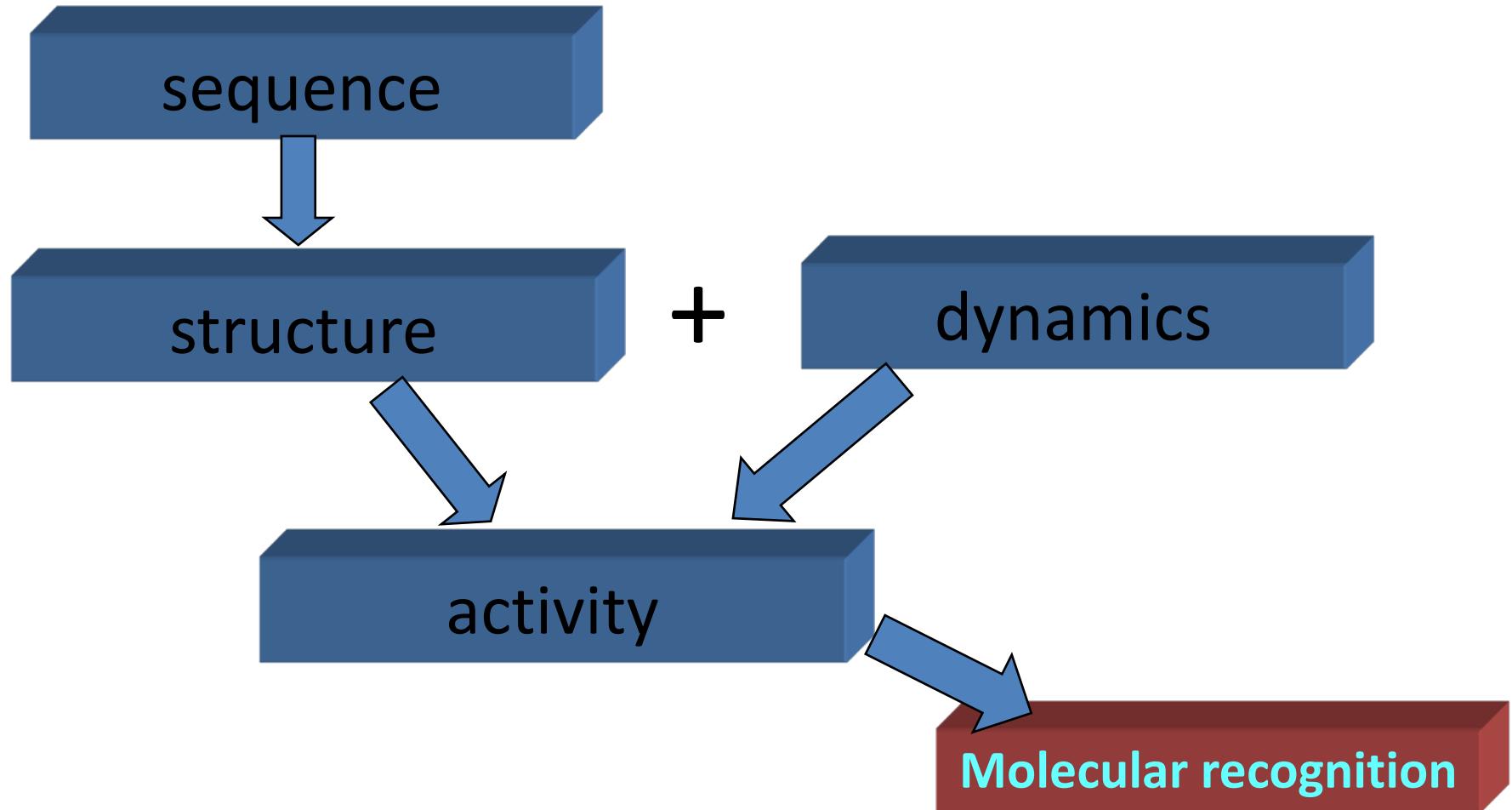


Protein Structure

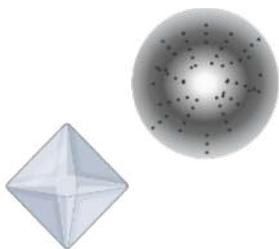
: a hands-on approach

Part 1: Some justification

Why does it matter?



Experimental Structure Determination

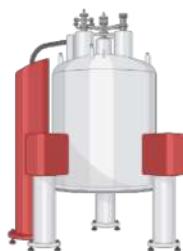


X-ray diffraction
(protein crystallography)

1960-

In principle, size is not a limitation

Crystals required



NMR spectroscopy

1980-

Size IS a limiting factor

Solution measurements



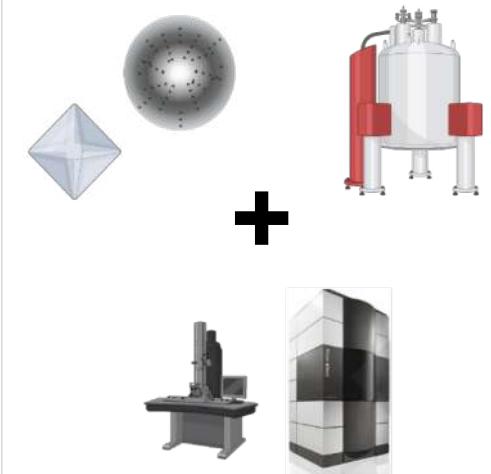
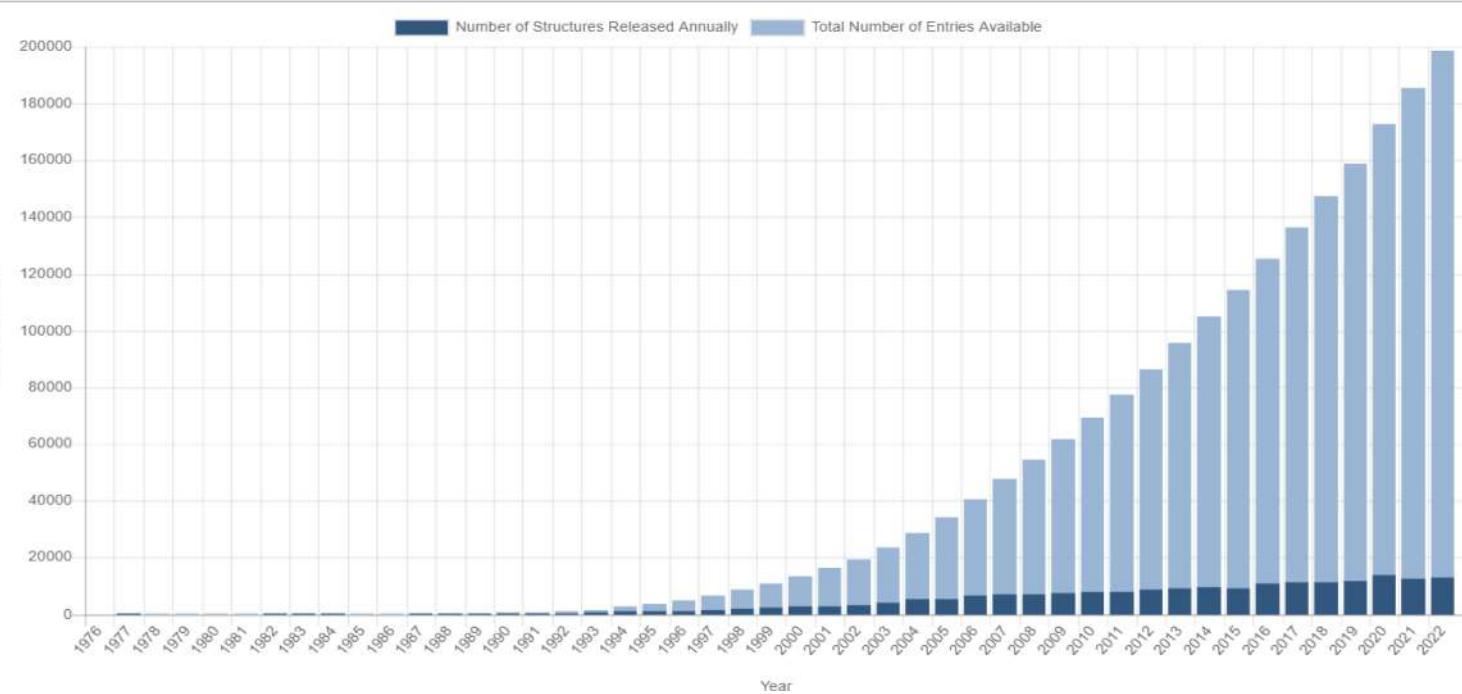
Cryo-EM
(single particle analysis)

2010-

Large Particles (but changing rapidly)

Single Particles
Different conformational States

The state of structural biology



AlphaFold
Protein Structure Database
Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism BETA Search

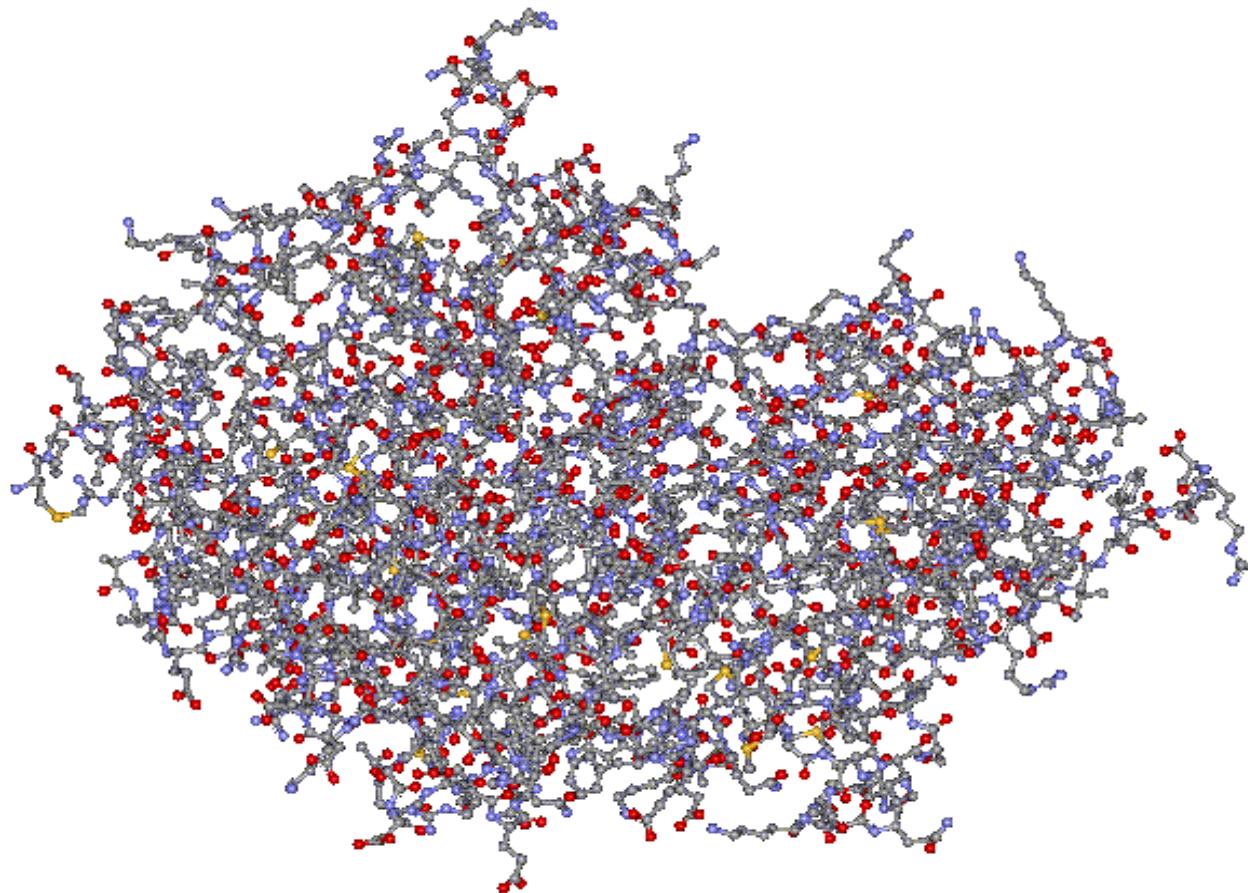
Examples: Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli Help: AlphaFold DB search help

Feedback on structure: Contact DeepMind

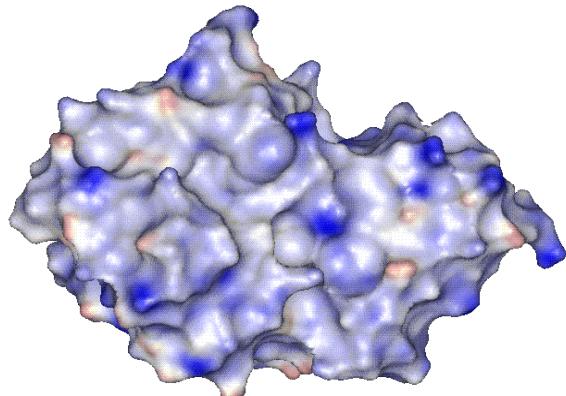
200 million !!!

What matters is to understand the biology behind the structures

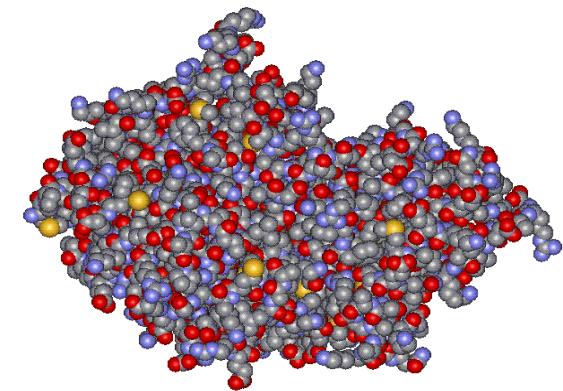
The problem we face is complexity!



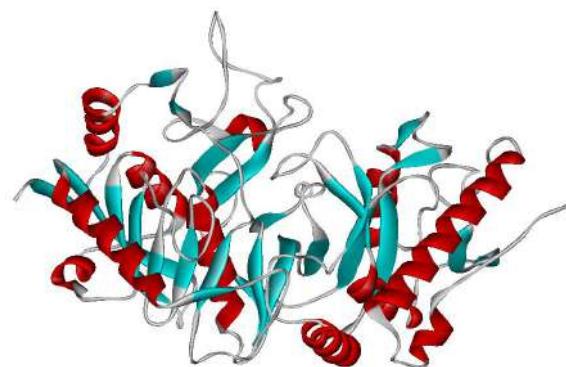
How to represent a structure?



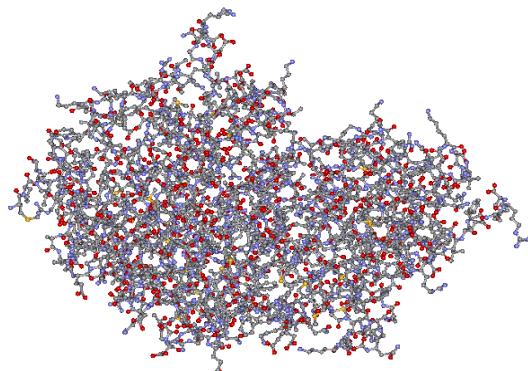
Accessible surface



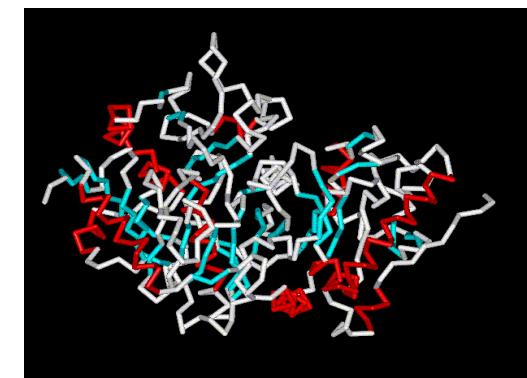
Space filling



Ribbon Cartoon



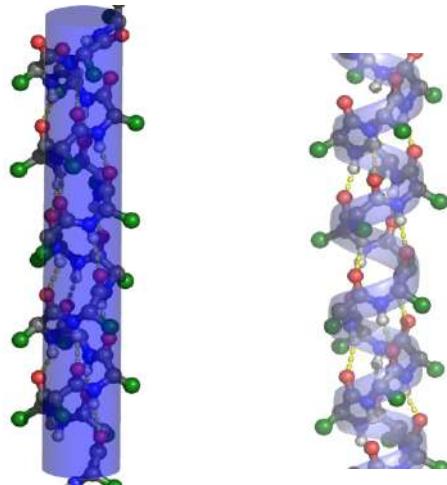
Ball and Stick



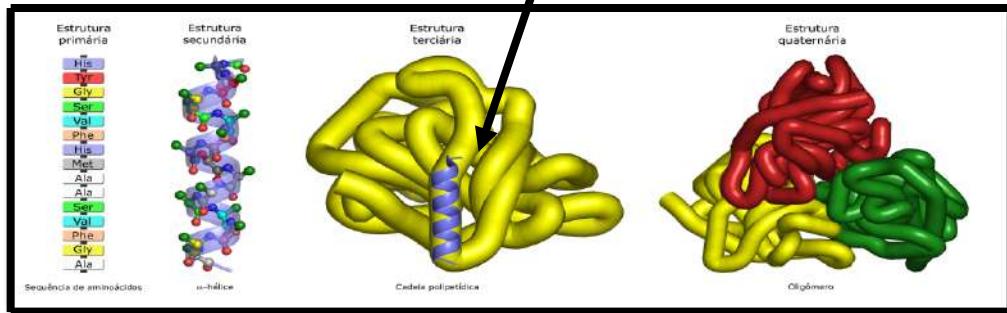
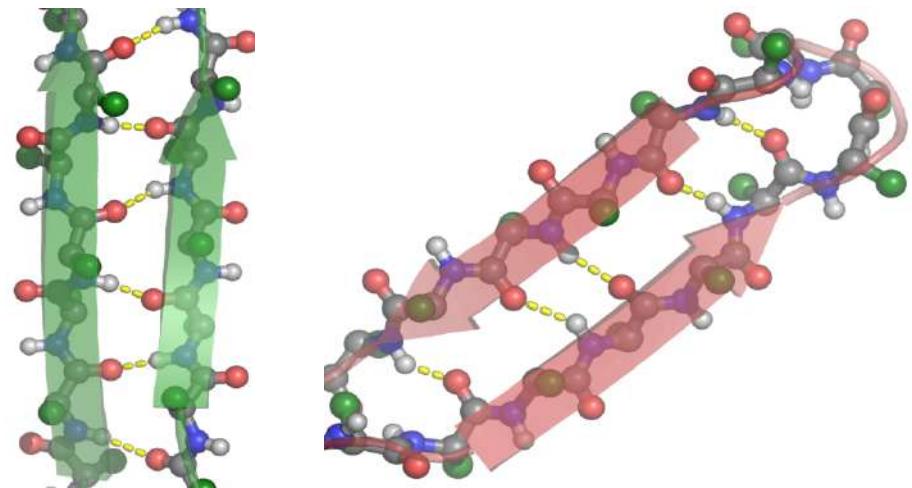
C α Trace

Secondary Structures

α -Helices



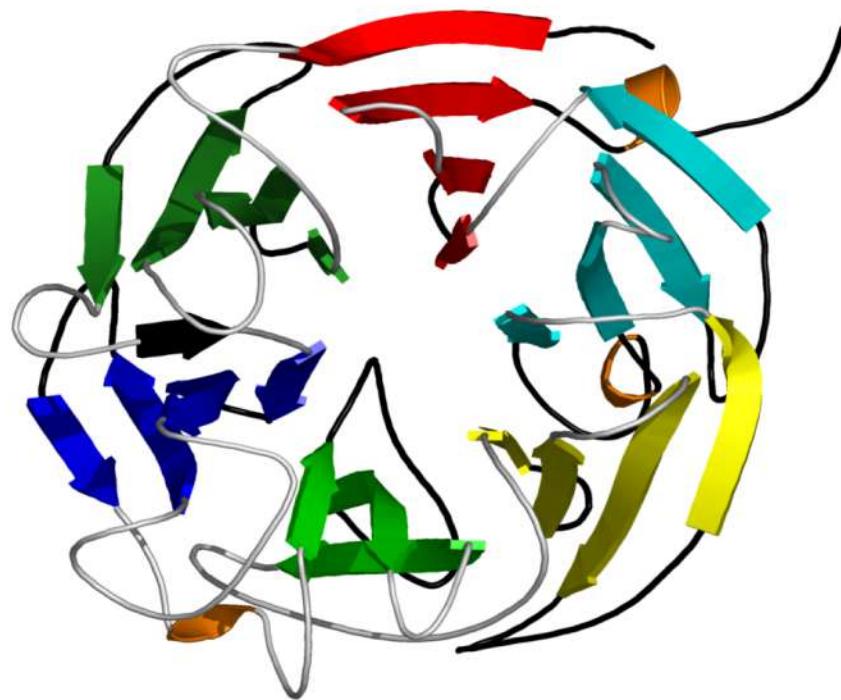
β -sheets



Protein Hierarchy

Cartoon Representation

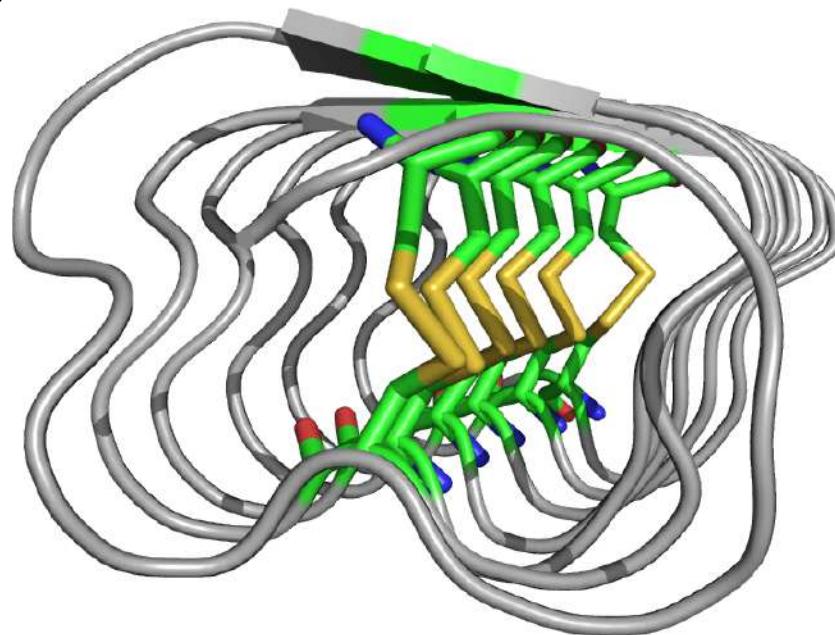
Six-bladed β -propeller



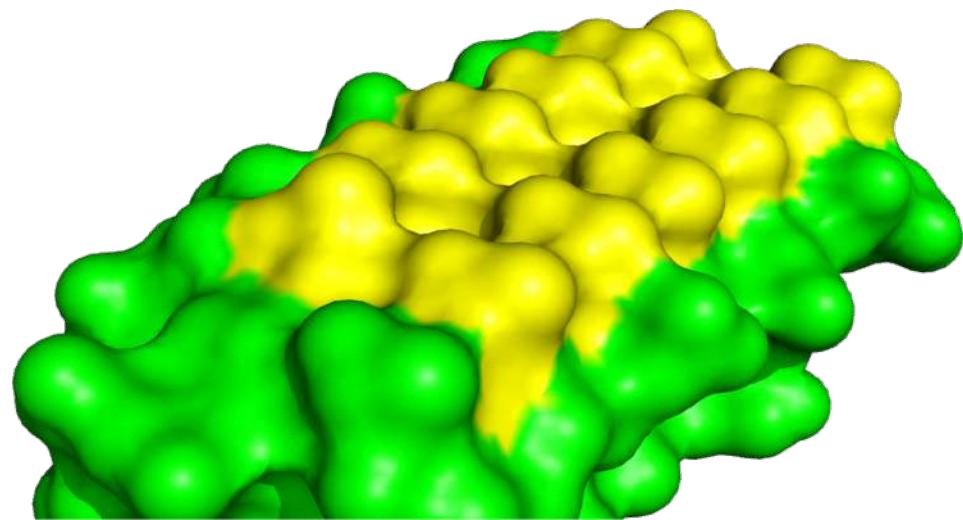
Structure-Function : AFP de Tenebrio molitor

Anti-freeze protein (AFP)

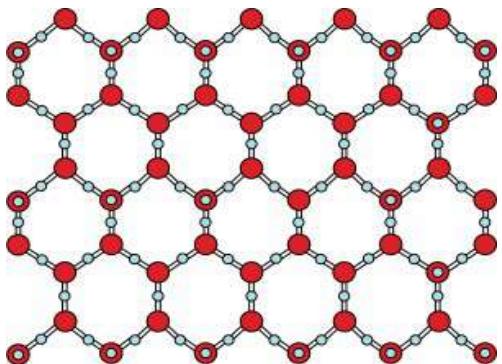
TCTxSxxCxxAx



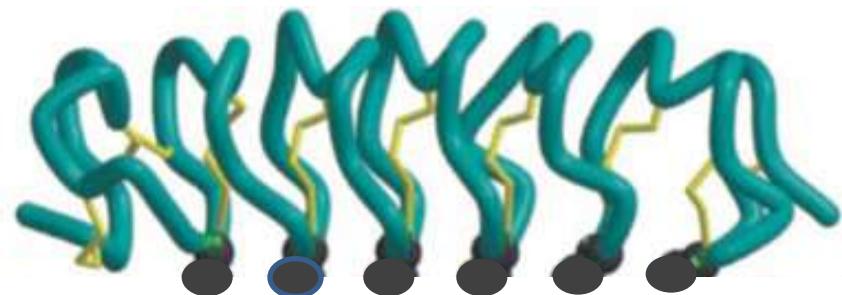
How does the protein aid in avoiding ice formation?



The hydroxyls are arranged such that they coincide with water molecules in an ice crystal.



The surface of AFP is covered by a regular array of hydroxyl groups coming from threonine sidechains



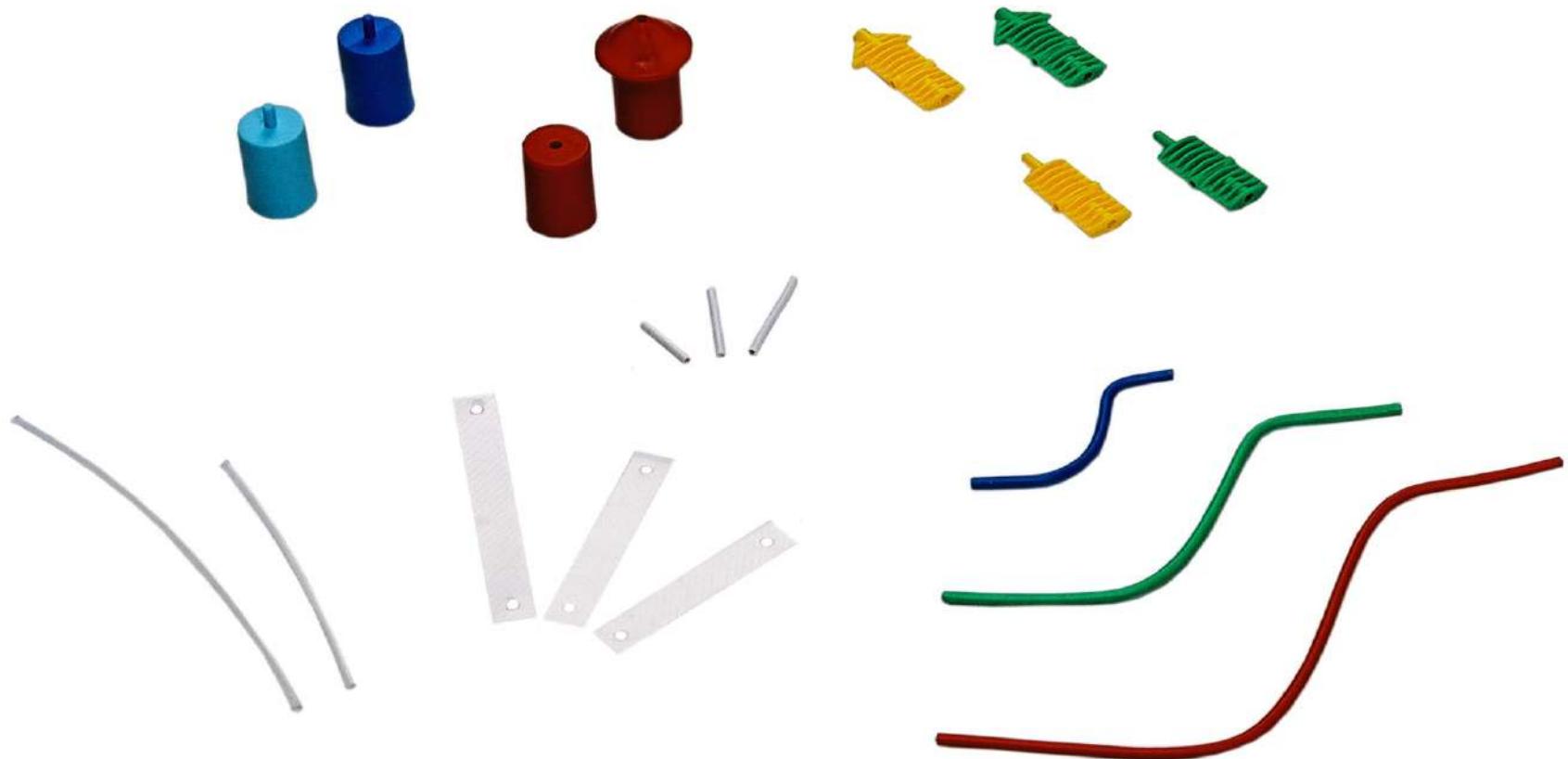
Protein Structure – Arid and boring?



Going to try to change this!!!!!!



Cartoon models

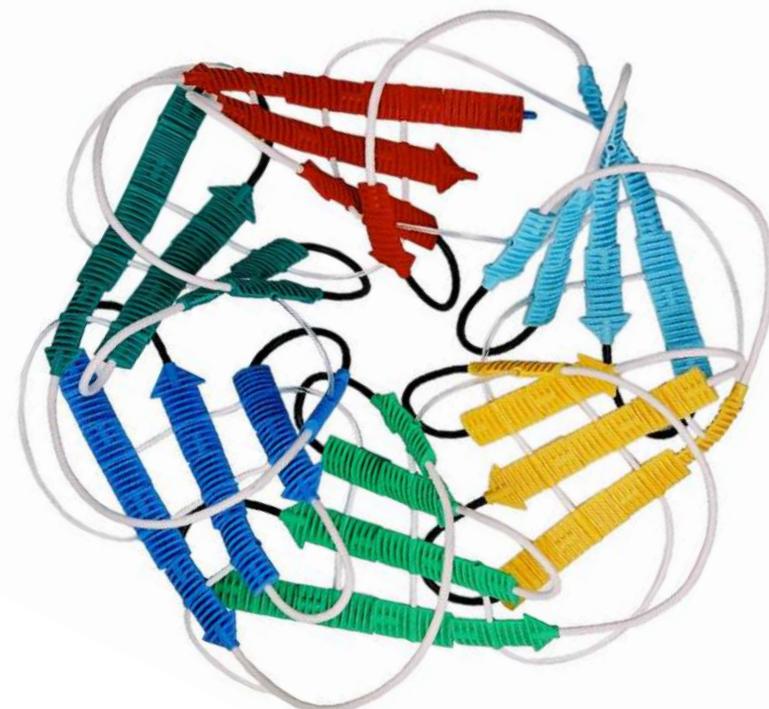
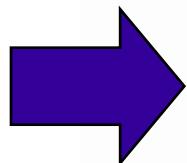


Transforming an image into a physical object

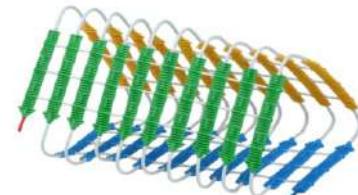
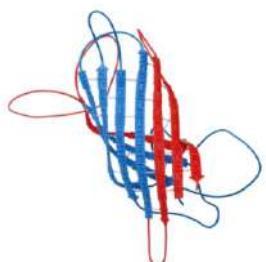
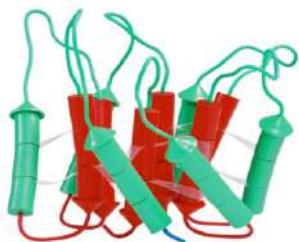
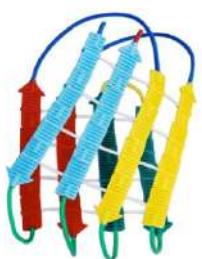
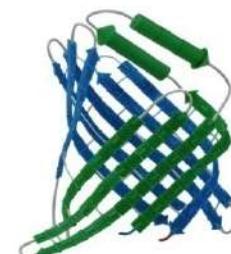
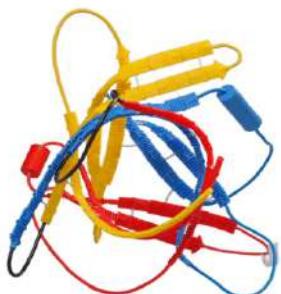
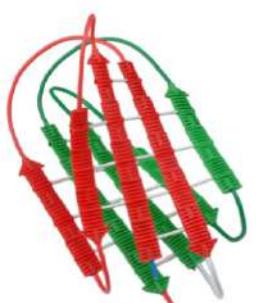
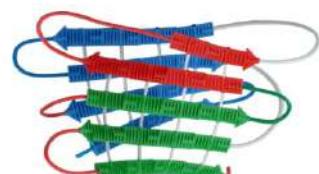
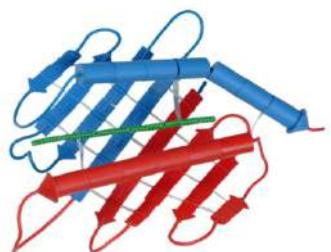
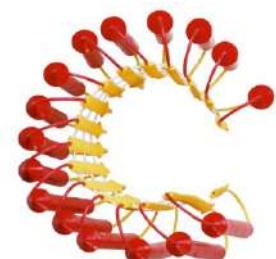
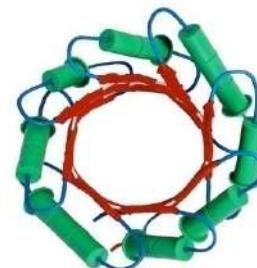
Six-bladed β -propeller



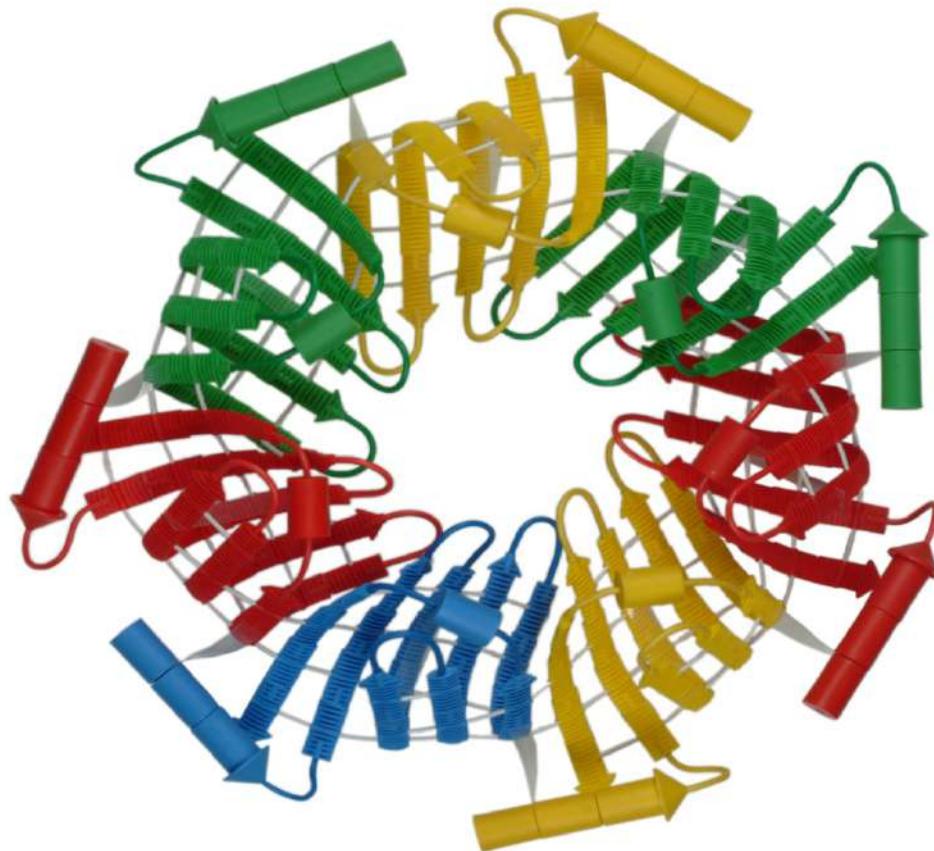
Ribbon diagram



Physical Model



Our final objective



Richard C. Garratt
Christine A. Orengo

The Protein Chart

The "periodic table" of proteins illustrates the beauty, diversity, and complexity of proteins in one place. It is a remarkable teaching and learning tool.

(Donald Voet, author of "Biochemistry")

The ideal tool to understand and to teach the principles of protein structure, and beautifully designed. I wholeheartedly recommend it.

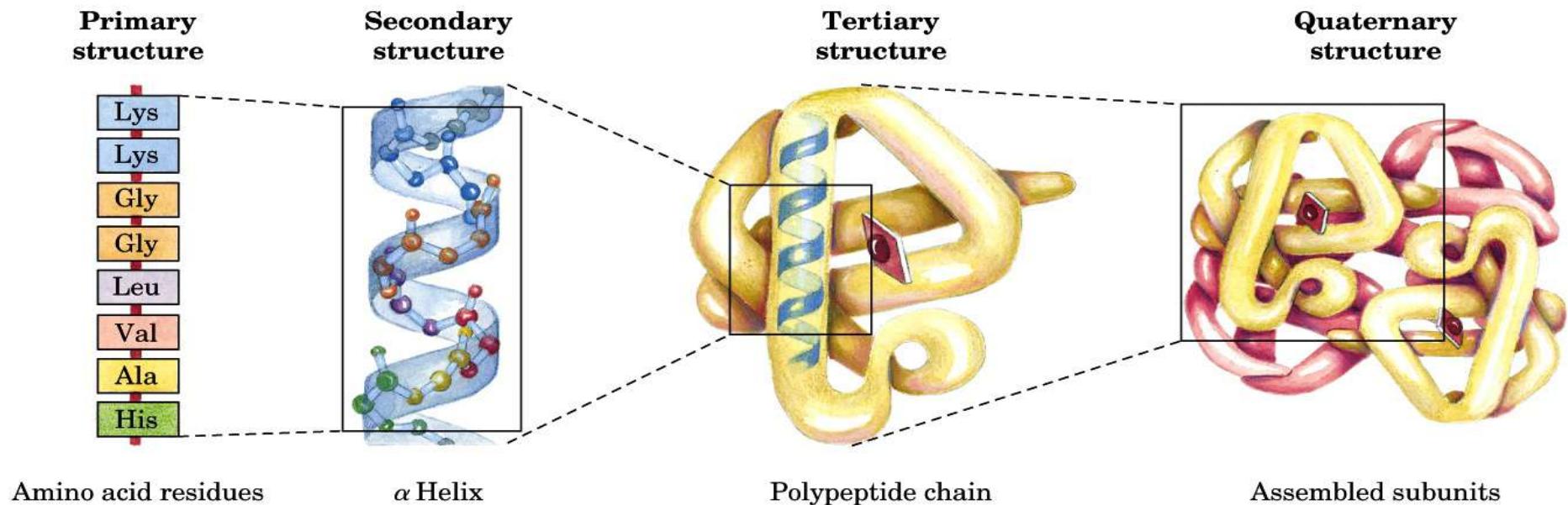
(Robert Huber, Nobel Prize in Chemistry 1988)

 WILEY-VCH

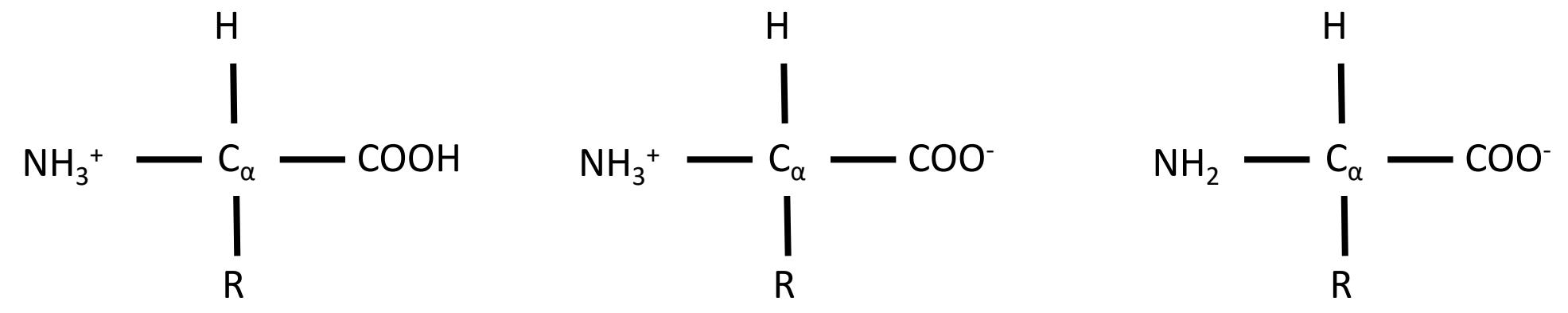
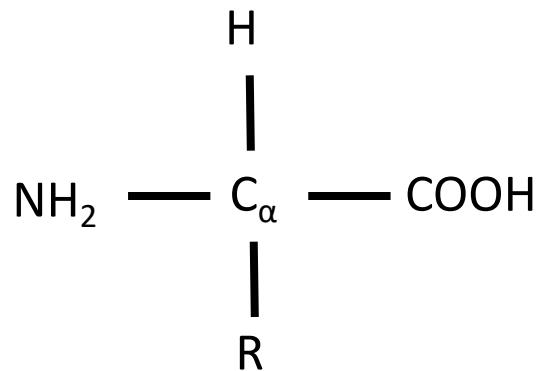
Part 2:

Some basic concepts

The levels of protein structure (a hierarchical description)



Amino acids

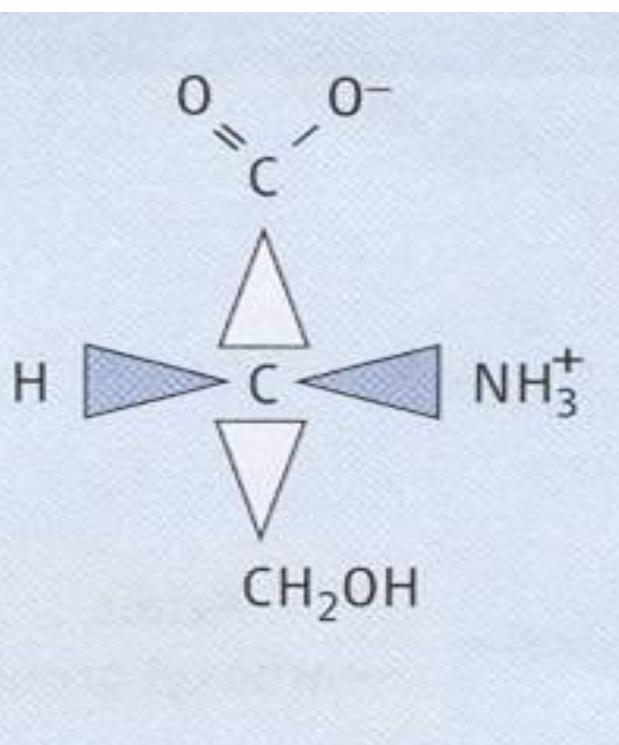


acidic pH

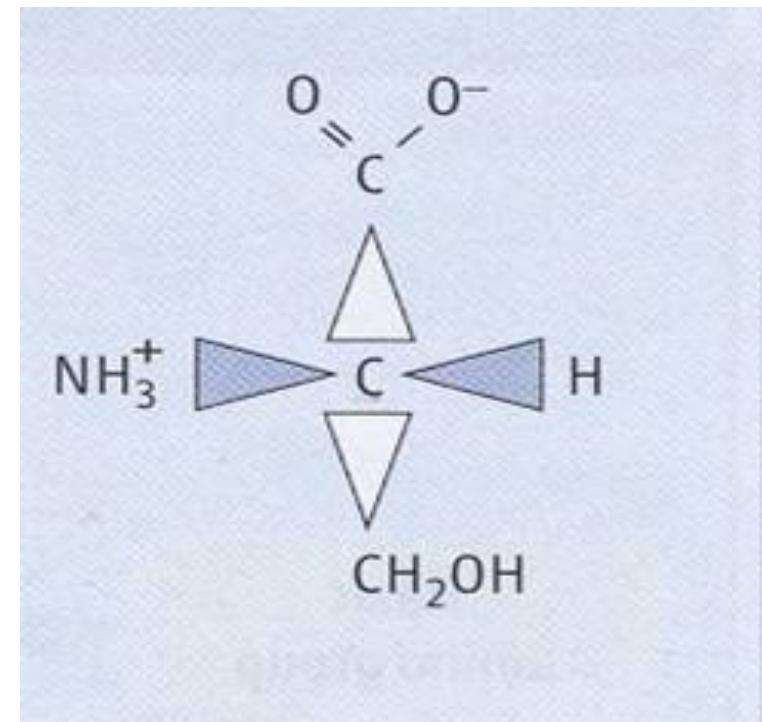
neutral pH

basic pH

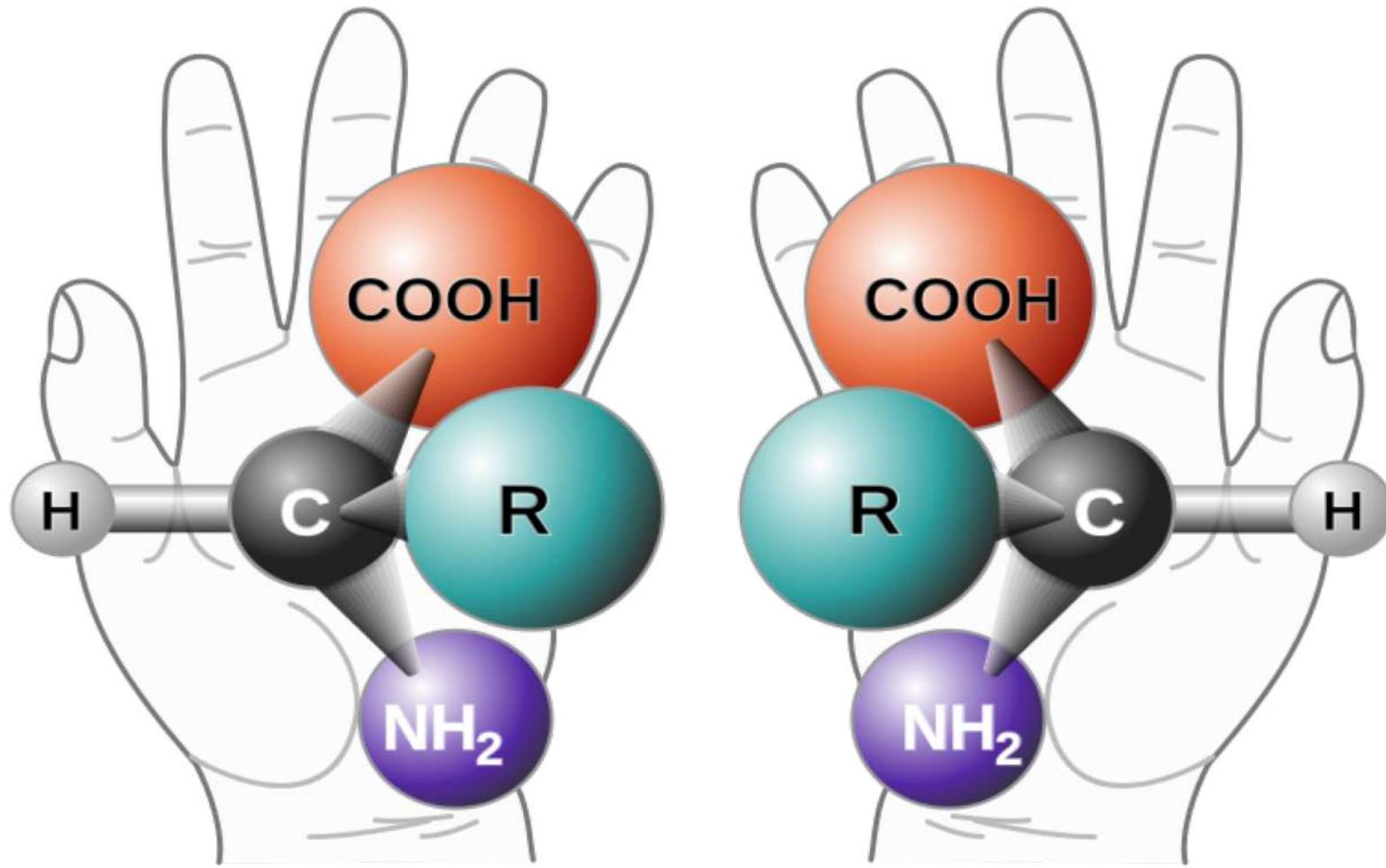
The structure of an α -amino acid



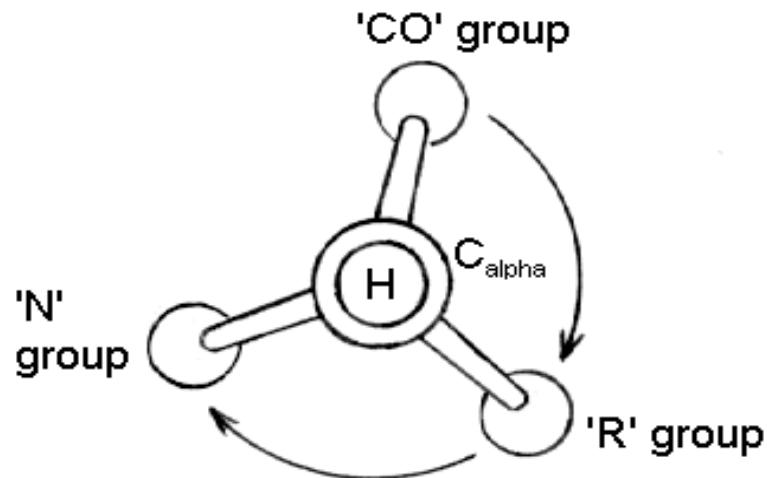
D



L

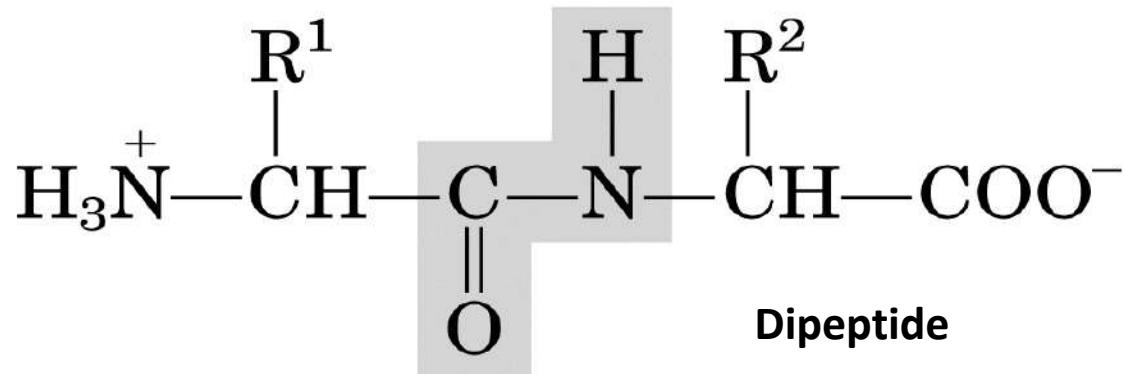
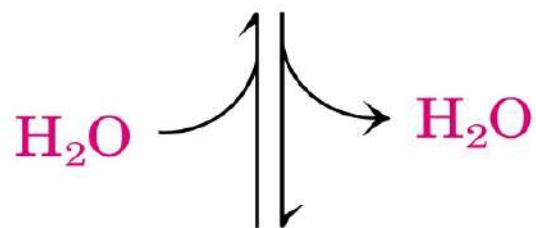
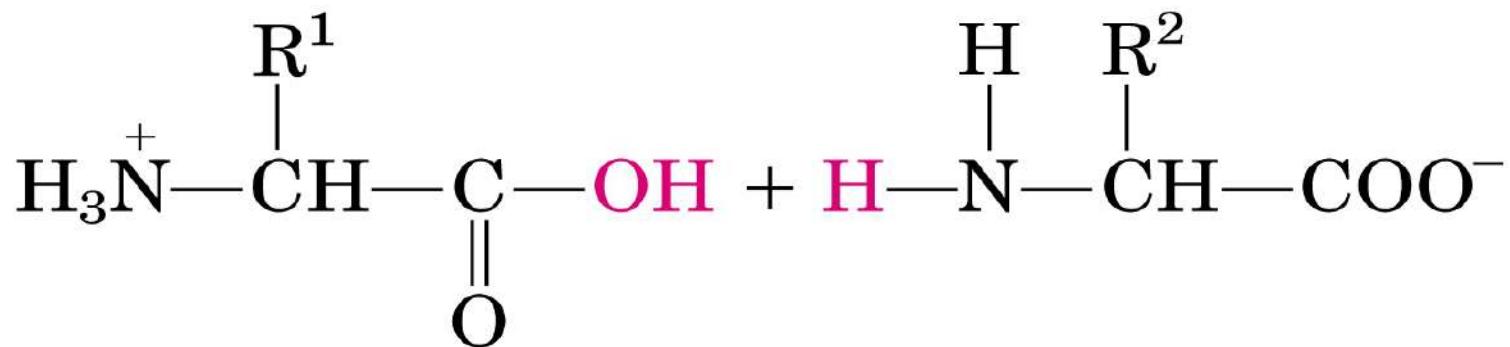


“The CORN Law” (C=O, R, N)

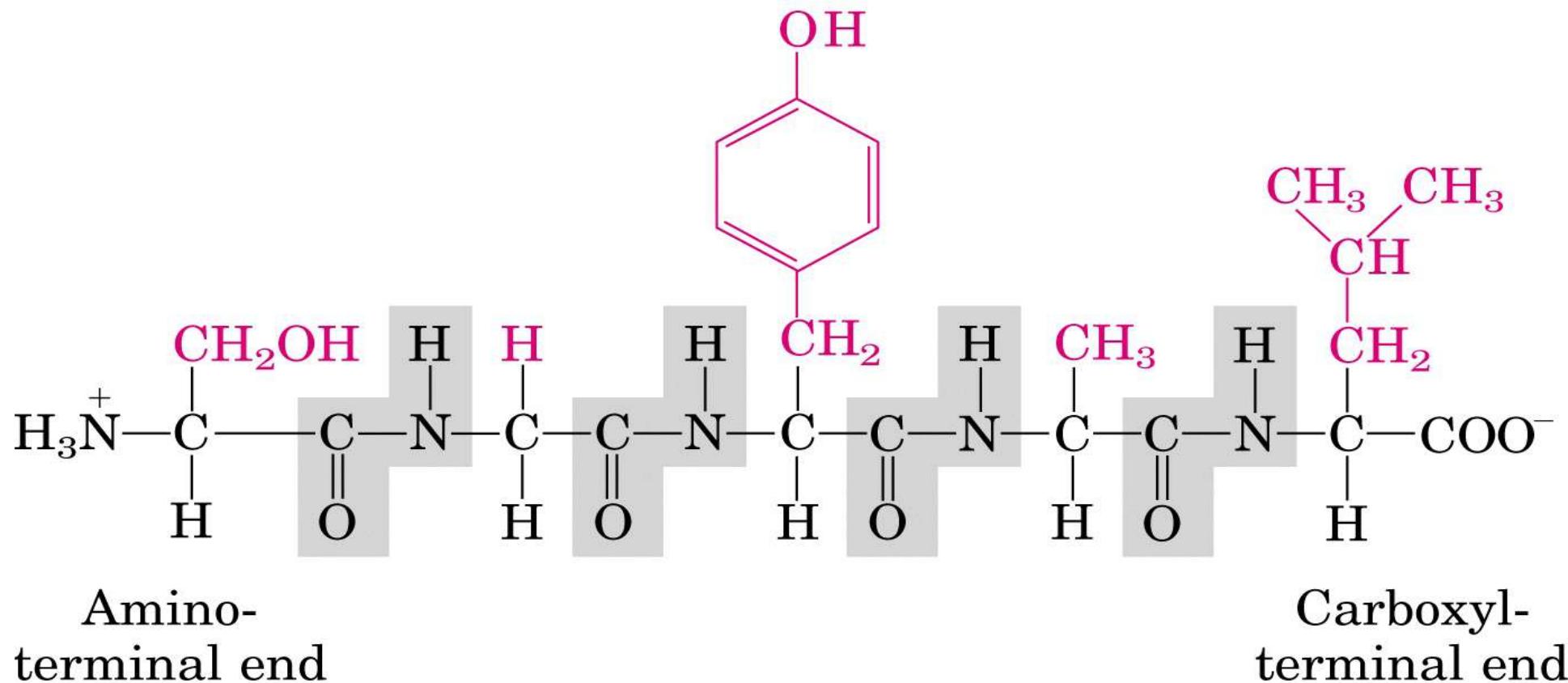


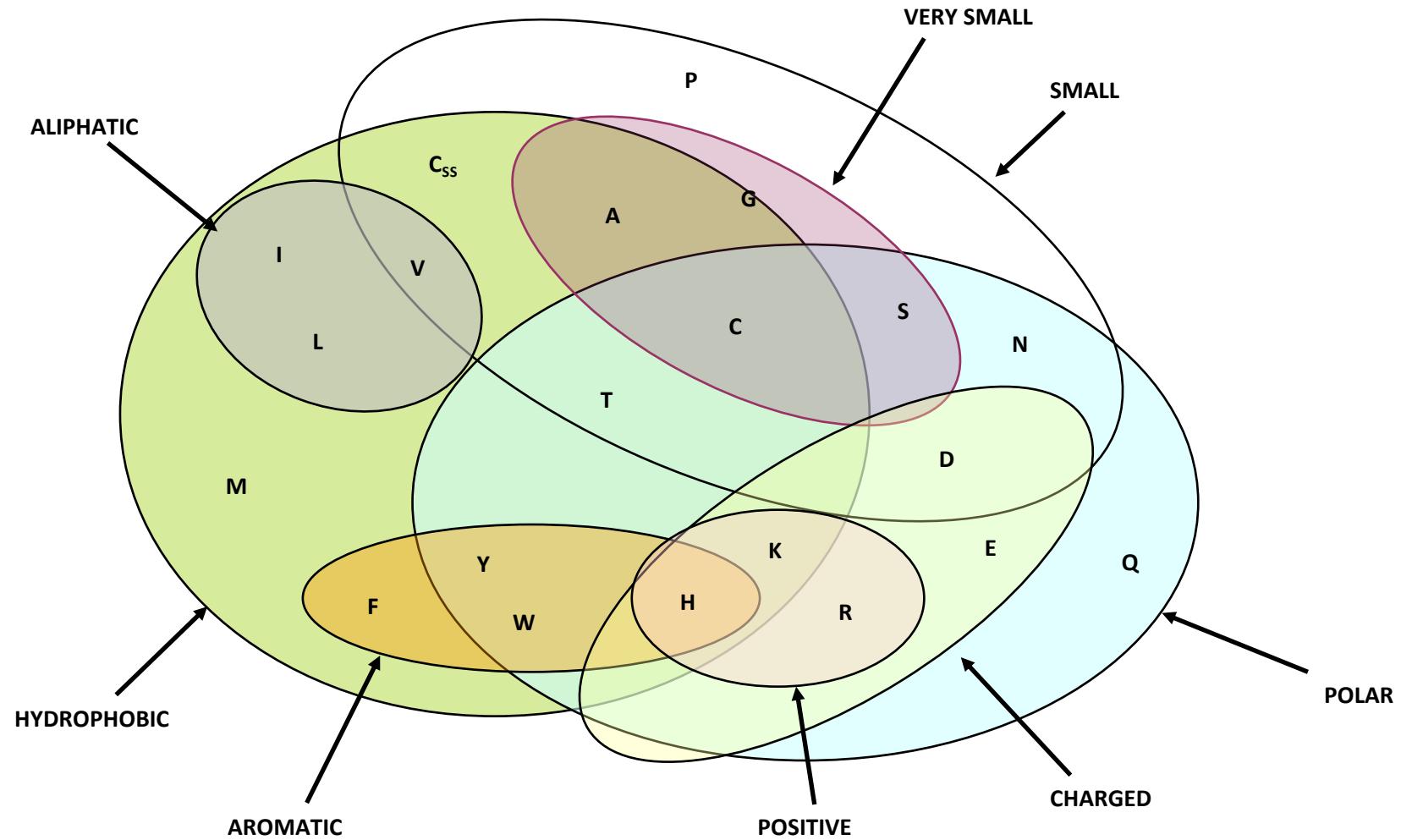
L- amino acid

The peptide bond



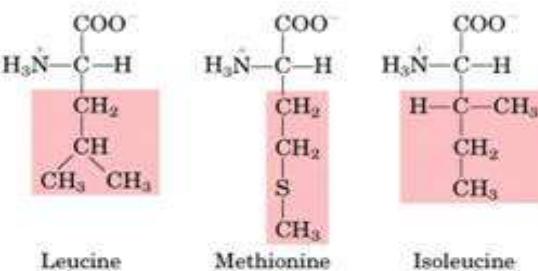
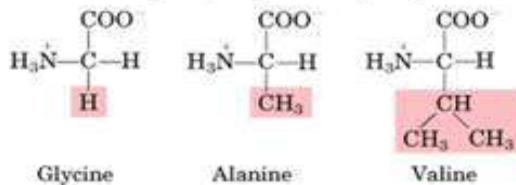
The polarity (direction) of the chain



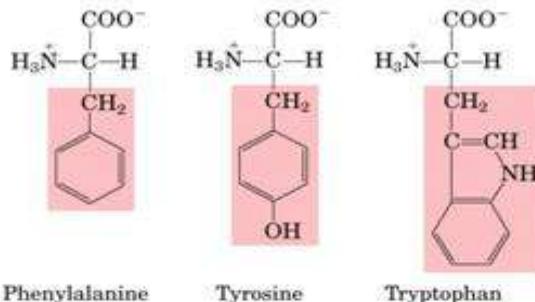


Twenty standard Amino Acids

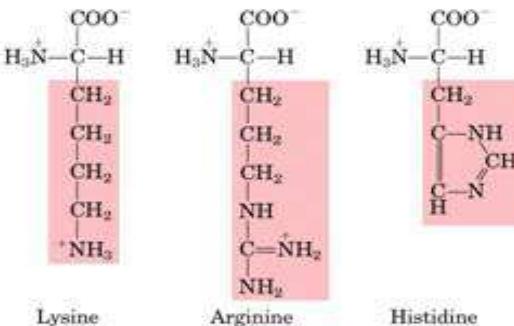
Nonpolar, aliphatic R groups



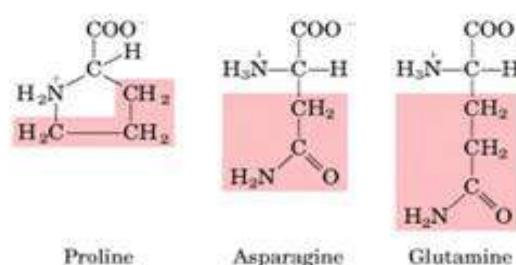
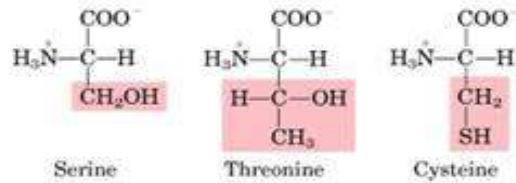
Aromatic R groups



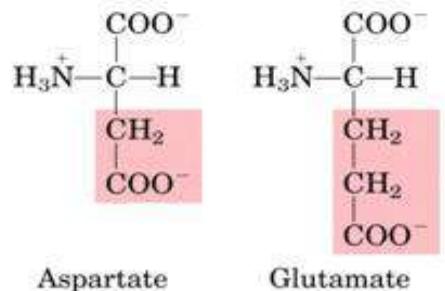
Positively charged R groups



Polar, uncharged R groups



Negatively charged R groups



of residues (n)

of possible sequences(20^n)

2

400

3

8000

4

160.000

5

3.200.000

10

1.024×10^{13}

50

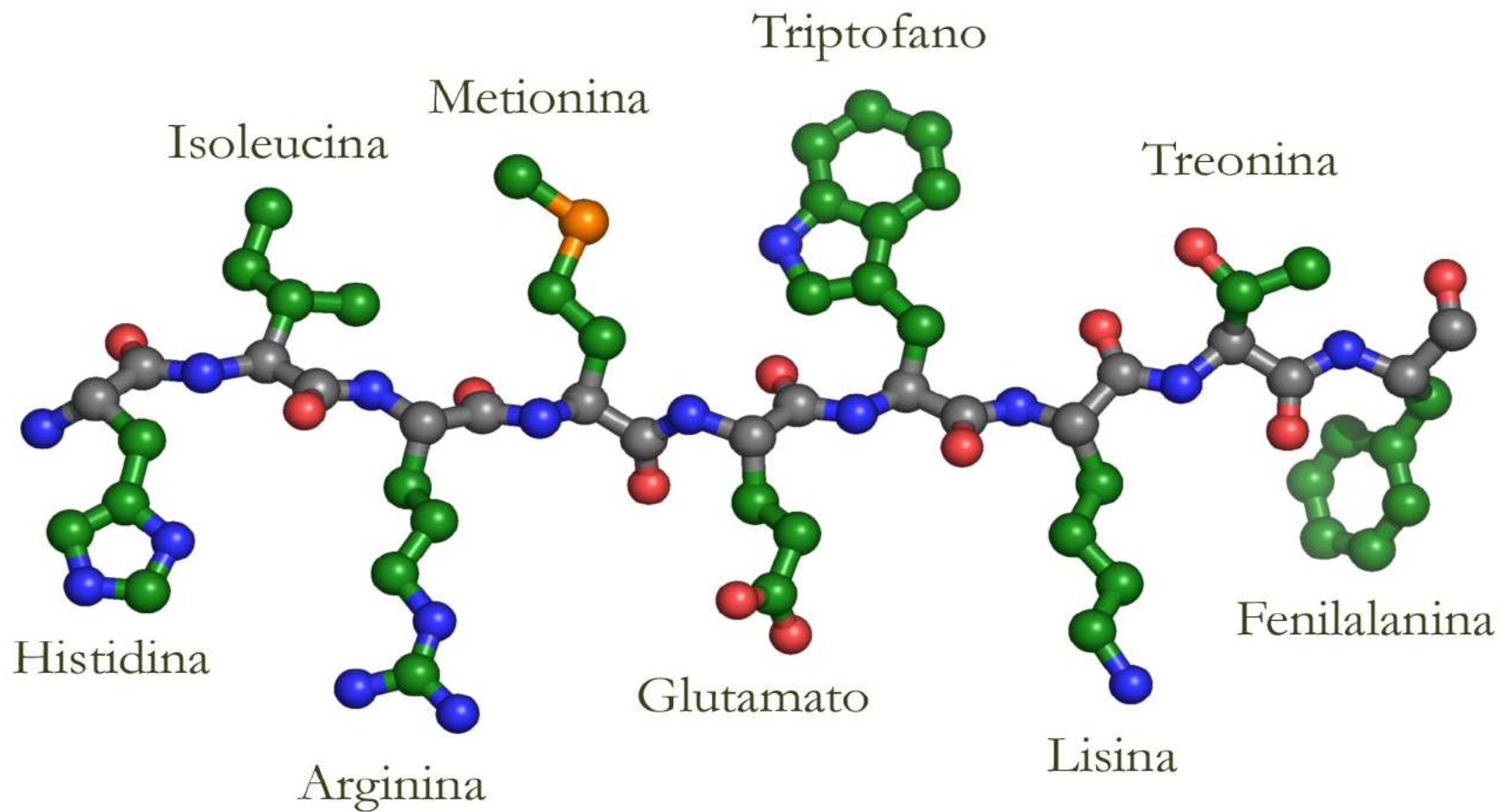
1.125×10^{65}

80

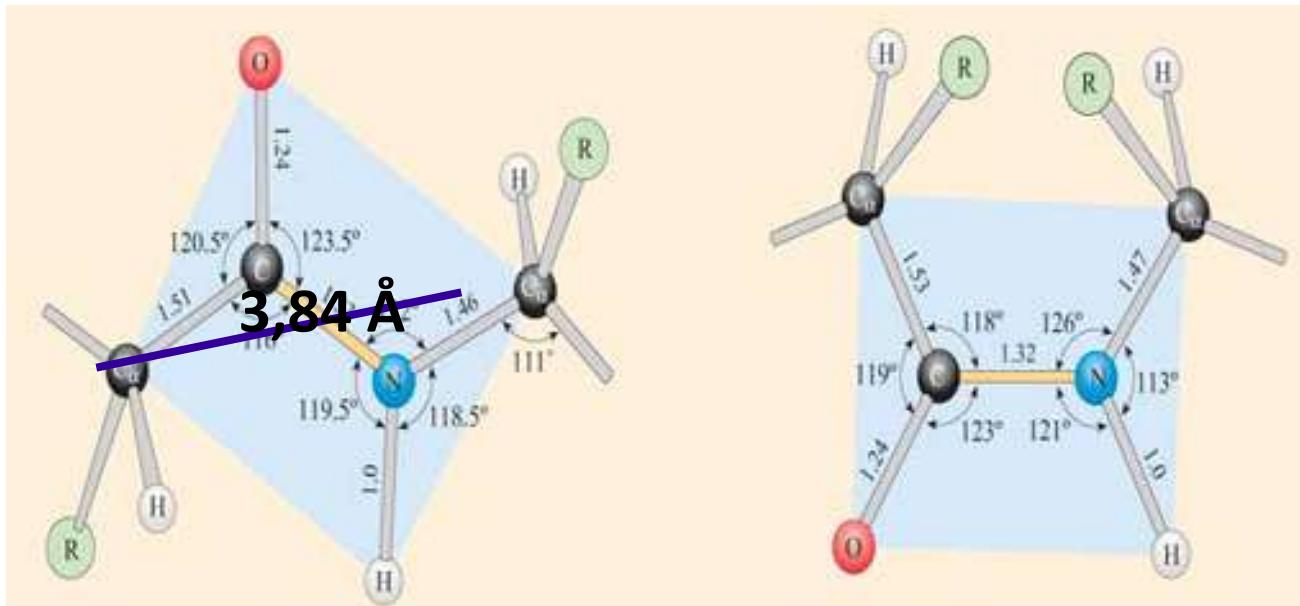
1.12×10^{78}

100

What is the conformational flexibility of the polypeptide chain?

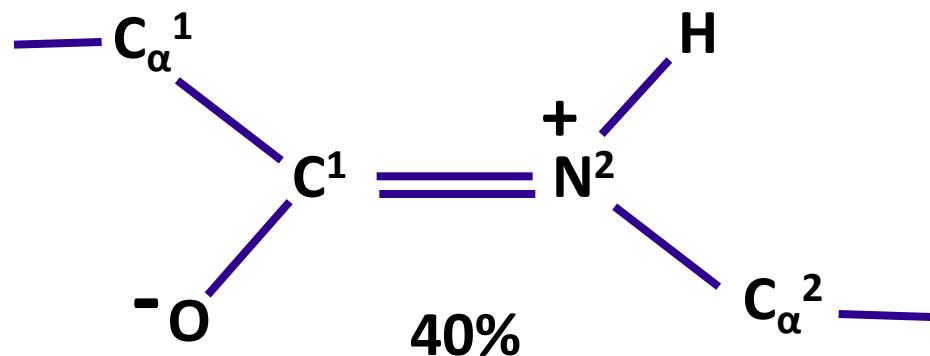
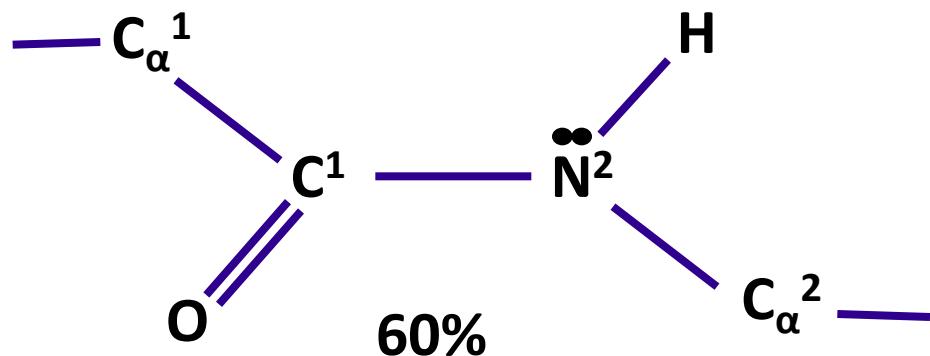


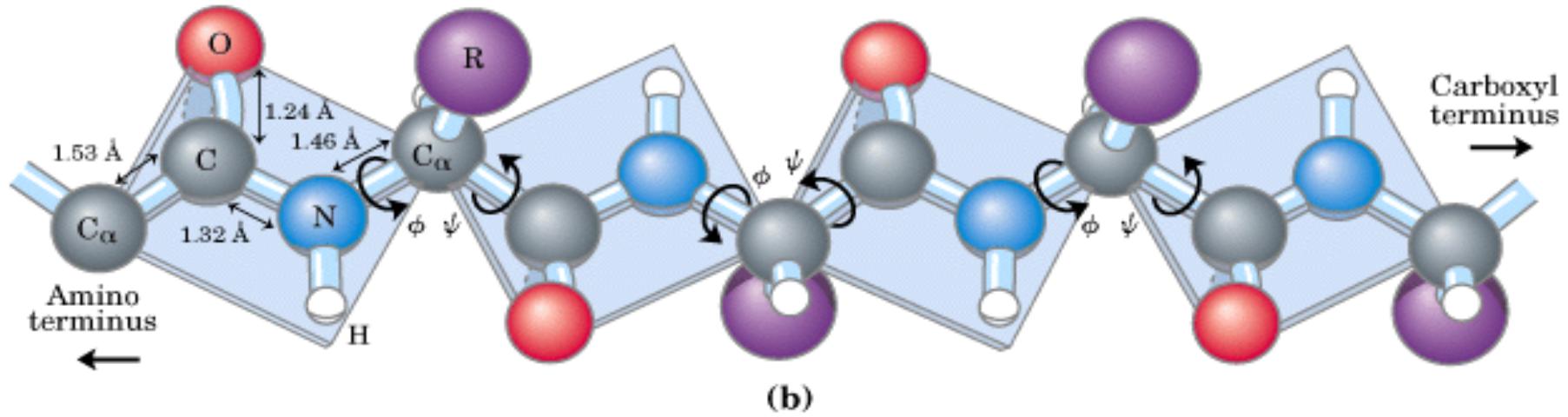
The peptide group is planar



Trans

Cis





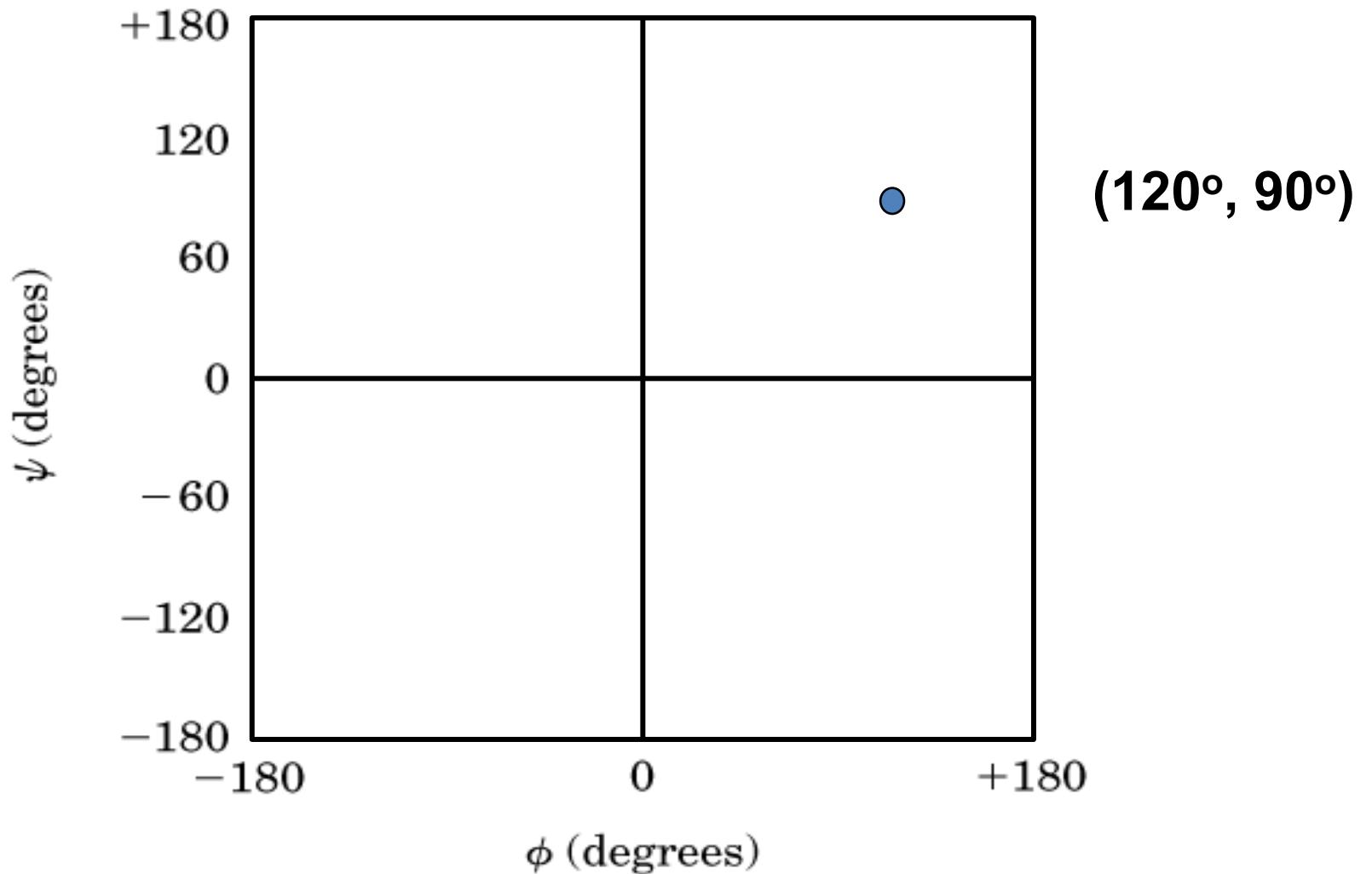
(b)

Only the ϕ and ψ angles are free to rotate

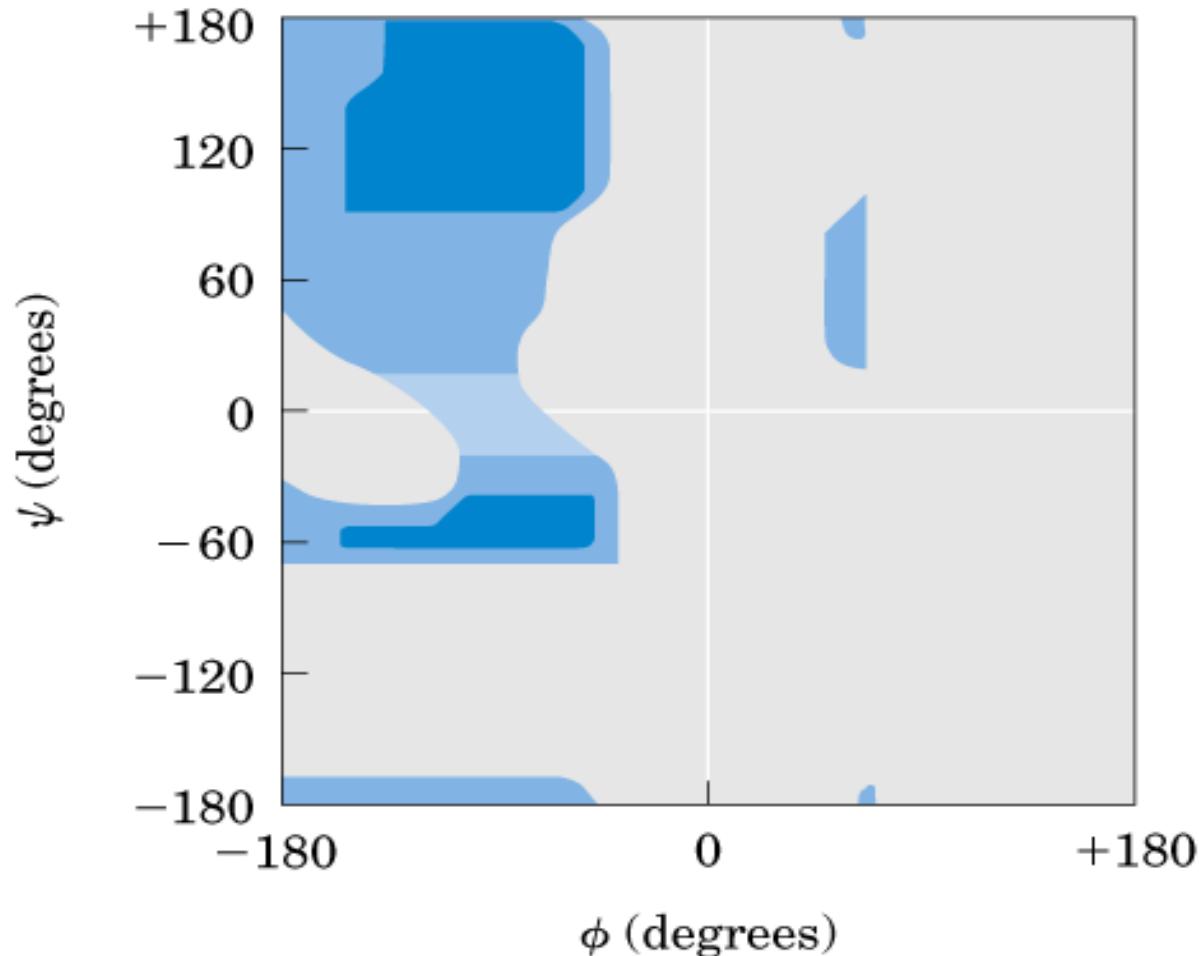
Where does the flexibility come from?

- 1) Bond lengths are fixed
- 2) Bond angles are fixed
- 3) Three torsion angles per residue are left (ω , ϕ e ψ) and one of them (ω) is also fixed.
- 4) All conformational variation comes from the combination of ϕ e ψ (which can vary between -180° and +180°).

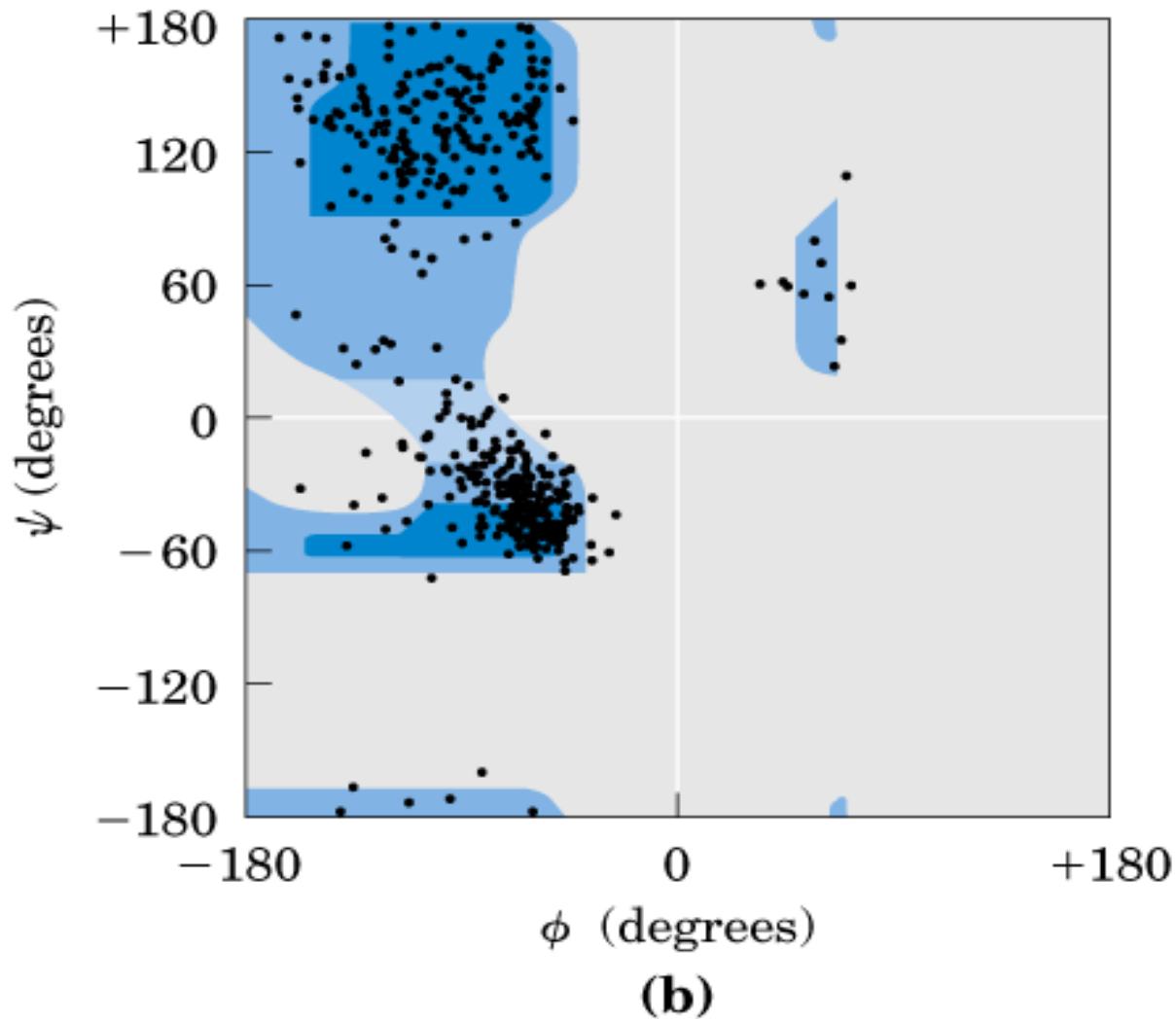
A system of cartesian coordinates

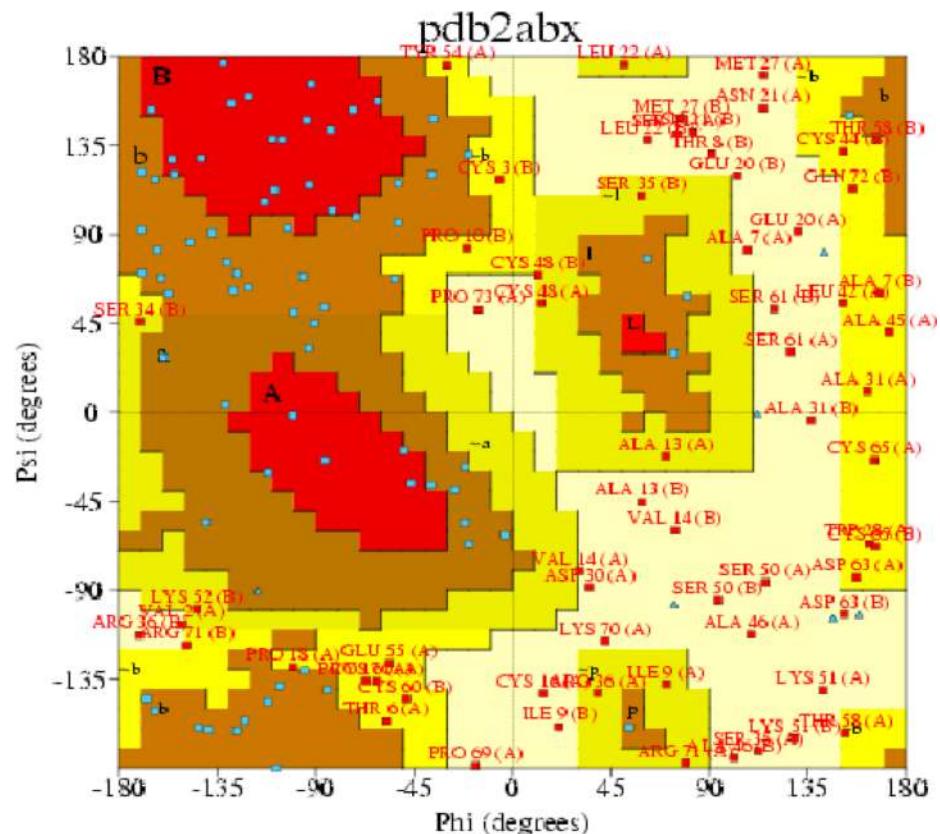
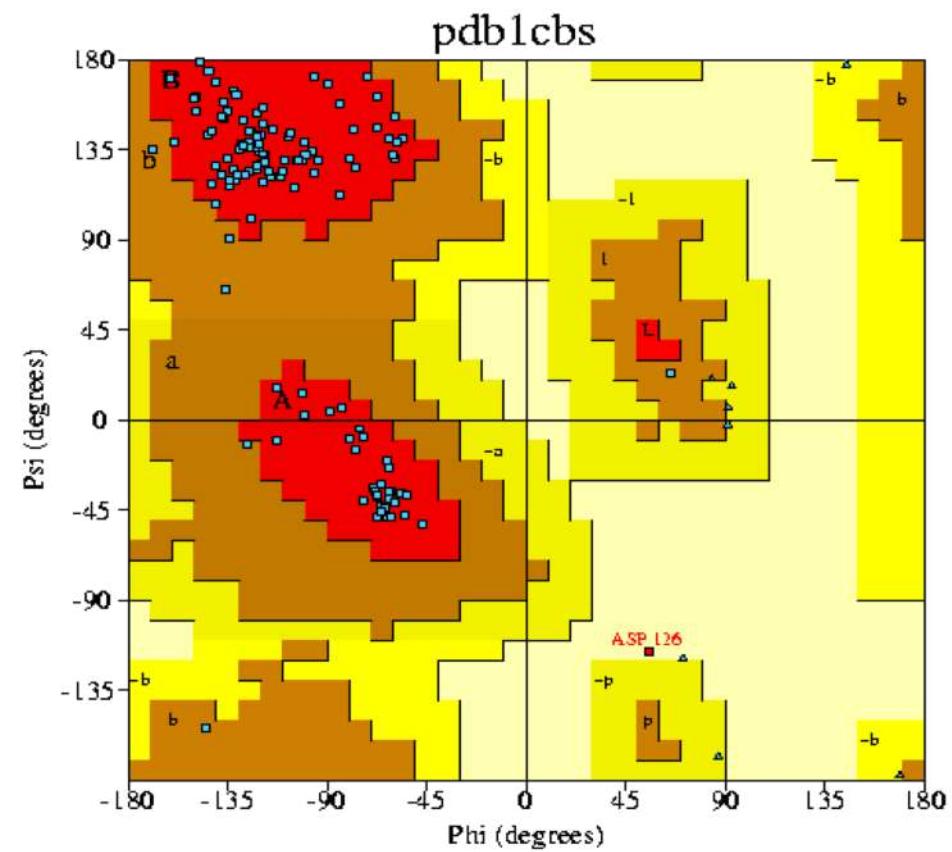
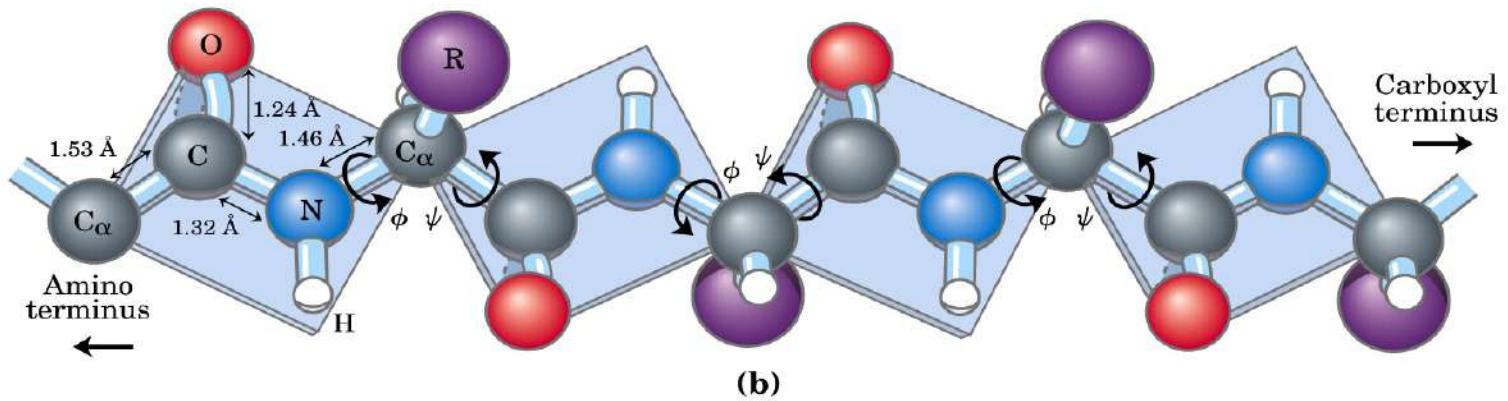


The RAMACHANDRAN plot – not all combinations of ϕ e ψ are possible

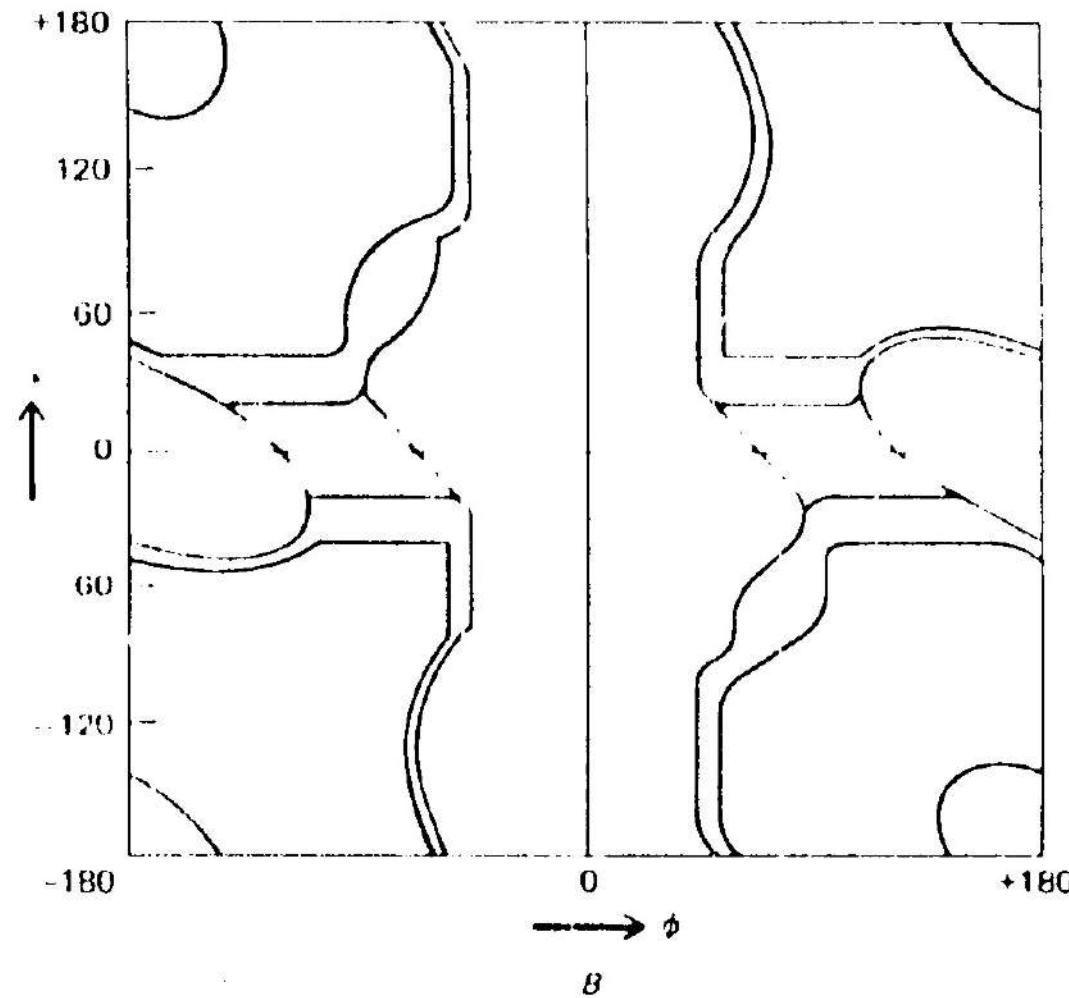


A well-determined structure (Rabbit Pyruvate Kinase)

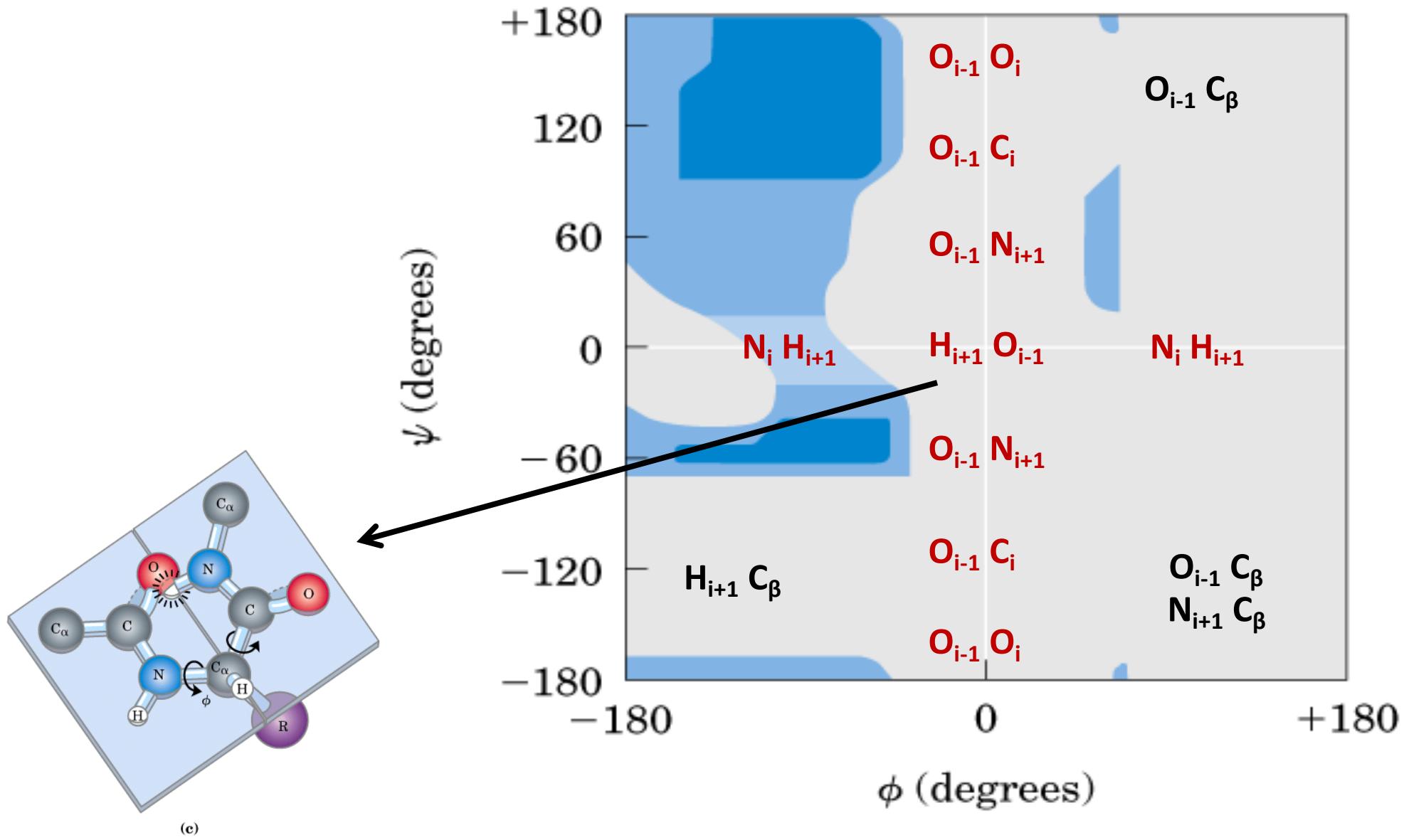




Glycine is different!



What are the steric impediments?



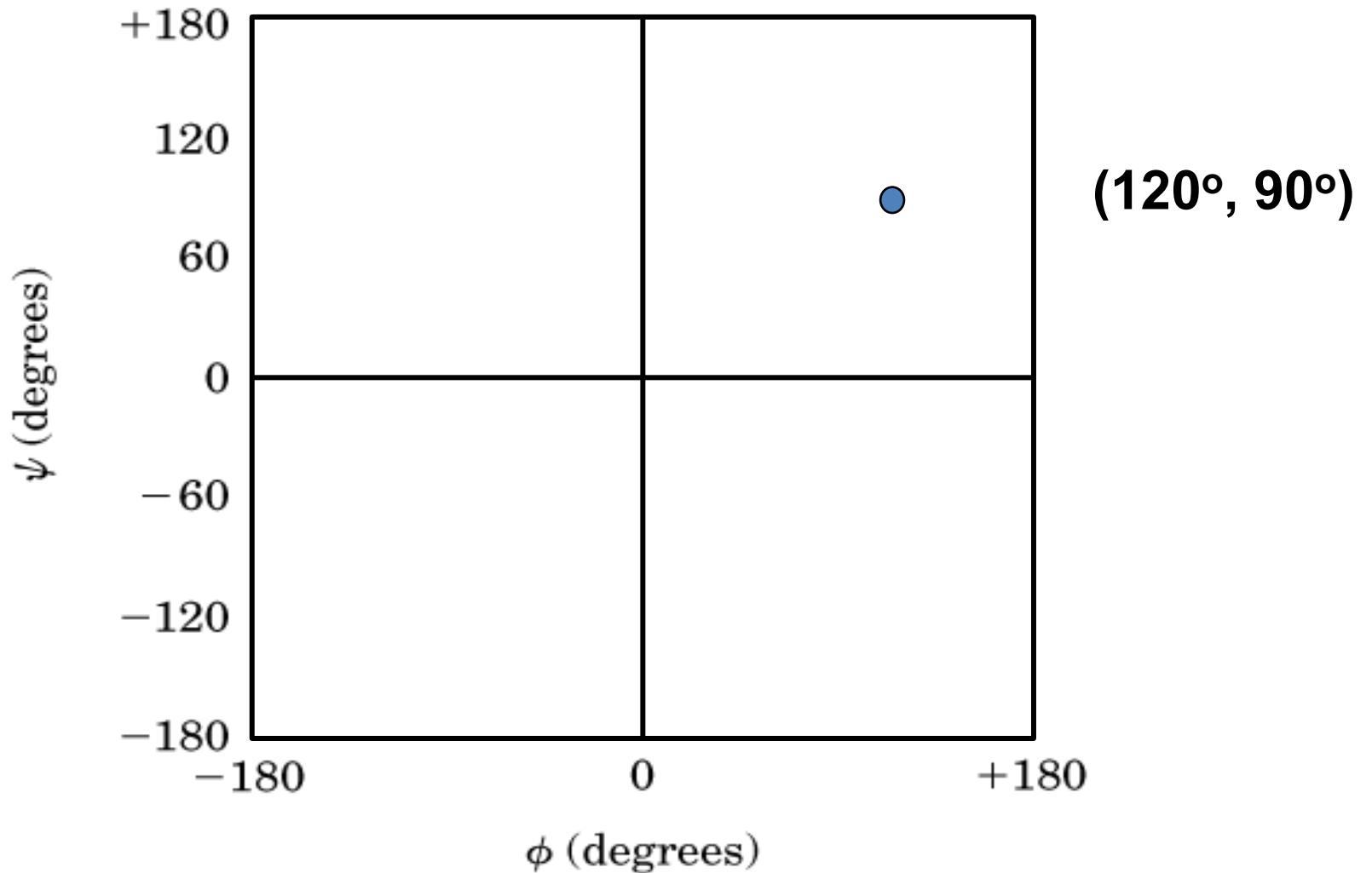
Part 3: Secondary Structures

What happens if we repeat Φ e Ψ
systematically?

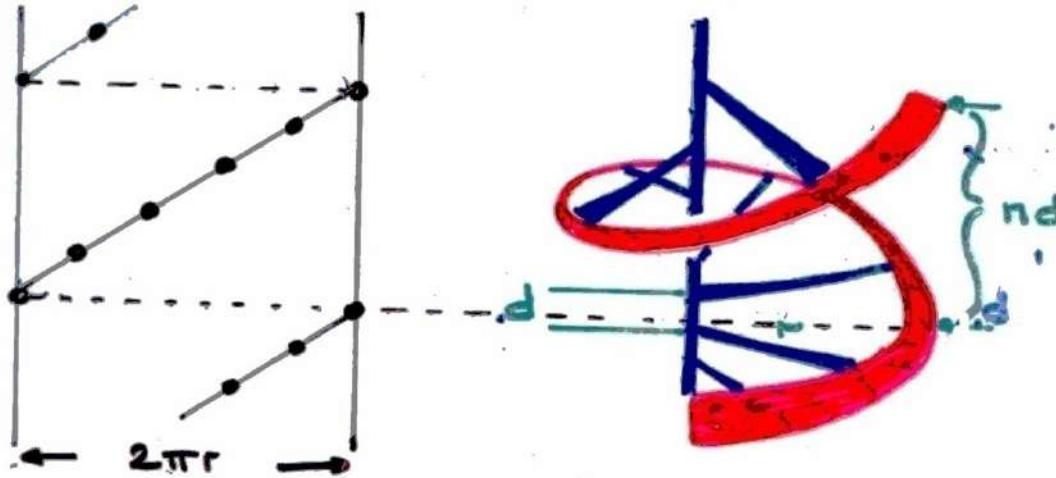
Randomly select values for Φ e Ψ

Do It!

A system of cartesian coordinates



Hélices em geral



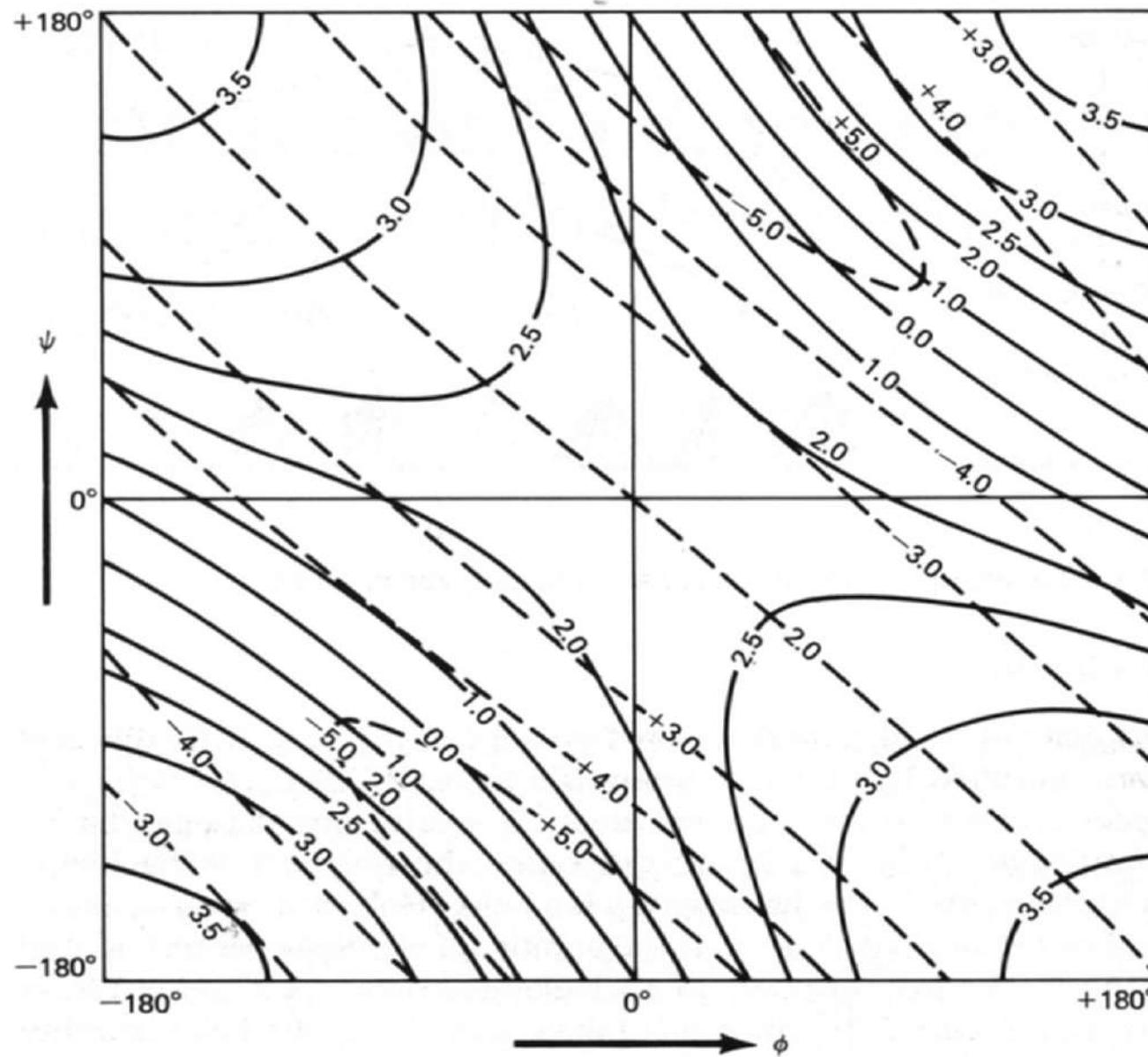
When ϕ and ψ are repeated the result is ALWAYS a helix.

n = number of residues per turn

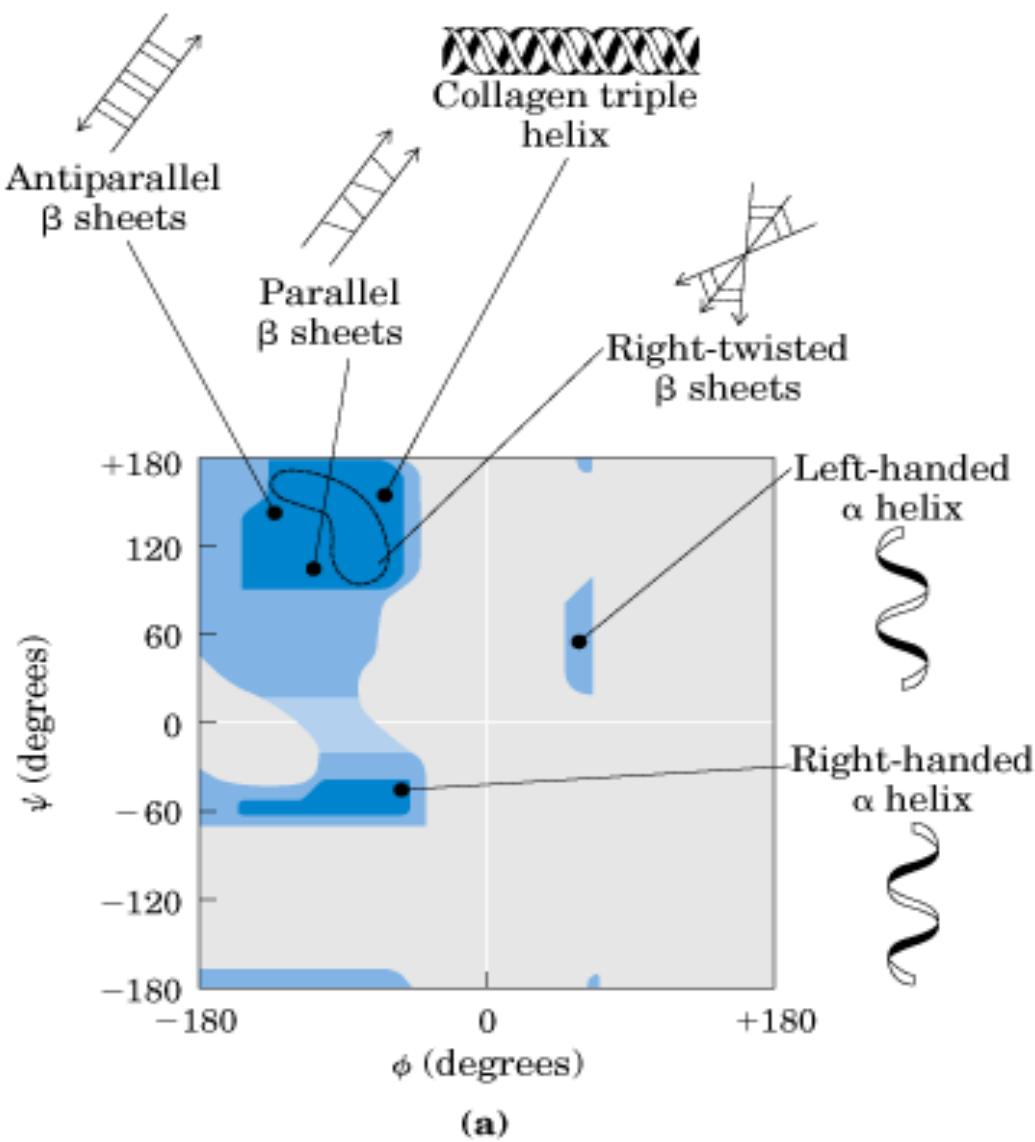
d = axial shift per residue

r = radius of the helix

$P = n \times d$ (pitch- axial shift per turn)



d = full lines; n = dashed lines



Amino terminus

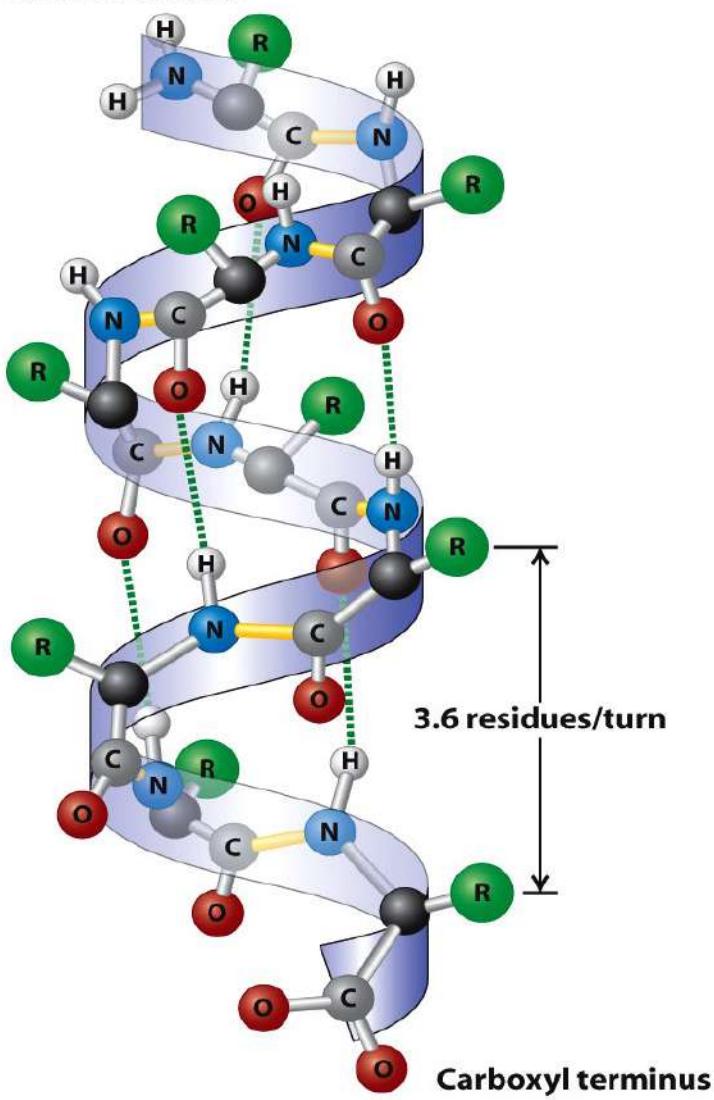
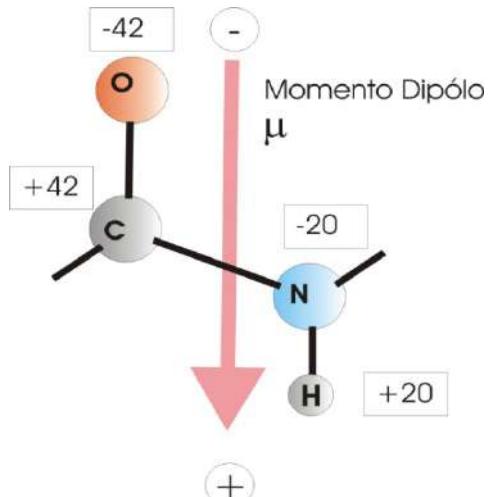


Figure 3-4

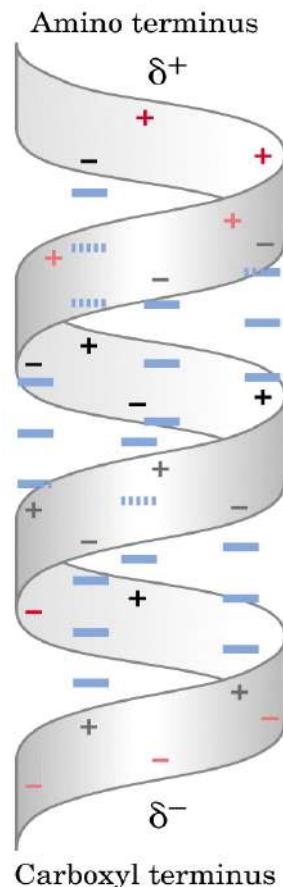
Molecular Cell Biology, Sixth Edition

© 2008 W.H. Freeman and Company

Electrical dipole



$$\begin{aligned}\mu &= 0.72 \text{ e}\text{\AA} \\ &= 3.46 \text{ D/res} \\ &= 1.155 \times 10^{-29} \text{ Cm}\end{aligned}$$

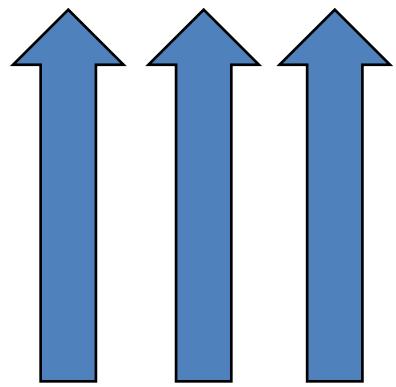


Typically equivalent to $1/2 \epsilon$ at Each end of the helix.

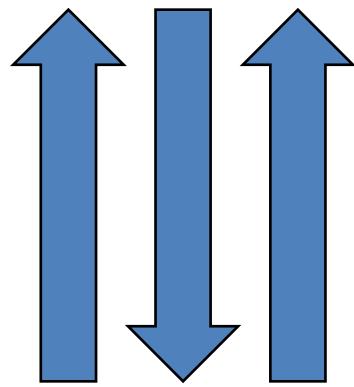
Biological Function

- Help in binding charged cofactors
- Long range attraction (K^+ and Cl^- channels)
- Change the nucleophilic properties of neighbouring residues for catalysis

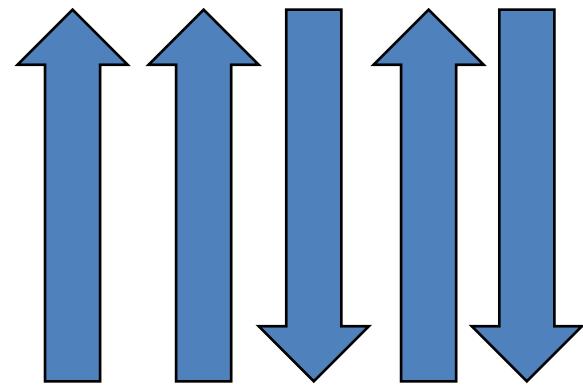
β -sheets



parallel

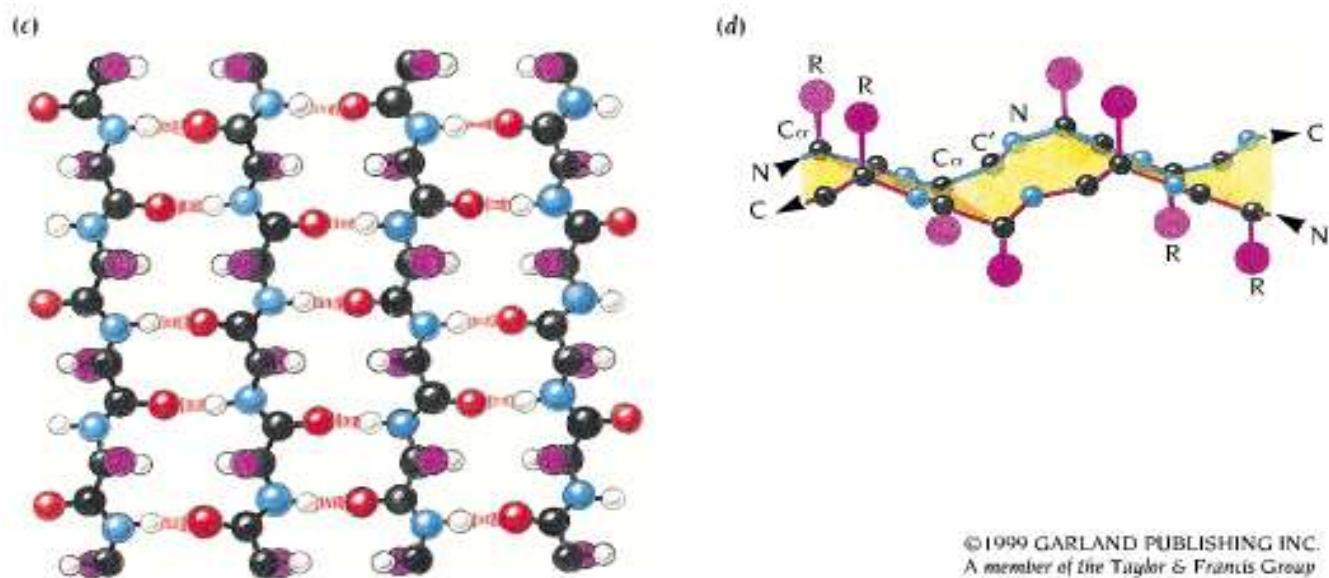


anti-parallel

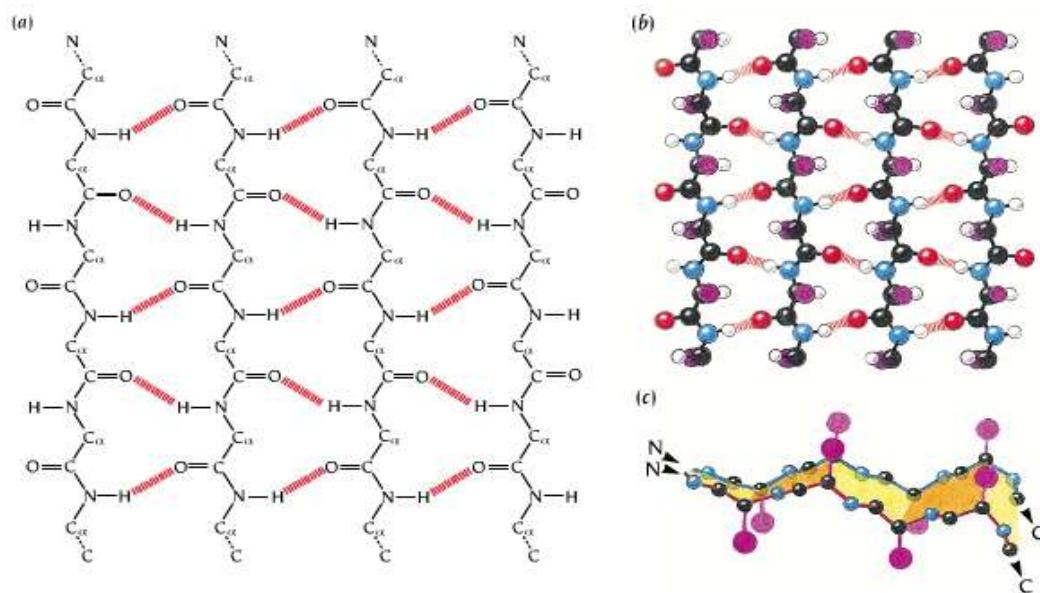


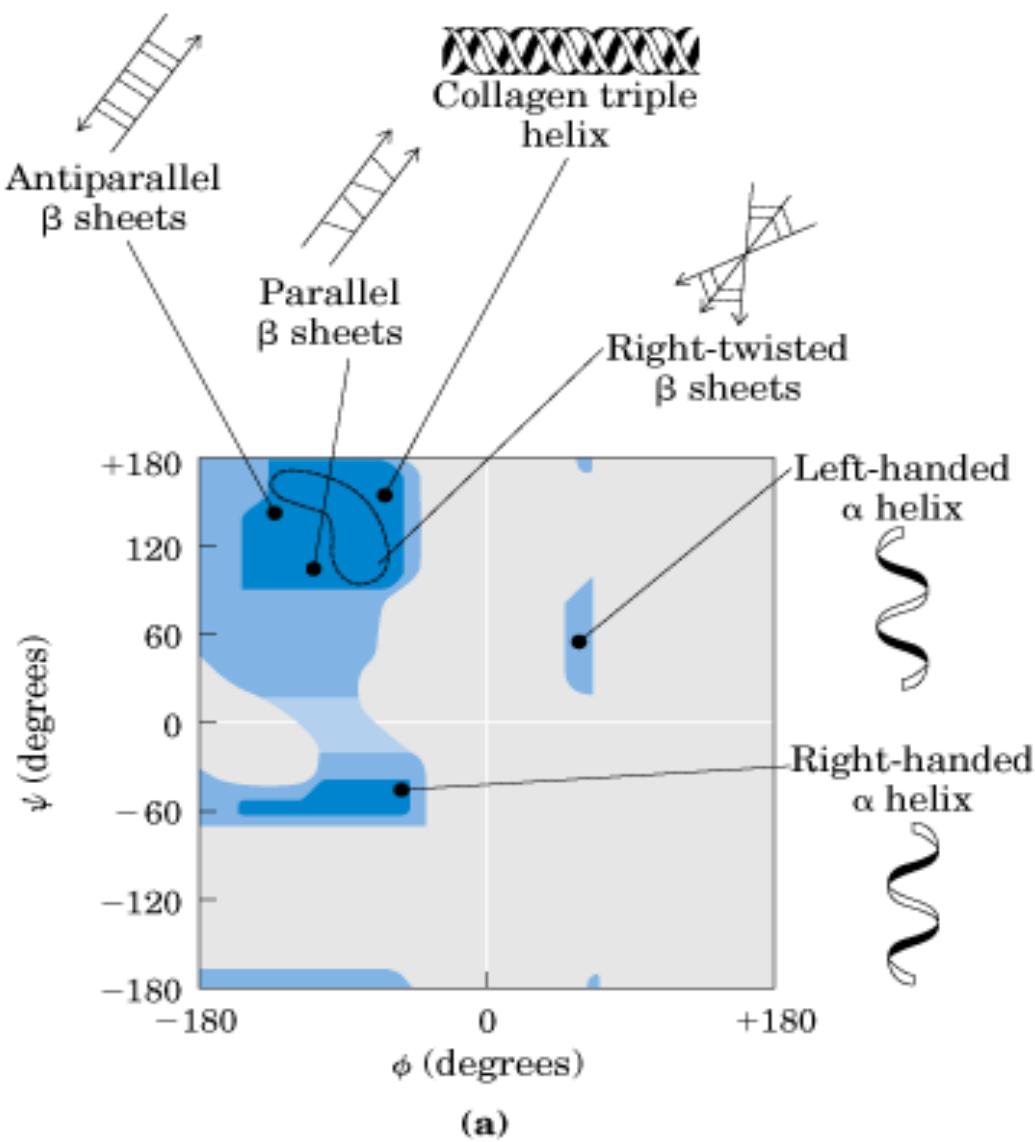
mixed

Anti-parallel
 β -sheet

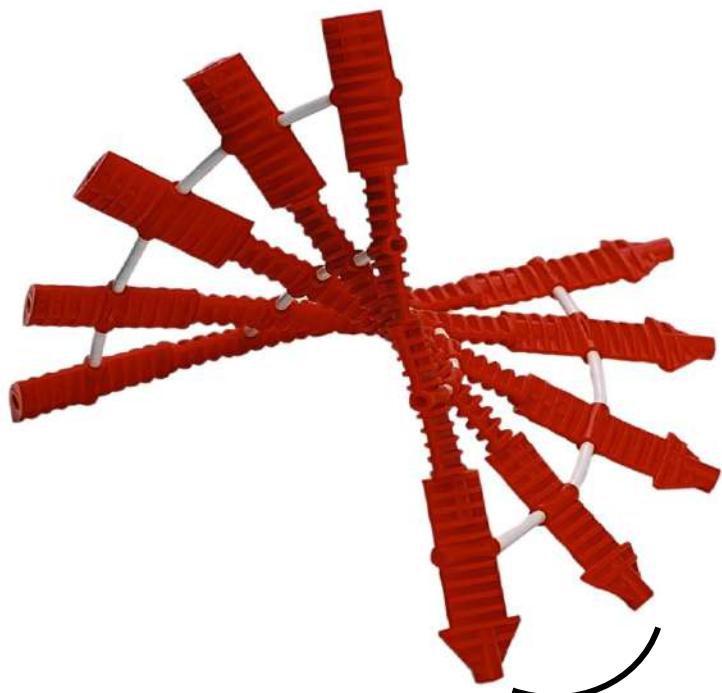


parallel
 β -sheet





The twist of the β -sheet



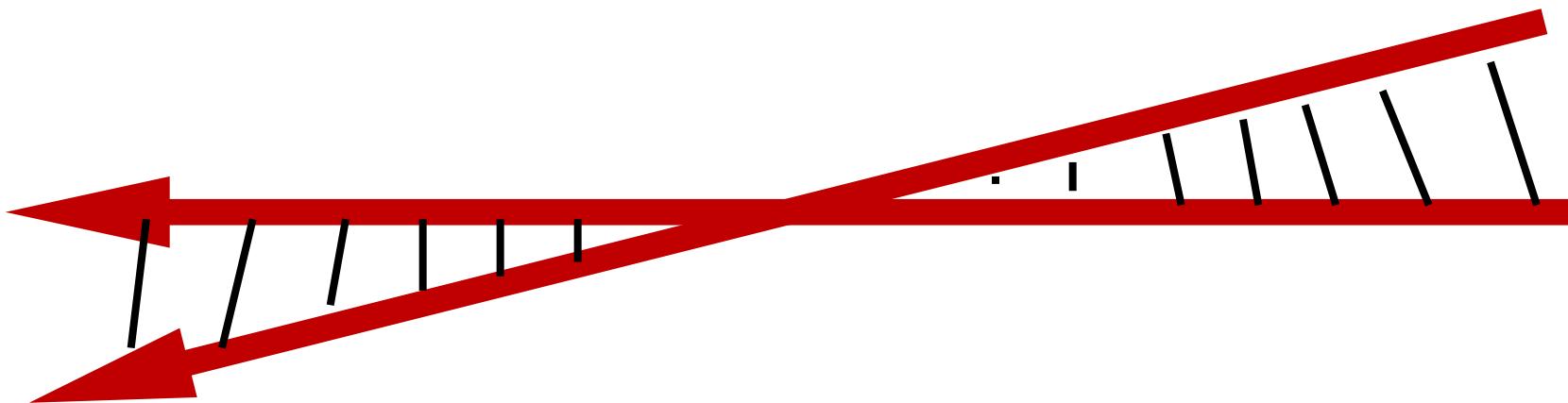
Left-handed when seen
perpendicular to the strands

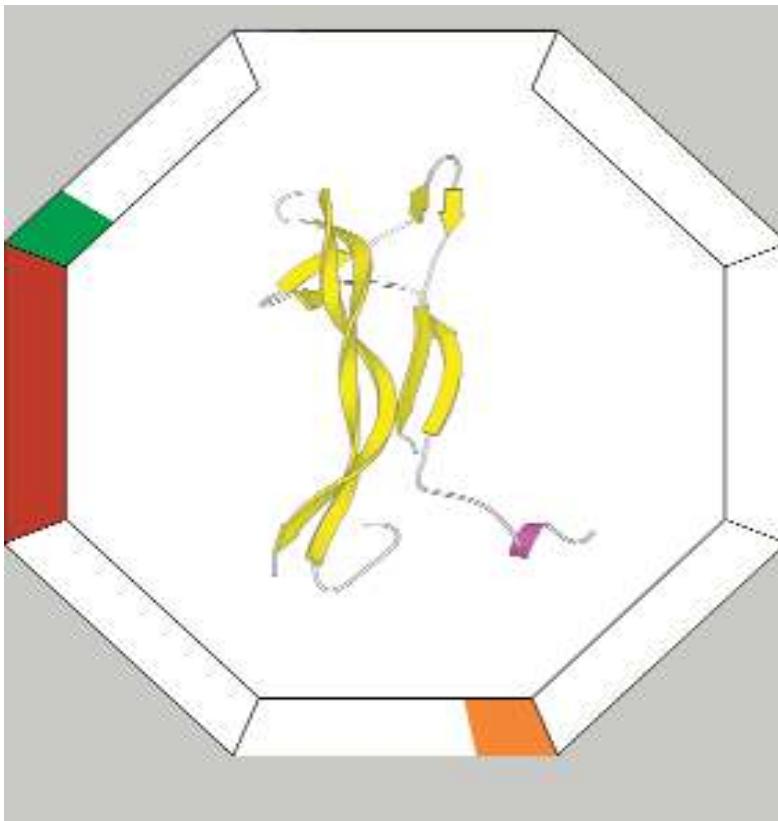
Right-handed when seen
along the strands

Typically $\sim 20\text{-}25^\circ$

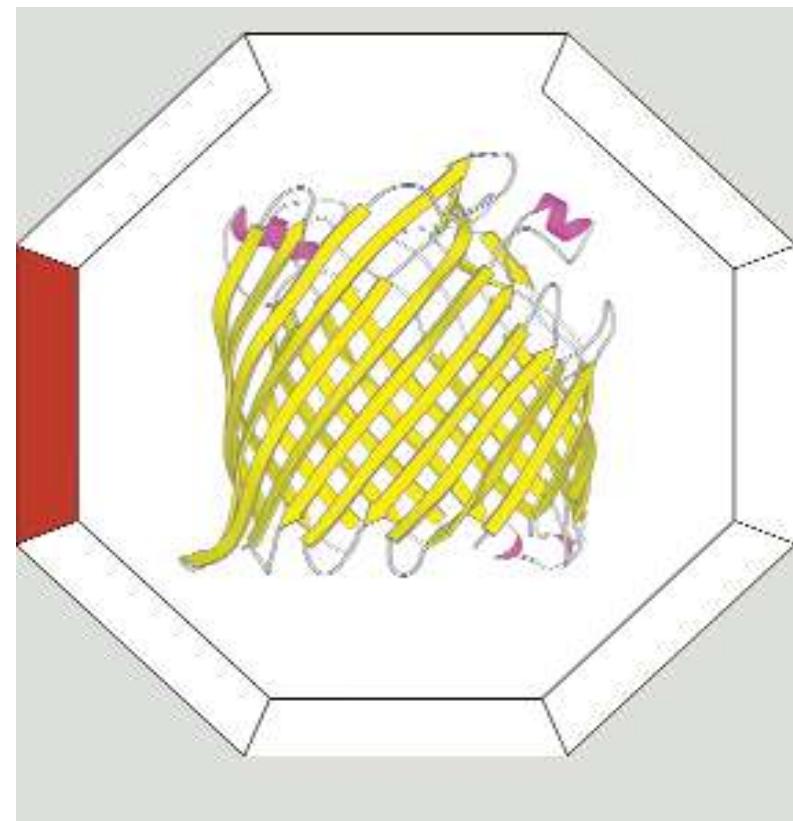
Would we expect long strands to be twisted or planar?

The H-bonds become longer at then ends



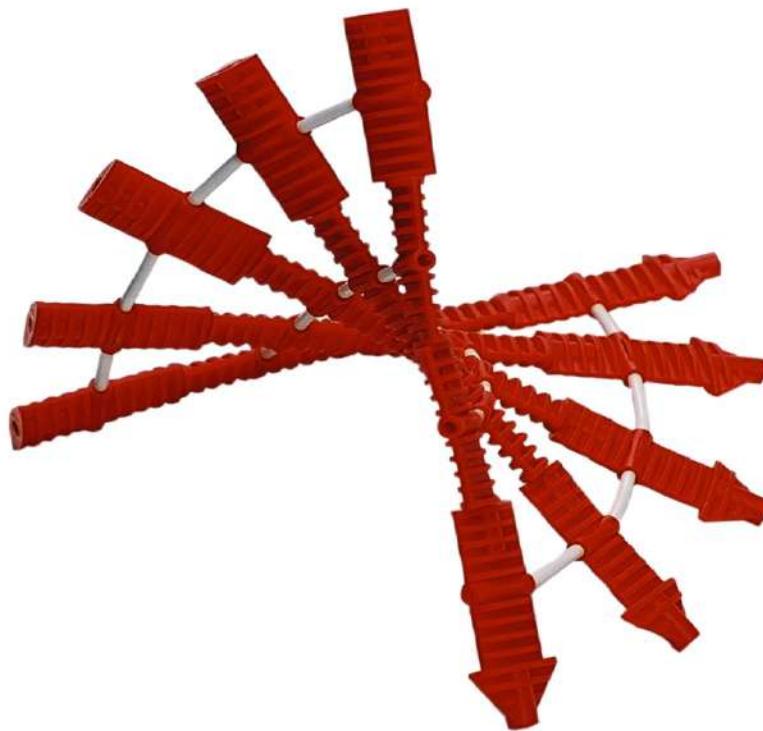


2.10.90



2.40.170

Building a twisted sheet

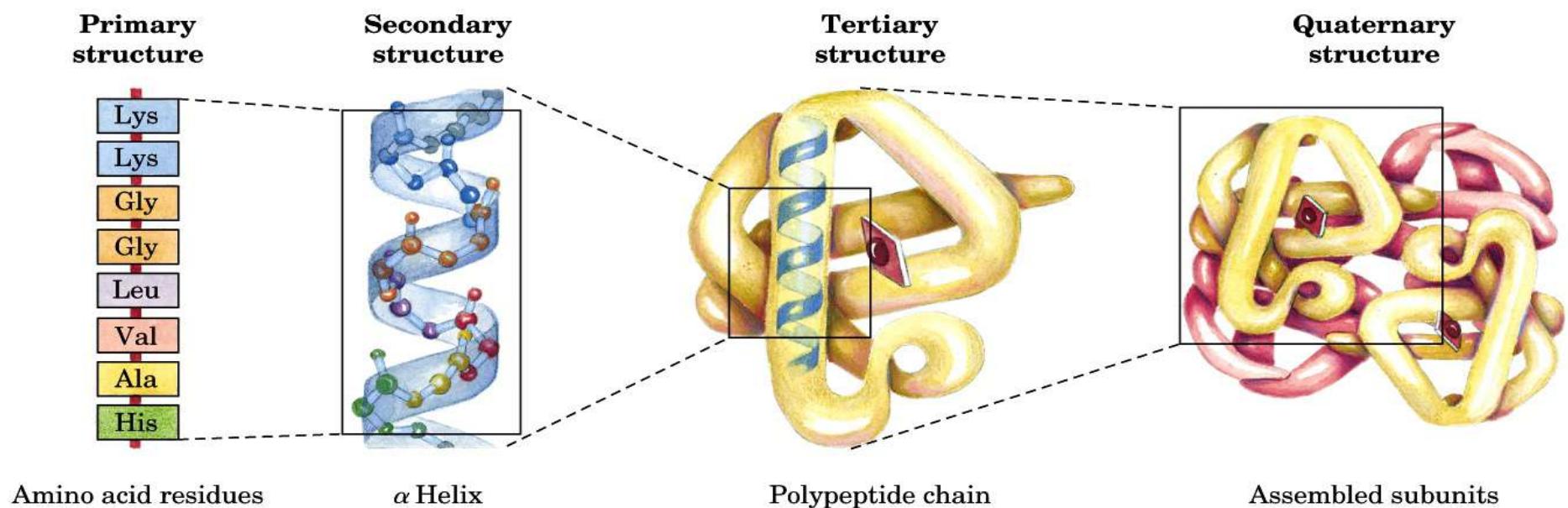


Part 4: Protein Folds

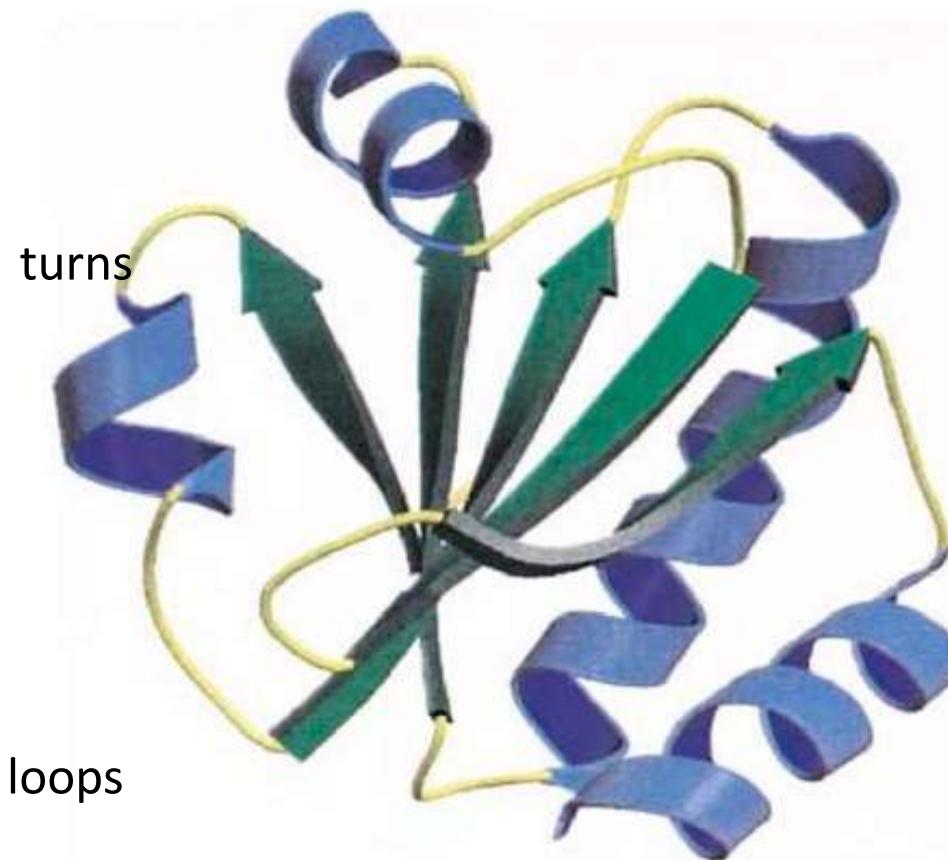
A summary of everything



Protein hierarchy



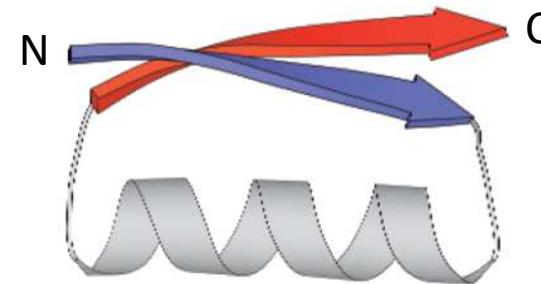
We require connections between secondary structures



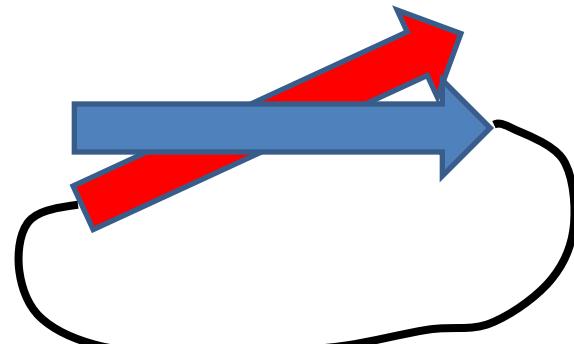
The loops are often more flexible and functionally important.

Motifs

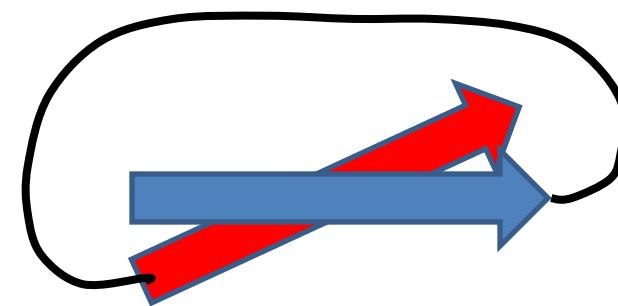
$\beta\alpha\beta$



The connection is “always” right-handed

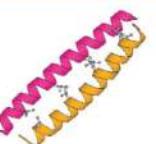
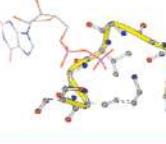
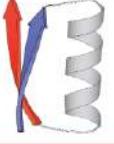
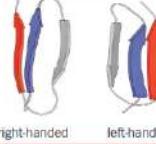


clockwise



anti-clockwise

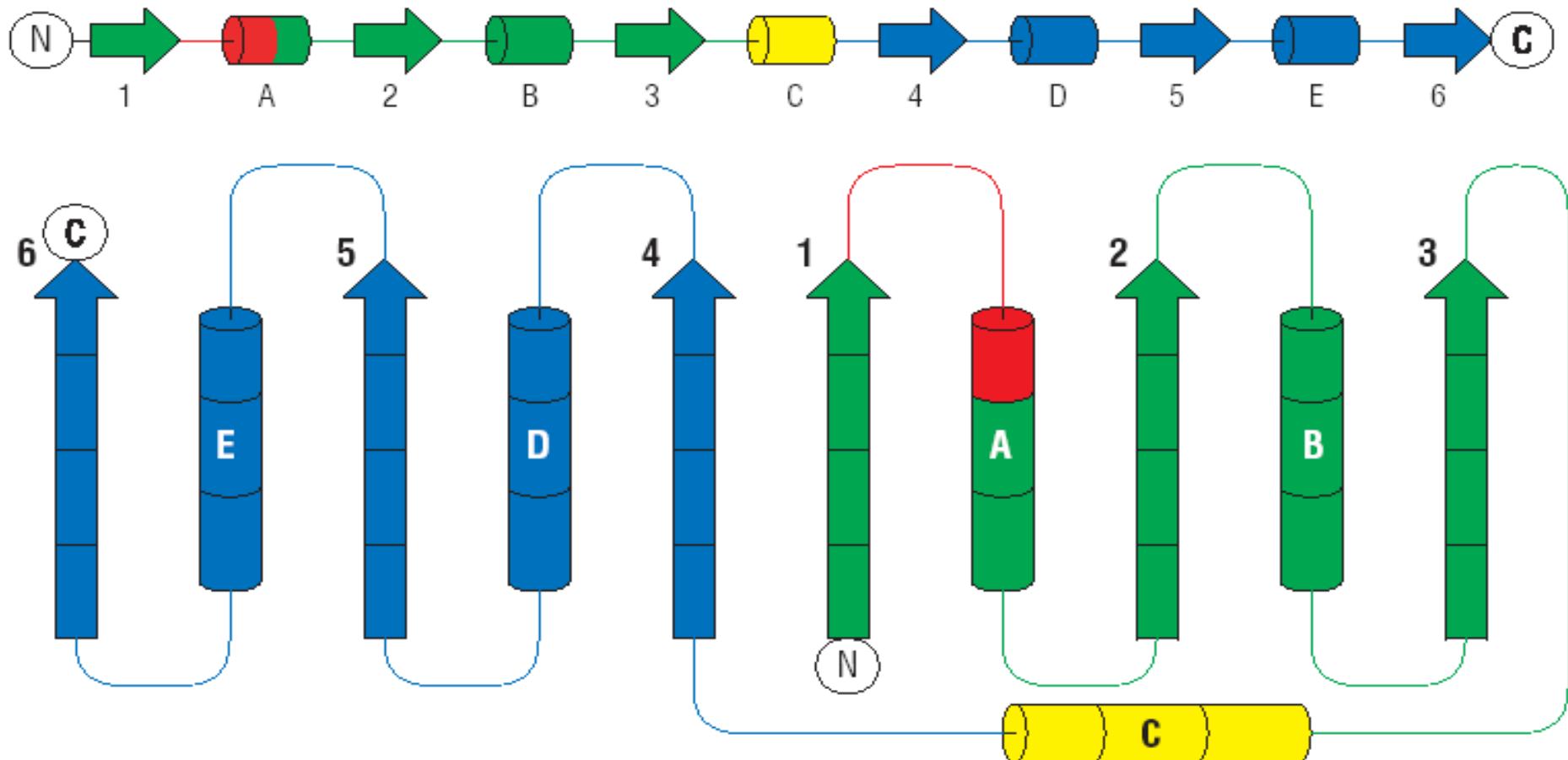
Important structural motifs

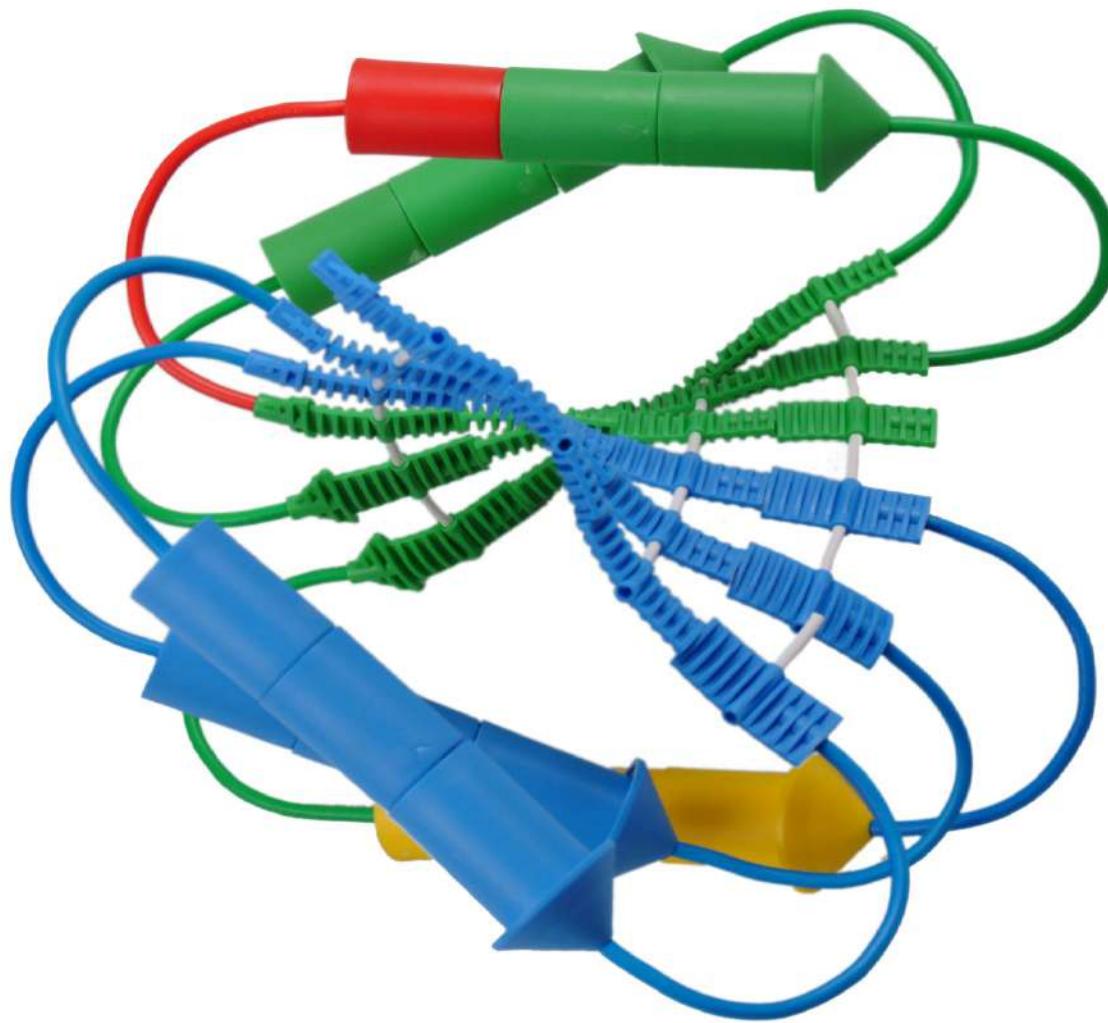
 <p>HTH motif</p>	<p>The helix-turn-helix and EF-hand motifs are both characterized by two orthogonal α-helices. The former is a specific example of an α/α corner and is found in DNA binding proteins, where the second (recognition) helix inserts into the major groove. The EF-hand is observed in Ca^{2+} binding proteins, where the Ca^{2+} is bound by a loop between the two helices.</p>	 <p>Greek key</p>	<p>The Greek key and the interdigitated β-arcs are two of the most commonly observed motifs (or sub-structures) in β-proteins. The former is predominantly observed at the edges of antiparallel β-sheets where the motif is often divided between two such sheets. The latter has been described as the most common sub-structure observed in β-sandwiches.</p>
 <p>Leucine zipper</p>	<p>Both the leucine zipper and the helix-loop-helix are dimerization motifs. In the former case this occurs via the formation of a classical left-handed coiled coil, with leucines at every 7th position (the d position of the coiled coil). In the case of the HLH, the two helices of the motif come together with those of the second monomer to form a 4-helix bundle.</p>	 <p>Zinc finger</p>	<p>Zinc fingers are metal binding motifs involved in DNA recognition. They differ in their Zn^{2+} ligands, 3D structures and DNA binding modes. The example shown is a 'classical' Zinc finger involving two His and two Cys ligands. The P-loop is a glycine rich motif involved in nucleotide binding, where it interacts directly with the α and β phosphate moieties.</p>
 <p>HLH motif</p>		 <p>P-loop</p>	
 <p>Right handed $\beta\beta\beta$</p>	<p>Most connections between parallel β-strands are right-handed, but exceptions are to be seen in the left-handed β-helices of the main table. The $\beta\alpha\beta\beta$ motif includes an additional intervening anti-parallel strand.</p>	 <p>left-handed $\beta\beta\beta$ motif</p>	<p>The Leucine rich repeat (LRR) is characterized by a sequence motif which typically contains 6 leucines. They form a structural motif of a β-strand, α-helix and connecting loop. Several examples of proteins, containing different numbers of repeats, can be seen in the $\alpha\beta$-horseshoes of the main table.</p>
 <p>$\beta\alpha\beta\beta$ motif</p>			

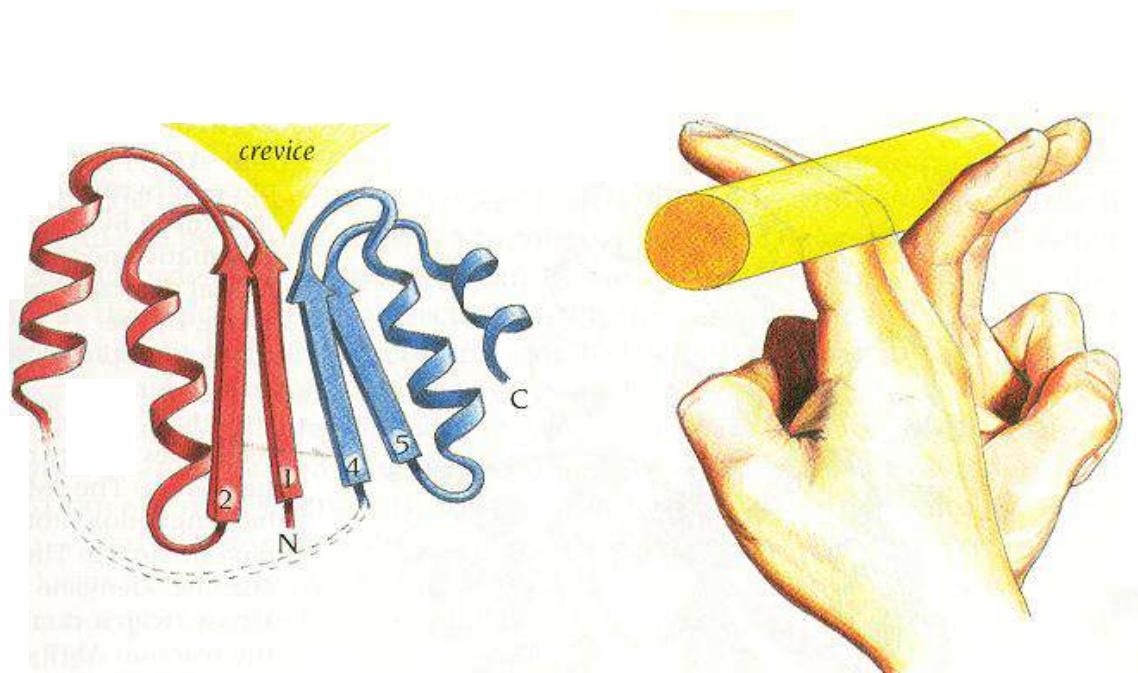
3.1 O Domínio de Ligação ao NAD⁺

The NAD⁺ Binding Domain

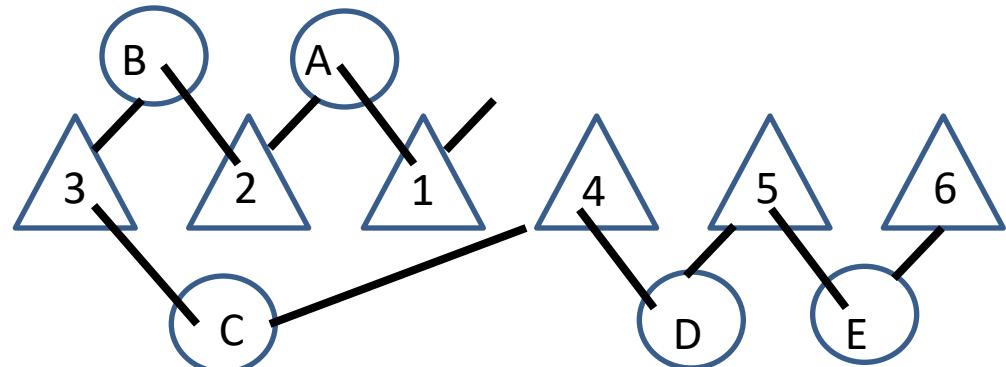
Domínio de Ligação ao NAD⁺



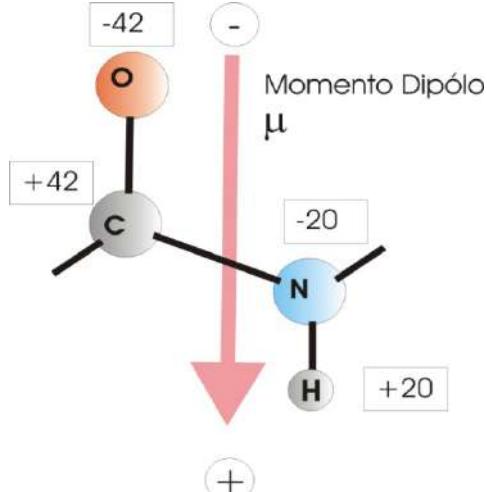




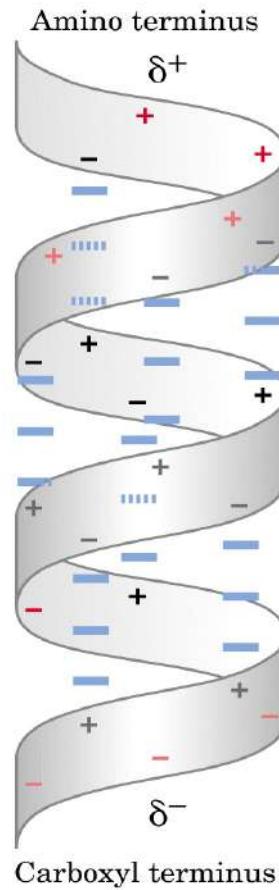
The connection between $\beta 1$ and αA turns to one side of the sheet and that between $\beta 4$ e αD turns to the other, generating a cleft



Electrical dipole



$$\begin{aligned}\mu &= 0.72 \text{ e}\text{\AA} \\ &= 3.46 \text{ D/res} \\ &= 1.155 \times 10^{-29} \text{ Cm}\end{aligned}$$

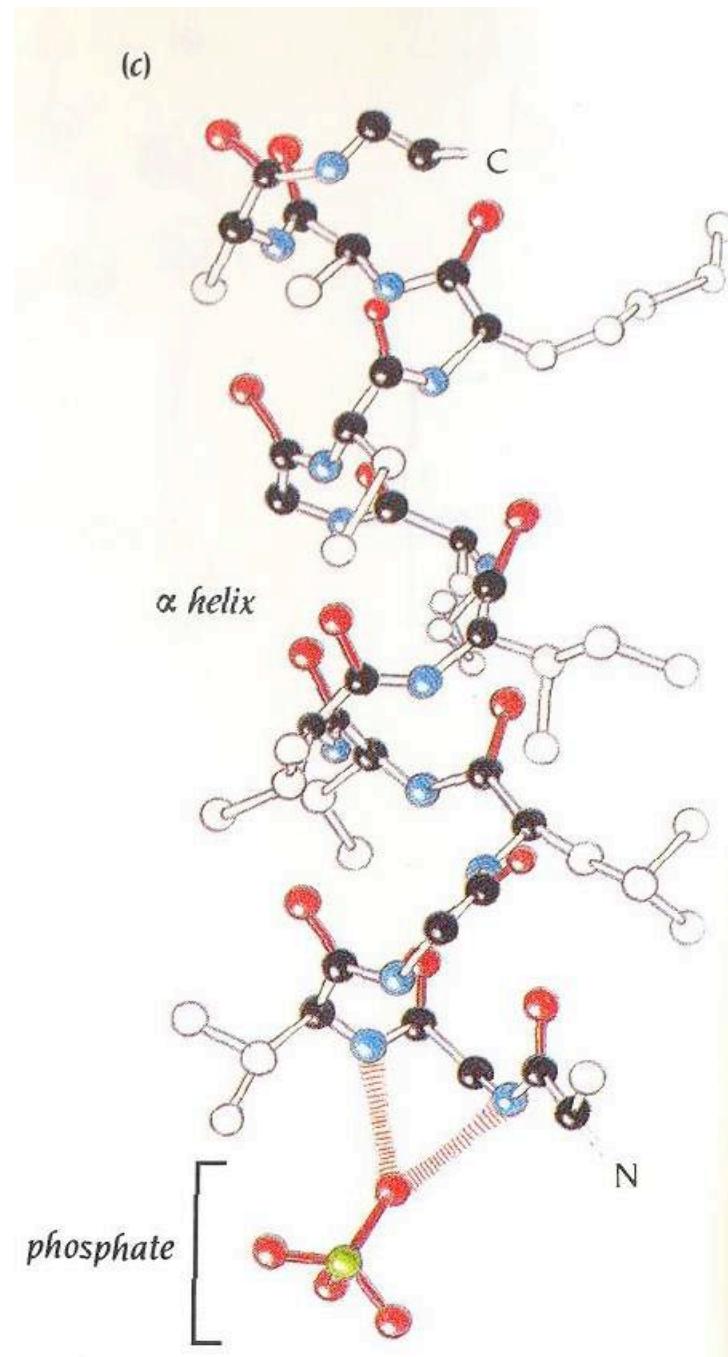


Typically equivalent to $1/2 \epsilon$ at Each end of the helix.

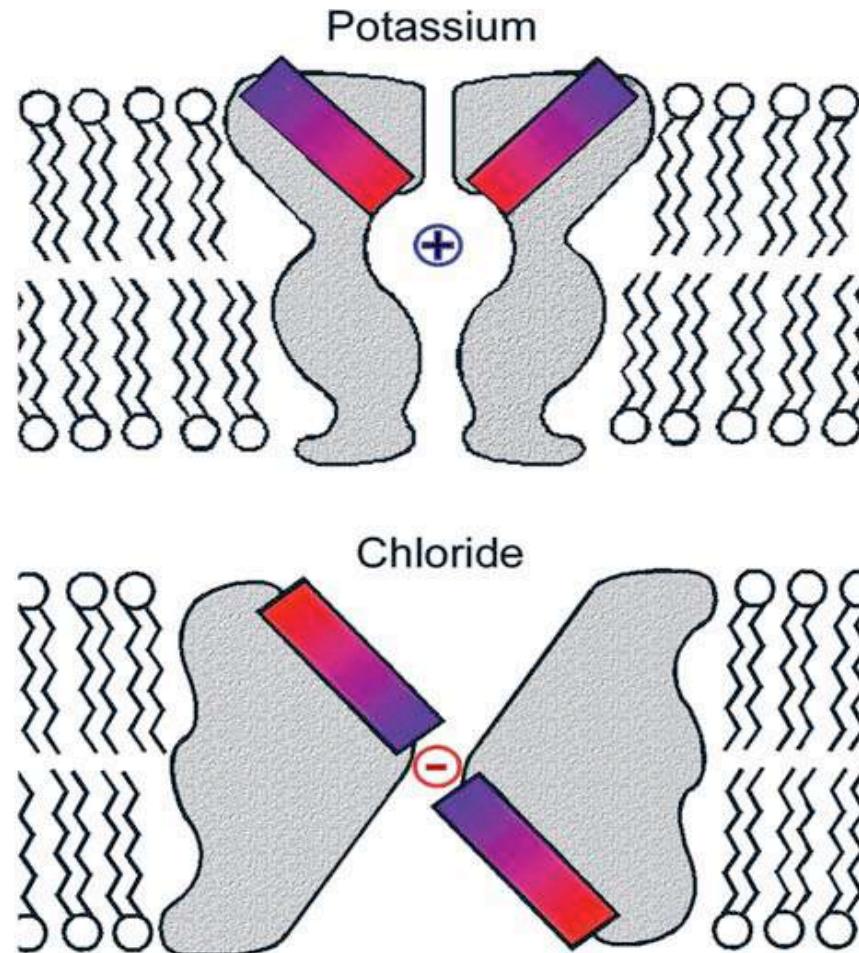
Biological Function

- Help in binding charged cofactors
- Long range attraction (K^+ and Cl^- channels)
- Change the nucleophilic properties of neighbouring residues for catalysis

Phosphate binding site at the N-terminus of the helix.



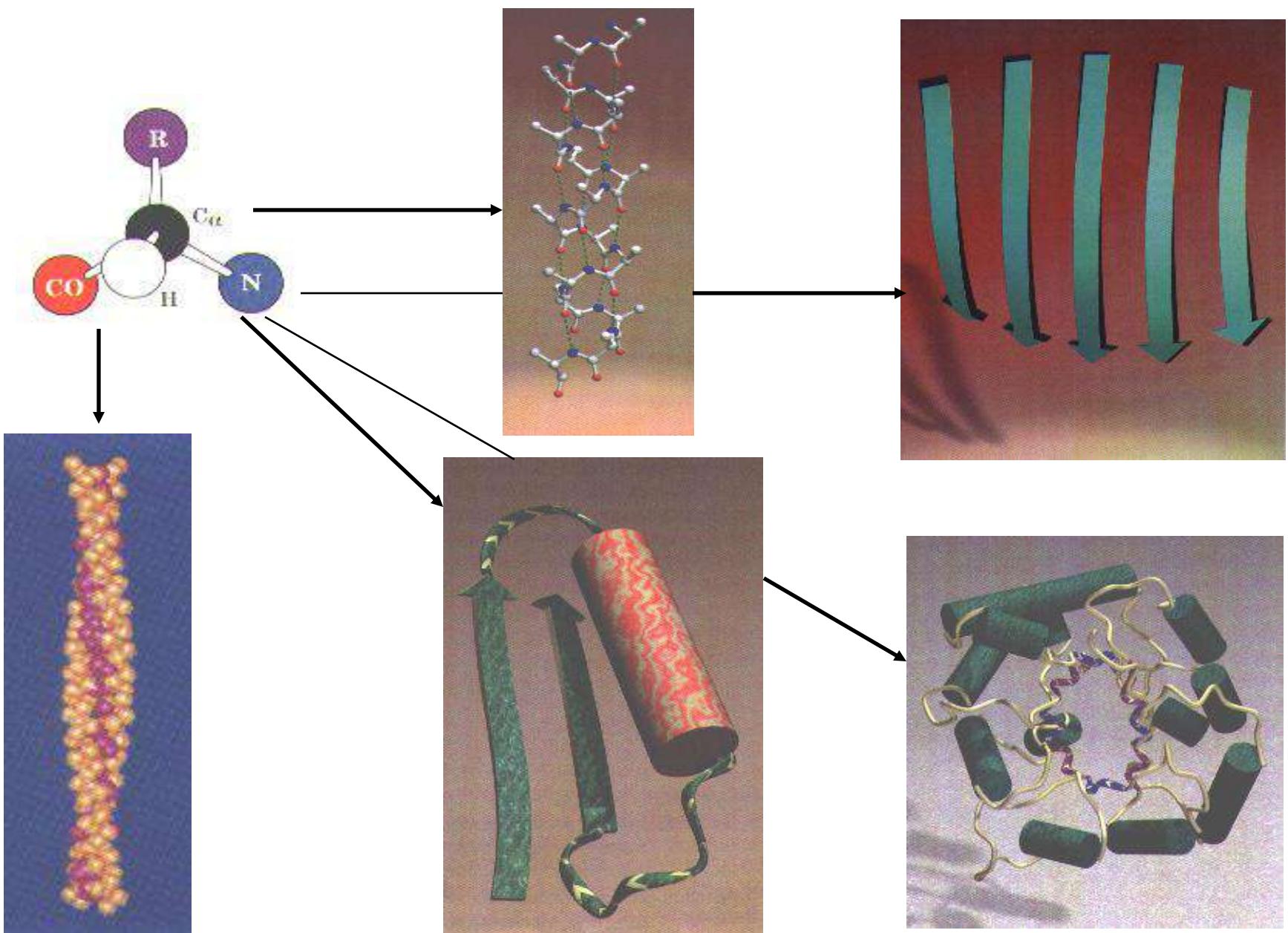
The K⁺ channel uses the C-terminus of α -helices and the Cl⁻ channel the N-terminus to bind their respective ions.

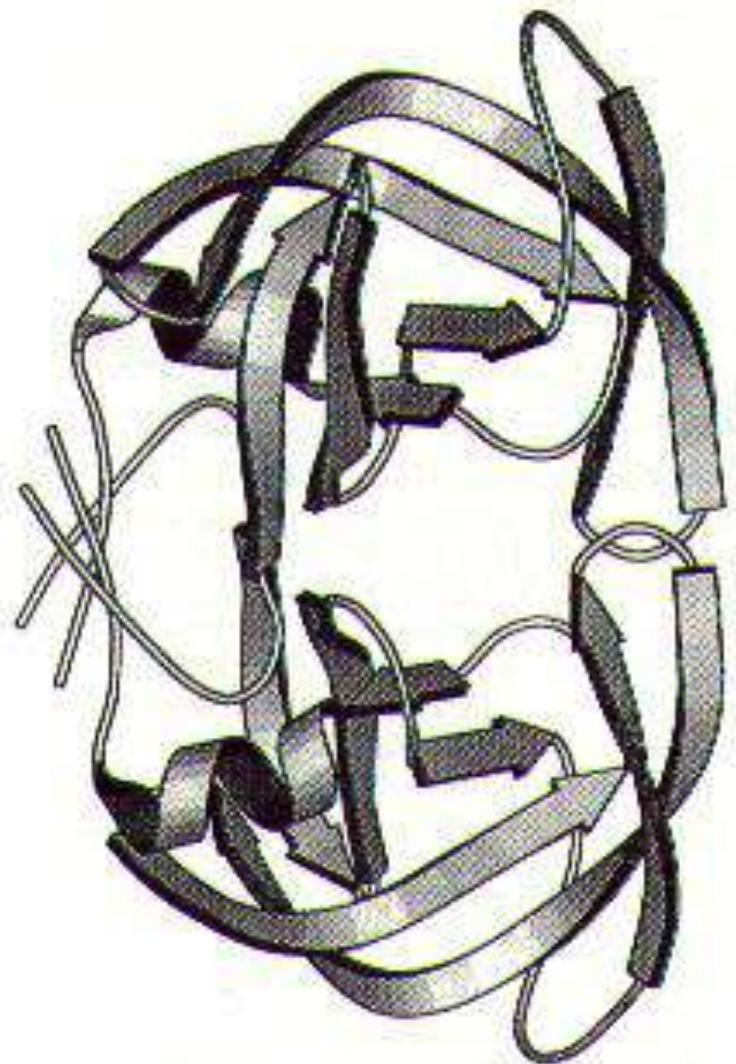


Algumas perguntas adicionais...

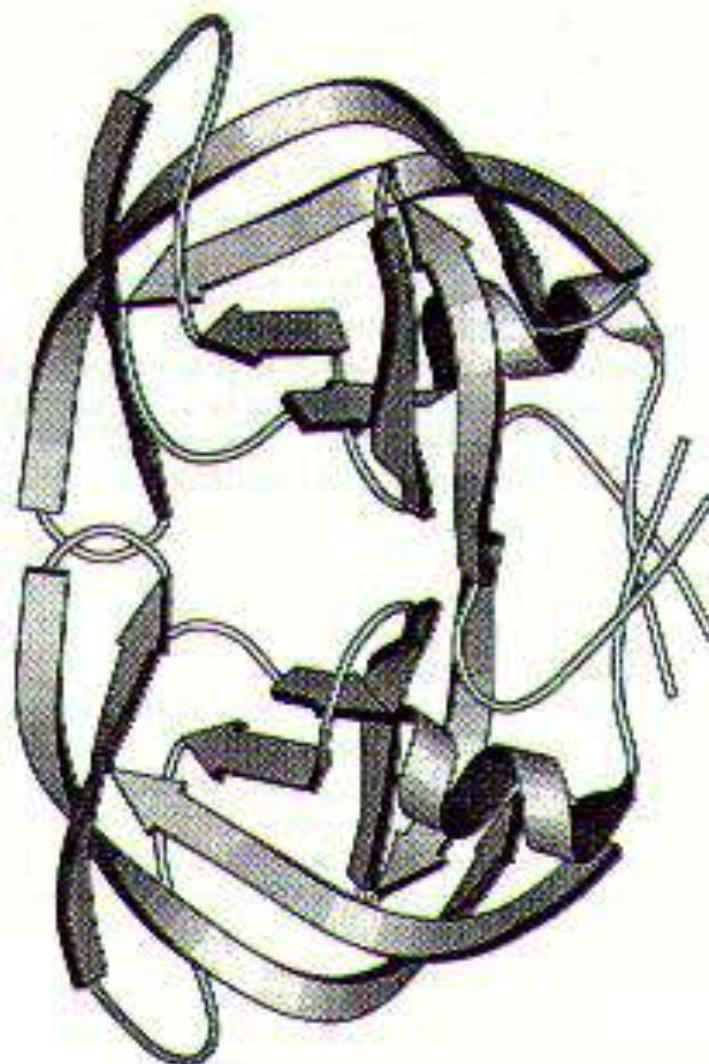
- 1) Usando a tabela principal do *Protein Chart*, qual a arquitetura do enovelamento construído?
- 2) Há algum tipo de simetria (ou pseudo-simetria) na estrutura?
- 3) Usando a tabela de “Basic Topologies of Secondary Structures” no verso do *Protein Chart*, identifique a estrutura construída (tem um nome diferentes!) Localize também na tabela principal a identifique as duas funções mais comumente associadas com este enovelamento.
- 4) É um enovelamento comum ou não?
- 5) Re-desenhe o enovelamento na forma de uma figura tipo “triângulos e círculos”

The importance of chirality





L-HIV

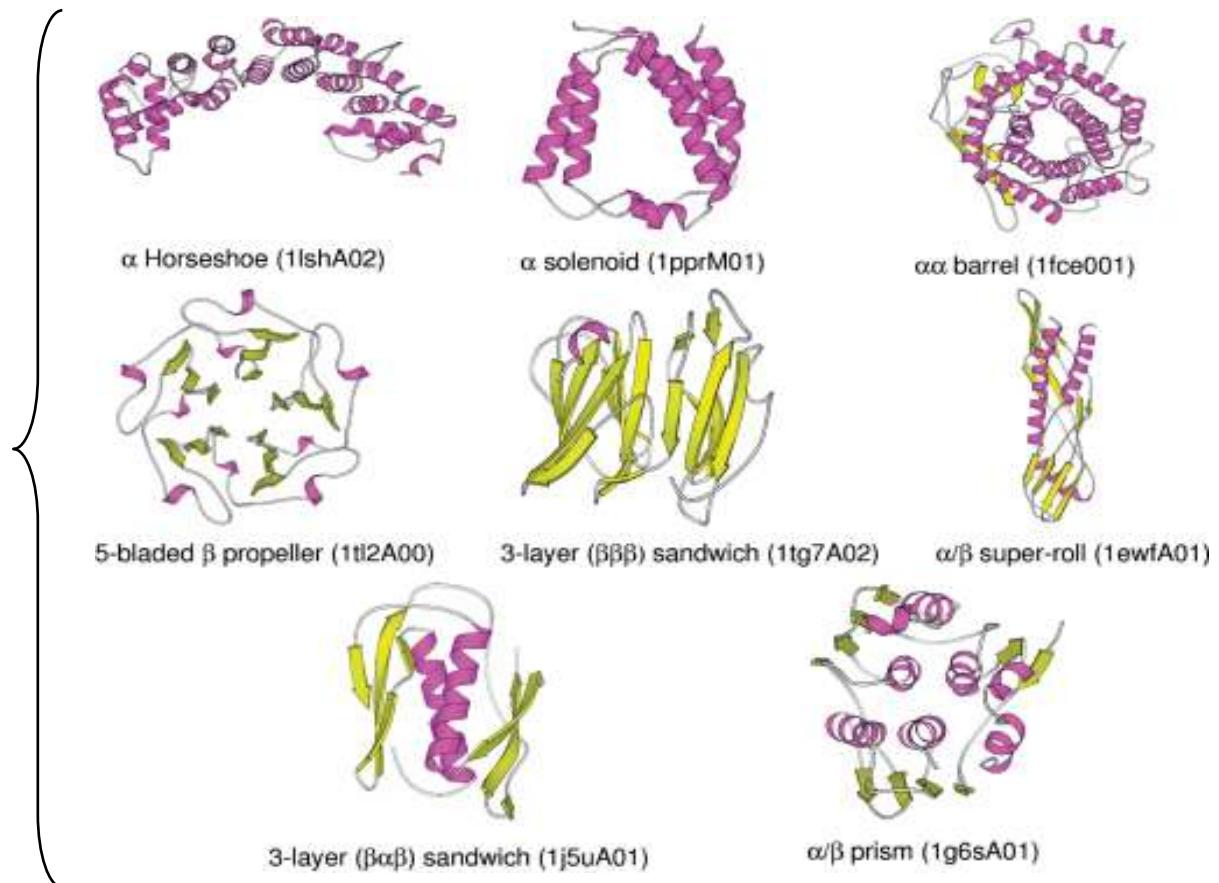


D-HIV

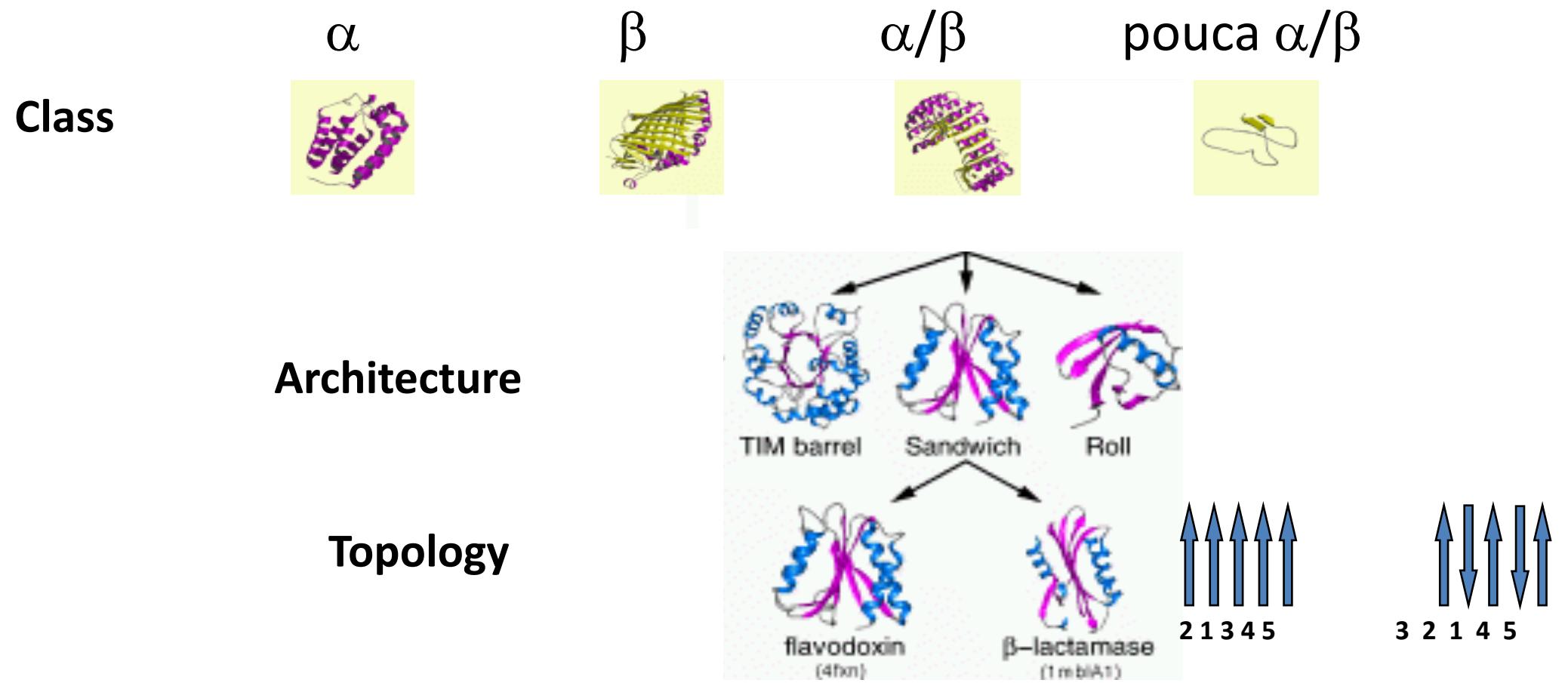
Part 5: Tertiary Structure

Classification of 3D structures

How to describe the diversity?



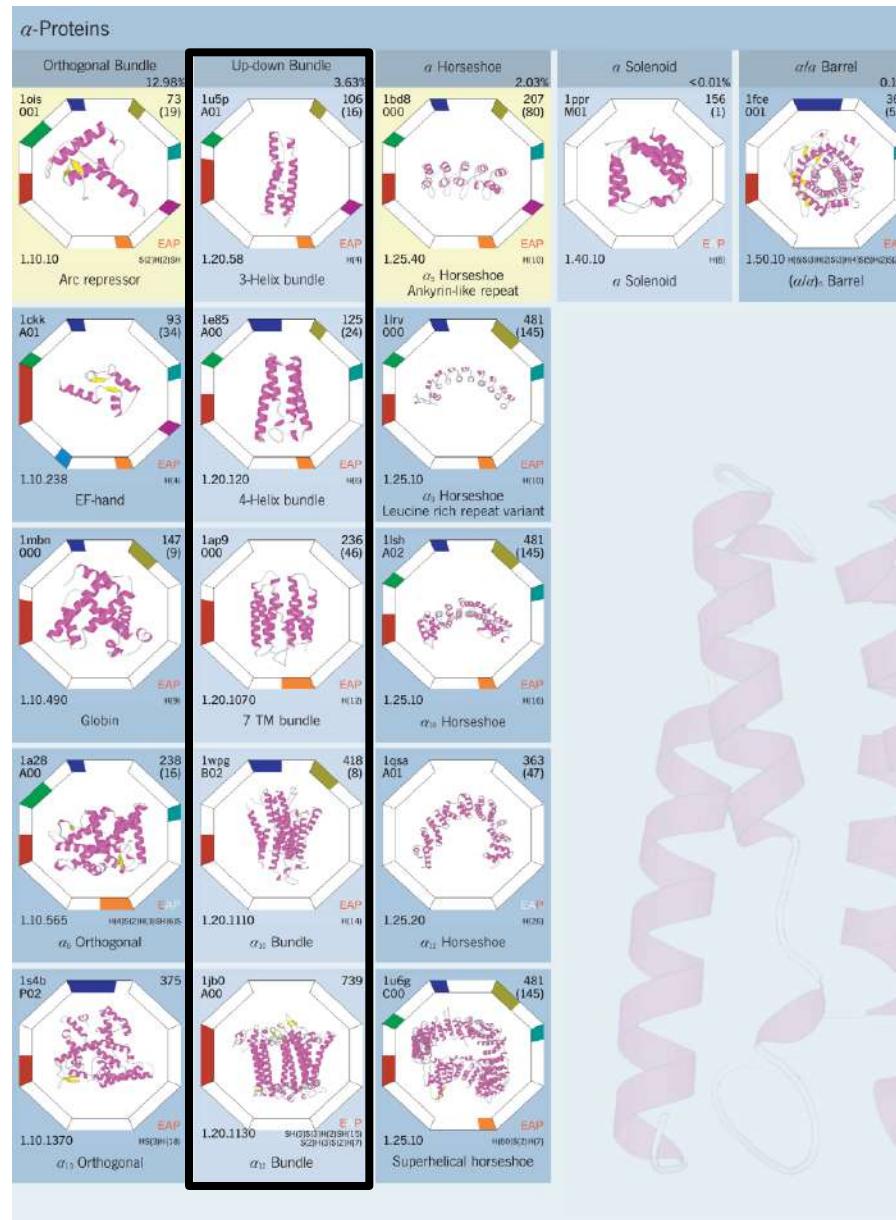
C.A.T.H



A summary of everything

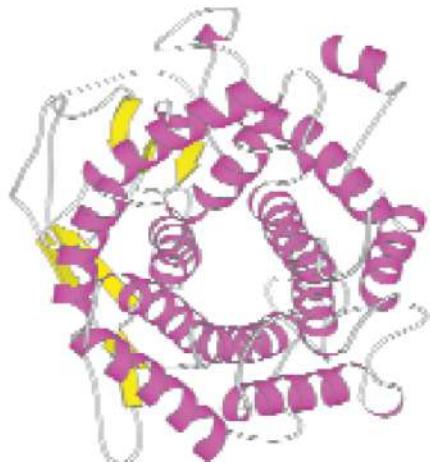


α -proteins

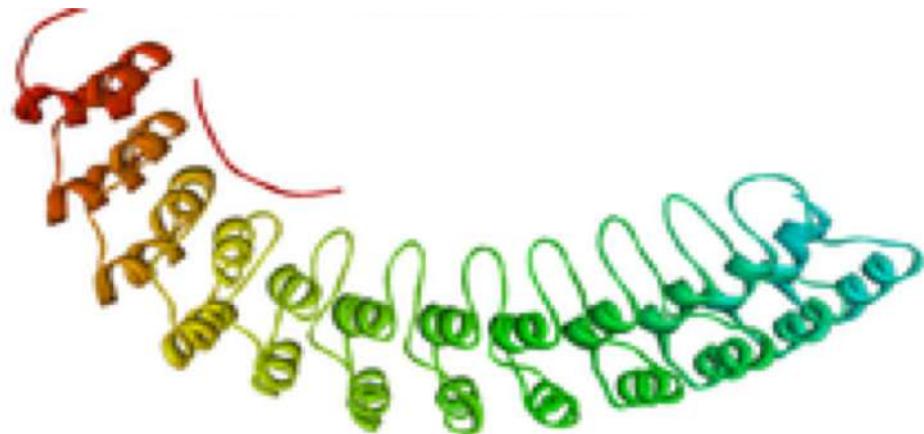




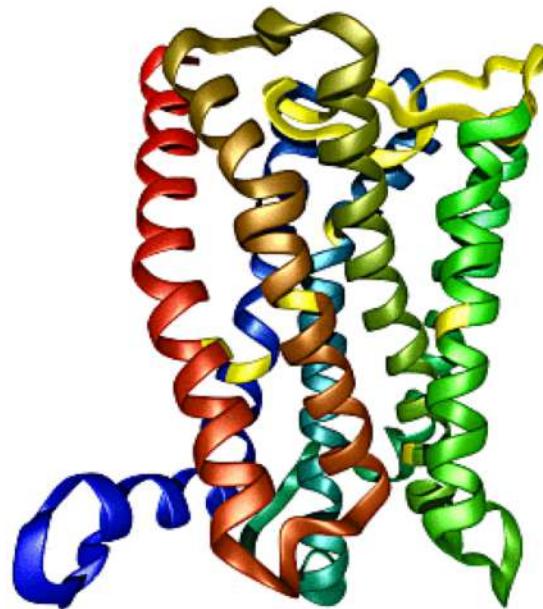
Feixe de 4 hélices



Barril (α/α)₆

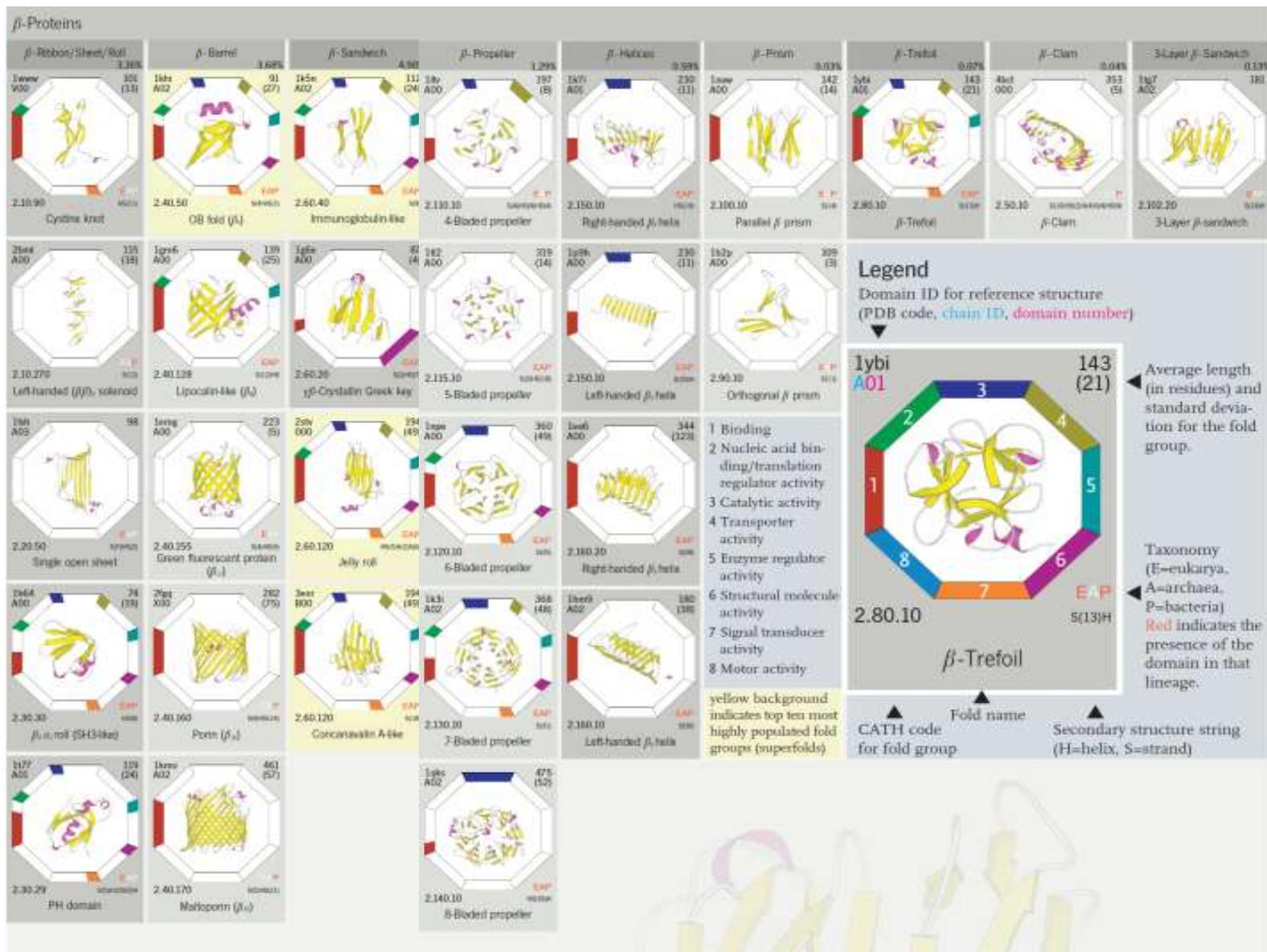


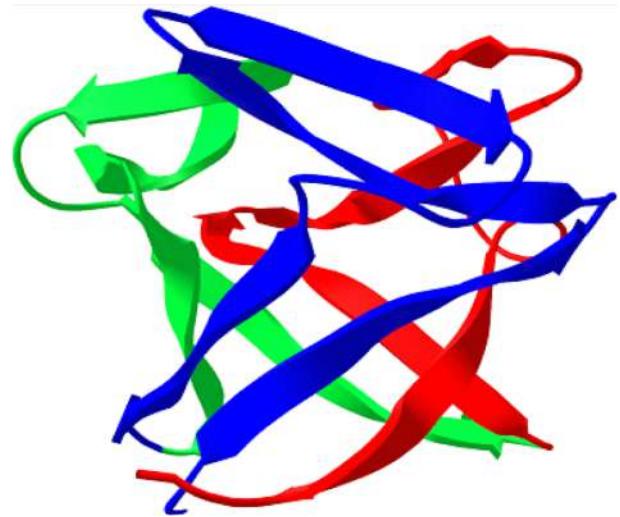
Ferradura



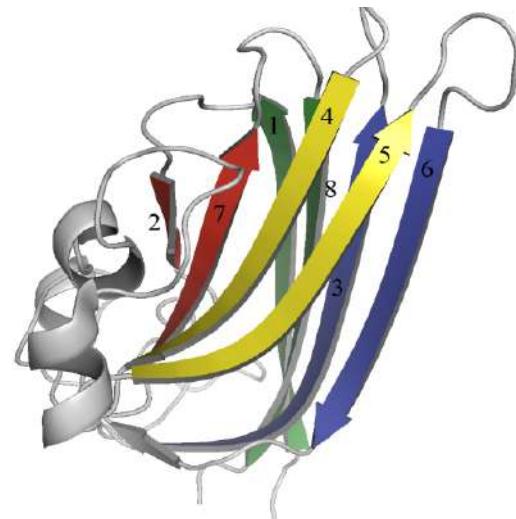
7TM

β -proteins

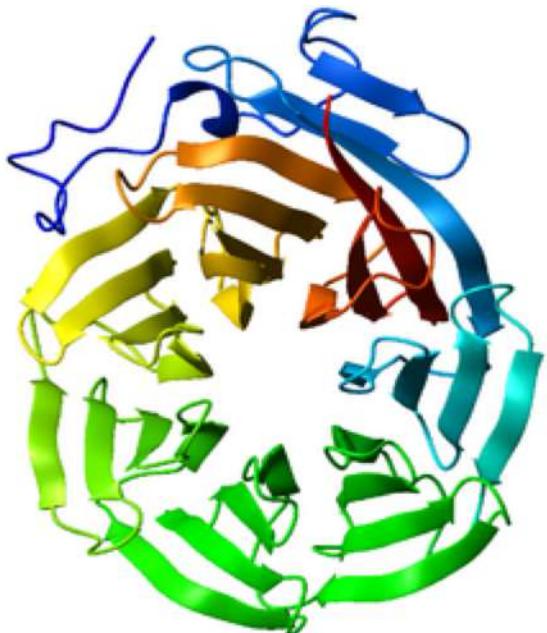




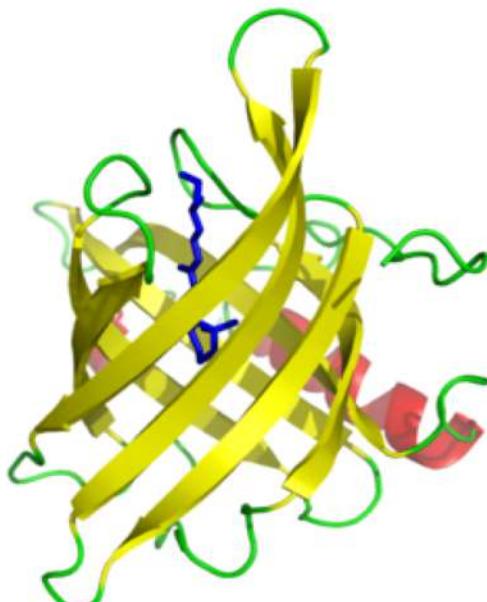
Trefoil-β



Rocambole



Hélice de avião



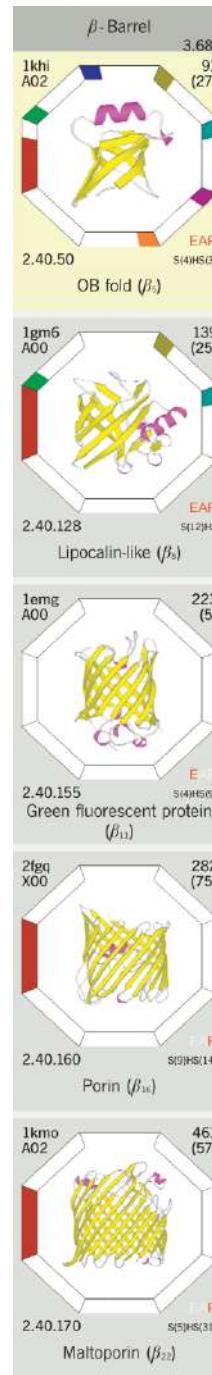
Barril-β



Hélice-β

Structure/Function

- How β -barrel dimensions are related to function



However, it should be remembered that the barrel radius does not only depend on strand number but also *barrel shear*.

OB Fold: Small and common. Compact hydrophobic core and diverse functionality on its outer surface.

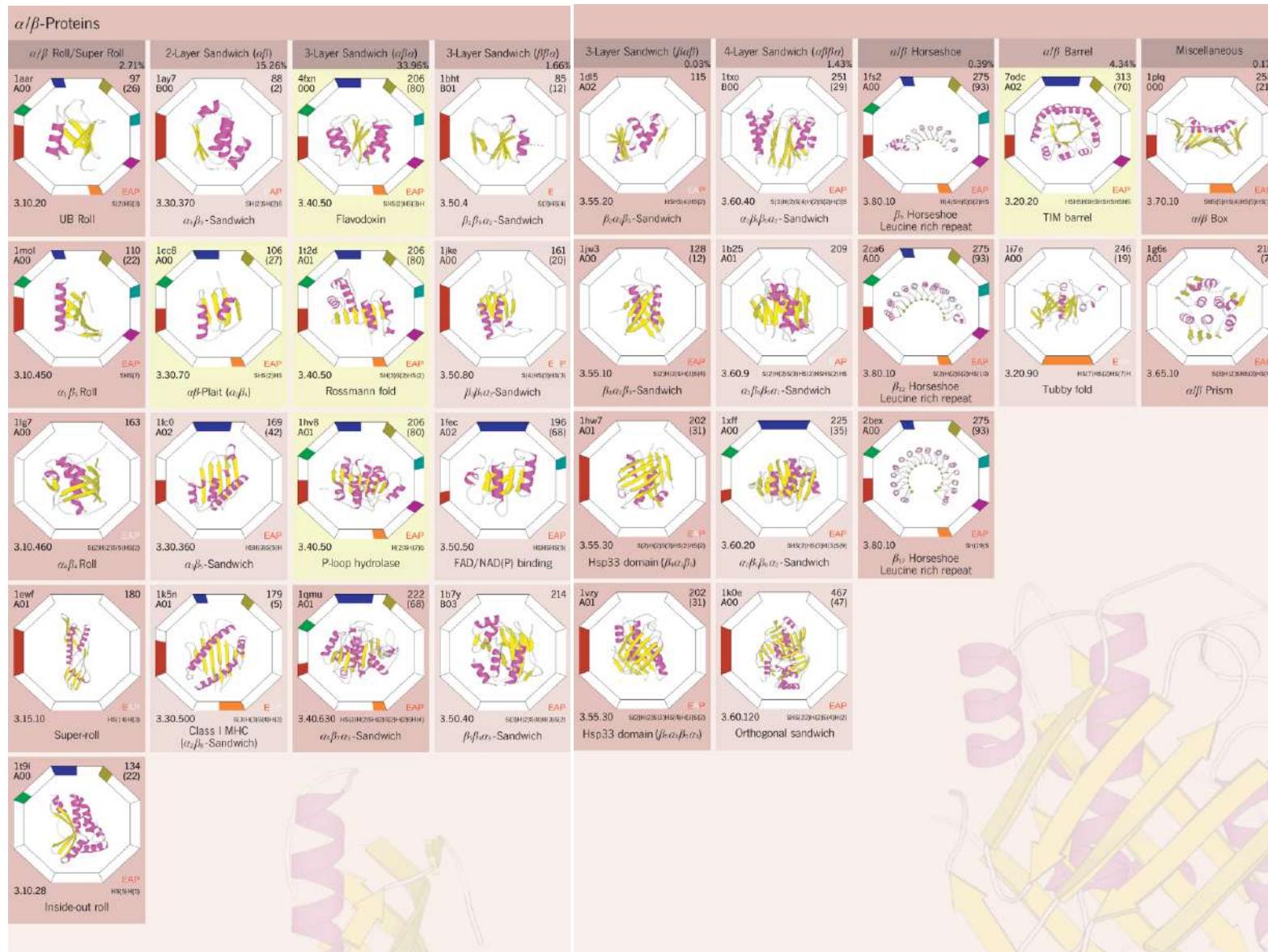
Lipocalins: Barrel is now large enough to bind a small hydrophobic ligand in its internal cavity

GFP: The larger barrel is now sufficient for part of the polypeptide chain to pass right through its center.

Porin: The barrel radius is now so large that it is impossible to fill it with sidechains. The result is a pore used for transport across membranes

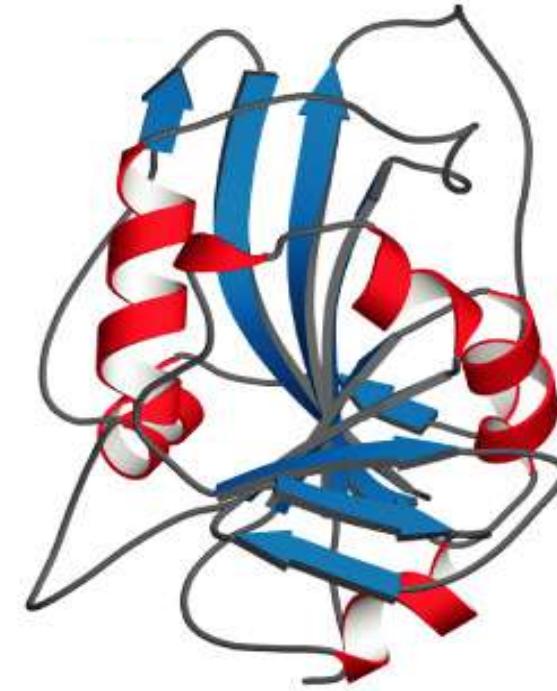
Maltoporins: The pore is even larger permitting transport of larger molecules

Proteínas α/β

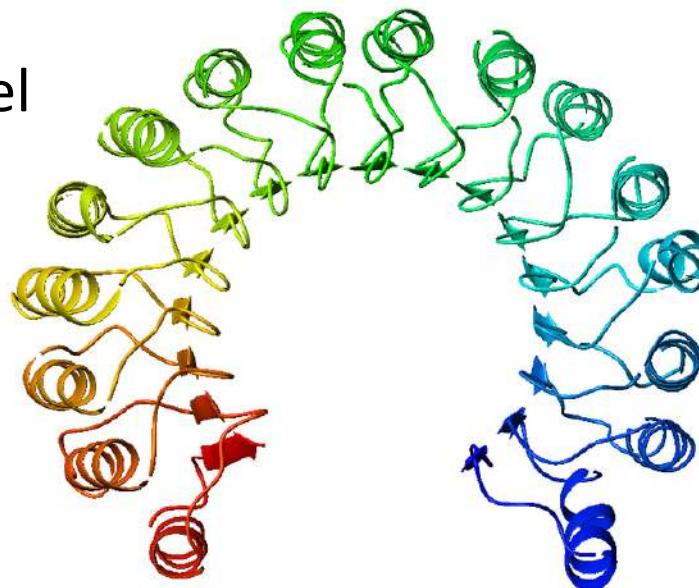




TIM-barrel

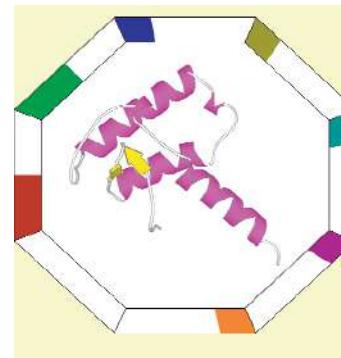


$\alpha\beta\alpha$ sandwich

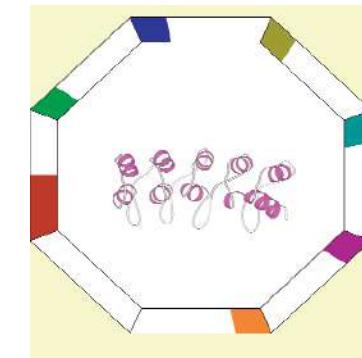


$\alpha\beta$ horseshoe

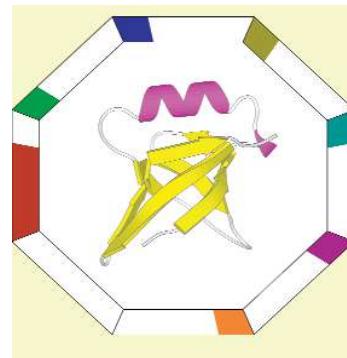
Superfolds



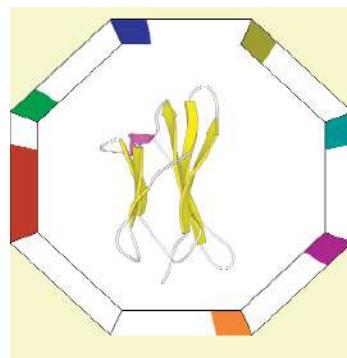
ARC repressor



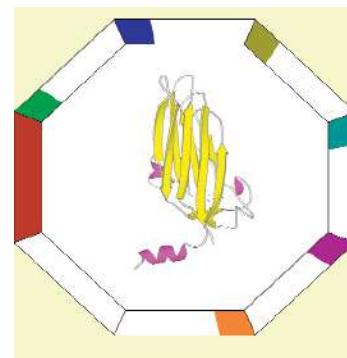
α 5 horseshoe



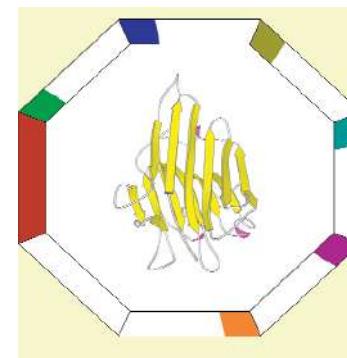
OB fold



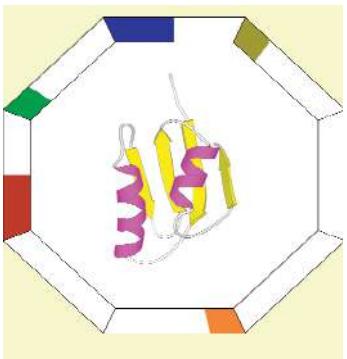
Ig like



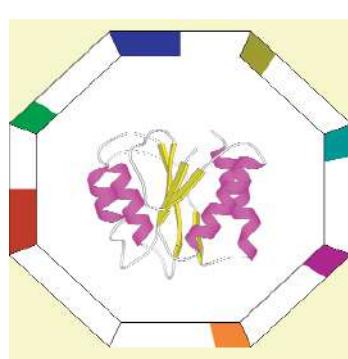
Jelly Roll



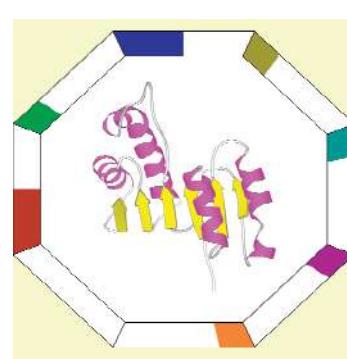
ConA like



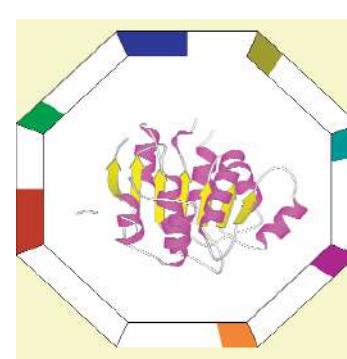
$\alpha\beta$ -plait



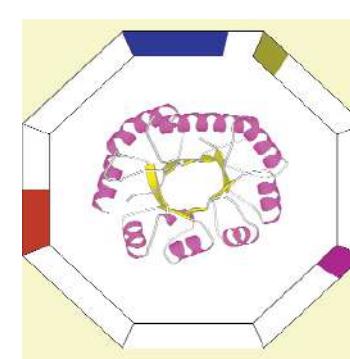
Flavodoxin



Rossmann fold



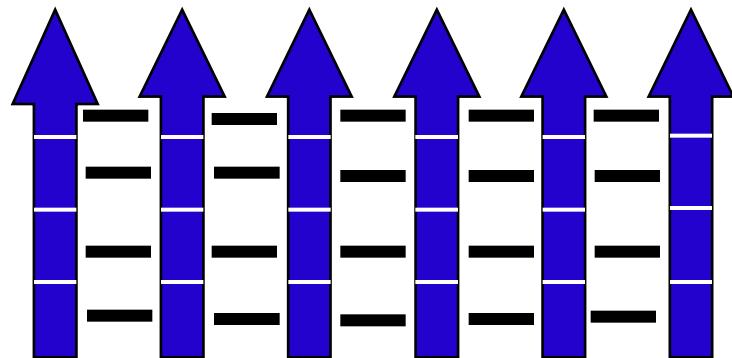
P-loop hydrolase



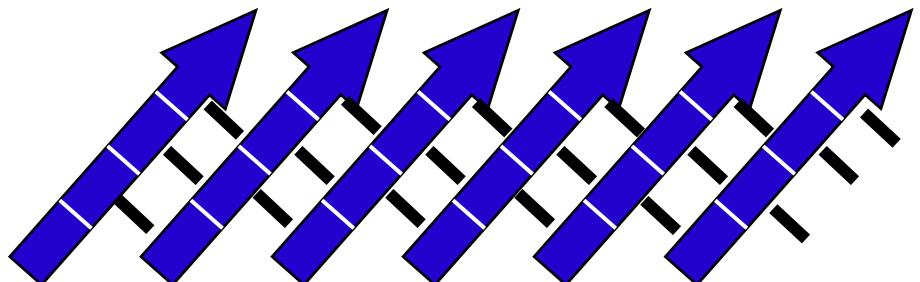
TIM barrel

Part 6: Barrels – why are they different?

Barrels have shear!

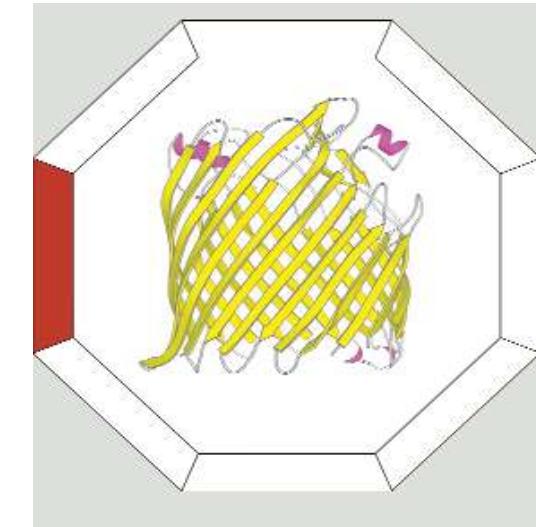
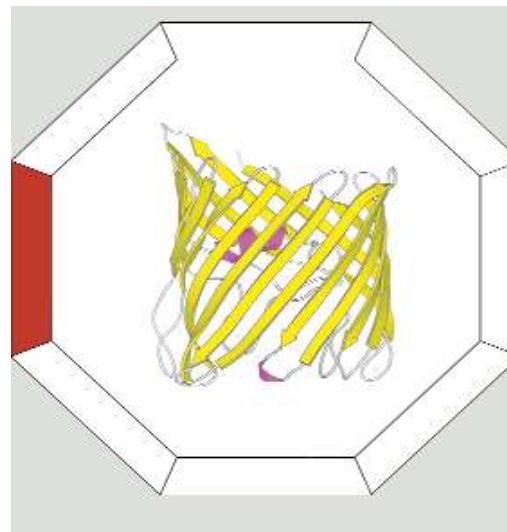
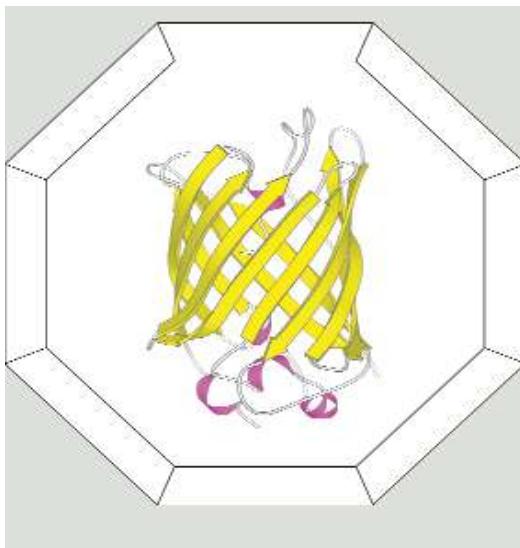
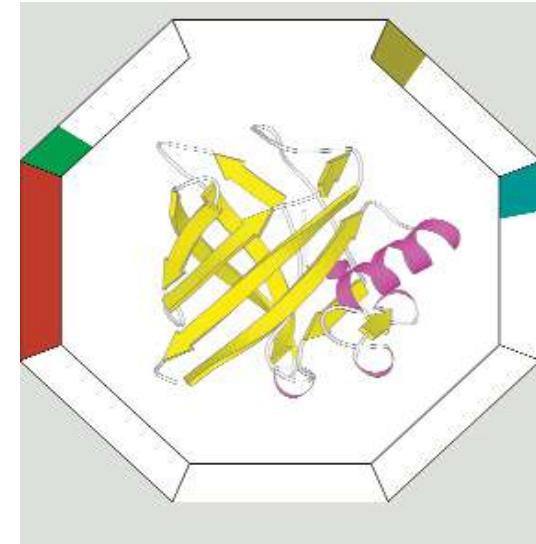
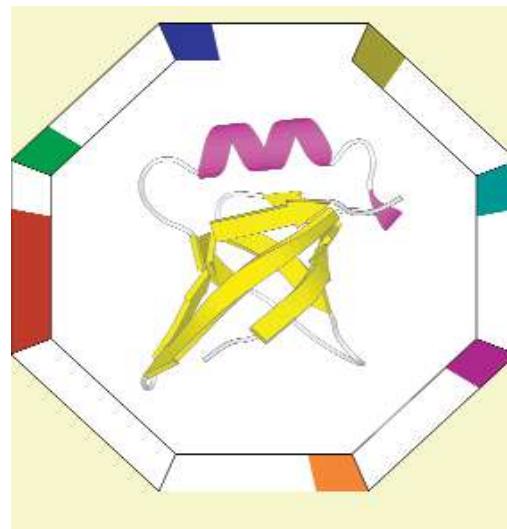


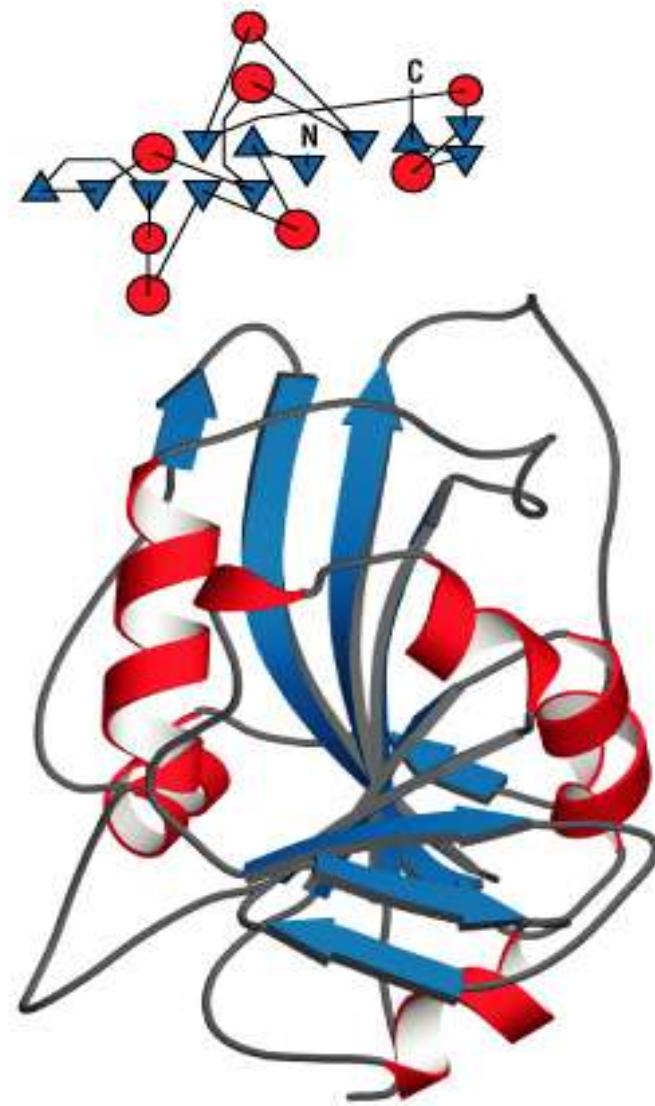
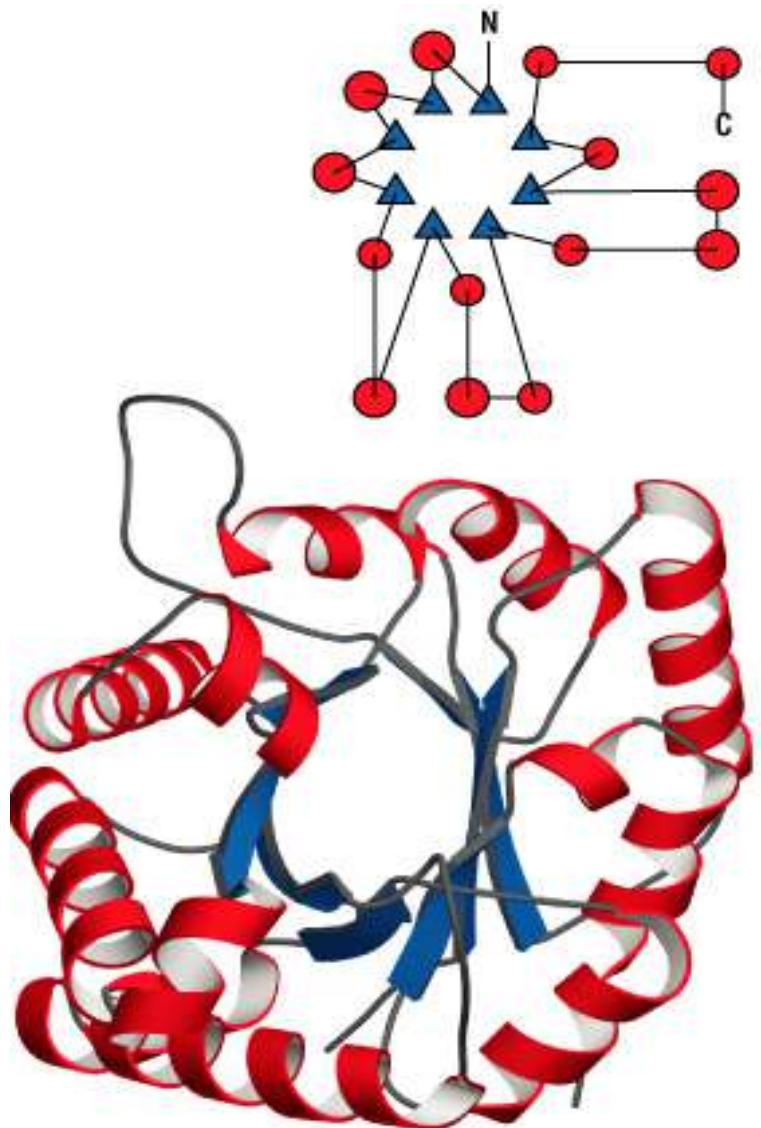
β -sheet without shear
Saddle



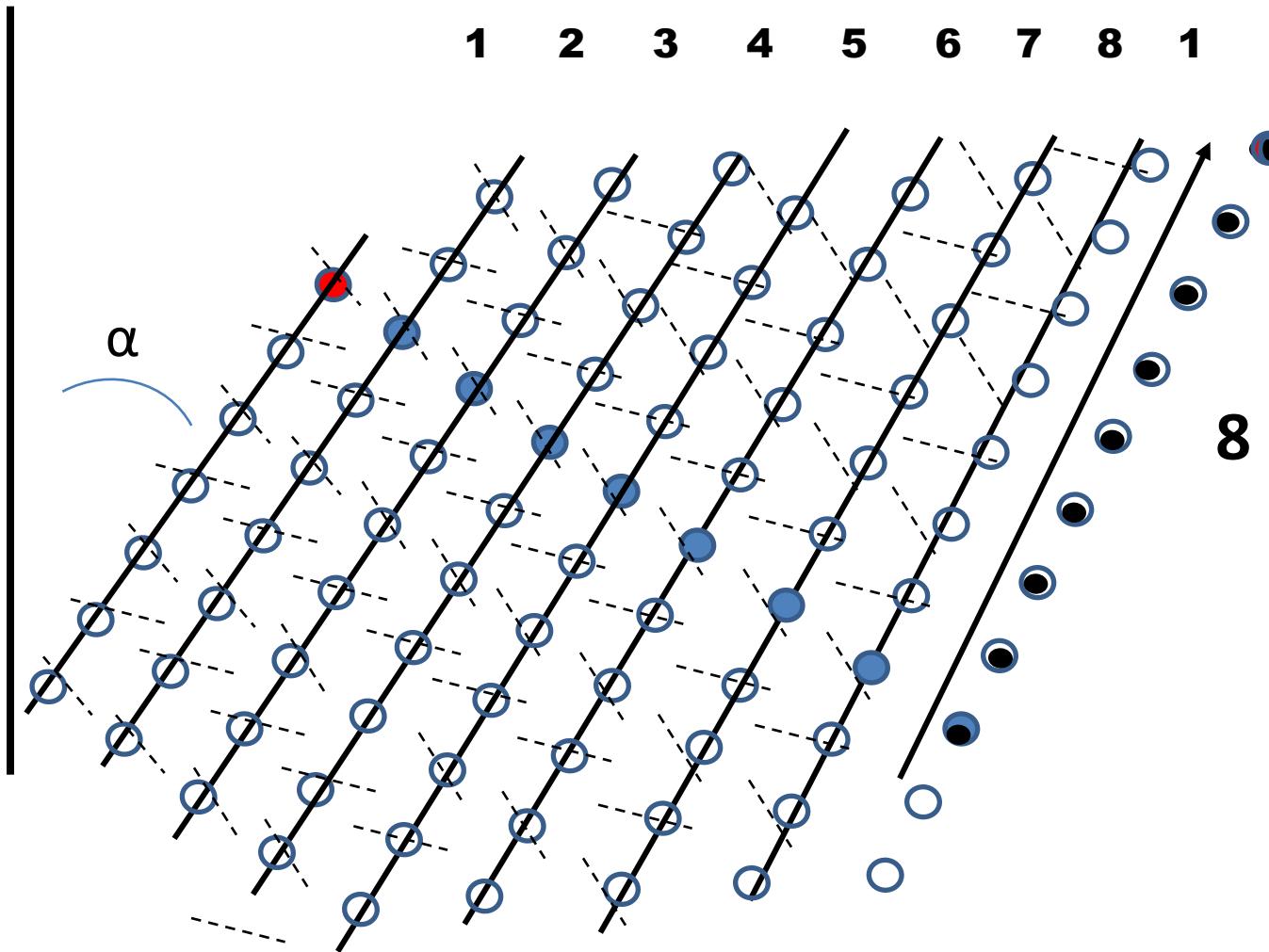
β -sheet with shear
Barrel

As fitas de um barril estão sempre inclinadas em relação ao eixo





Definition of the Shear number (S)



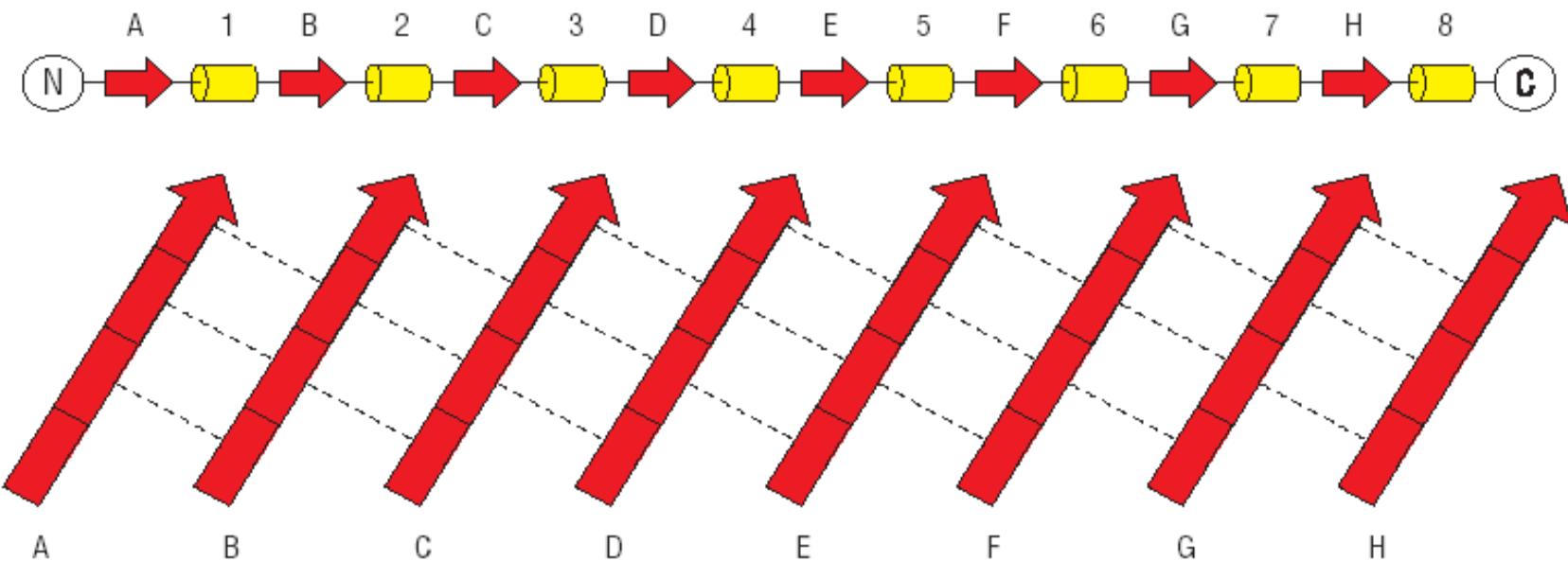
3.2

O Barril TIM (Triose Fosfato Isomerase) ou Barril $(\beta\alpha)_8$

The TIM (Triose Phosphate Isomerase) Barrel or $(\beta\alpha)_8$ Barrel

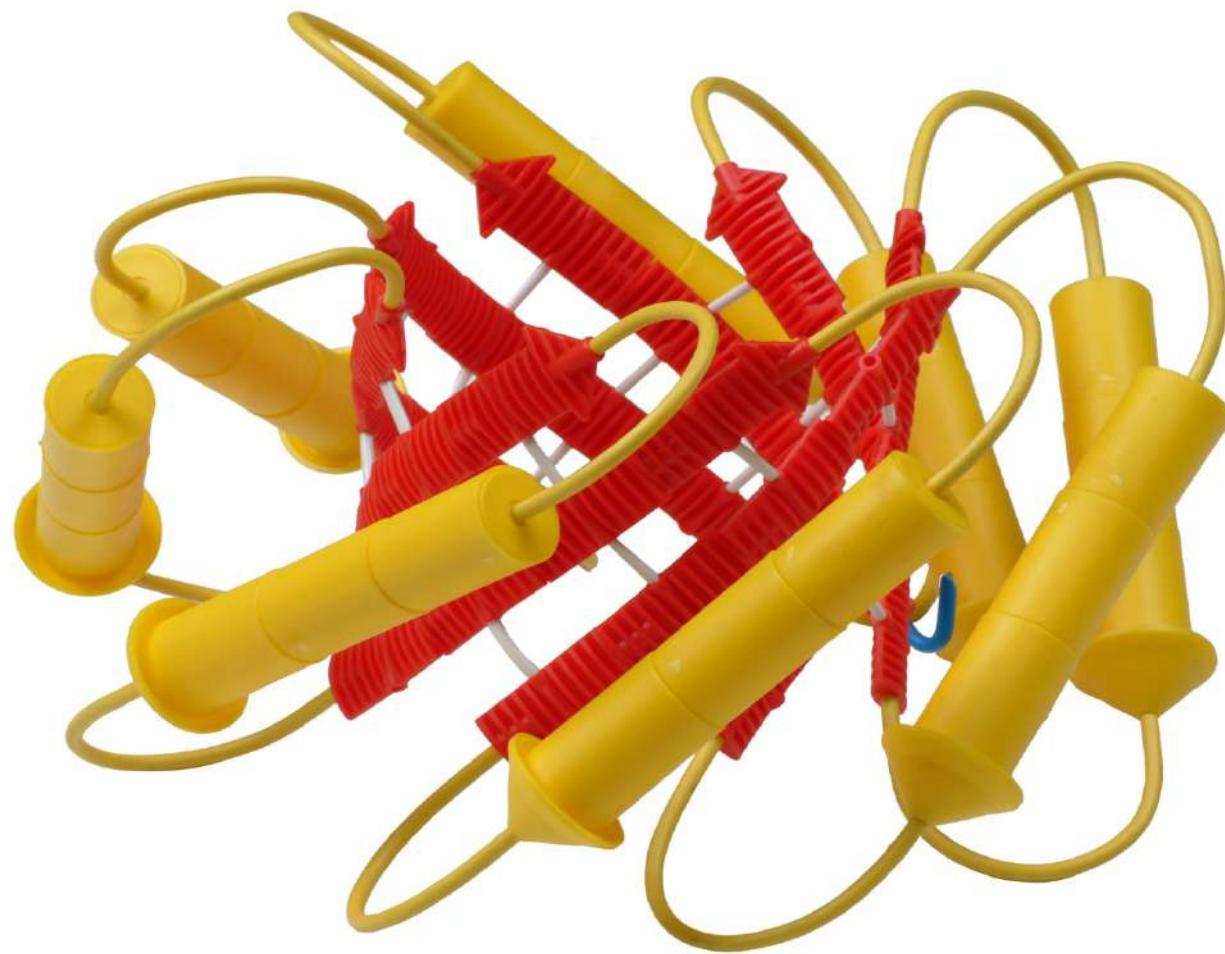


Barril TIM



Painel 2
Panel 2

PROTEIN FOLDER
TEACHER'S GUIDE



Many different enzymes have the TIM Barrel structure

flavocitocromo b2

glicolato oxidase

trimetilamino desidrogenase

RUBISCO

mandelato racemase

muconato cicloisomerase

α -amilase

cyclodextrina glicosiltransferase

piruvato quinase

enolase

xilose isomeraseN-(5'fosforibosil)antramilato isomerase

Triptofano sintase

triose fosfato isomerase (TIM)

Frutose bisfosfato aldolase

adenosina desaminase

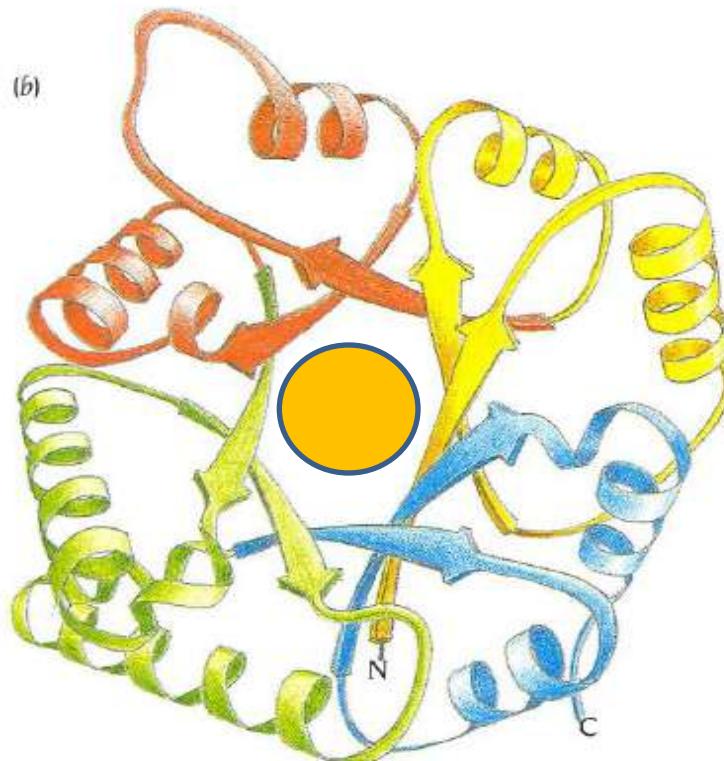
R1 de ribonucleotídeo redutase $((\beta\alpha)_4\beta\beta(\beta\alpha)_4)$

Funções diversas mas todas são enzimas

- isomerização de açúcares
- oxidação por cofatores ntipo flavina
- transferência de fosfato
- degradação de polímeros de açúcares

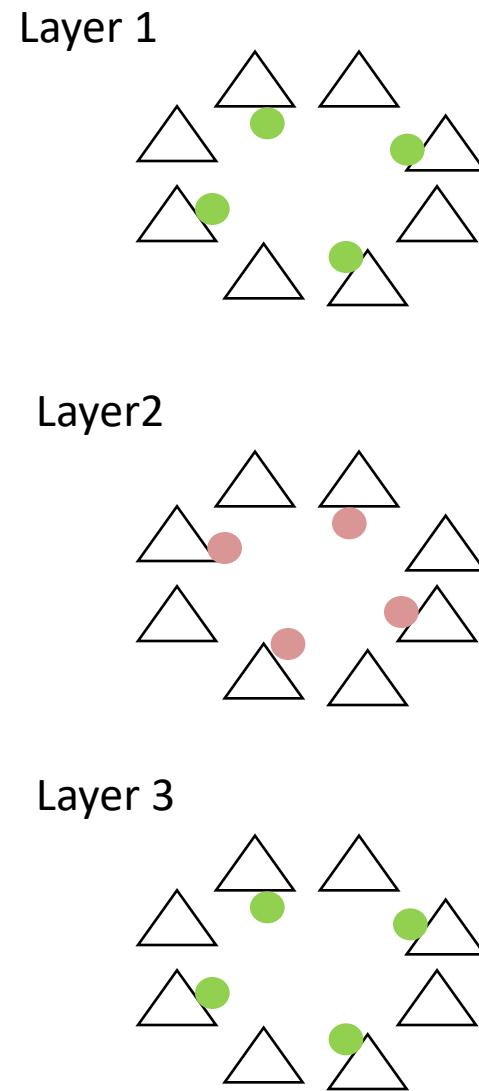
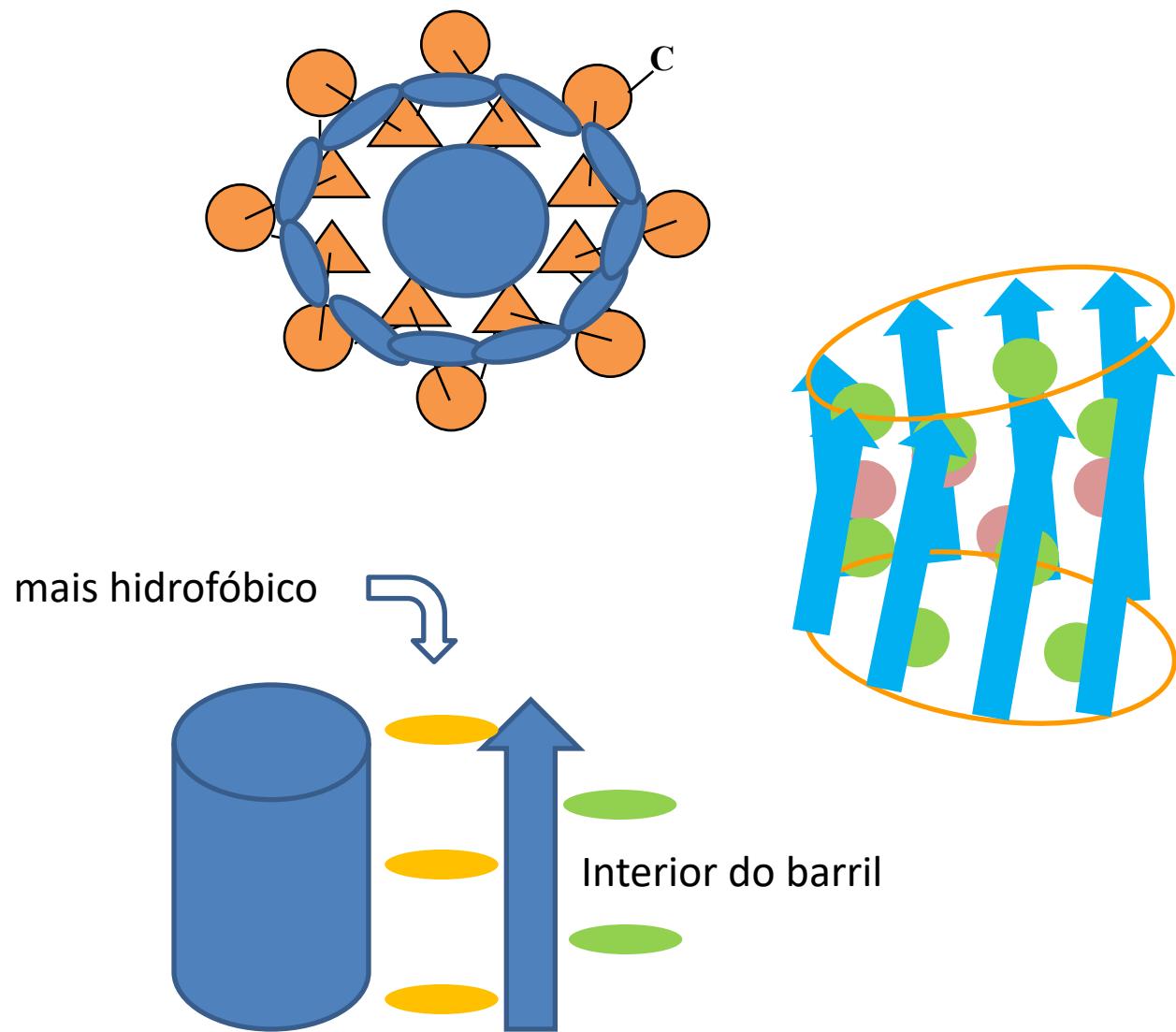
pelo menos famílias diferentes

Localização de sítios ativos

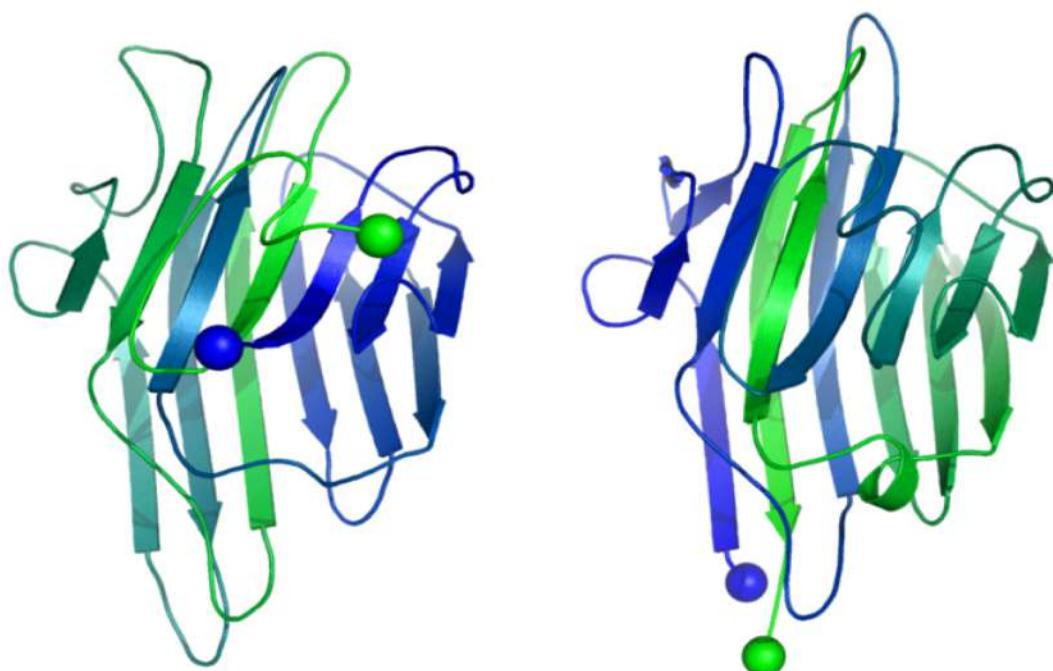


The active site is always at
the C-terminal end of the
barrel

There are two ways of packing sidechains inside the barrel

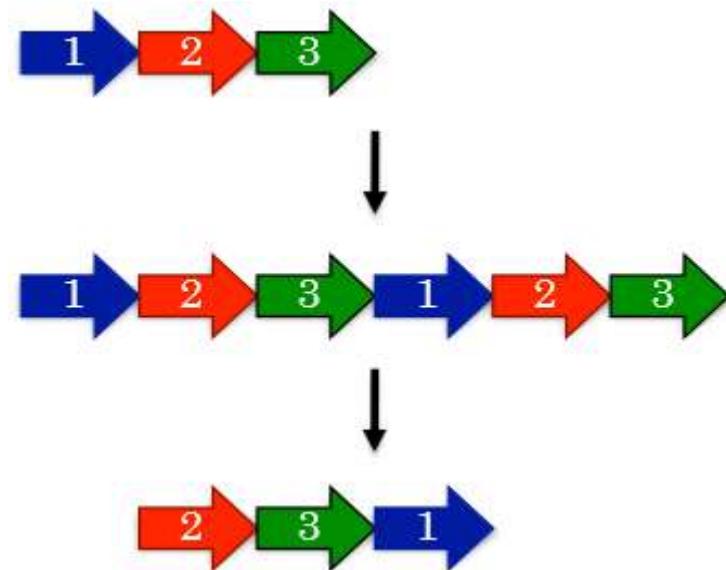


Different packings may occur via cyclic permutation



Folds where the N- and C-termini
are spatially close

May arise from gene duplication
and subsequent deletion



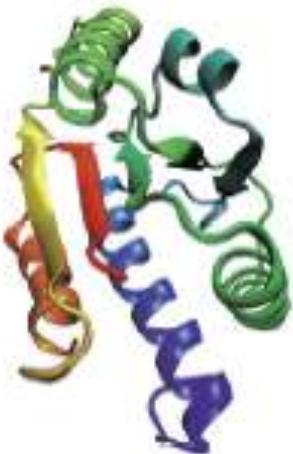
Part 7: Some unusual features

Knots

Knots		
1mxi A00	160 (13)	
3.40.1280	SH(3)S(2)HSH	EAP
3 ₁ Trefoil knot		

Knots		
1xd3 A00	217 (9)	
3.40.532	HS(2)H(5)S(3)HS	EAP
4 ₁ (Figure-of-eight) knot		

Knots		
1yve I02	152 (76)	
1.10.1040	H(22)	EAP
5 ₂ Knot		



Trefoil knot (3_1)

Methyl transferase from *E. coli*

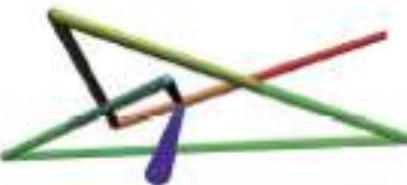
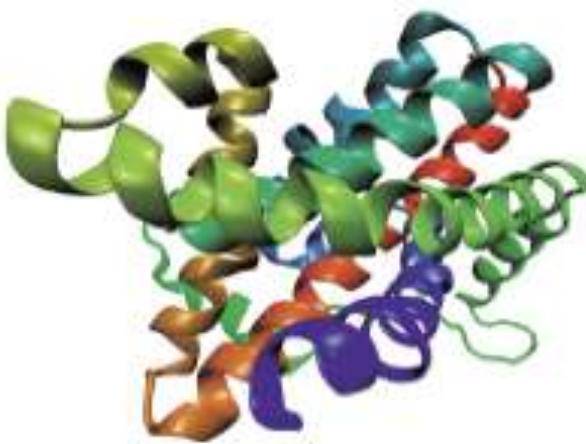
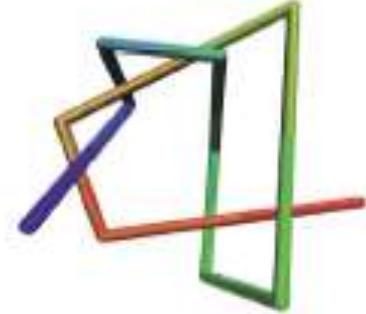


Figure-of-eight knot (4_1)

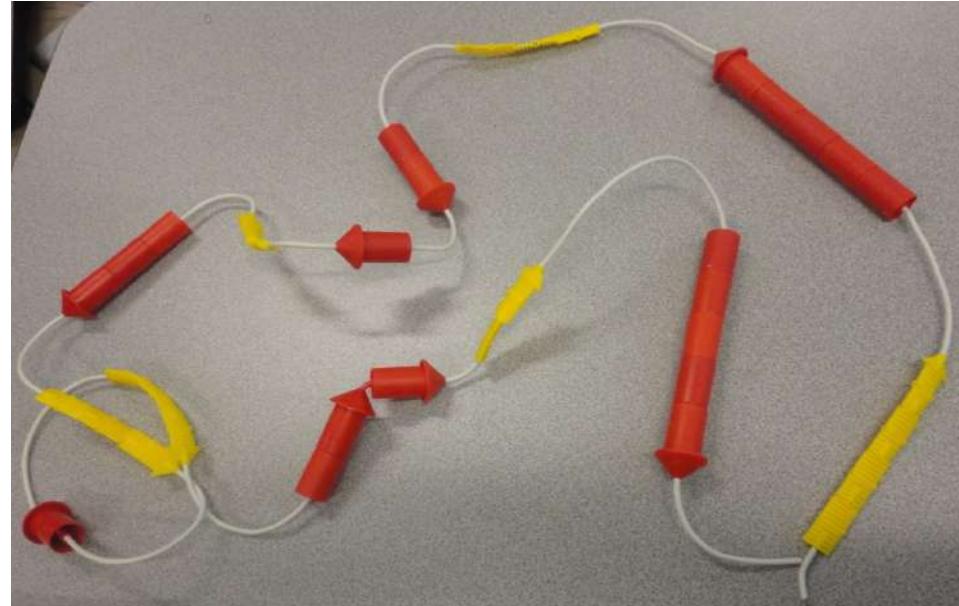
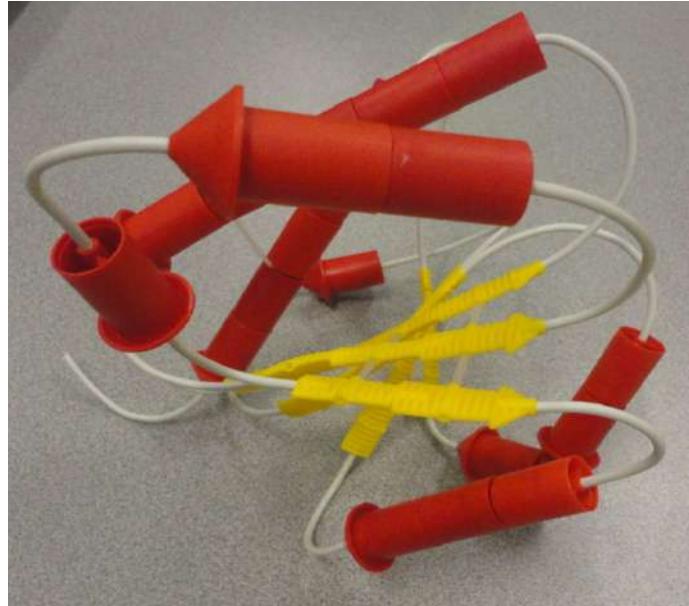
Class II keto-acid reductoisomerase
from *Spinacia oleracea*



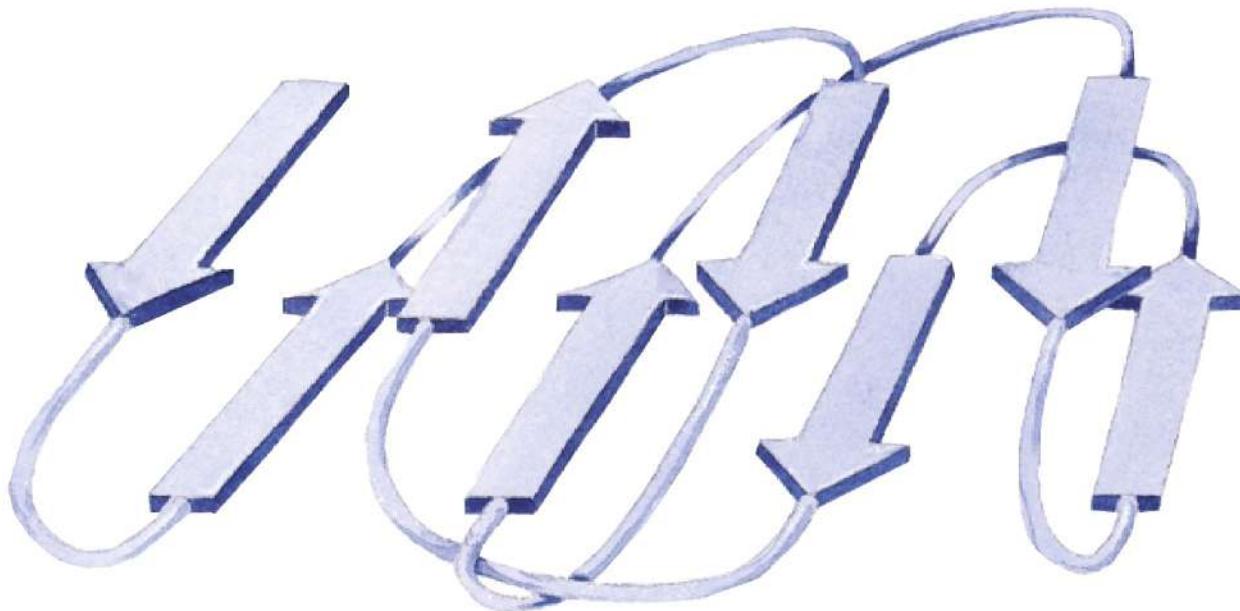
5_2 knot

Ubiquitin hydrolase UCH-3L

Knots



Knots are rare and corssed conetcions also



Crossover connection

3.17

Topologias permitidas e não permitidas para folhas- β
Allowed and Disallowed β -sheet Topologies



1. A D E B C F
2. B A F C D E
3. B E D C F A



1

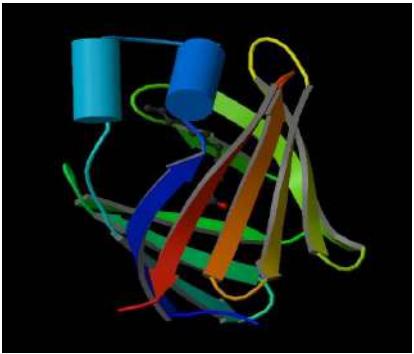


2

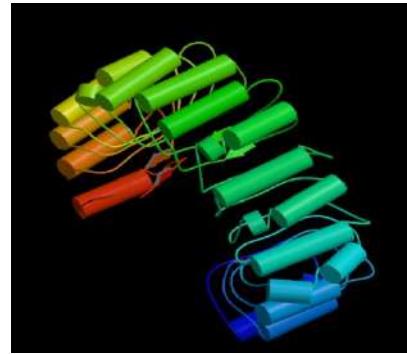


3

Matematically sepaking...



Fatty acid binding protein



$\alpha\beta$ Horseshoe



N-terminus

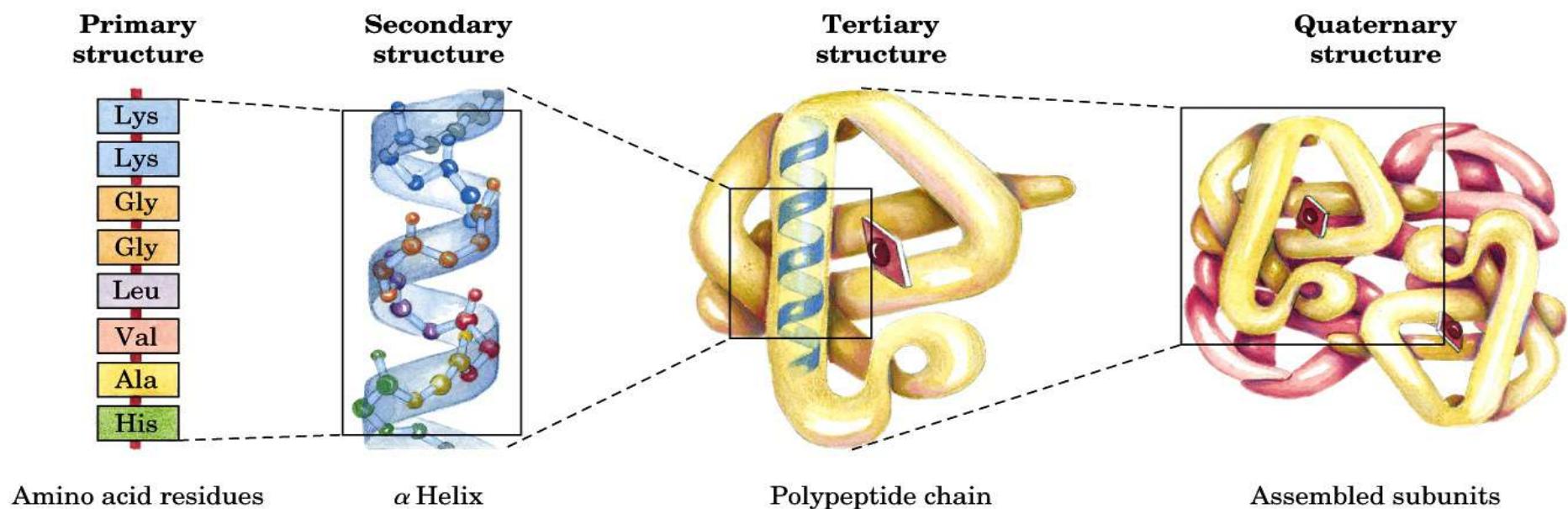


C-terminus

...all proteins have the same topology and there are no knots!

Part 8: Quaternary Structure

Protein Hierarchy

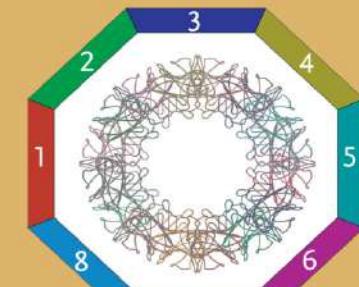


Legend

PDB code for reference structure
▼

Number of subunits
▼

1qtj 16



D₈ 822
SAP from horseshoe crab

▲ Protein name

Point group symmetry (Schoenflies nomenclature)

▲ Protein name

Point group symmetry (International nomenclature)

1 Multiple sites, cross-linking, membrane association

2 Cooperativity/allosterism

3 Cavities, channels and pores

4 Functional (active) site formation

5 Size and stability

6 Economy of genetic material

7 "Rulers" (exact separation between binding sites)

8 Multiple functions (in hetero-oligomers)

Yellow background indicates dihedral symmetry

Gray background indicates cubic/icosahedral symmetry

WILEY-VCH

Oligomeric Proteins

Highest Order Rotation Axis



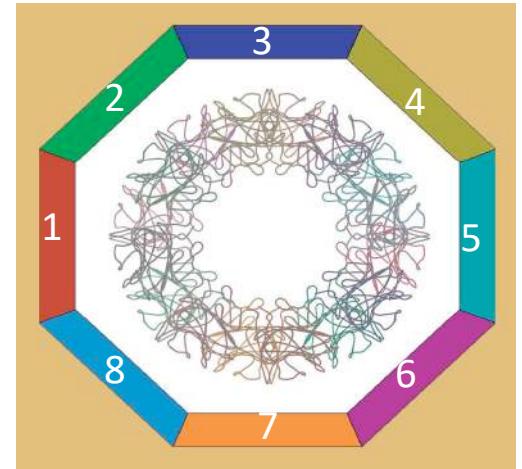
ISBN 978-3-527-31963-3

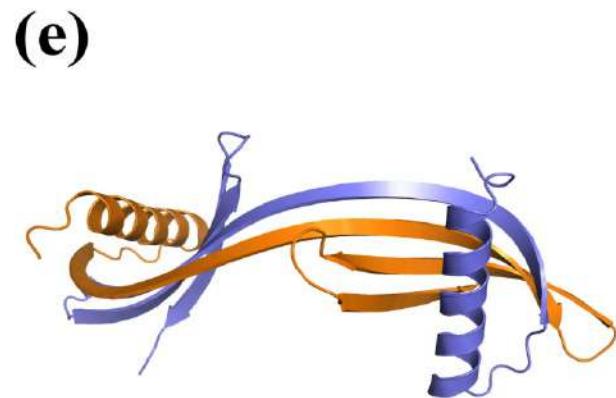
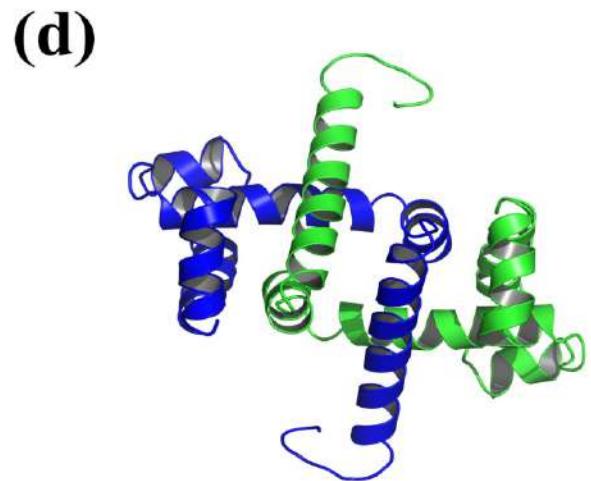
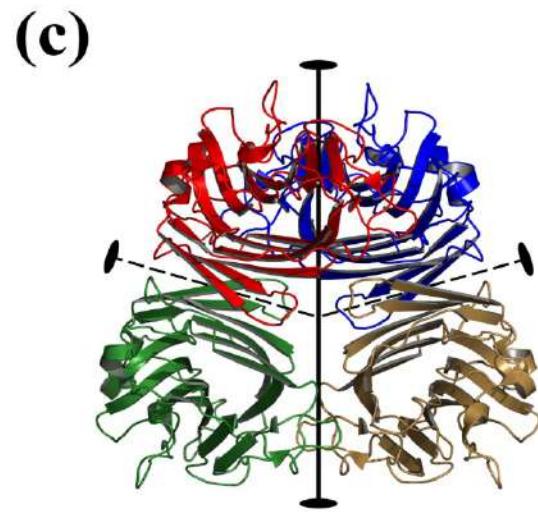
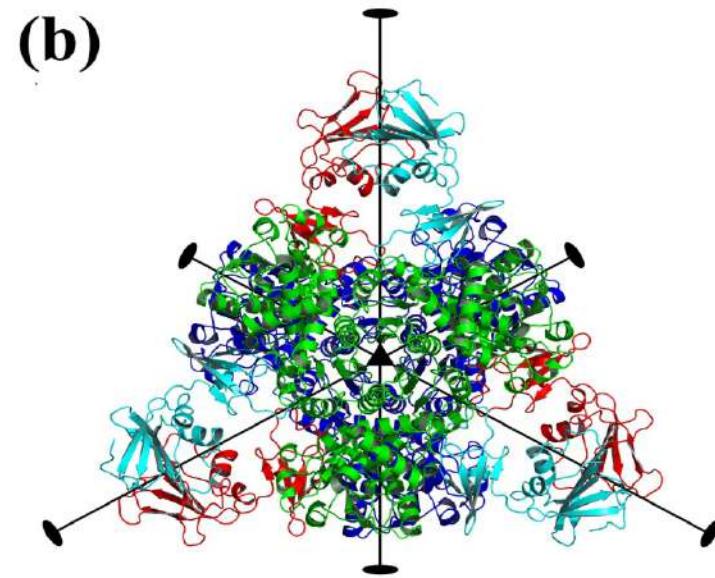
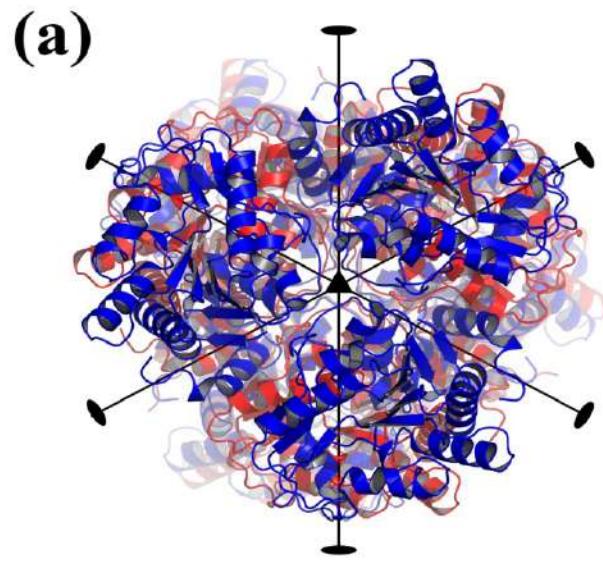


9 783527 319633

Why be oligomeric?

- 1) Multiple sites, cross-links, membrane association
- 2) Cooperativity/Alosterism
- 3) Cavities, channels and pores
- 4) Formation of functional sites (ative sites)
- 5) Size and Stability
- 6) Economy of genetic material
- 7) “Rulers” - exact eperation between active sites
- 8) Multiple functions (in the case of hetero-oligomers)



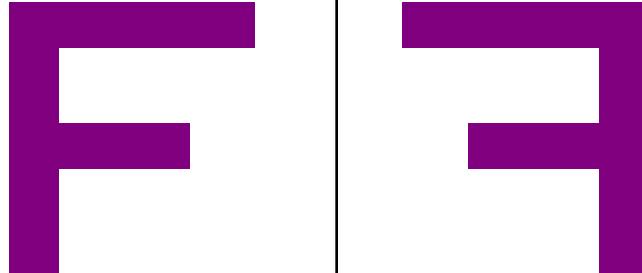


Distribuição de oligomeros em *E. coli*

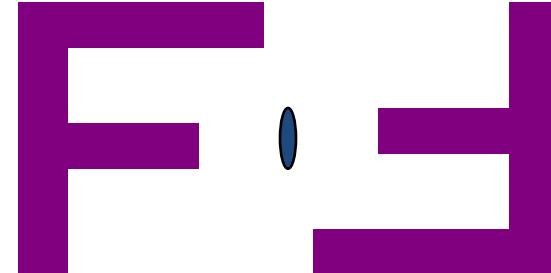
# subunits	% protenas
1	19.4
2	38.2
3	5.4
4	21.0
5	0.1
6	5.6
7	0.1
8	2.4
9	0.0
10	0.0
11	0.0
12	1.6
>12	2.2
Polimers	2.7

The majority of oligomers present some type of symmetry

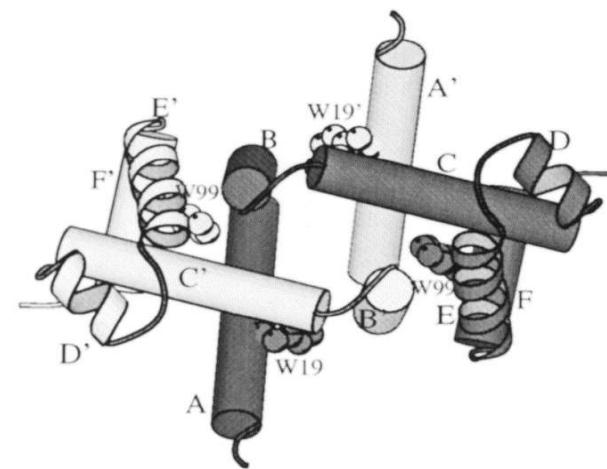
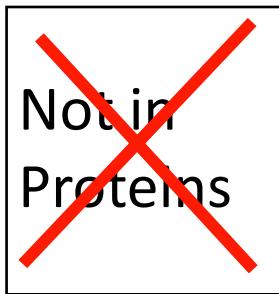
Types of symmetry



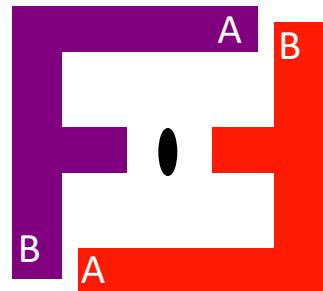
Reflection (mirros)



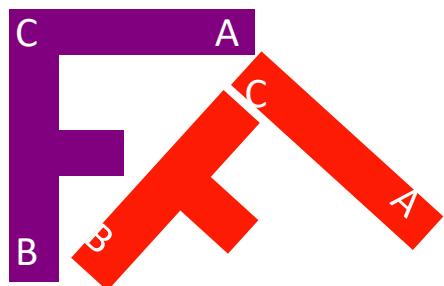
Rotation about an axis



Why be symmetric?



Arranjos simétricos geram
mais contatos favoráveis



The simplest type of symmetry is rotacional symmetry (polar)

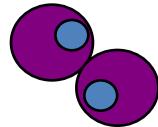


180°

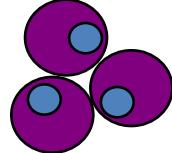
120°

90°

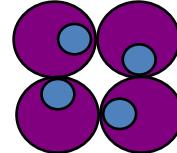
etc. etc.



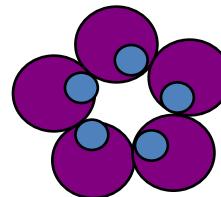
2-fold



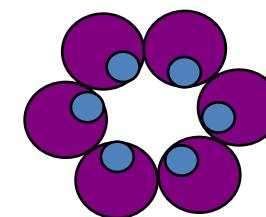
3-fold



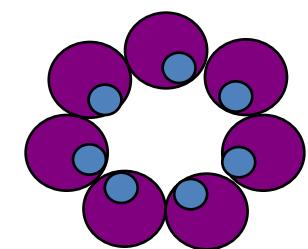
4-fold



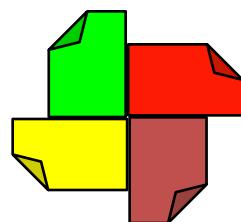
5-fold



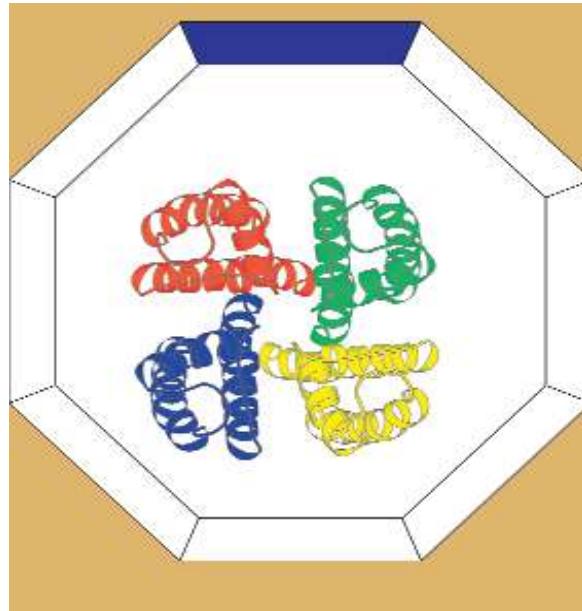
6-fold



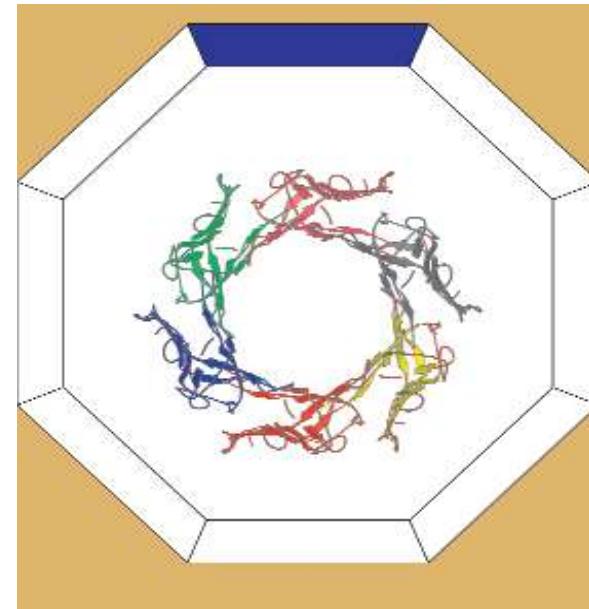
7-fold



An example - channels

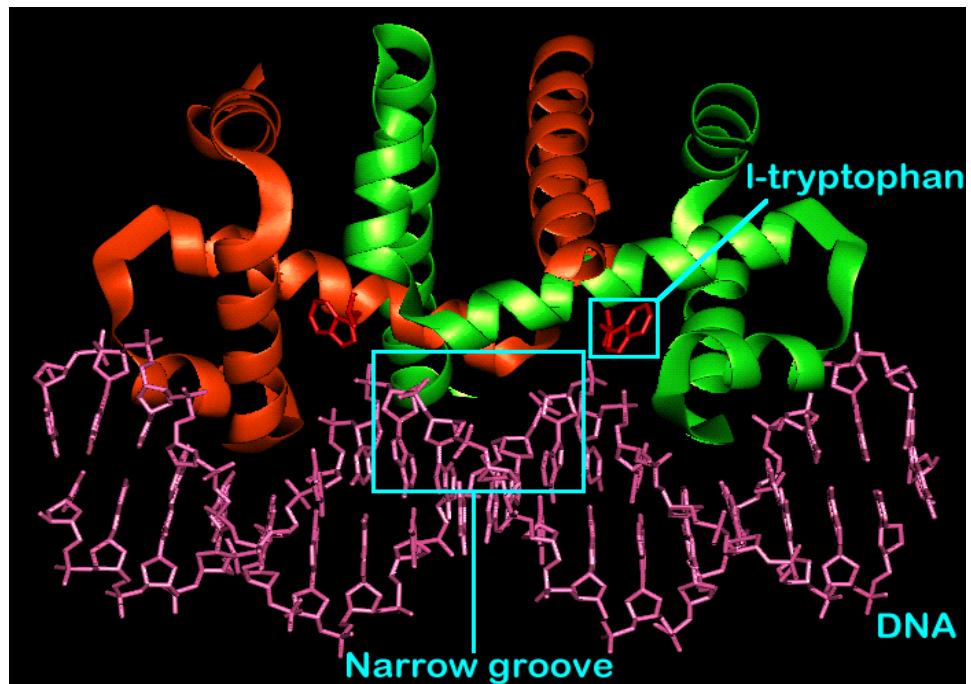
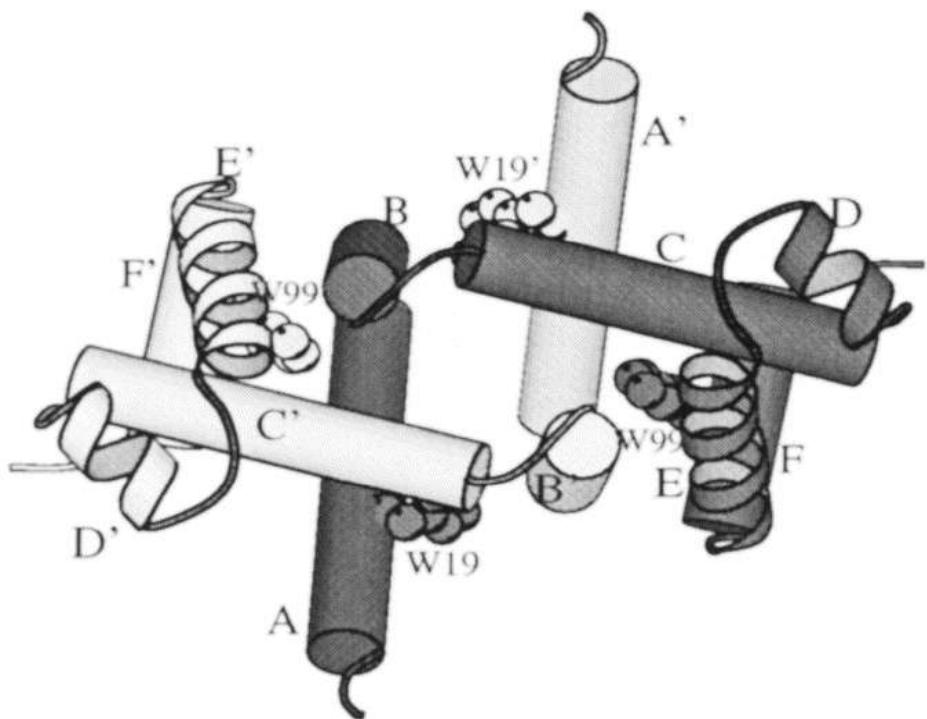


K^+ channel

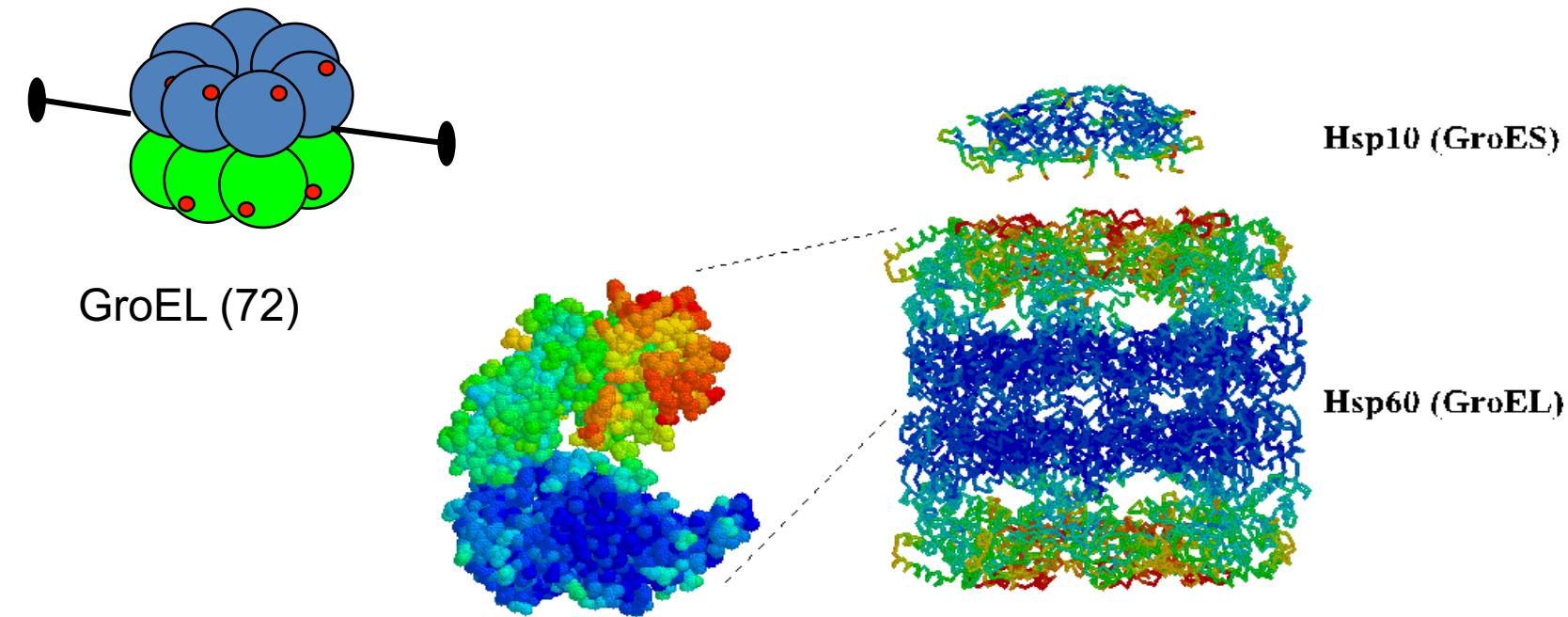


Protein secretion apparatus

Trp repressor

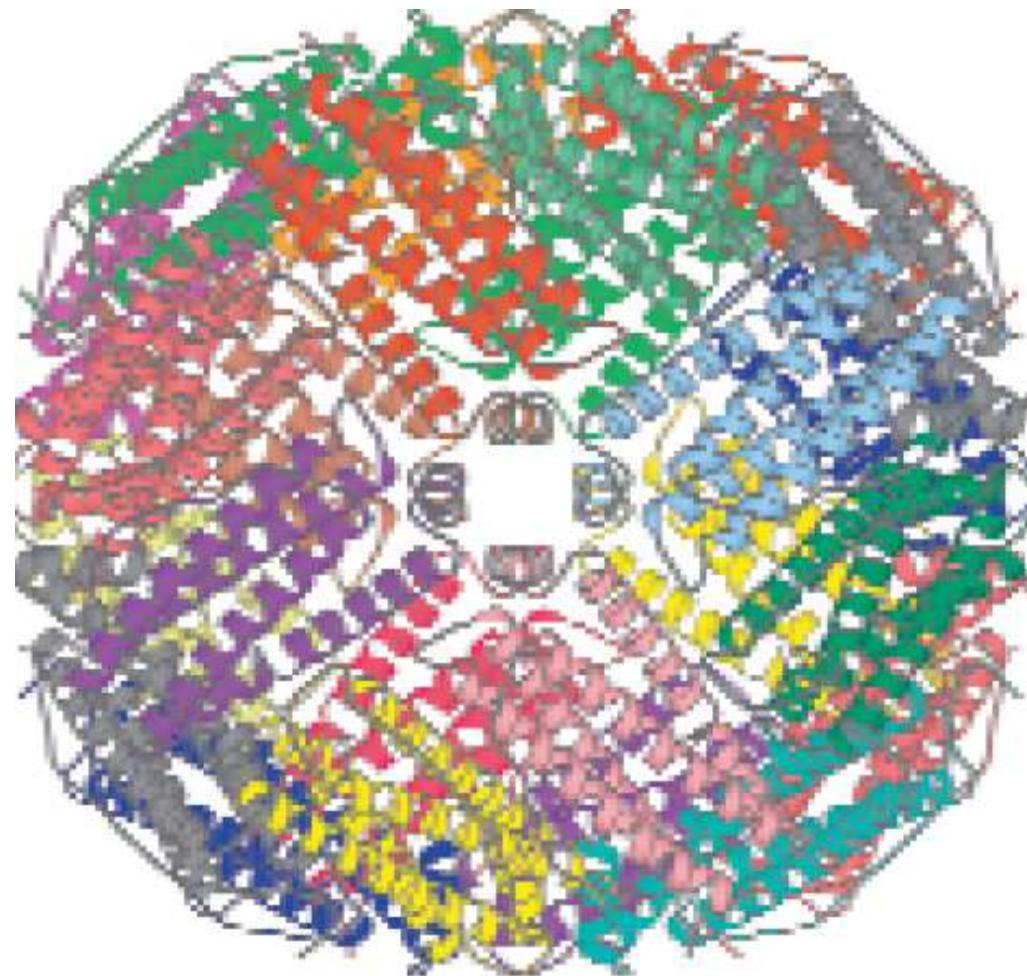


Dihedral symmetry

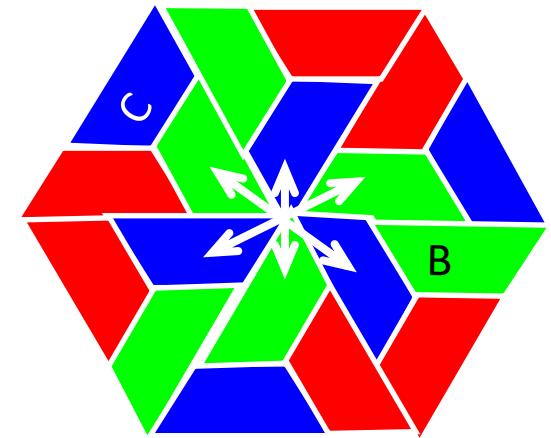
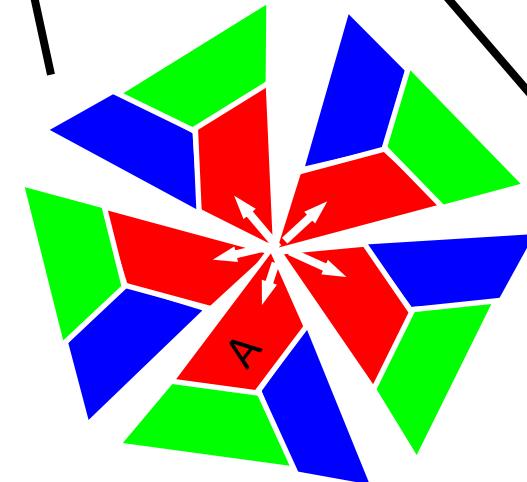
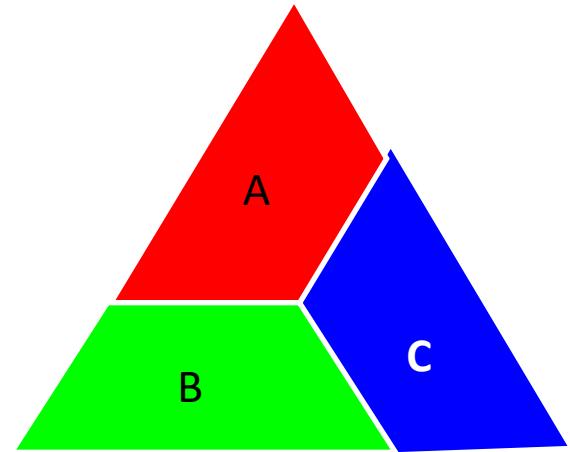
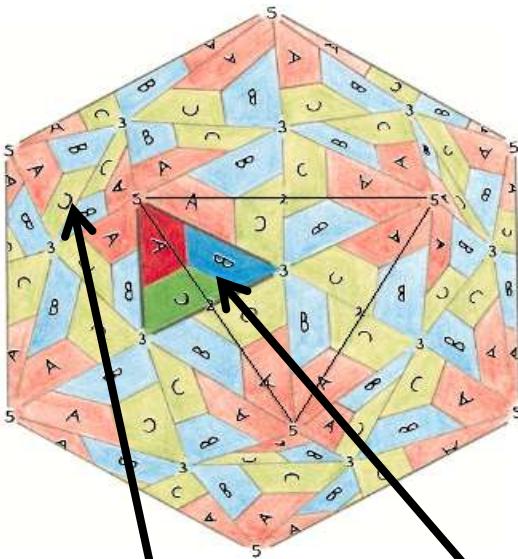
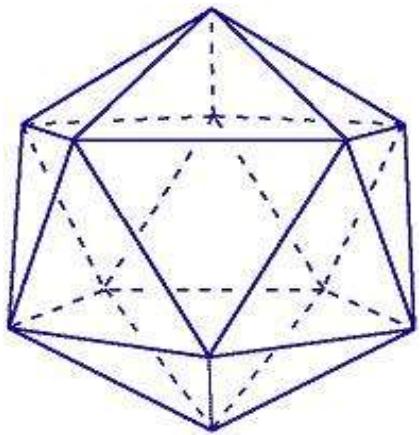


This can result in cylindrical structures with internal cavities

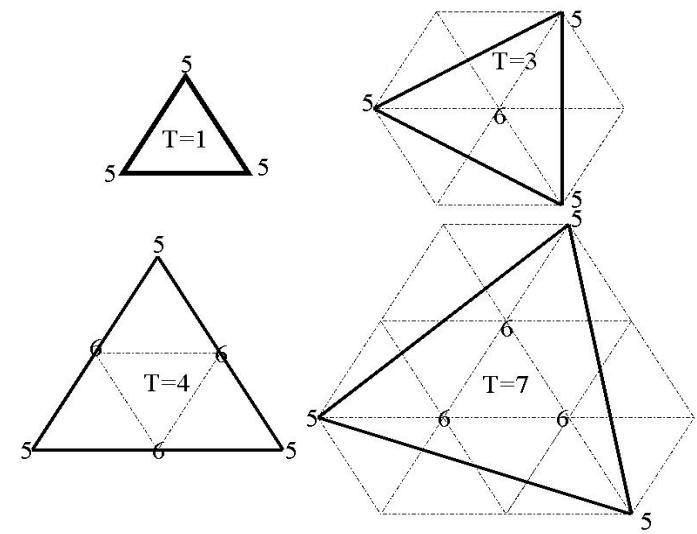
A totally closed cavity can be formed using cubic or icosahedral symmetry?



Ferritin



Número de triangulação

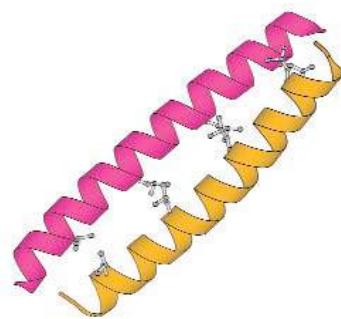


Quase simetria

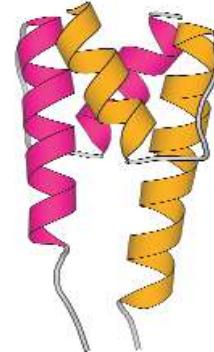
What maintains the oligomeric structure?

All types of non-covalente or even covante (in the case od disulphide bridges)

Some commun motifs:



Leucine zipper



helix-turn-helix



TRAP

Both are used as dimerization motifs

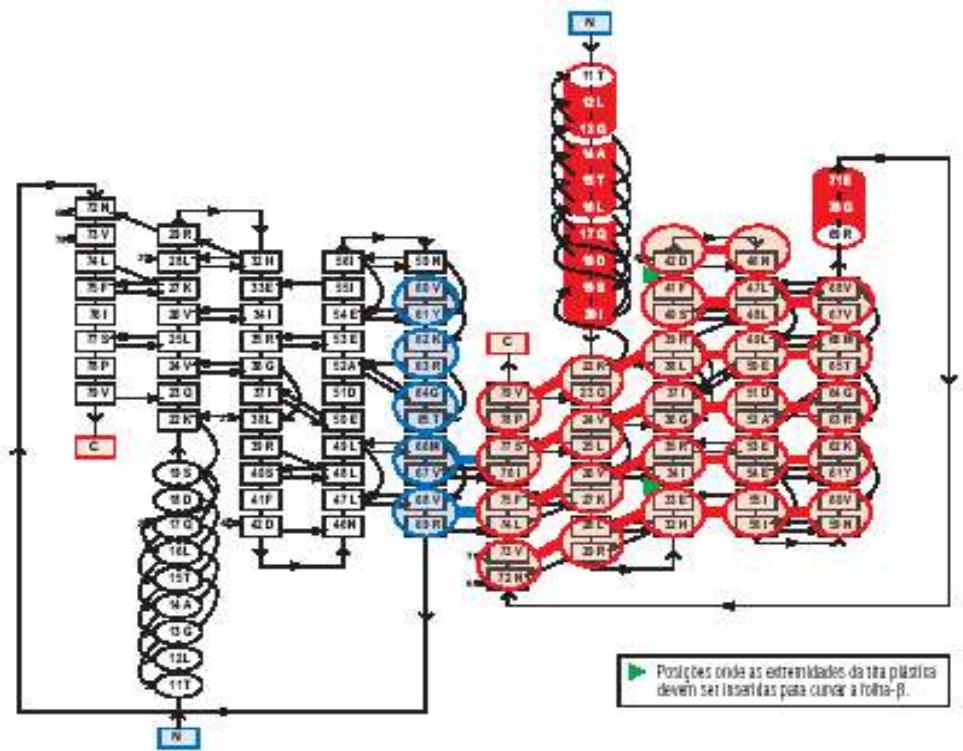
Uses β -strands from adjacent subunits to form a continuous sheet

Part 9: Final exercise

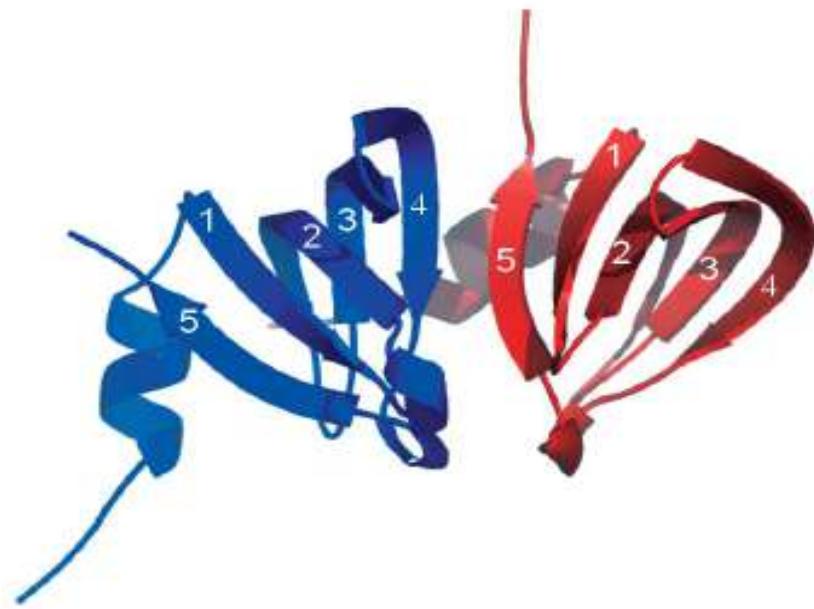
3.11

SmAP (Sm-like archaeal proteins) - Uma proteína multimérica

SmAP (Sm-like archaeal proteins) – a multimeric protein

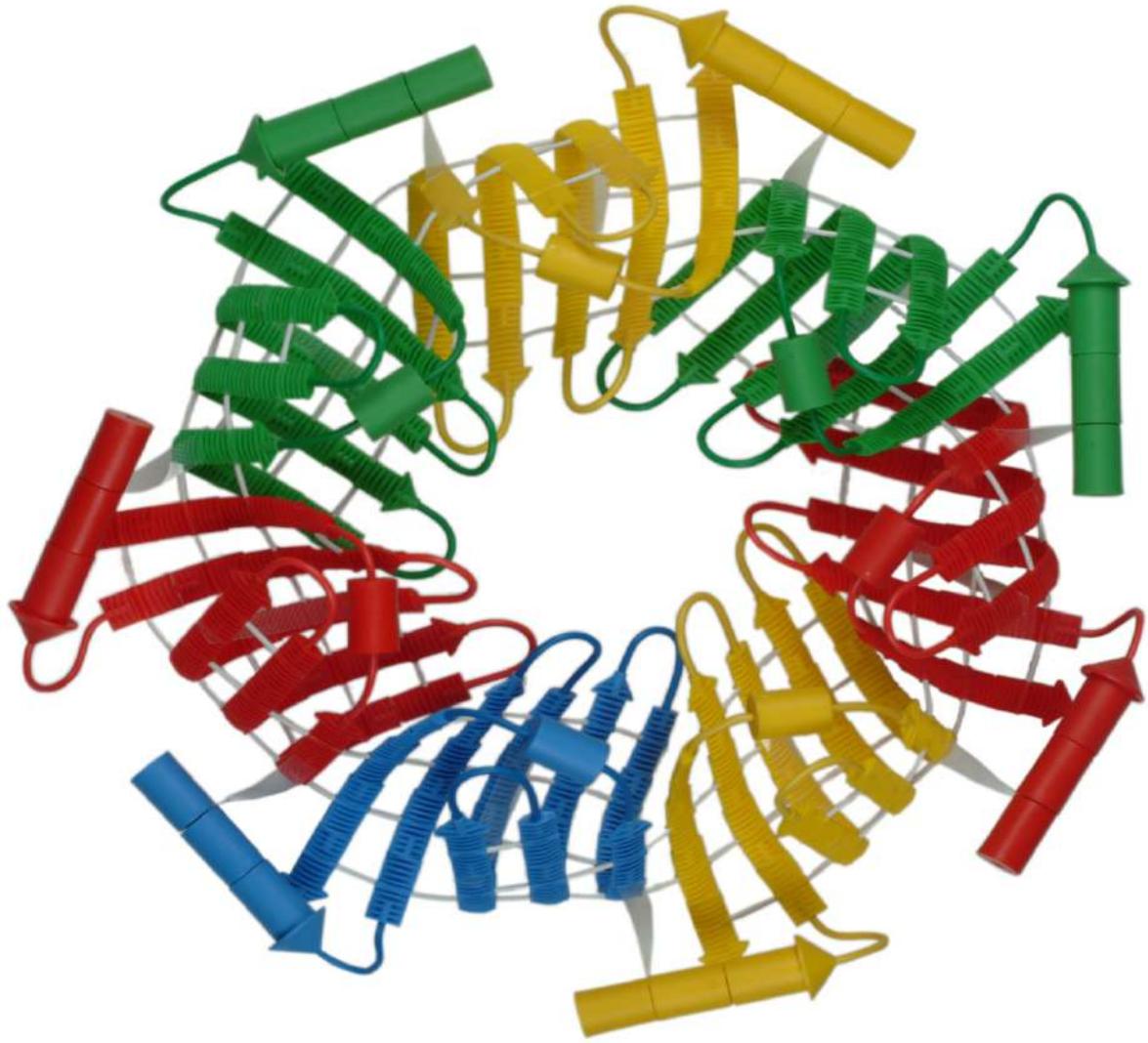


(a)



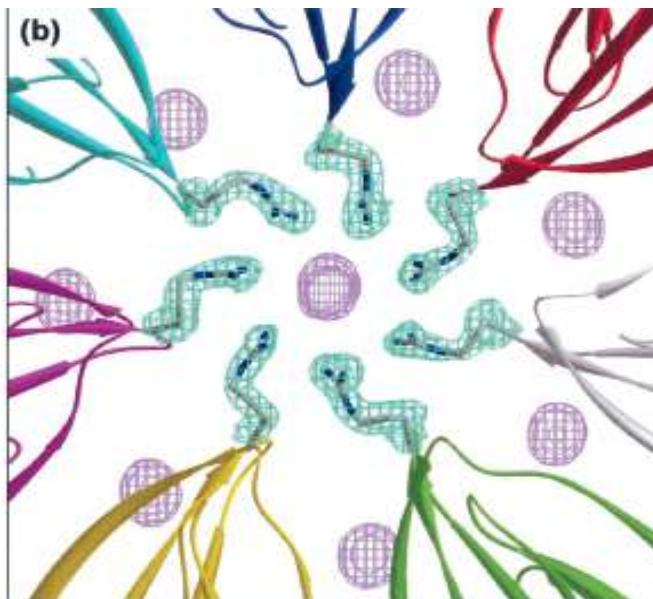
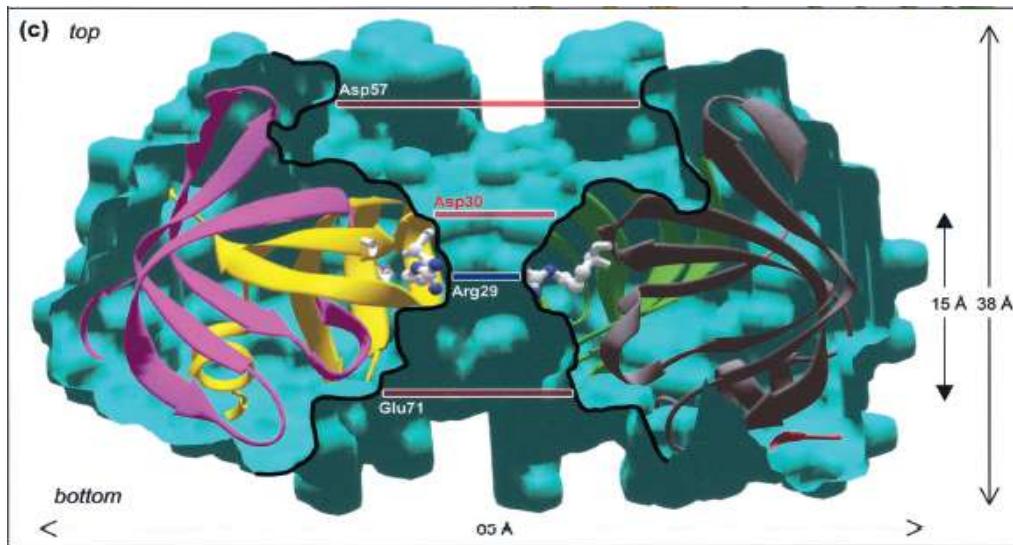
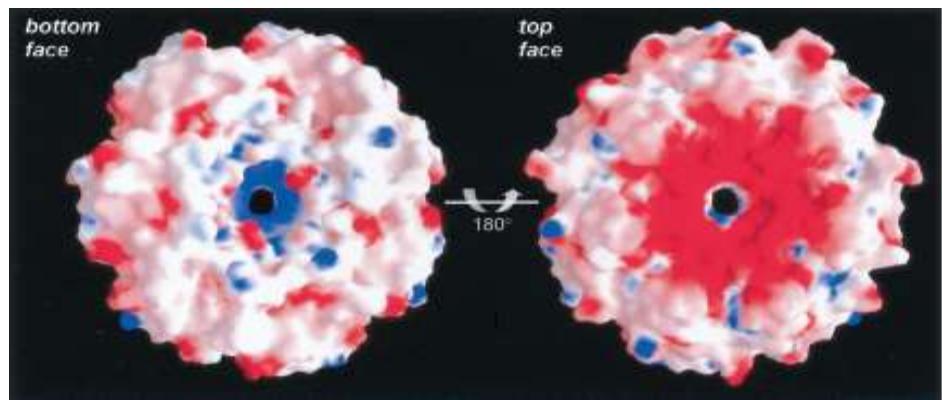
(b)

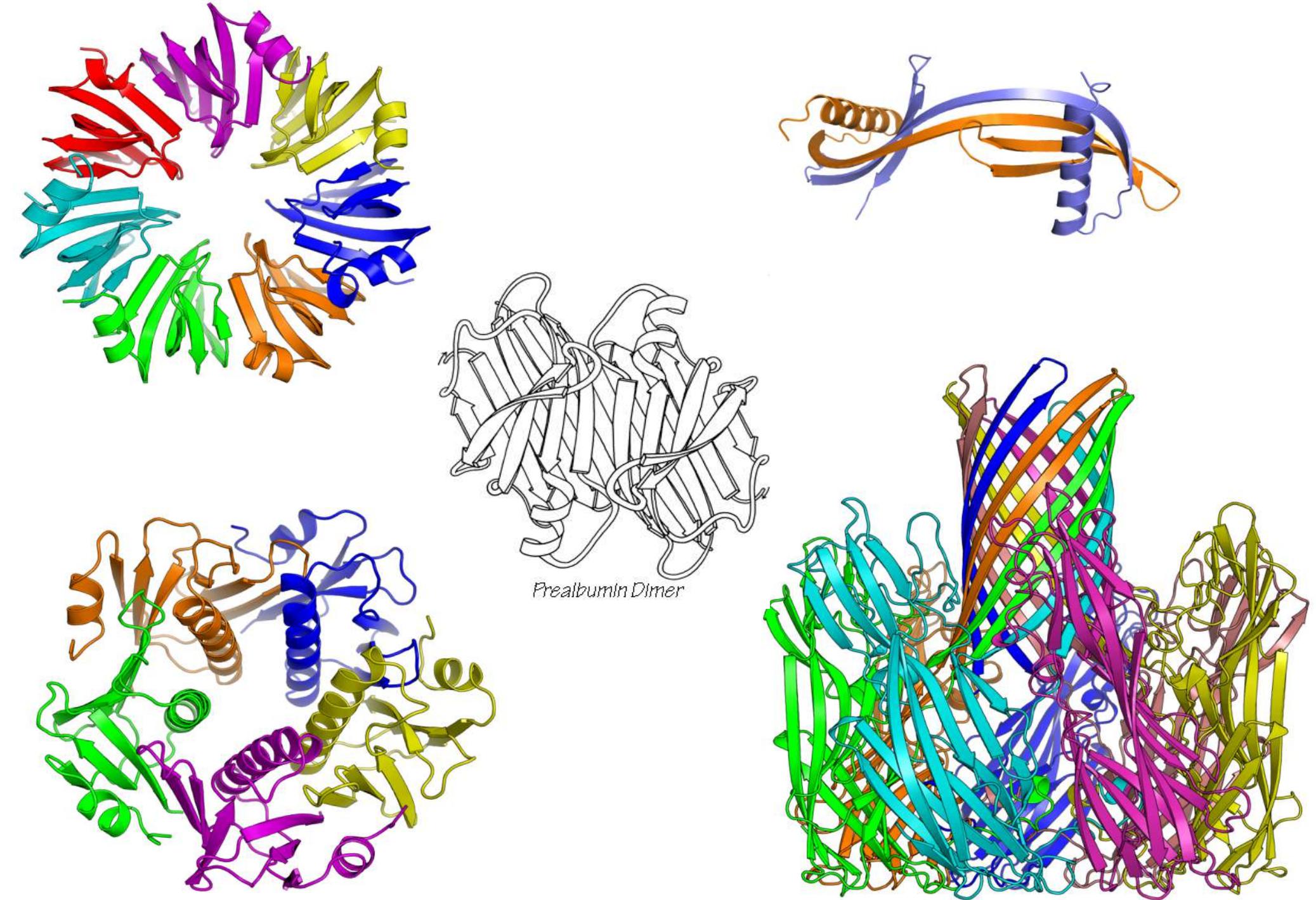
Painel 11a / 11b
Panel 11a / 11b



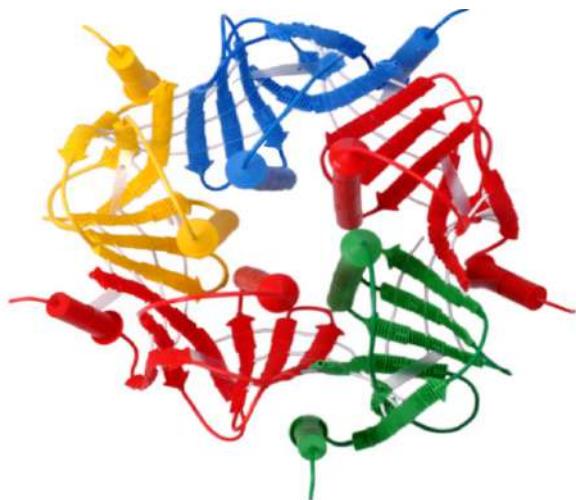
The SmAP structure

- The high degree of bending is unusual. If the β 2 strand was not bent (but was fully extended) the distance between the N- and C-termini would be 39.7 \AA , but in fact is 23.7 \AA .
- The bending seems to require conserved Gly. For example in the β 2 strand it is Gly36.
- The final heptamer has one single 35-strand β -sheet.
- There is an asymmetric distribution of charge on the two faces of the disk indicating the way in which snRNA may bind to the SmAP complex.
- A single Arg residue (Arg29) is responsible for the constriction of the hour-glass like pore and the concentration of positive charge.



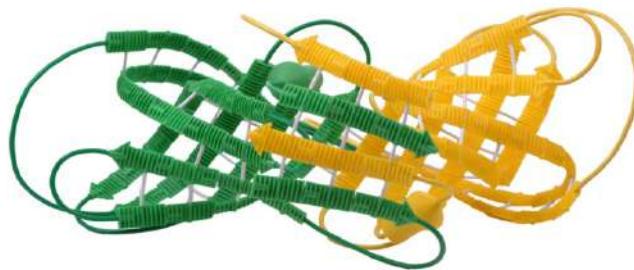


Proteínas oligoméricas : exemplos



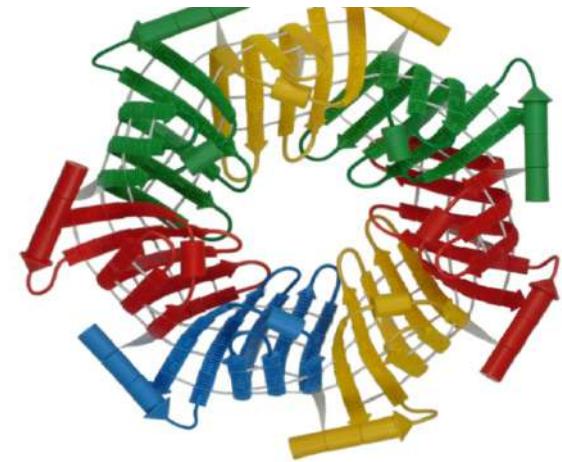
E. coli Heat Labile
Enterotoxin

Pentamer



HIV Protease

Dimer

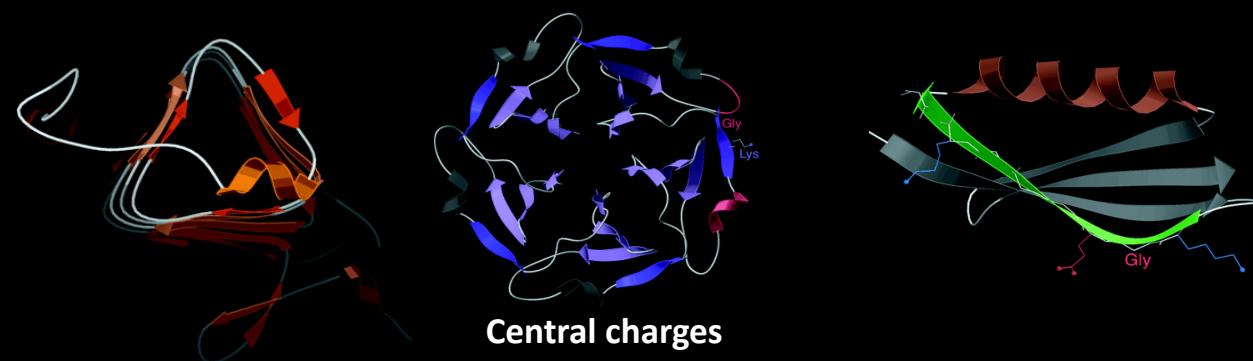
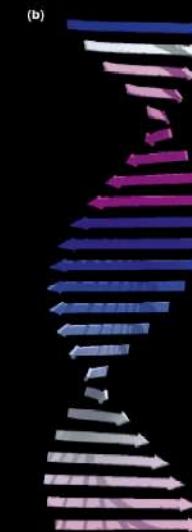
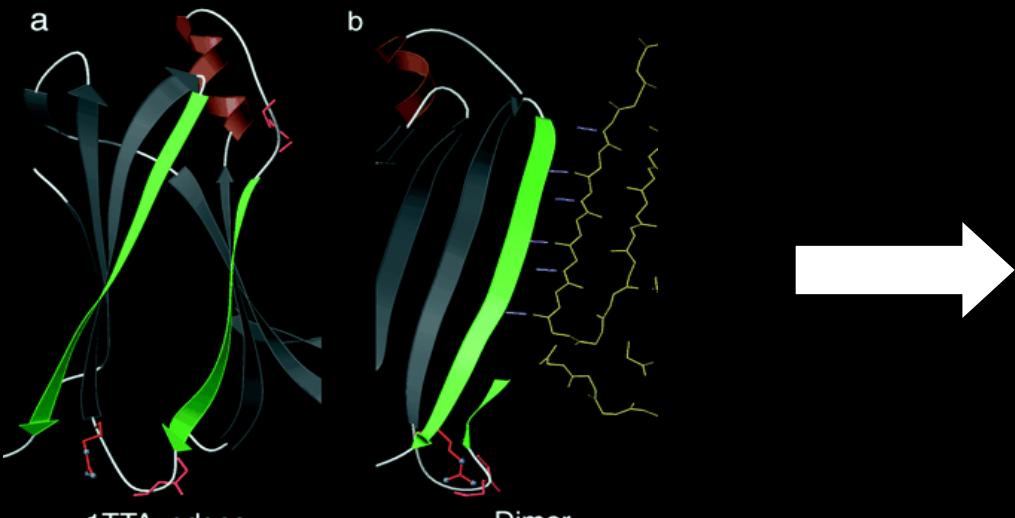


SmAP from *P. aerophilum*

Heptamer

How to avoid β -sheet aggregation and why?

Como evitar agregação

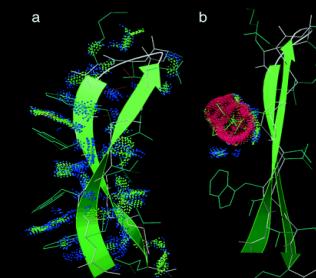
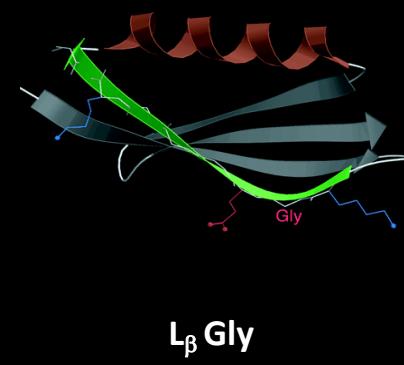


Loop Coverage

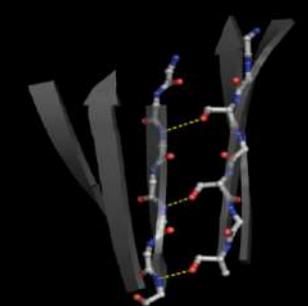
β -Bulges

Prolines

Short edge strands



Strand coil



Serine ladder