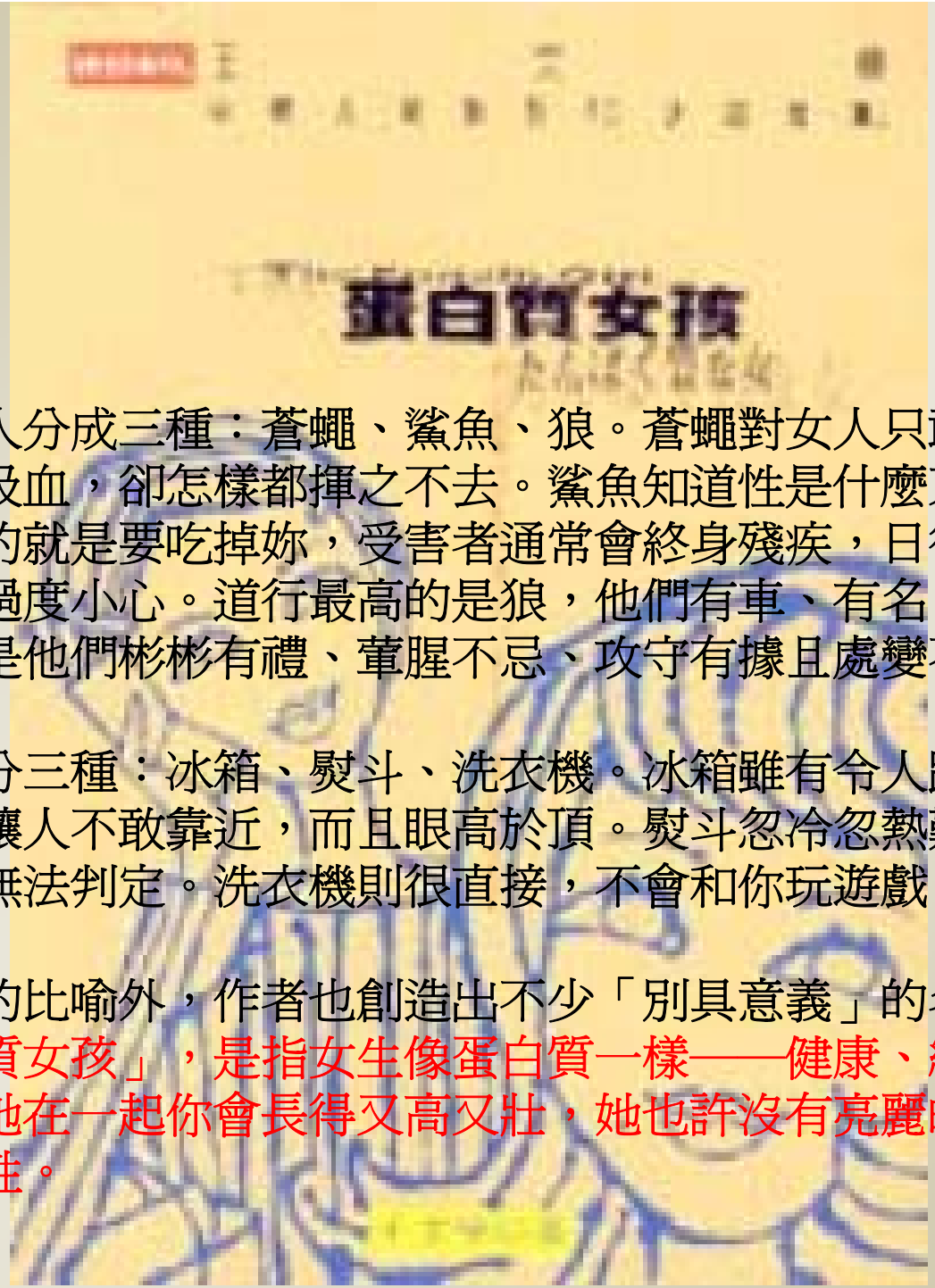


由蛋白質看生物物理

清大物理系
牟中瑜


THE FALL
(WE WISH YOU) A PROTEIN CHRISTMAS



作者將男人分成三種：蒼蠅、鯊魚、狼。蒼蠅對女人只敢繞圈飛行，不咬人不吸血，卻怎樣都揮之不去。鯊魚知道性是什麼東西，他們的目的就是要吃掉妳，受害者通常會終身殘疾，日後對好男人也會過度小心。道行最高的是狼，他們有車、有名、有時間，最重要的是他們彬彬有禮、軟腥不忌、攻守有據且處變不驚。

而女人也分三種：冰箱、熨斗、洗衣機。冰箱雖有令人跌倒的美麗，卻冷酷地讓人不敢靠近，而且眼高於頂。熨斗忽冷忽熱難以捉摸，外表完全無法判定。洗衣機則很直接，不會和你玩遊戲。

除了巧妙的比喻外，作者也創造出不少「別具意義」的名詞，如「蛋白質女孩」，是指女生像蛋白質一樣——健康、純淨、營養、圓滿，跟她在一起你會長得又高又壯，她也許沒有亮麗的外表，卻有很好的個性。

蛋白質折疊是一個典型的生物物理問題

由此問題來看：

* Why do we care about proteins?
  and what is the physics involved?
* What is the challenge?
* What can physicists contribute?

--an experience from past researches

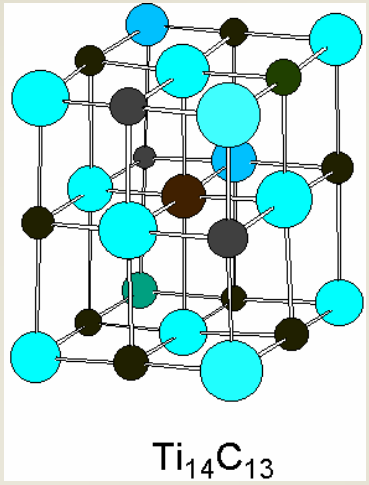
* Summary and outlook

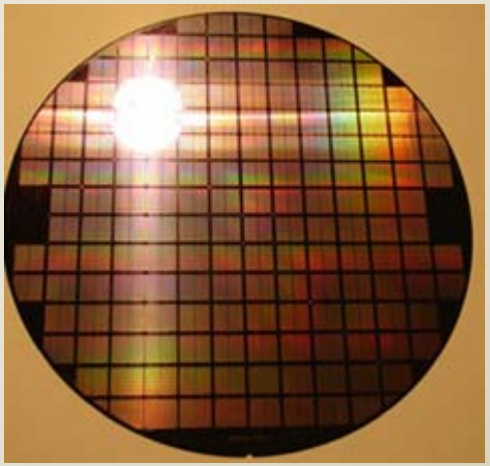
  (何以21世紀是生物的世紀?)

Physics ⇔ Biology

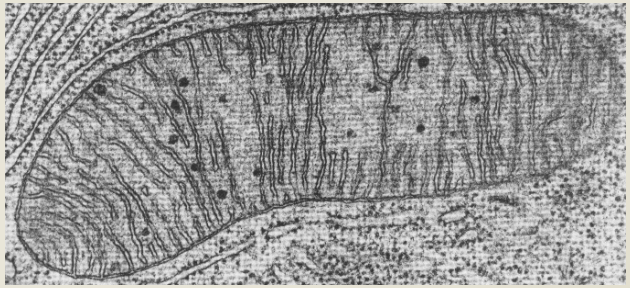
Physics: mechanical universe

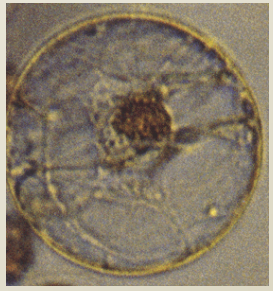
Typical physical matter: regular, crystals, ..





Biological matter: proteins, membrane, ..

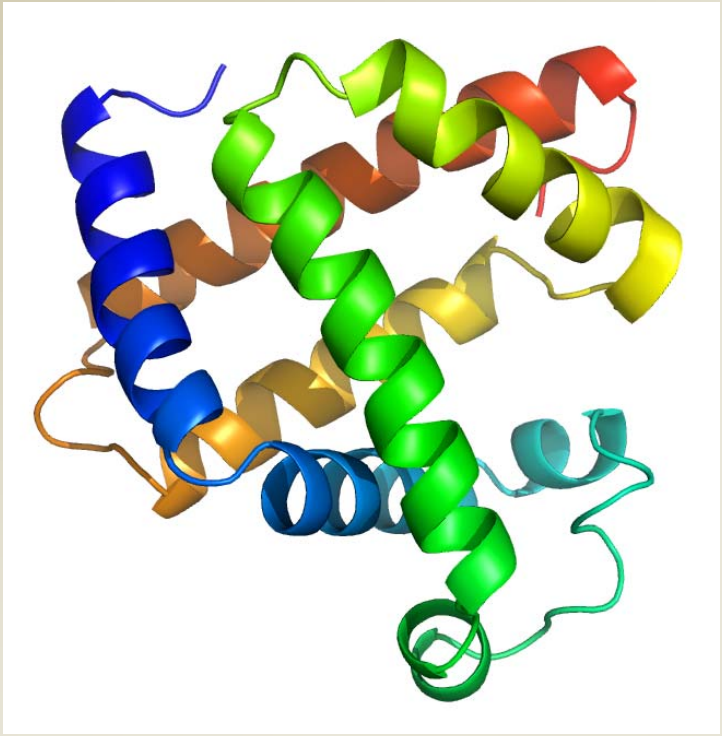


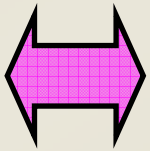


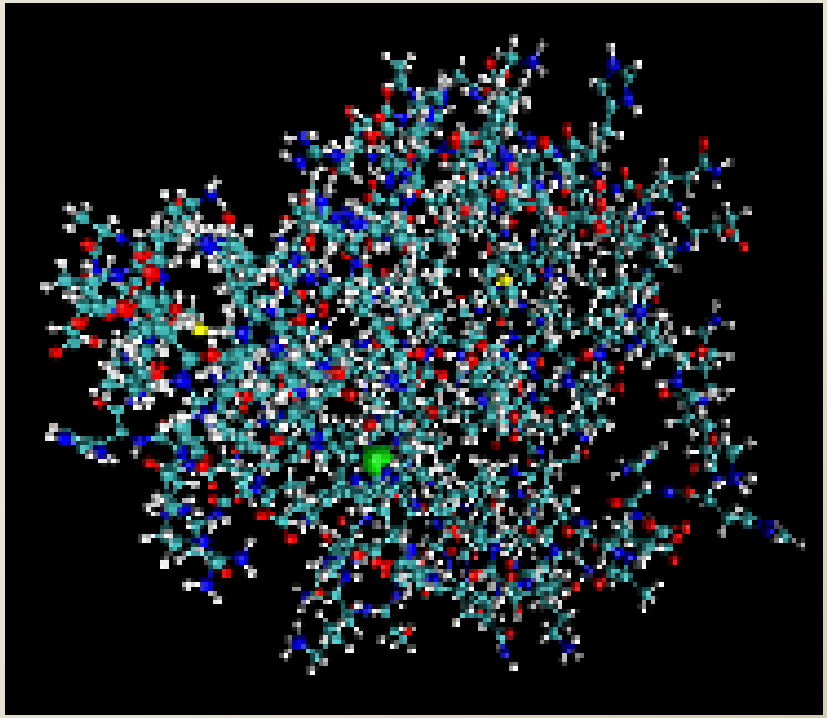
irregular, in water, ..

What are proteins?

Proteins are macromolecules





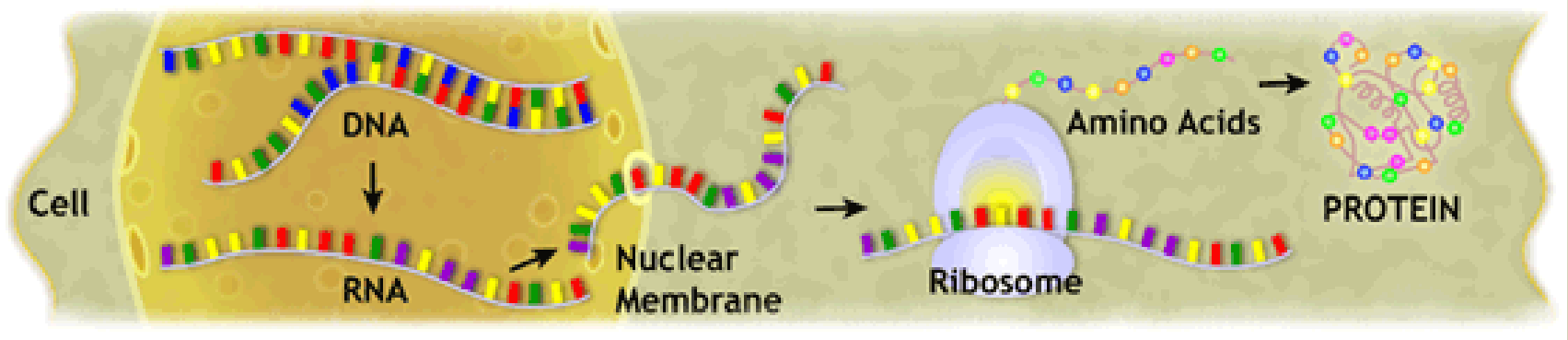


第一個結構被解出來的蛋白質: Myoglobin(肌紅蛋白,存在於紅色的肌肉中):
http://en.wikipedia.org/wiki/Image:Myoglobin.png

Why should we care about proteins?

生物學的基本教條

DNA 轉錄 RNA 再製造蛋白質



http://learn.genetics.utah.edu/units/basics/protein/

蛋白質負責執行生命的主要功能

Building block of structure

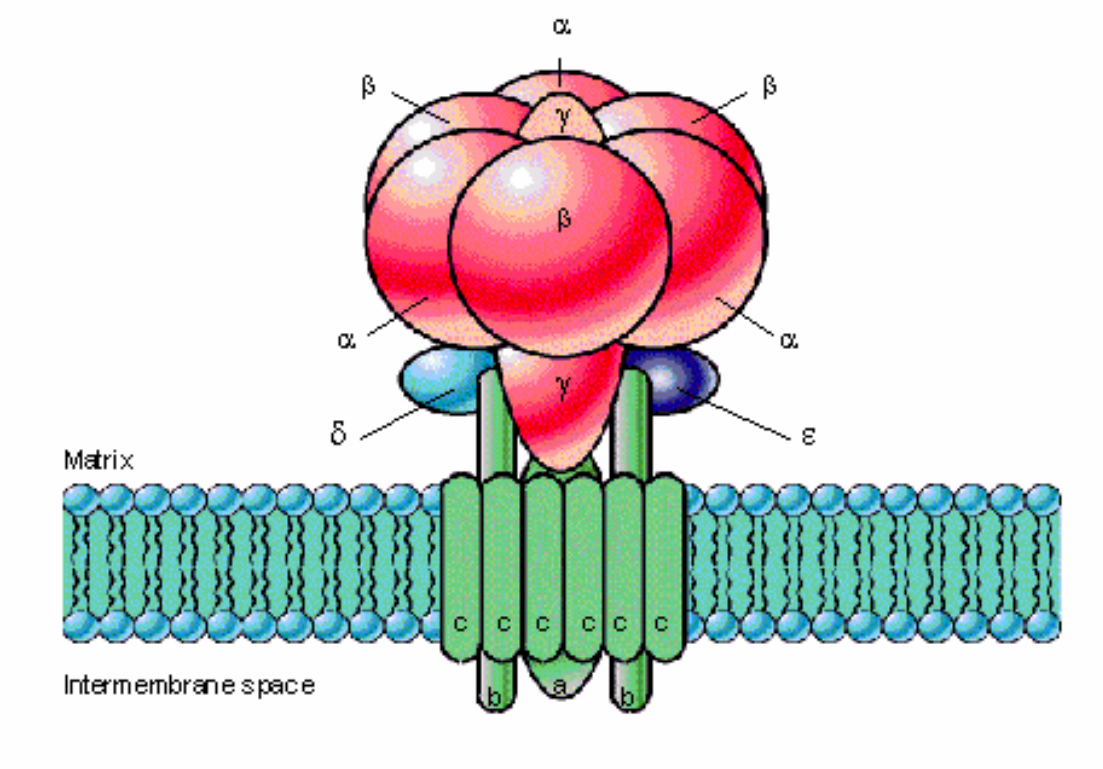
Cell membrane, blood, muscle, bone, etc

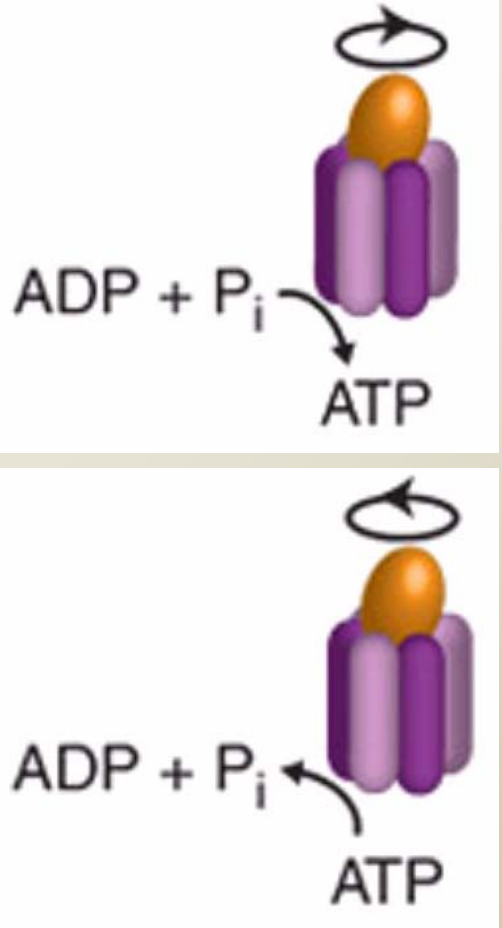
Function

酶(催化)、新陳代謝(如傳輸)、免疫 、荷爾蒙等

許多藥也是蛋白質（如胰島素、退黑激素）

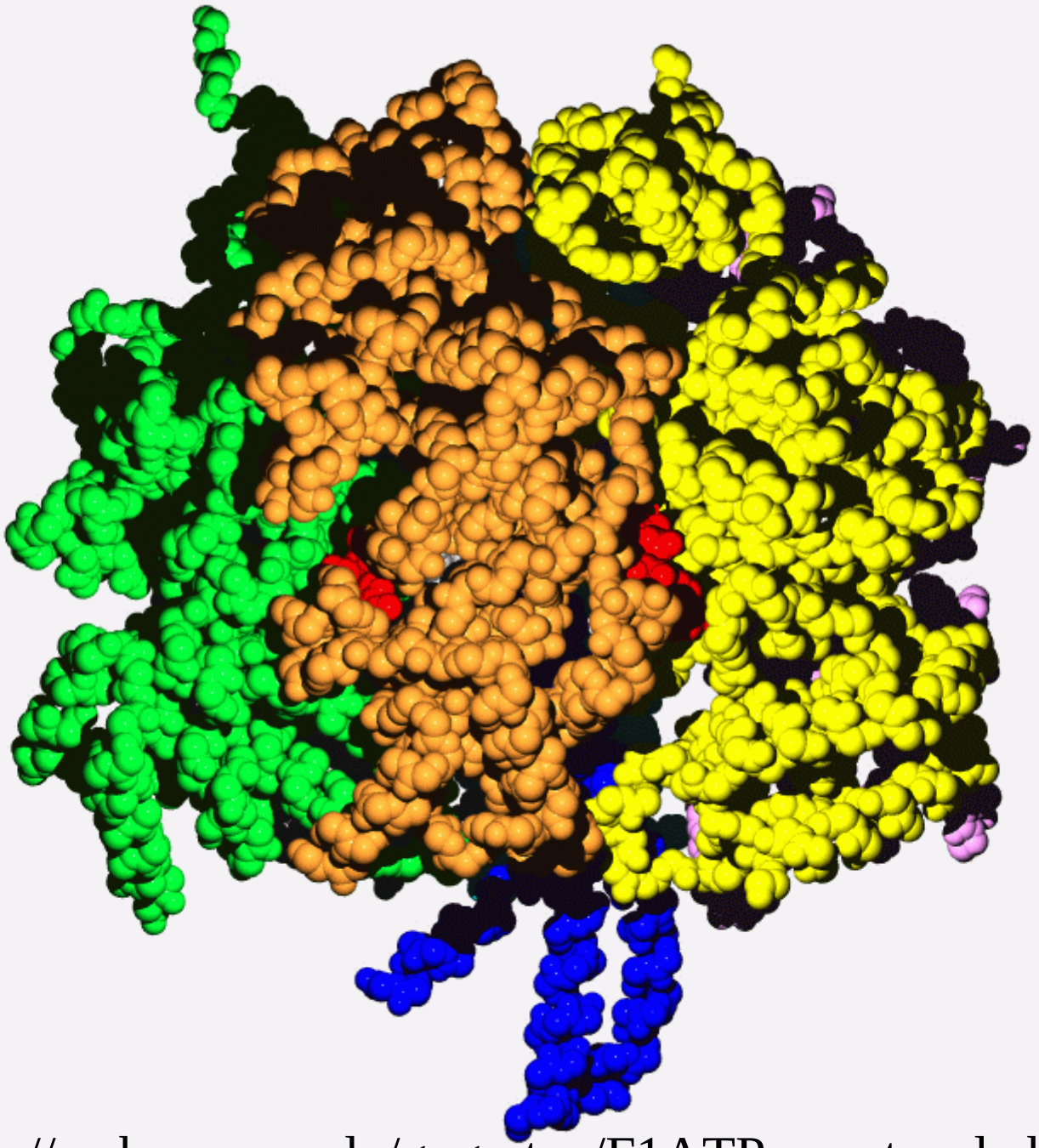
Biomachines: F1-ATPase motor





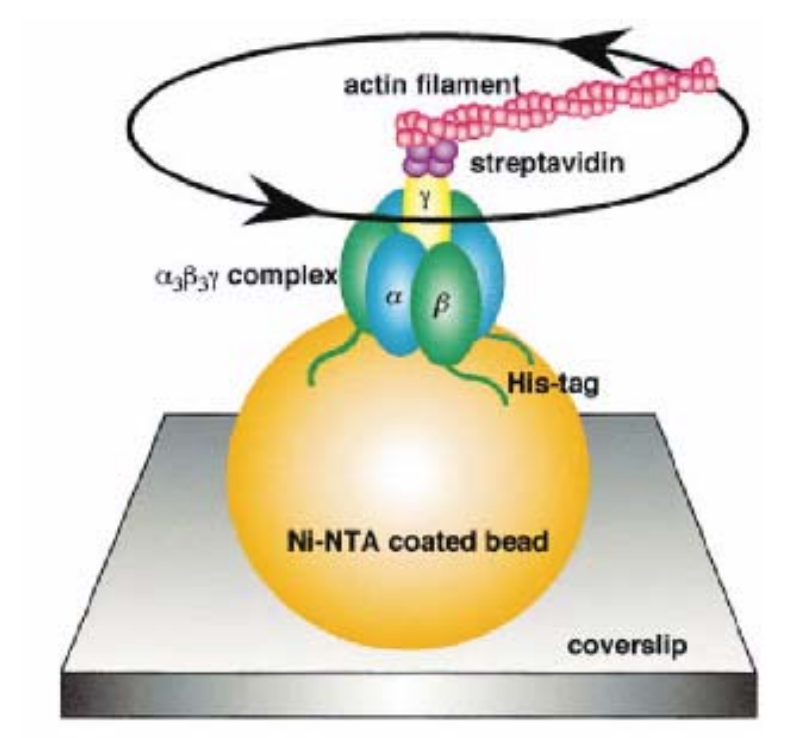
http://www.biomed.metu.edu.tr/courses/term_papers/prost-med-apll-polm_sagay.htm

http://www.k2.phys.waseda.ac.jp/Movies.html



http://web.uconn.edu/gogarten/F1ATPasecatcycle.htm

Single-molecule physiology is now possible:



Noji H. et al. Nature 386, 299 (1997); Yasuda R, Cell 93, 1117 (1998)

http://www.k2.phys.waseda.ac.jp/Movies.html

http://www.k2.phys.waseda.ac.jp/ResearchPDF/Winter03.pdf

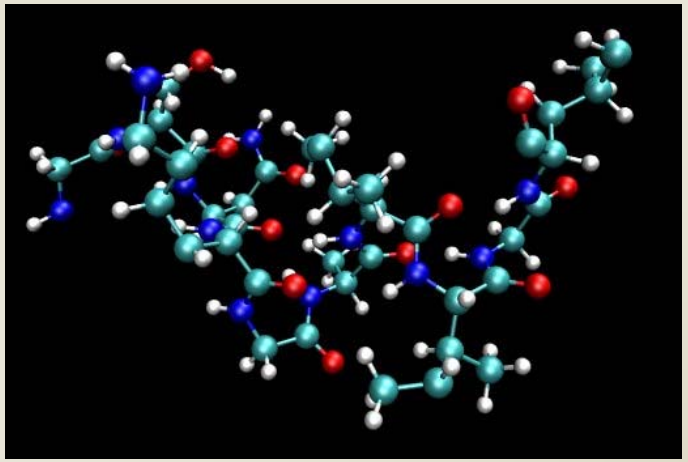
許多疾病起因於蛋白質的不正常功能

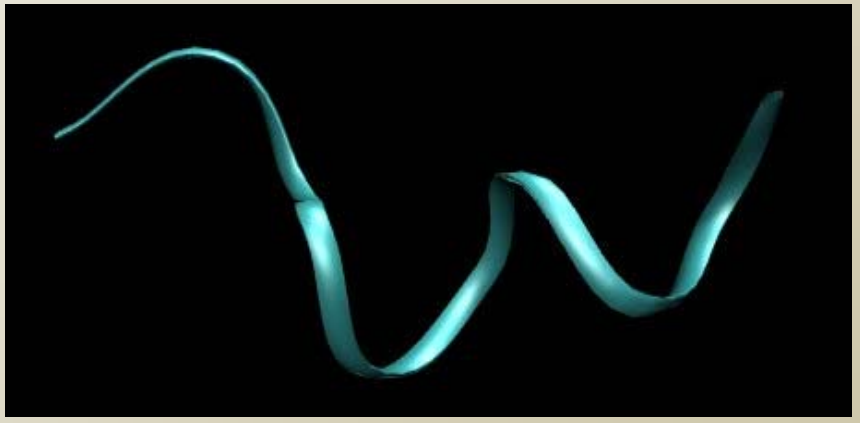
Cystic Fibrosis (囊狀纖維化, protein misfolding)

Aggregation of proteins

Amyloid beta (Alzheimer)

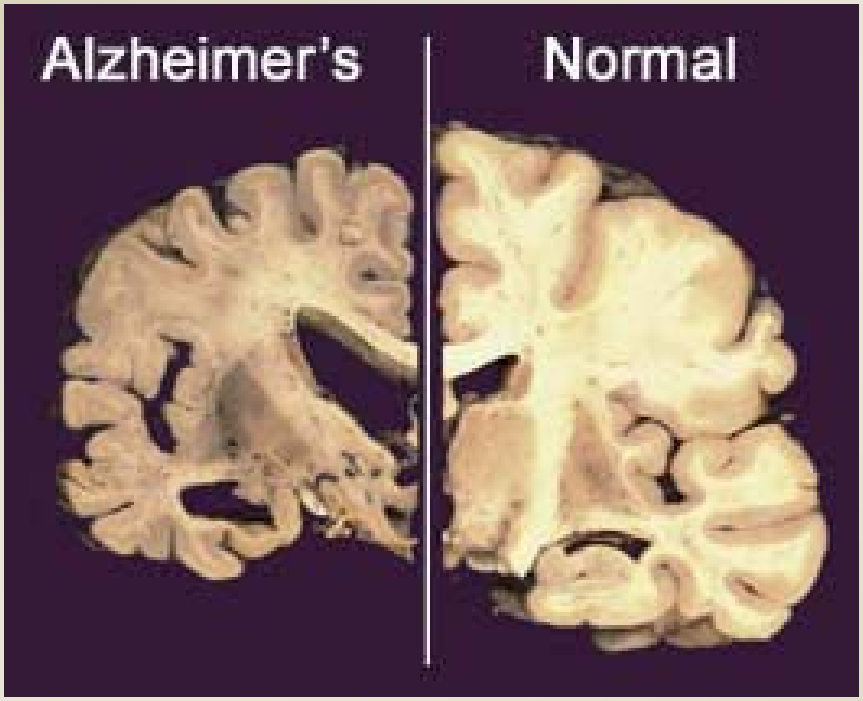
Prions (mad-cow disease)

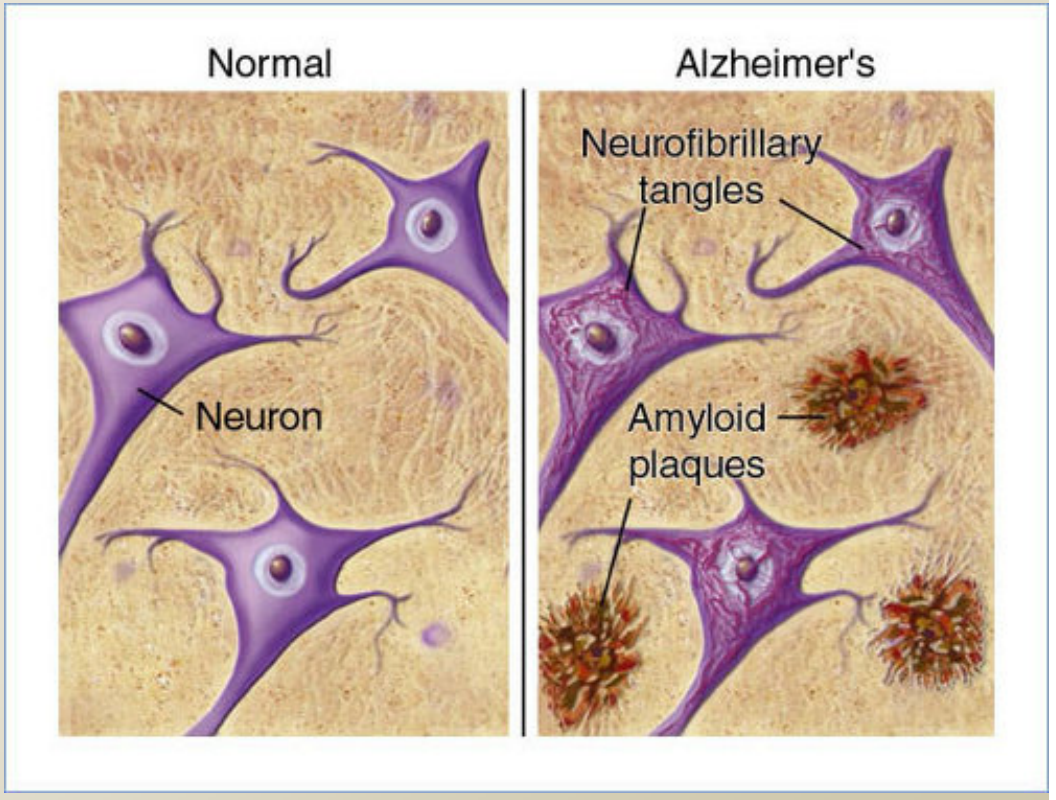


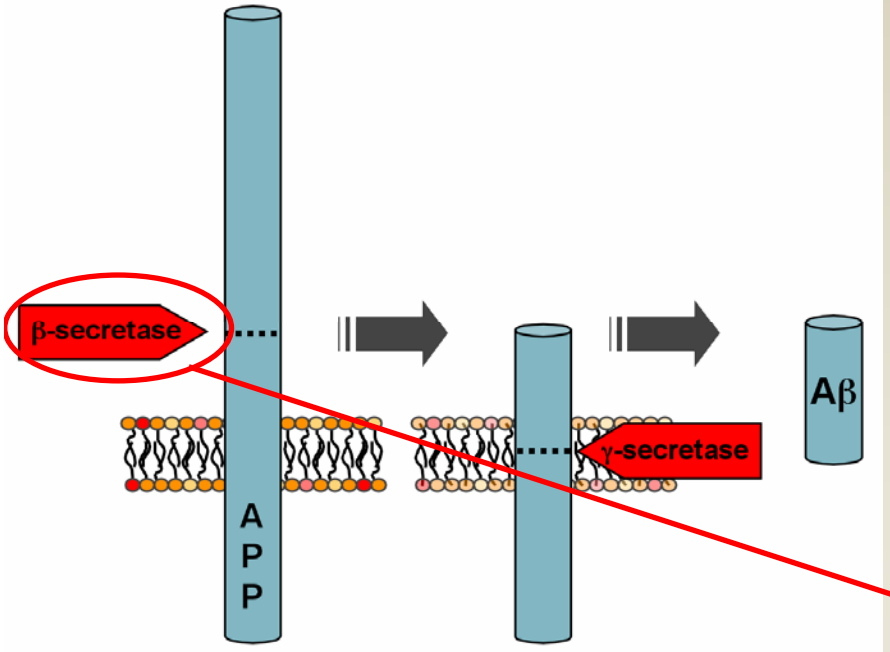


Understanding how proteins work leads to cure many diseases

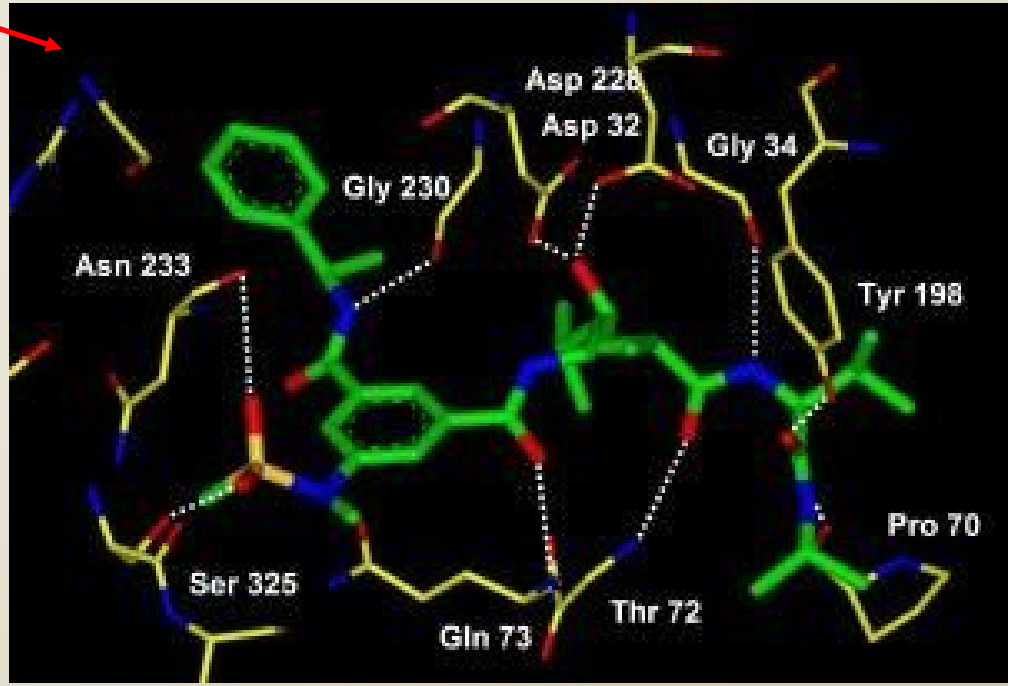
Example: Alzheimer - amyloid beta protein cascade







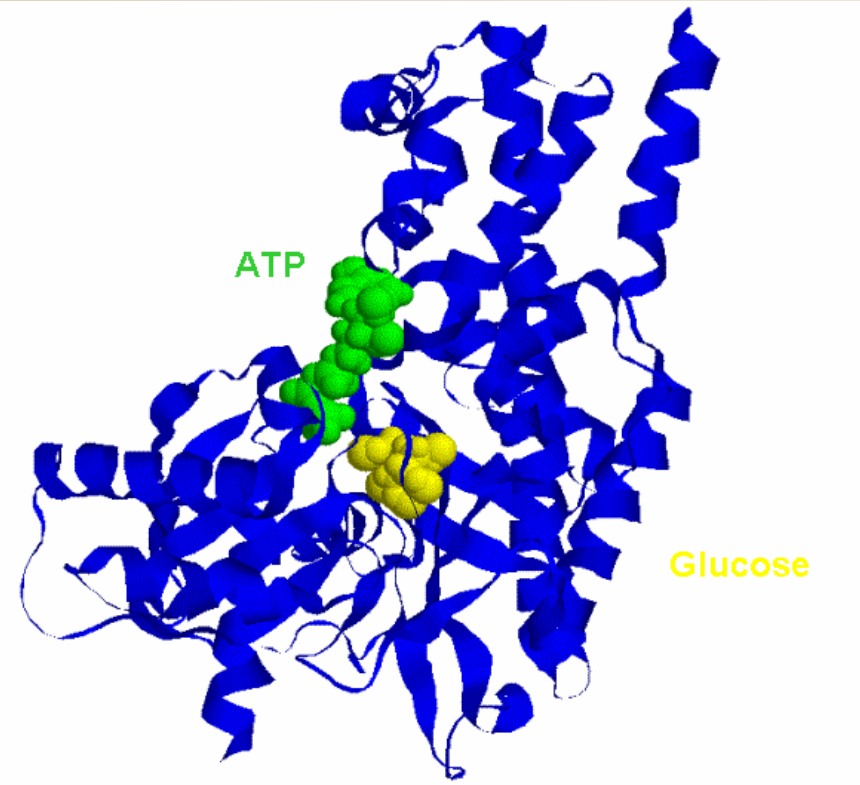
http://www.lbl.gov/Science-Articles/Archive/LSD-Alzheimers-switch.html



http://www.chem.purdue.edu/NewsFeed/newsstory.asp?itemID=214

Arun Ghosh's group

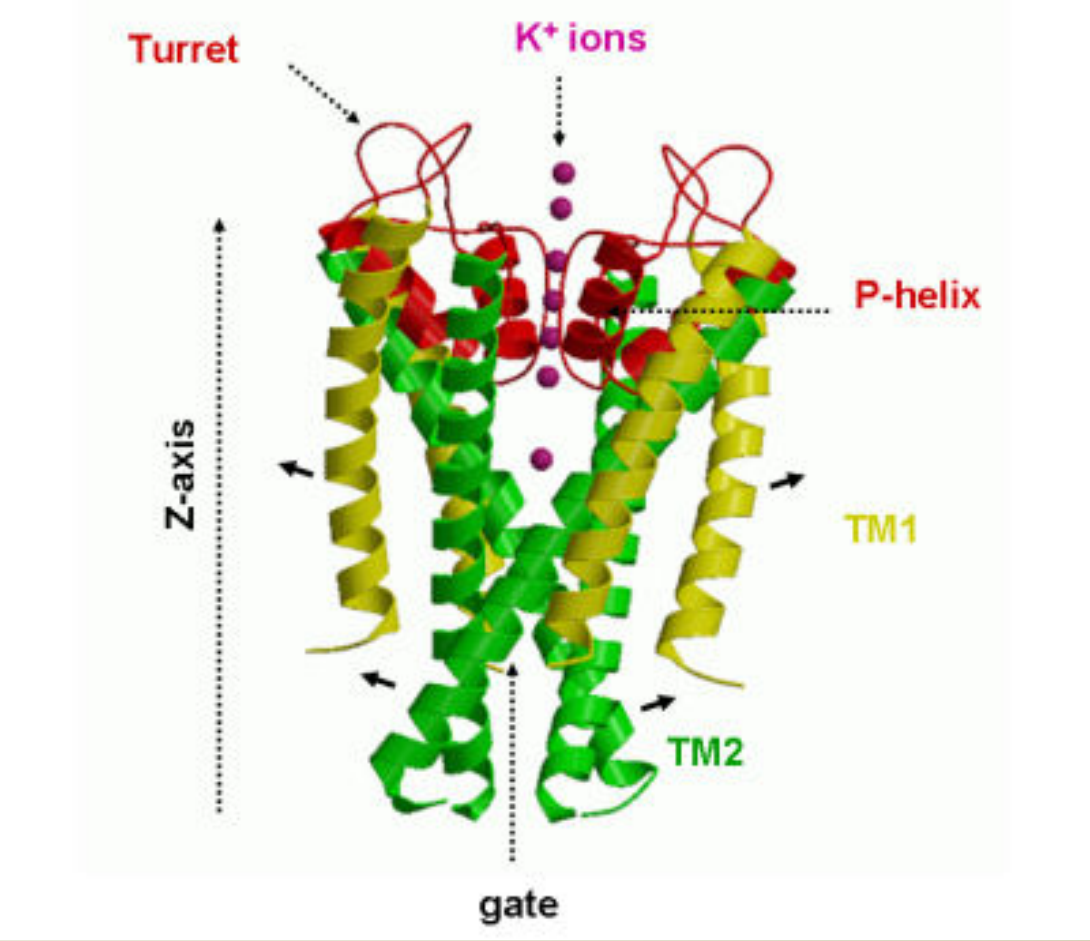
The function of a protein depends on its physical shape



http://asteris.cs.gsu.edu/~weber/GK-glc-atp03.gif

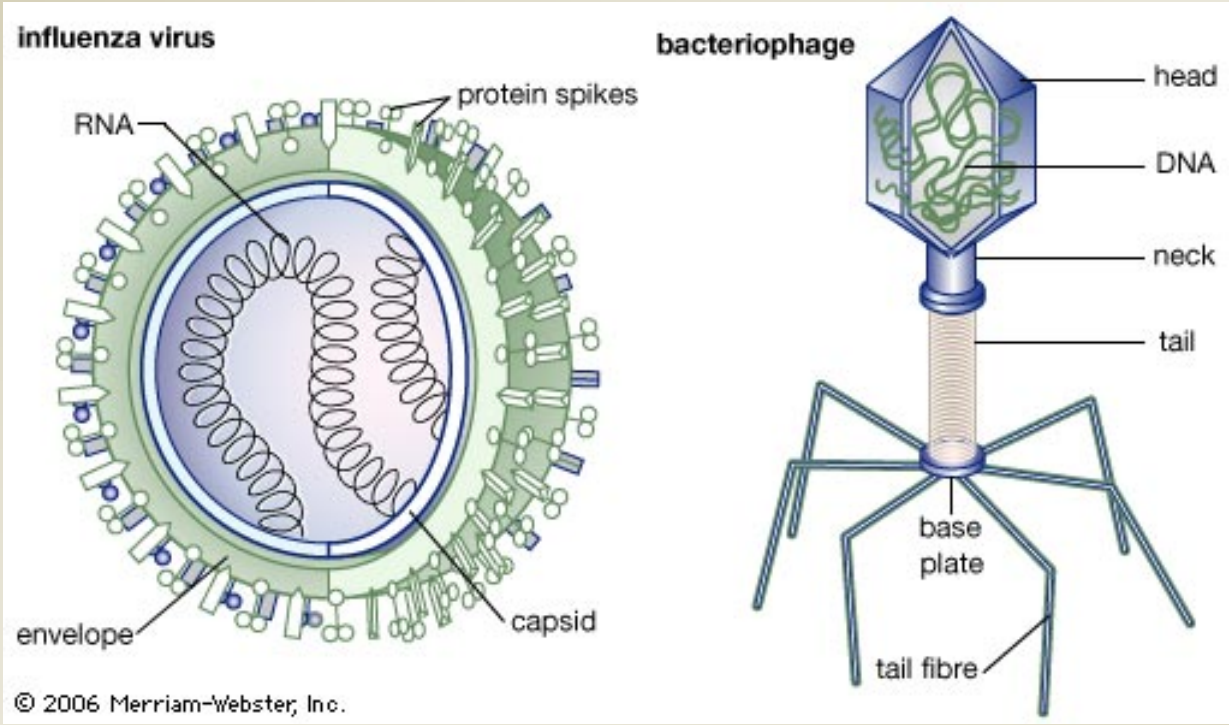
Structure of glucokinase when binding ATP and glucose
Glucokinase is the glucose sensing enzyme in the liver and the pancreas

Membrance proteins control ion channels



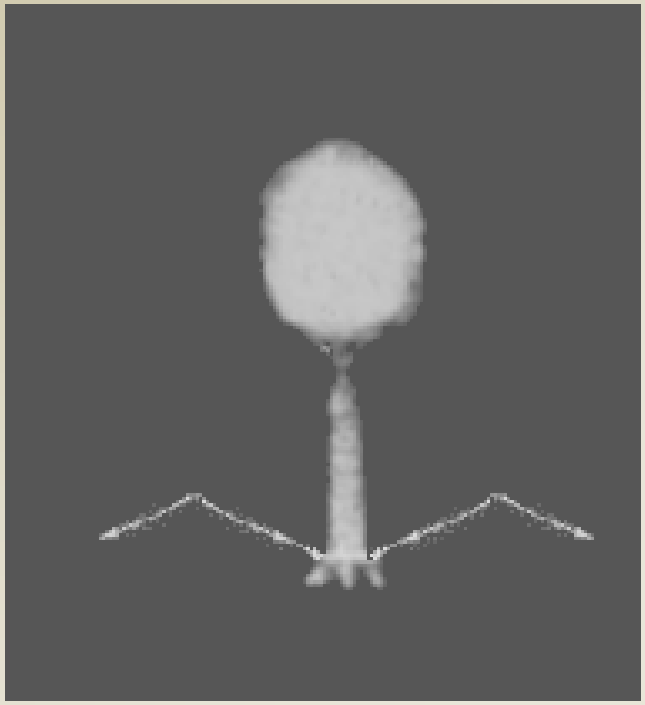
http://www.ccbb.pitt.edu/research/bahar_lab/Common_Mechanism_Pore_Opening/

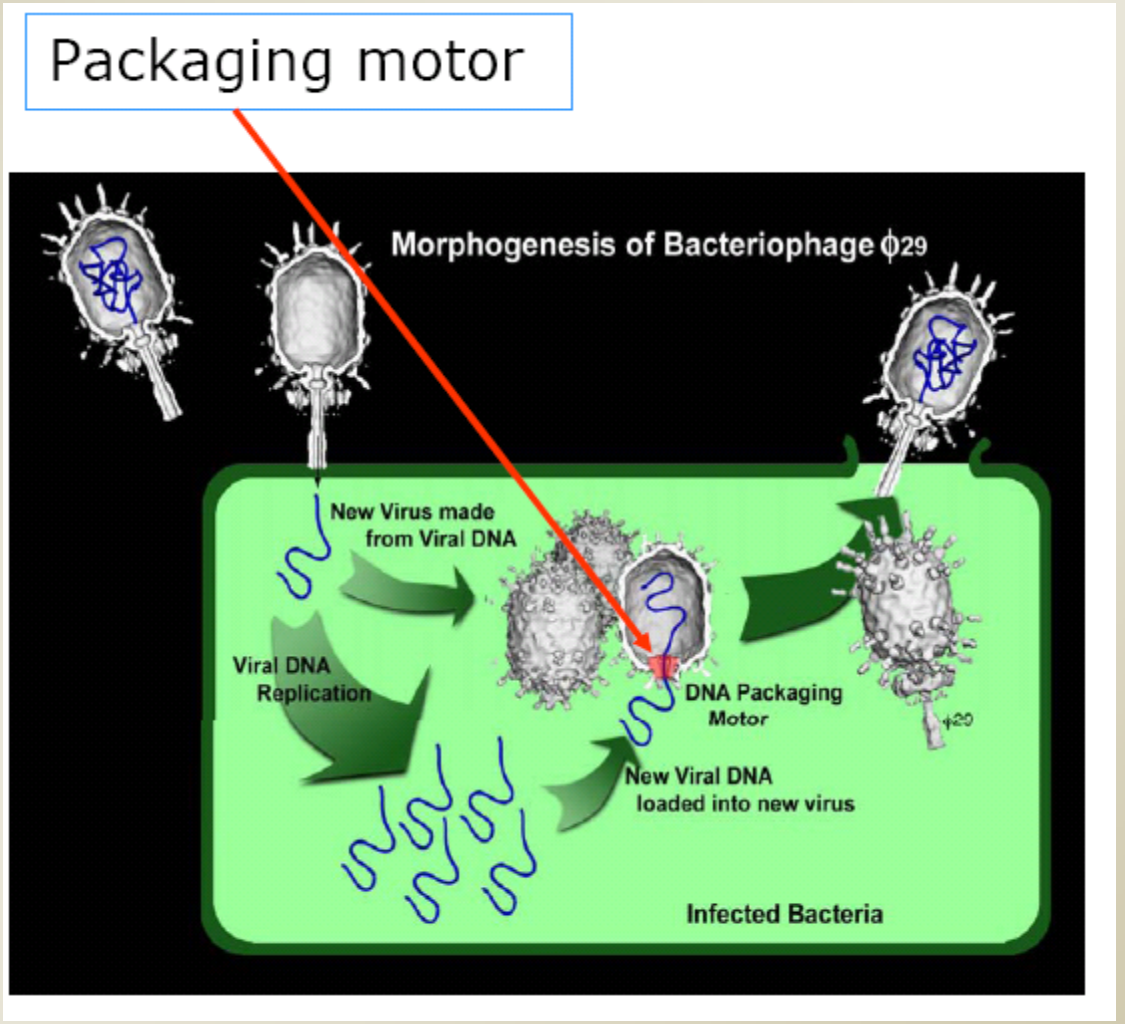
Proteins also help virus to function



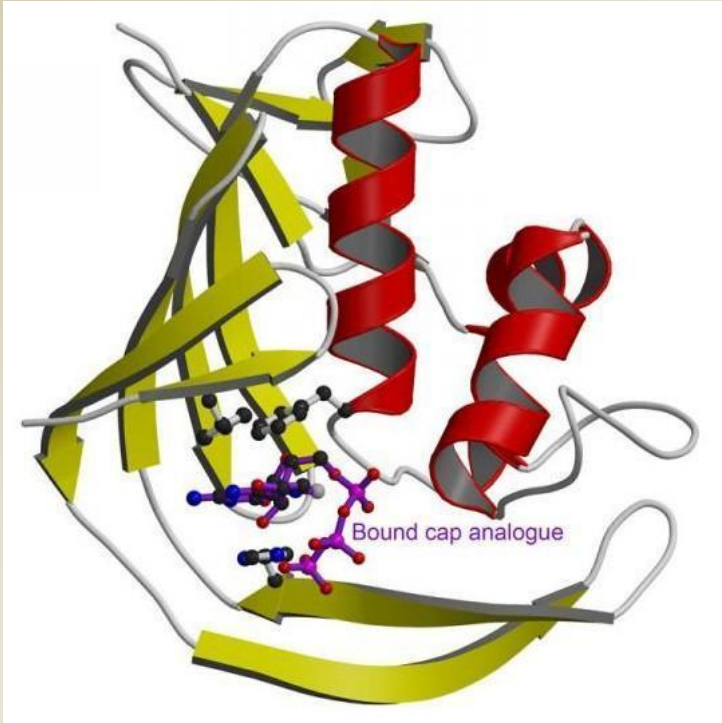
The influenza virus possesses both a protein shell (capsid) and a lipid and protein envelope. The protein spikes of the envelope facilitate adherence and entry into the host cell. The capsid proteins determine the influenza virus type (A, B, C), and the highly variable proteins of the spikes and envelope determine the different strains within each type. The bacteriophage (bacterial virus, 噬菌體) shown here has a head shaped like an icosahedron (with 20 sides).

http://student.britannica.com/comptons/art/print?id=66138&articleTypeId=0





http://www.mcb.uct.ac.za/tutorial/virusentbacteria.htm



Upon infection the influenza virus starts multiplying in the cells of its host.
One protein that is crucial in this process is the viral polymerase –
the enzyme that copies its genetic material and helps to produce more viruses.
One component of the polymerase, called PB2, plays a key role in stealing an important tag
from host cell RNA molecules to direct the protein production machinery
towards the synthesis of viral proteins.

The virus steals a 'cap' molecule from its host to take over the protein production machinery and multiply. PB2 binds the cap (purple) by sandwiching it between aromatic amino acids.

http://www.sciencedaily.com/releases/2008/05/080504153820.htm

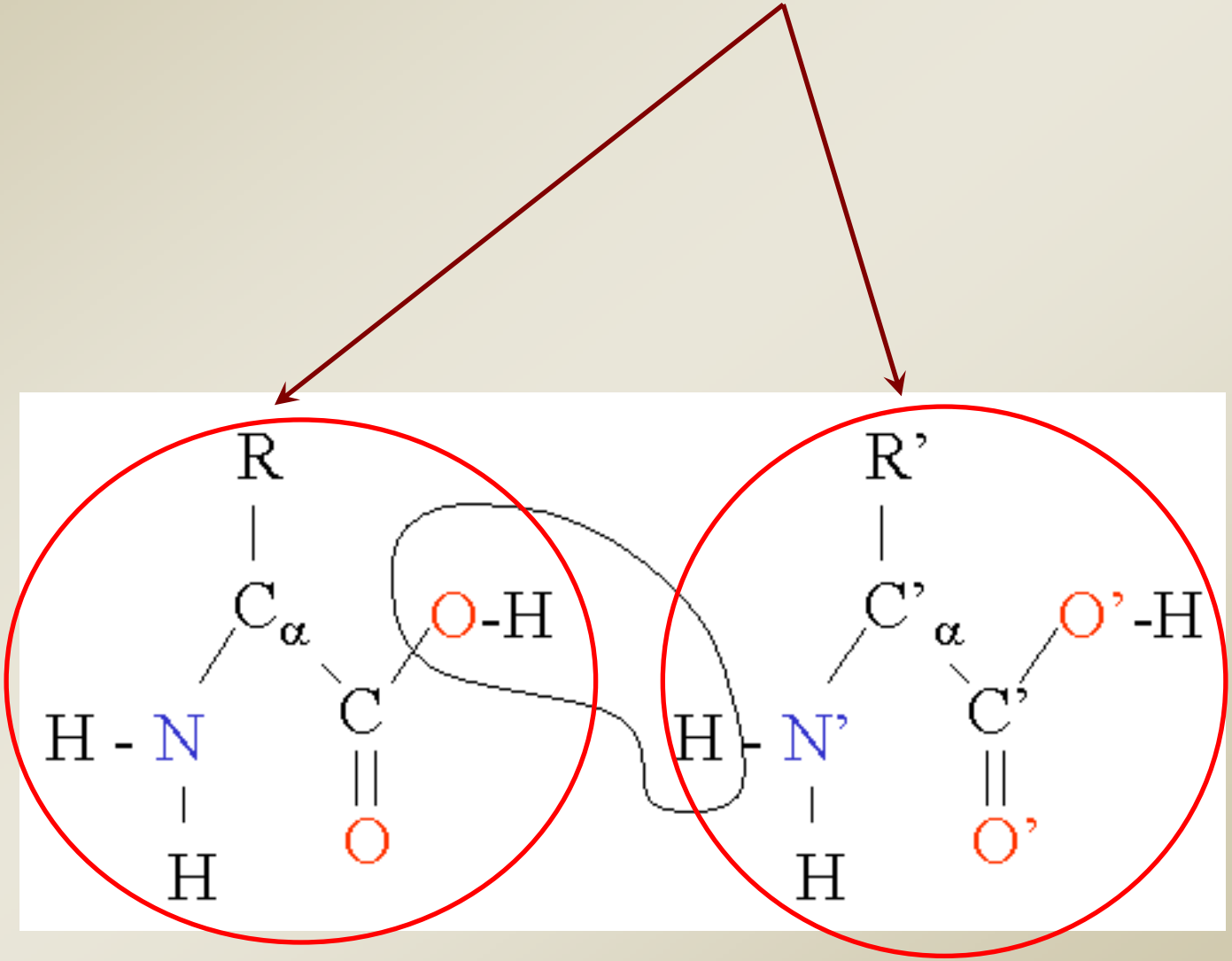
What is the challenge?

蛋白質是典型的 "Biological Matter"
與典型的Solid-State 有很大的差異
--沒有規則的幾何形狀

如果知道如何設計適當形狀及化學性質的蛋白質

⇒ 可在理論上直接設計新藥
(Theoretical Drug Design)

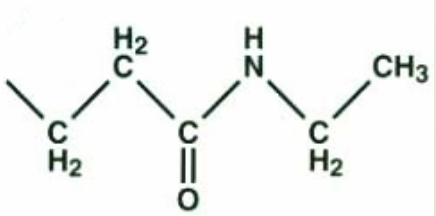
蛋白質由氨基酸組成

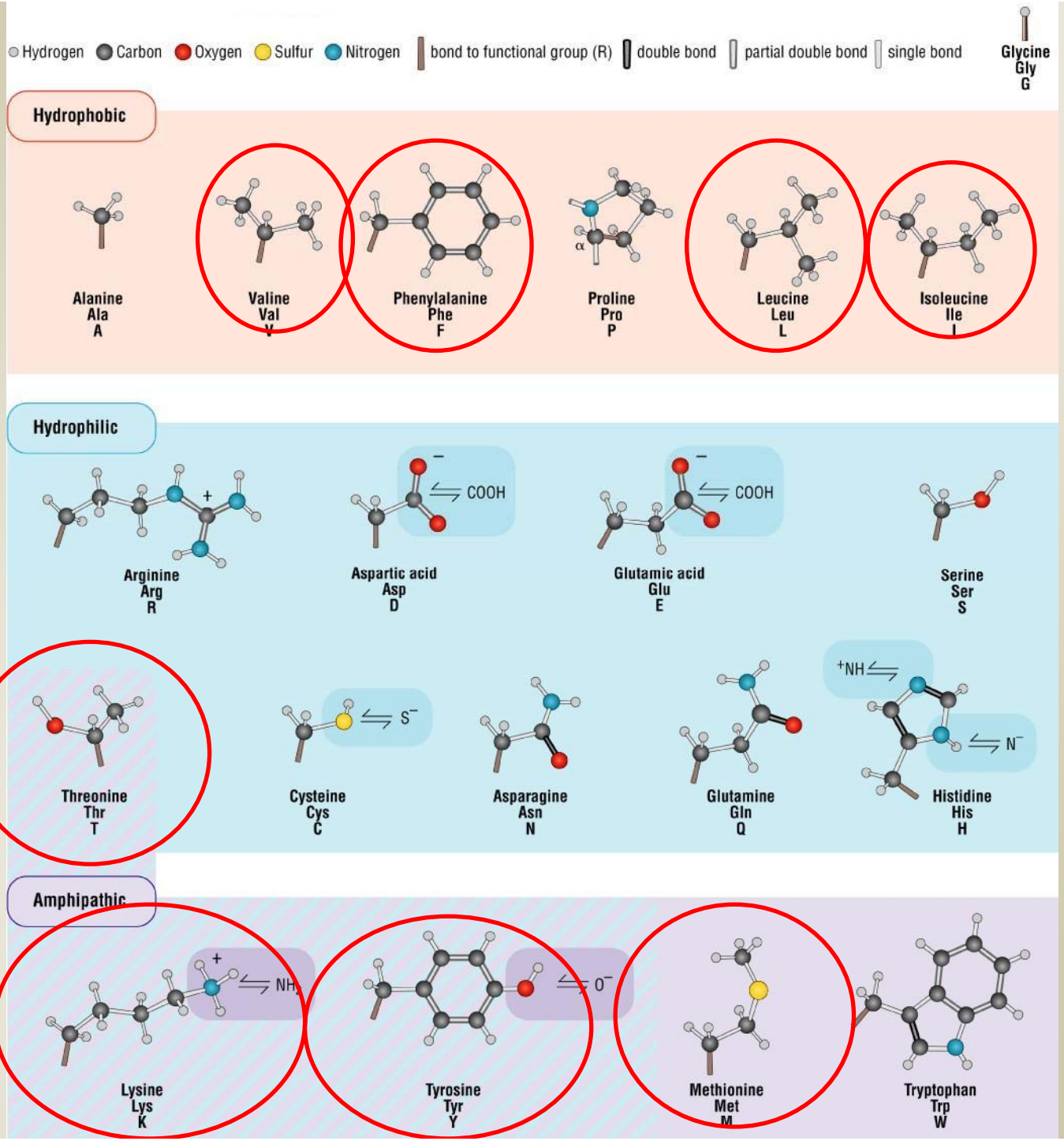


人體中有20種氨基酸

並不是都能自行合成

此外也有一些其他的氨基酸：茶氨酸(theanine)爲Glutamine變形





蛋白質的四級結構：

一級結構 – amino acid sequence

Arg–Pro–Tyr–His–Cys–Ser–Tyr–Cys–Asn–Phe–Ser–
Phe–Lys–Thr–Lys–Gly–Asn–Leu–Thr–Lys–His–Met–
Lys–Ser–Lys–Ala–His–Ser–Lys–Lys

↓

二 級 結構 – local structures, α helices

or β sheets

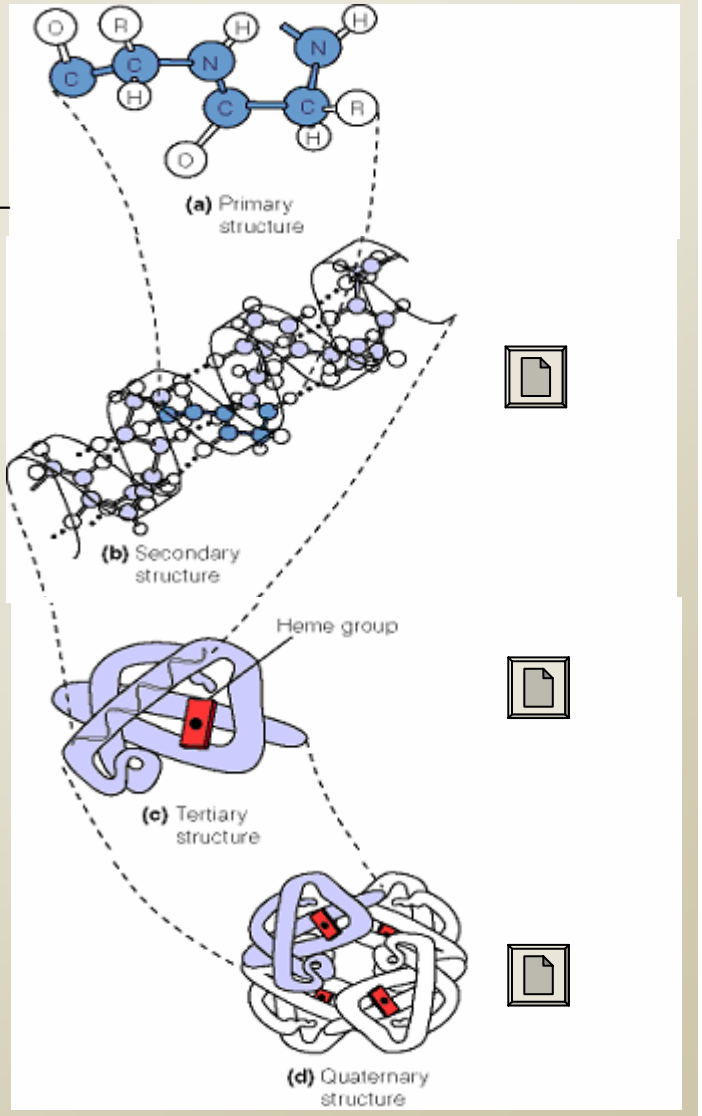
↓

三級結構 – overall 3-D structure

↓

四級結構 – subunit organization

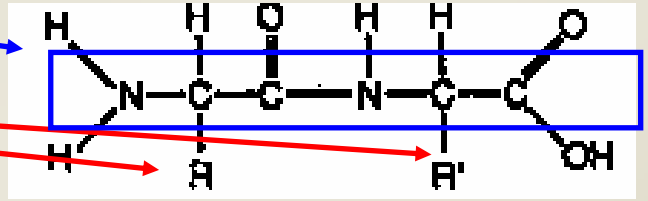
protein-protein interactions



Interesting physics lies in how these ordered structures form for a given sequence

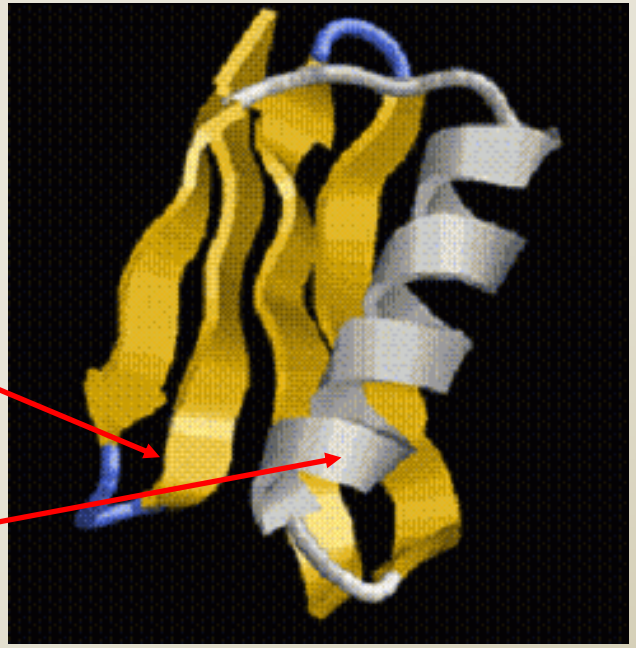
backbone

Side chains



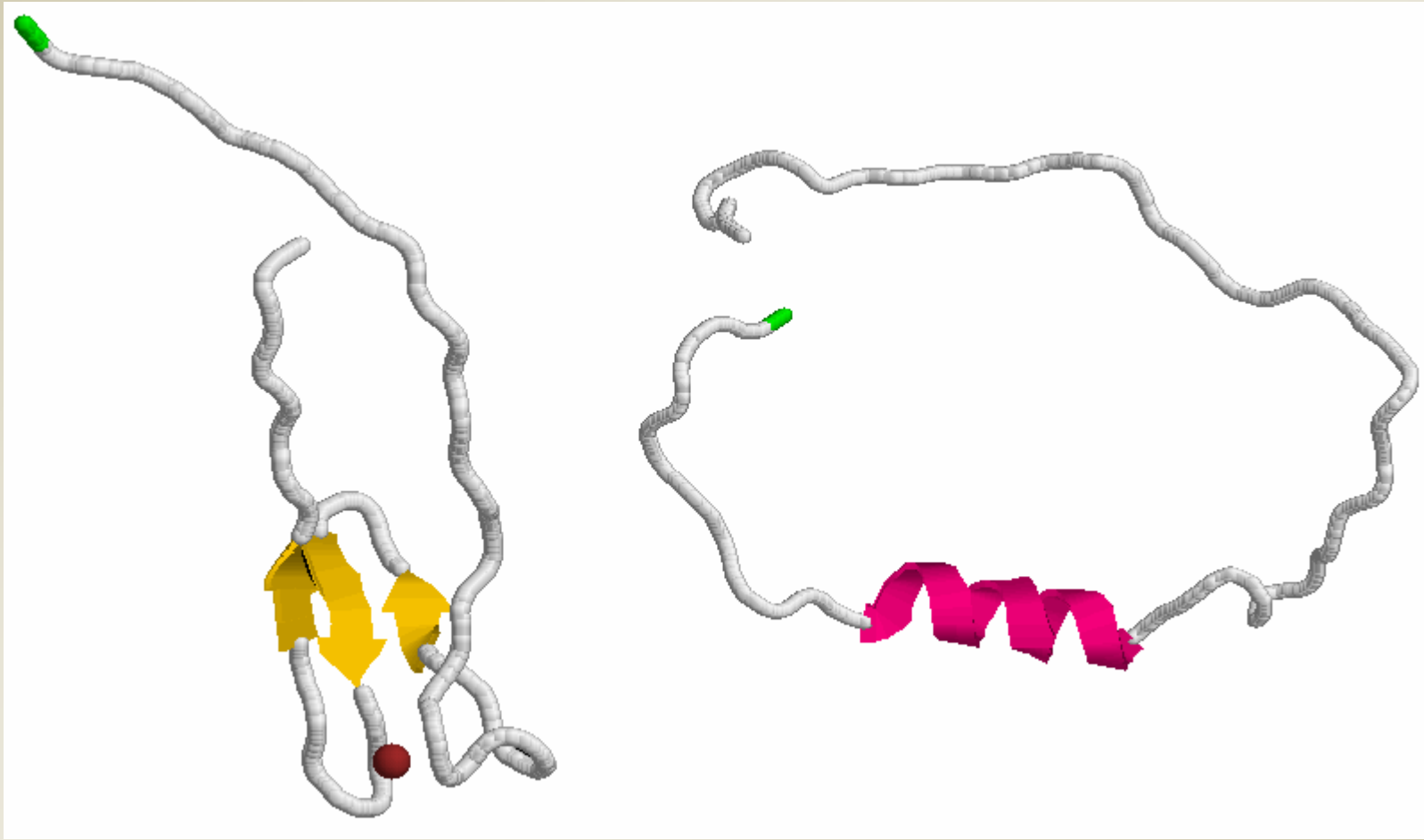
β

α



Native structure

Disorder proteins also exist



2K4X

2K5K

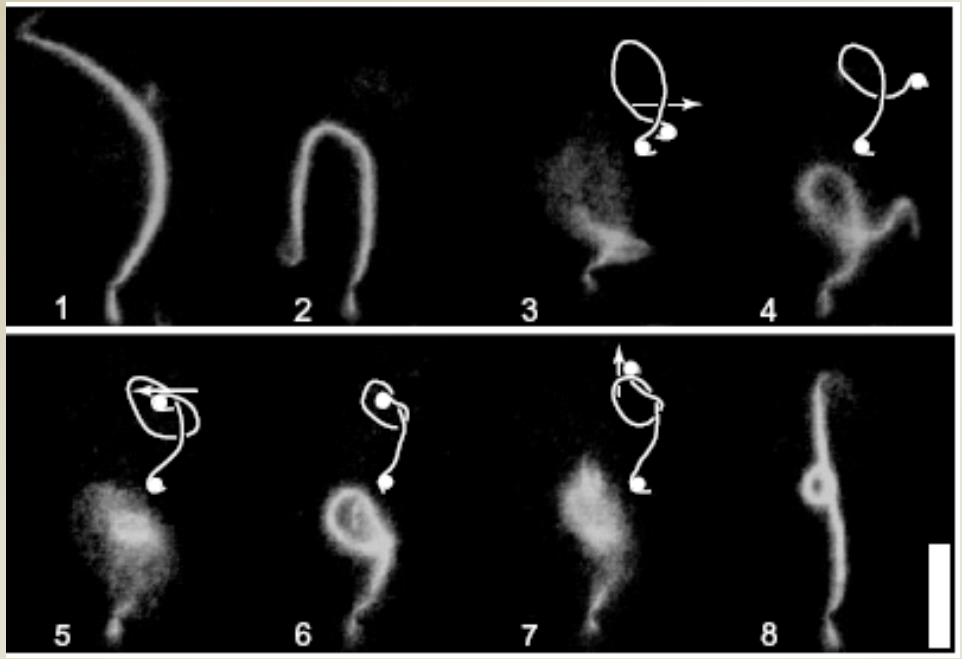
蛋白質折疊問題

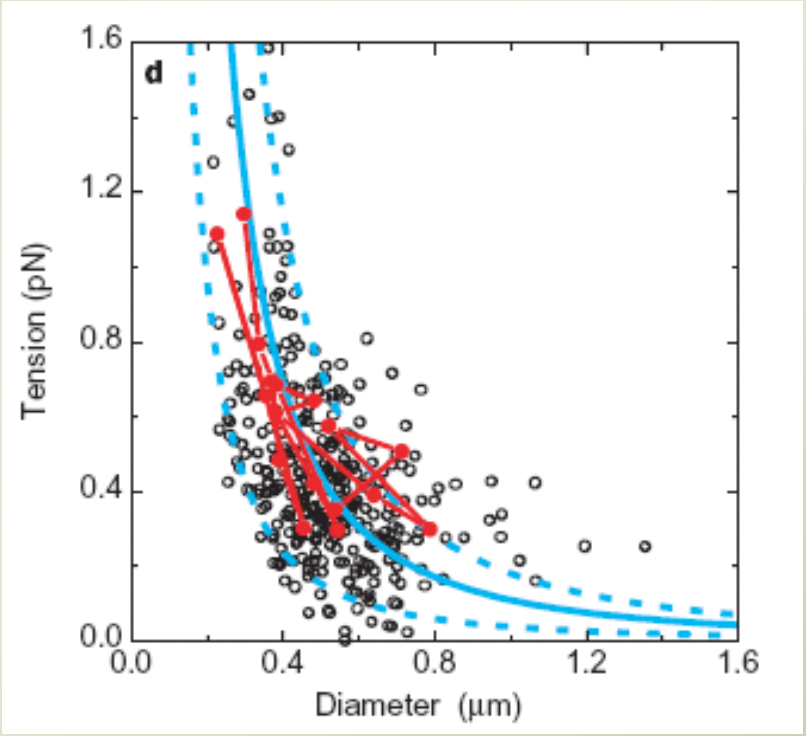
Two important issues from physics point of view

1. What and how to determine the 3D structure of a protein?
(protein structure prediction)

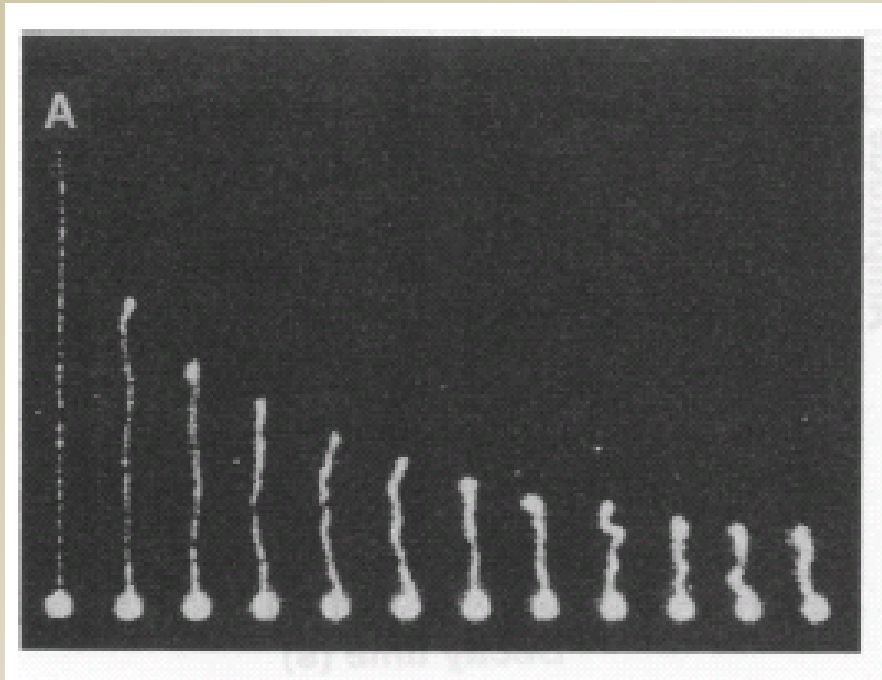
2. How could proteins find their shapes in such short time?
(Dynamical Nature)

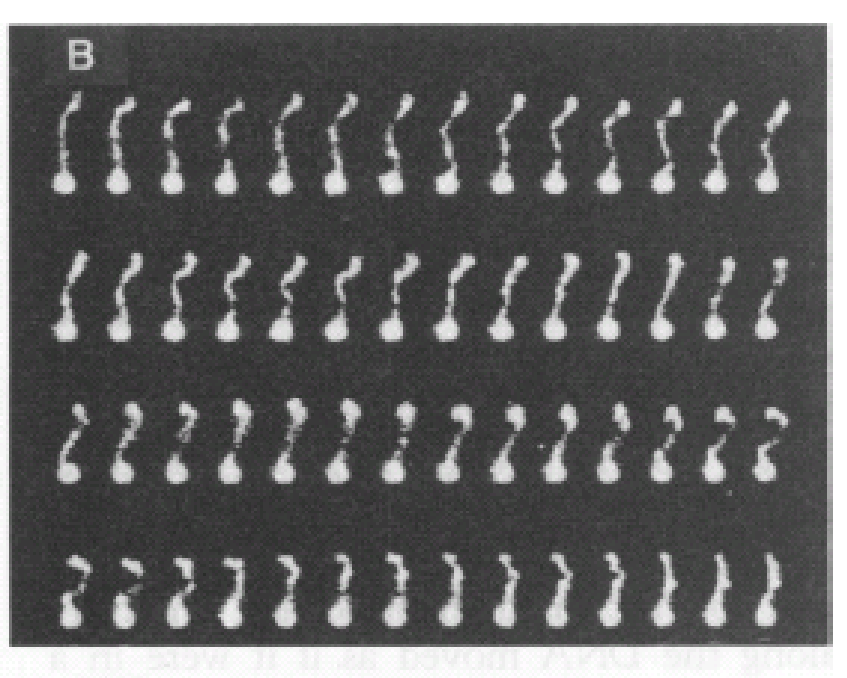
Modern experimental tools: Single-molecule physiology is now possible by using optical tweezers





Tying a knot in the protein actin. By using optical tweezers to manipulate the ends of the molecule, the chain is curled into a hoop and threaded on itself.
Y. Arai et al., Nature 399, 446, 1999



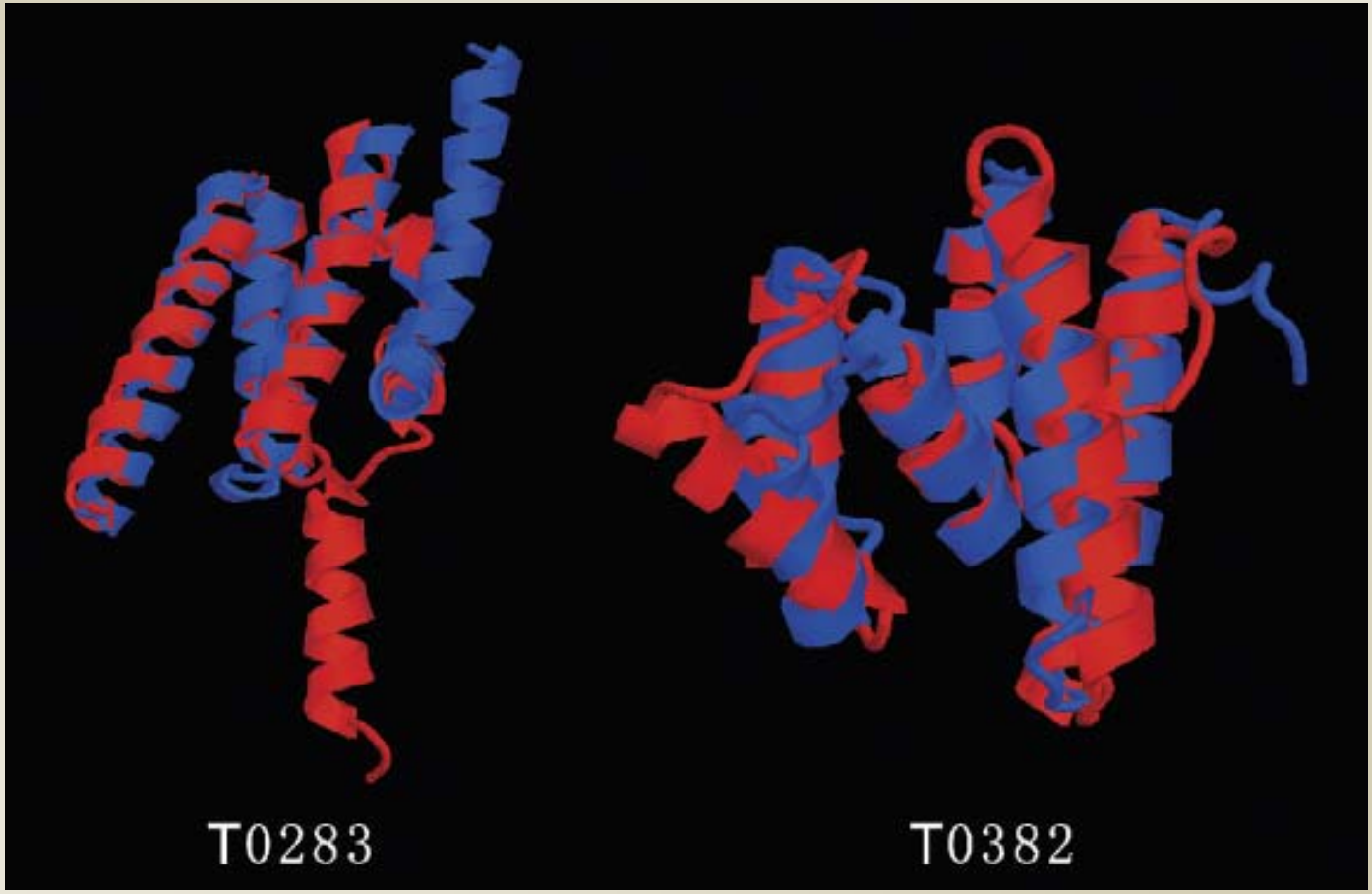


DNA folding: Perkins et al. Science 264, 822 (1994)

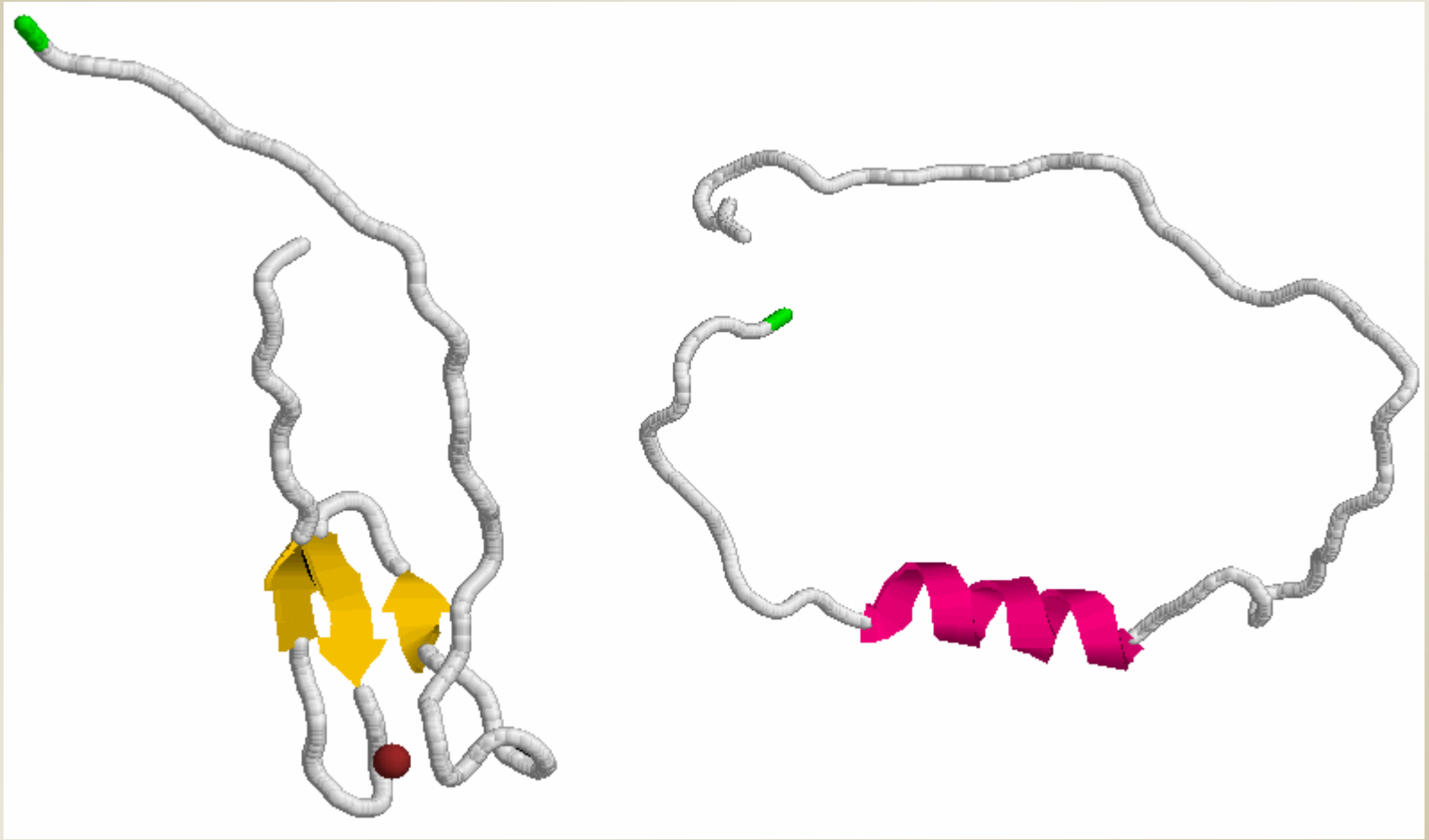
But only the protein structure prediction seems to be relevant to biologists.

Two approaches for structure prediction:
Template based vs Template free

Template based approach (based on database search)



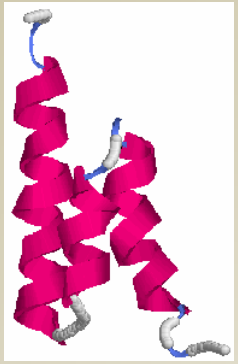
Have difficulty in treating disorder proteins

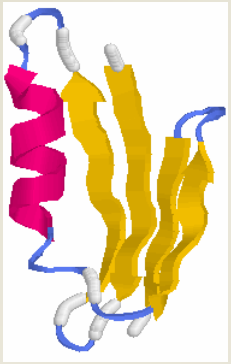


2K4X

2K5K

Other difficulties





```
-------HHHHHHHHHHHHHHHHHH-HHHHHHHHHH--HHHHHHHHHHHHHHHHH----
MEAVDANSLAQAKEAAIKELKQYGIGDYYIKLINNAKTVEGVESLKNEILKALPTE
Ga = 2FS1
```

94.64% ↓

16.07% ↓

```
-------------------------**-------------*------------
MEAVDANSLAQAKEAAIKELKQYGIGEKYIKLINNAKTVEGVWSLKNEILKALPTE
Ga30
```

66.07% ↓

30.36% ↓

```
********-*----*-----***--*-------*---------*--*-----**--
TTYKLILNLKQAKEEAIKELVDAGIAEKYIKLIANAKTVEGVWTLKDEILKATVTE
Ga77
```

96.43% ↓

76.79% ↓

```
-------------------------------------------------**----
TTYKLILNLKQAKEEAIKELVDAGIAEKYIKLIANAKTVEGVWTLKDEILTFTVTE
Ga88 = 2JWS
```

96.43% ↓

87.50% ↓

```
-----------------------*-----------------------*------
TTYKLILNLKQAKEEAIKELVDAGTAEKYIKLIANAKTVEGVWTLKDEIKTFTVTE
Ga95
```

94.64%

```
EEEEEEEEE--EEEEEEEEE--HHHHHHHHHHHHHH-----EEEEE---EEEEEEE
MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE
Gb = 1PGA
```

91.07% ↓

```
----------------------------*--*----**-*----------------
MTYKLILNGKTLKGETTTEAVDAATAEKYFKLYANDKTVEGEWTYDDATKTFTVTE
Gb30
```

85.71% ↓

```
*---------*--*-**------------------*-----*---*-*--------
TTYKLILNGKQLKEEAITEAVDAATAEKYFKLYANAKTVEGVWTYKDETKTFTVTE
Gb77
```

92.86% ↓

```
--------*--*-----*-*-------------------------------------
TTYKLILNLKQAKEEAIKELVDAATAEKYFKLYANAKTVEGVWTYKDETKTFTVTE
Gb88 = 2JWU
```

92.86% ↓

```
-------------------*---*--------*---------------*-------
TTYKLILNLKQAKEEAIKEAVDAGTAEKYFKLIANAKTVEGVWTYKDEIKTFTVTE
Gb95
```

Template free approach

Hints from experiments:

Anfinsen (Science,181, 223-224, 1973)

The native structure

=> minimum of free energy

$$e^{-\beta F} \propto \iiint dr_1 dr_2 \ldots dr_n e^{-\beta V(r_1, r_2, r_3, \ldots)}$$

Two technical issues

* **What is V?**

* **How to find the global minimum of free energy efficiently?**

Whole protein ~ milliseconds to seconds
Hinge motion in proteins ~ nanosecond
} $10^6 - 10^9$

=> Need Optimization algorithms

Various approaches

All-atom approach: molecular dynamics

remarkable results for short peptides but lots of problems

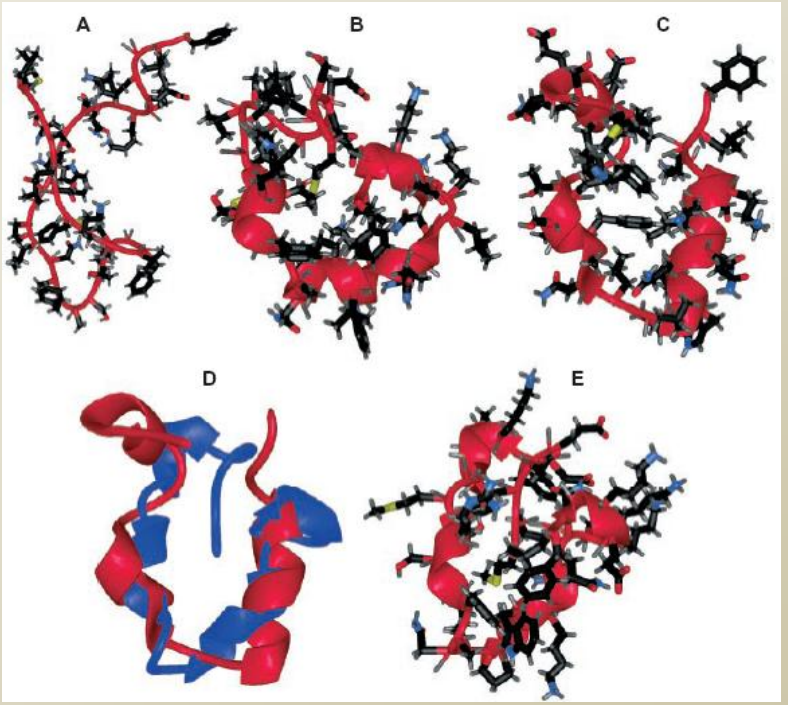
for longer proteins (time-consuming & accumulation error, not feasible)

Duan and kollman, Science 2**82**, 740, 1998

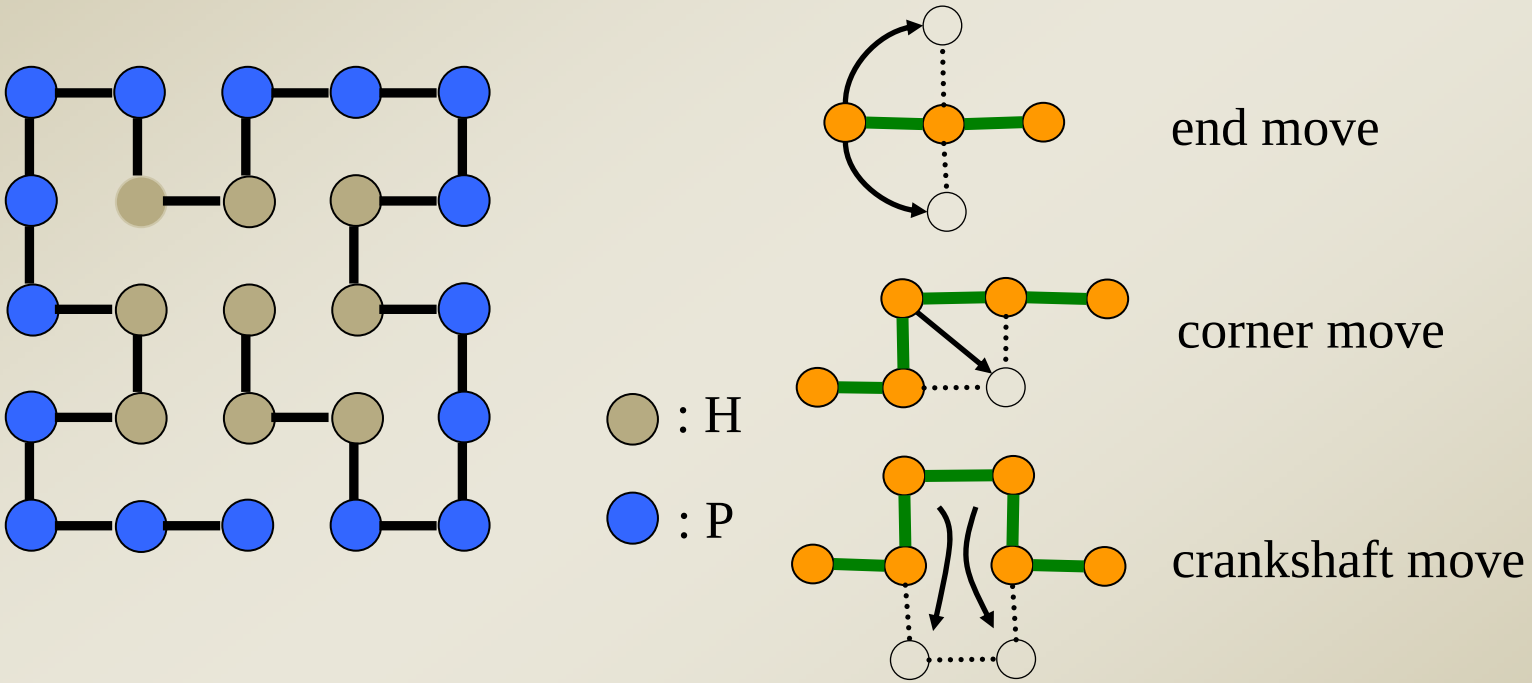
36-residue (1VII)
with 3000 water molecules

Cray3TE (hundreds of CPU in parallel)
1μsec simulation

Pathways and marginally stable states are found!



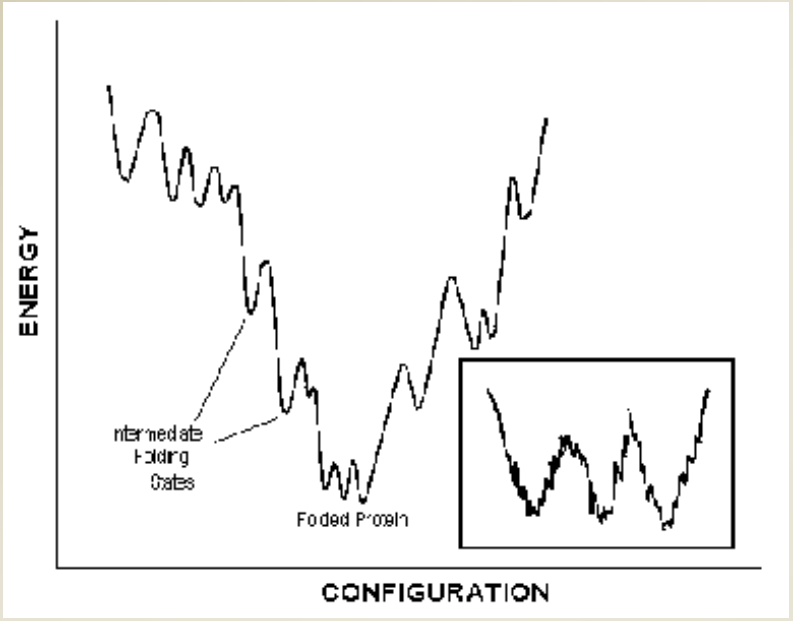
Model approaches: lattice model (oversimplified)

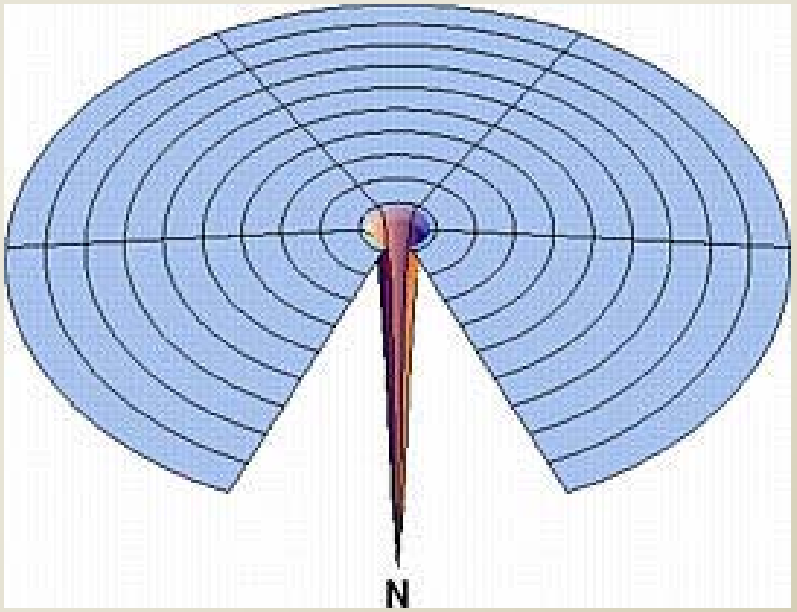


HP model: H. Li, R. Helling, C. Tang and N. Wingreen, Science **273,** 666 (1996)

Hints from Lattice model:
Real dynamics: Side chains move cooperatively

* Nontrivial energy landscape: e.g. funnel (Peter Wolynes)





Drawn by Ken Dill UC at San Francisco

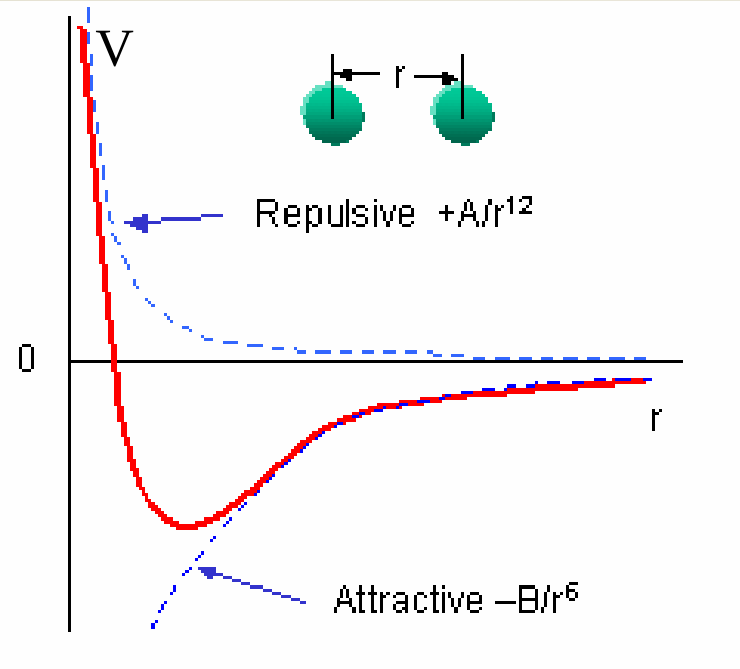
- **May exist kinetic folding pathways or intermediate states**

(Levinthal, J. Chim. Phys.,**65**, 44-45,1968)

Model approaches: based on destination state (biased)

non-bonded atomic-pair $V_{ij}(r) = \varepsilon_{ij}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$

native contact (r<6A) $\Rightarrow \varepsilon_{ij} \times 10$ (biased)

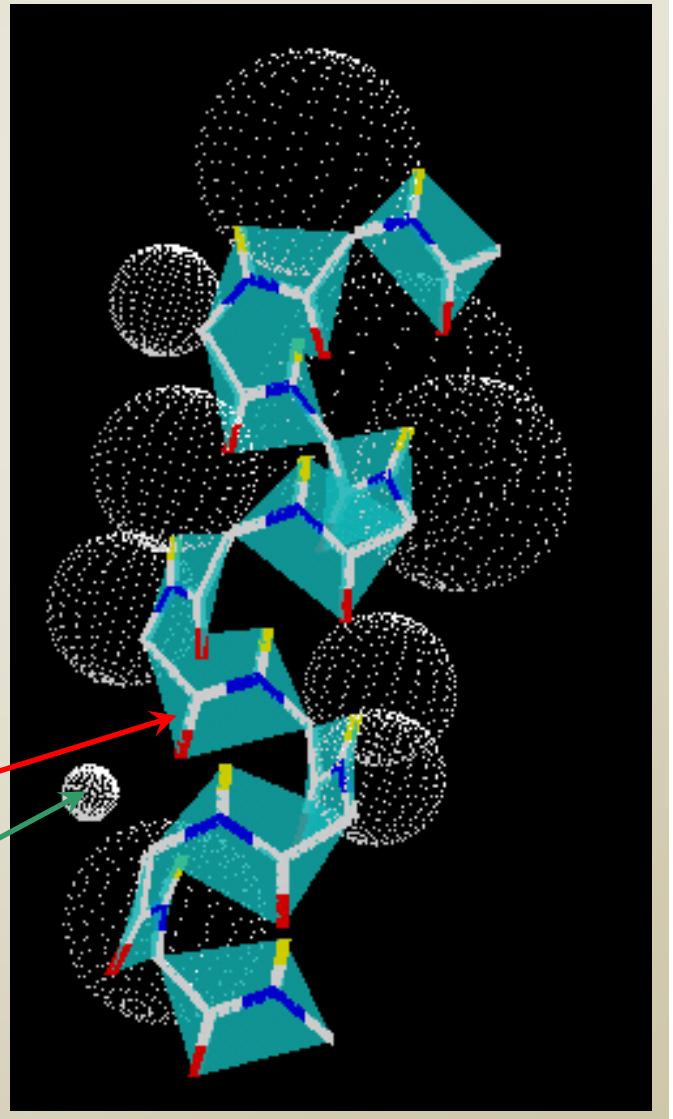


Gō-type model: A. Garg, Stanford Univ.

Coarse-grained approach

Treat back-bone and side-chain as units

$$V_{eff} = \sum_{i} V_i(\text{on-site}) + \sum_{(i,j)} V_{ij}(\text{two-body}) + \sum_{(i,j,k)} V_{ijk}(\text{3-body}) + \cdots$$

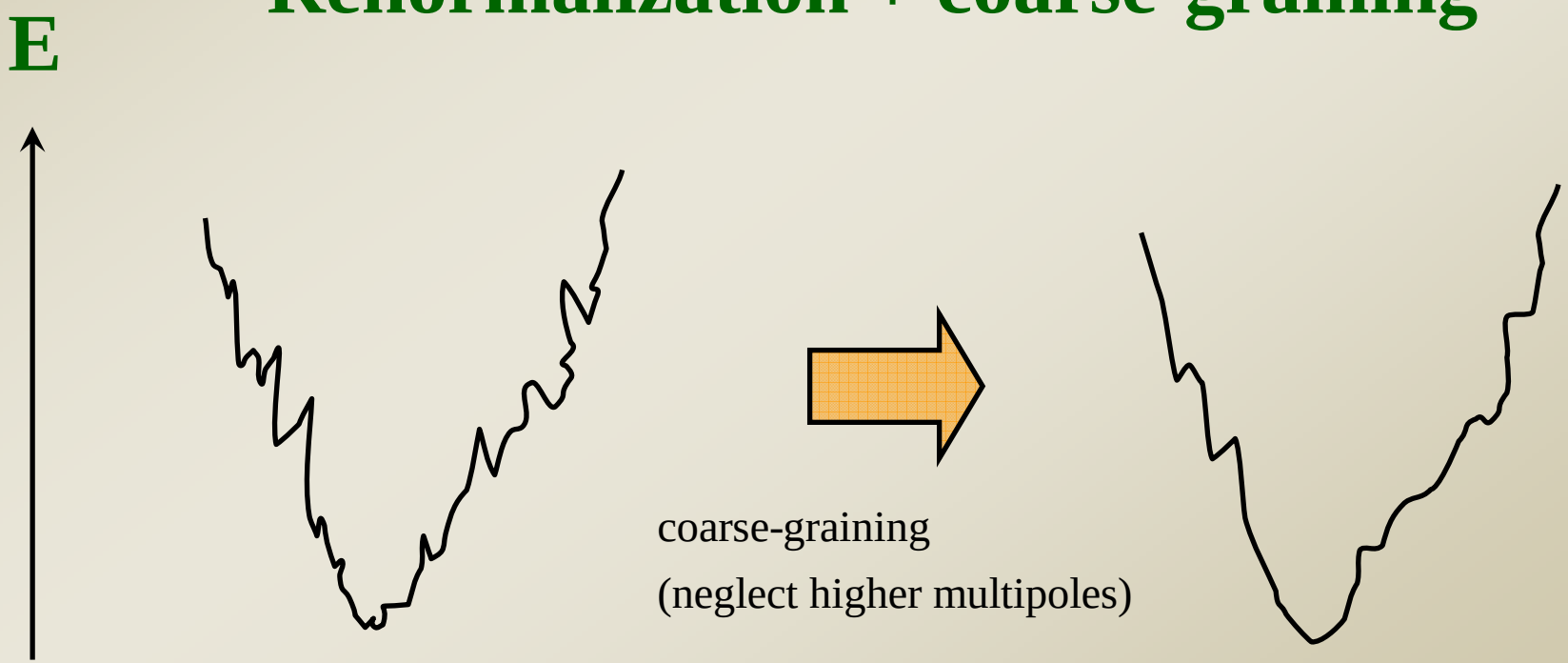


i = back-bone or side-chain

Feasible procedure

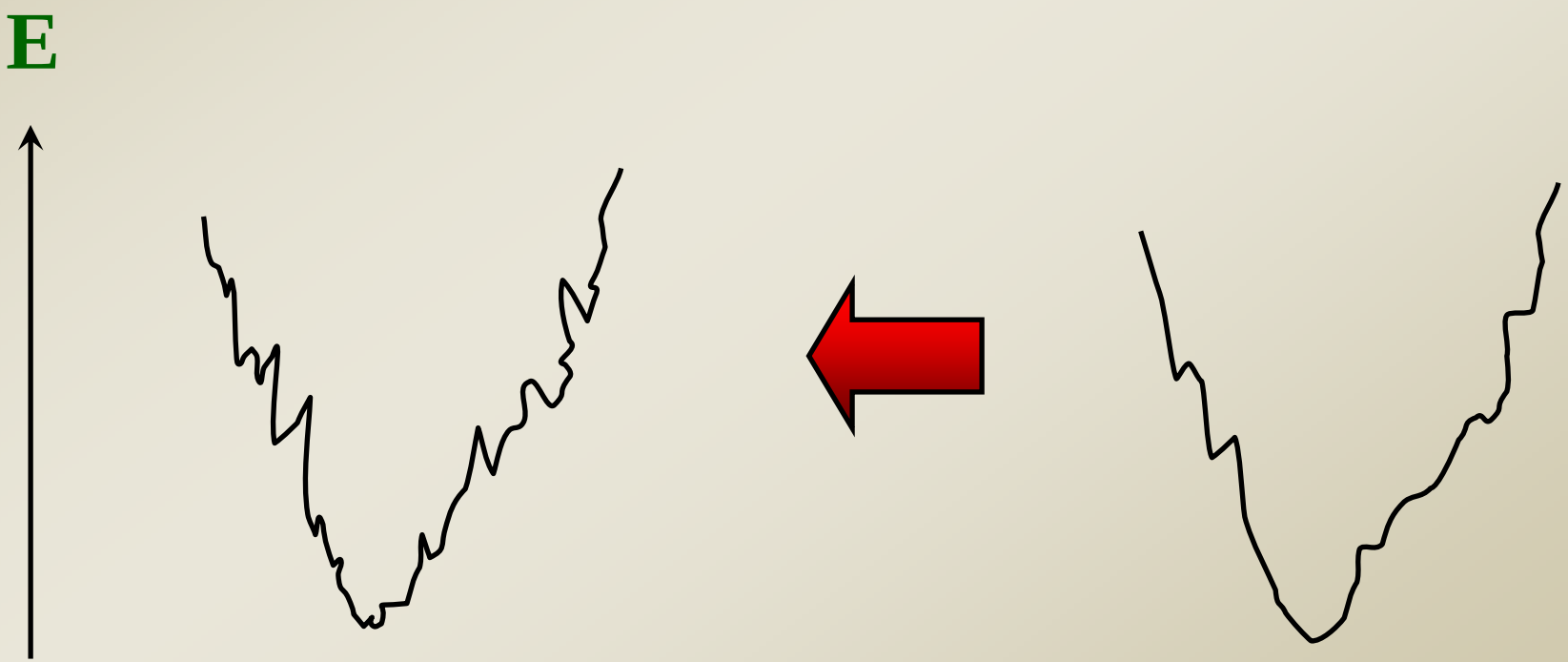
1st step: Find $V_{eff}^{R} = \sum_{i} V_{i}^{R} + \sum_{(i,j)} V_{ij}^{R}$

Renormalization + coarse-graining



2nd step

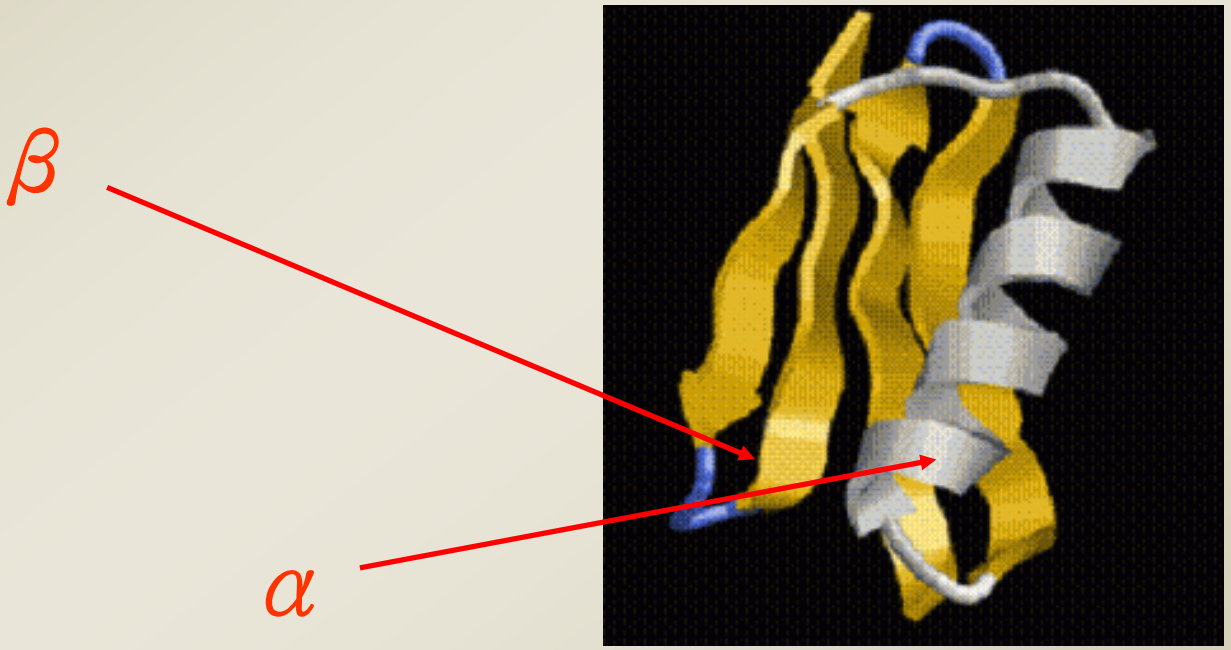
Starting from coarse-grained configuration, put back all atoms in side chains



What can physicists contribute?
--an experience from past researches

What are essential interactions (V) responsible for the folding?

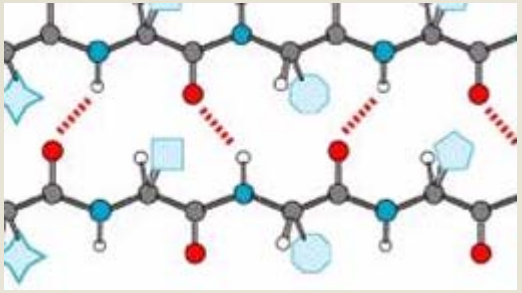
Must have built-in α & β instability



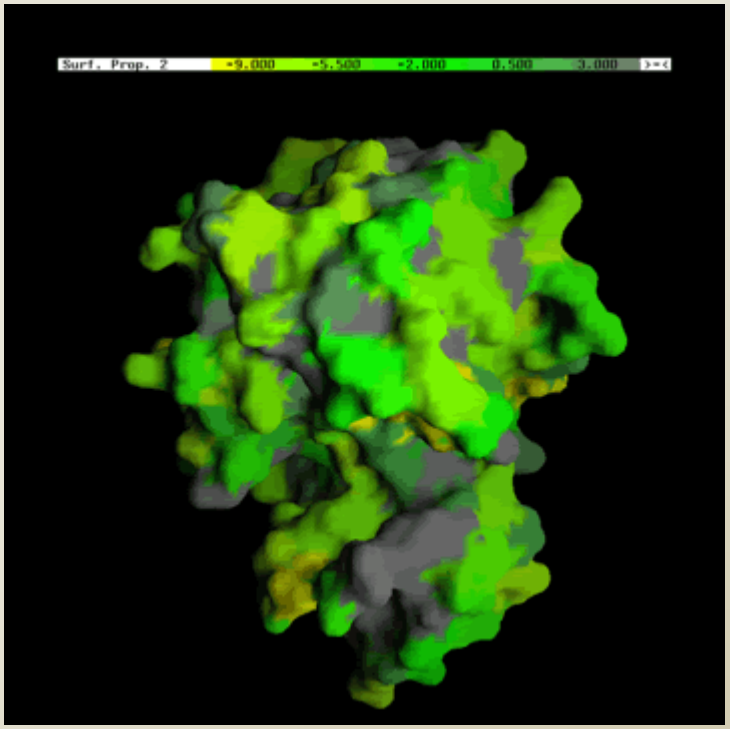
Conventional Wisdom

Hydrogen Bonding

O—H...:N
O—H...:O
N—H...:N
N—H...:O



Hydrophobicity(親、厭水性）cause long-range contact

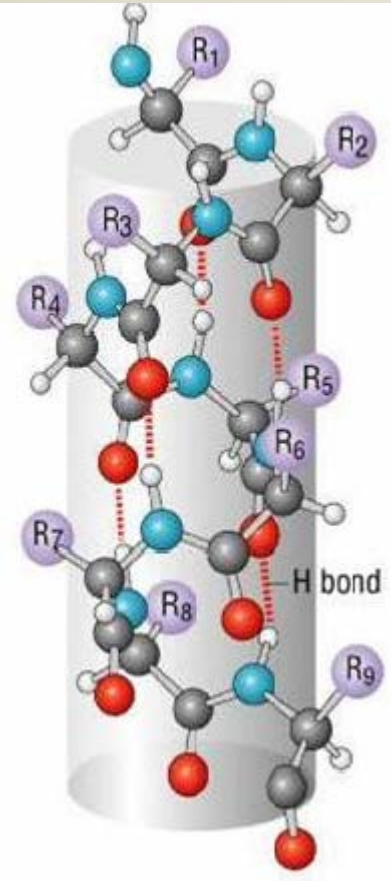


http://trantor.bioc.columbia.edu/SMS/STINGm/help/img/megahelp19.gif

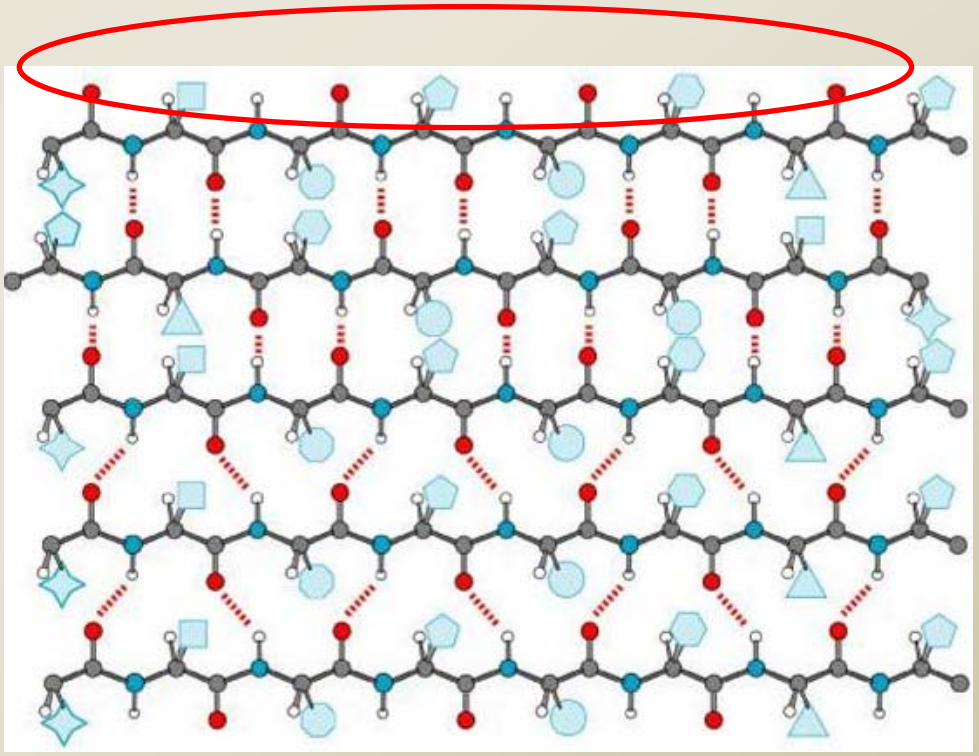
Our observations:

(1) hydrogen bonding and long-range hydrophobicity are not sufficient, lack of mechanism for stabilizing beta

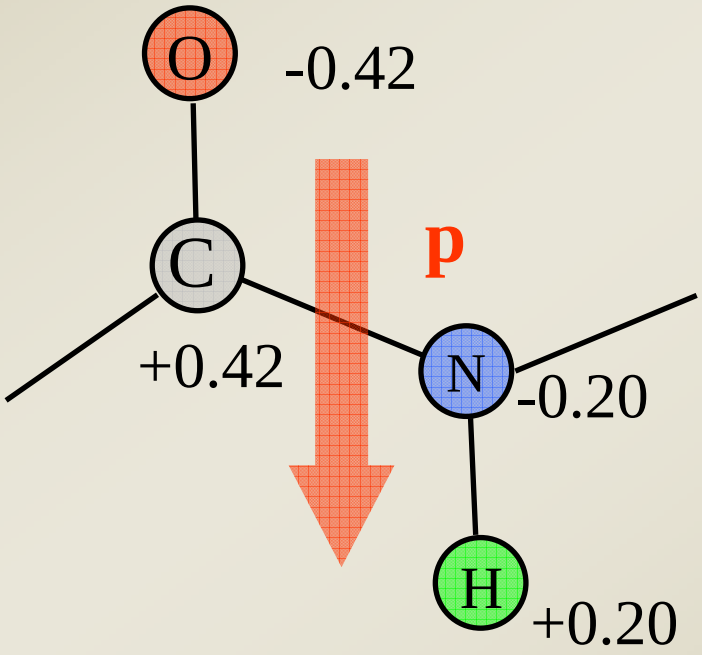
α



β



(2)There exist electric dipoles on amide planes and their configurations are strongly correlated with the structures.

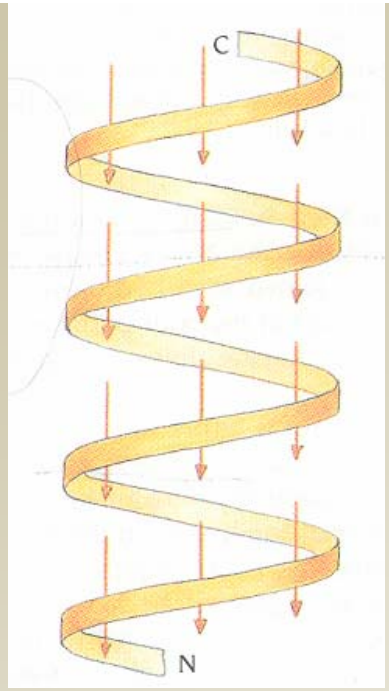


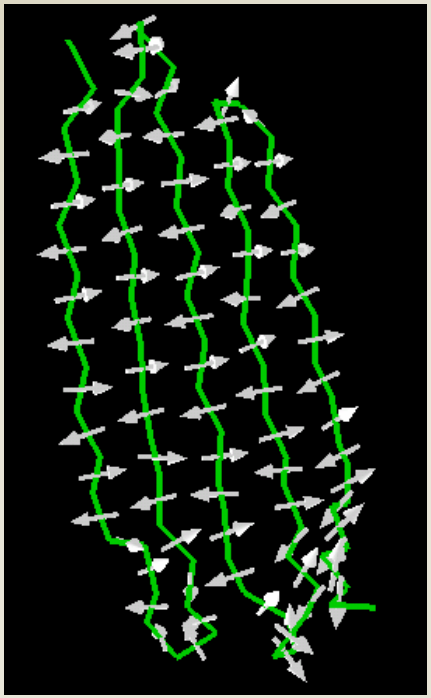
$$p = 1.15 \times 10^{-29} Cm$$

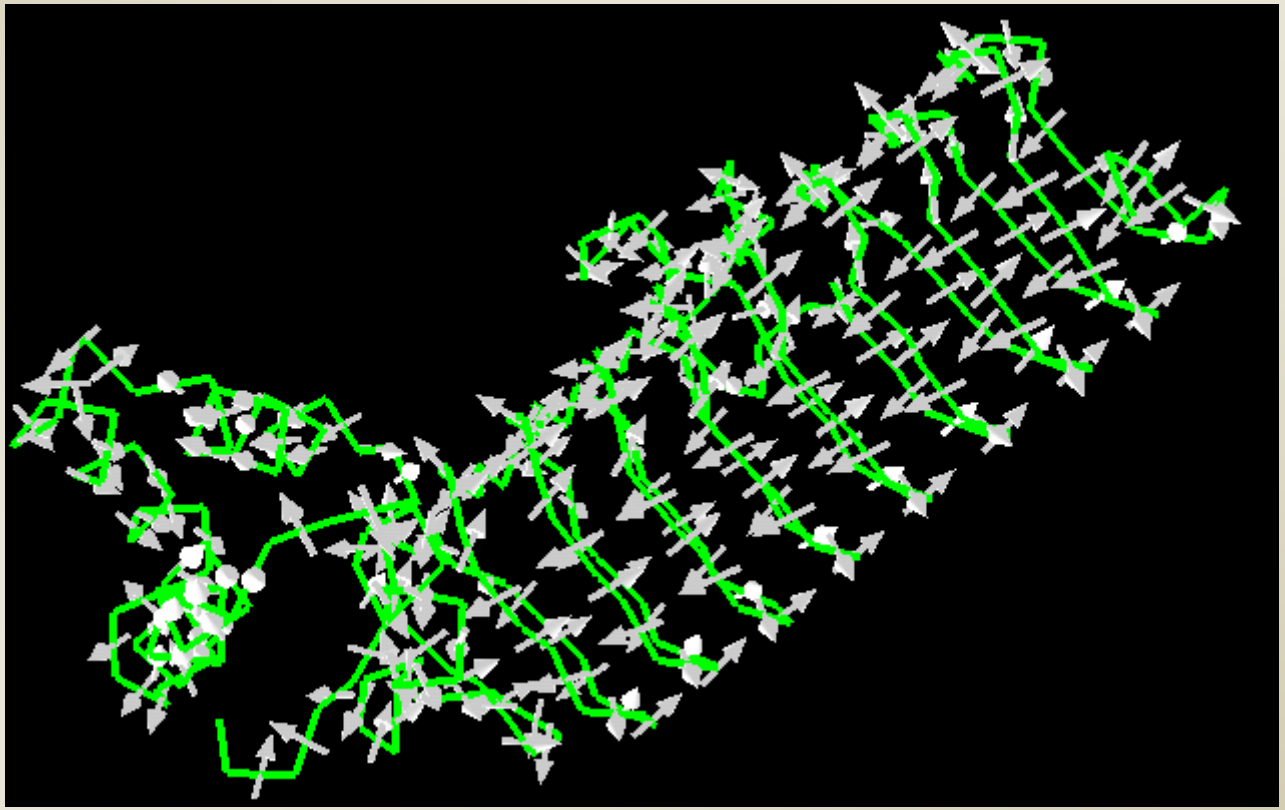
HB ~ 4.8Kcal/mol

hydrophobic ~1-3Kcal/mol

dipole ~ 2Kcal/mol


C
N





Mean-field theory:
bare V_{dipole} still favors helices

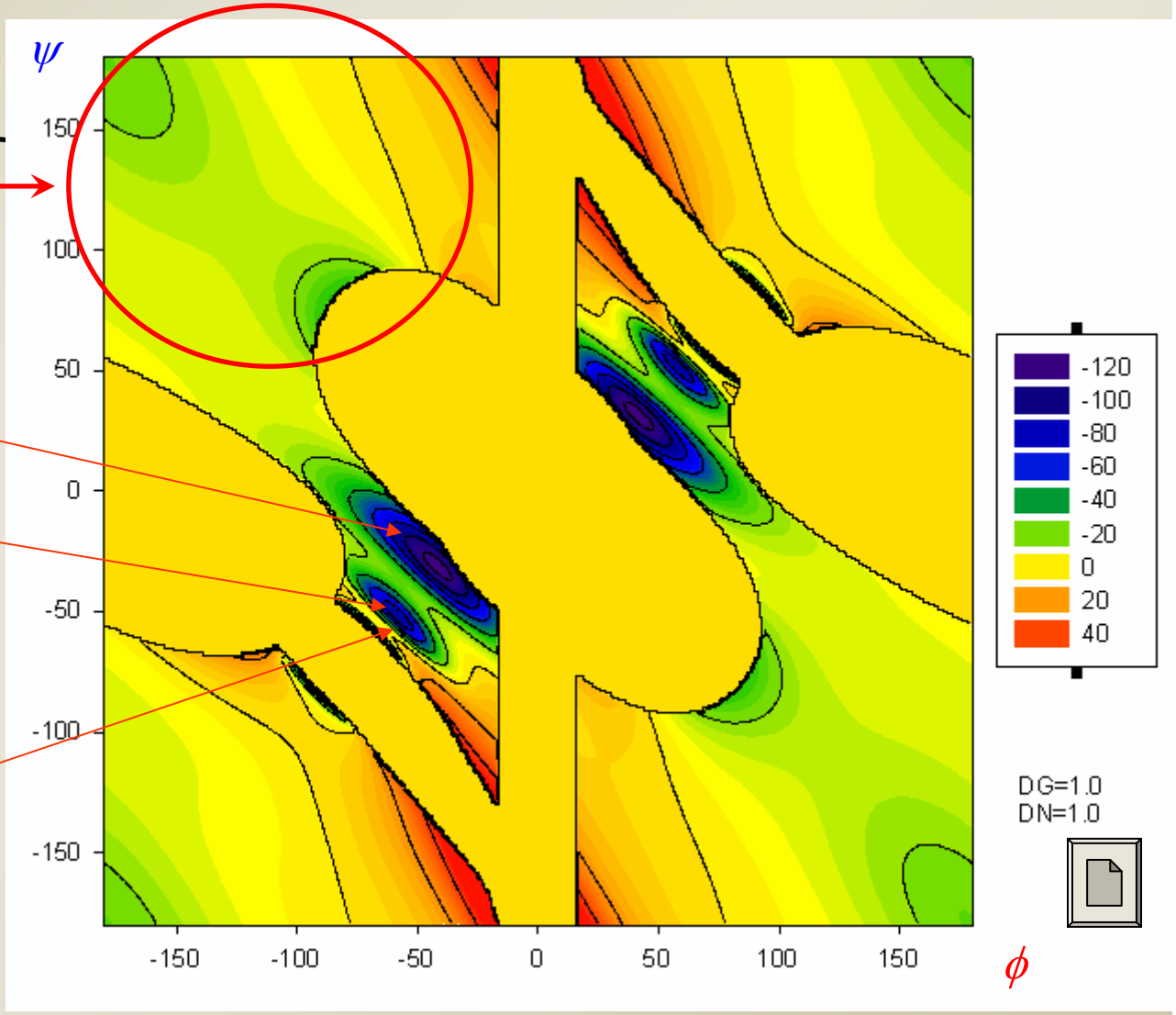
Expt data except glycine

Beta strands

3-10 Helix
ϕ = -49
ψ = -26

Standard Alpha Helix
ϕ = -57
ψ = -47

π Helix
ϕ = -57
ψ = -70

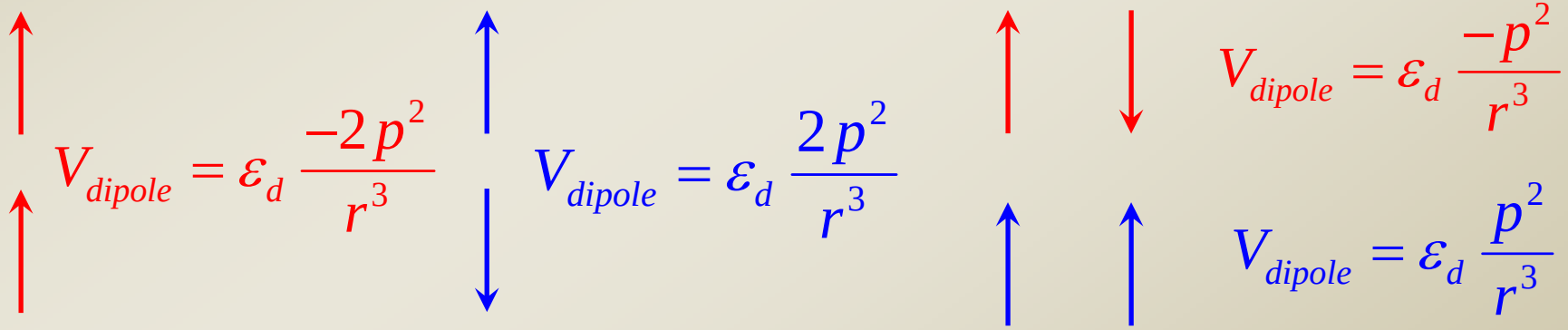


What is the rationale for the existence of beta?

Perhaps, the strong correlation between the configuration of electric dipoles and the structures is the consequence not the cause?

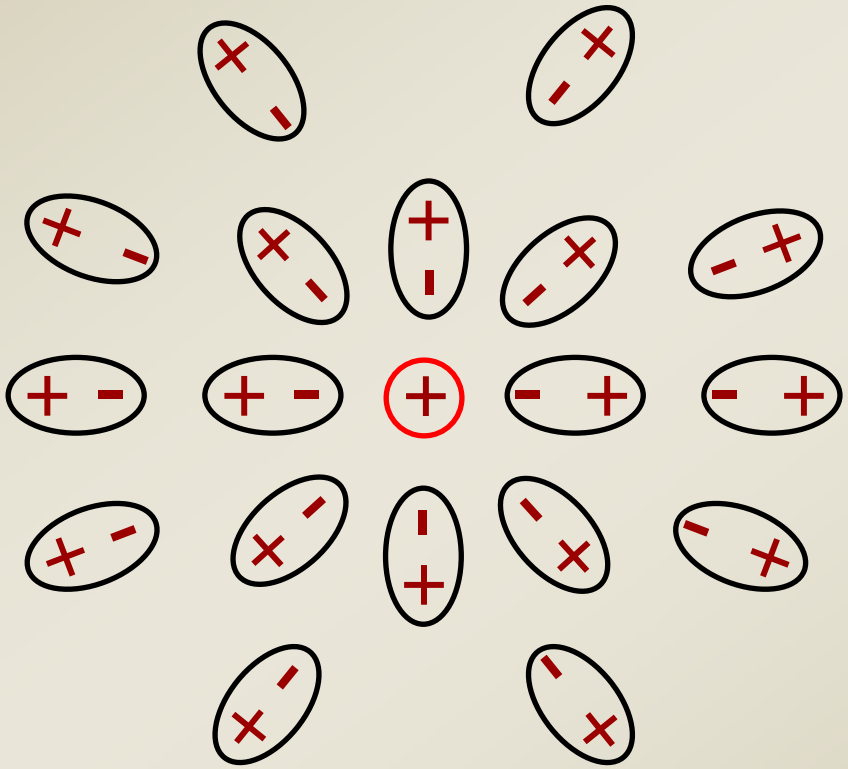
Local energetics of dipole-dipole interaction

$$V_{dipole} = \varepsilon_d \sum_{ij} \left(\frac{\mathrm{p}_i \cdot \mathrm{p}_j}{r_{ij}^3} - \frac{3(\mathrm{p}_i \cdot \mathrm{r}_{ij})(\mathrm{p}_j \cdot \mathrm{r}_{ij})}{r_{ij}^5} \right)$$



Mechanism for stabilizing β : effects of screening

$$\varepsilon_d = \frac{1}{4\pi\varepsilon}$$



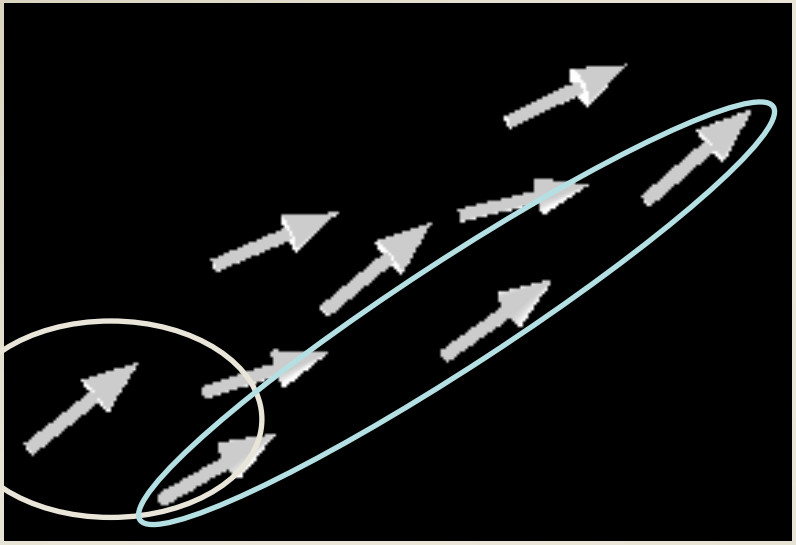
water: $\varepsilon \approx 80\varepsilon_0$

protein: $\varepsilon \approx 2-4\varepsilon_0$

α **helix:**

small r – unfavored

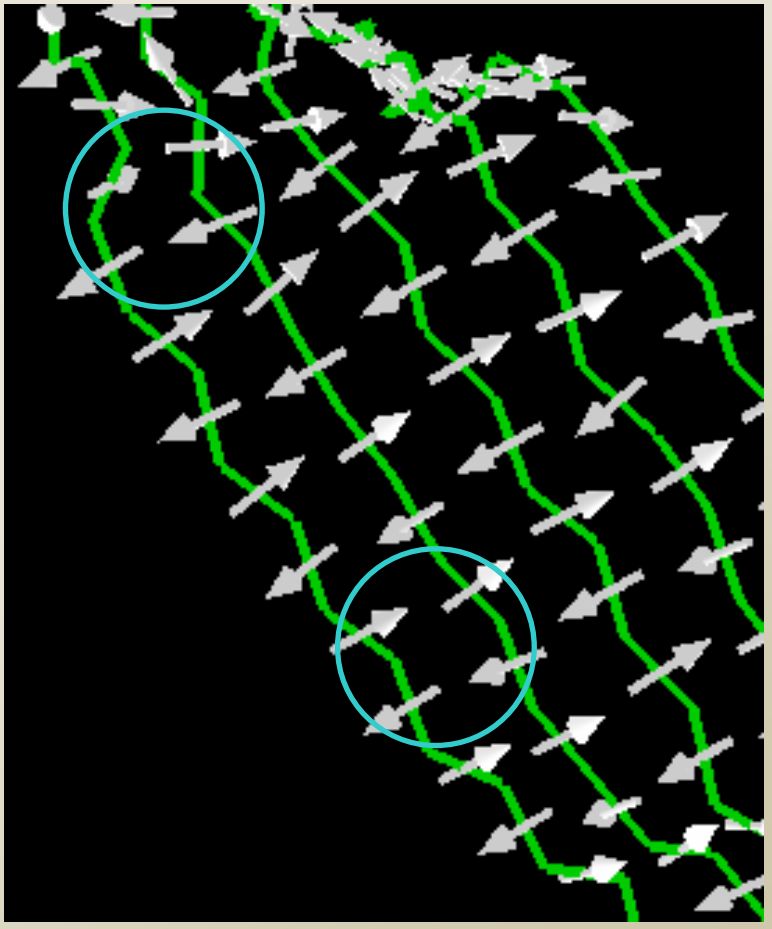
large r -- favored



β **sheets:**

small r – favoured

large r --neutralized



Different Screening effect for different lengths

large r: $\varepsilon_d = DG$

small r: $\varepsilon_d = DN$

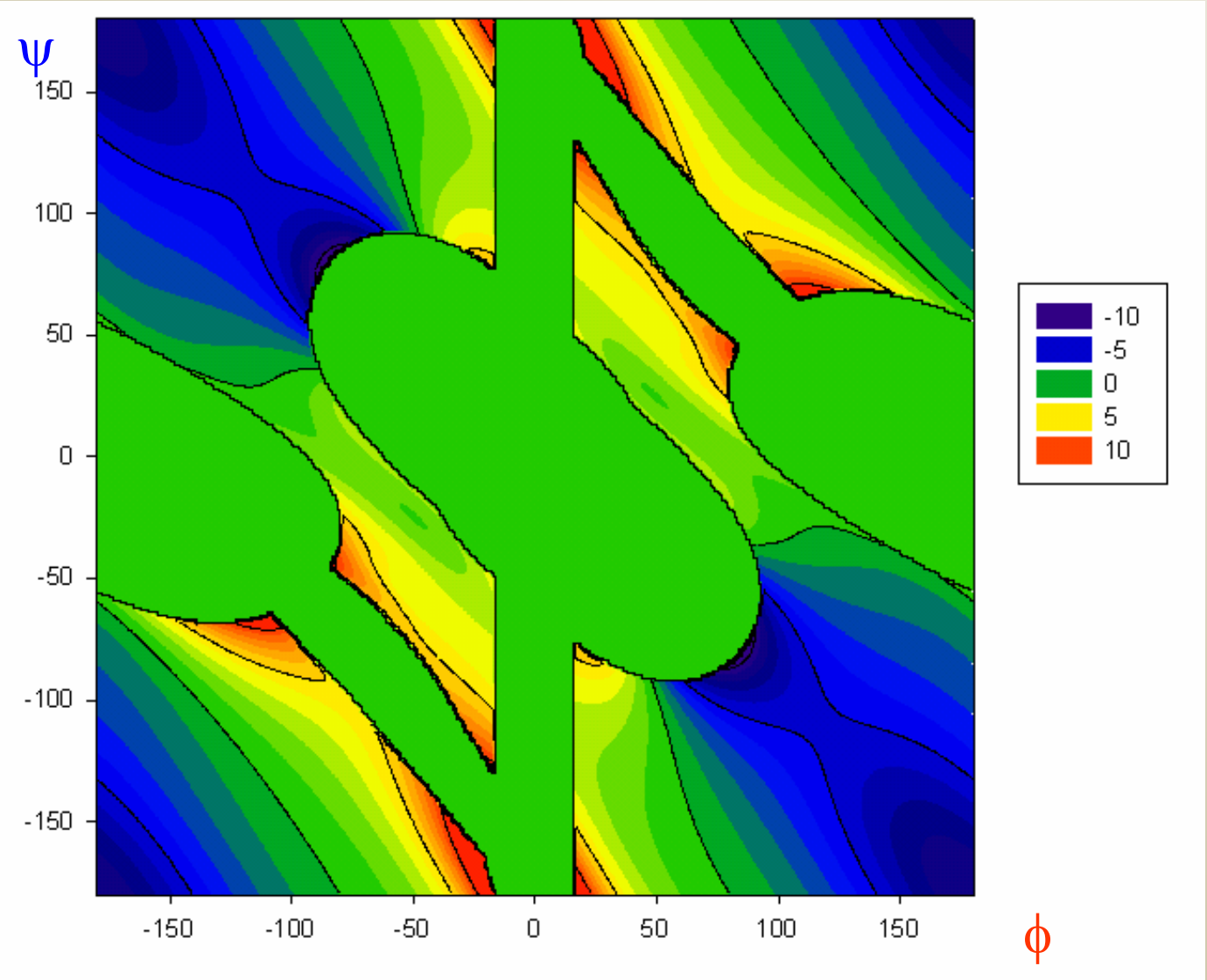
Global:

$$V_{DG} = \varepsilon_{DG} \times \sum_{ij} \left(\frac{\vec{p}_i \cdot \vec{p}_j}{r_{ij}^3} - \frac{3 \times (\vec{p}_i \cdot \vec{r}_{ij})(\vec{p}_j \cdot \vec{r}_{ij})}{r_{ij}^5} \right)$$

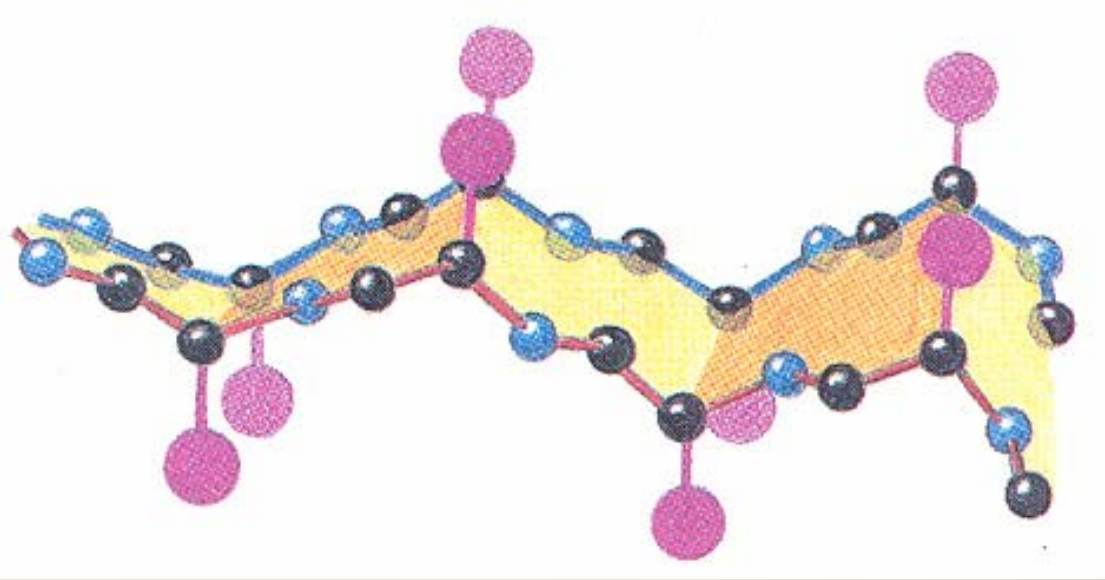
Local:

$$V_{DN} = \varepsilon_{DN} \times \sum_{i,i+1} \left(\frac{\vec{P}_i \cdot \vec{P}_{i+1}}{\left|\vec{P}_i\right| \times \left|\vec{P}_{i+1}\right|} - 1 \right) \times \frac{1}{2}$$

β strands is stabilized by screening effects!
DG=0.0125<DN=0.333



(3) Local hydrophobicity is strongly correlated with the structure



Beta strand: HPHPHP….

Helix: HHPPHHPPHHPP….

Classification of side-chain units

Amino acids are divided into five classes:

- **Hydrophobic (H) :**

  Ala, Val, Cys, Leu, Ile, Met, Phe, Tyr , Trp
- **Polar (P) :**

  Ser, Thr, Asn, Gln, His
- **Neutral (N) : Gly, Pro**
- **Positive (+) : Lys, Arg**
- **Negative (-) : Asp, Glu**

Local hydrophobic interaction

Attraction: V_{HH} , V_{PP} , **and** V_{+-} for ↑↑

Repulsion: V_{HP} , V_{++} , **and** V_{--} for ↑↓

	Neutral	Hydrophobic	Polar	Positive	Negative
Neutral	0	0	0	0	0
Hydrophobic	0	V_{HH}	V_{HP}	V_{HP}	V_{HP}
Polar	0	V_{HP}	V_{PP}	V_{PP}	V_{PP}
Positive	0	V_{HP}	V_{PP}	V_{++}	V_{+-}
Negative	0	V_{HP}	V_{PP}	V_{+-}	V_{--}

Other ingredients

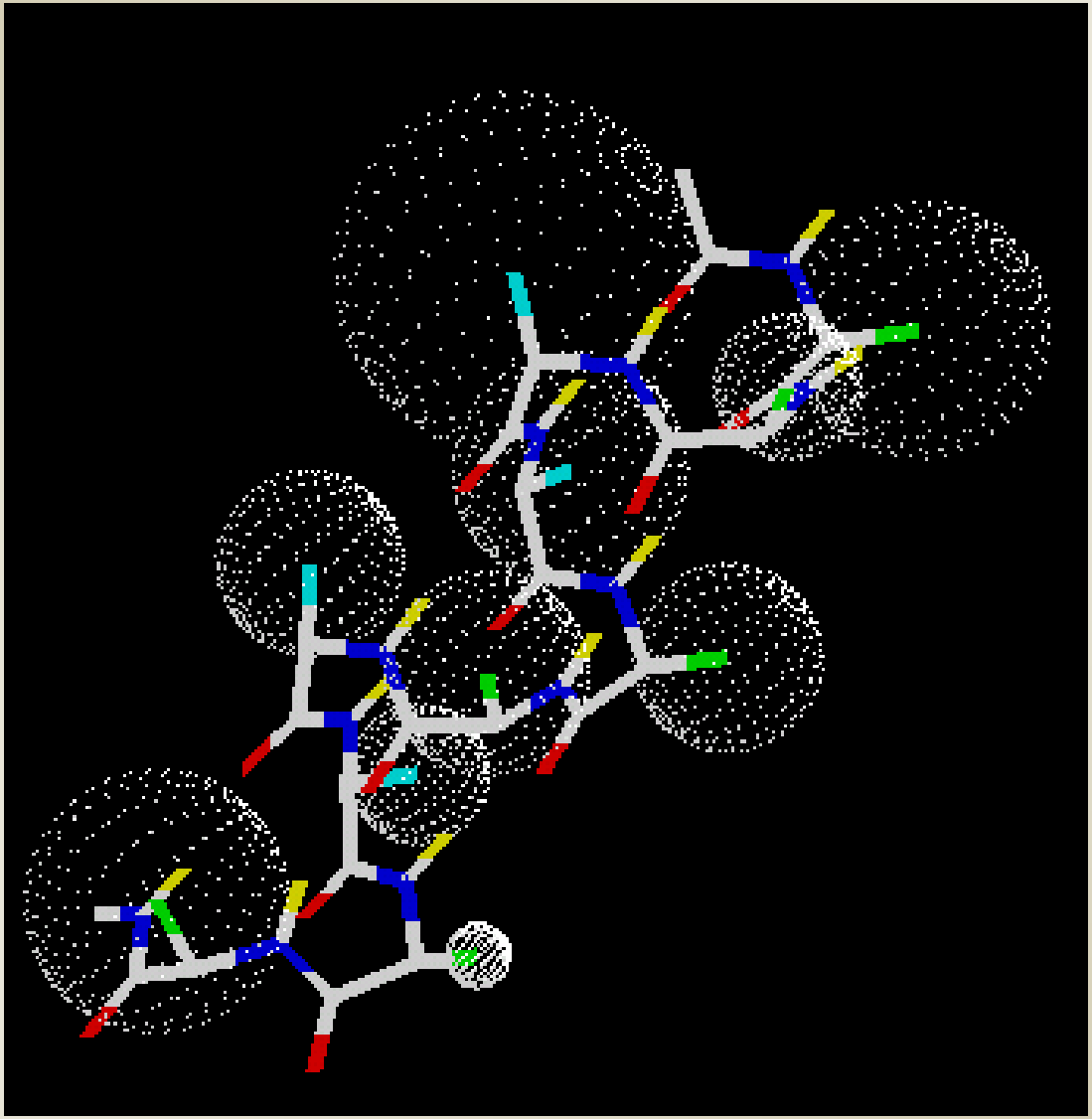
$$V_{total} = V_{steric} + V_{GlobalHP} + V_{HB} + V_{DD} + V_{LocalHP}$$

$$V_{steric} = V_{hardcore} + V_{dihedral}$$

$$V_{GlobalHP} = V_{MJ} + V_{A}$$

$V_A \propto$ solvent-accessible-surface-area

Protein representation



Numerical Method

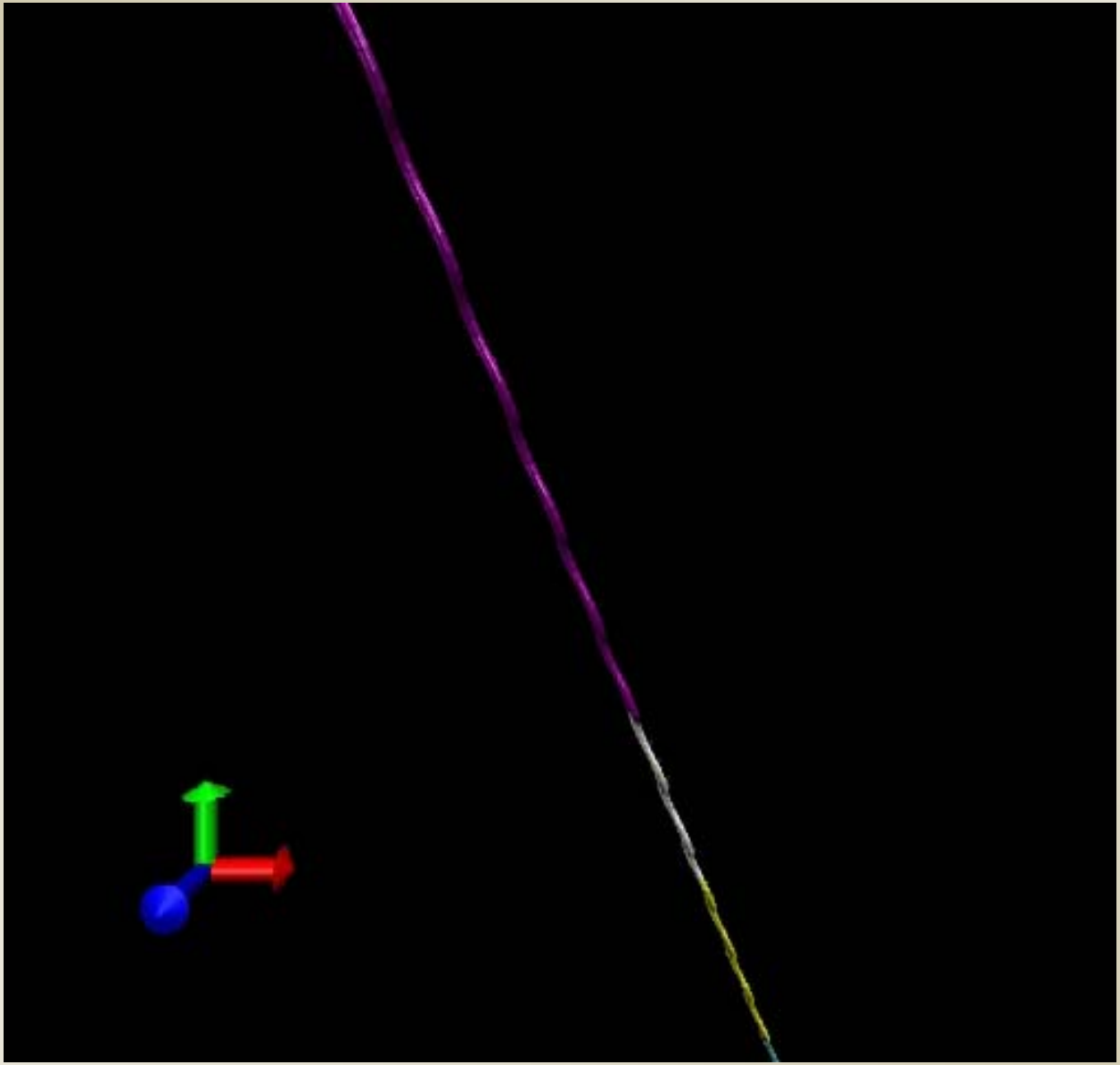
- Monte Carlo method with Metropolis algorithm

$$w_{i\to j} = \min\left\{1, e^{-\beta\,\Delta E}\right\}$$

$$\Delta E = E_j - E_i,\ \beta = \frac{1}{kT}$$

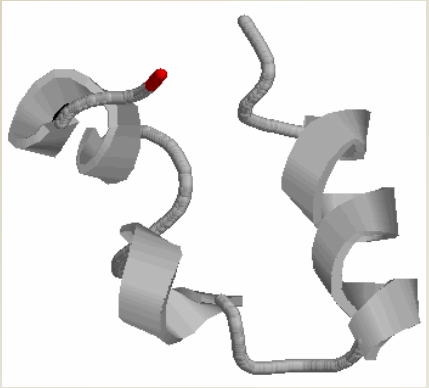
$w_{i\to j}$ is the probability from state *i* to state *j*.

- Initial structures: line structures, random generating structures, or unfolded structures



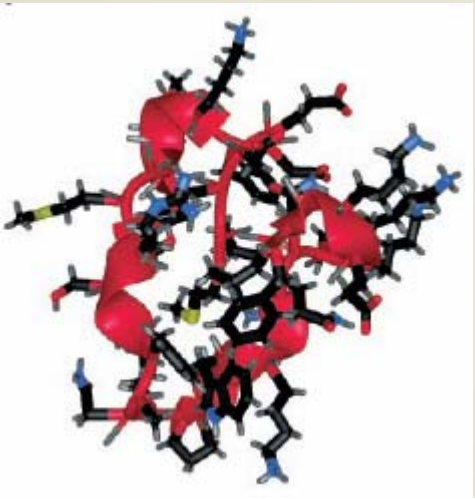
Results

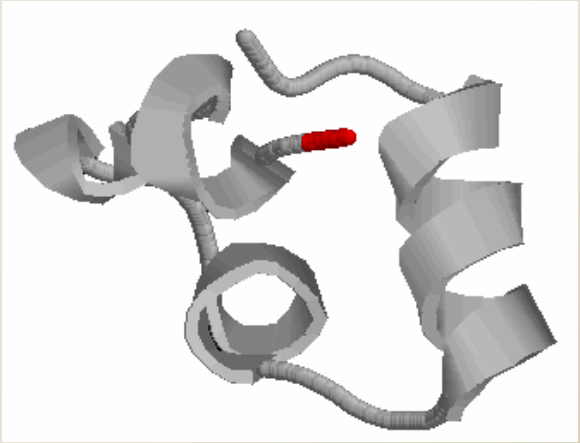
1VII

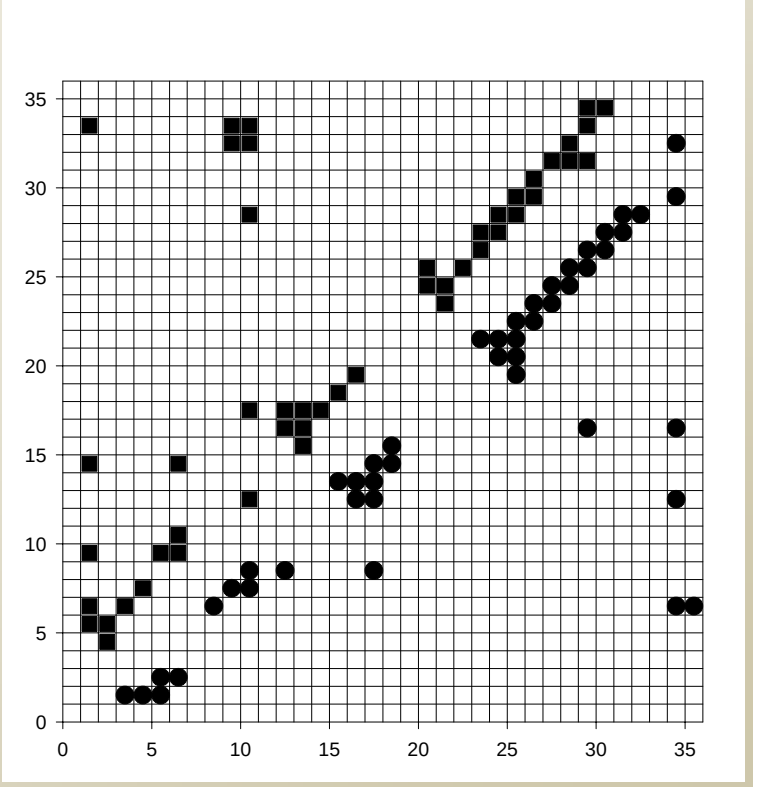


reference

Duan and kollman

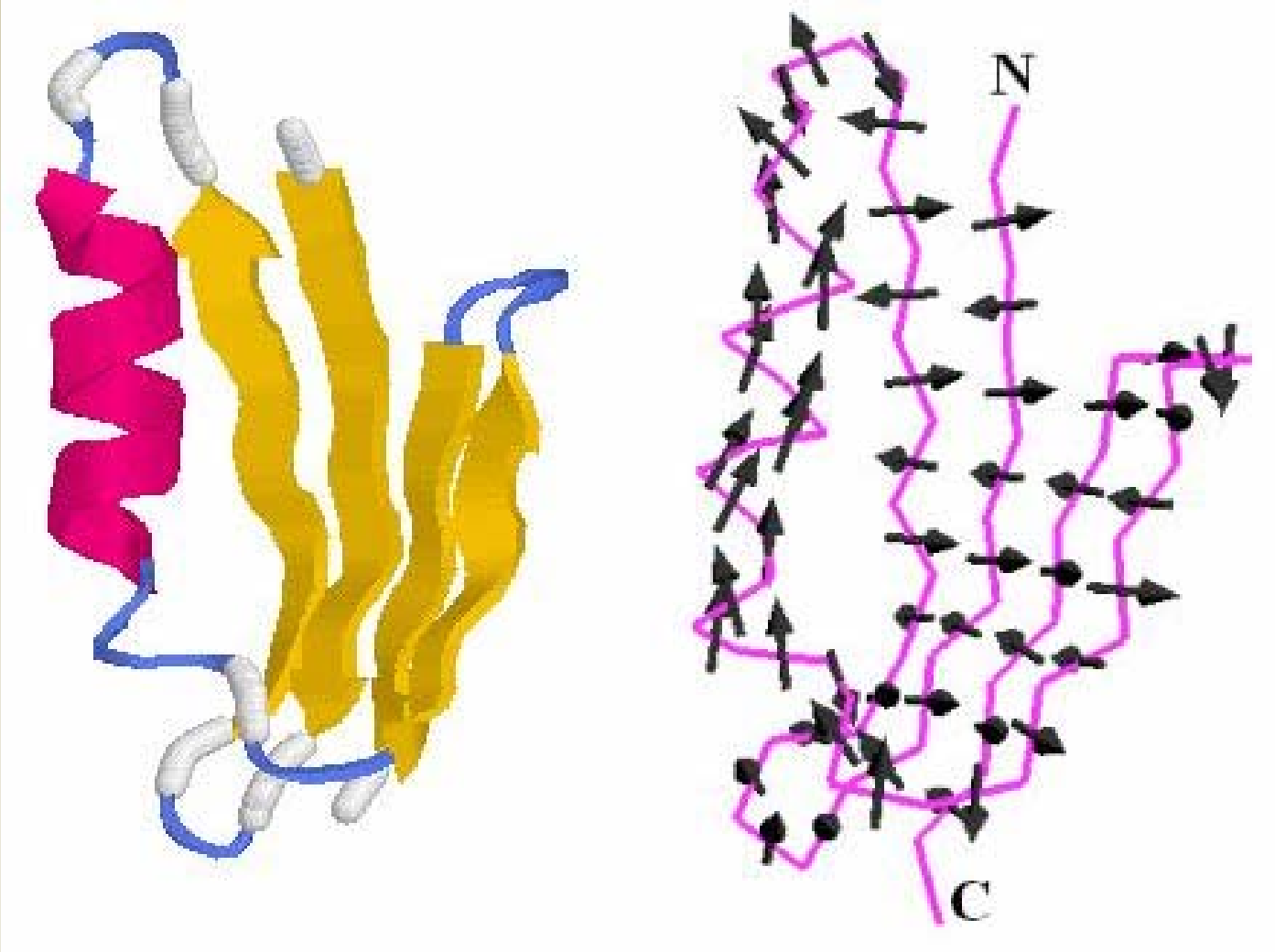






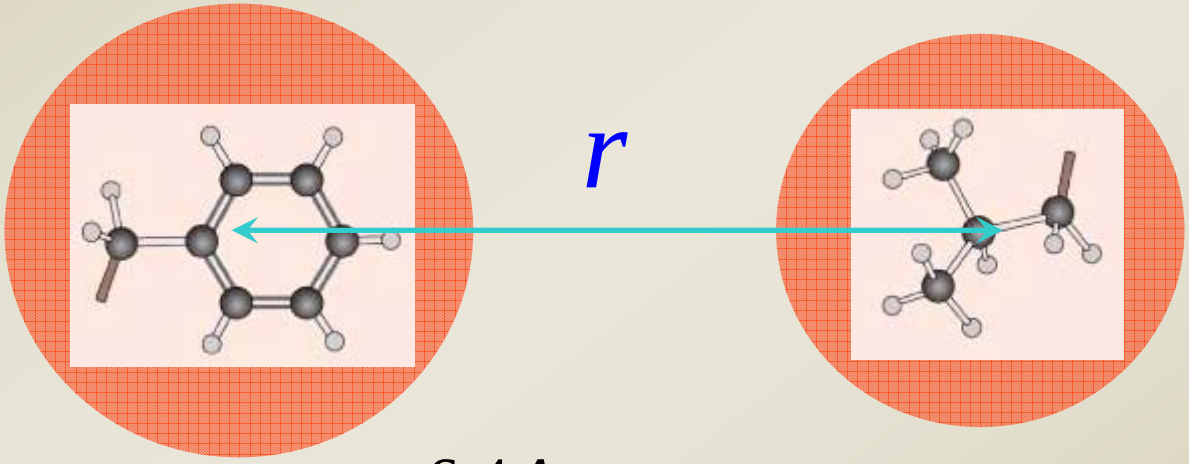
The protein G

- Immunoglobulin binding protein
- PDB ID : 1GB4
- 56 residues
- 15 different types of amino acids
- One α helix
- Two β sheets
- Sequence = M-T-Y-K-L-I-L-N-G-K-T-L-K-G-E-T-T-T-E-A-V-D-A-A-T-A-E-K-V-F-K-Q-Y-A-N-D-N-G-V-D-G-E-W-T-Y-D-D-A-T-K-T-F-T-V-T-E
- 3-4 hours (fastest), 5-8 hours (average) on a P4-3GHz PC


N
C

An efficient way of describing shapes

Contact configuration

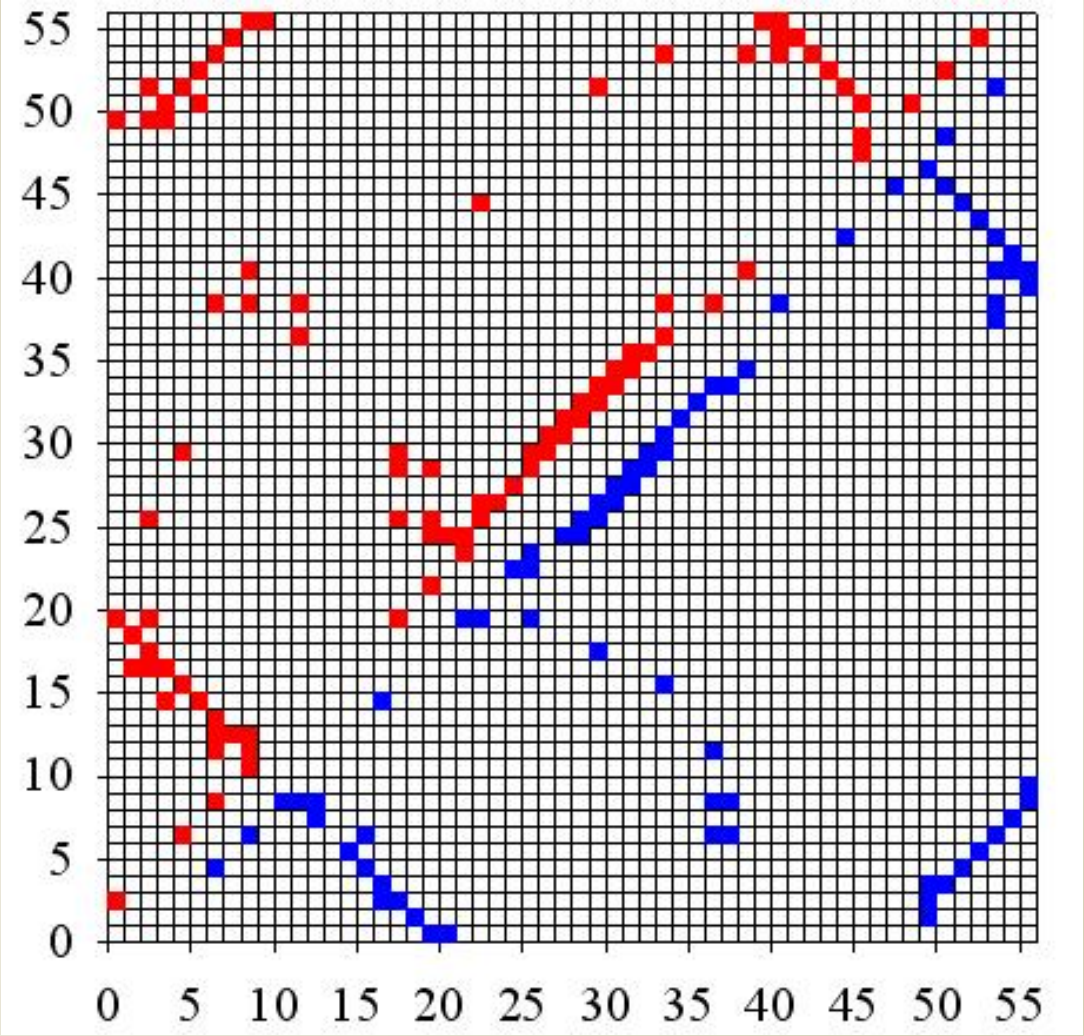


$r < 6.4A$ = one contact

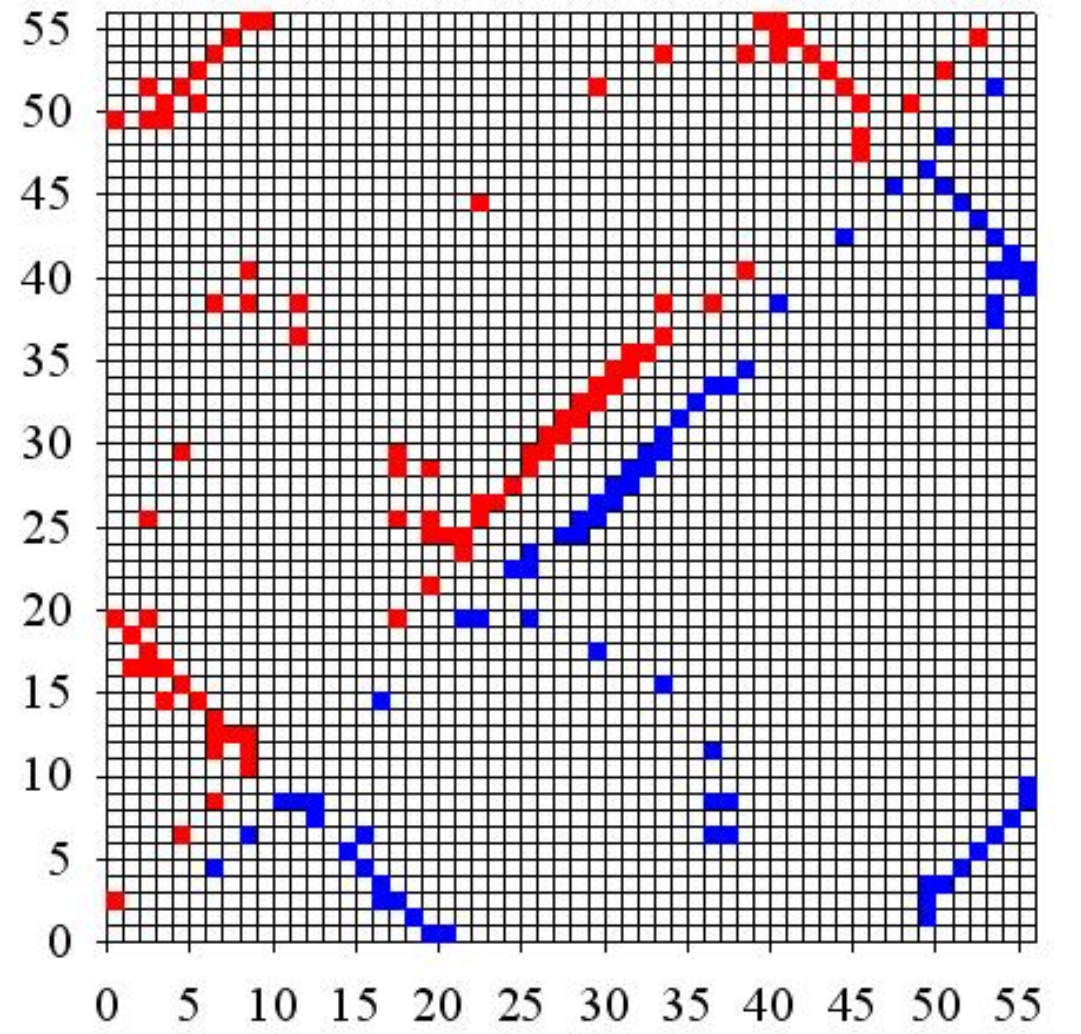
Each nonlocal contact reduces entropy greatly!

David Baker, Nature 405, 39, 2000

Contact configuration almost fixes the shape



Contact map



$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}} = 2.97A$$

Red : PDB data

Blue : Simulation result

Capturing structures of proteins G and L with only 25% similarity

Only 25% similarity in sequence but have similar structure

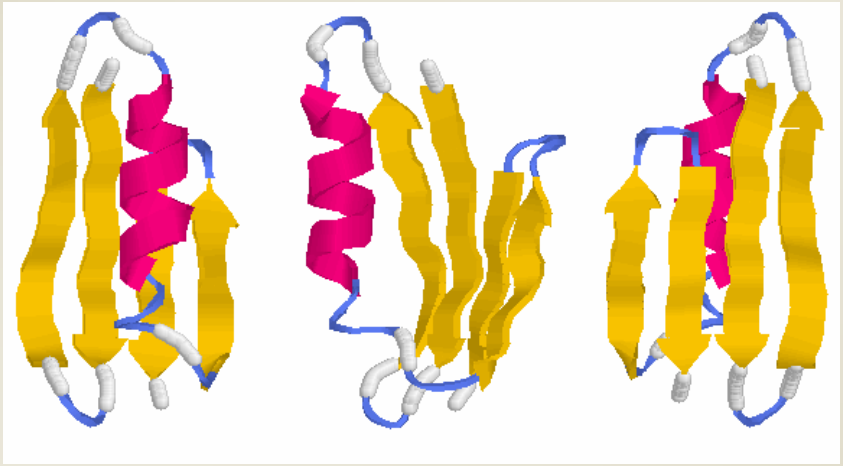
Protein G (1GB4 with 56 residues)

TTFKLIINGKTLKGEITIEAVDAAEAEKIFKQYANDNGIDGEWTYDDATKTFTVTE

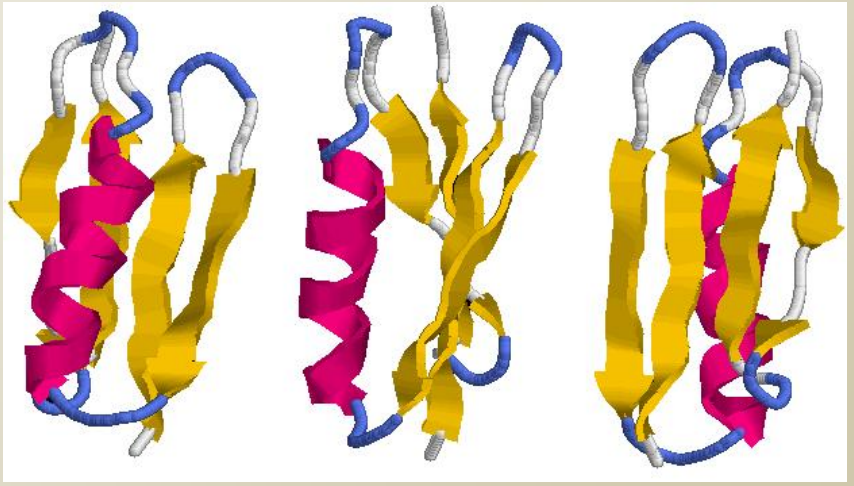
Protein L (2PTL with 61 residues)

VTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEYTVDVADKGYTLNIKFAG

Protein G



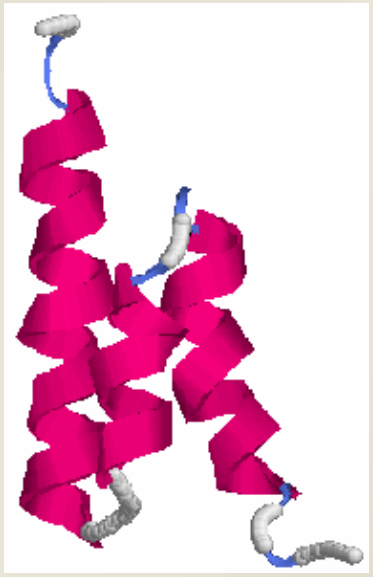
Protein L

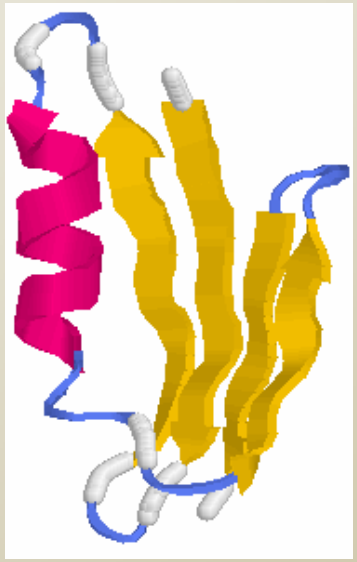


Proteins 1Gb88 and 2Ga88 with only seven amino acids in difference

TTYKLILNLKQAKEEAIKELVDAATAEKYFKLYANAKTVEGVWTYKDETKTFTVTE
Gb88 = 2JWU

TTYKLILNLKQAKEEAIKELVDAGIAEKYIKLIANAKTVEGVWTLKDEILTFTVTE
Ga88 = 2JWS





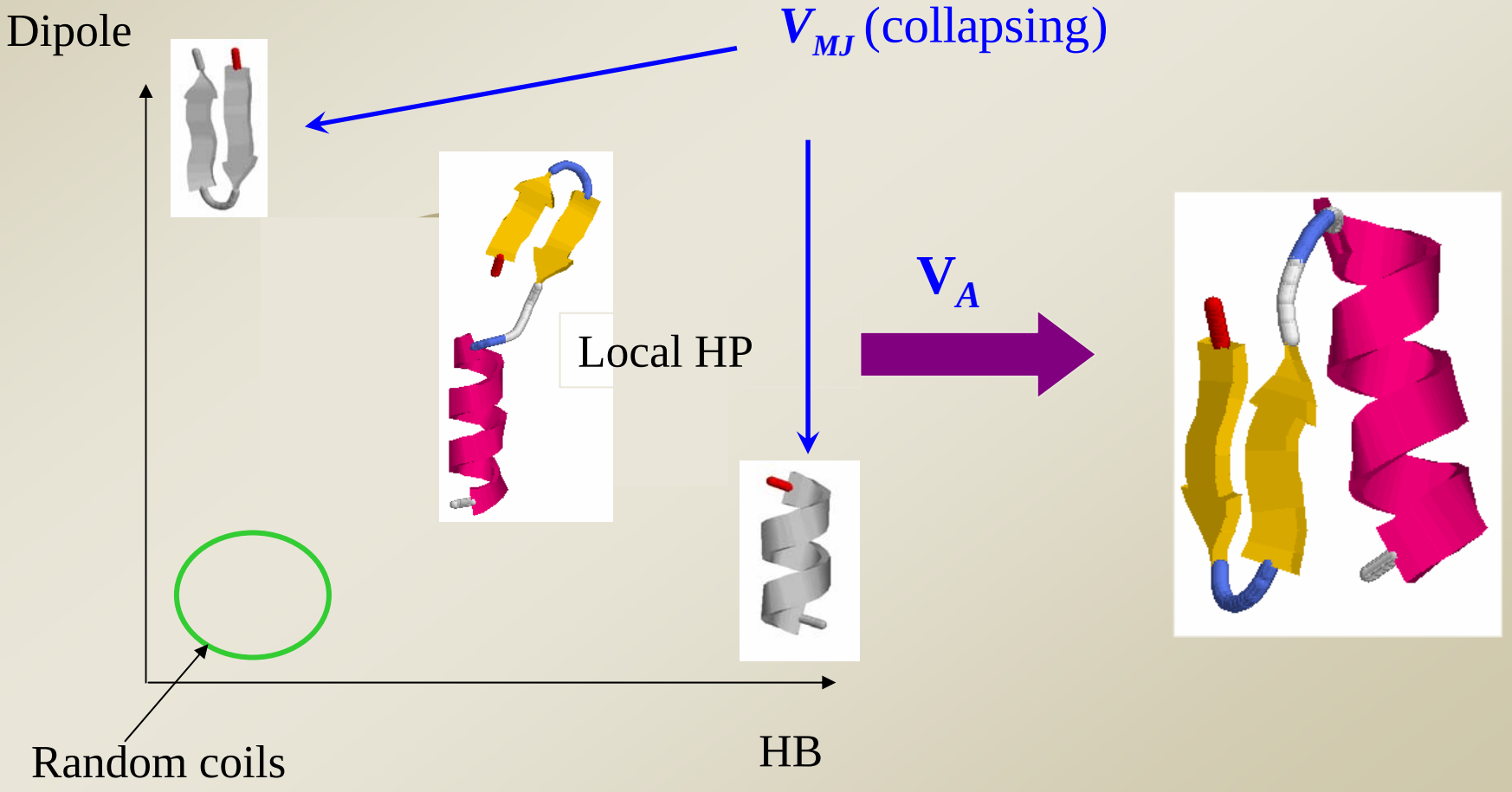
Real proteins

PDB code	Seq. #	Sequences	Main structures
1BYZ	12	ELLKKLLEELKG	Helix
1DJF	15	QAPAYKKAAKKLAES	Helix
1L4X	15	DELERAIRELAARIK	Helix
1NJ0	16	RKRIHIGPGRAFYTTK	Sheet
1IBN	20	GLFGAIAGFIENGWEGMIDG	Helixes
1PEI	22	VEEKSIDLIQKWEEKSREFIGS	Helix
1OEG	23	PLVEDMQRQWAGLVEKVQAAVGT	Helix
3ZNF	30	RPYHCSYCNFSFKTKGNLTKHMKSKAHSKK	Helix,Sheet
1PIQ	31	RMKQIEDKIEEILSKQYHIENEIARIKKLIG	Helix
1LYP	32	GLRKRLRKFRNKIKEKLKKIGQKIQGLLPKLA	Helix
1BB1	34	AEIAAIEYEQAAIKEEIAAIKDKIAAIKEYIAAI	Helix
1ZDD	34	FNMQCQRRFYEALHDPNLNEEQRNAKIKSIRDDC	Helixes
1PPT	35	GPSQPTYPGDDAPVEDLIRFYDNLQQYLNVVTRHRY	Helix
1VII	36	MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF	Helixes

PDB code	GSE, *kcal/mol*	R_G, Å		*WAV*, Å^3		*Q*	*RMSD*, Å	*MCsteps* / 10^7
		NS	Sim	NS	Sim			
1BYZ	–96.2	5.853	6.031	5774	6084	0.923077	1.4650	24.0
1DJF	–117.2	8.782	7.930	7436	7171	0.75	2.6396	3.4
1L4X	–129.3	7.042	7.951	7332	7556	0.75	2.2158	11.0
1NJ0	–115.5	8.469	7.427	8120	7765	0.470588	3.5714	11.0
1IBN	–138.8	8.943	7.266	7926	7667	0.714286	3.7758	1.8
1PEI	–191.2	10.484	9.933	10243	9585	0.863636	2.2651	4.4
1OEG	–176.1	10.662	10.971	10355	10778	0.681818	2.7816	5.4
3ZNF	–234.7	8.439	9.296	10879	12510	0.2	5.7044	17.0
1PIQ	–262.0	14.065	11.606	12420	12823	0.543478	6.6259	5.4
1LYP	–262.5	15.175	12.293	13928	13064	0.794118	2.7162	25.0
1BB1	–342.5	14.637	14.973	11590	13248	0.893617	2.0816	20.0
1ZDD	–306.8	9.243	9.271	11943	11736	0.704918	4.5887	25.0
1PPT	–260.4	10.863	12.281	14241	15209	0.509804	5.9307	11.0
1VII	–293.8	8.939	8.922	11599	12290	0.541667	5.2003	8.0

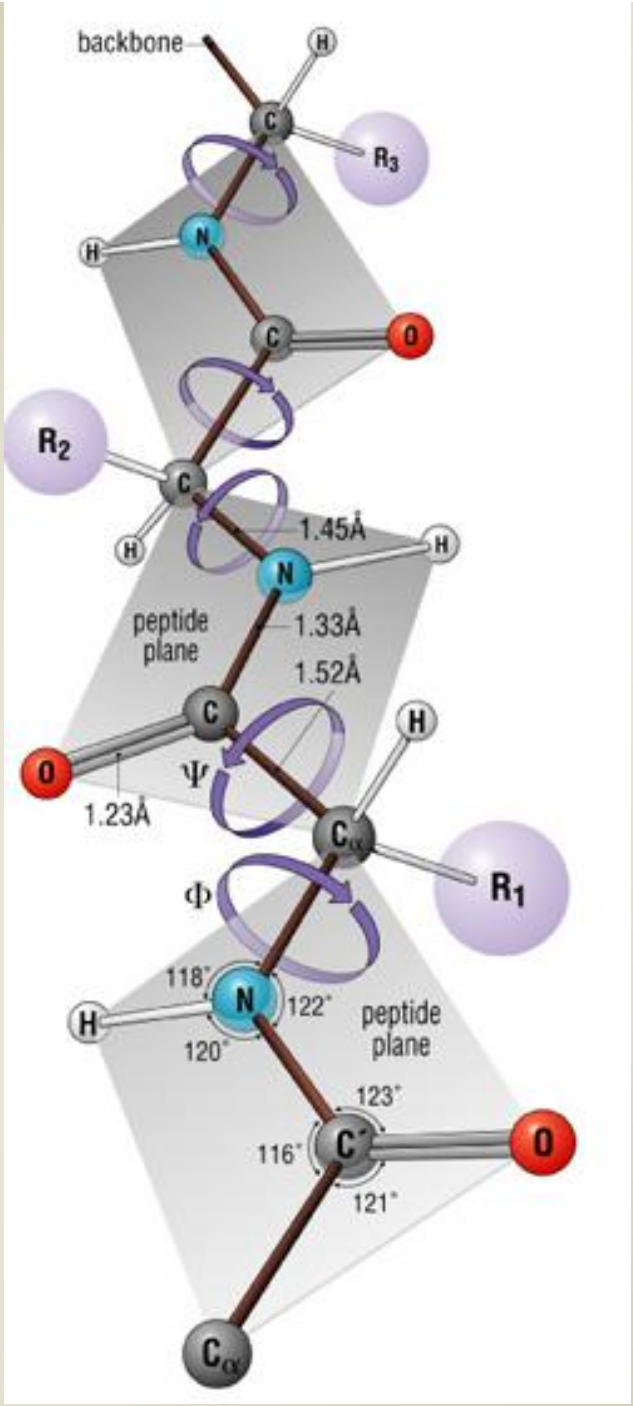
Summary and outlook

Phase diagram of protein structures



*多數物理學家所關心的生物問題本質上是物理問題與生物學家極需要解決的實際問題有一段差距

*物理在分子尺度上技術的成熟是促成21世紀是生物世紀的主要因素


backbone
H
C
R3
N
H
C
O
R2
C
1.45Å
H
H
N
peptide
plane
1.33Å
1.52Å
C
H
O
Ψ
1.23Å
Cα
R1
Φ
118°
N
122°
H
peptide
plane
120°
123°
116°
C
O
121°
Cα

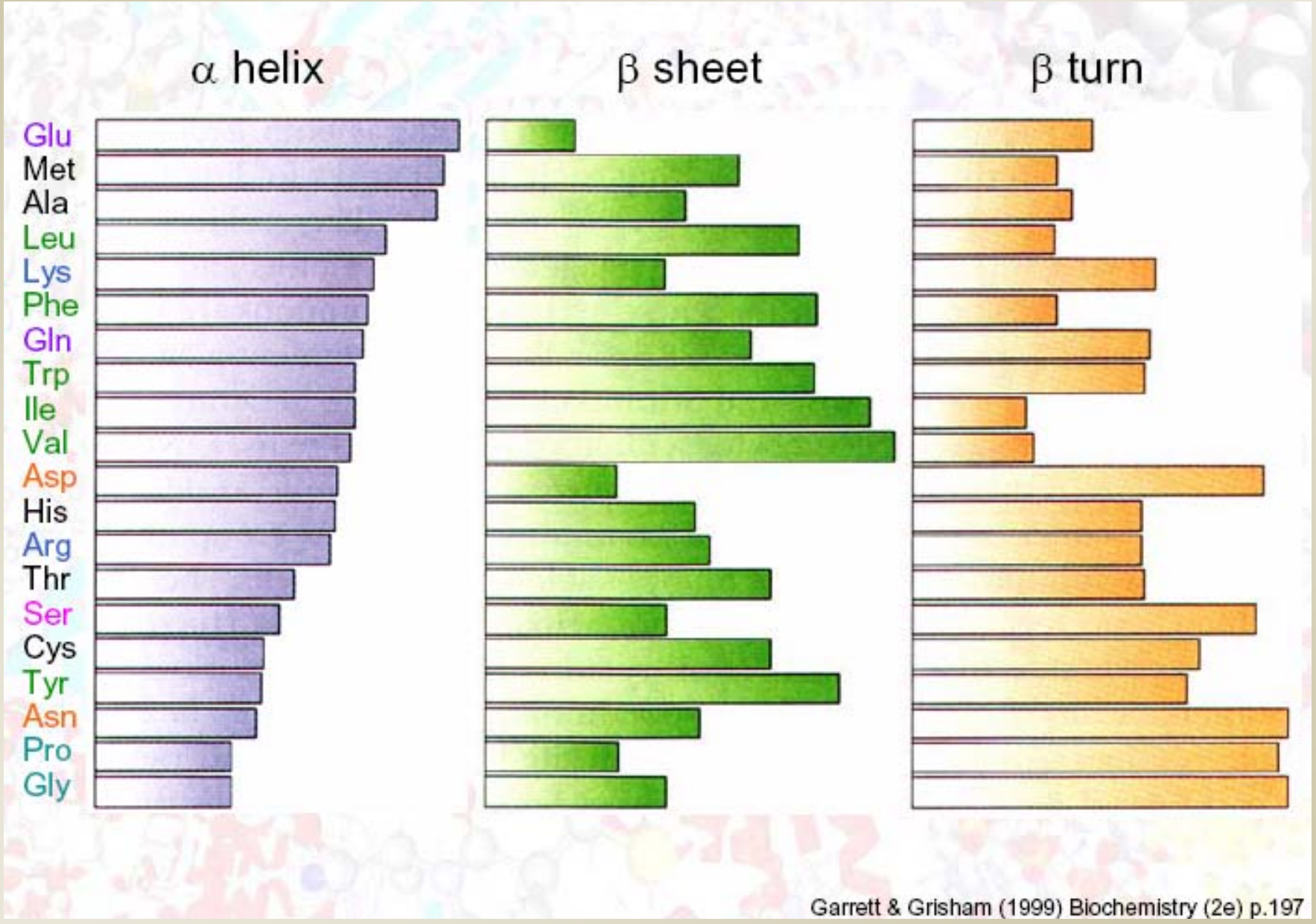
For a given protein sequence, what is the native structure?

Anfinsen, Science 181, 223-224 (1973):

For a given protein sequence and condition, the protein always folds to the same structure (native structure).

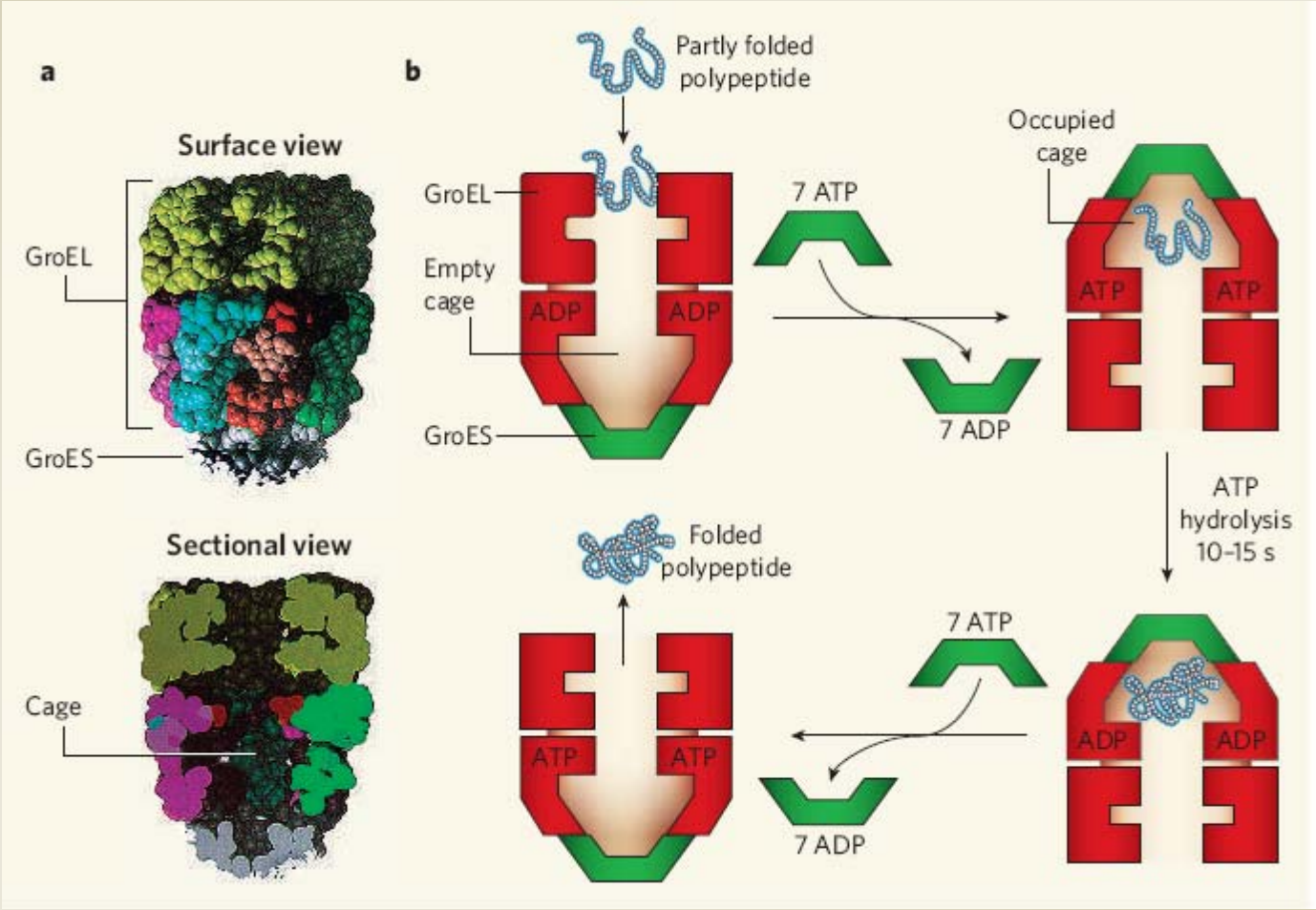
[primarily based on 核醣核酸水解酶A (ribonuclease A, RNaseA)]

The whole sequence matters



Garrett & Grisham (1999) Biochemistry (2e) p.197

It turns out that for large, multi-domain proteins, one needs chaperones to help folding proteins:



R. J. Ellis, Nature 442, 360, 2006

Levinthal paradox:

Random search:

100 amino acids, each link with two configuration (at least)

$\Rightarrow 2^{100} \times 10^{-9}\text{sec} \sim 10^{16}$ years

Whole protein ~ milliseconds to seconds

Hinge motion in proteins ~ nanosecond

} $10^6 - 10^9$

The Miyazawa-Jernigan Statistical Interaction

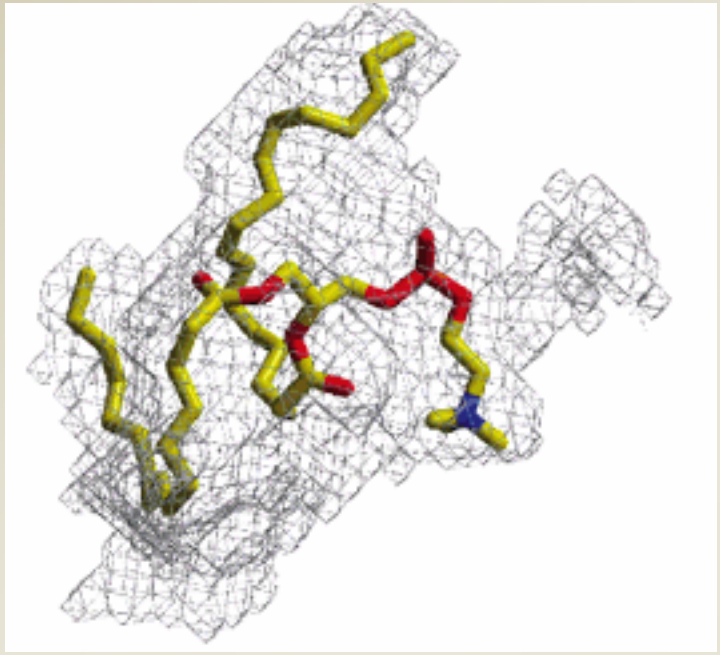
ε_{ij}: MJ matrix (20×20)

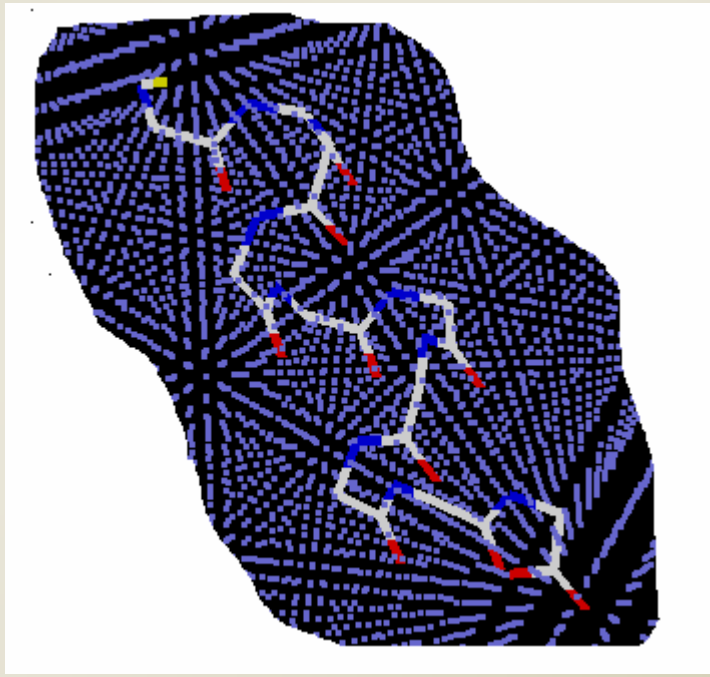
$$\frac{N_{ij}}{\sum N_{ij}} = e^{-\frac{\varepsilon_{ij}-\bar{\varepsilon}}{k_B T}}$$

N_{ij}: number of contact (<6.5A) between the ith and jth amino acides

S. Miyazawa and R.L. Jernigan, Macromolecules 18, 534, (1985) (42 proteins)
S. Miyazawa and R.L. Jernigan, J. Mol. Bio., 256, 623, (1996) (42 proteins)

V_A : On-site potential





$$V_A = \sum_i (\varepsilon_{ii} - \overline{\varepsilon_{ii}}) A_i + A_{backbone}$$

Solvent-accessible-surface-area of ith side-chain