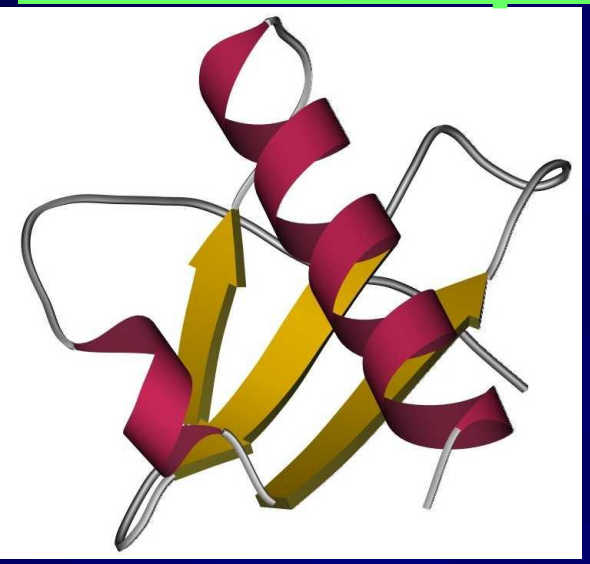
PROTEIN PHYSICS

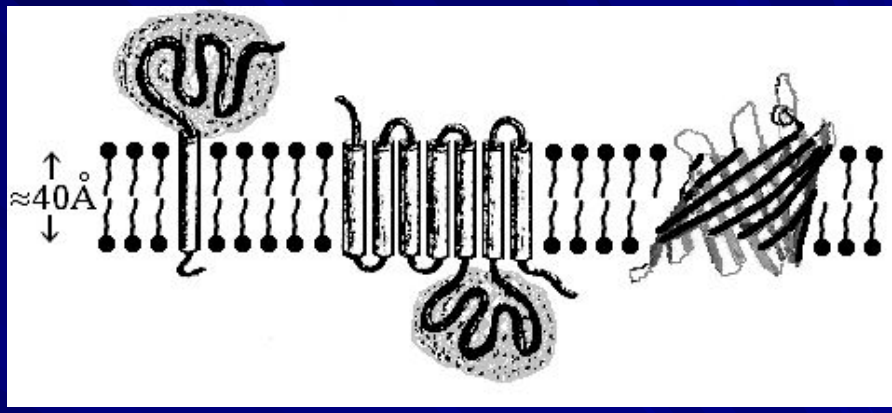
LECTURE 13-16

- Structures of water-soluble globular proteins
- Physical selection of protein structures
- Structural classification of proteins

Globular proteins (water-soluble)



Membrane



Fibrous

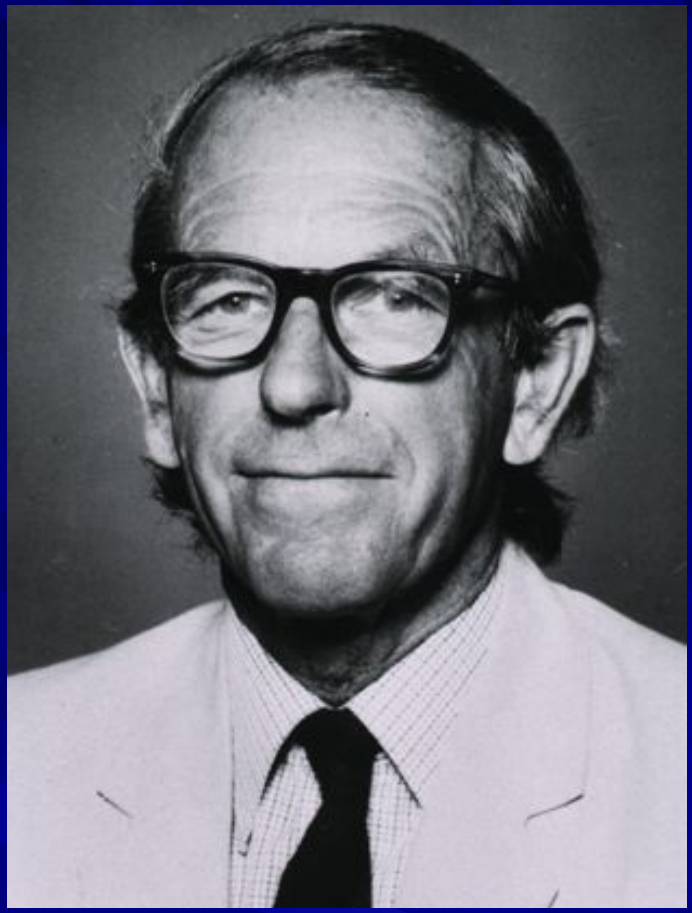
H-bonds & hydrophobics

PROTEINS	SEQUENCES	
Globular	○○●○○●●○○●○●●○●○○●○○○●●○●○○●○●○○●●○●○○○○○●○○○●	*quasi-random*
Membrane	●●●●●●●●●●●○○●○○○●○○○○○●●●●○●●●●●○○○○●●○○○●○●○	*blocks*
	| Hydro-phobic ‖ Hydro-philic |	
Fibrous	●○○●○○○●○○●○○○●○○●○○○●○○●○○○●○○●○○○●○○●○○○●○○●	*repeats*
	| repeat |	

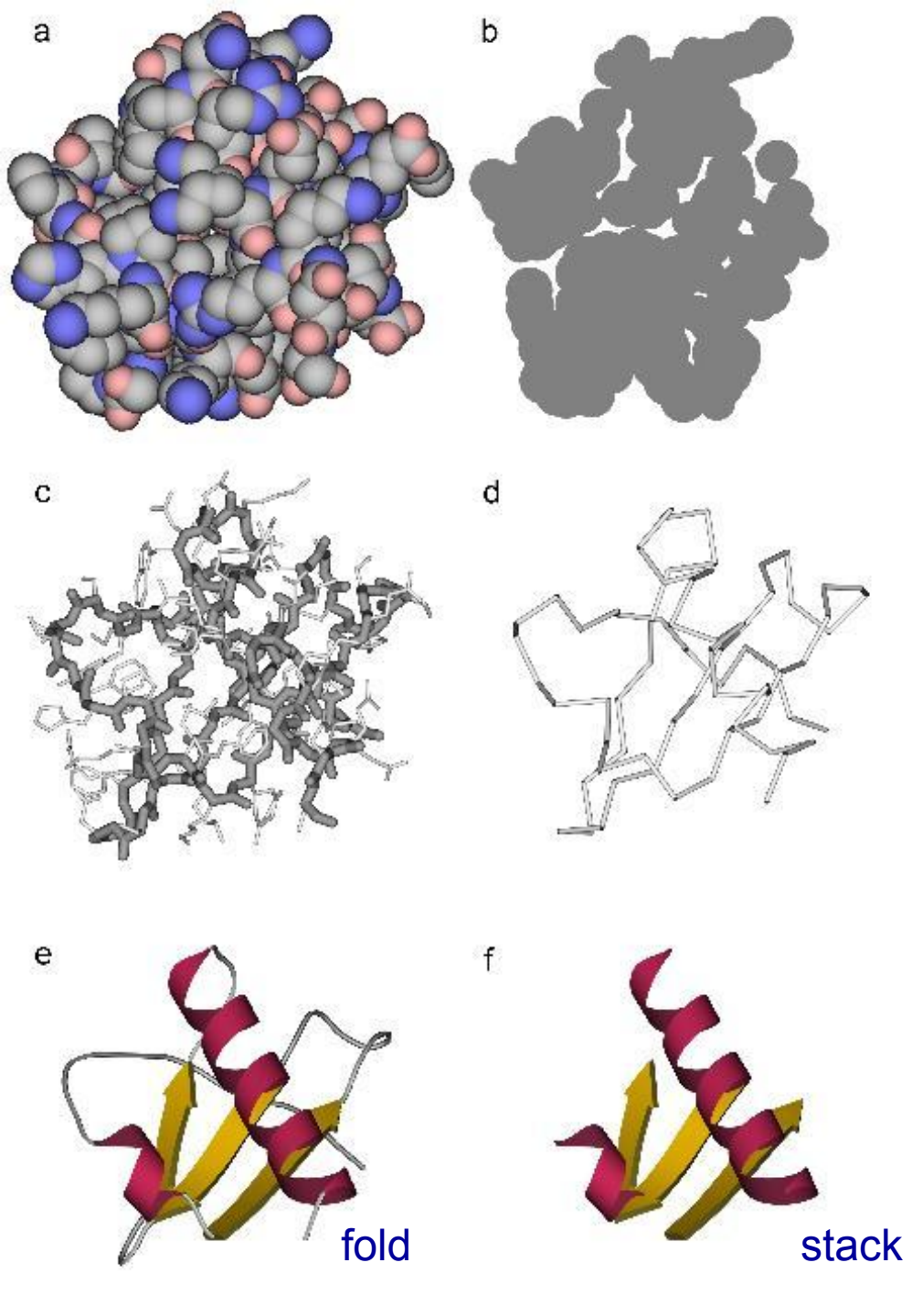
Protein chain

Hermann Emil Louis
Fischer
(1852 –1919)
Nobel Prize 1902

Protein sequence

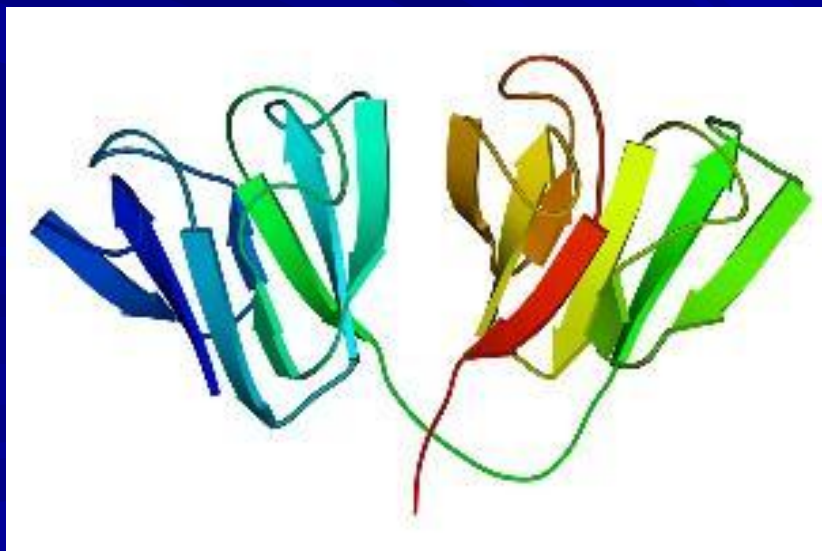


Frederick **Sanger**
(1918 –2013)
Nobel Prizes: 1958, 1980



← single-domain
globular protein

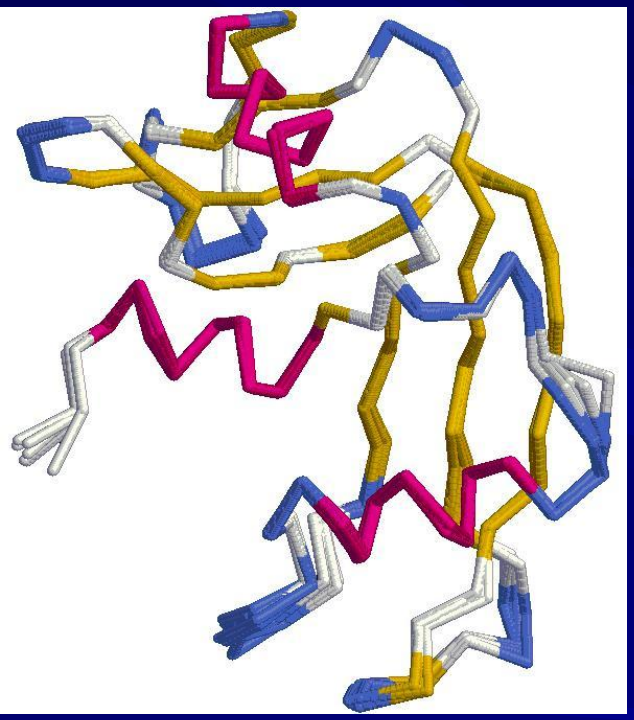
domain 1 domain 2

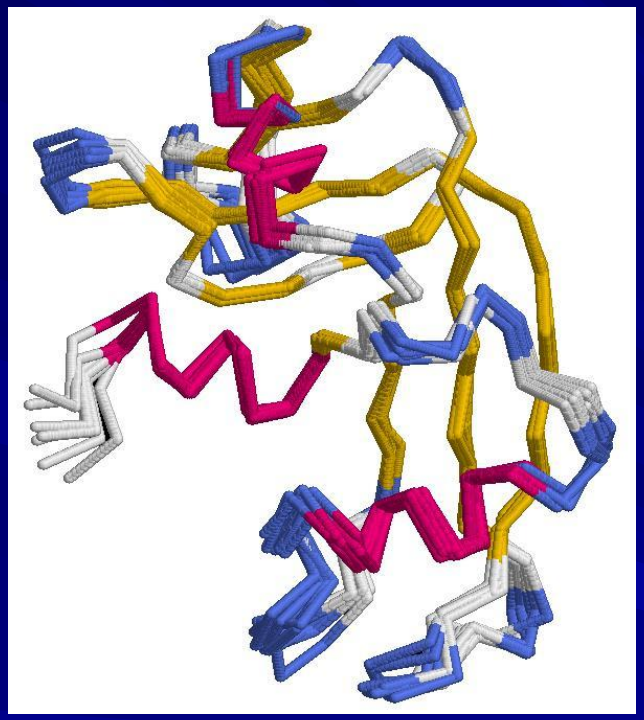


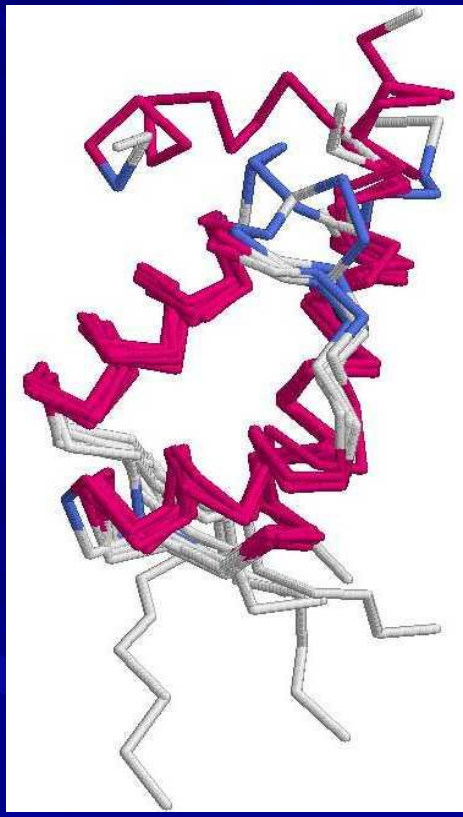
X-RAY
One protein, various crystallizations

NMR
Structures, compatible with one NMR experiment

Homologous
(closely related) proteins





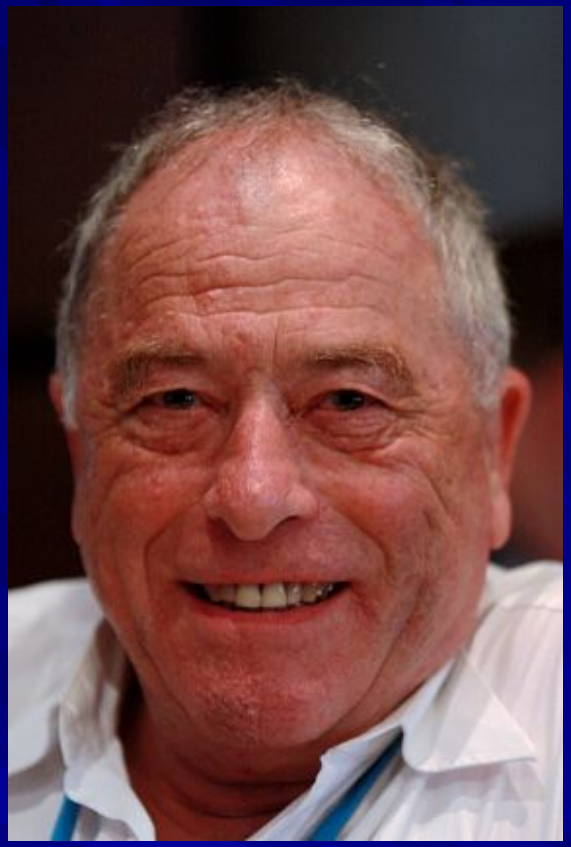


Secondary structures (α-helices, β-strands) are the most rigid and conserved details of proteins; they are determined with the smallest errors and form a basis of protein classification

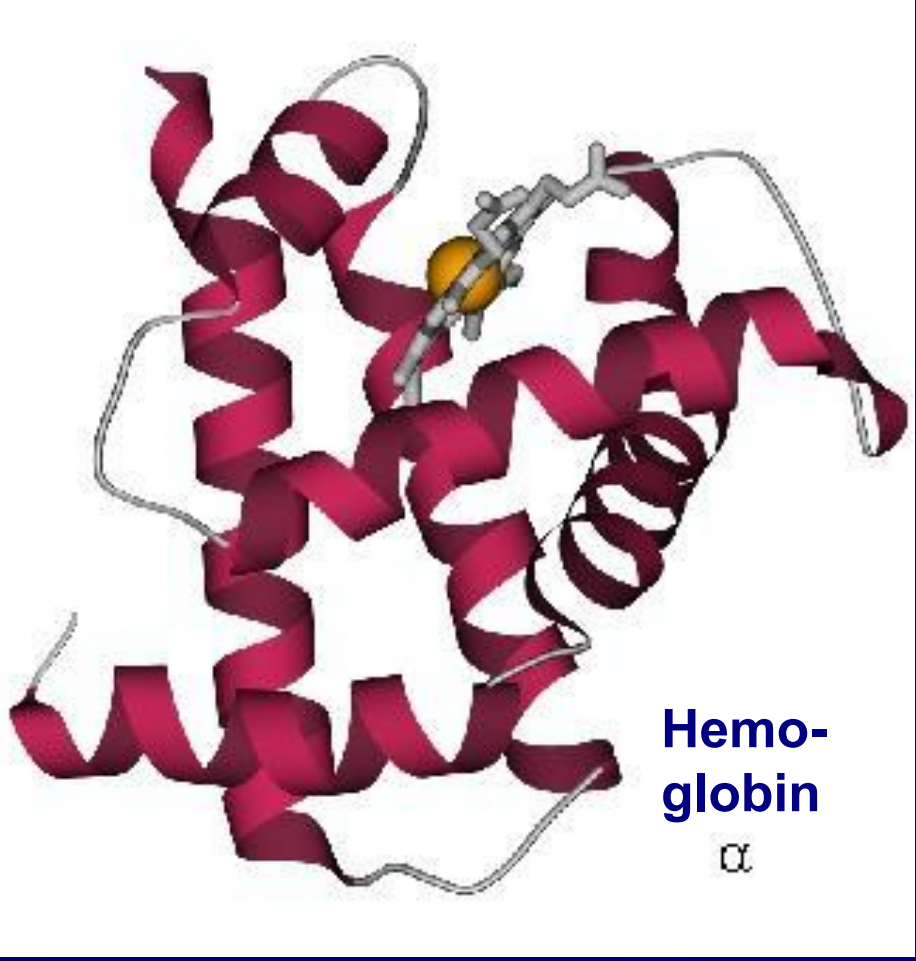
X-ray 3D protein structure

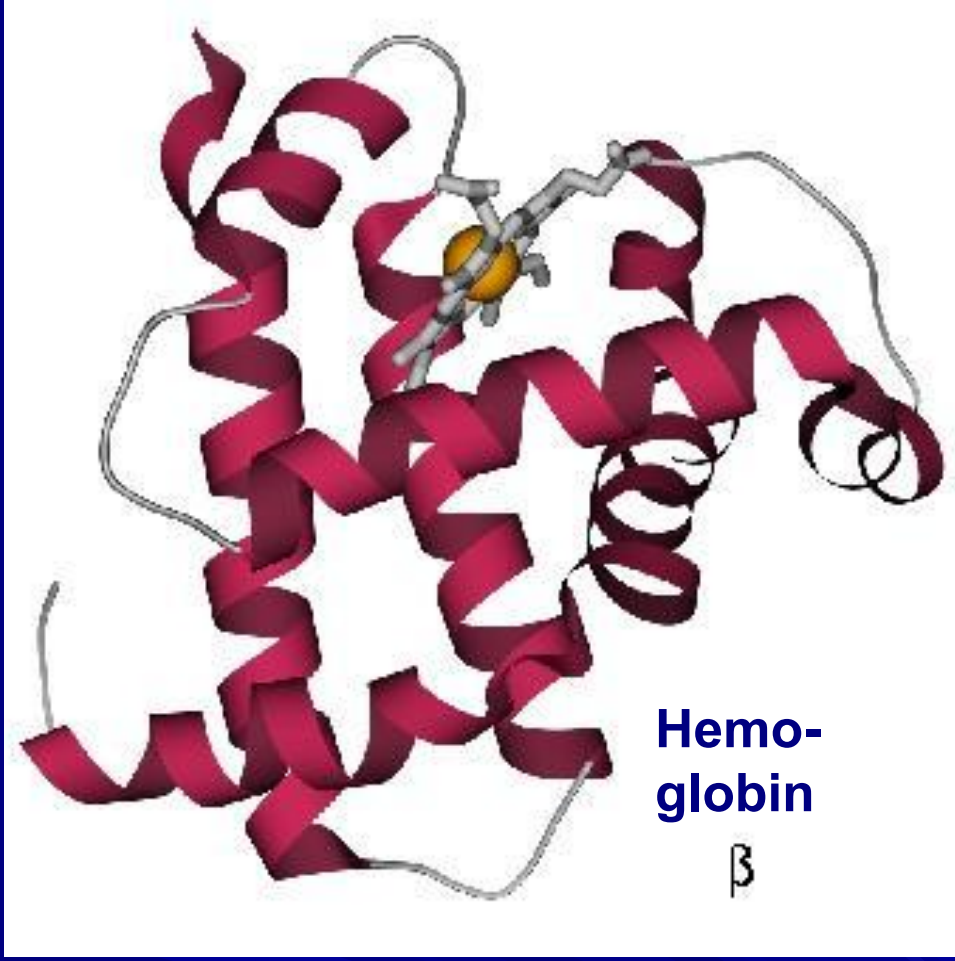
Max Ferdinand **Perutz**
(1914 –2002)
Nobel Prize 1962

NMR 3D protein structure



Kurt **Wüthrich**, 1938
Nobel Prize 2002



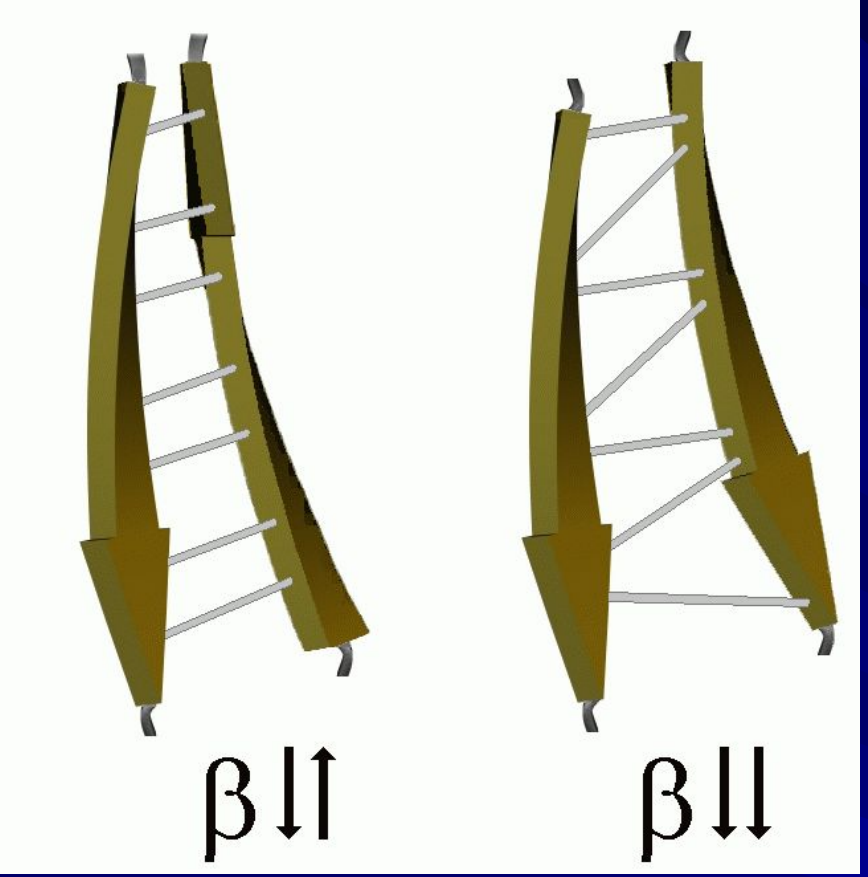


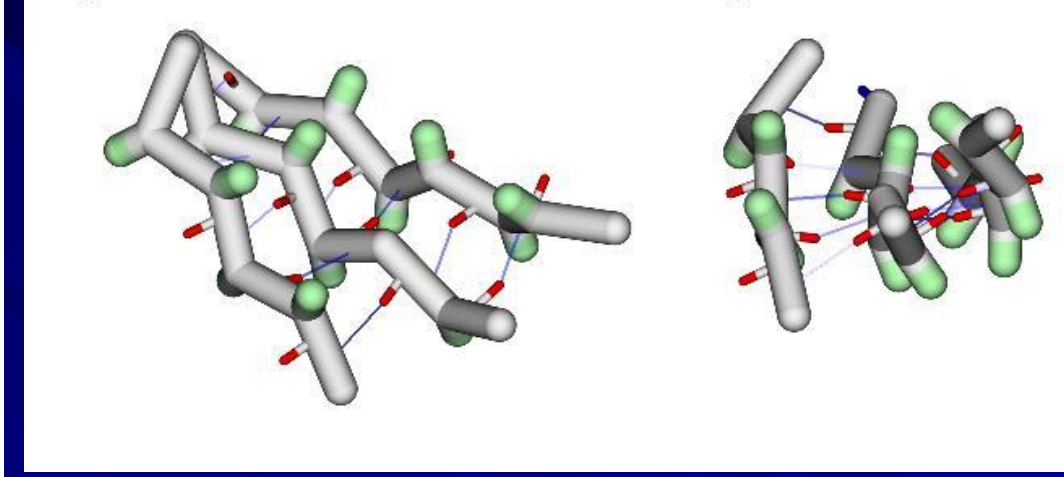
Homologous proteins have similar folds.

True, but trivial.

NON-trivial:

Many NON-homologous proteins have similar folds.



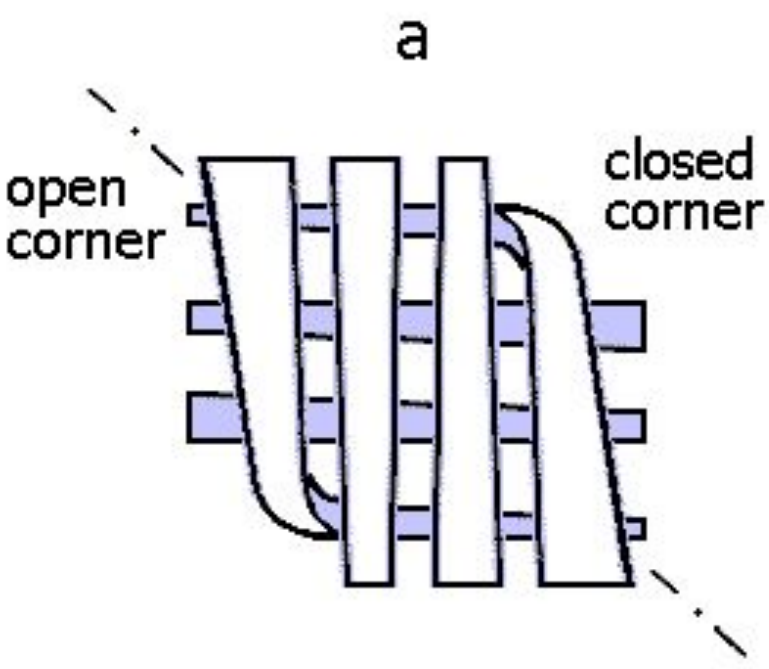


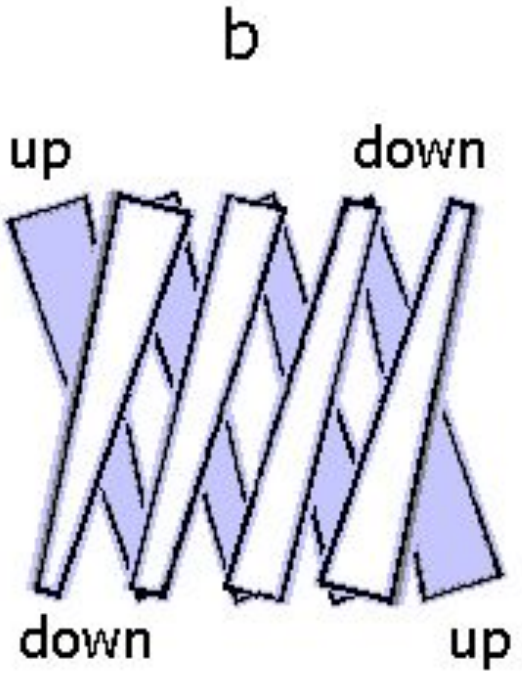
β-sheets: usually, twisted
(usually, right-) ↑

β-proteins

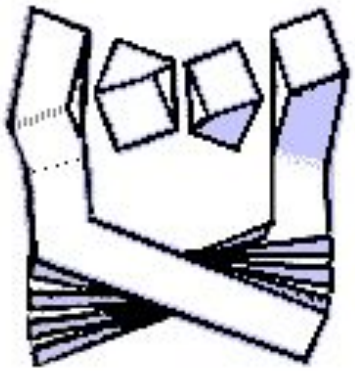
H-bonds: within sheets

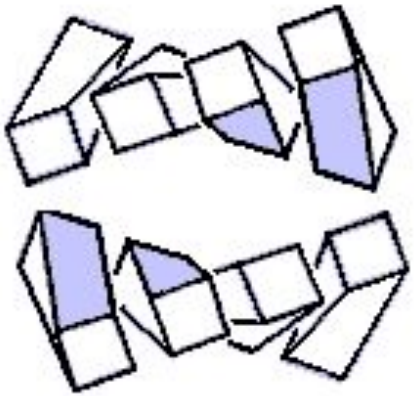
Hydrophobics: between sheets





sandwiches & cylinders

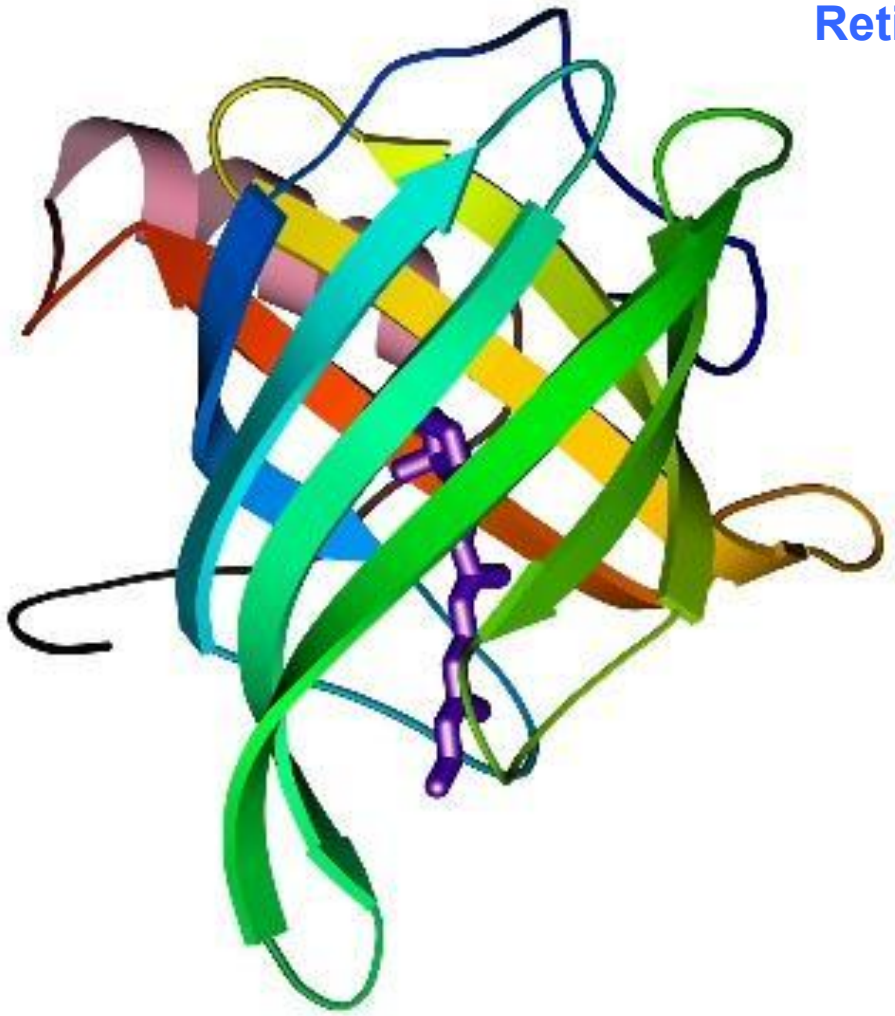


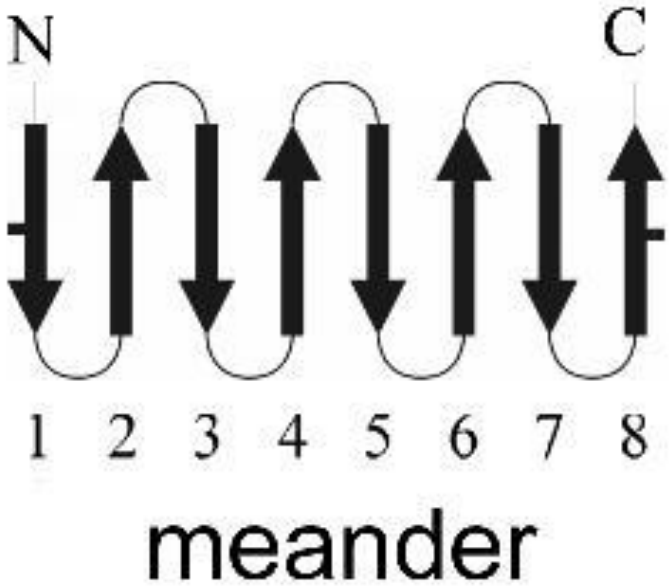


Orthogonal packing of β-sheets

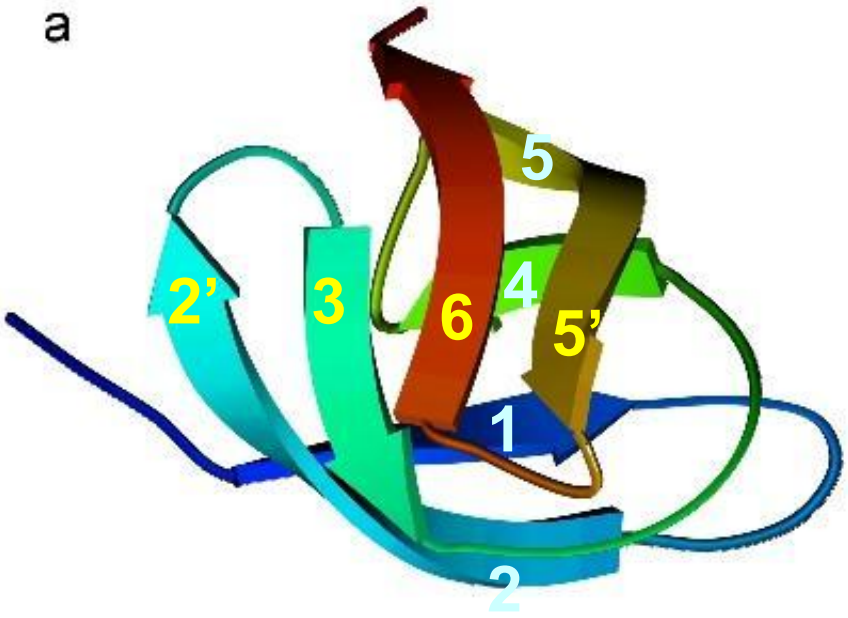
Aligned packing of β-sheets

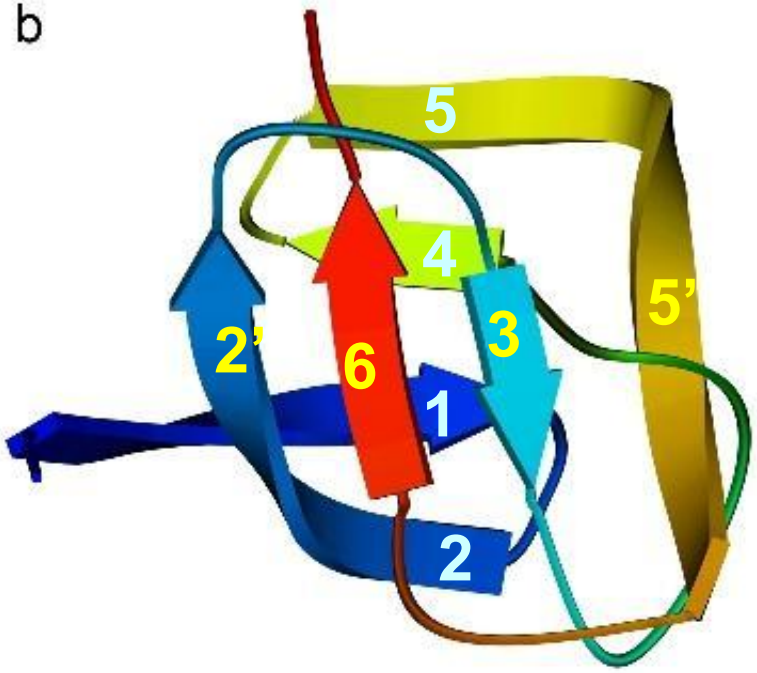
Retinol-binding protein





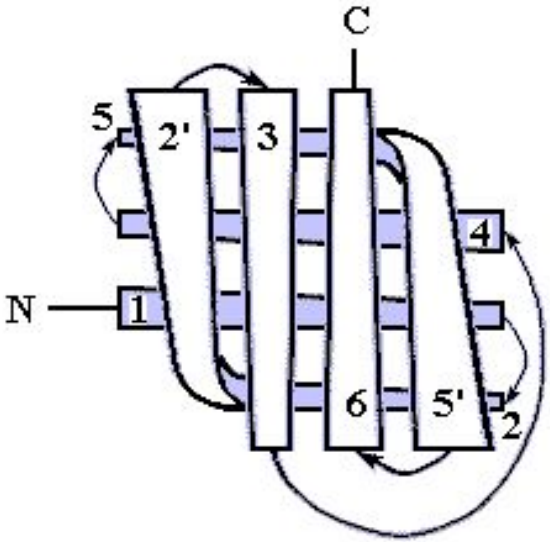
orthogonal packing of one rolled β-sheet

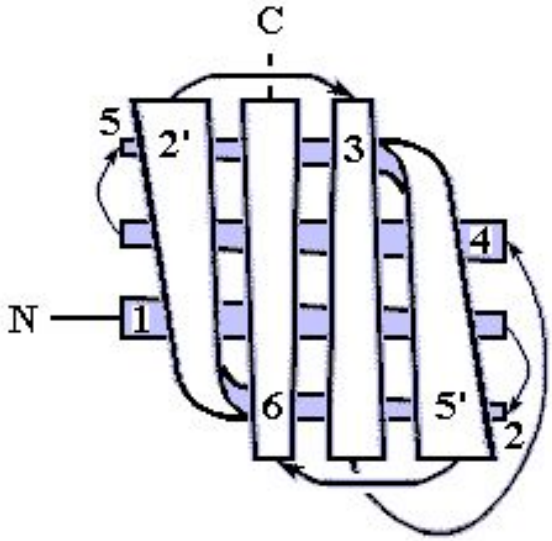


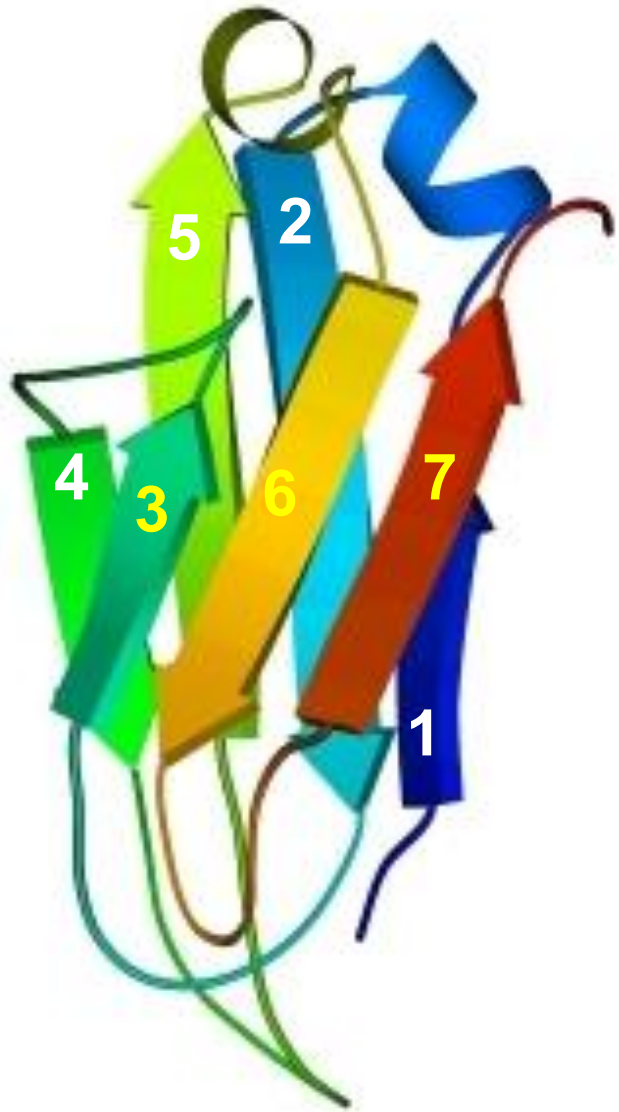


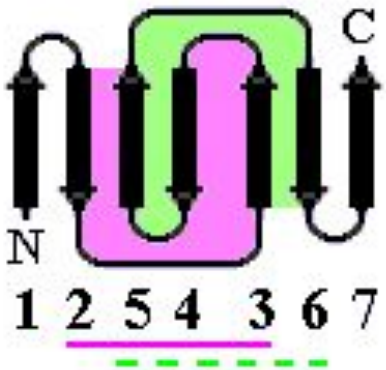
Trypsin-like SER-protease **Acid-protease**

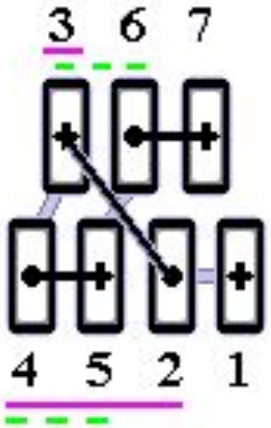
orthogonal packings of β-sheets











Greek key 2::5

Greek key 3::6

non-crossed loops

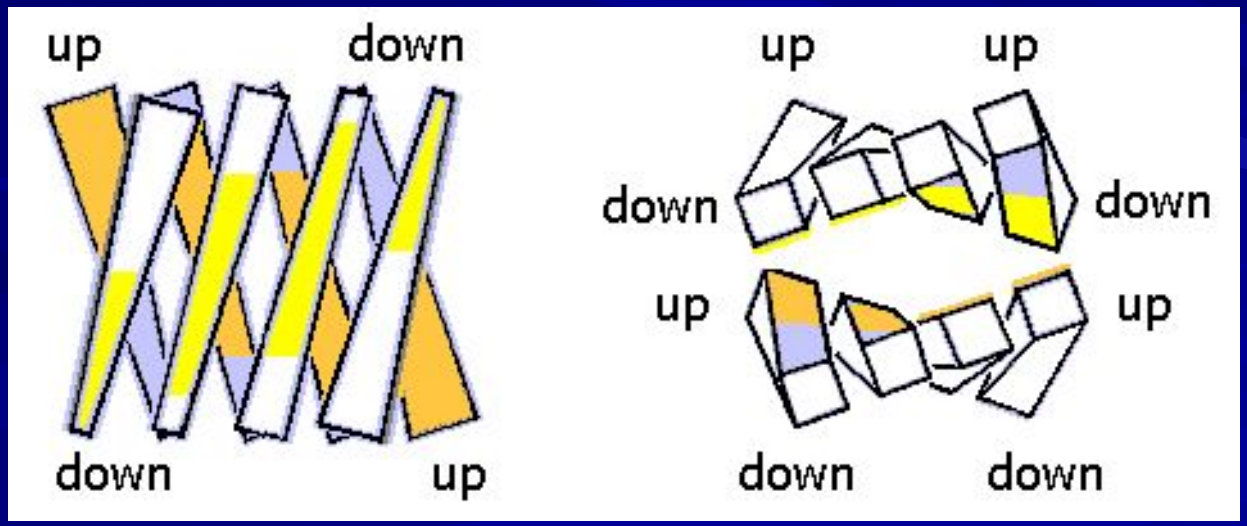
IG-fold: **aligned packing of β-sheets**

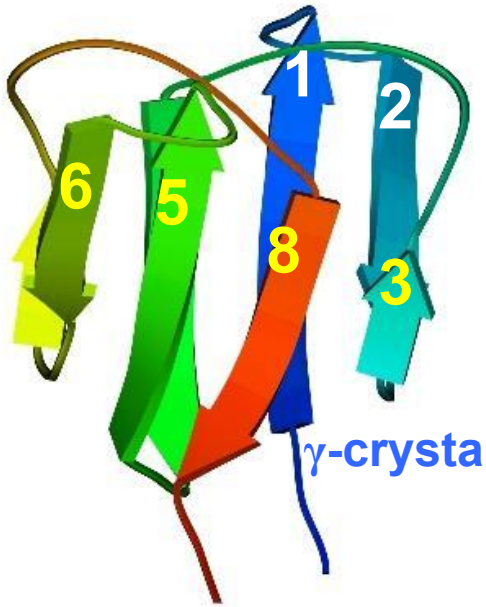
β-sandwich

Greek key:
edge of sandwich

Interlocked pairs:
center of sandwich

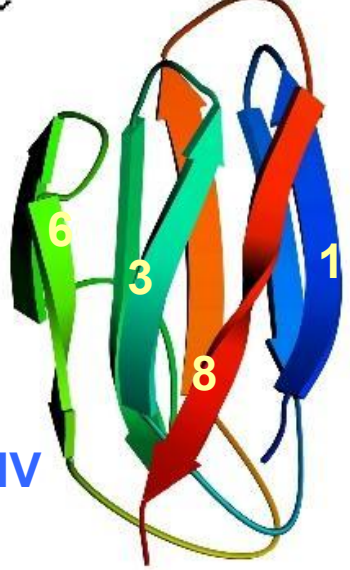
Hydrophobic surfaces
of sheets of the sandwich





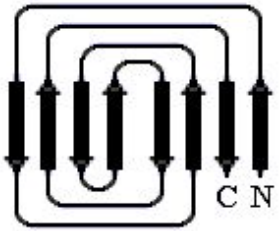
γ-crystallin

βCAB



cpSTNV

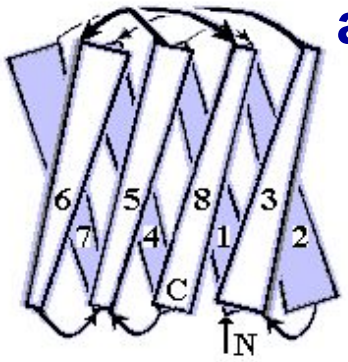
GREEK KEY



JELLYROLL

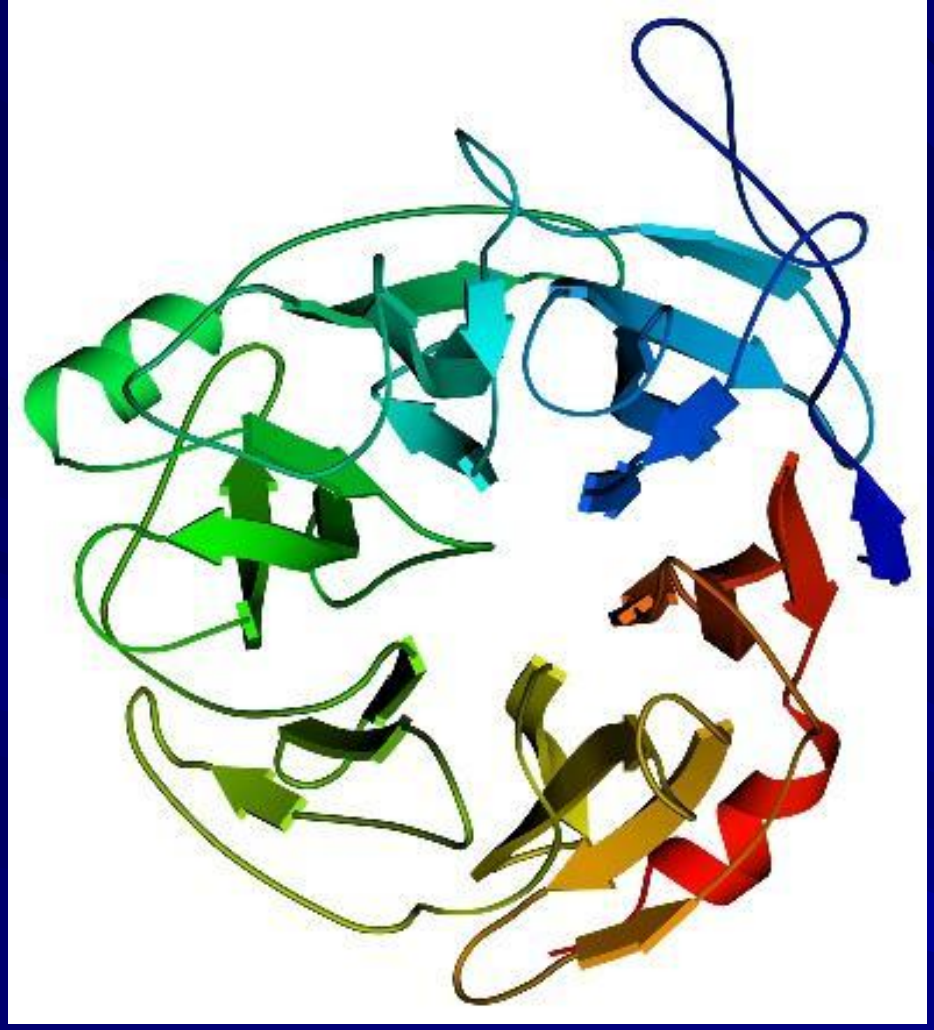
JELLYROLL

aligned packings of β-sheets



a) different: only topologies

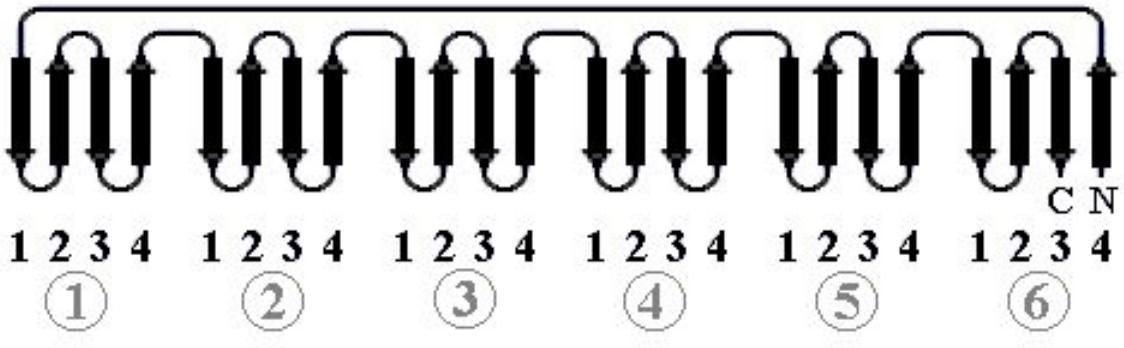
b) equal: even topology



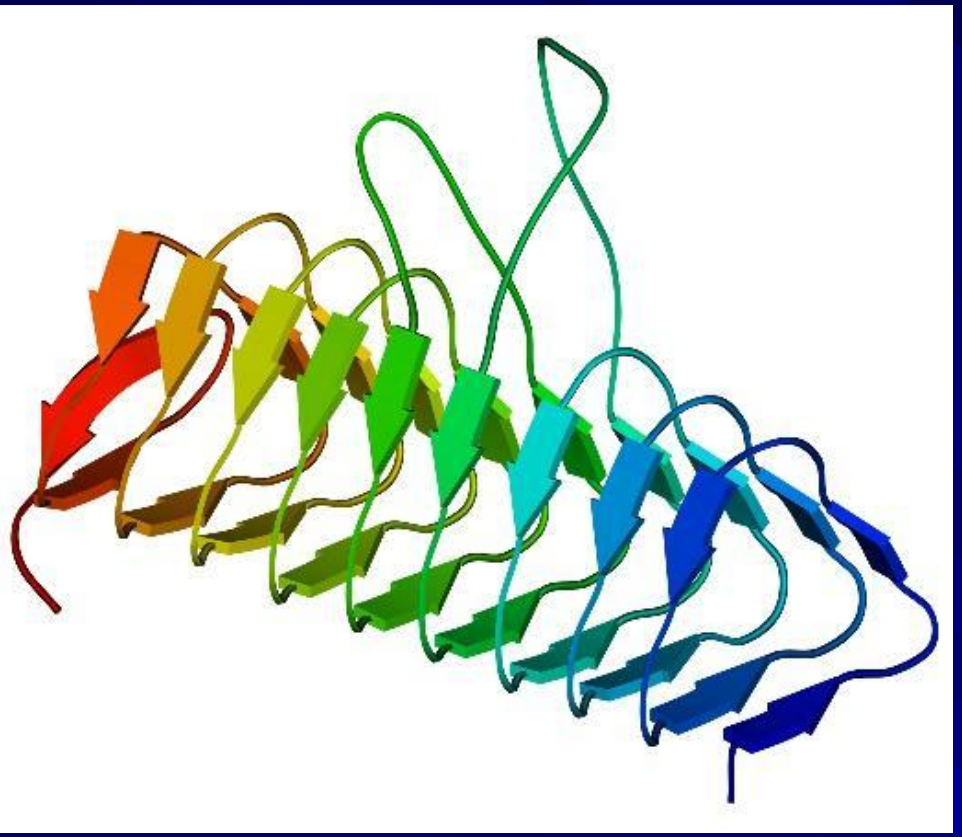
aligned packing
of β-sheets

6-bladed propeller

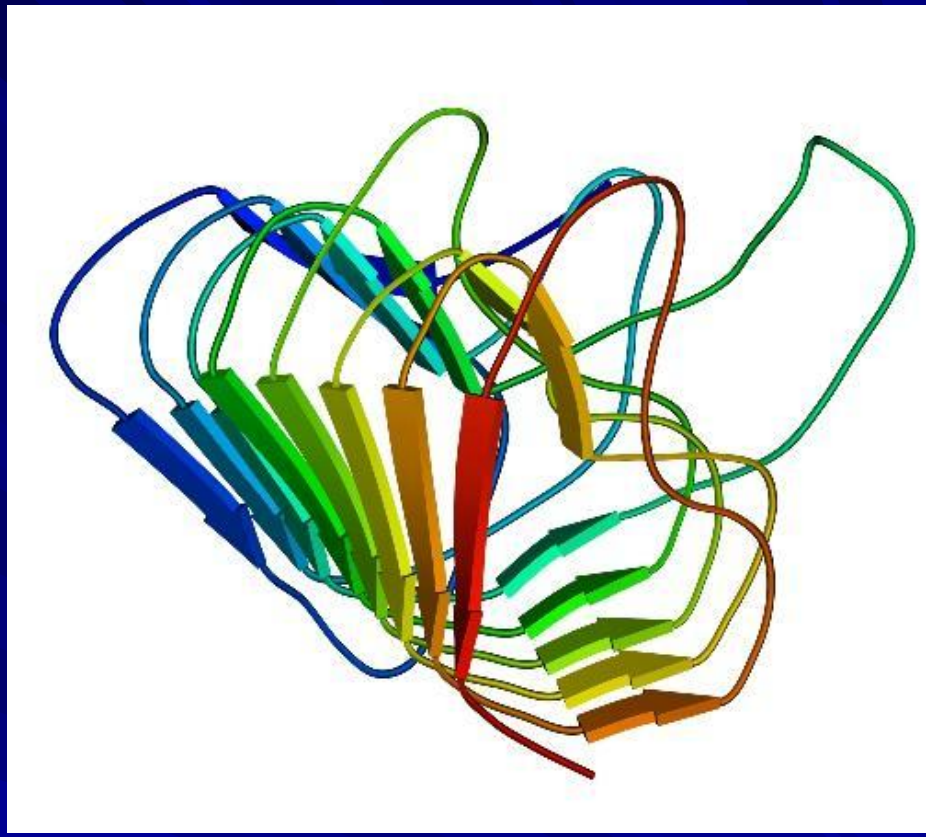
neuraminidase



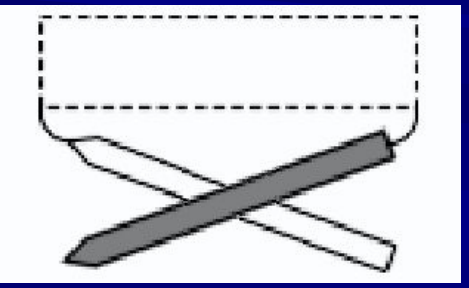
Left-handed β-prism: Acyl transferase



Right-handed β-prism: Pectate lyase

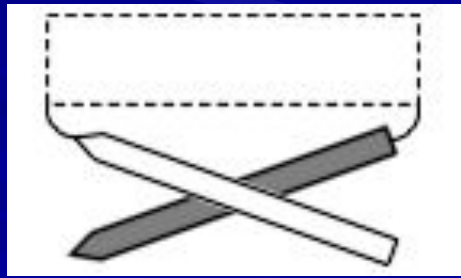


TOPOLOGY of chain turns between parallel β-strands

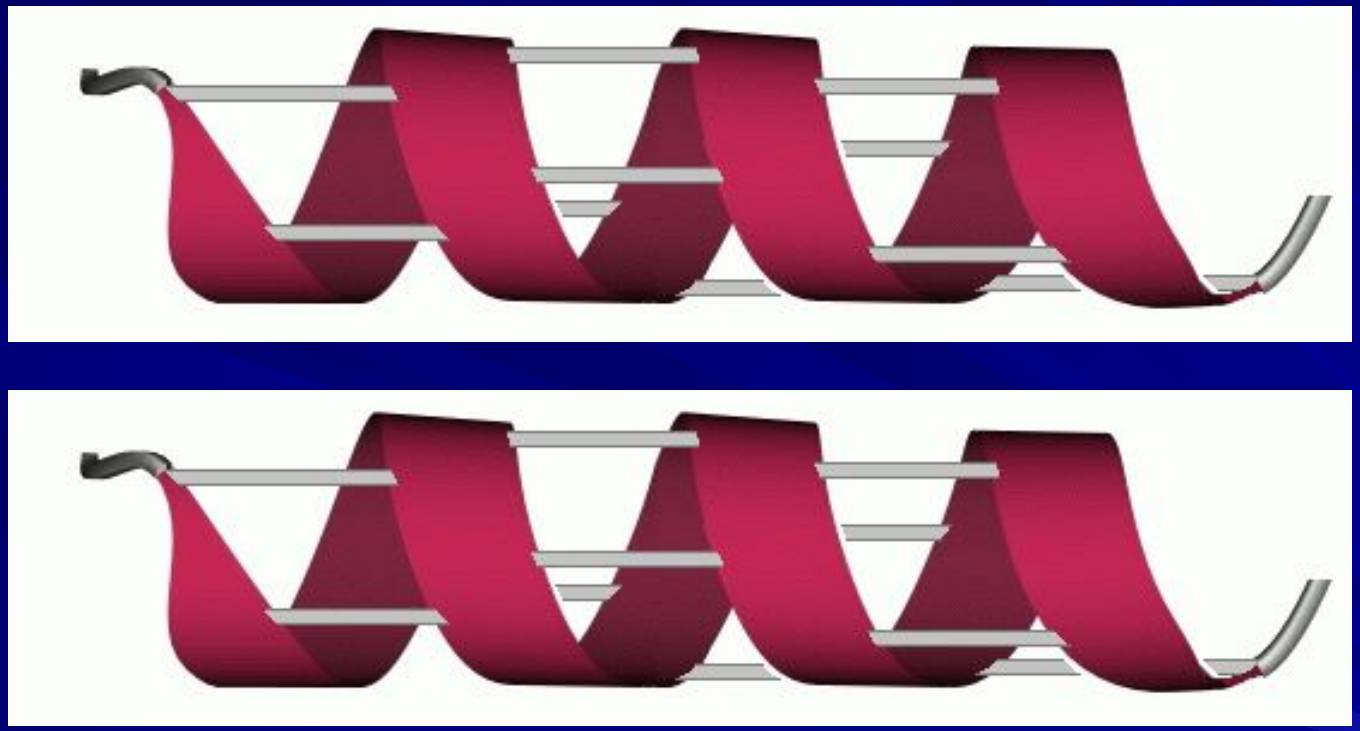


UNusual
LEFT-HANDED
chain turns
(AND NO
β−TWIST!)

Usual
RIGHT-HANDED
chain turns
(AND RIGHT
β-TWIST!)



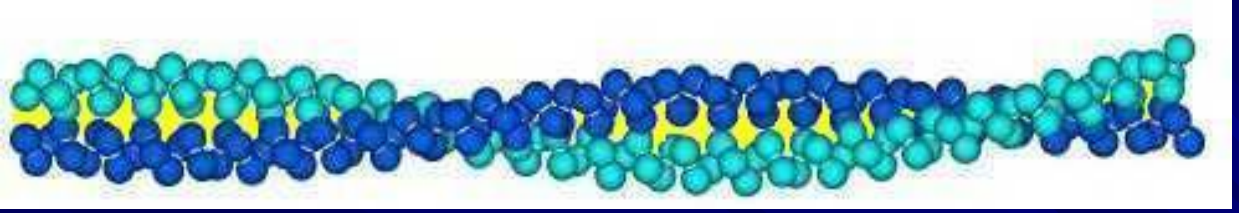
α-proteins

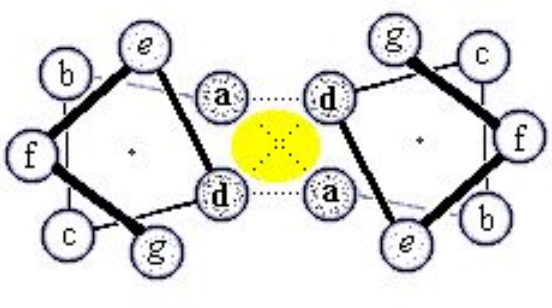


H-bonds: within helices

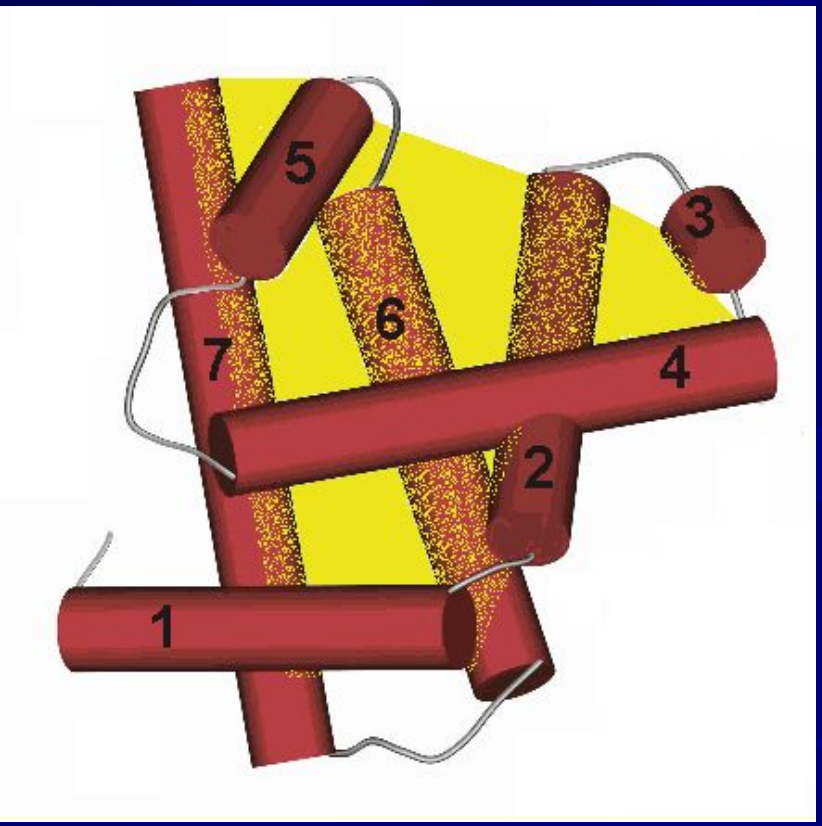
&

Hydrophobics: between helices

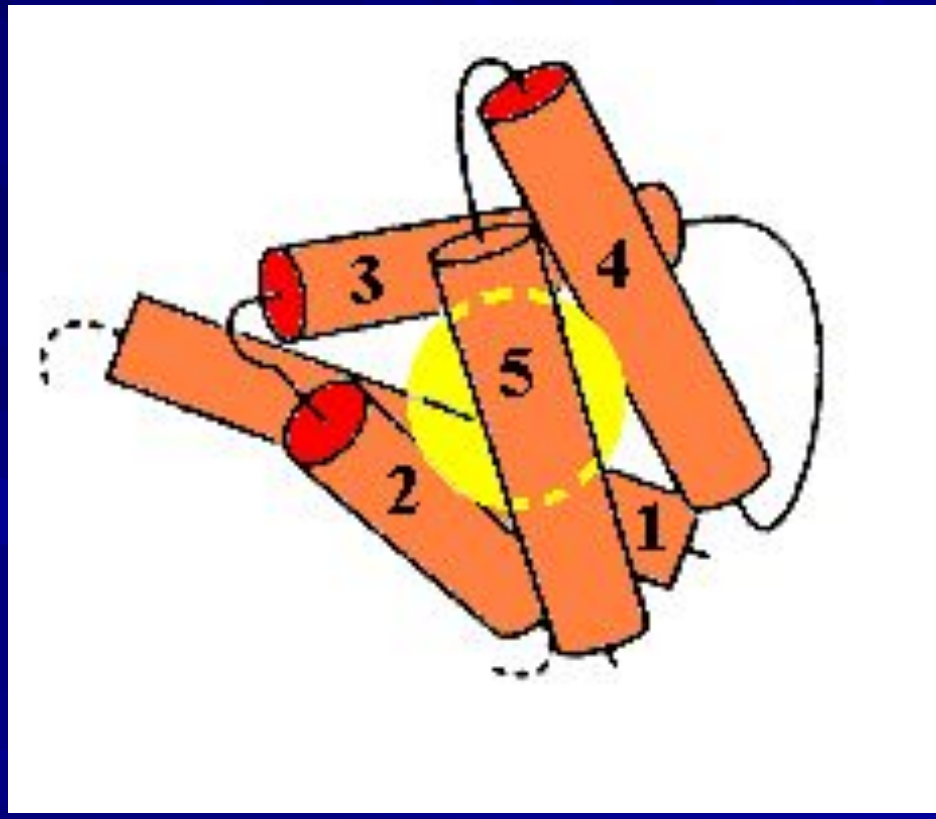




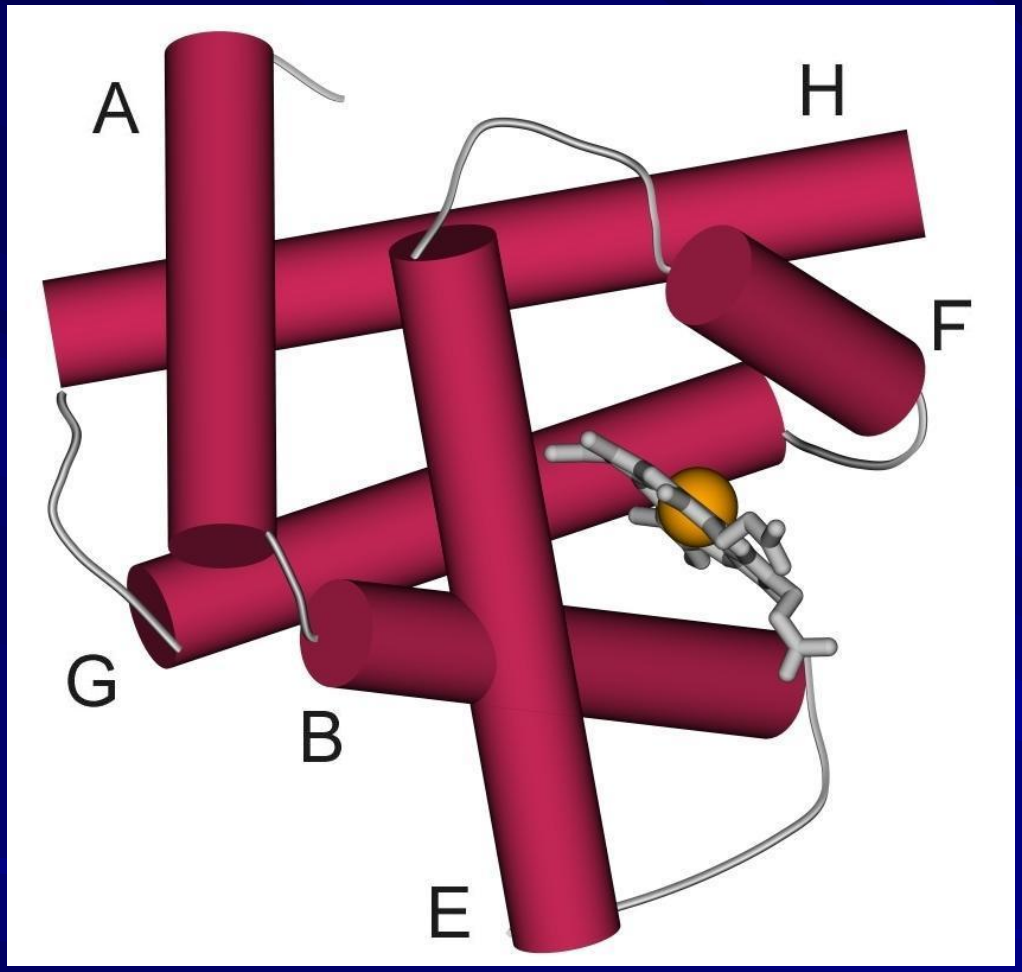
Quasi-cylindrical core (in fibrous)

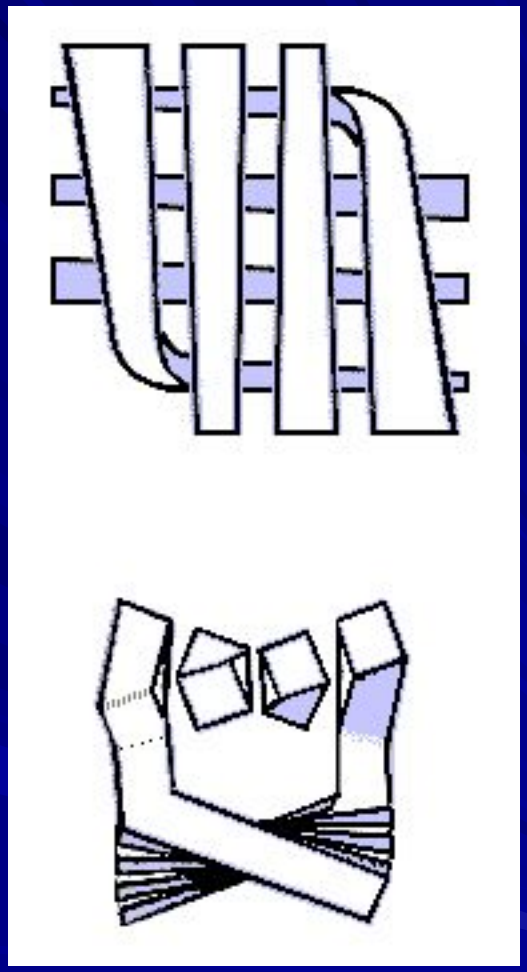


Quasi-flat core



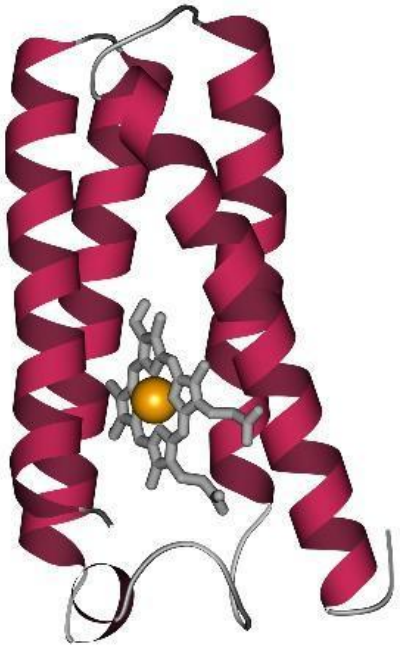
Quasi-spherical core
MOST COMMON



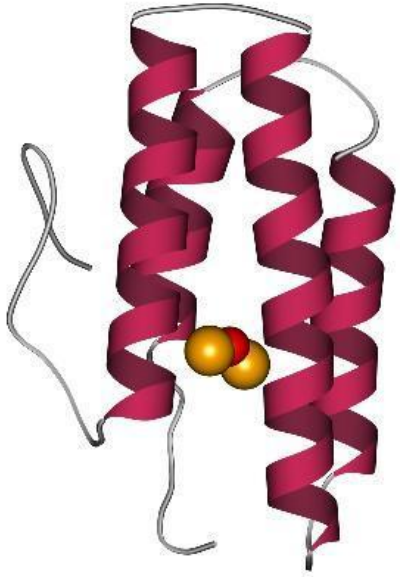


Orthogonal packing of LONG α-helices

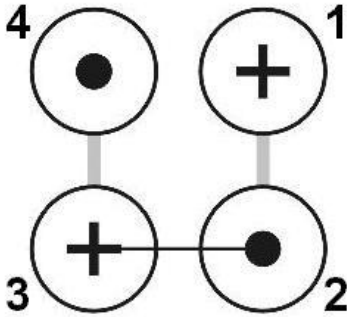
Similar to orthogonal packing of β-sheets



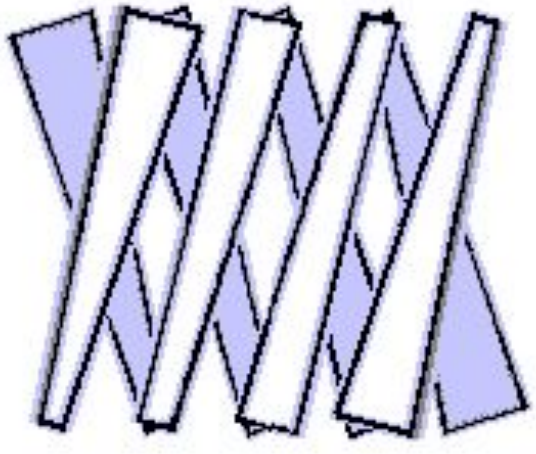
Cytochrome c'

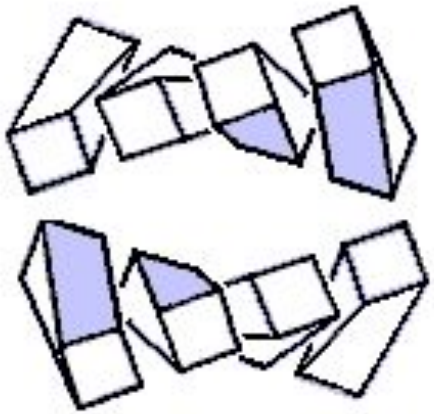


Hemerythrin

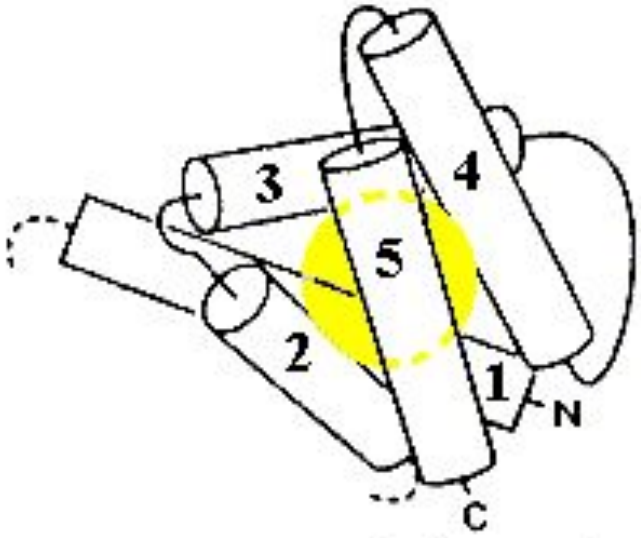


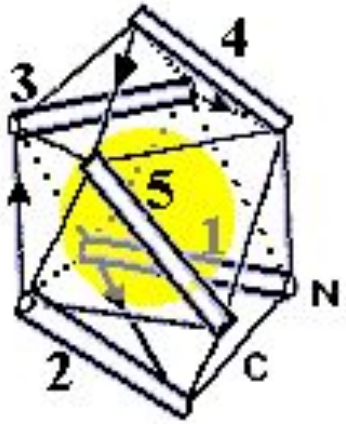
Aligned packing of LONG α-helices



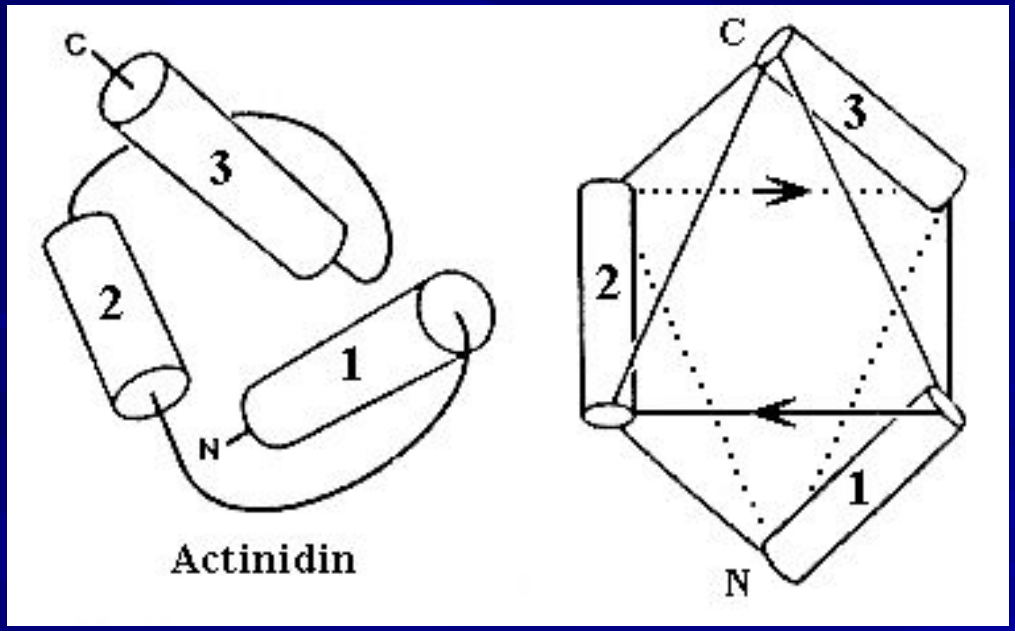


Similar to aligned packing of β-sheets





Thermolysin, domain 2



Actinidin

Quasi-spherical core:

MOST COMMON

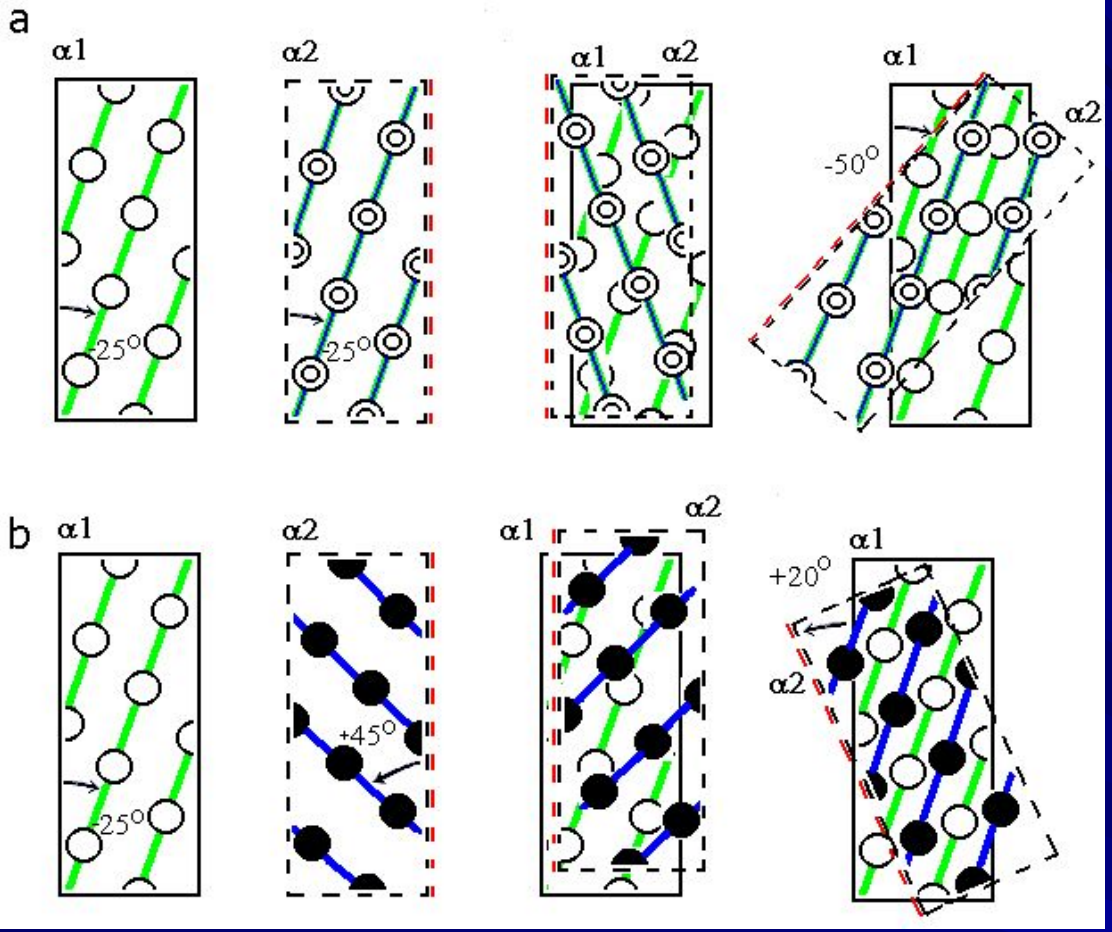
Quasi-spherical polyhedra

no loop turns of ~360°

no loop crossings

CLOSE PACKING

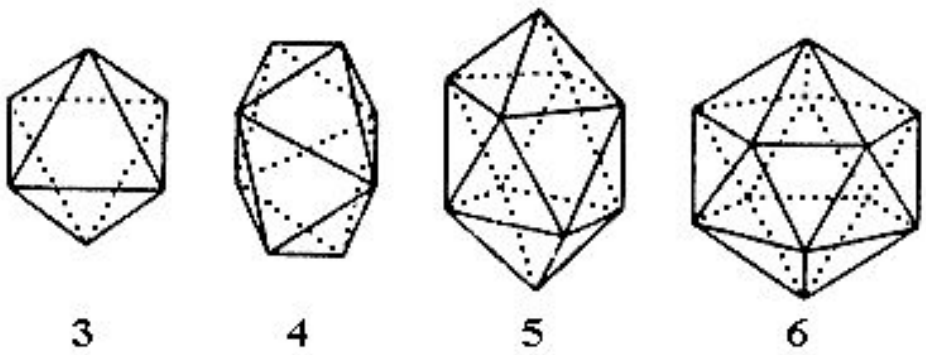
Packing of ridges:



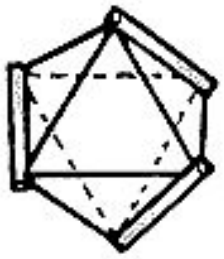
"0-4" & "0-4": -50^0

"0-4" & "1-4": $+20^0$

$-60^0 \approx -50^0$ $+60^0 \neq +20^0$



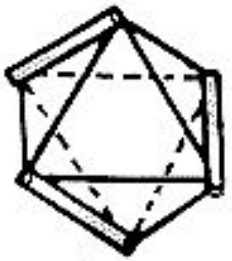
IDEAL POLYHEDRA



right-

(-60° between helices)

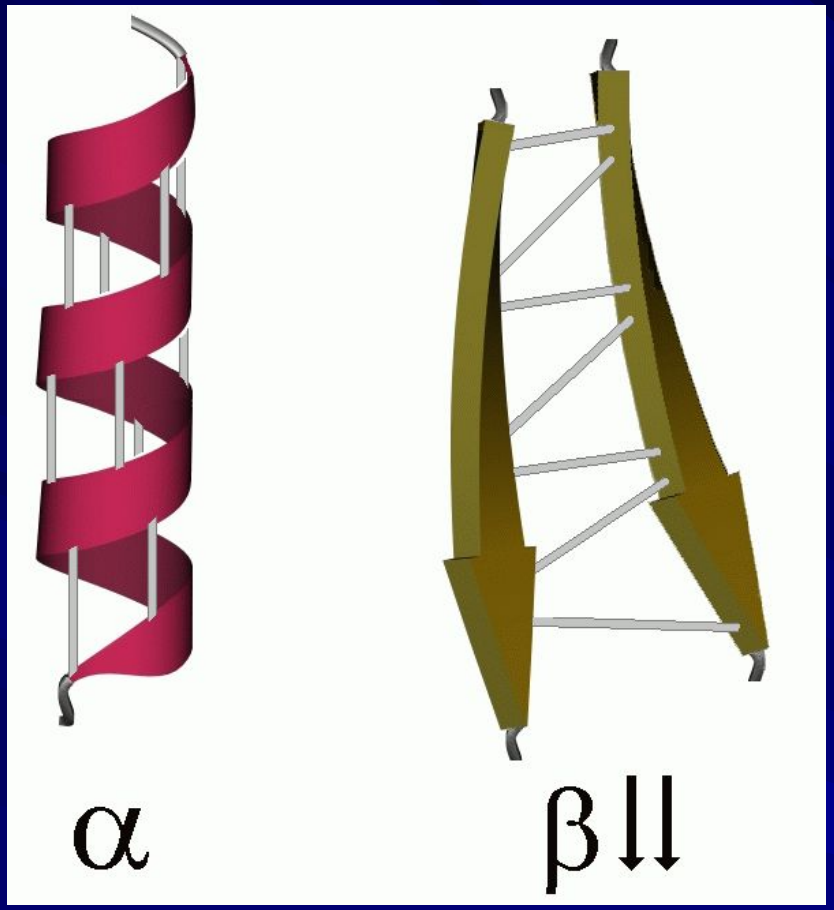
* often *



left-

(+60° between helices)

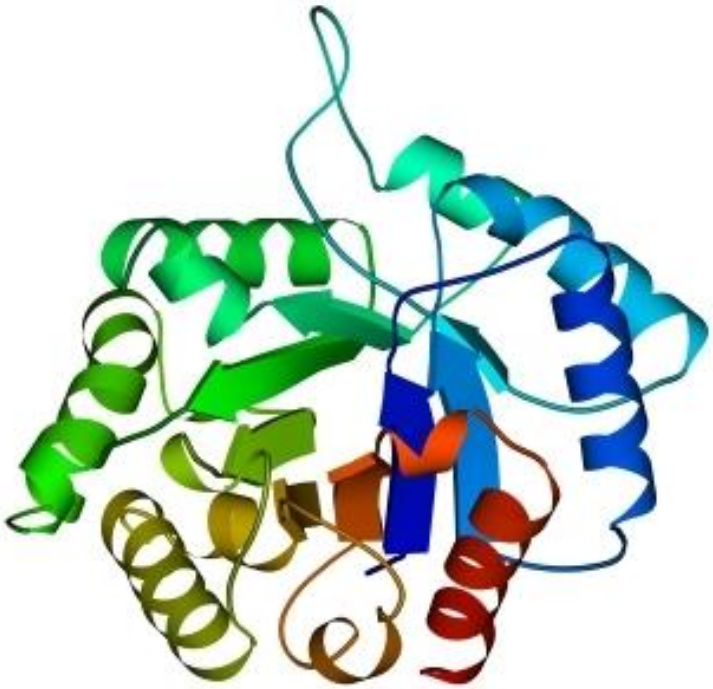
rare

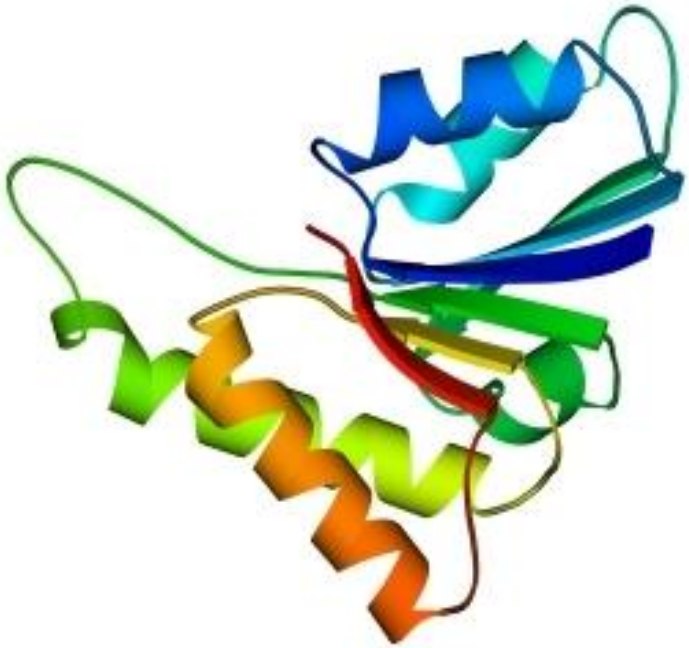


α/β proteins

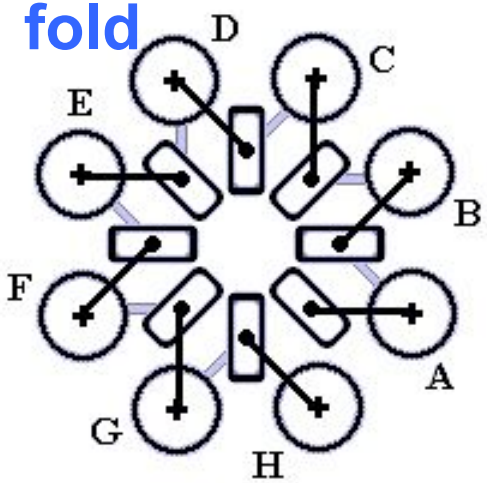
H-bonds: within helices & sheets

Hydrophobics: between helices & sheets

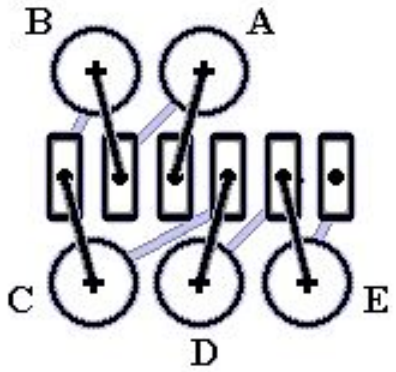




a **TIM barrel fold**

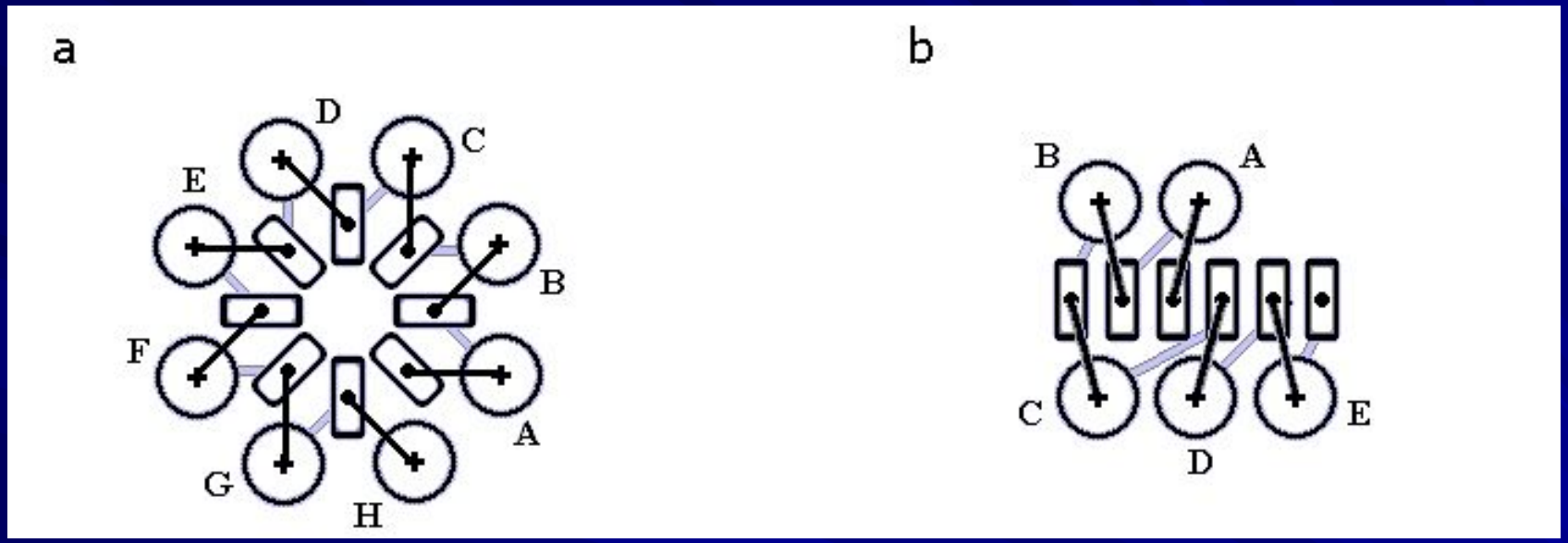


b **Rossmann**



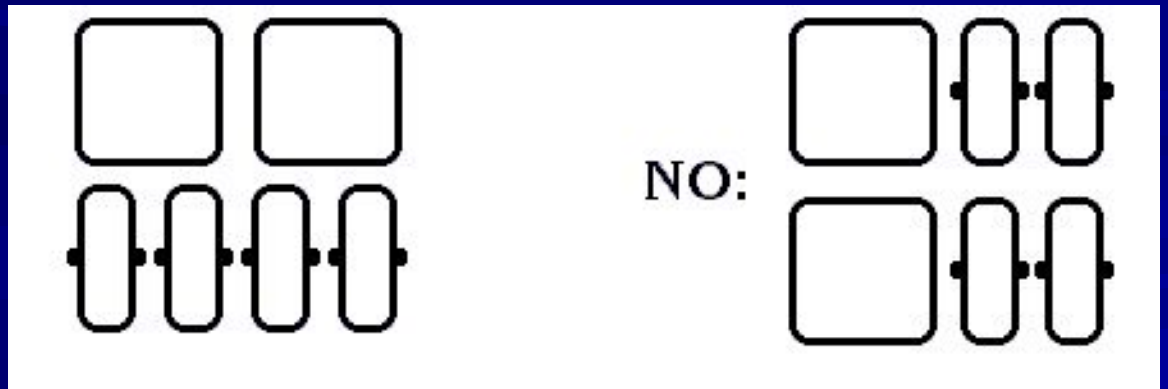
Regular secondary structure sequence:

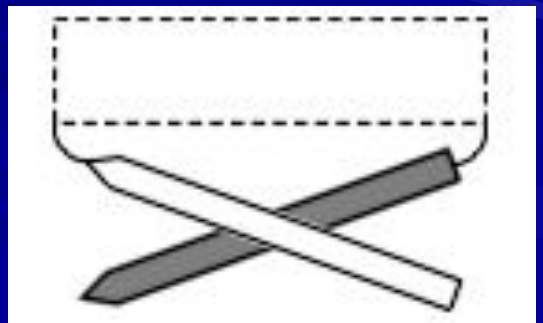
β – α – β – α – β – α – β – α – β – ...

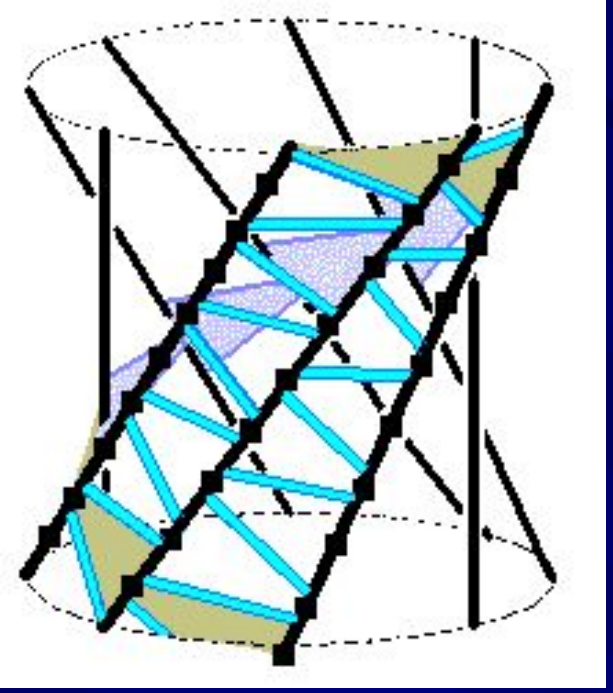


α and β layers

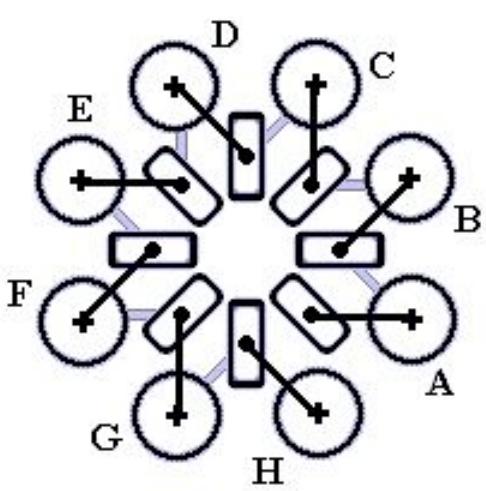
right-handed superhelices



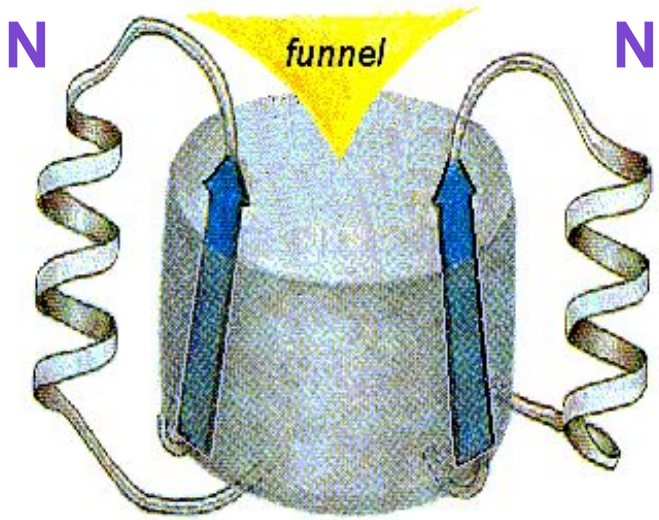


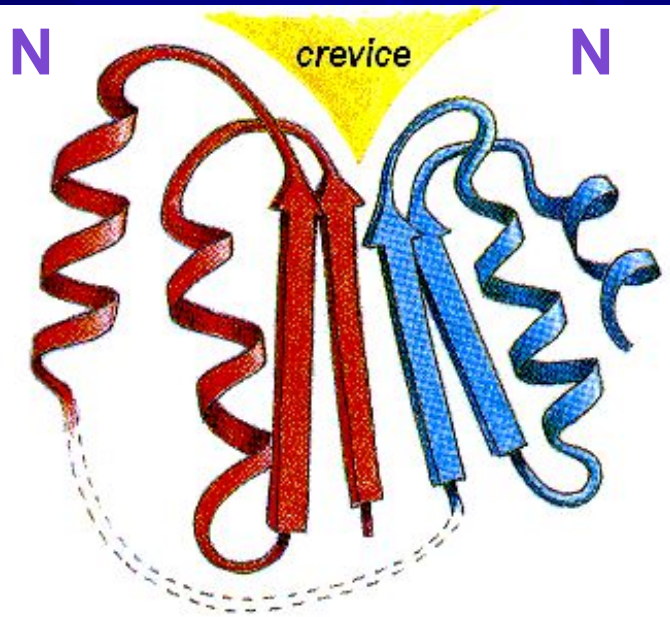


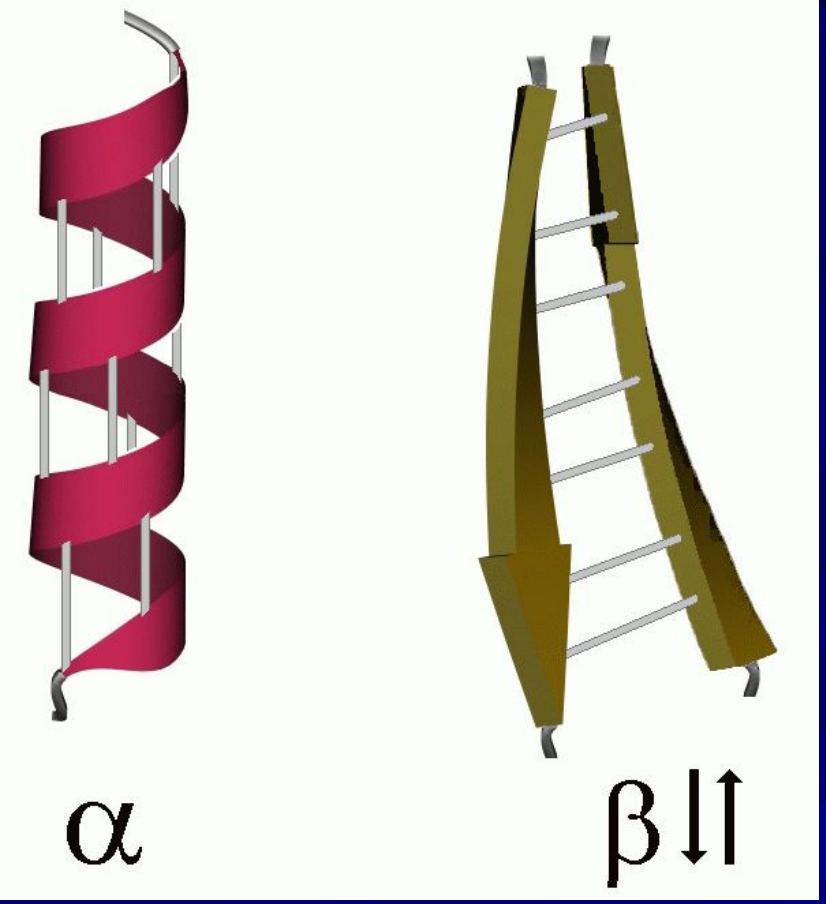
Classification of β-barrels:
"share number" S and
strand number N.
Here: S=8, N=8



Standard active site position is given by the archi-tecture







α+β proteins

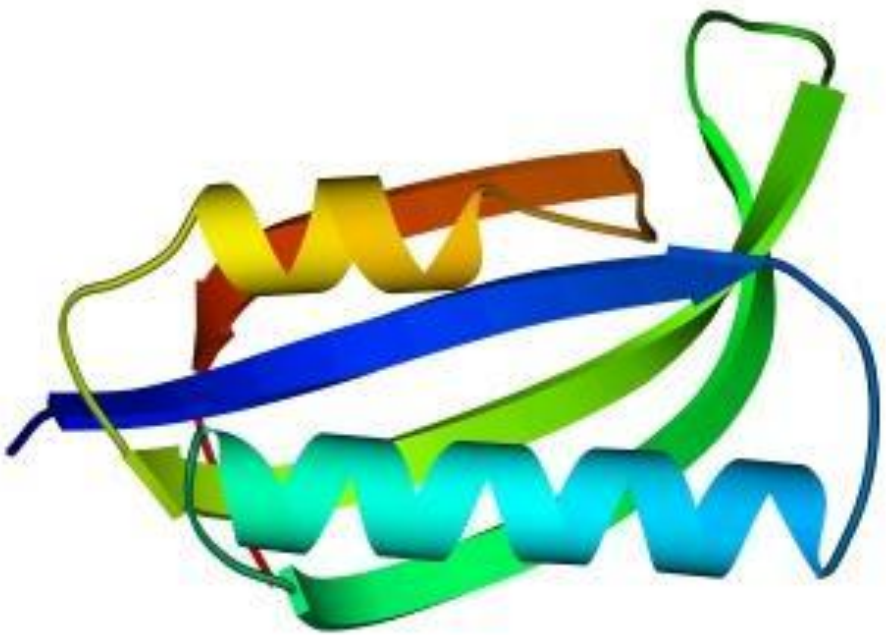
H-bonds: within helices & sheets

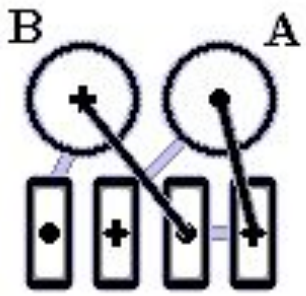
Hydrophobics: between helices & sheets

α+β:

a) A kind of regularity in the secondary structure sequence:

β – α – β – β – α – β ...





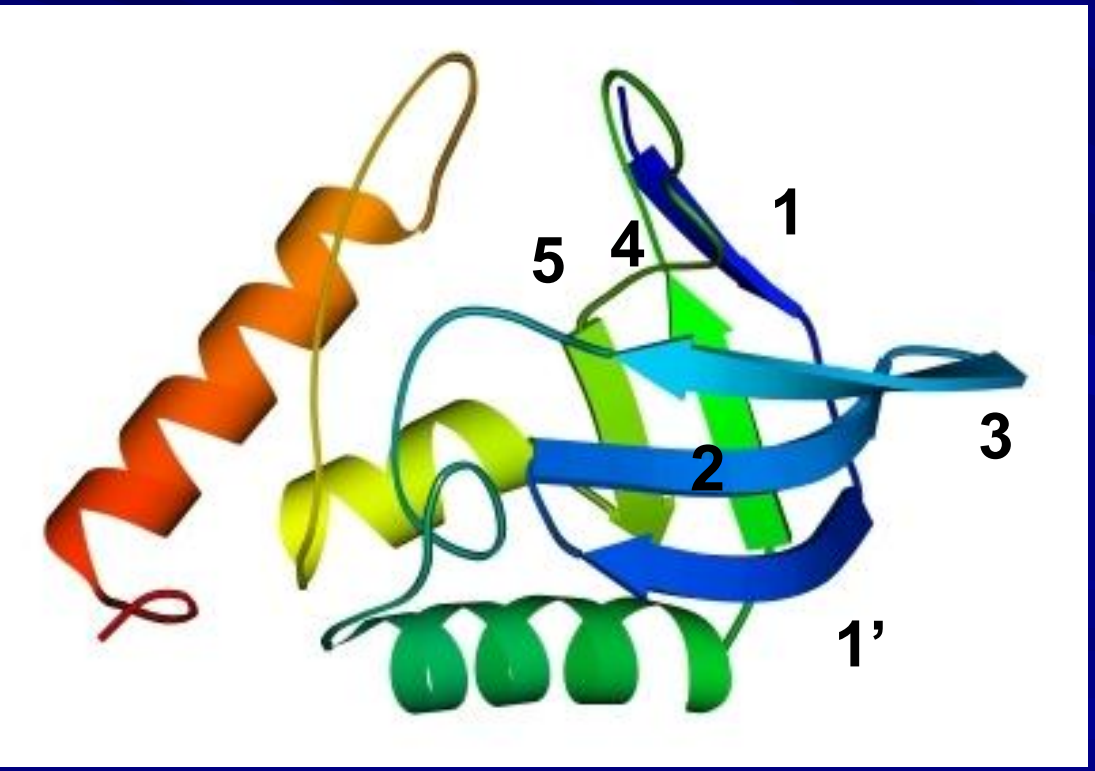
Ferridoxin fold

α+β:

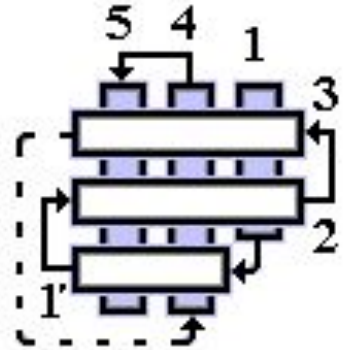
b) Secondary structure sequence:

composed of irregular blocks, e.g.:

β – β – β – β – β – α – β – β – α – α ...



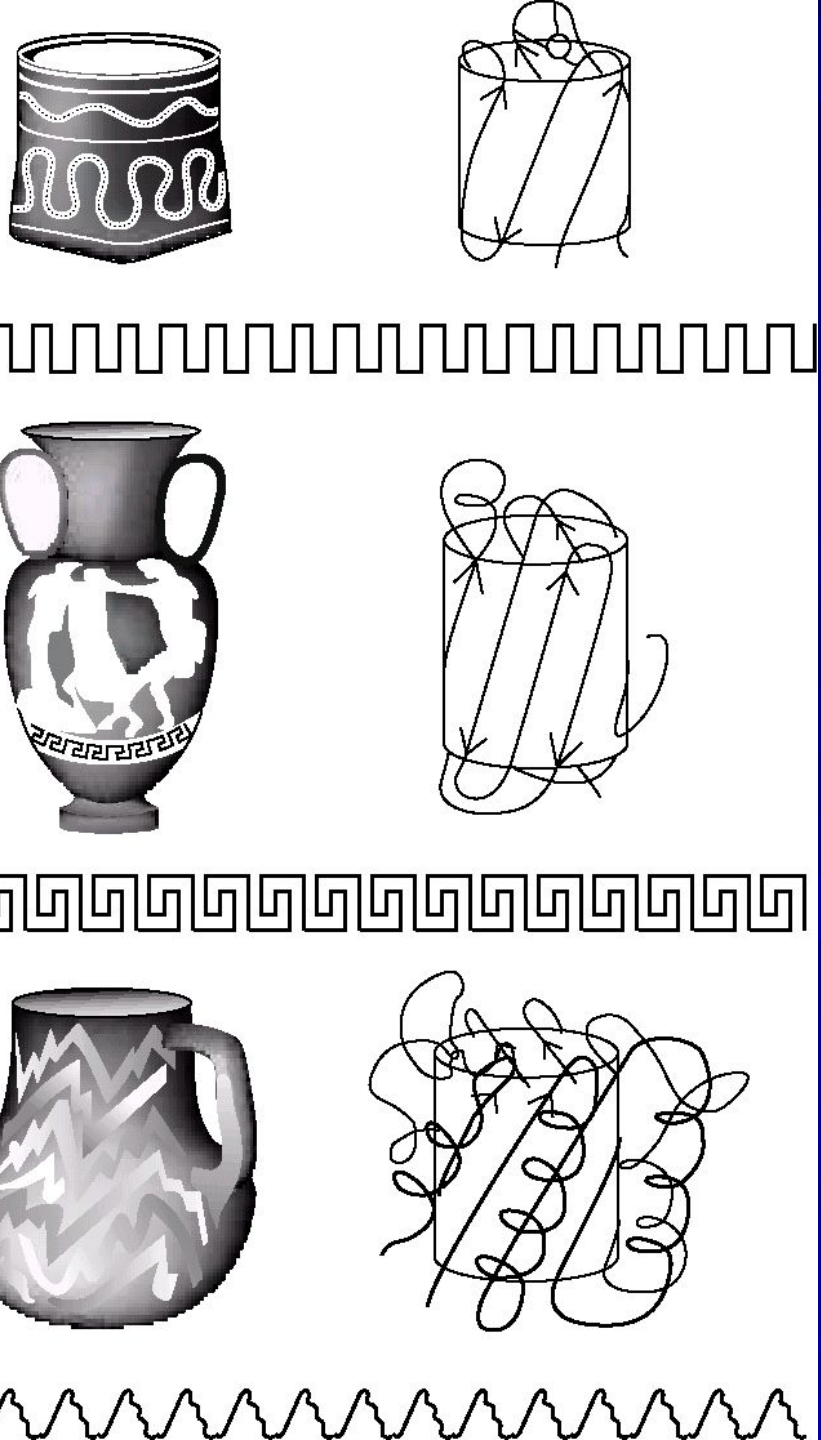
Nuclease fold



OB-fold

of the β-subdomain of nuclease

("Russian doll effect")

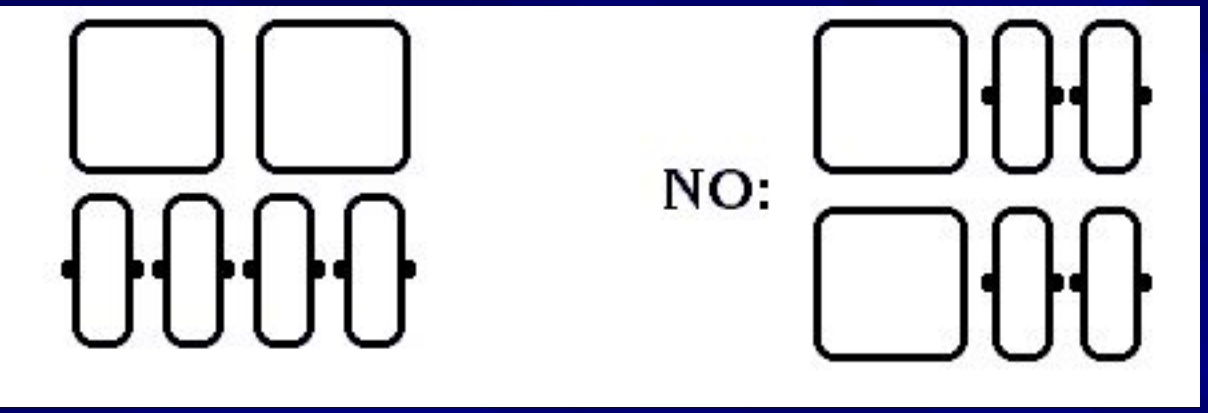


TYPICAL FOLDING PATTERNS (1977)

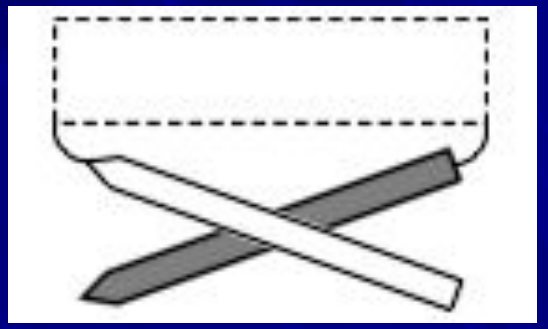
Jane Shelby **Richardson**, 1941

EMPIRICAL RULES

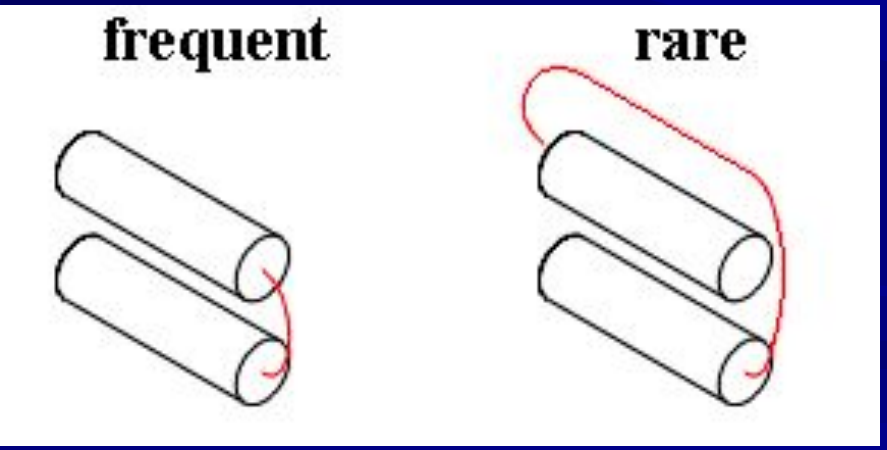
separate α and β layers

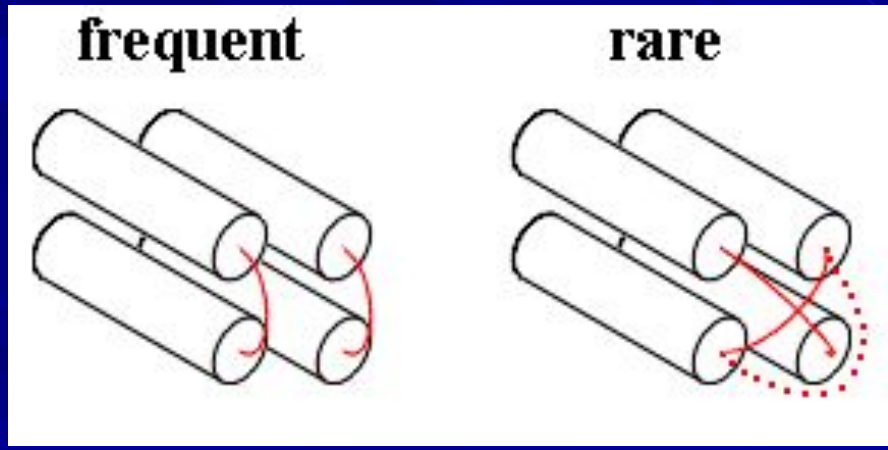


right-handed superhelices



Lost H-bonds: defect!





no large, ~360º turns

NO 'defects'

no loop crossings

RESULT:

NARROW SET
OF PREDOMINANT FOLDING PATTERNS

these are those that have no ‘defects’

ALSO,

these are “natively disordered proteins”,

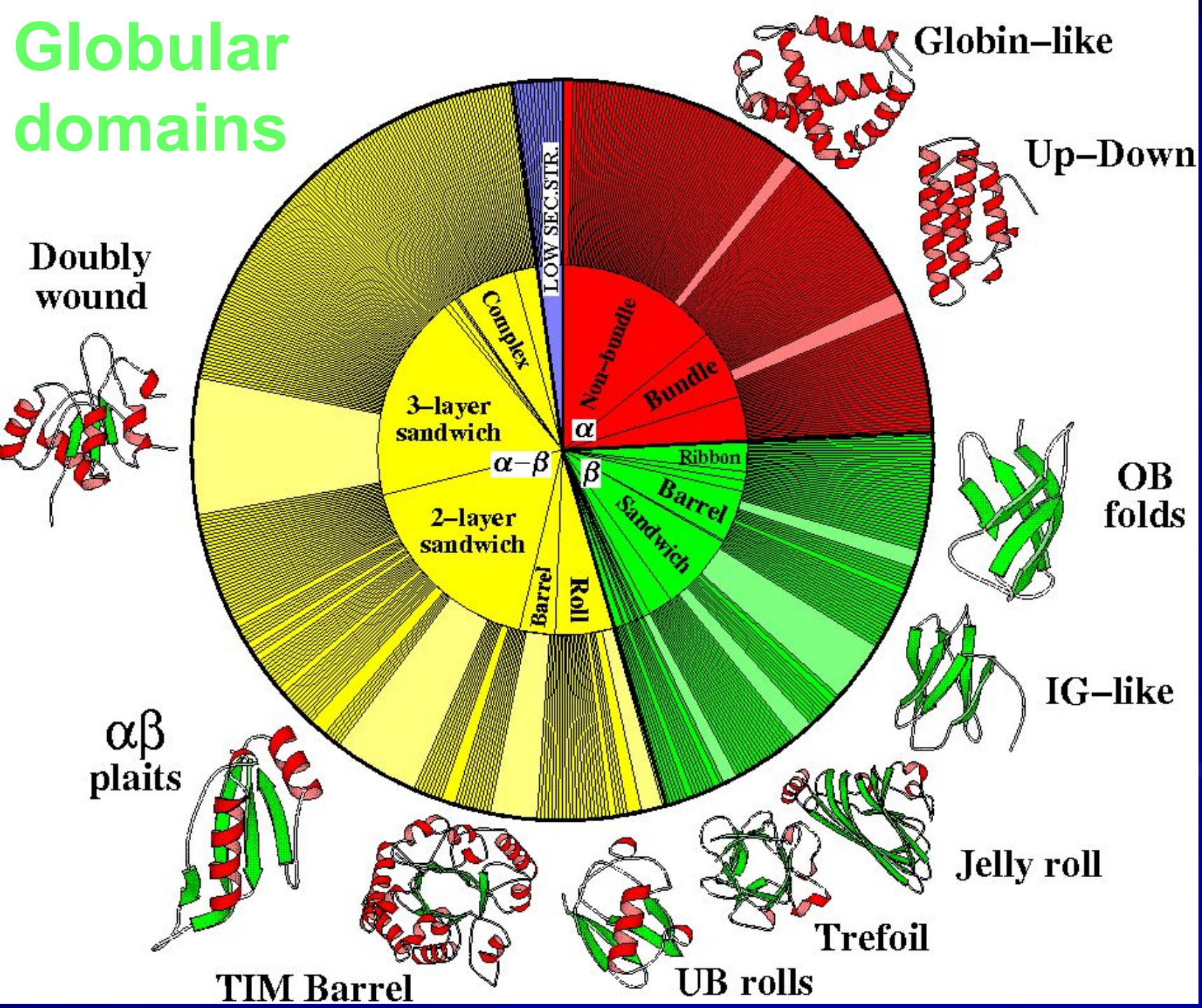
which form a definite structure

only when bound

to some another molecule

(ligand, DNA, protein…)

Globular domains



CATH ≈ SCOP

Classification of 3D protein folds

Алексей Григорьевич **Мурзин**, 1956

SCOP

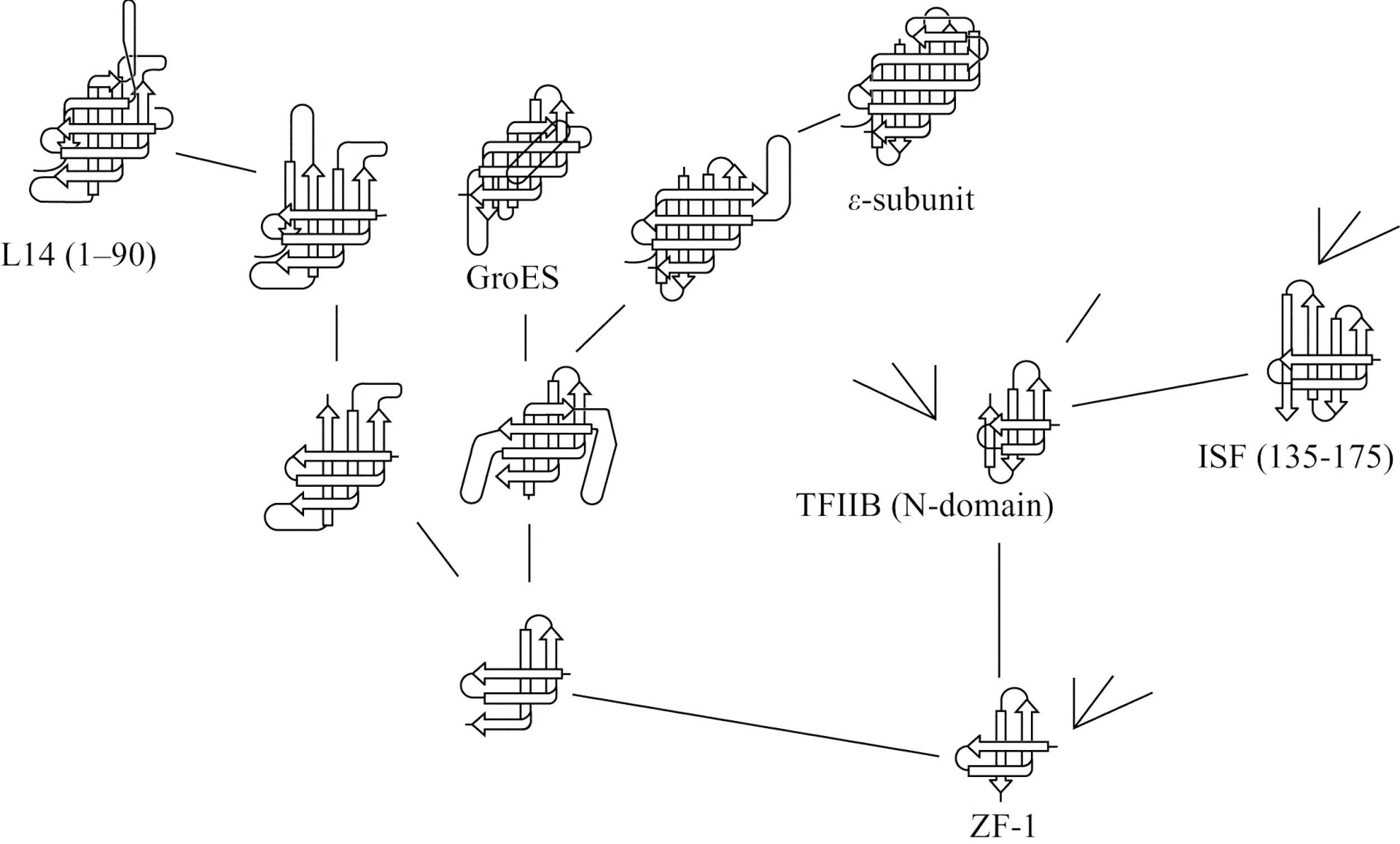
Cyrus Homi **Chothia**, 1942

CATH

Dame Janet Maureen **Thornton**, 1949

«Деревья»

Александр Васильевич **Ефимов**, 1954



Efimov's "trees"

80/20 LAW:

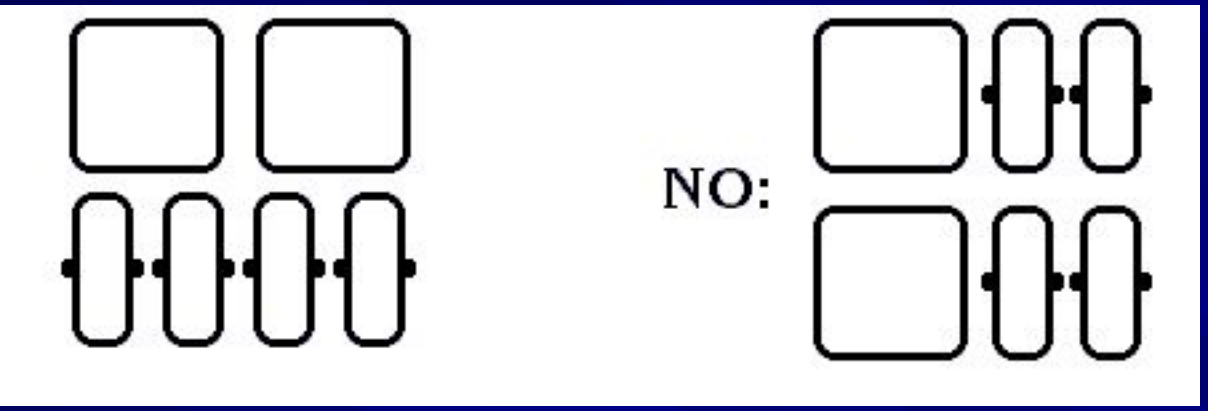
80% OF BEER IS CONSUMED BY ONLY
20% OF THE POPULATION

80% OF PROTEINS BELONG TO ONLY
20% OF OBSERVED FOLDS

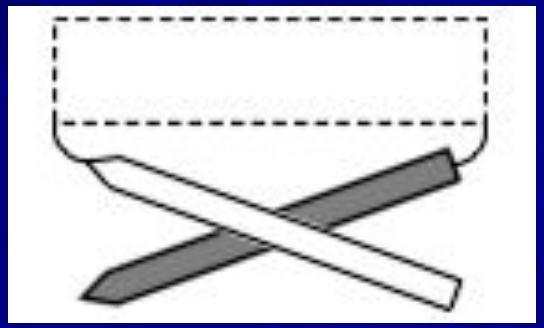
These folds are "typical" for proteins.
The remaining 20% of proteins are scattered over "unusual" folds, which form 80% of observed folds

EMPIRICAL RULES for FREQUENT FOLDS

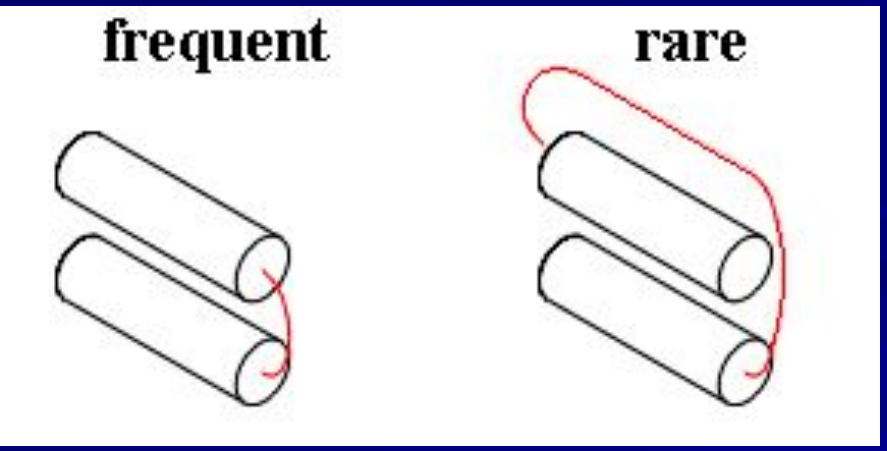
α and β structures,
separate α and β layers



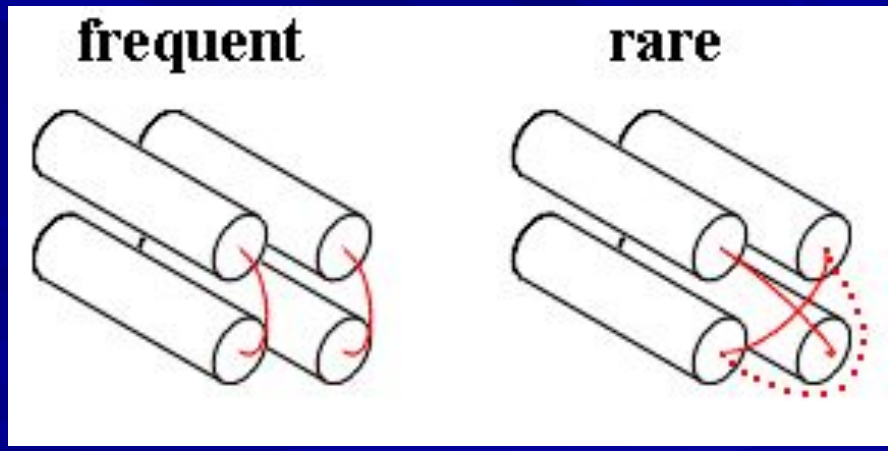
right-handed
superhelices



Lost H-bonds: defect!

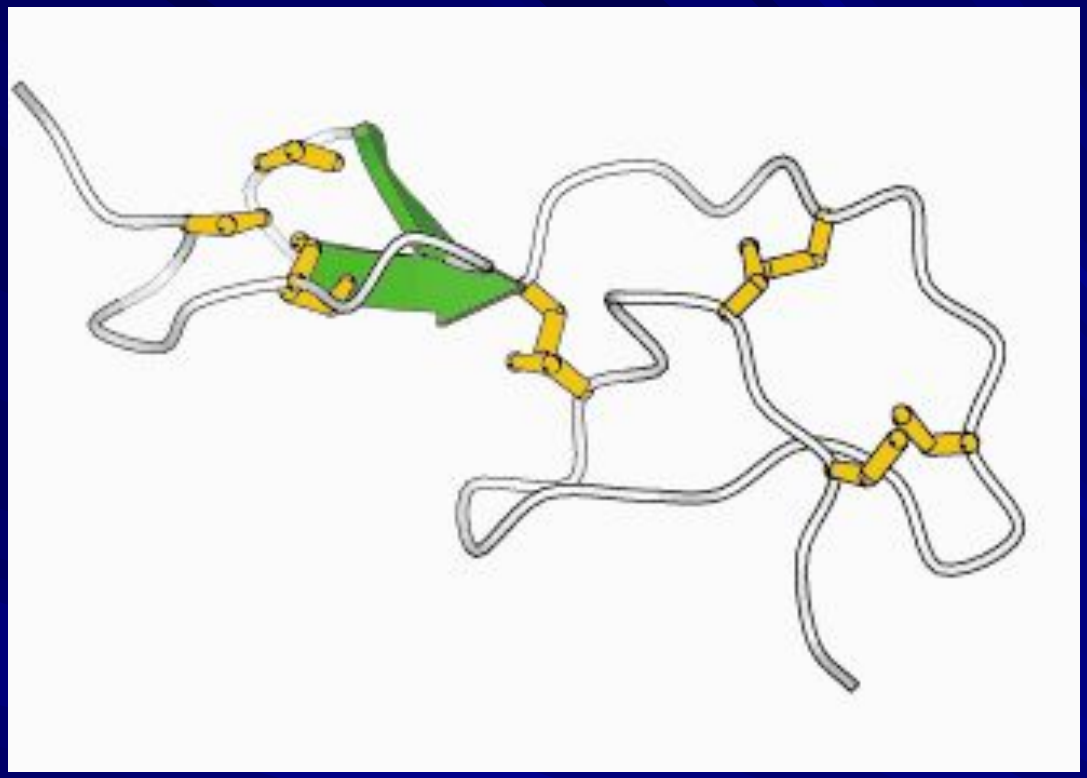


no large (360-degree) turns



no loop crossing

e.g.:

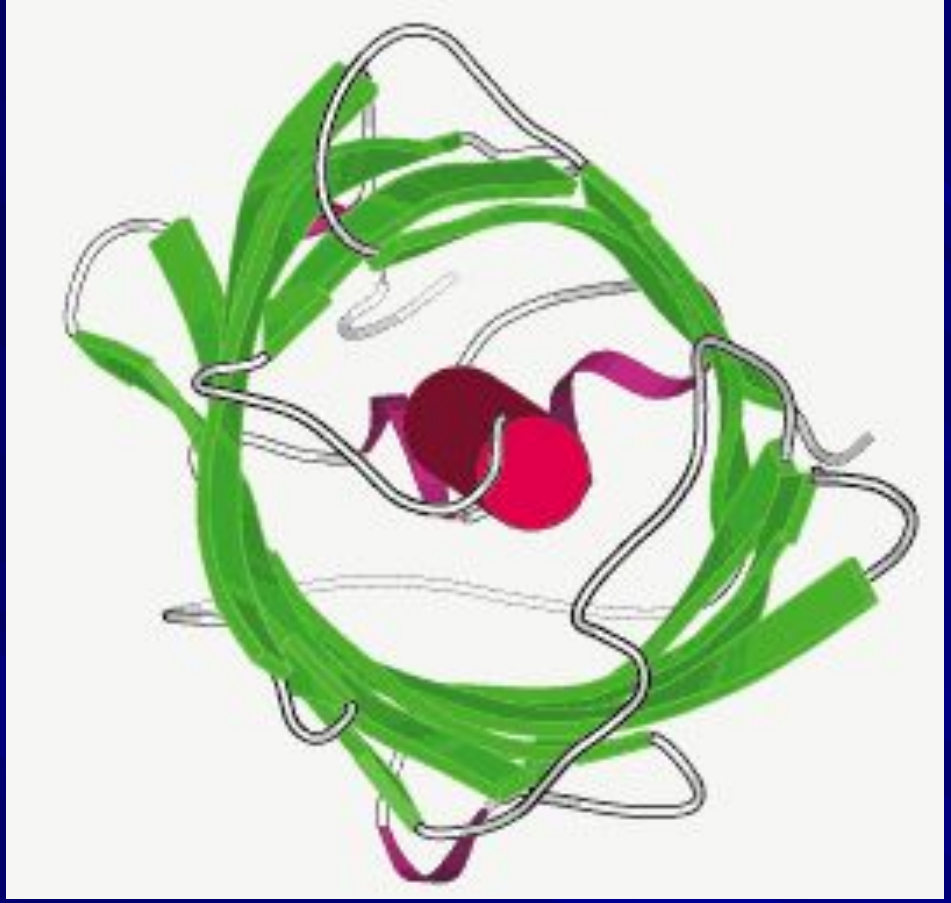


Unusual fold

(no α, almost no β structure: bad for stability) -

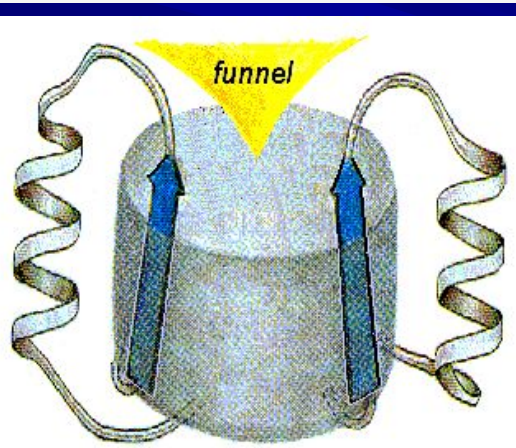
BUT: very special sequence

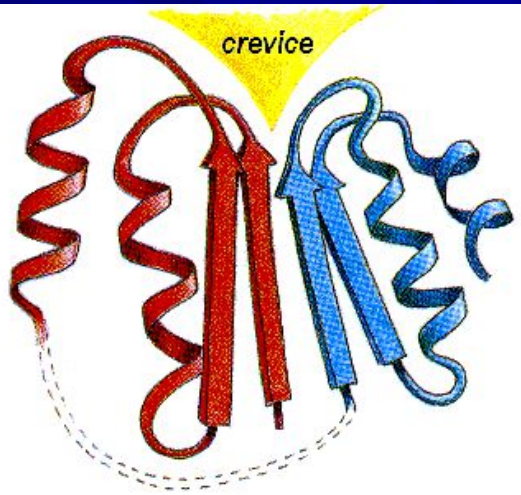
(very many Cysteins, and therefore very many S-S bonds)



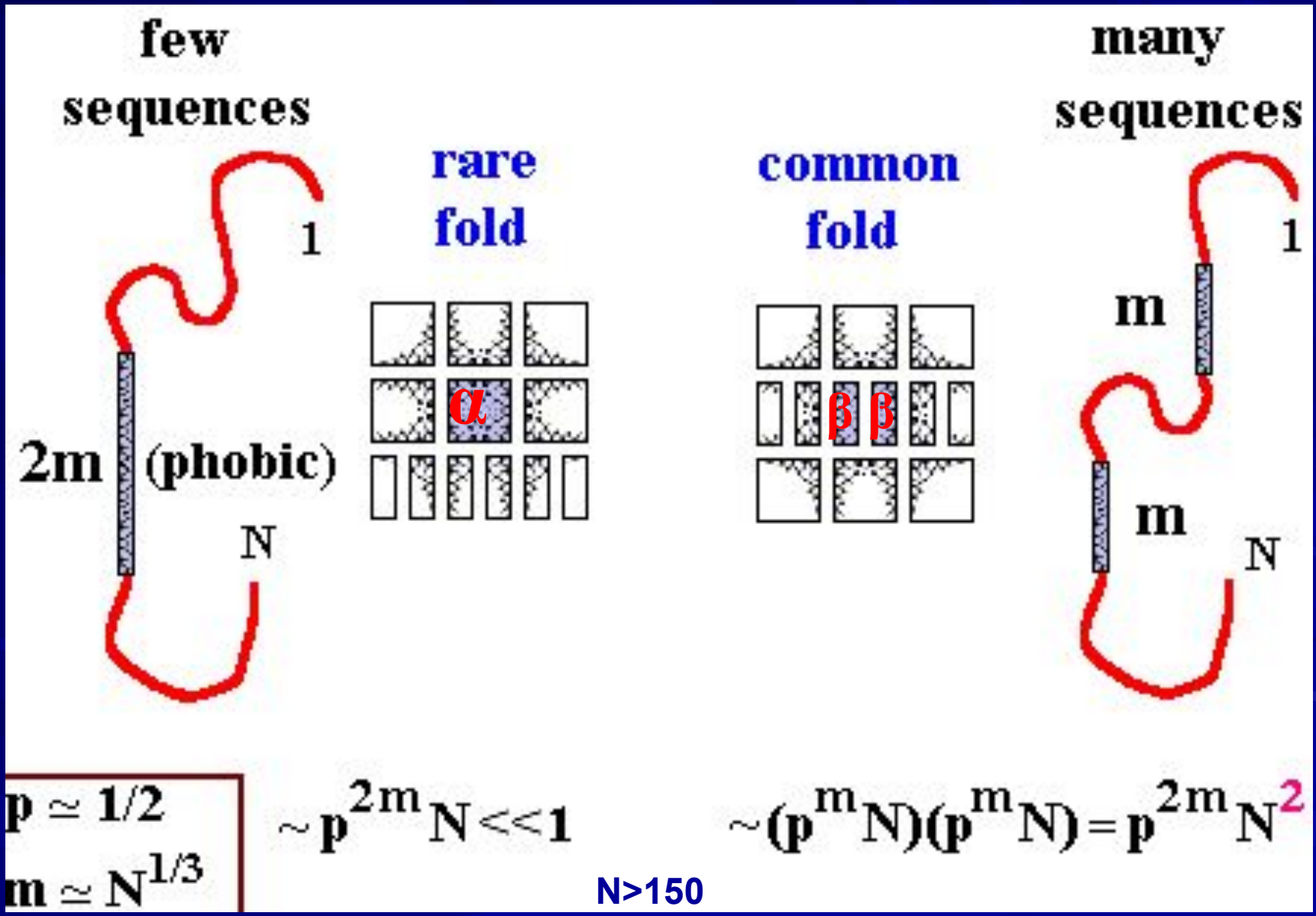
Unusual fold (GFP): helix inside

Usual folds: helices outside



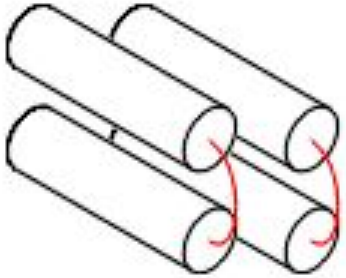


What is more usual: sequence providing α inside or β β inside?

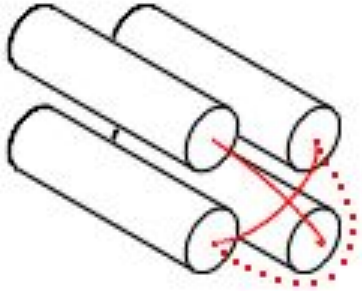


Selected elements

common



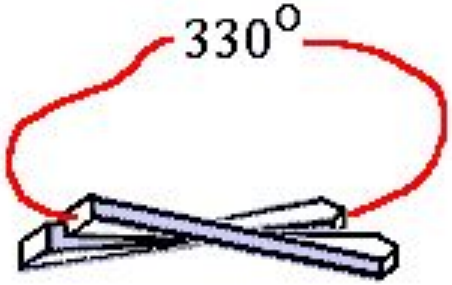
rare



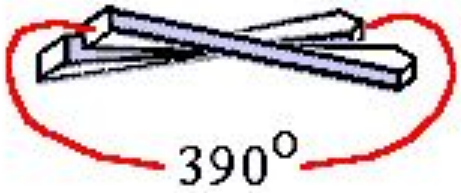
Defect: Loop crossing.

Either lost H-bond (energy defect, $\simeq +2$ kcal/mol $\simeq +3$kT), or additional bending, or special sequence

common



rare



Defect: additional bending.

$\simeq +3$kT. Entropic defect ?!

~ 20 times less conformations

= 20 times less chance to include the lowest-enegy conformation

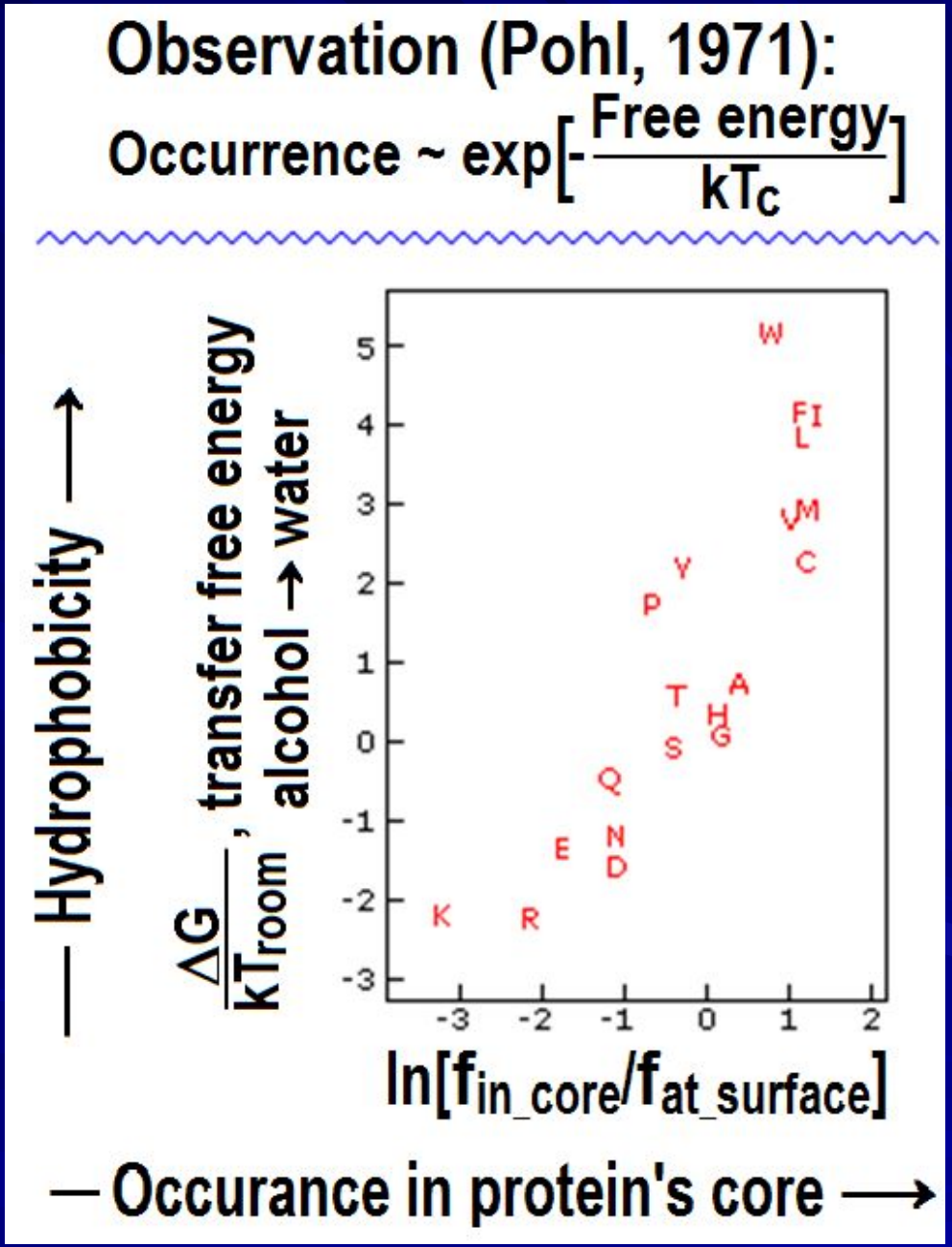
free energy DEFECTS

common folds — NO DEFECTS — common sequences?

rare folds — WITH DEFECTS — rare sequences?

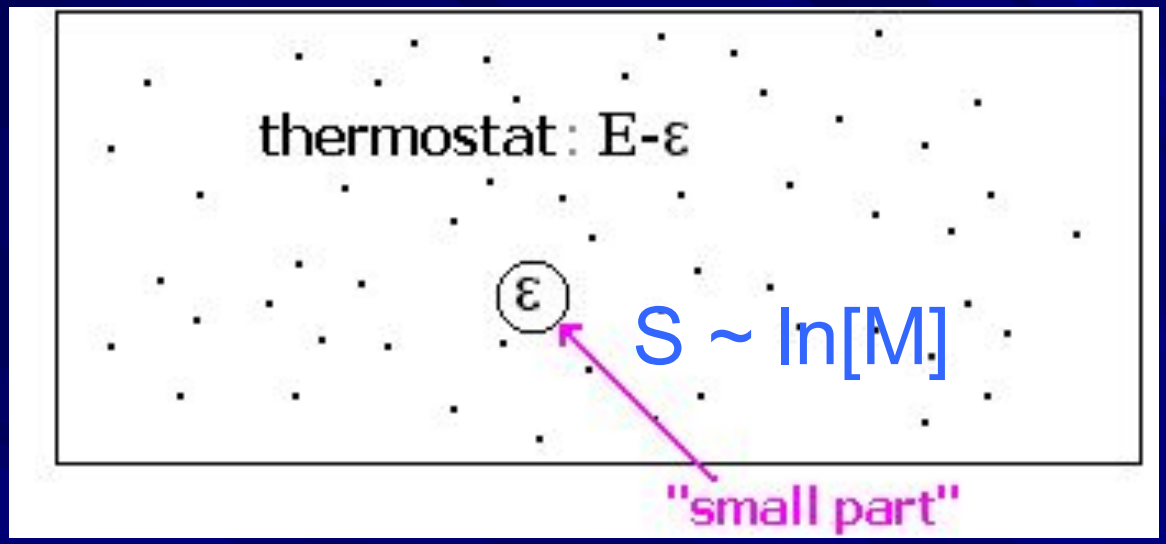
Small protein details

Example:



Miller, Janin, Chothia 1984

WHAT IS "TEMPERATURE"?



THEORY

Closed system: energy E = const

CONSIDER: 1 state of "small part" with ε & all states of thermostat with E-ε. $M(E-\varepsilon) = 1 \bullet M_{th}(E-\varepsilon)$

$S_t(E-\varepsilon) = k \bullet \ln[M_t(E-\varepsilon)] \cong S_t(E) - \varepsilon \bullet (dS_t/dE)|_E$

$M_t(E-\varepsilon) = \exp[S_t(E)/k] \bullet \exp[-\varepsilon \bullet (dS_t/dE)|_E/k]$

Thus: $d[\ln(M_t)]/dE = 1/kT$

Gibbs: $\frac{1}{kT} = \frac{d}{dE} \ln[\#states]$

'state' = *configuration*
E - *its energy*

as well:

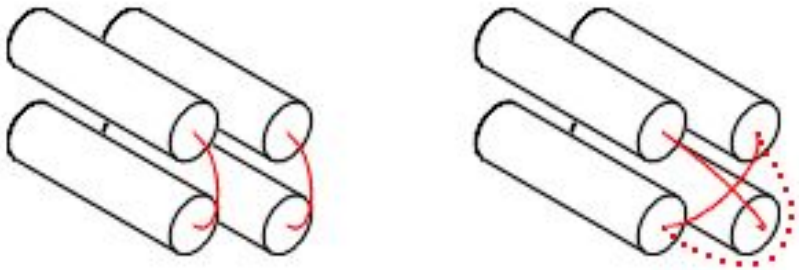
'state' = *a. a. sequence*
E - *the structure's stability for this sequence*

Protein structure is stable,
if its free energy is below some threshold

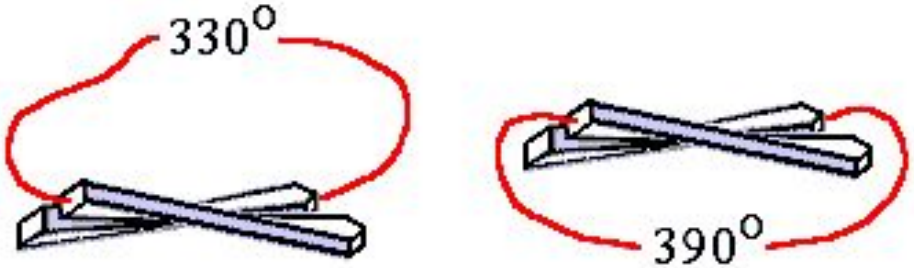
For example:
below that of completely unfolded chain;
or:
below that of any other globular structure

PHYSICAL SELECTION OF FOLDS

common **rare**



common **rare**



free energy DEFECTS

common folds — NO DEFECTS — — common sequences.

rare folds — WITH DEFECTS — — rare sequences.

More stable detail –
more random sequences
Less stable detail –
less random sequences

What's good for protein's detail is good for the whole protein structure

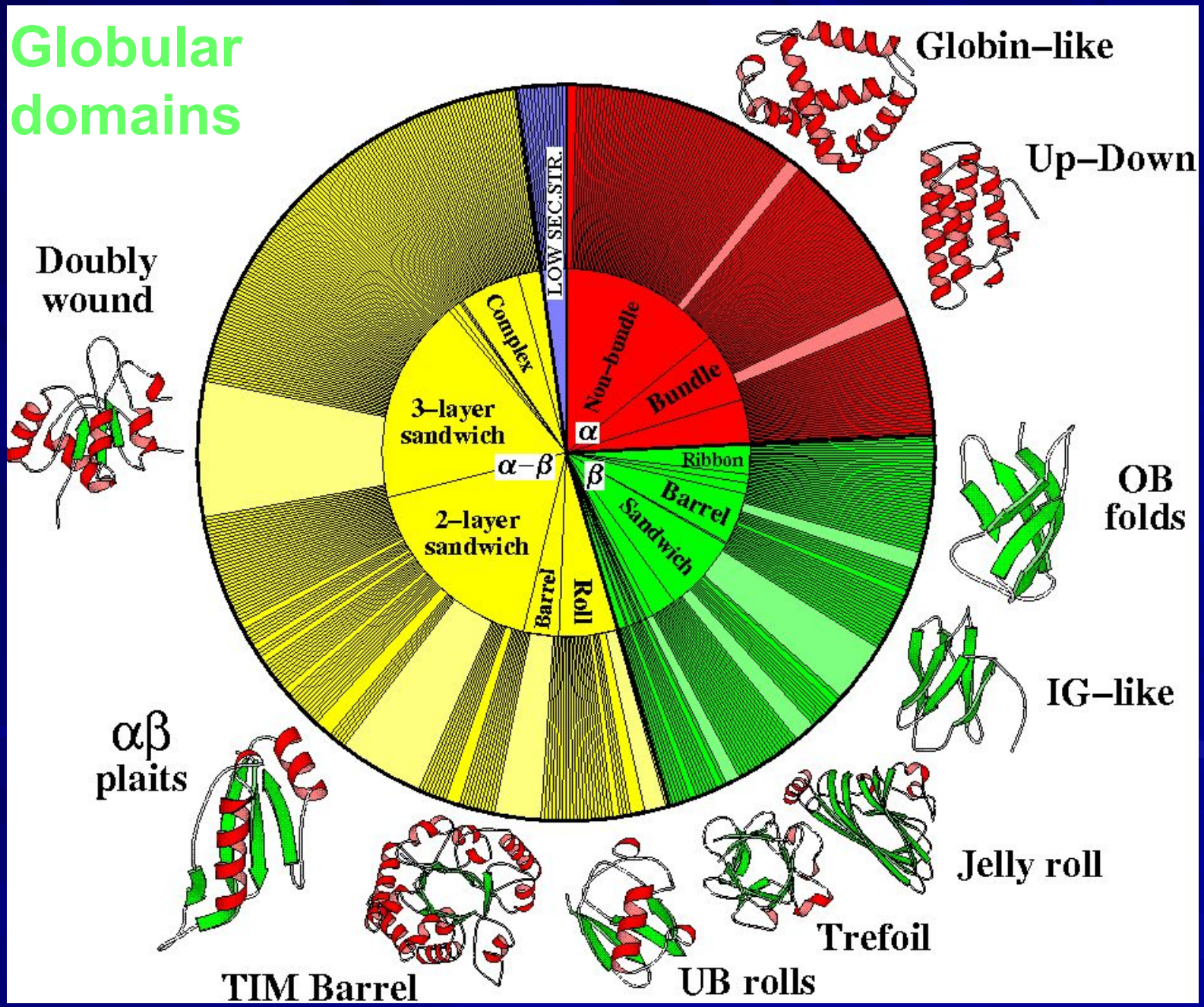
"What's good for General Motors is good for America"

*(a famous misquote of Charles Erwin **Wilson**)*

“Multitude principle”
for physical selection of folds
of globular proteins (*now*: “designability”):

the more sequences fit the given architecture without destroying its stability, the higher the occurrence of this architecture in natural proteins.

RATIONAL STRUCTURAL CLASSIFICATION OF PROTEINS



CATH ≈ SCOP

- Structures of water-soluble globular proteins
- Physical selection of protein structures: *min. of defects*!
- Rational structural classification of proteins