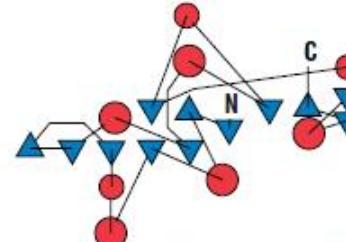
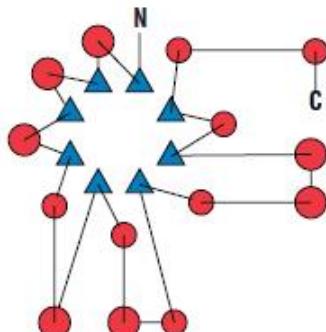


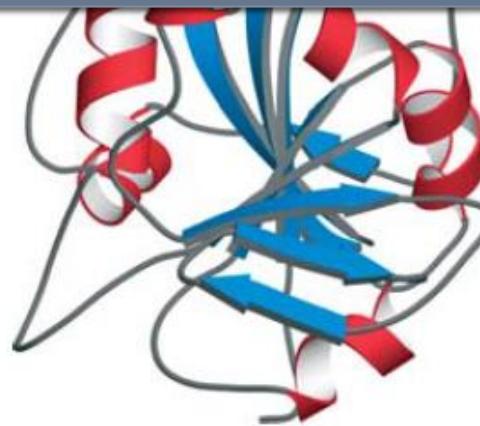
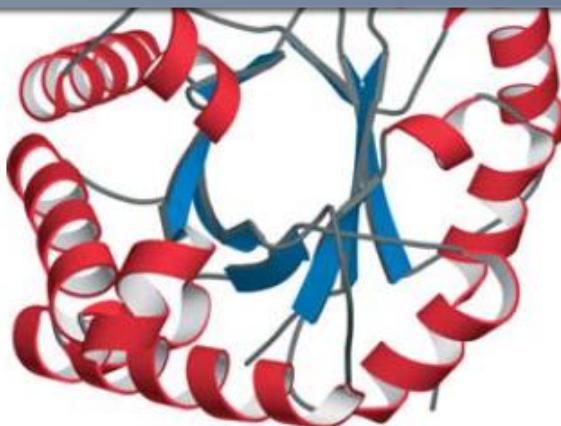
TERTIARY STRUCTURE, DOMAIN, AND MOTIF

Dr. Zhiyi Wei
SUSTC

Tertiary Structure

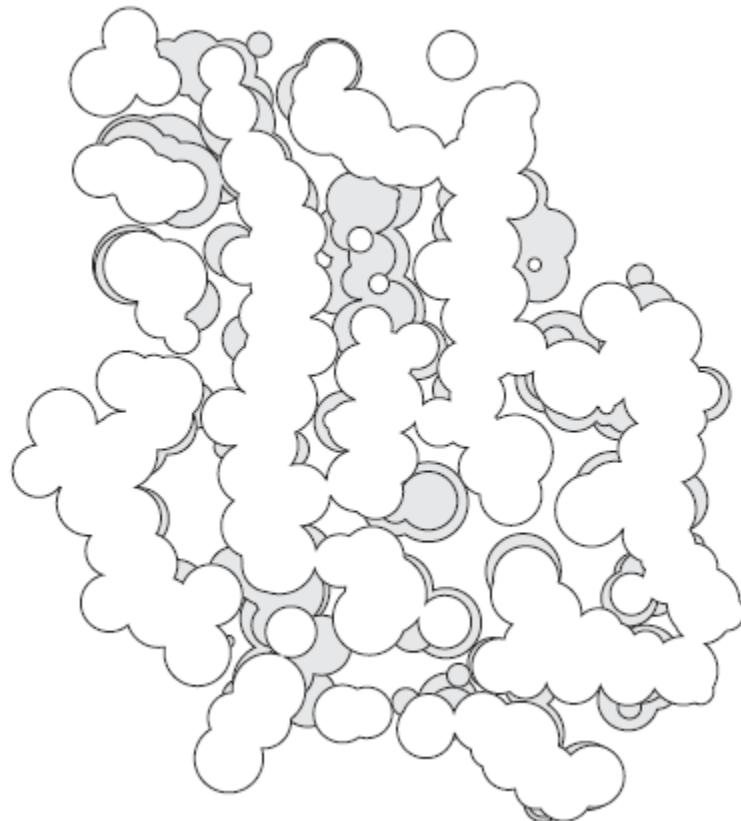


How can we read these structures?



Comparison of the structures of triosephosphate isomerase (PDB **1TIM**) and dihydrofolate reductase (PDB **1AI9**). Two proteins with similar secondary structure elements but different tertiary structures.

Efficient packing of atoms in the protein interior



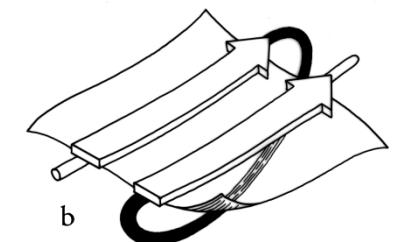
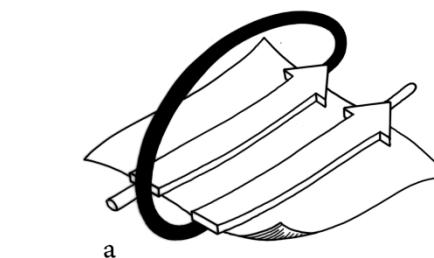
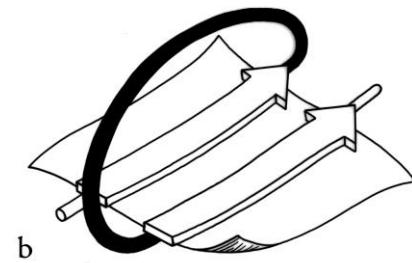
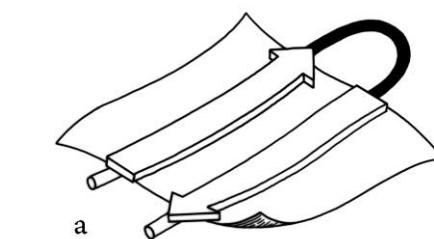
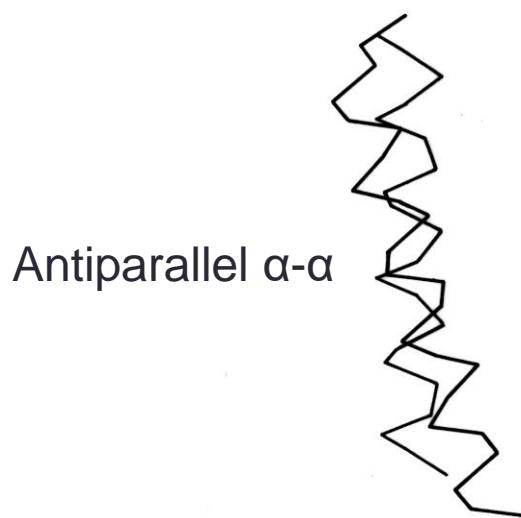
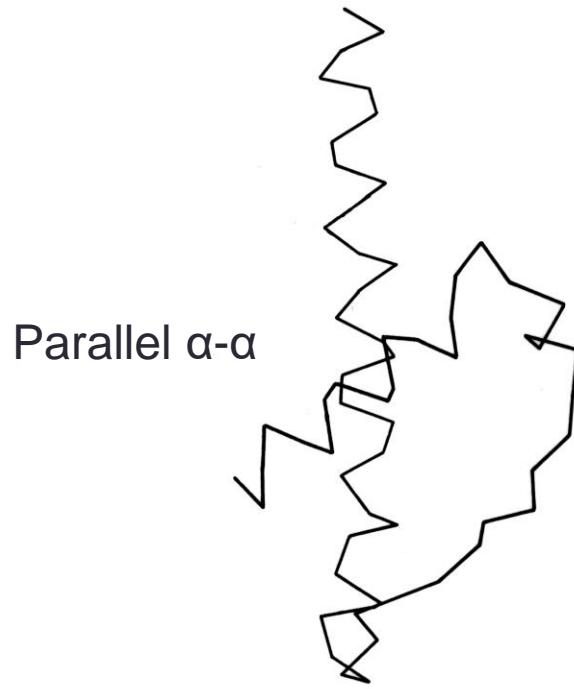
- Packing of individual secondary structure elements together in a protein

- burying the hydrophobic side chains
- forming a compact molecule
- very little empty space in the interior.

- Packing modes

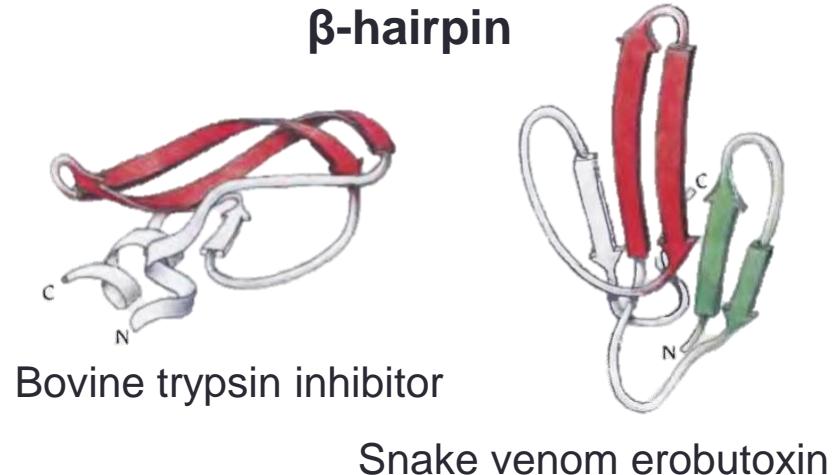
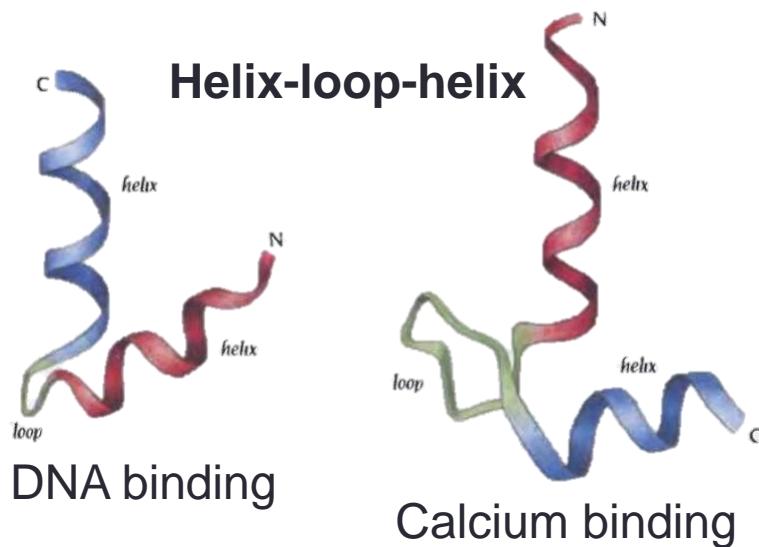
- α - α
- β - β
- α - β





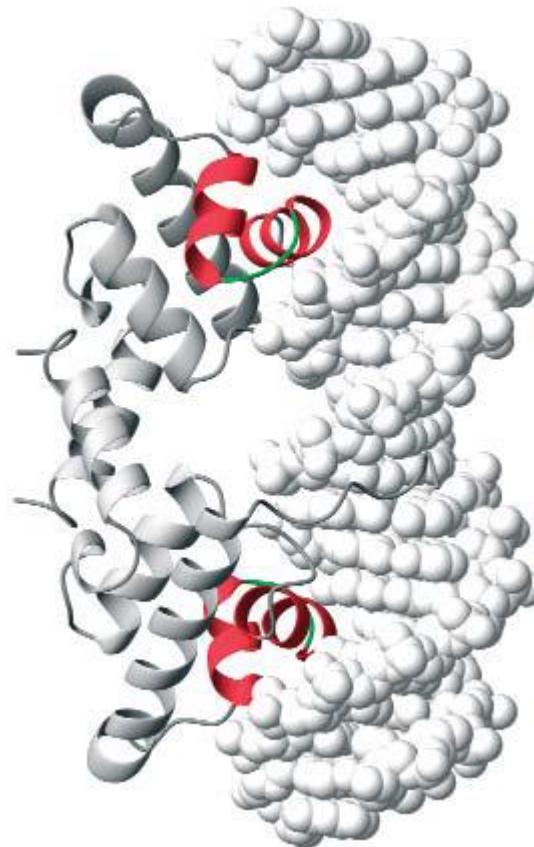
Protein Motif

- Sequence motif
- **Structural motif** (super secondary structure)
 - Secondary structure elements are connected to form simple motif.
 - Some have biological function, others are part of larger structural assembly (domain).



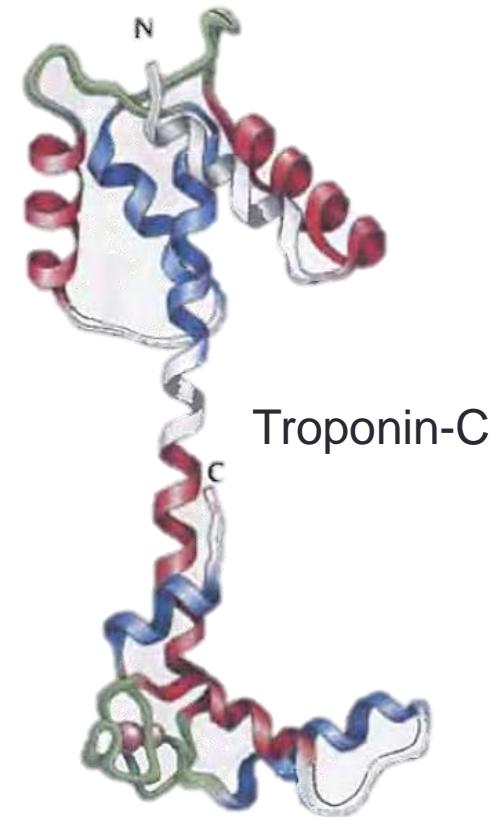
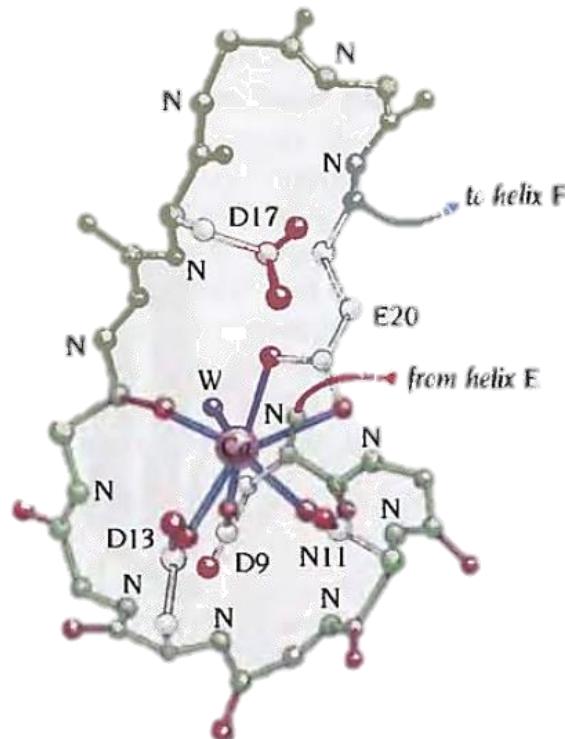
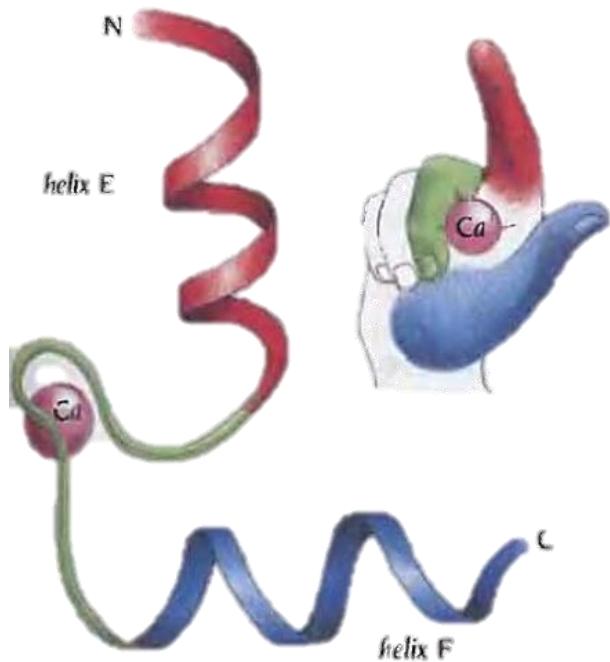
- Prediction of motifs is more difficult than that of secondary structure.

Helix-turn-helix (HTH)



The DNA-binding domain of the bacterial gene regulatory protein λ-Repressor, binding in the major groove and recognize specific gene regulatory sequences in the DNA. (PDB [1LMB](#))

EF-hand



Parvalbumin V K K A P A I I D Q D K S G F I E E D E L K L P L Q N F

Calmodulin F K E A F S L F D K D G D G T I T T K E L G T V M R S L

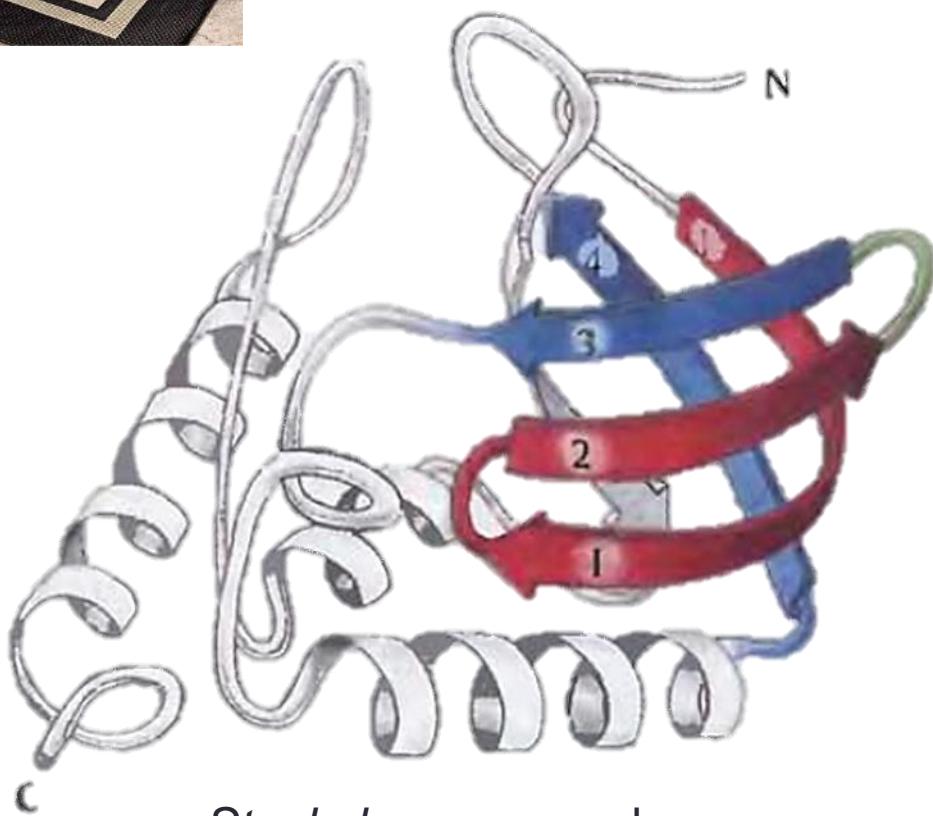
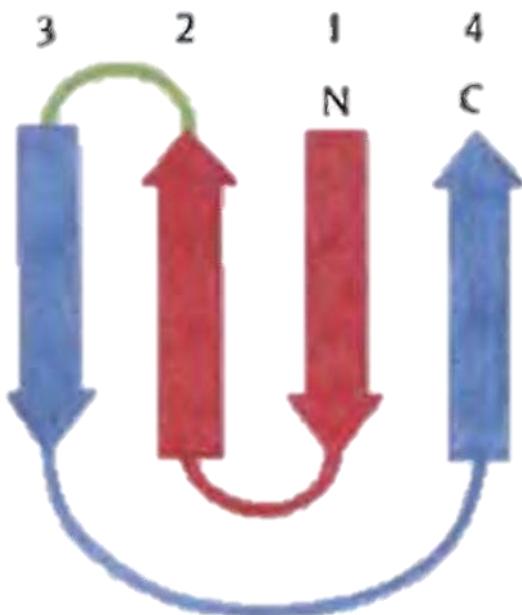
Troponin-C L A D C F E R I F D K N A D G F I D I E E L G E I L R A T

E helix

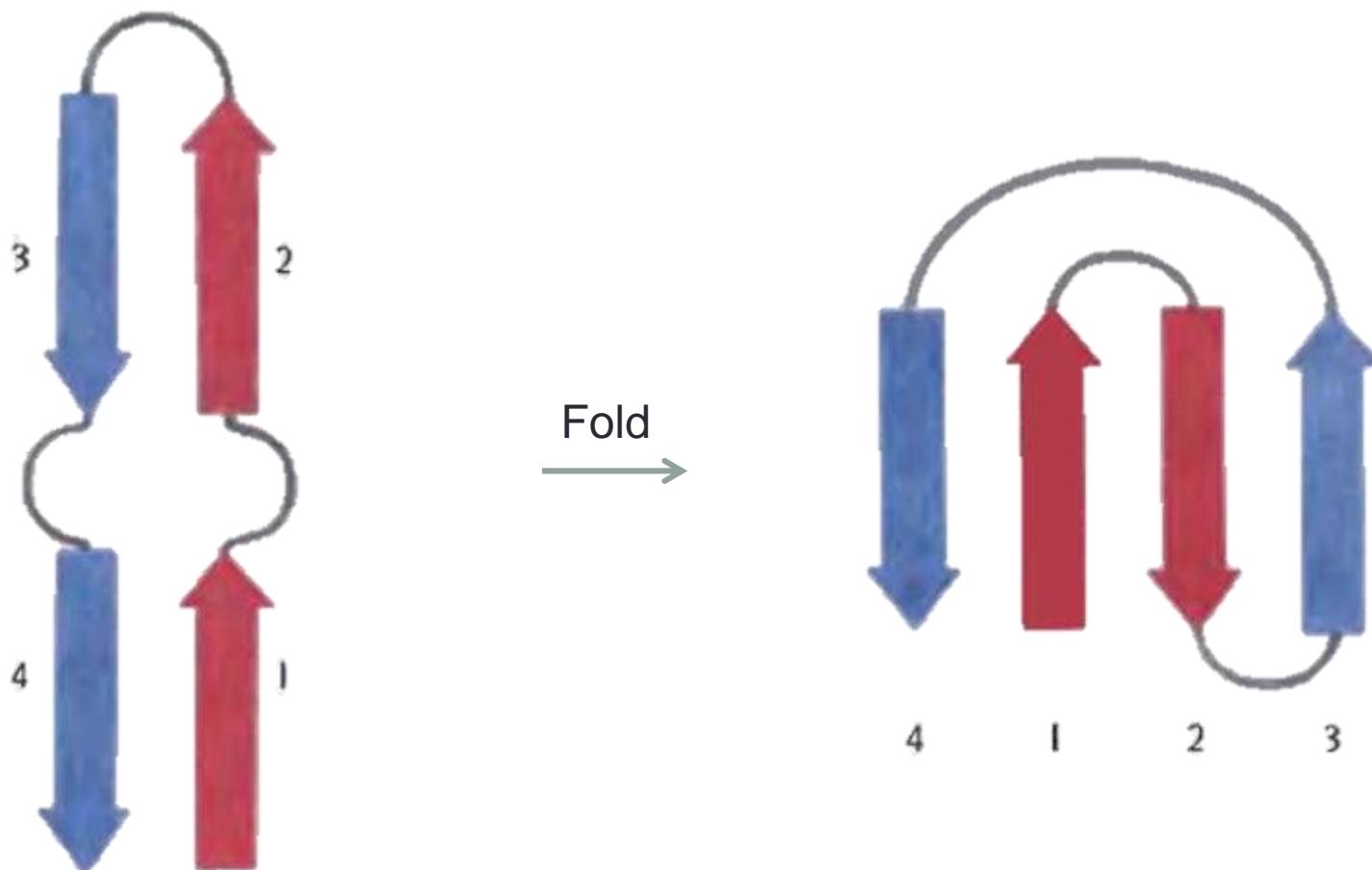
loop

F helix

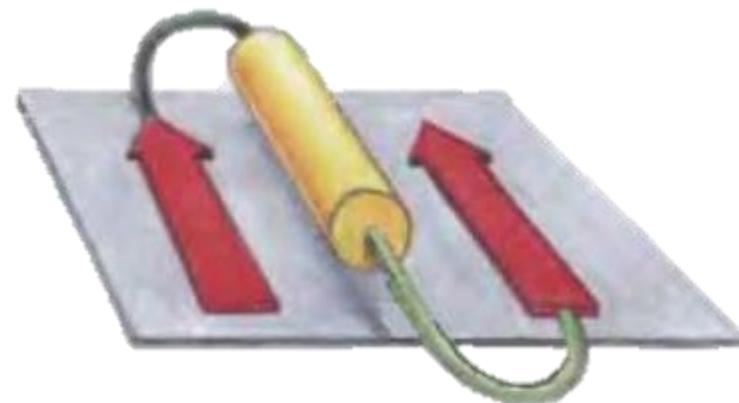
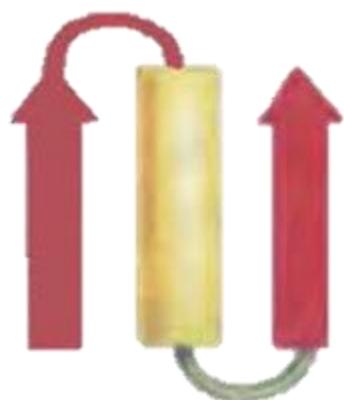
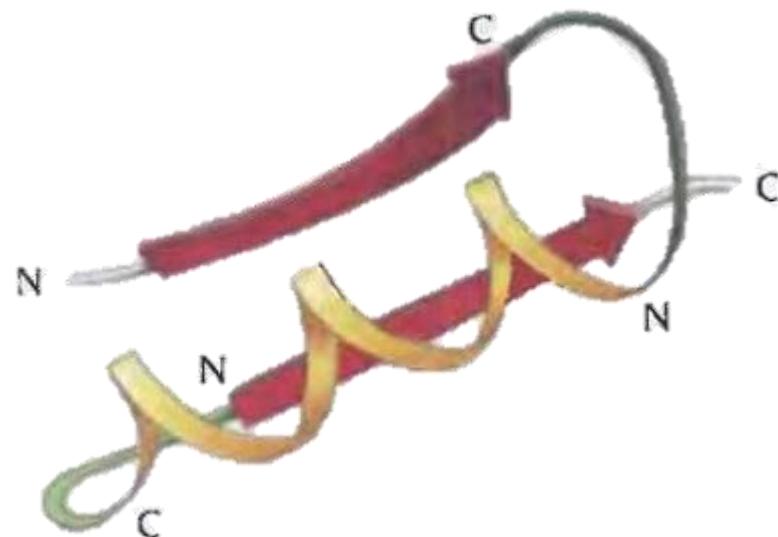
Greek key

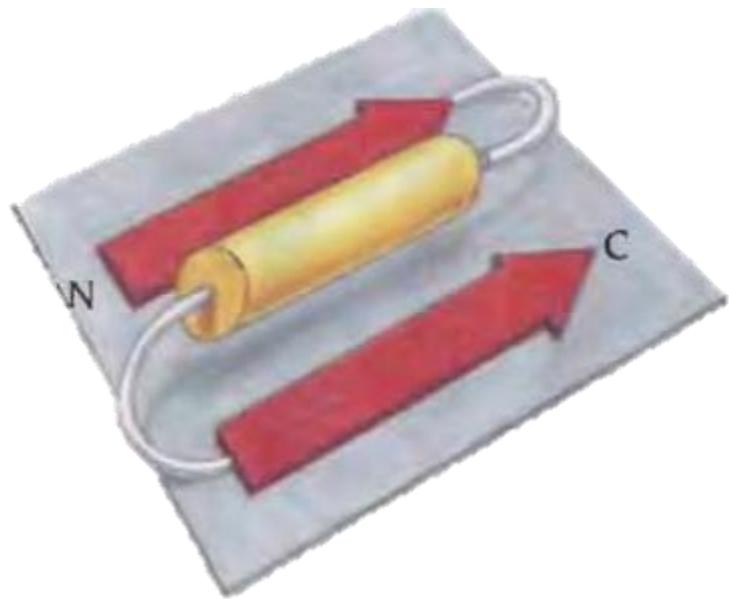


Staphylococcus nuclease

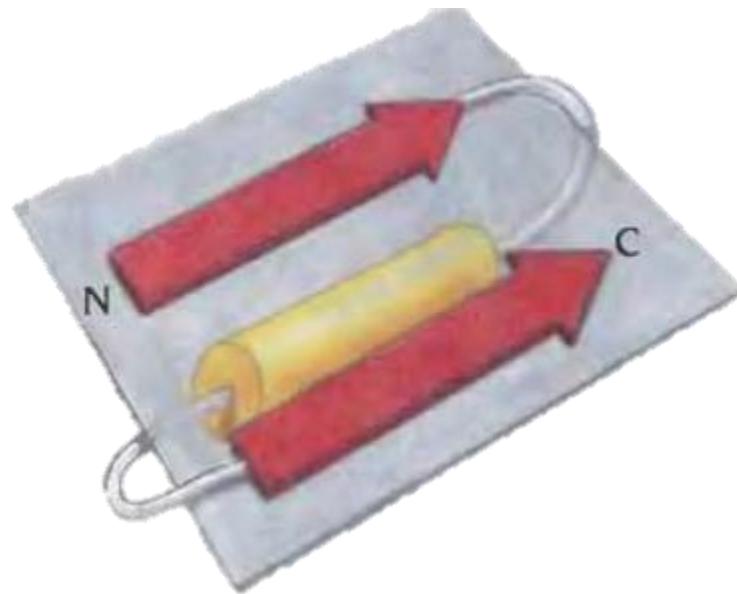


β - α - β motif





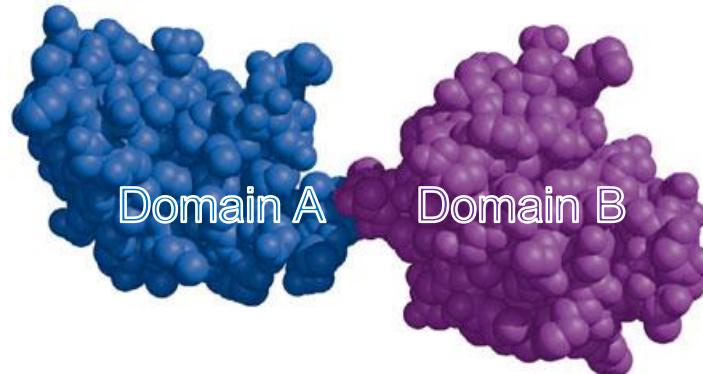
Right handed
(>95%)



Left handed

Protein Domain

- Domains are built from structural motifs.
- A domain is a **compact** region of protein structure.
- Usually, domain folding is **stable** enough to exist on its own in aqueous solution.
- Domain often retain part of the **biochemical function** of the larger protein from which they are derived.



Function A + Function B = Protein Function

- **Hydrophobic cores** are essential for the stability of domains.

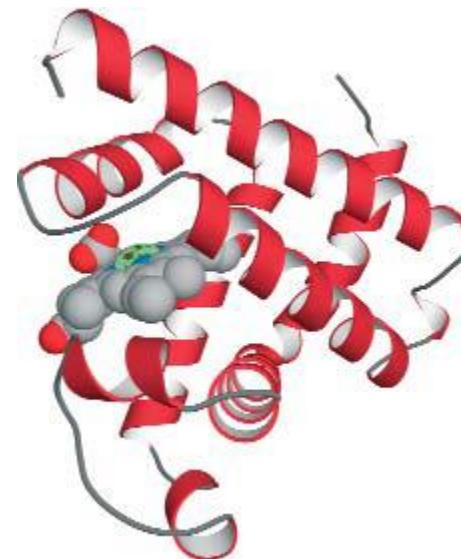
Classification of domains

- α domain
 - α helices connected by loops
- β domain
 - Antiparallel β -sheet
- α/β domain
 - β - α - β motif
 - Predominant parallel β -sheet
- $\alpha+\beta$ domain
 - Combination of discrete α and β motifs
- Cross-linked domain
 - Disulfide bonds
 - Metal ions

α domain

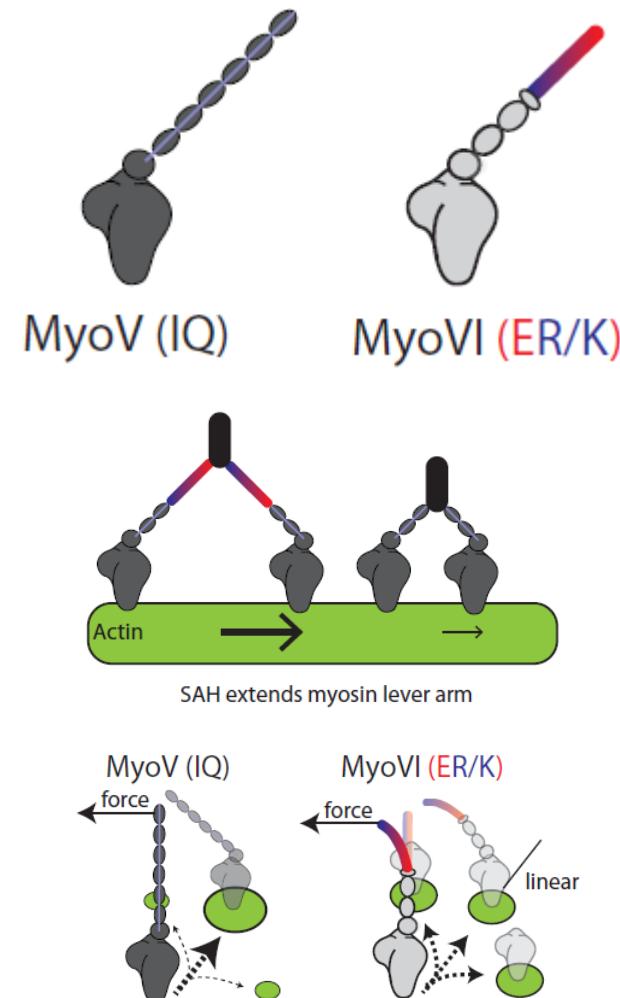
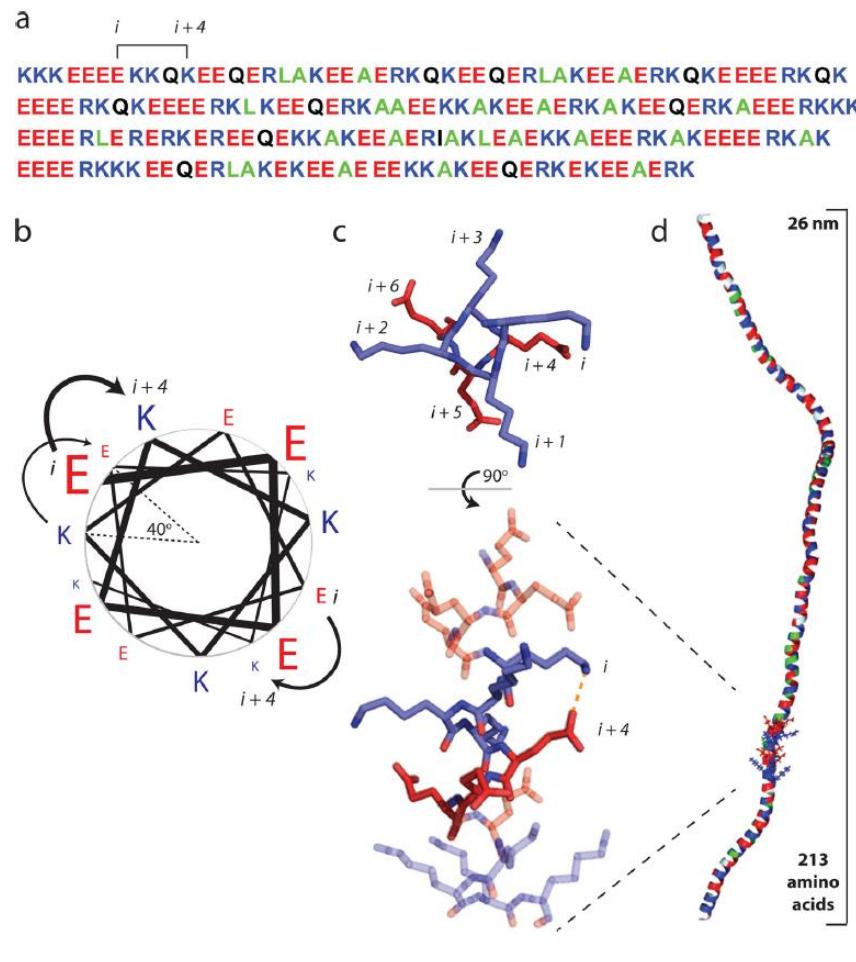


Myohemerythrin, a protein composed of a single four-helical bundle domain. (PDB **2MHR**)

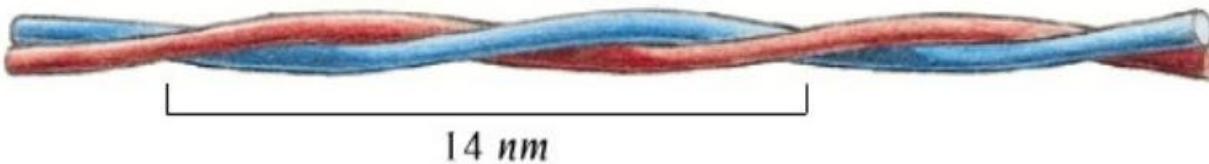


Myoglobin, a protein composed of a single globin fold domain. (PDB **1A6K**)

α domain: single α -helix (SAH)



α domain: coiled-coil

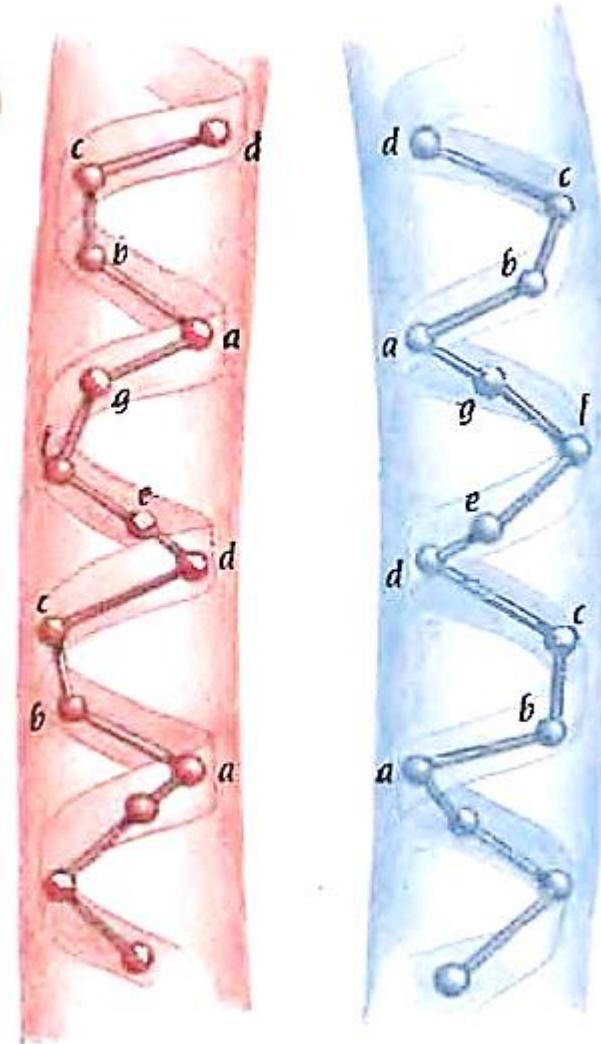


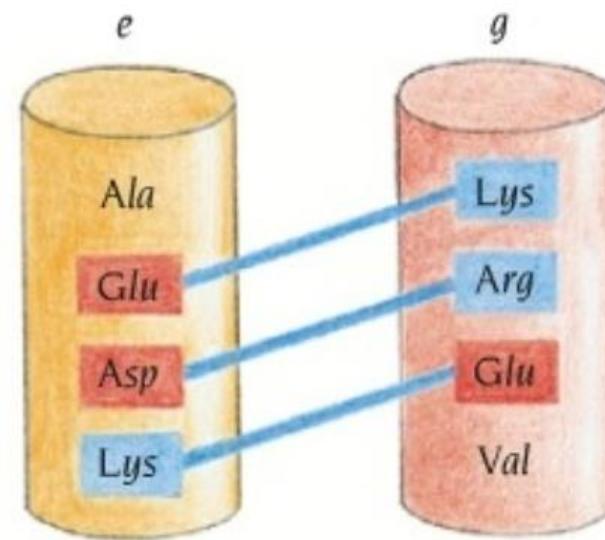
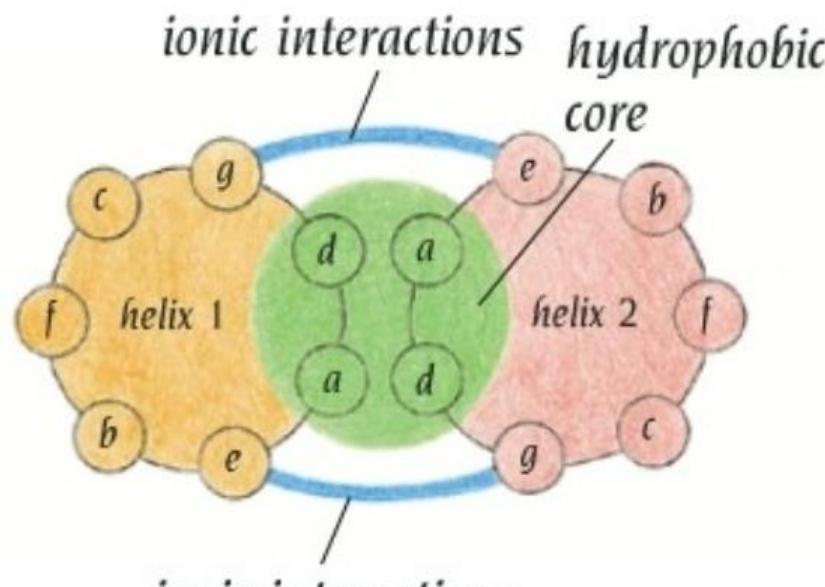
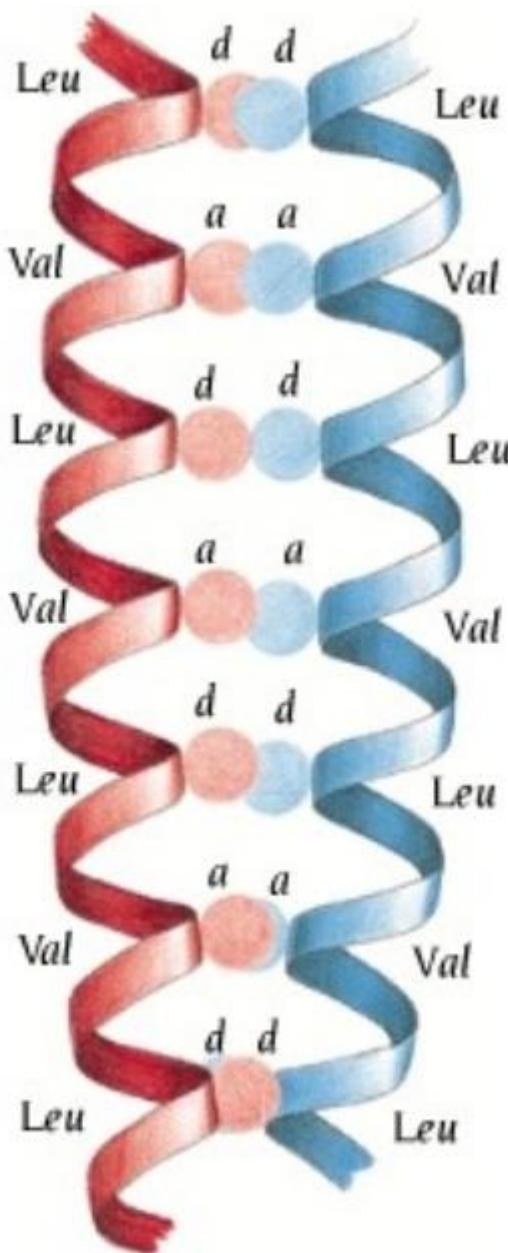
- Coiled-coil helices contain a **repetitive heptad** amino acid sequence pattern.
 - d: often leucine (leucine zipper)
 - a: often hydrophobic
 - e,g: often charged
 - b,c,f: charged or polar
 - Prediction

Why a heptad?

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
NH ₂ - Met	- Lys	- Gln	- Leu	- Glu	- Asp	- Lys -
Val	- Glu	- Glu	- Leu	- Leu	- Ser	- Lys -
Asn	- Tyr	- His	- Leu	- Glu	- Asn	- Glu -
Val	- Ala	- Arg	- Leu	- Lys	- Lys	- Leu - COOH

Transcription factor GCN4



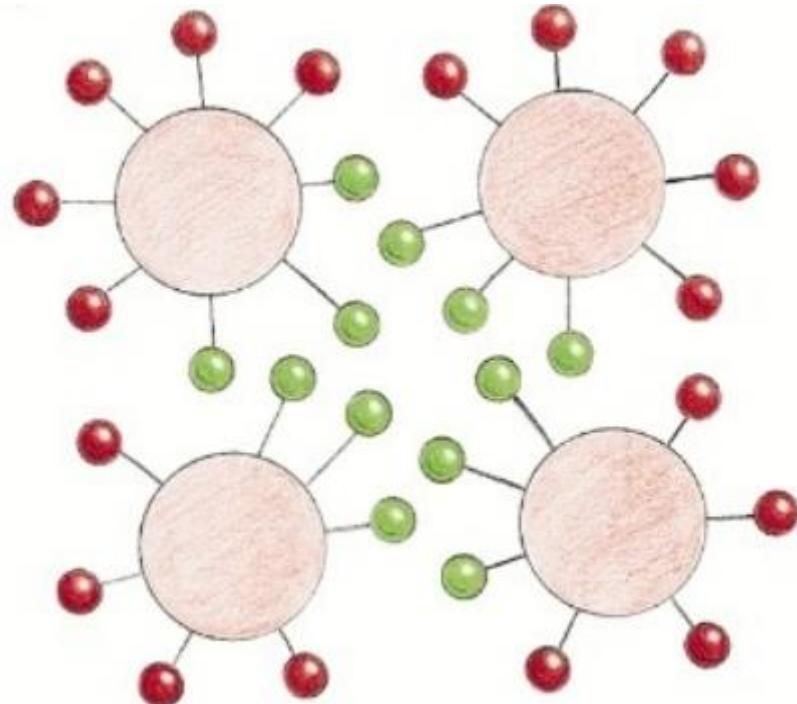




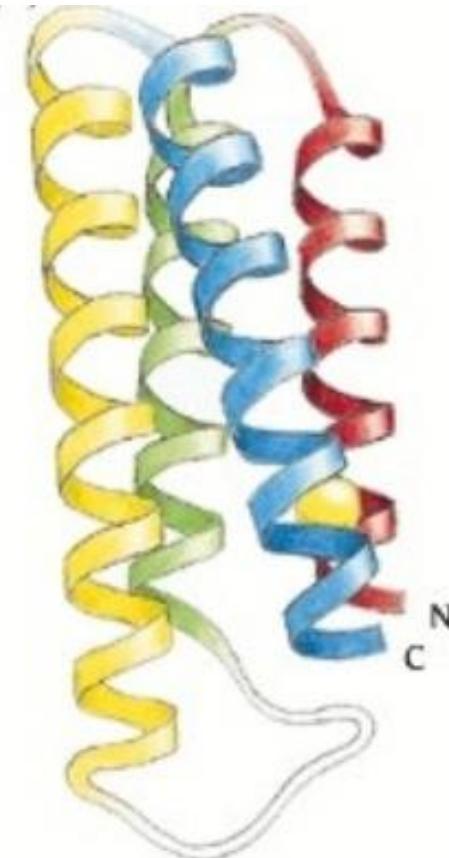
α domain: four-helix bundle



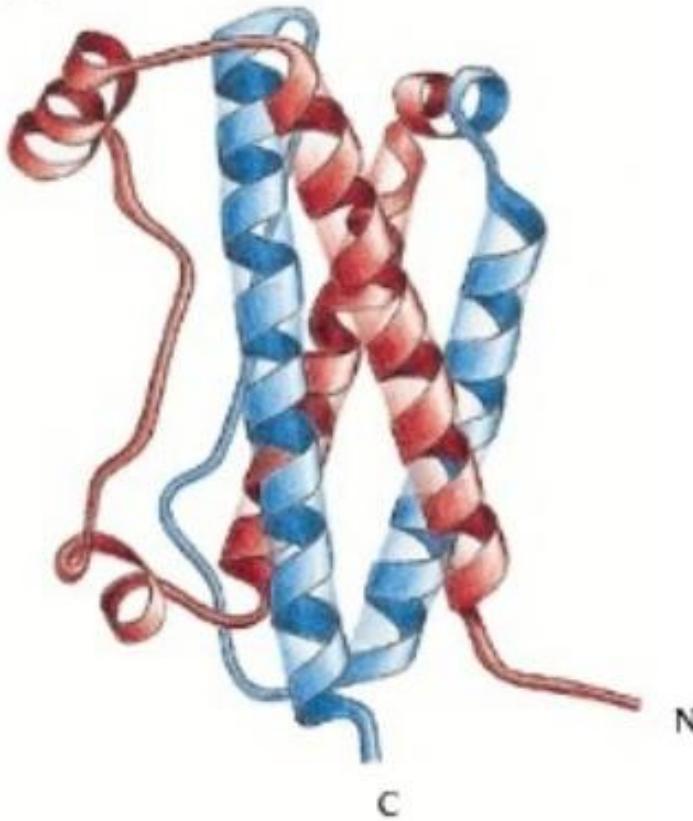
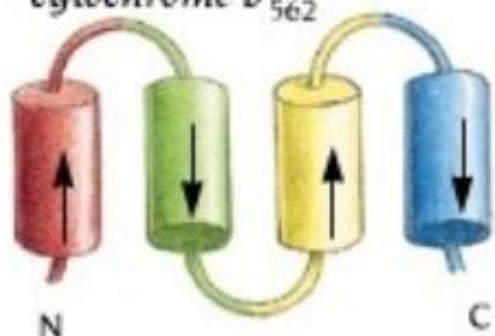
A common domain fold in α proteins



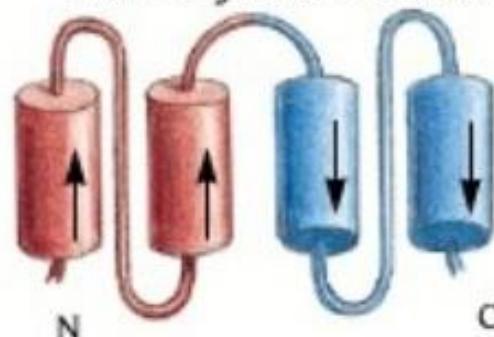
Creating a hydrophobic core in the middle of the bundle along its length

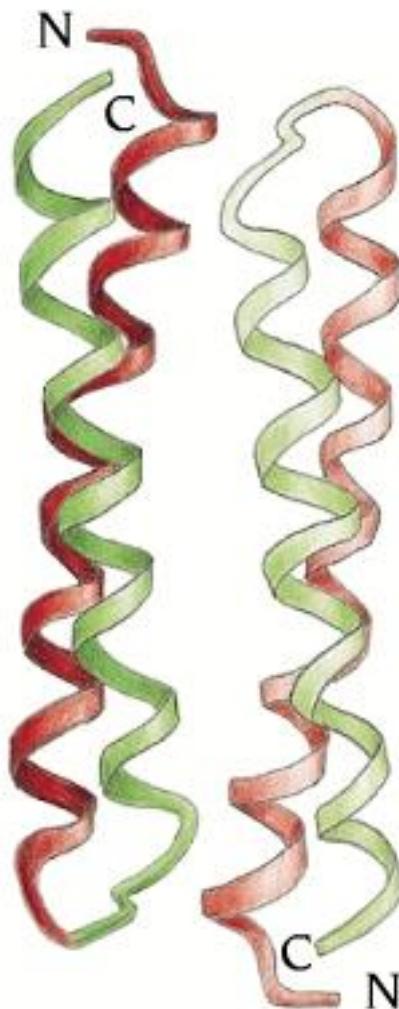


*cytochrome b*₅₆₂ (PDB 256B)



human growth hormone (PDB 1HGU)

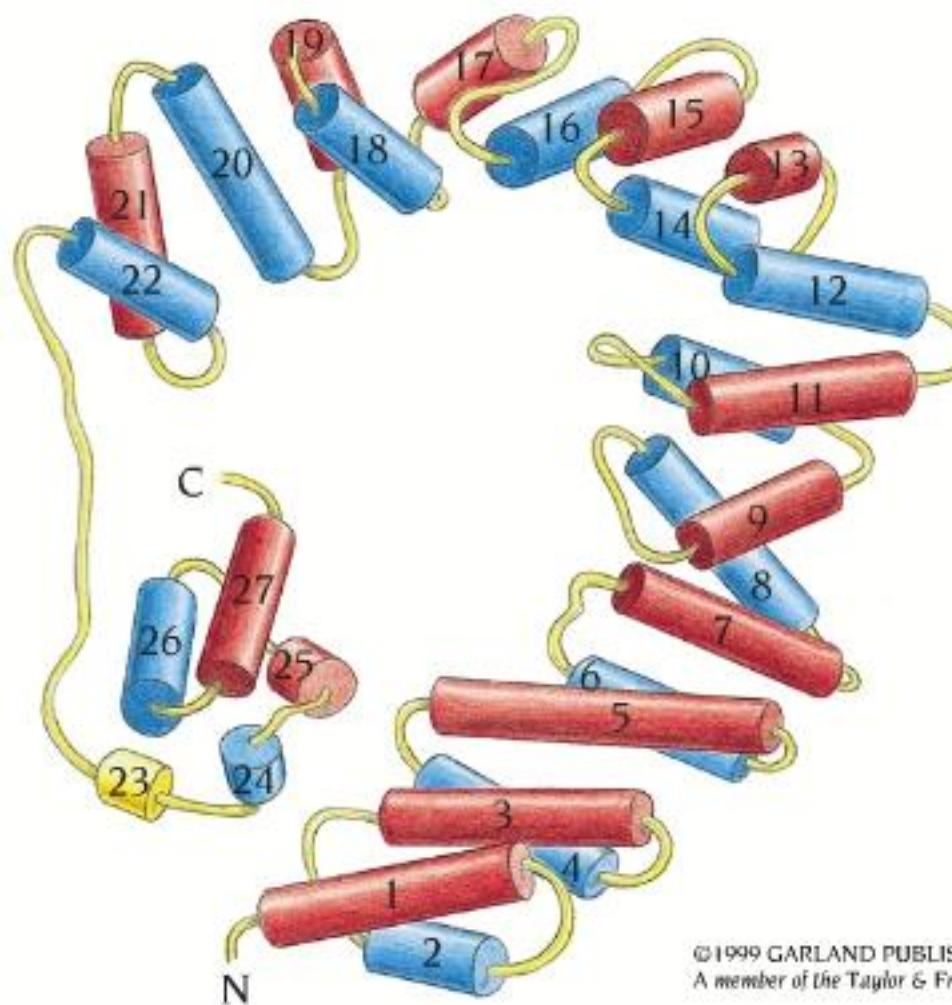




Dimeric Rop

A special case: two leucine zippers can form a four-helix bundle.

α domain can be large and complex.

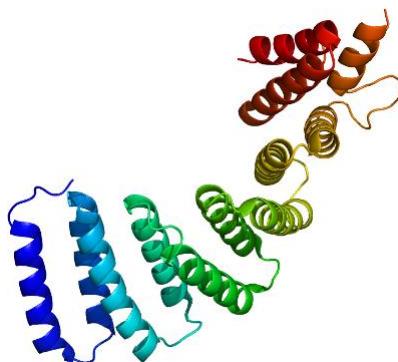


©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

Bacterial muramidase (PDB 1SLY)

α domain: repeats

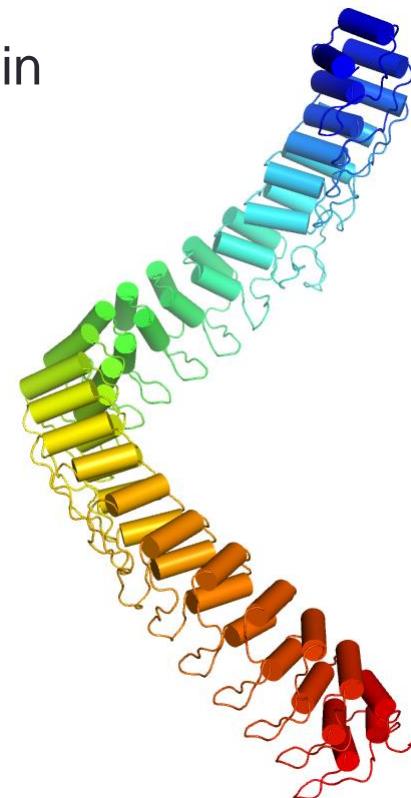
Versatile protein binding domain



TPR repeat
Human kinesin light chain 2

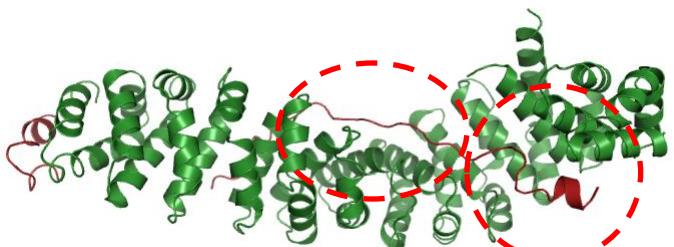


ARM repeat
Importin α

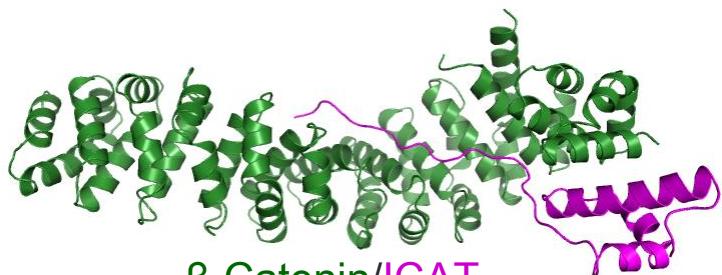


ANK repeat
Ankyrin-B

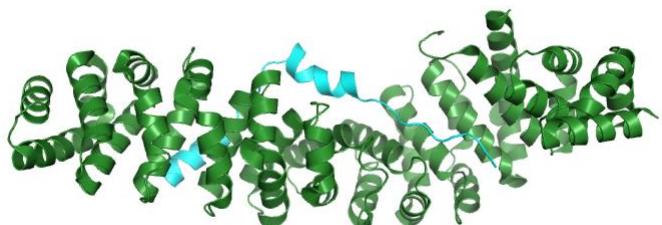
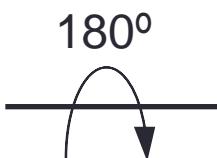
tion



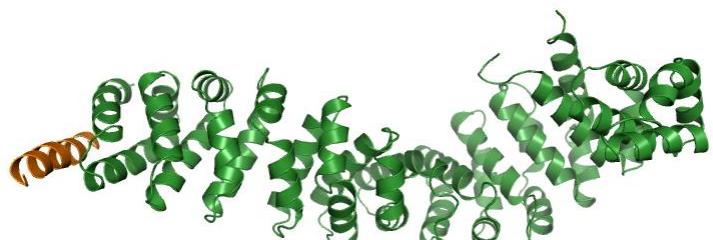
β-Catenin/E-Cadherin
(PDB id: 1I7W) Huber & Weis. Cell. 2001



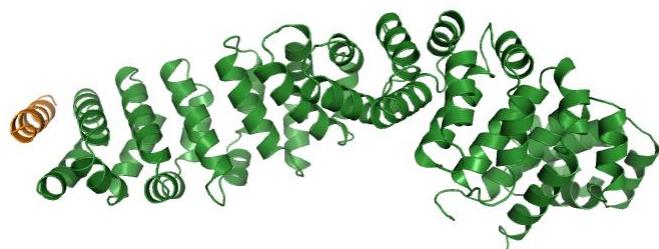
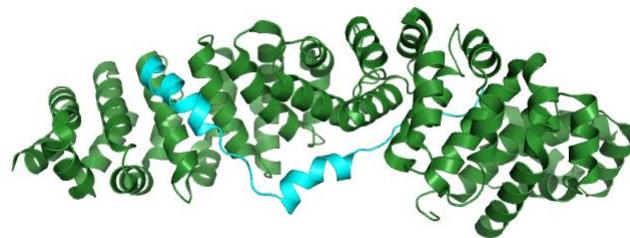
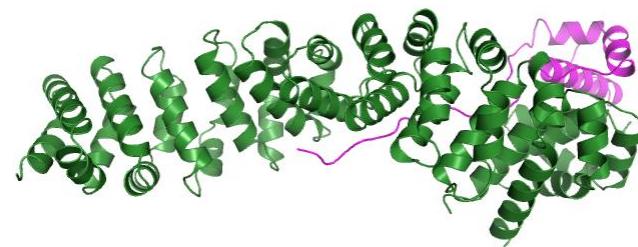
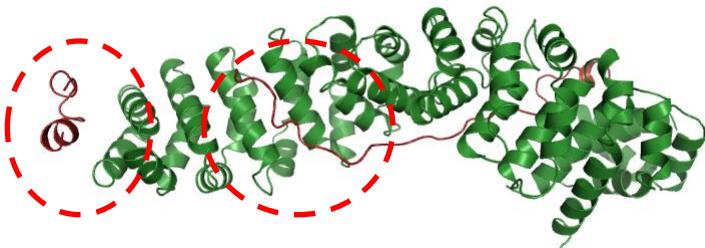
β-Catenin/ICAT
(PDB id: 1LUJ) Graham et al. Mol Cell. 2002



β-Catenin/TCF4
(PDB id: 1JDH) Graham et al. Nat Struct Biol. 2001

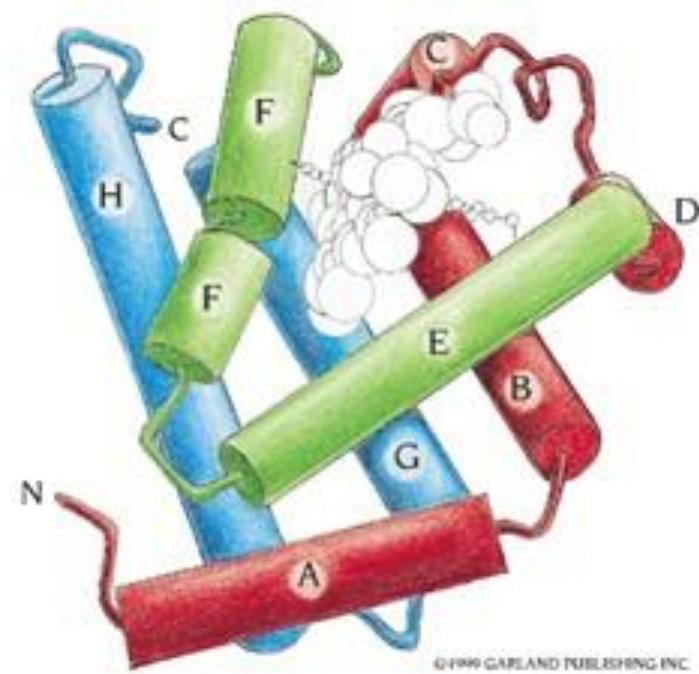


β-Catenin/BCL9
(PDB id: 2GL7) Sampietro et al. Mol Cell. 2006

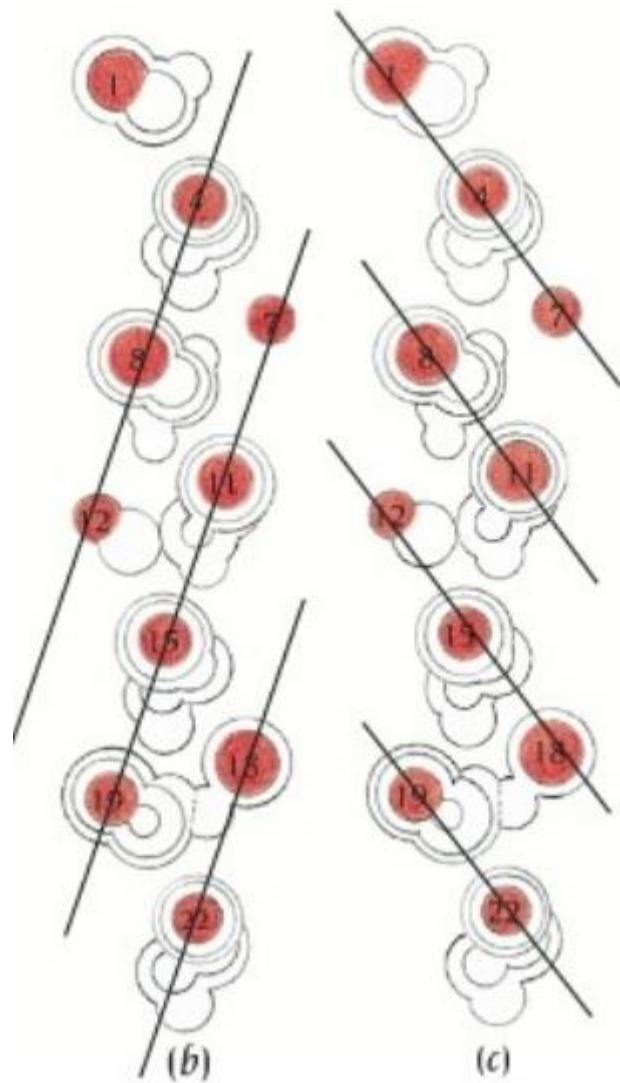
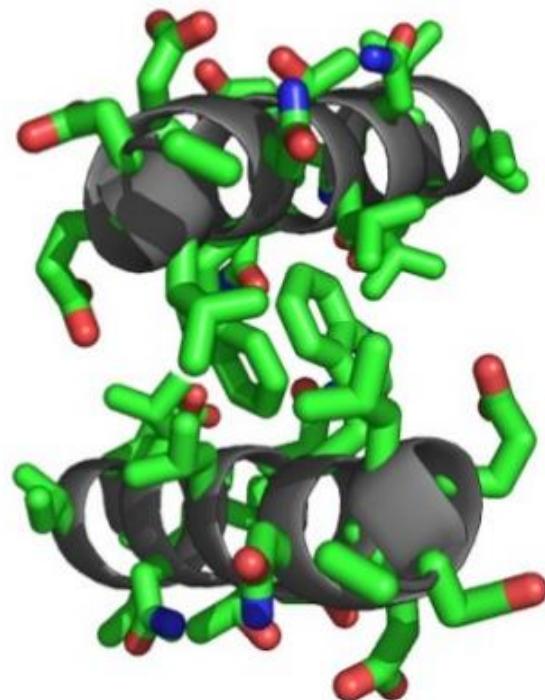
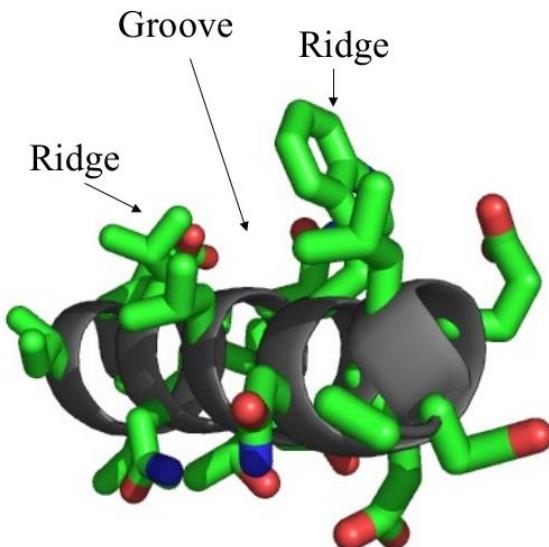


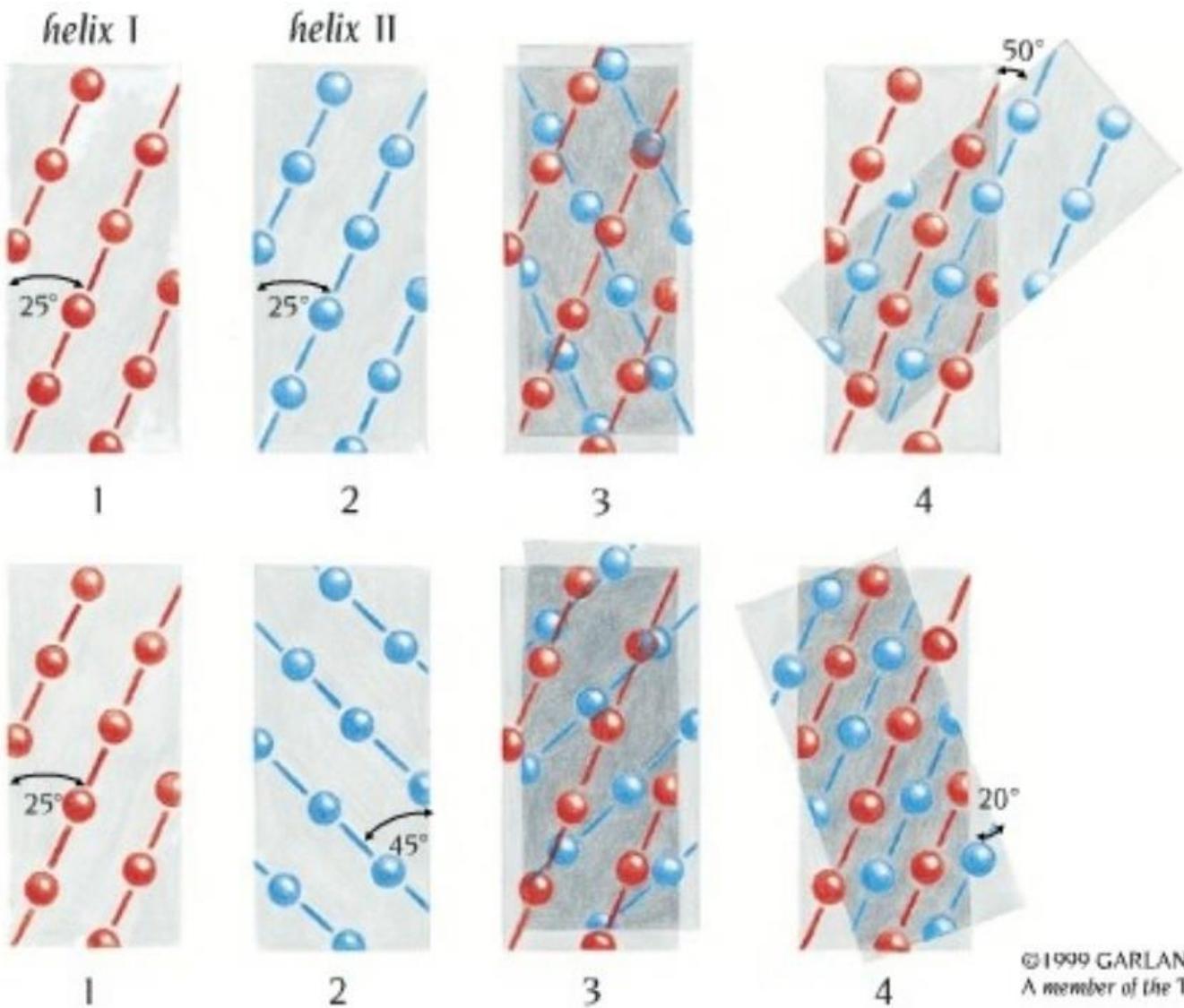
α domain: Globin fold

- A large protein family
 - Hemoglobin (α , β)
 - Myoglobin
 - Neuroglobin
 - Cytoglobin
 - Flavohemoglobin
 - ...
- One of the most important α structure
- All organisms contain proteins with globin fold
- Contains 8 helices, forming a hydrophobic pocket for heme binding (active site)

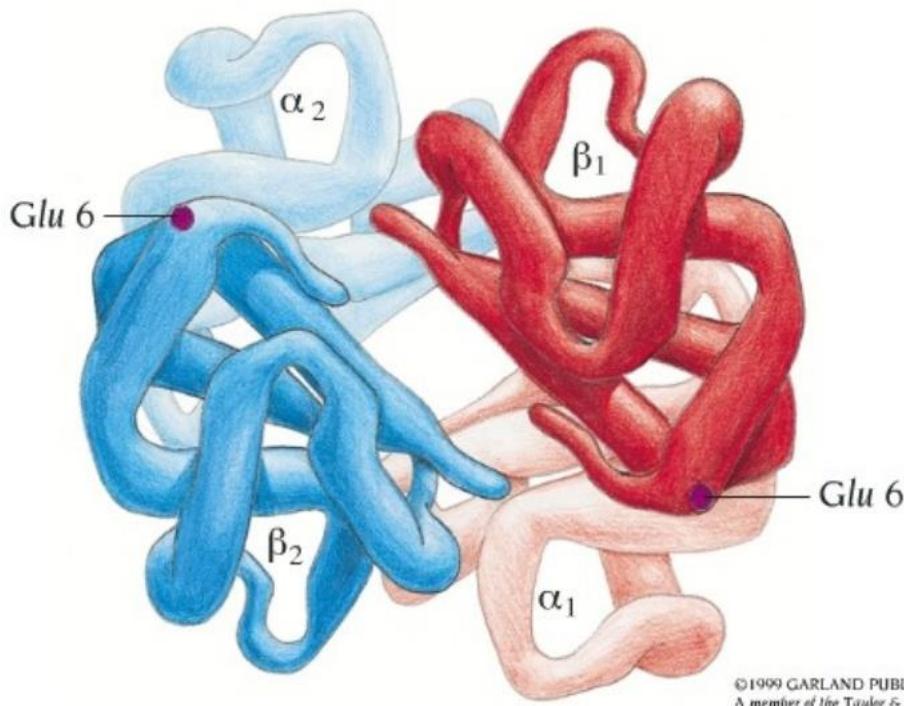


“Ridges in grooves” model

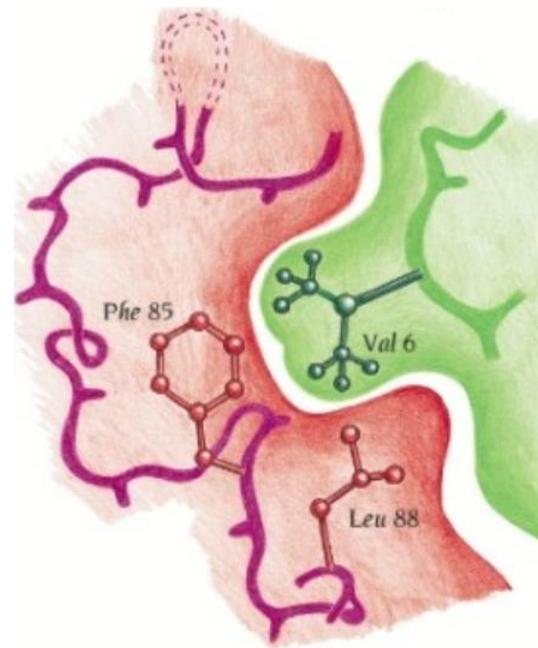




Understand sickle cell disease from structure



Glu6Val mutation



Why has this deadly mutation survived during evolution?

Malaria



Sickle cell anemia



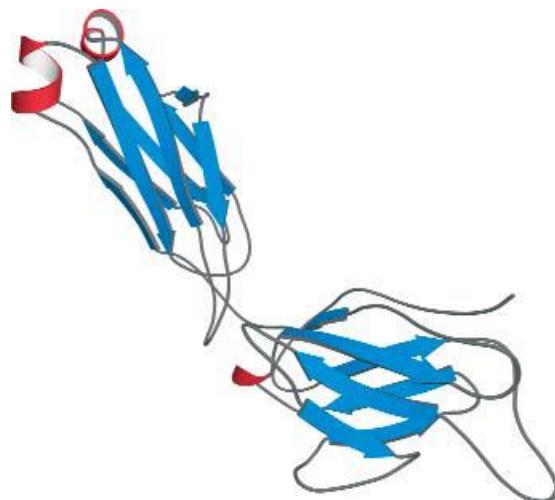
VS.



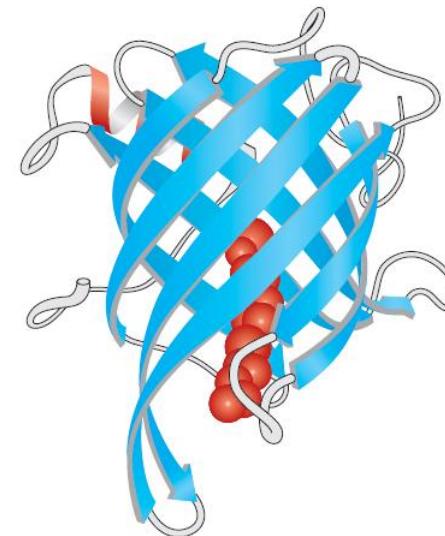
Evolution's choice

β domain

- β -strands are arranged in predominantly **antiparallel** fashion.
- Usually two β -sheets are formed, which pack each against other, resembling barrel or distorted barrel (also **β sandwich**).
- Multi-layered β sandwich can be found in fibrous proteins, like fibroin.
- Highly diverse functions

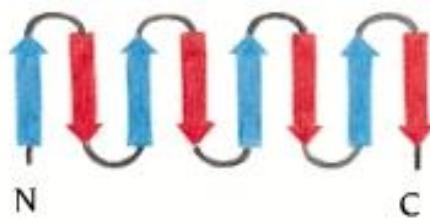
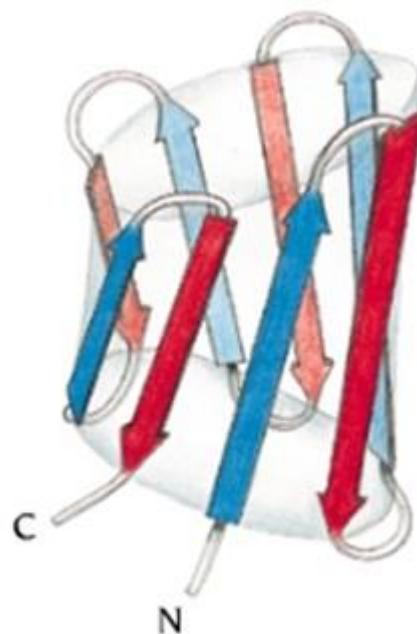


Immunoglobulin, each IG fold is comprised of two sandwiched β -sheet. (PDB **1A3I**)



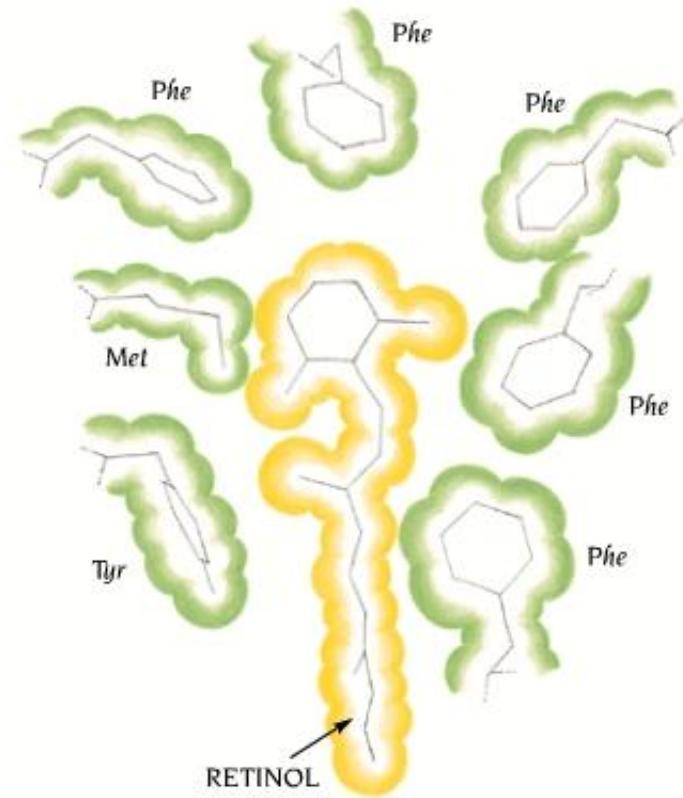
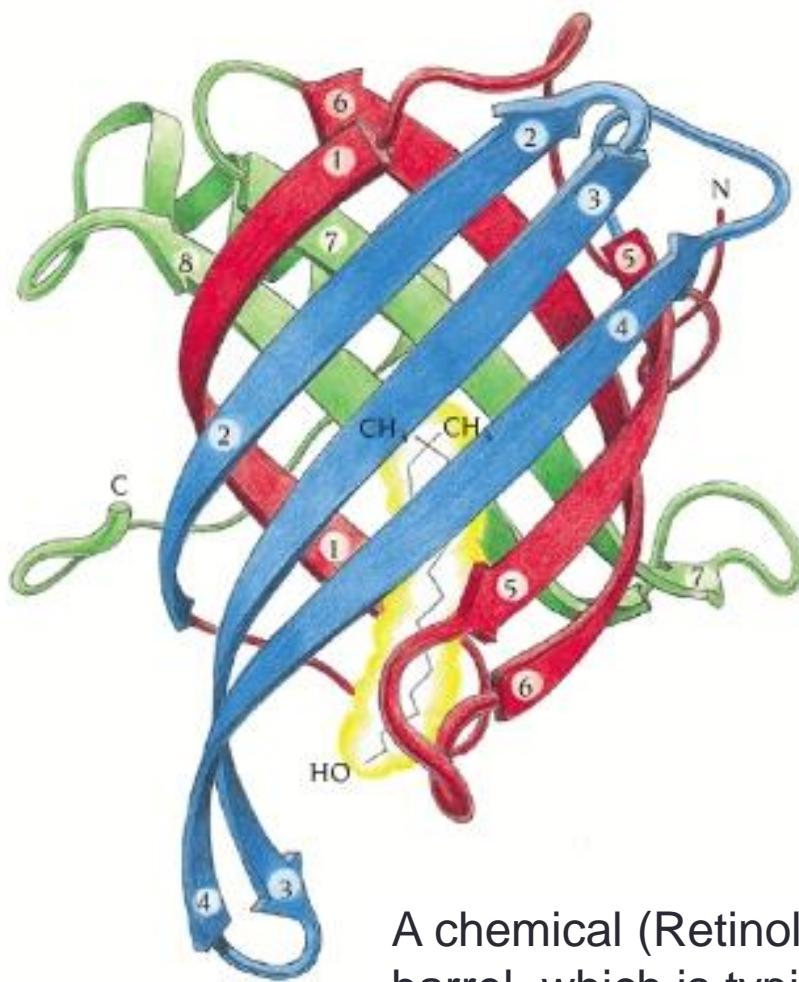
Plasma retinol-binding protein, a typical β -barrel fold. (PDB **1RLB**)

β domain: Up-and-down β -barrel



- The simplest way to fold a barrel-like structure
- All connections forming hairpins
- Versatile structure and functions

Retinol-binding protein (RBP)

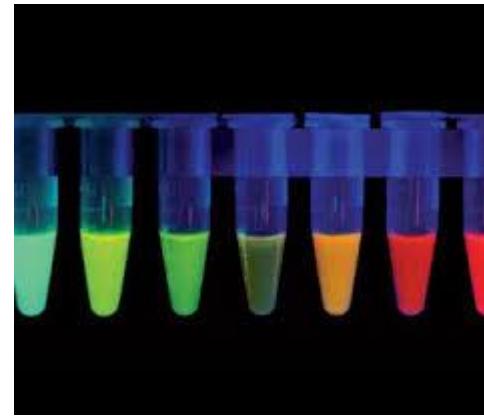
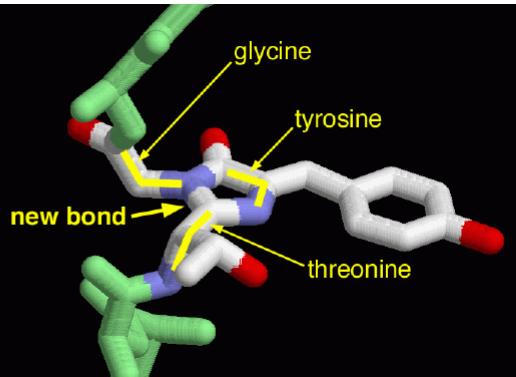
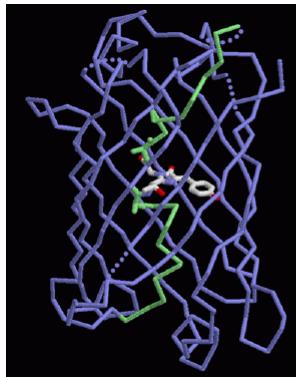
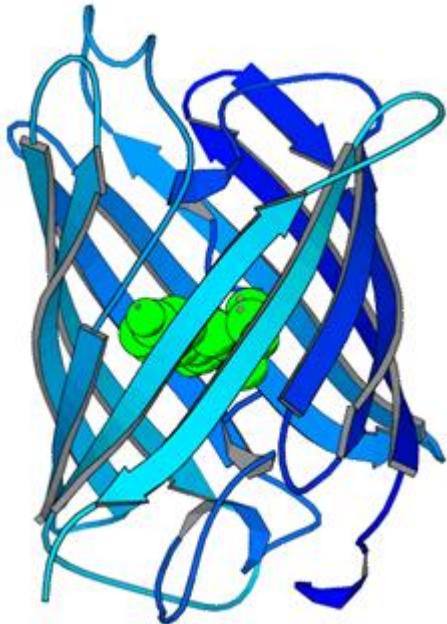


A chemical (Retinol) binds in the inside of barrel, which is typical for up-and-down barrels.

<i>strand no.</i>	<i>residue no.</i>	<i>amino acid sequence</i>
2	41–48	– Ile – Val – Ala – Glu – Phe – Ser – Val – Asp –
3	53–60	– Met – Ser – Ala – Thr – Ala – Lys – Gly – Arg –
4	71–78	– Ala – Asp – Met – Val – Gly – Thr – Phe – Thr –

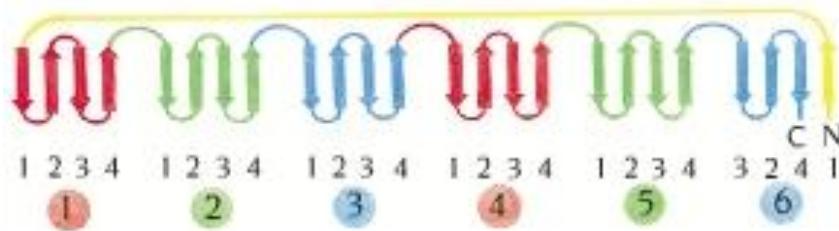
- The β -barrel **structure** is reflected by amino acid **sequence**
- Hydrophobic residues facing inside
- Hydrophilic residues exposed to solvent

Green fluorescence protein (GFP)

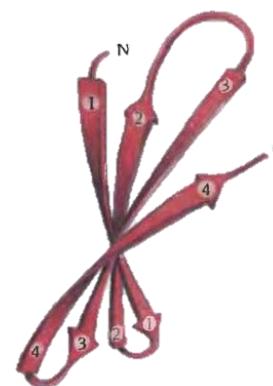


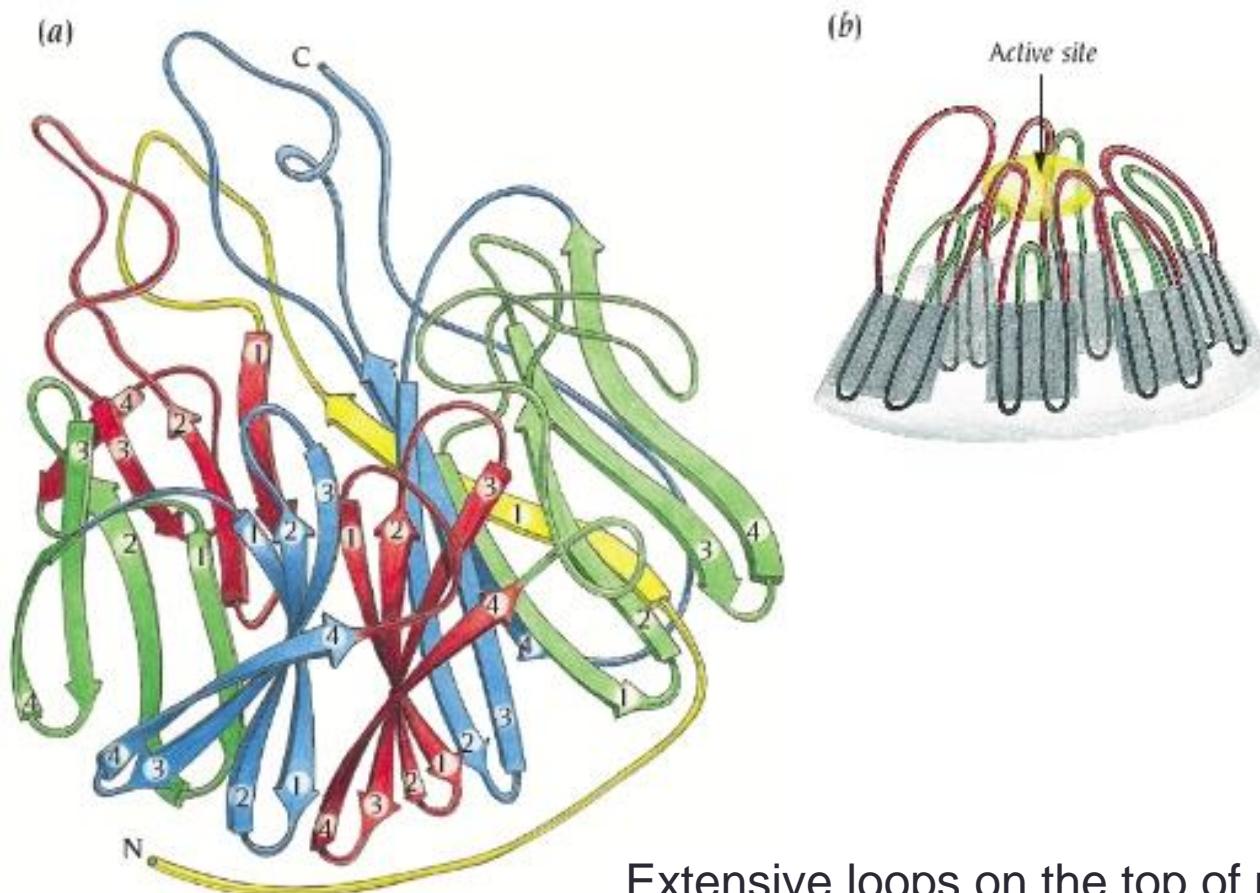
Roger Y. Tsien
钱永健

β domain: β -propeller



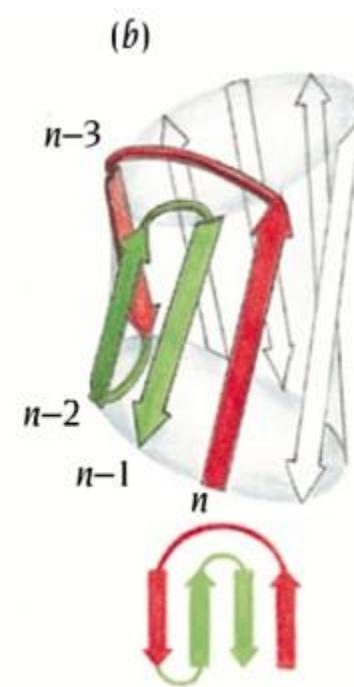
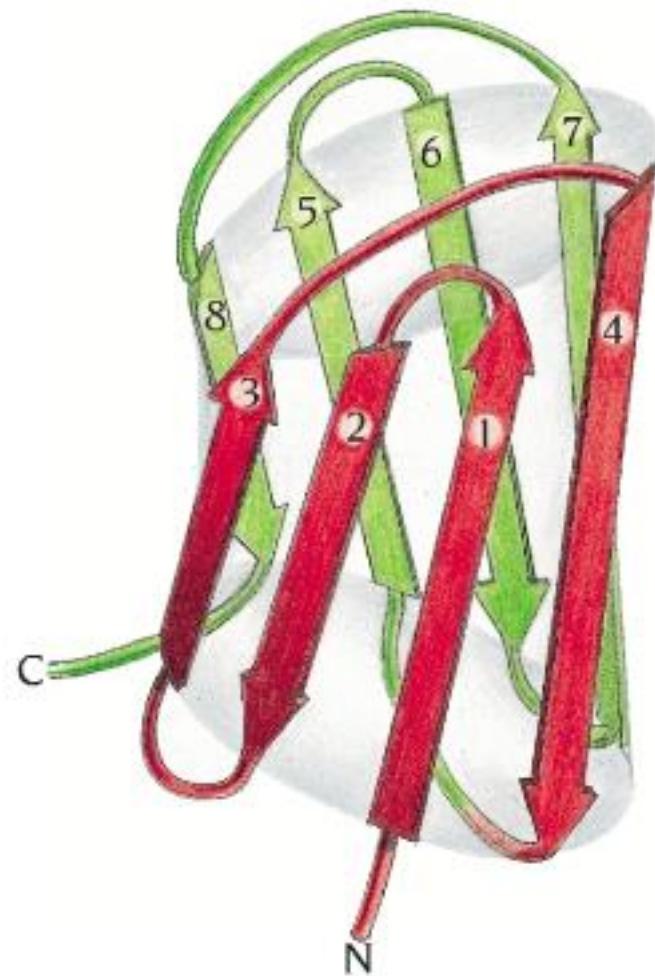
- Neuraminidase, an influenza virus protein, involved in virion release from cells
- Containing 6 β -sheet
- Each sheet having 4 strands arranged in an up-and-down fashion
- Called “**six-bladed β -propeller**”
- A unique up-and-down barrel





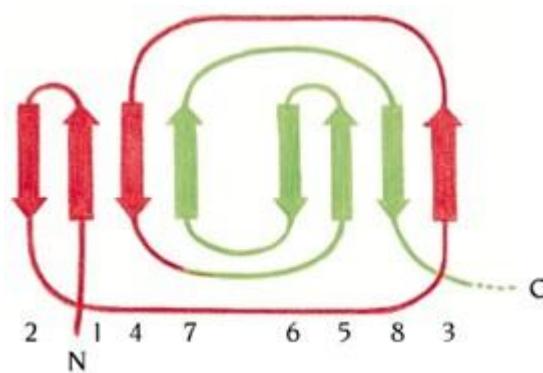
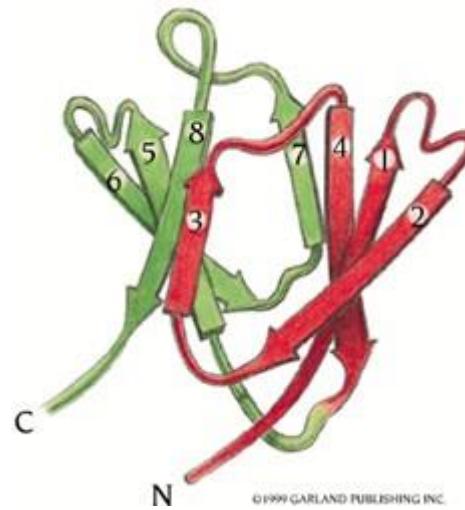
Extensive loops on the top of propeller form active site.

β domain: Greek key β -barrel



Two possible connection ways

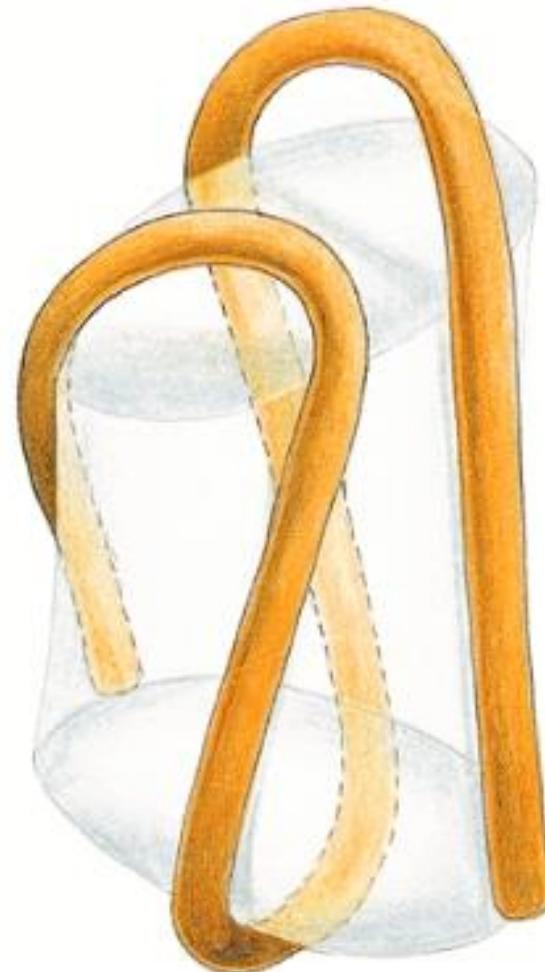
γ -crystallin

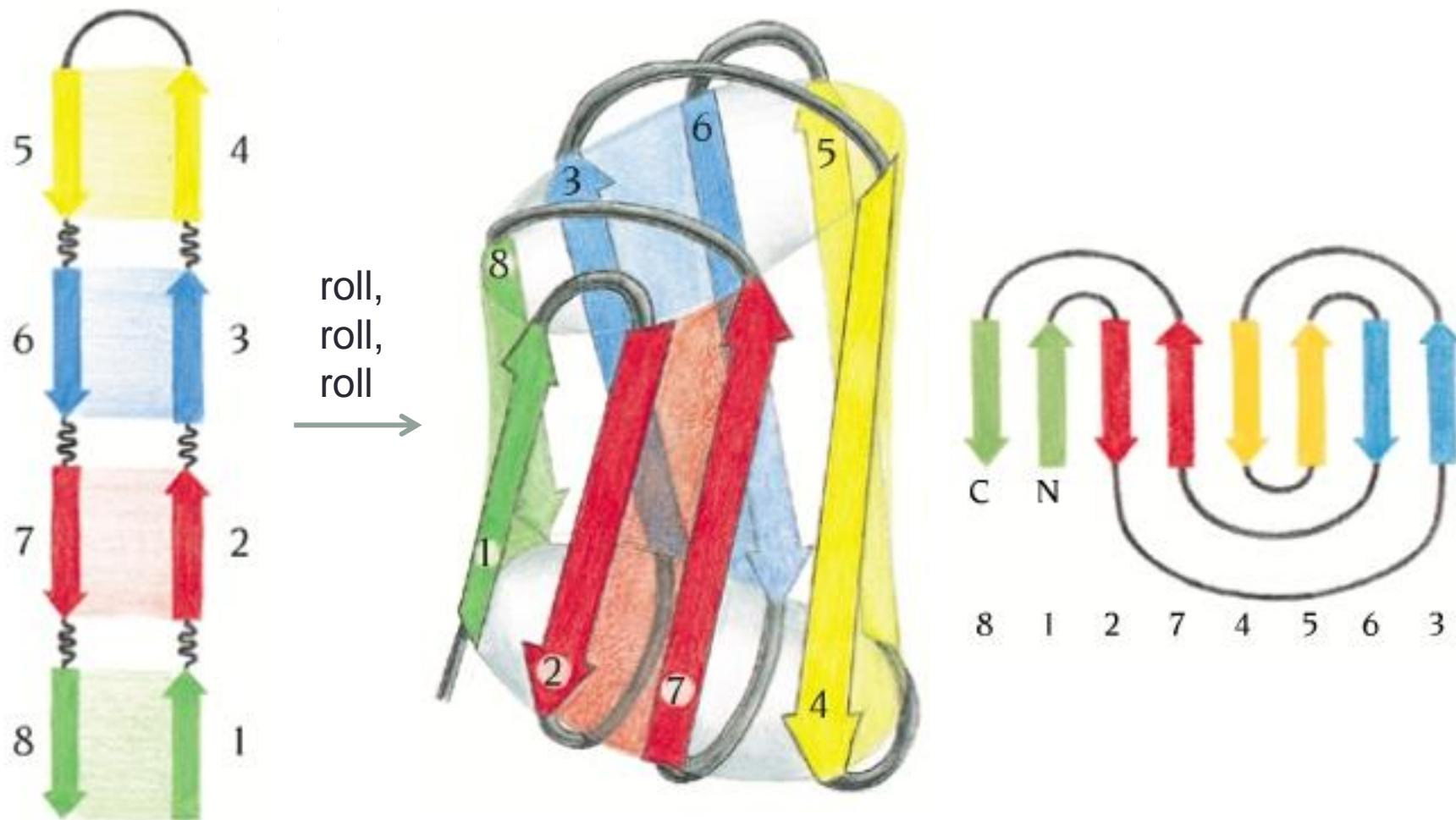


- Found in lenses of human eyes
- Each domain built from 2 greek key motifs
- One connection across the barrel between two motifs

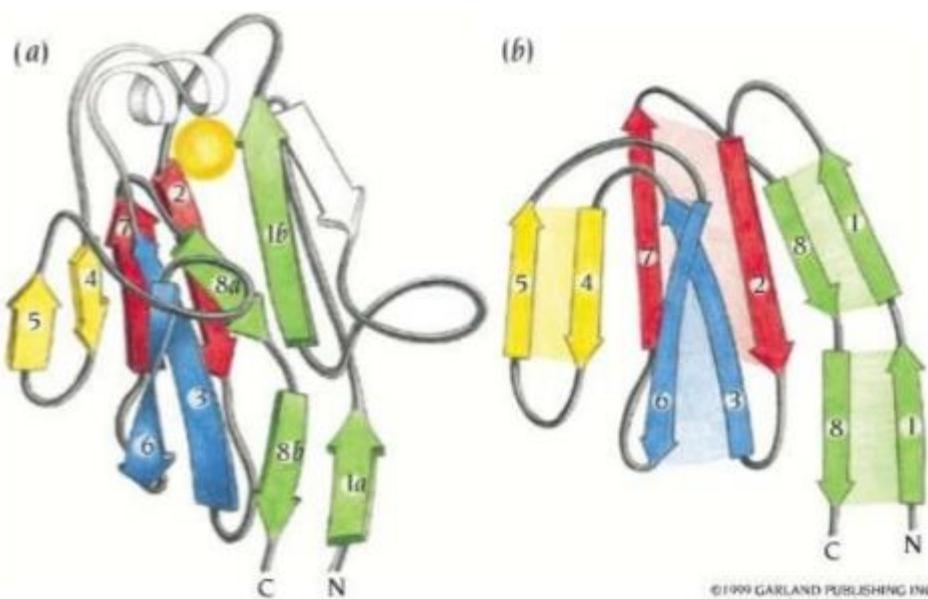
β domain: Jelly roll β -barrel

- Can be formed by Greek key motifs
- Wrapped around a barrel
- Very common in subunits of spherical viruses
- Distorted barrel / β -sandwich

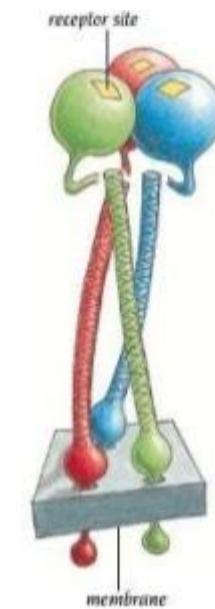




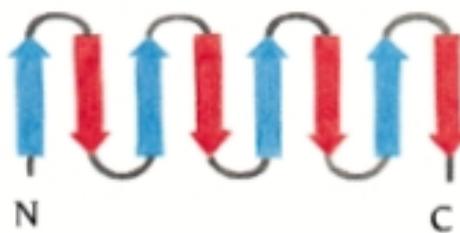
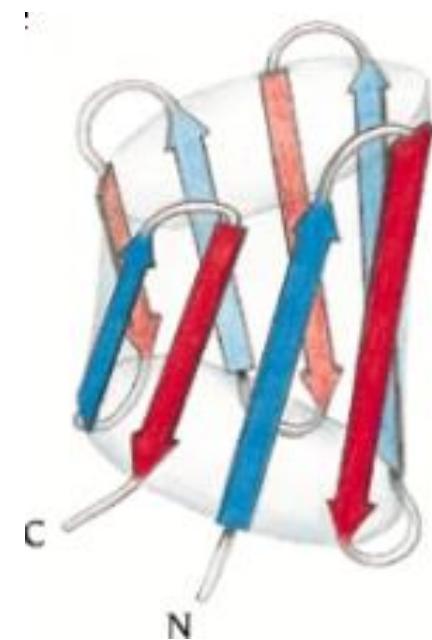
The receptor binding domain of virus coat protein hemagglutinin



©1999 GARLAND PUBLISHING INC



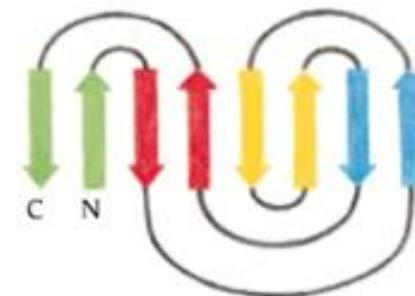
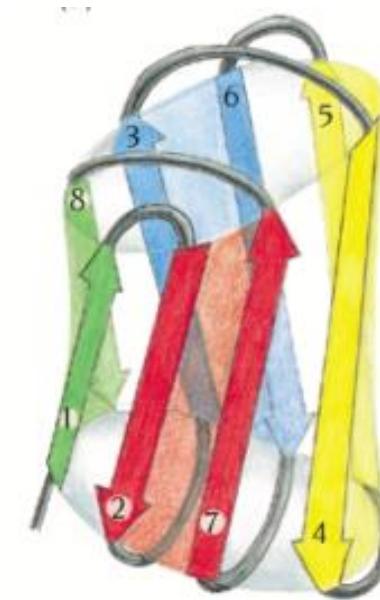
Summary of antiparallel β -barrel



Up-and-down

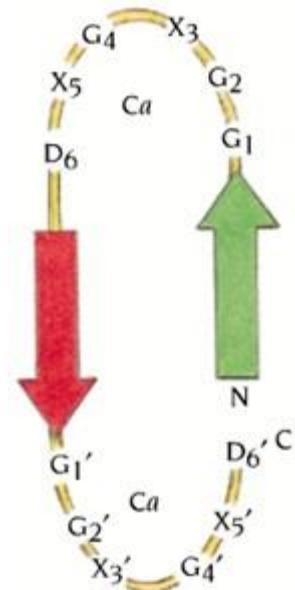
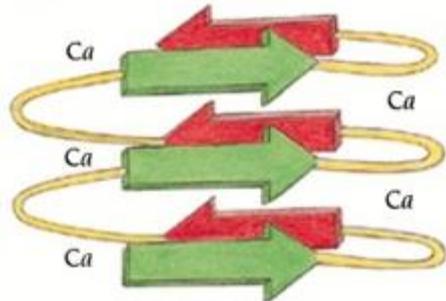


Greek-key/crystallin-like

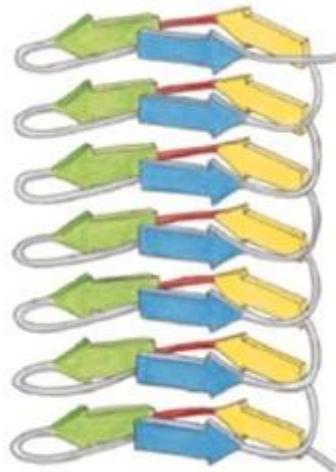


Jelly-roll

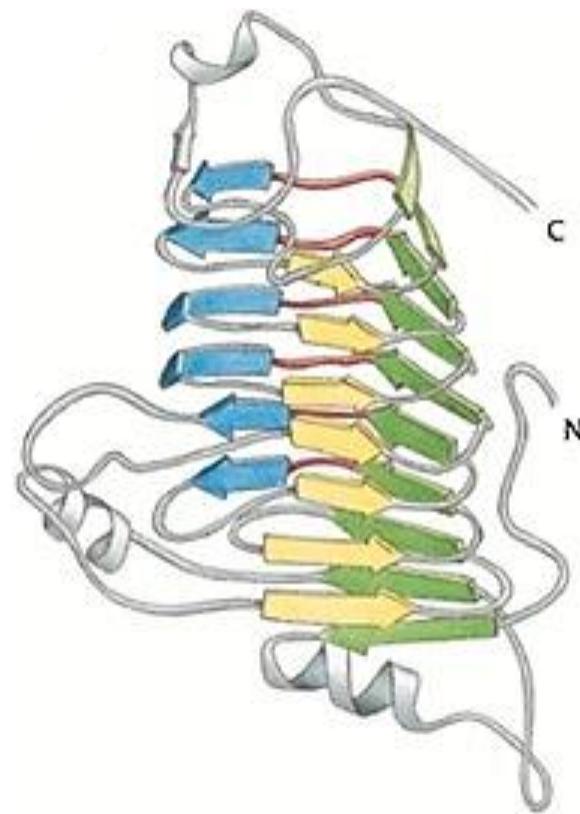
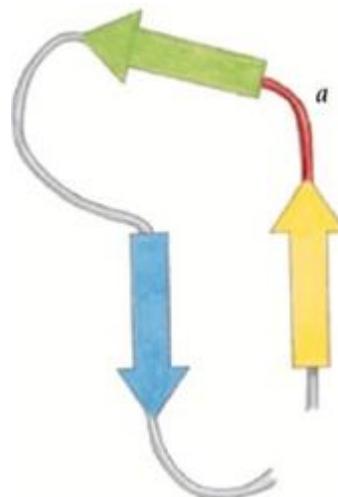
β domain: β -helix (parallel!)



Two sheet β -helix



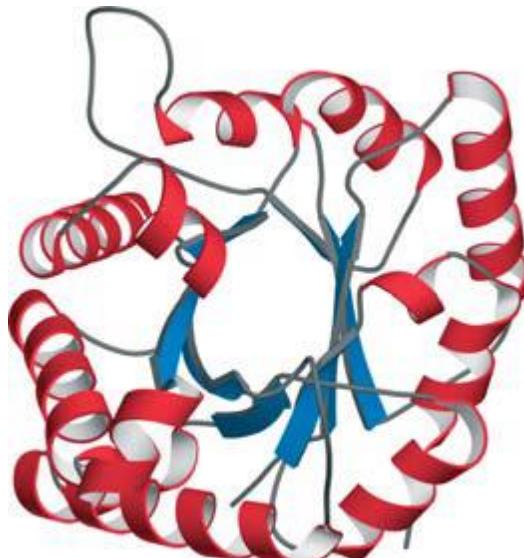
Three sheet β -helix



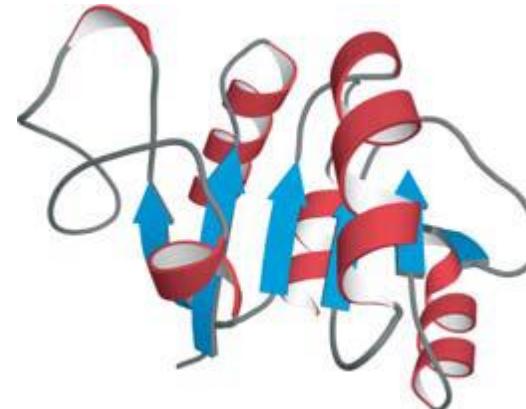
Pectate lyase C (PDB 1O88)

α/β domain

- The most frequent domain structure
- Parallel β strands arranged in barrels or sheets
- The connecting segments are usually α -helices.
- β - α - β motif (**Rossmann fold**) is a general building unit.

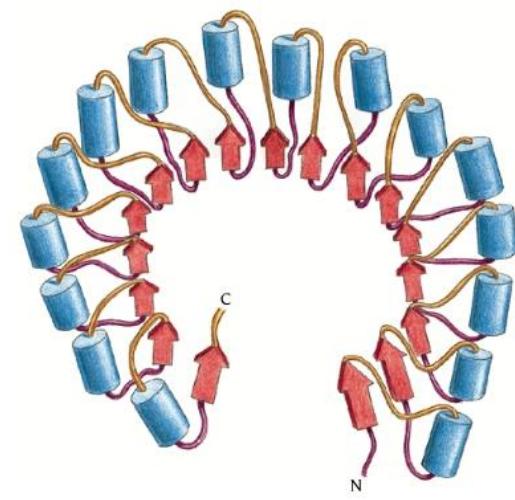
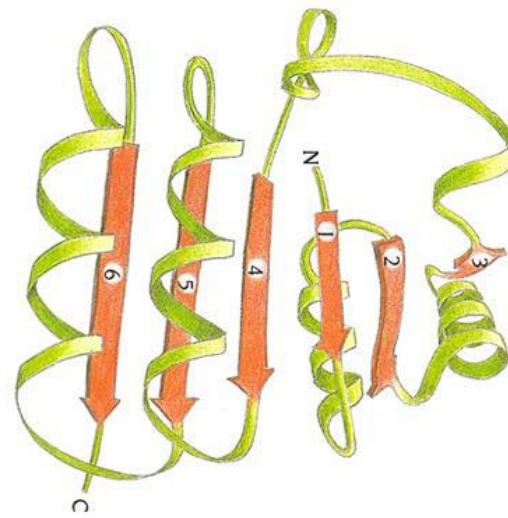
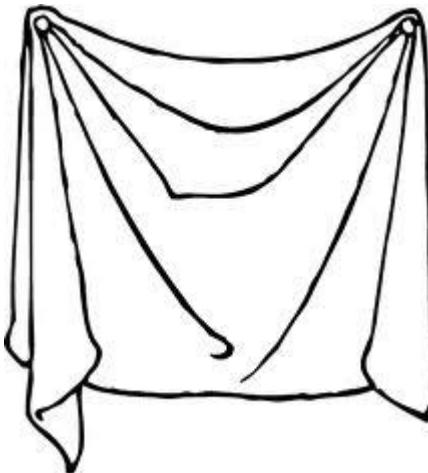


The TIM barrel (PDB **1TIM**)

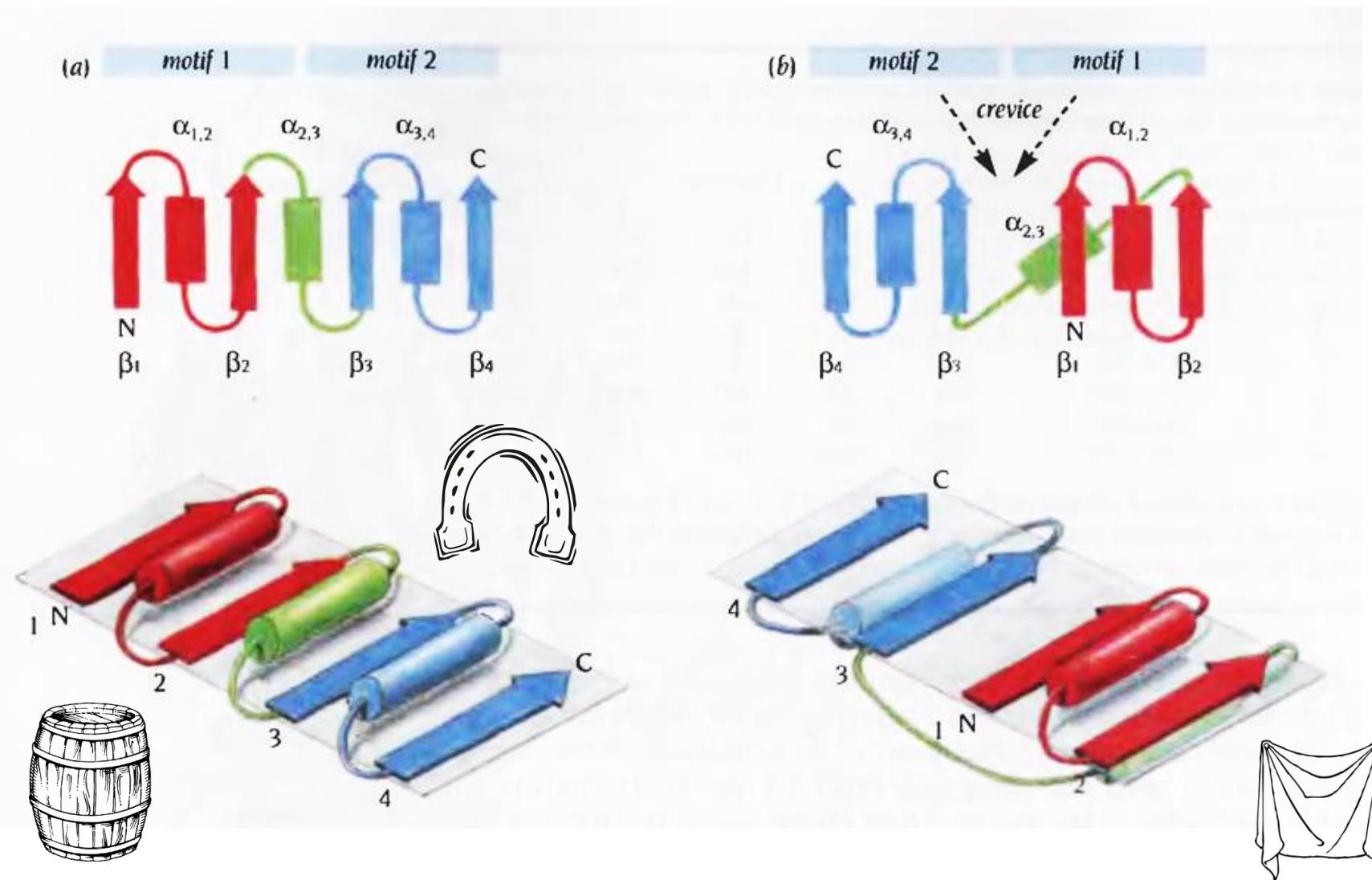


Aspartate semi-aldehyde dehydrogenase (PDB **1BRM**)

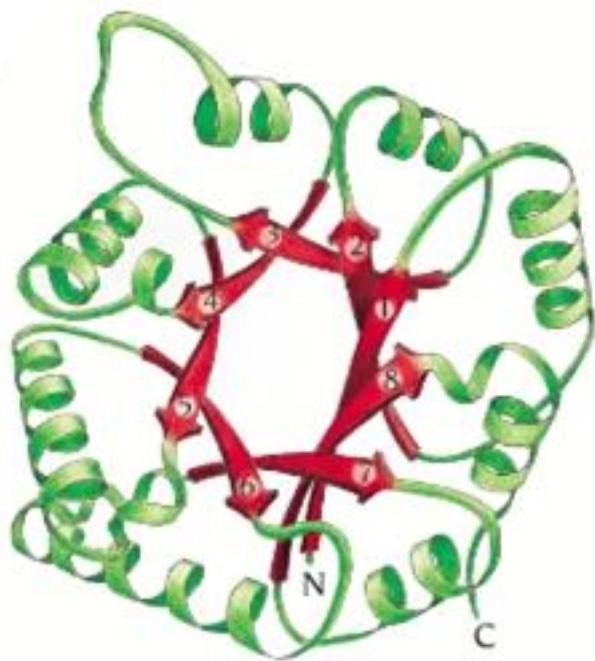
BIO446 Protein Structure and Function



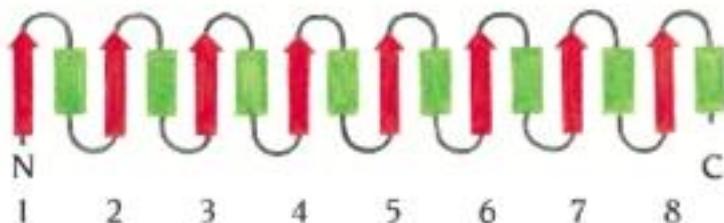
α/β domain: β - α - β motif



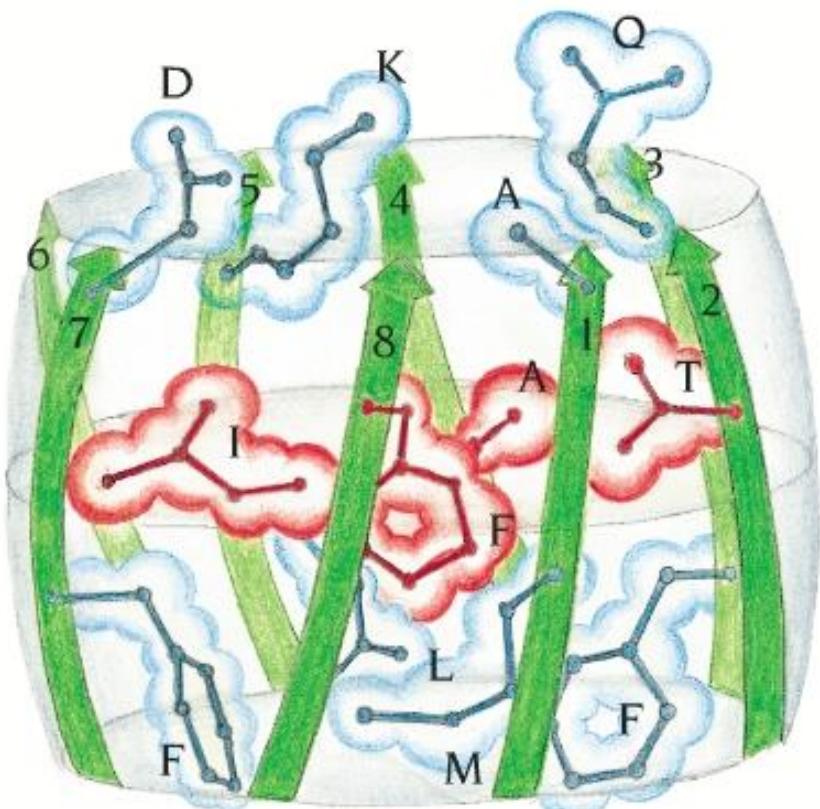
α/β domain: α/β barrel



- Also called **TIM barrel**, first discovered in enzyme **Triosephosphate Isomerase**
- Often found in many different enzymes with little sequence similarity
- The eight-stranded α/β barrel is the most frequent barrel.
- At least four strands are required to form a closed barrel.



The fold of α/β barrel

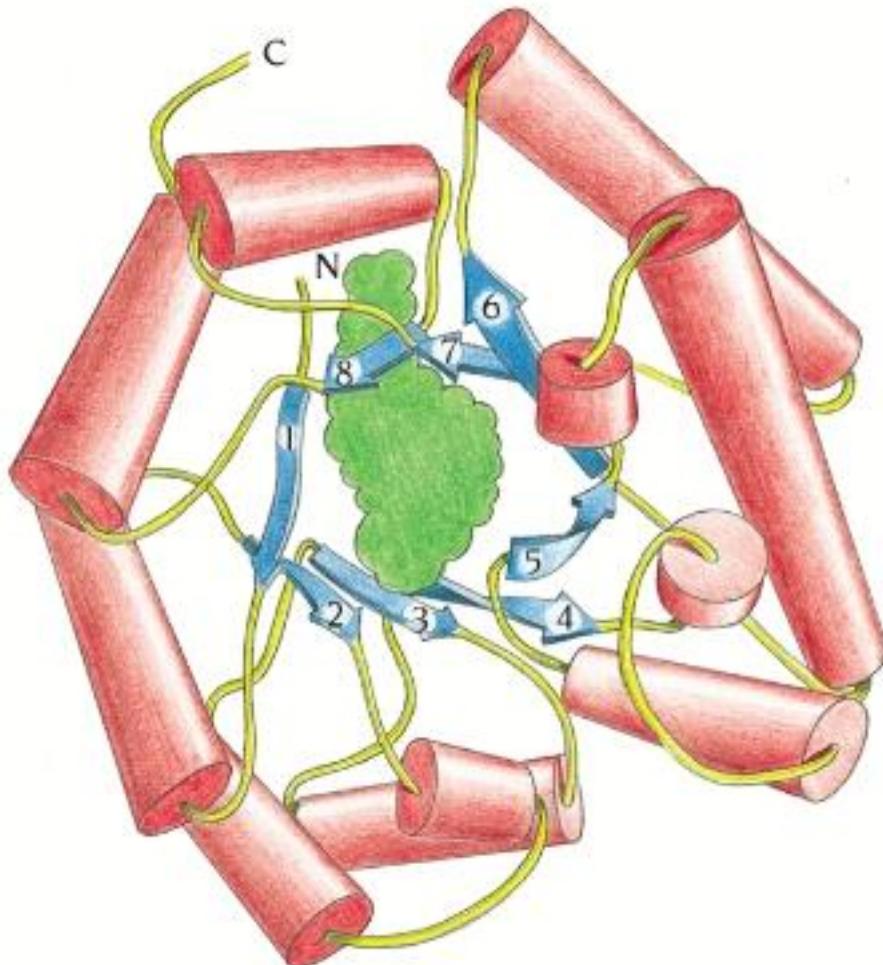


The **branched** hydrophobic side chains dominate the barrel core.

Strand no.	Residue no.	Positions				
		1	2	3	4	5
1	6–10	Phe	Val	Gly	Gly	Asn
2	37–41	Glu	Val	Val	Cys	Gly
3	59–63	Gly	Val	Ala	Ala	Gln
4	89–93	Trp	Val	Ile	Leu	Gly
5	121–125	Gly	Val	Ile	Ala	Cys
6	158–162	Lys	Val	Val	Leu	Ala
7	204–208	Arg	Ile	Ile	Tyr	Gly
8	227–231	Gly	Phe	Leu	Val	Gly

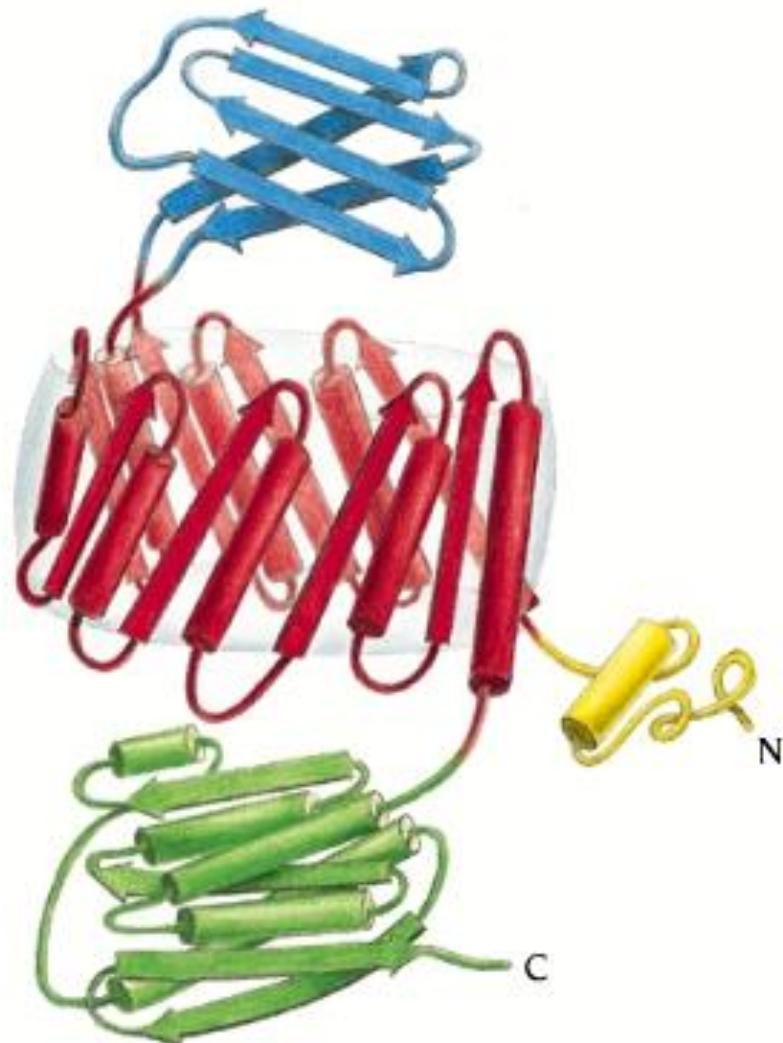
Triosephosphate Isomerase from chicken muscle

Methylmalonyl-Coenzyme A mutase, an exception



- Small hydrophilic residues (T or S) in the barrel core
- Resulting a hydrophilic tunnel
- For coenzyme-A binding and catalysis (hydrophilic environment)

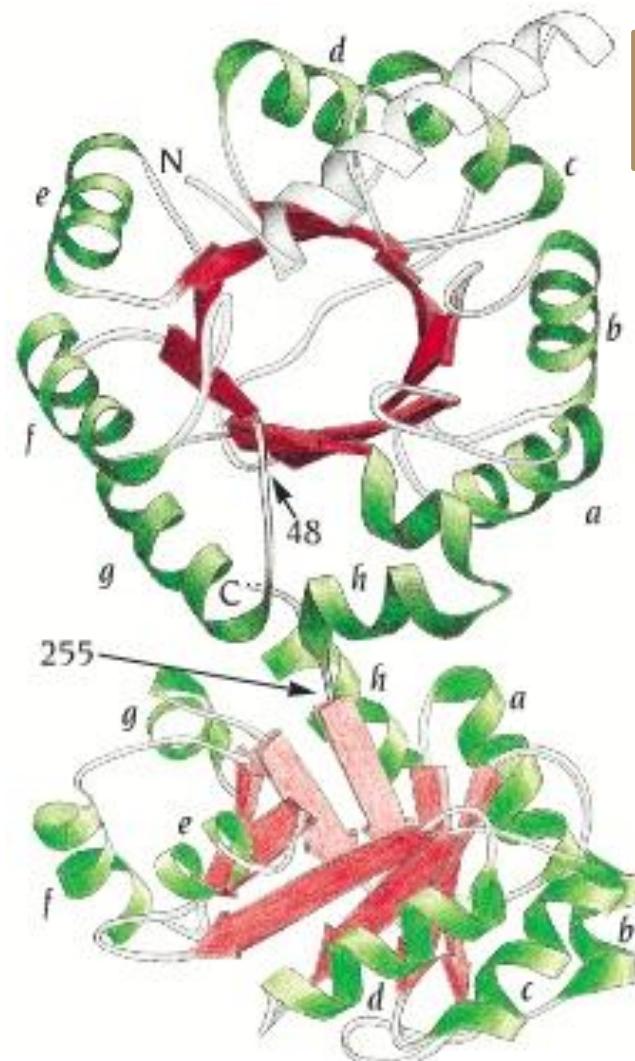
Pyruvate kinase, a multidomain enzyme



PRA-isomerase:IGP-synthase

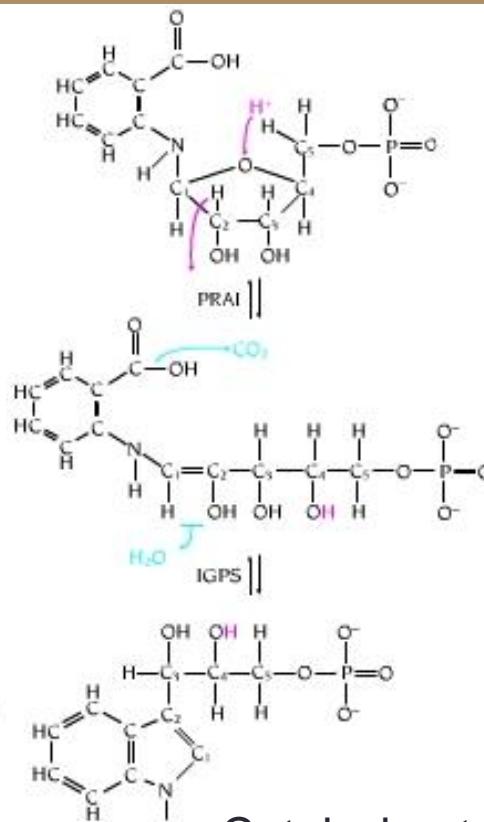
a bifunctional enzyme having double barrels by [gene fusion](#)

IGP-synthase



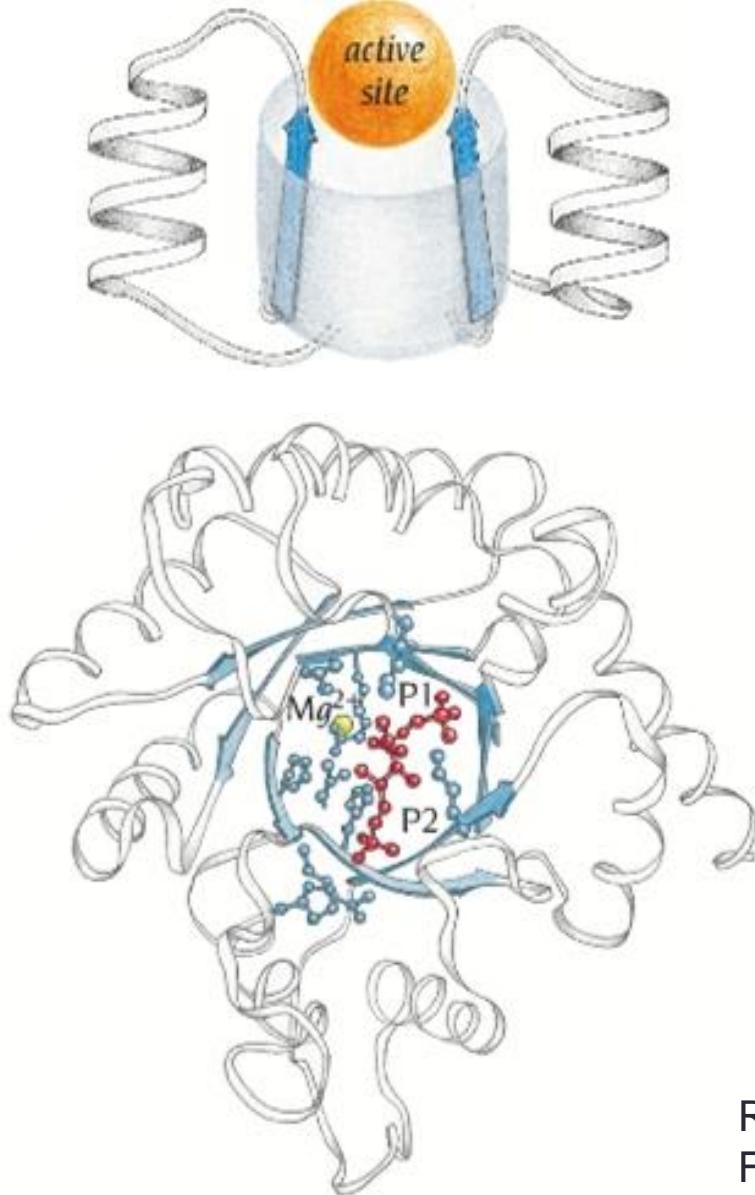
PRA-isomerase

Indicating that new enzymes can evolve by changing the active site of TIM barrel

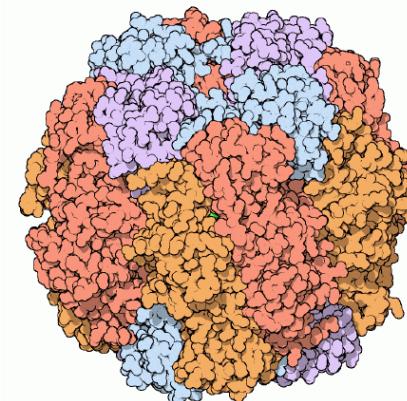


Catalysing two reactions in the tryptophan biosynthesis

The active site in α/β barrel enzyme



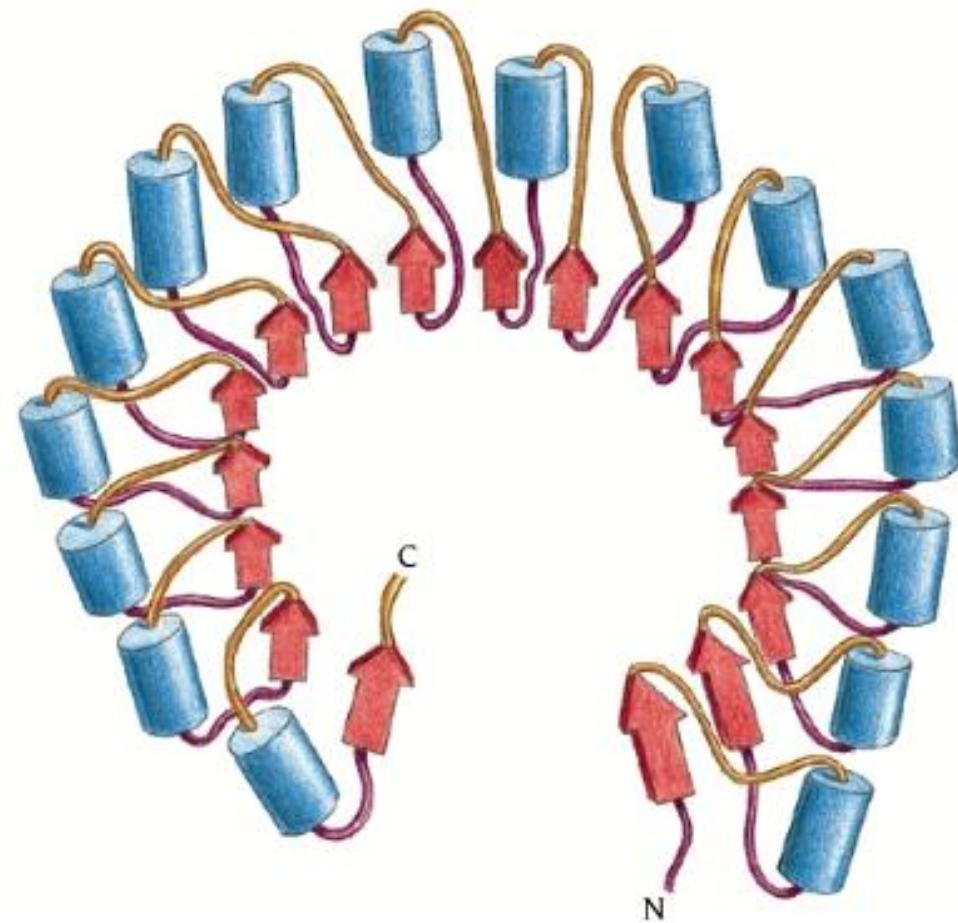
- The active site is formed by loops at one end of the α/β barrel.
- The specific enzymatic activity is determined by the loop regions at the carboxyl ends of the eight strands.
- The loops do not contribute to the structural stability.
- In some cases, loops from other regions come close to the active site and also participate in the binding and catalysis



RuBisCo (ribulose bisphosphate carboxylase), 500kDa
For CO_2 fixation

α/β domain: α/β -horseshoe

- Leucine Rich Repeat (LRR)
- 20-30 residues for each repeat
- Existing in receptors, cell adhesion molecules, bacterial virulence factors, and proteins involved in RNA splicing and DNA repair
- Right-handed β -loop- α repeats
- Form a curved open structure
- The hydrophobic core formed between the β -sheet and the α -helices
- The β -sheet exposed to solvent, a characteristic other α/β structures do not have



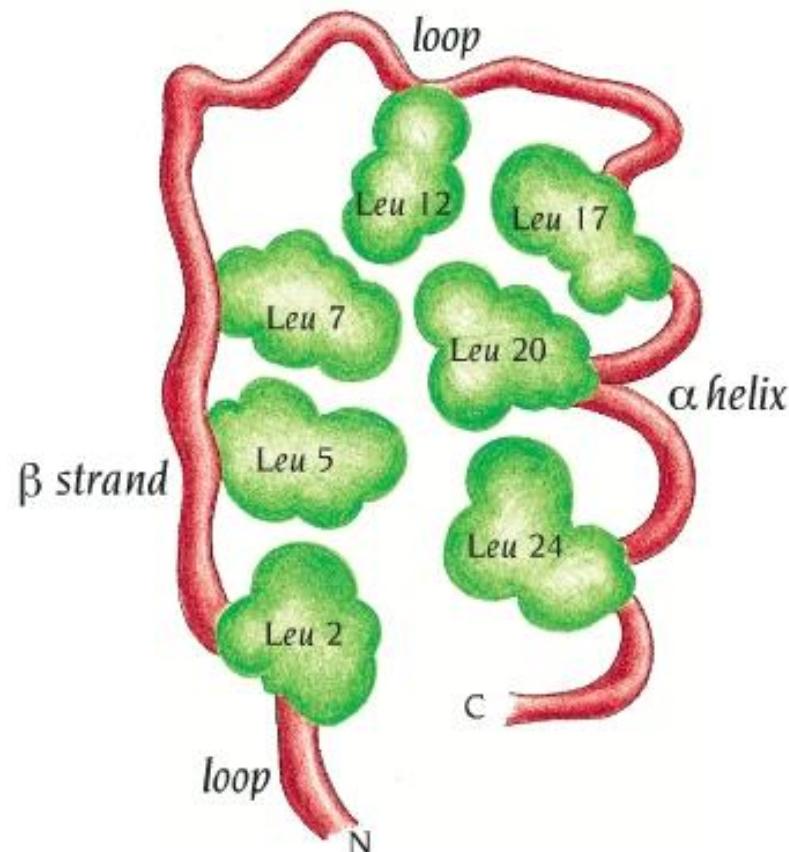
Sequence and structure feature of LRR

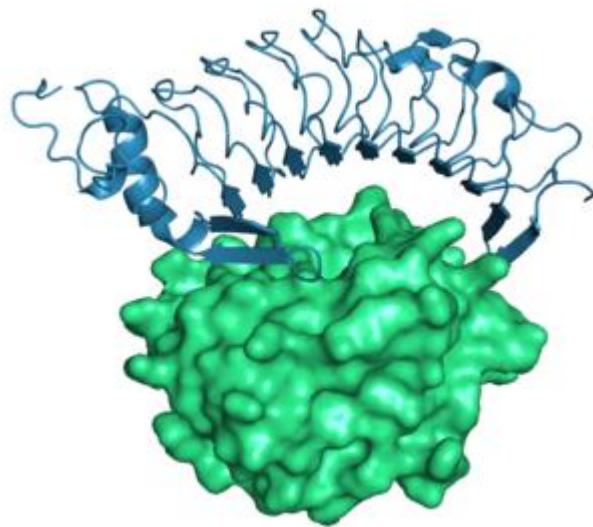
← β strand → loop ← α helix →

	2	5	7	12	17	20	24	
(type A) NH ₂ -	X -	L -	E -	X -	L -	X -	X -	-X -
	X -	C -	X -	X -	T -	X -	X -	X -
	X -	X -	L -	X -	X -	X -	a -	L -
	X -	X -	X -	X -	X -	X -	L -	X -
	X -	X -	X -	X -	X -	X -	X -	X -

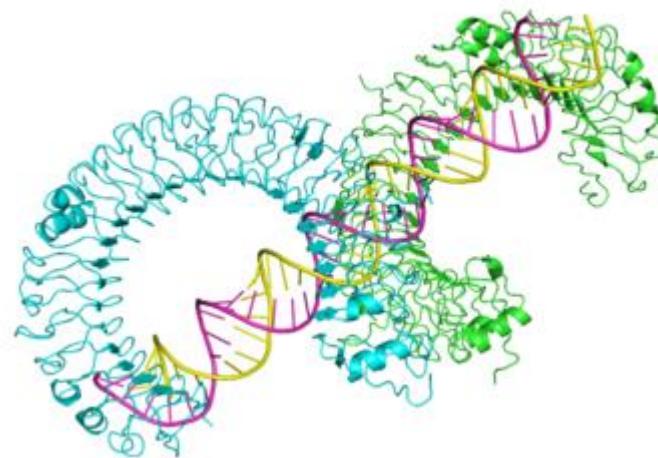
(type B) NH ₂ -	X -	L -	R -	E -	L -	X -	X -	N -	X -	L -	G -	D -	X -	G -	a -	X -	X -	L -	X -	X -	L -	X -	P -	X -	X -	-
----------------------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---

The hydrophobic core is formed by seven conserved leucine in β -strand, α -helix and loop of every repeat.



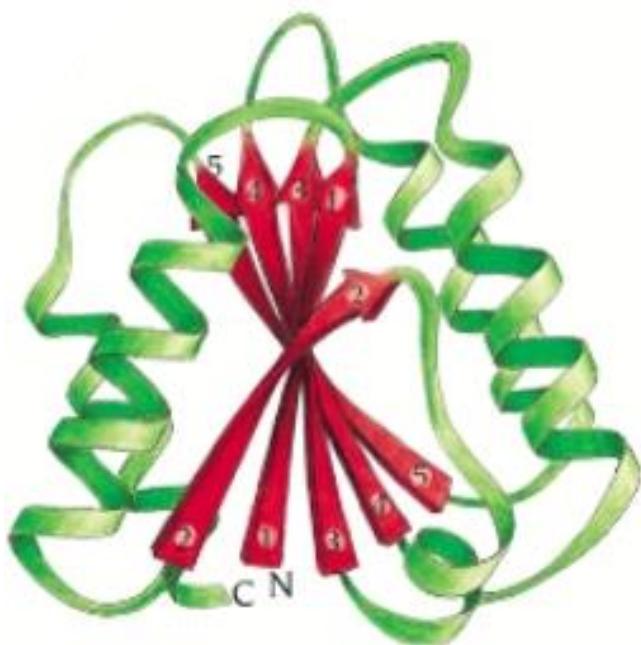


The human leucine-rich repeat family member Platelet-receptor glycoprotein Iba involved in Willebrand disease bound to the von Willebrand Factor A1 Domain. (PDB **1M10**)



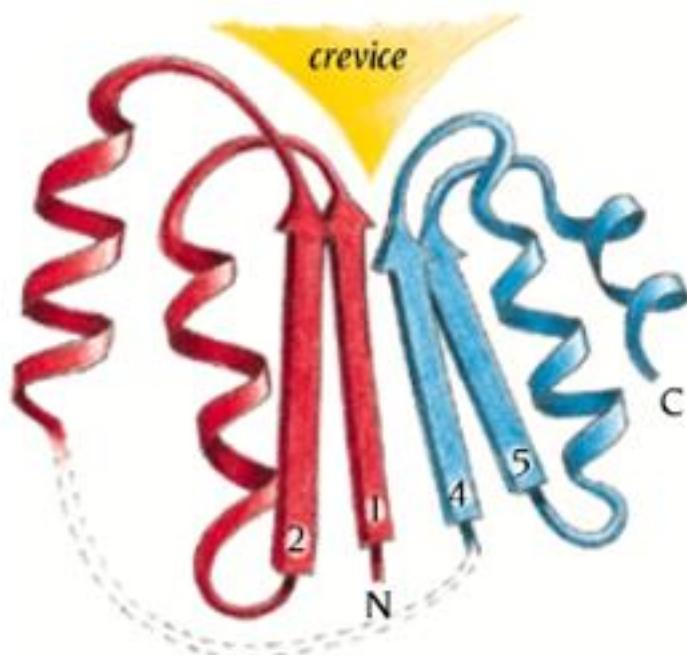
Extracellular domain of mouse Toll-like receptor bound to double-stranded RNA (a molecular signature of many viruses). (PDB **3CIY**)

α/β domain: α/β open twisted sheet



- Parallel or mixed β -sheet
- α -helices on both sides of the sheet
- Each side containing a hydrophobic core
- Each strand contributes to form the core
- Number of β -strands vary from 4 to 10
- Usually having a crevice outside the edge of the sheet between loops

Crevice in open sheet

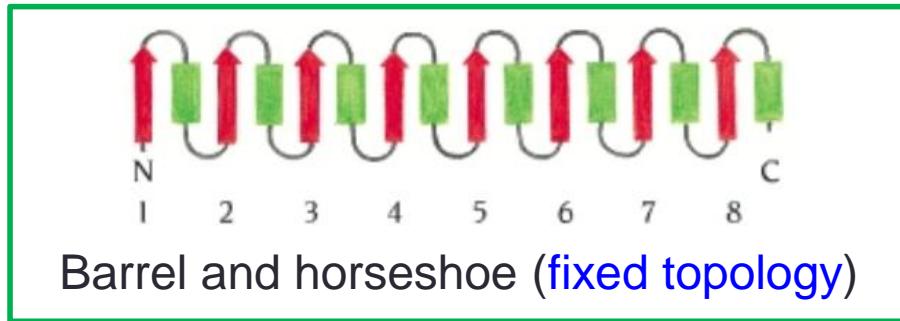


Located outside of the carboxyl end of the strands

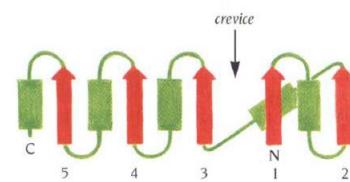
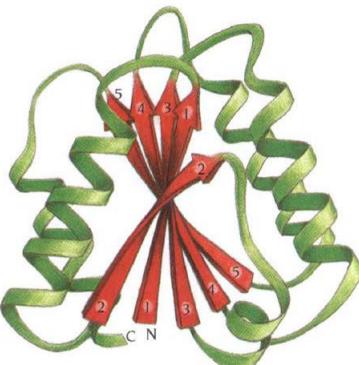


Mostly a binding site

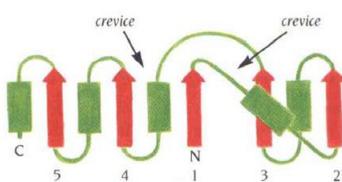
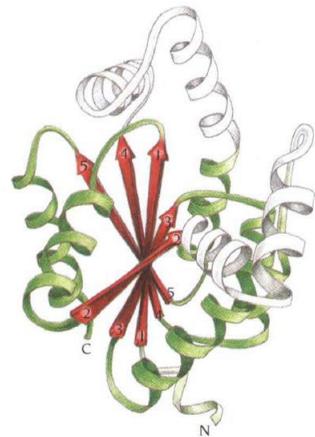
Varied topologies in open sheet structures



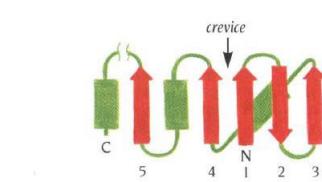
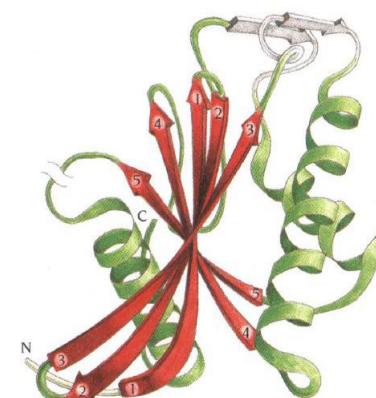
Open sheet (varied topology)



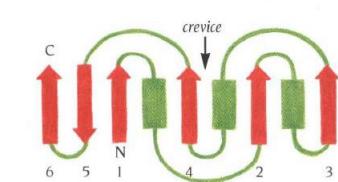
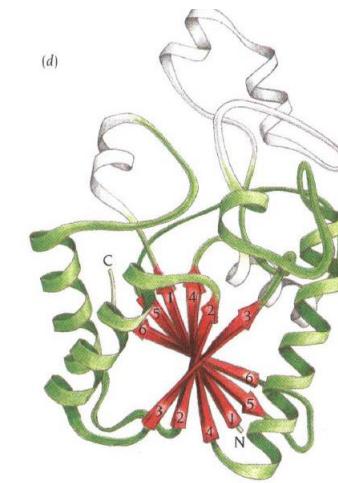
Flavodoxin



Adenylate kinase

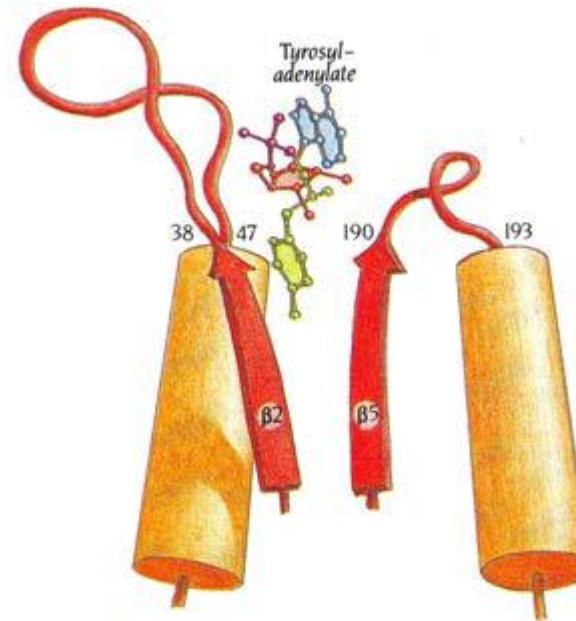
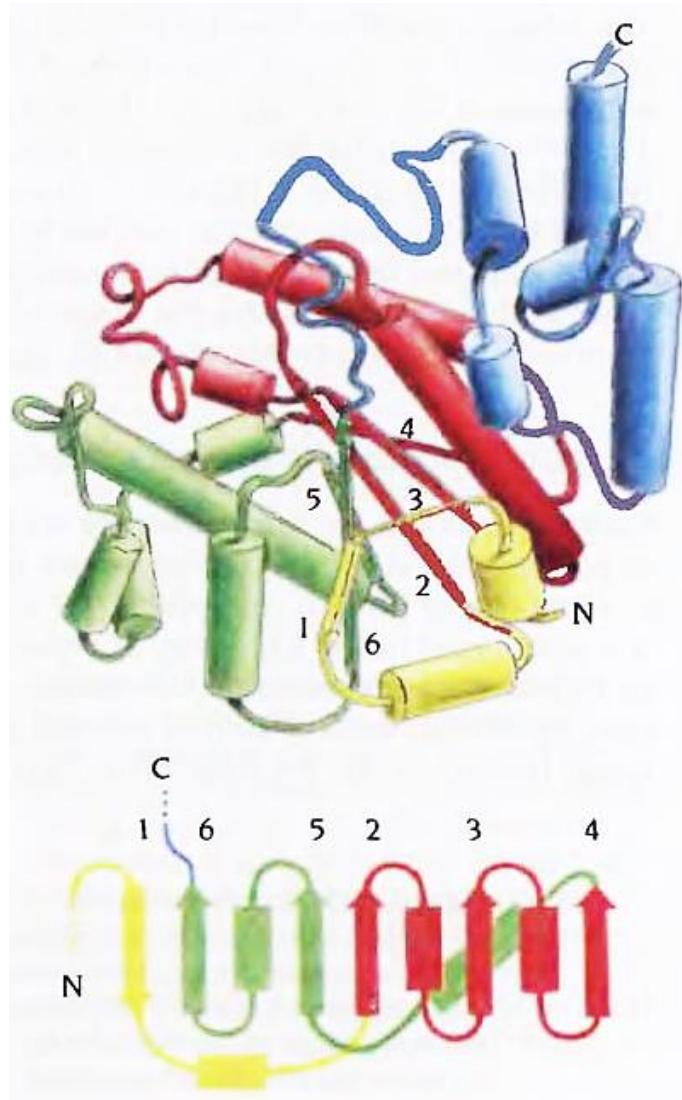


Hexokinase

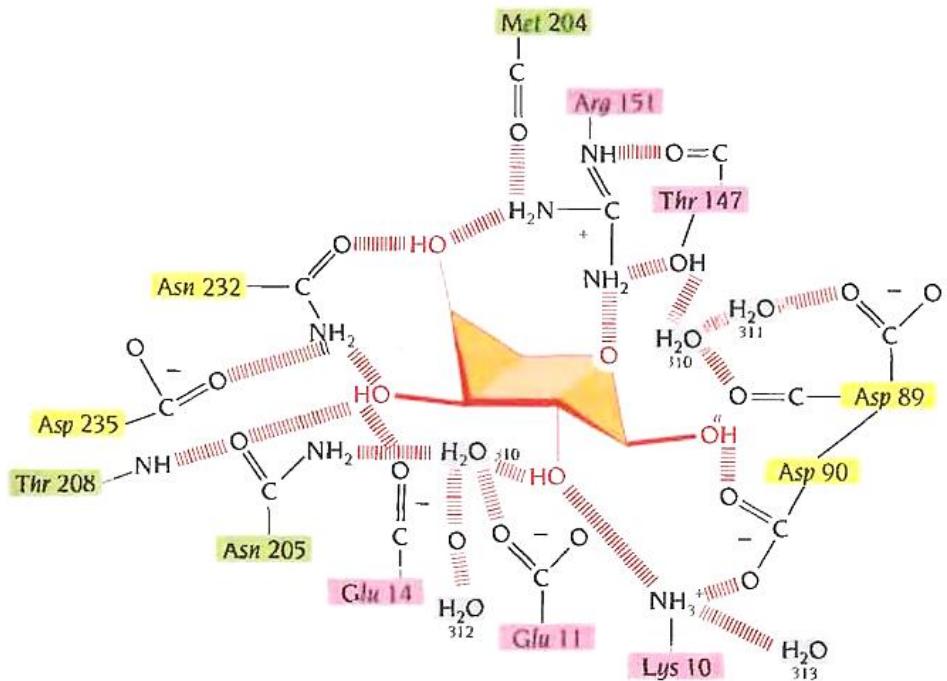
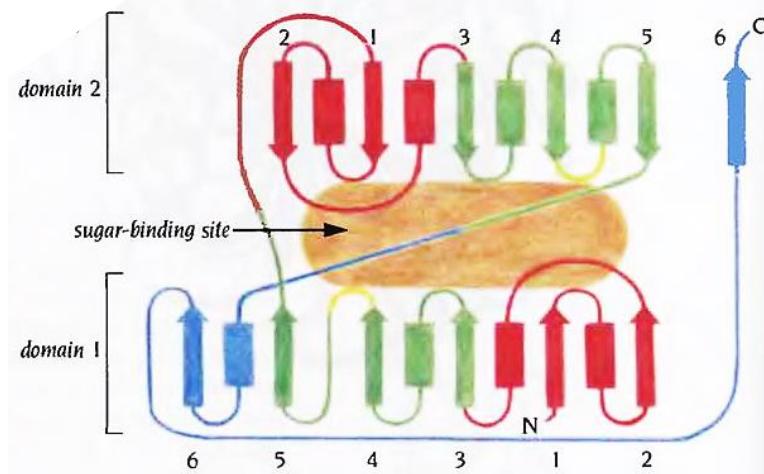
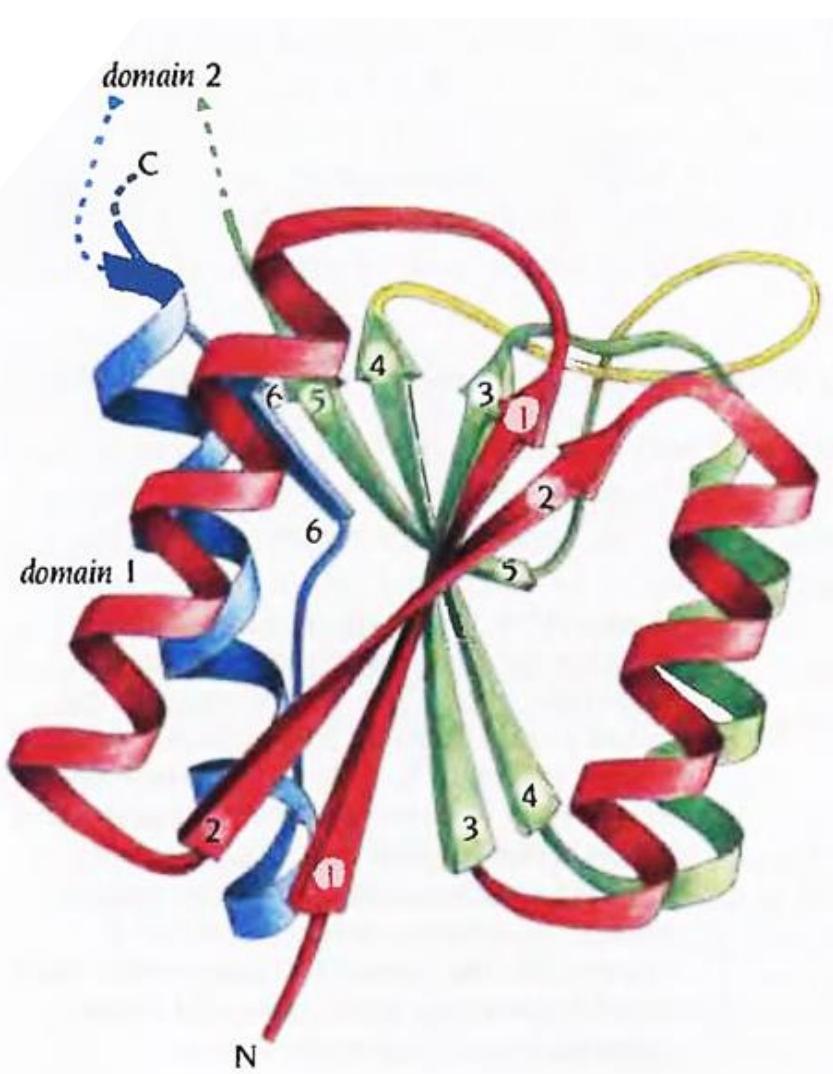


Phosphoglycerate mutase

Tyrosyl-tRNA synthetase

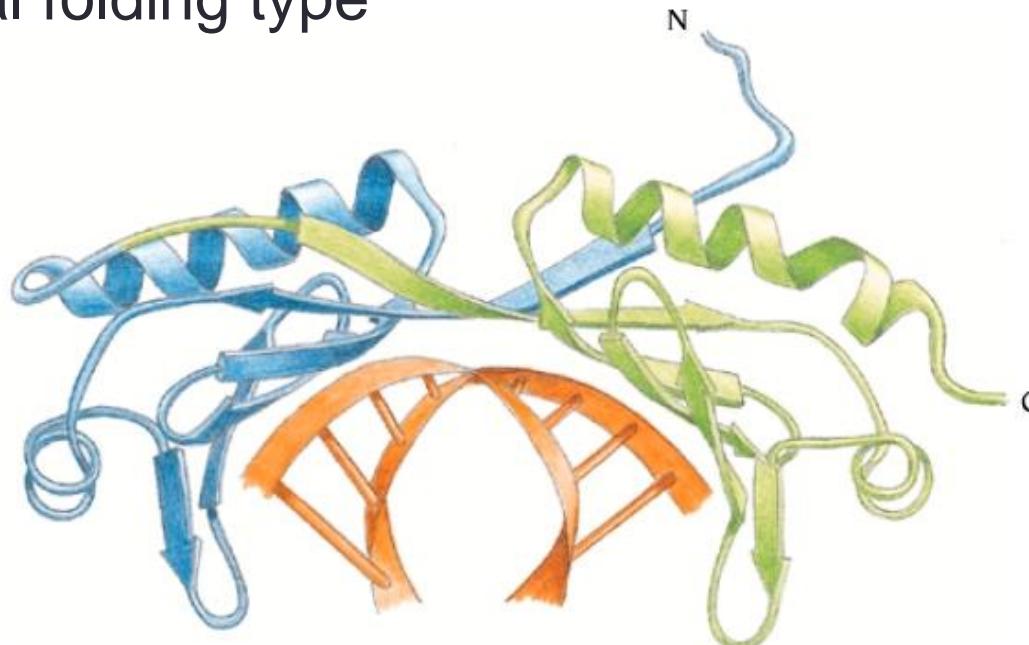


Arabinose-binding protein

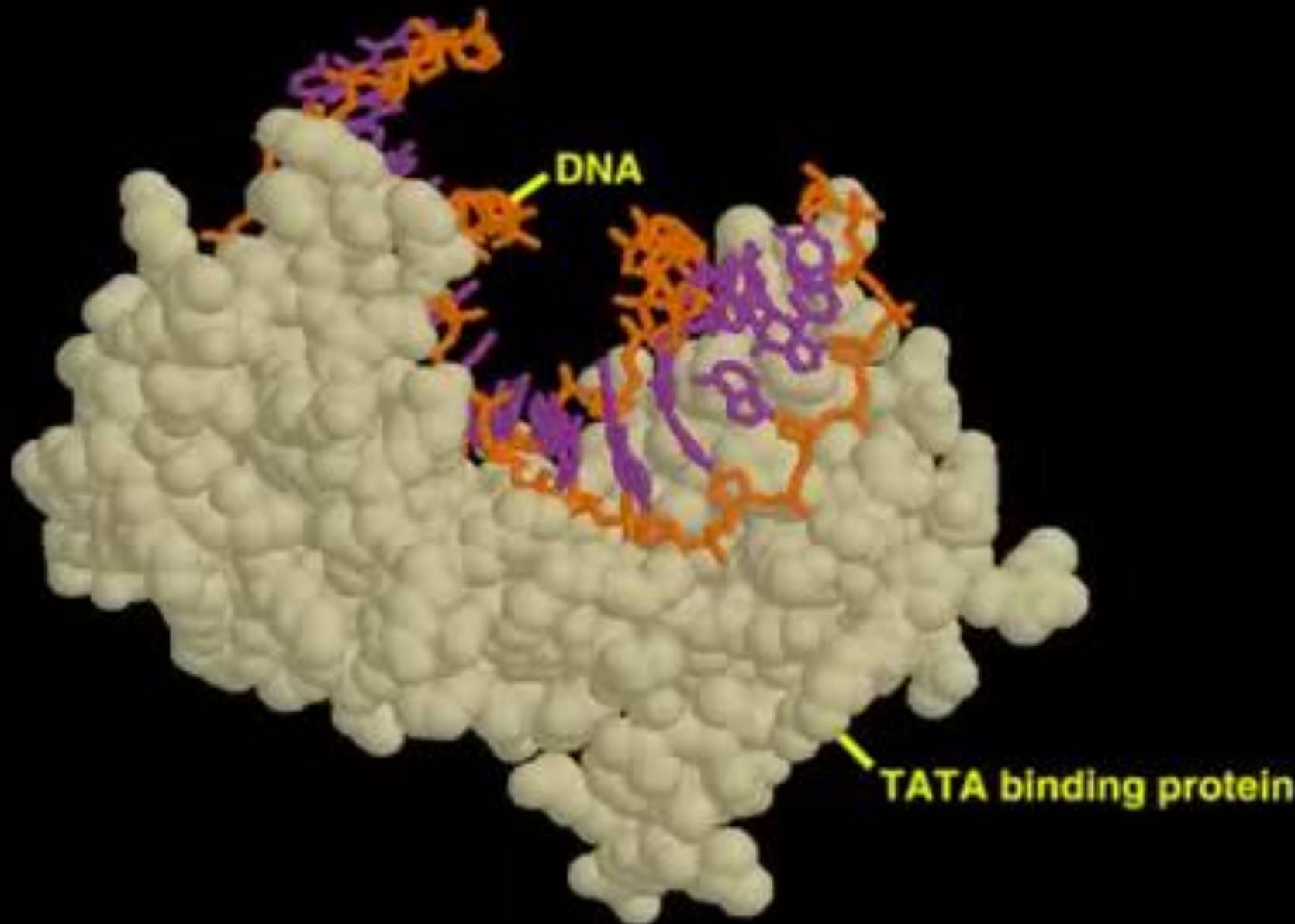


$\alpha+\beta$ domain

- Containing segregated sheets and helices
- Usually antiparallel or mixed sheets
- No general folding type



$\alpha+\beta$ saddle, the TATA-binding protein that binds to DNA at the TATA box. The sheet that forms the seat of the saddle binds in the minor groove of the DNA, bending it significantly. (PDB 1TGH)



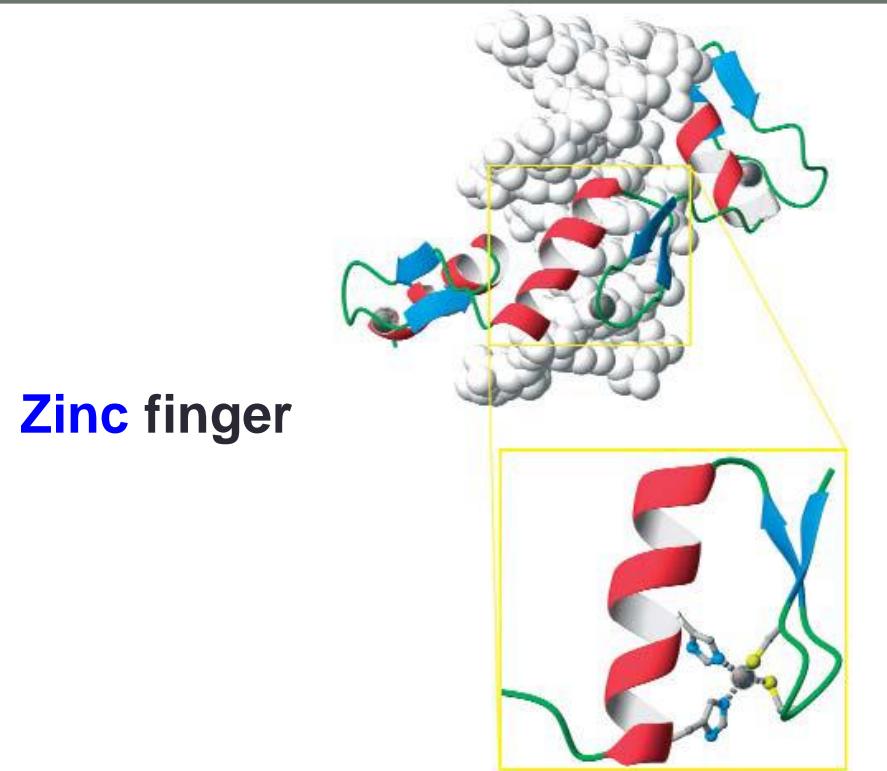
Cross-linked domain

- Metal ions and disulfide bridges
- Irregular domains
- Small size

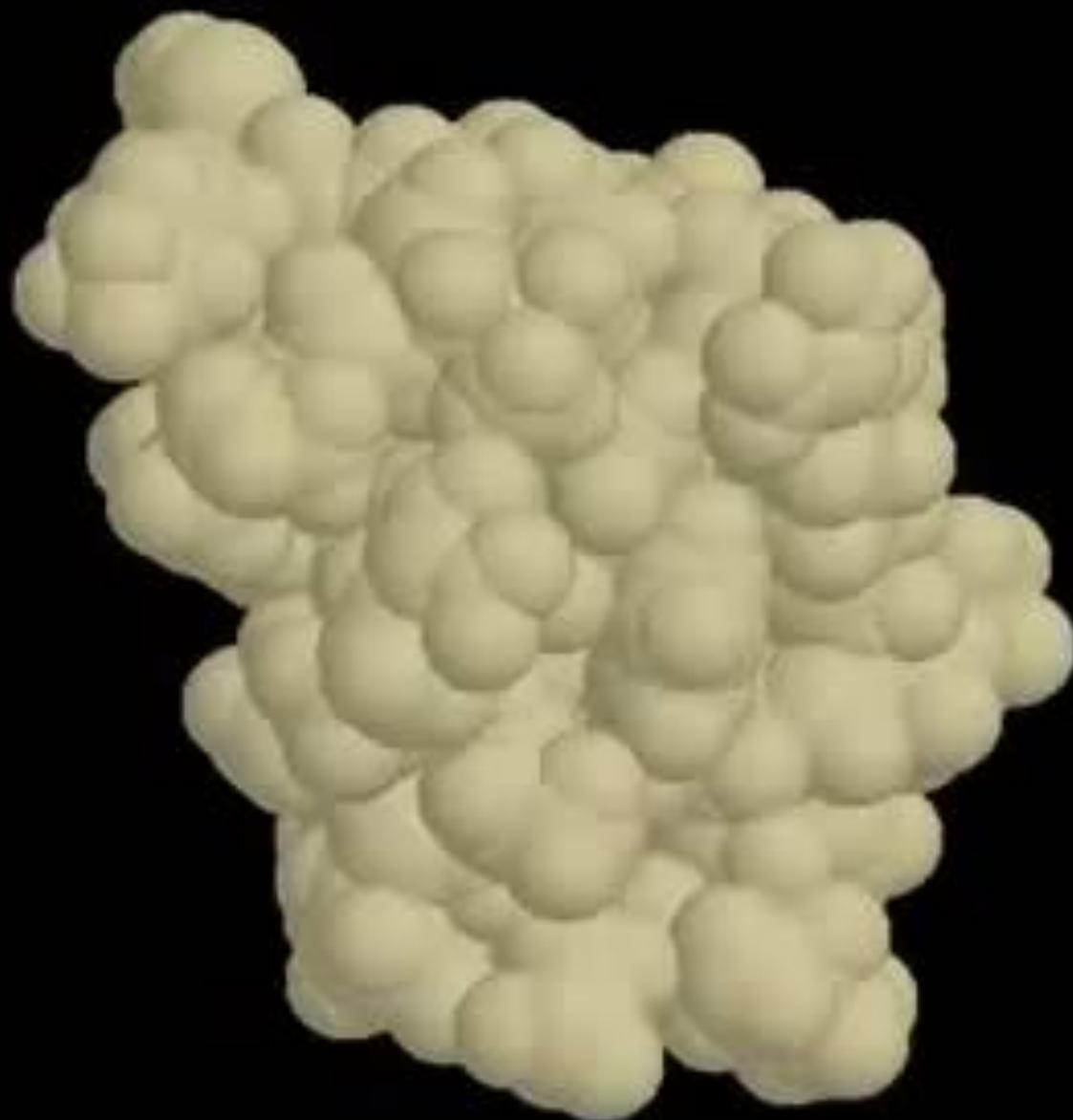


**Disulfide-linked
extracellular
small proteins**

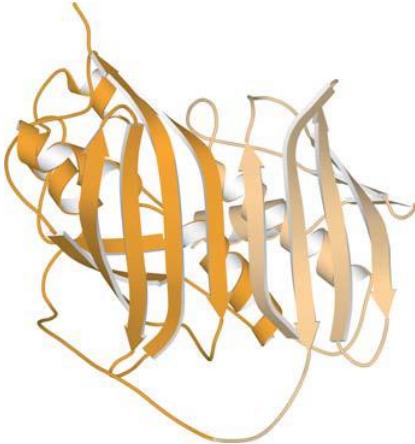
Scorpion toxin, a small irregular extracellular protein with no large hydrophobic core and minimal secondary structure. It is stabilized by four disulfide bridges. (PDB 1B7D)



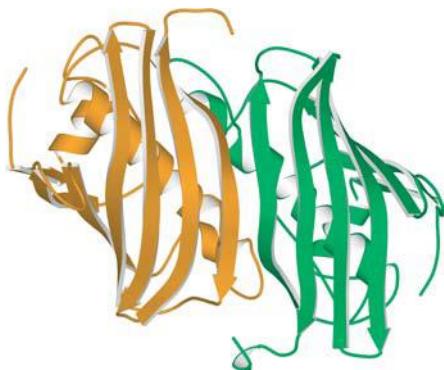
A mouse gene regulatory protein with three zinc fingers bound spirally in the major groove of a DNA molecule. (PDB 1AAY)



Multi-domain proteins

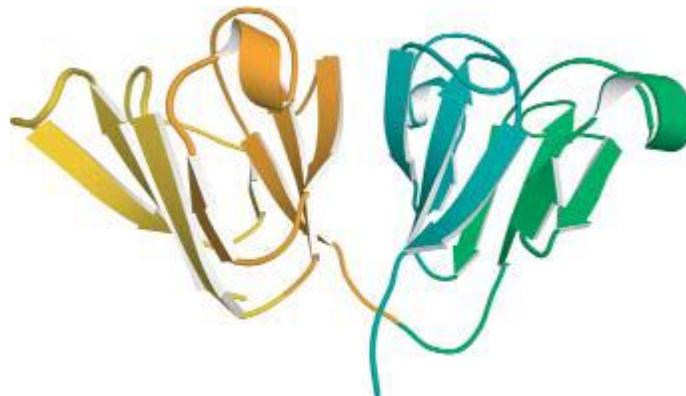


E. coli thioesterase is composed of two nearly identical domains (PDB 1C8U).



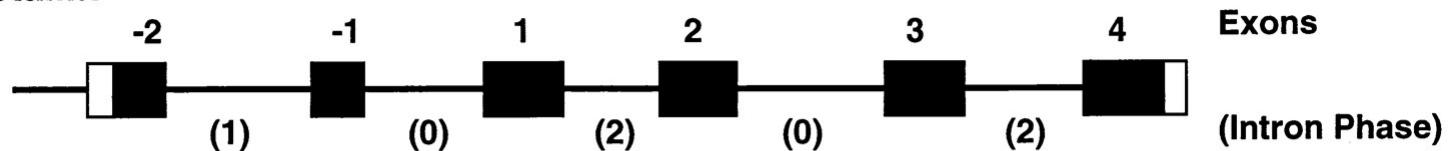
Thioester dehydrase is composed of two identical subunits (PDB 1MKB).

- **Evolutionary mechanism**
 - Multi-domain proteins may be evolved by the fusion of genes that once coded for separate proteins.
- **Inter-domain interaction**
 - In many cases, domains in one protein can interact with each other for some structural and functional purposes.

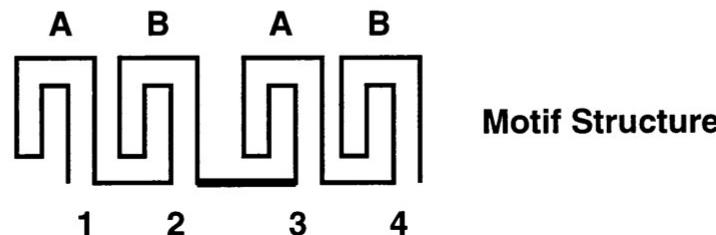
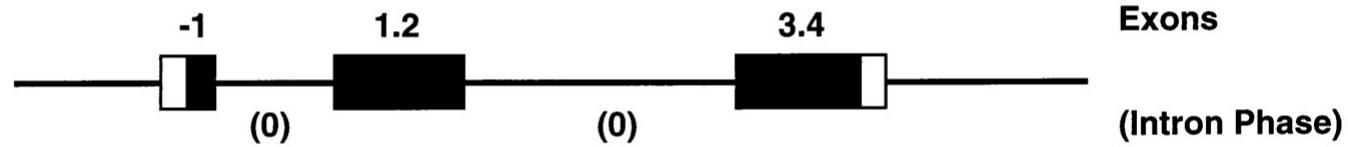


γ -crystallin is composed of two nearly identical domains. Each domain is also made up of two nearly identical halves. (PDB 1GCS)

β -crystallin



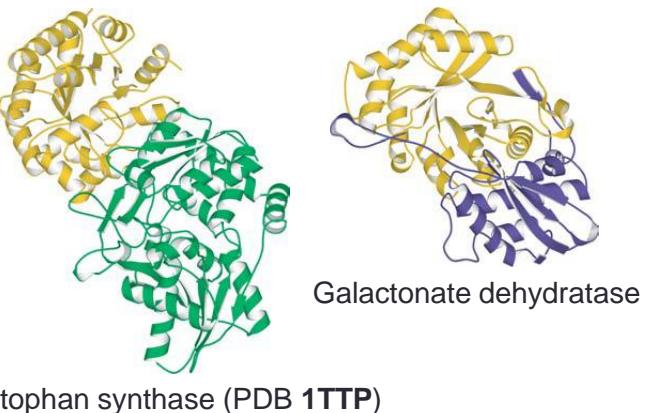
γ -crystallin



Protein Family

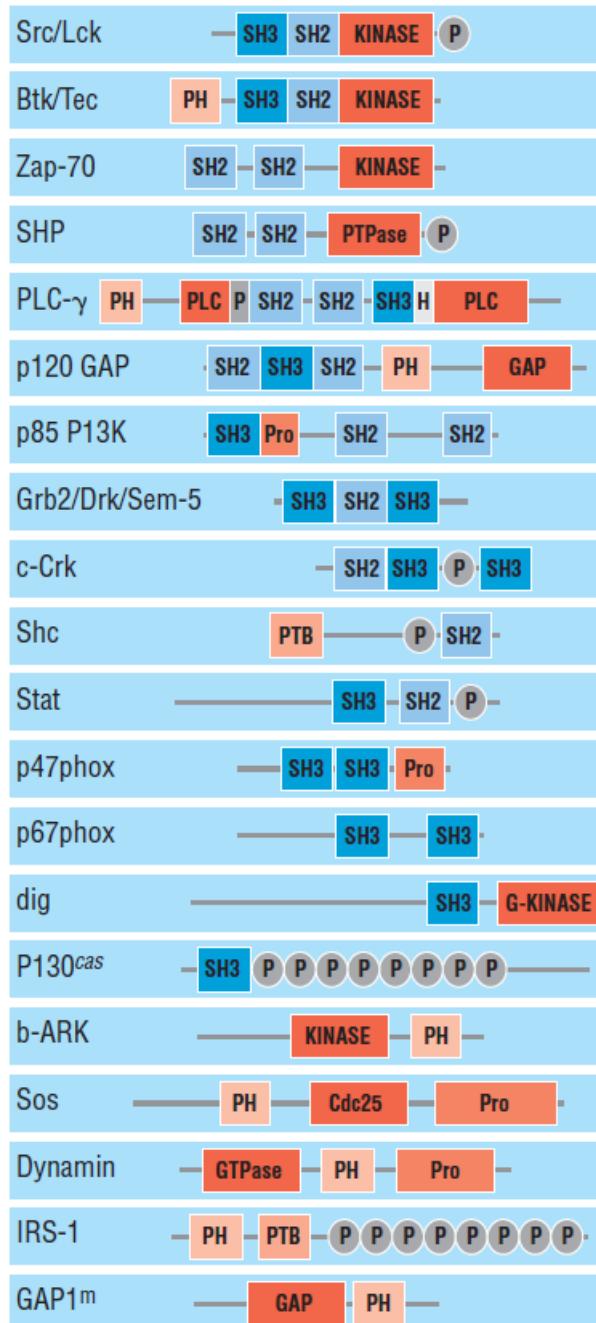
- Protein folds

- Limited number
- Domain fold
- Motif

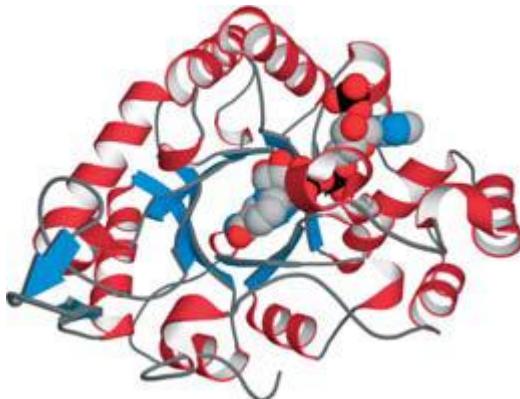


Tryptophan synthase (PDB 1TTP)

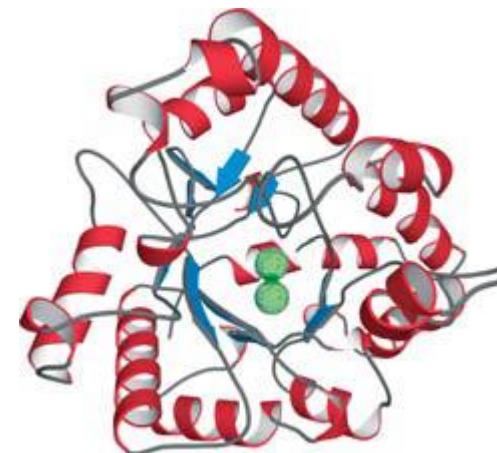
- Almost unlimited combinations
- Protein structures are modular.
- Proteins can be grouped into families **on the basis of the domains they contain.**



Similar fold does not mean similar function (Divergent evolution)



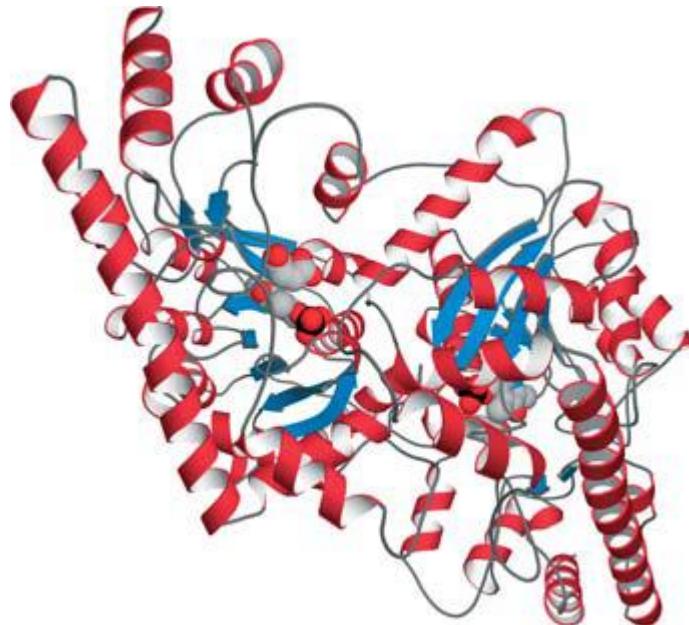
Aldose reductase
(PDB **1ADS**)



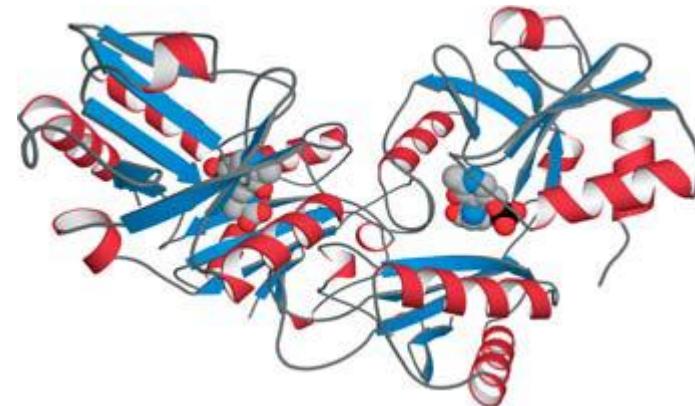
Phosphotriesterase
(PDB **1DPM**)

Although the overall folds are very similar, aldose reductase uses NADPH to reduce sugars, while phosphotriesterase hydrolyzes phosphate groups using a bimetallic cofactor.

Different fold does not mean different function (Convergent evolution)

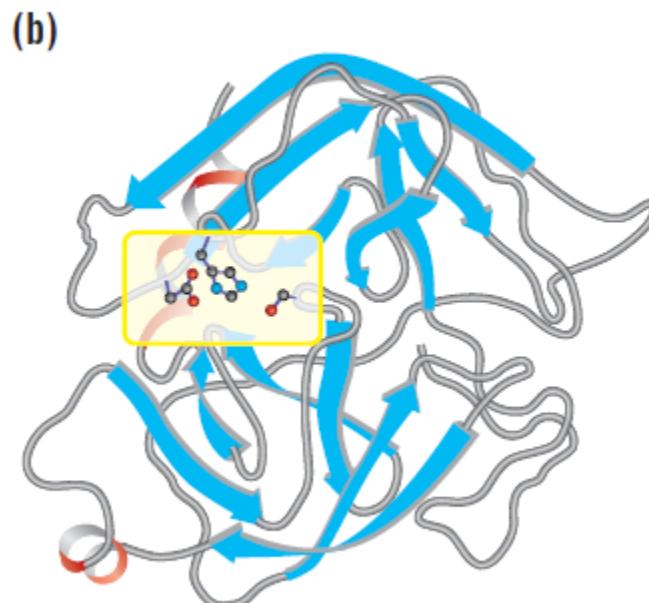
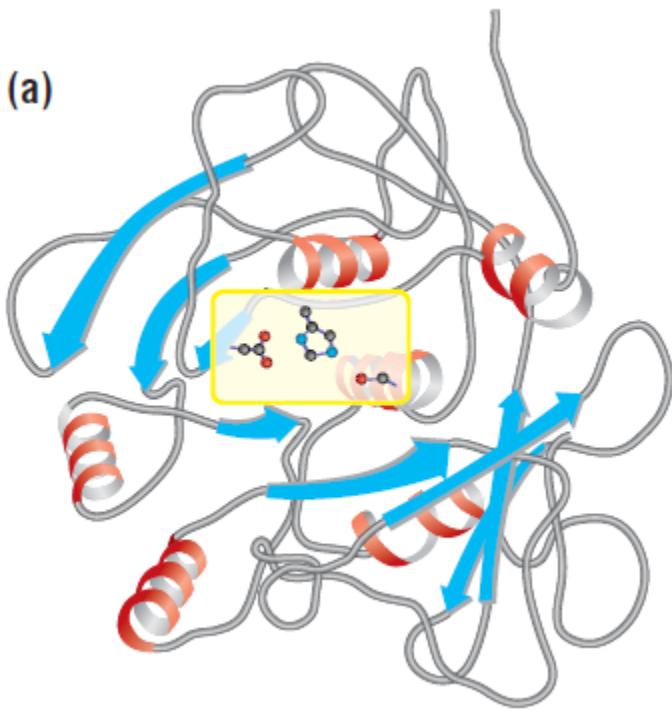


Aspartate aminotransferase
(PDB 1YAA)



D-amino acid aminotransferase
(PDB 3DAA)

Both enzymes catalyze the same reaction; but they have no structural similarity to each other at either the sequence or the tertiary level. Only the active sites, shown by the presence of the cofactor in space-filling representation, are very similar.



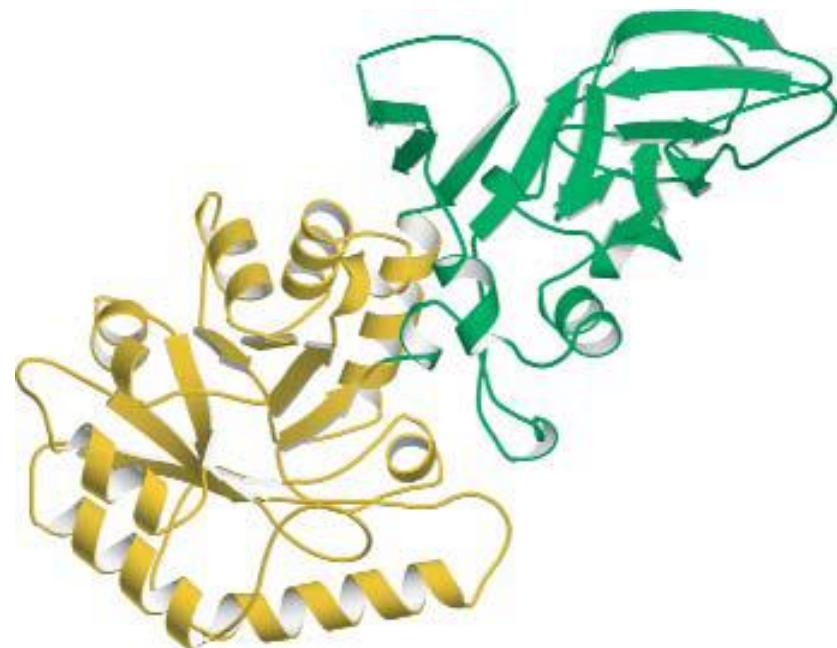
The catalytic triad of aspartic acid, histidine and serine in (a) subtilisin, a bacterial serine protease, and (b) chymotrypsin, a mammalian serine protease.

The modular nature of protein structure

- Sequence insertions and deletions
 - Evolution
 - Short sequence
 - Entire domain
- Within a given protein family, insertions and deletions mostly occur in **surface loops**.
 - Little effect on the packing of helices and sheets
 - An evolutionary way to generate a new protein fold and function

Domain insertion and split domain

- In some proteins, a domain is interrupted by a block of sequence.
 - The sequence usually forms another domain.
 - The insertion frequency is higher in eukaryotes.



Structure of alanine racemase having its structural domains (yellow) interrupted by insertion of another domain (green)