

# WHAT ARE PROTEINS?

Proteins are large, complex molecules that play many critical roles in the body.

- Proteins are made up of hundreds or thousands of smaller units called amino acids, (protein building blocks) which are attached to one another in long chains.
- There are mainly 20 different types of amino acids that can be combined to make a protein.
- The sequence of amino acids determines each protein's unique three - dimensional (3D) structure and its specific function.

The importance of proteins ?

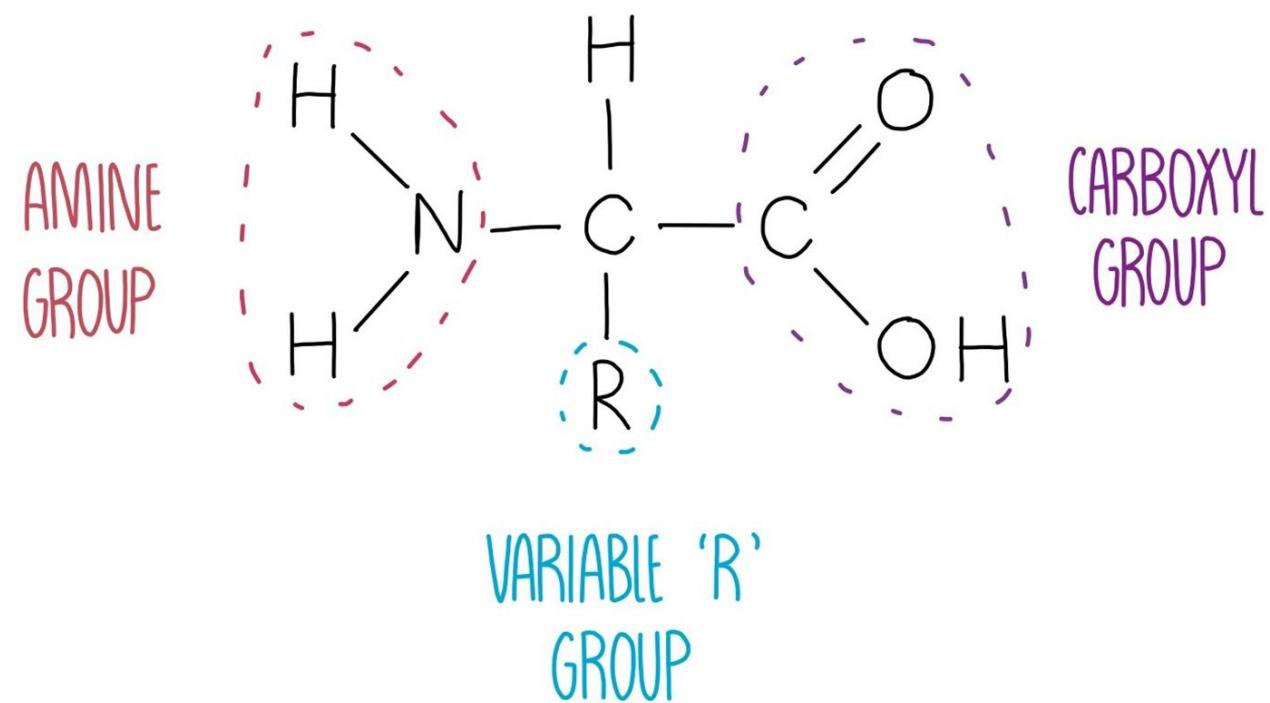
- They do most of the work in cells (structure, function, and regulation of the body's tissues and organs).

Proteins can be described according to their large range of functions in the body e.g antibody, enzyme, messenger, structural component and transport/storage.

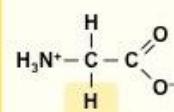
## Examples of protein functions

Function	Description
Antibody	Antibodies bind to specific foreign particles, such as viruses and bacteria, to help protect the body.
Enzyme	Enzymes carry out almost all of the thousands of chemical reactions that take place in cells. They also assist with the formation of new molecules by reading the genetic information stored in DNA.
Messenger	Messenger proteins, such as some types of hormones, transmit signals to coordinate biological processes between different cells, tissues, and organs.
Structural component	These proteins provide structure and support for cells. On a larger scale, they also allow the body to move.
Transport/storage	These proteins bind and carry atoms and small molecules within cells and throughout the body.

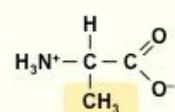
# AMINO ACID



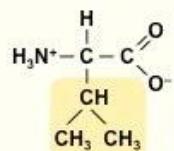
## NON-POLAR



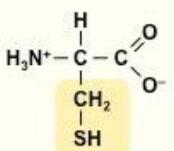
Glycine  
(Gly / G)



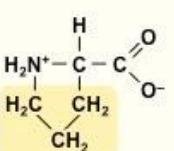
Alanine  
(Ala / A)



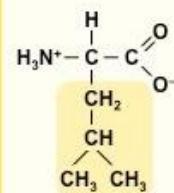
Valine  
(Val / V)



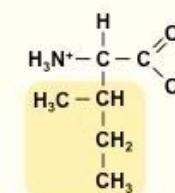
Cysteine  
(Cys / C)



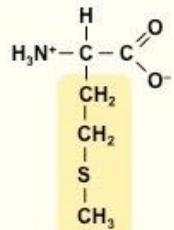
Proline  
(Pro / P)



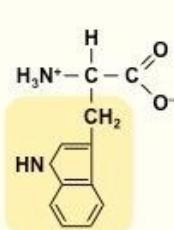
Leucine  
(Leu / L)



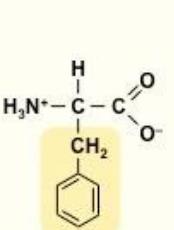
Isoleucine  
(Ile / I)



Methionine  
(Met / M)

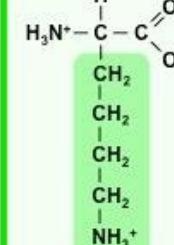


Tryptophan  
(Trp / W)

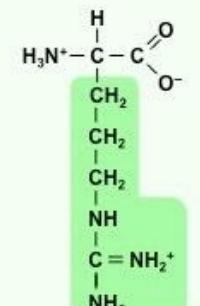


Phenylalanine  
(Phe / F)

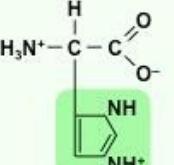
## + CHARGE



Lysine  
(Lys / K)

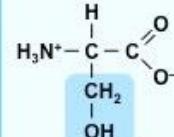


Arginine  
(Arg / R)

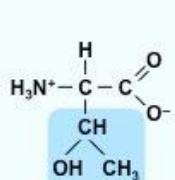


Histidine  
(His / H)

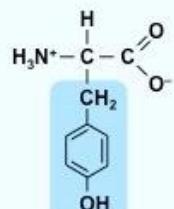
## POLAR



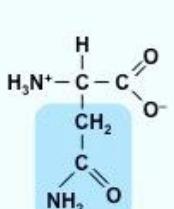
Serine  
(Ser / S)



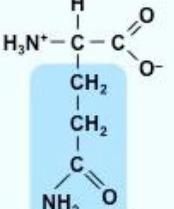
Threonine  
(Thr / T)



Tyrosine  
(Tyr / Y)

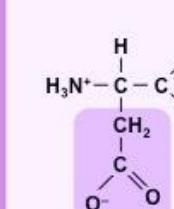


Asparagine  
(Asn / N)

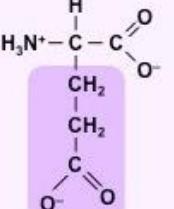


Glutamine  
(Gln / Q)

## - CHARGE



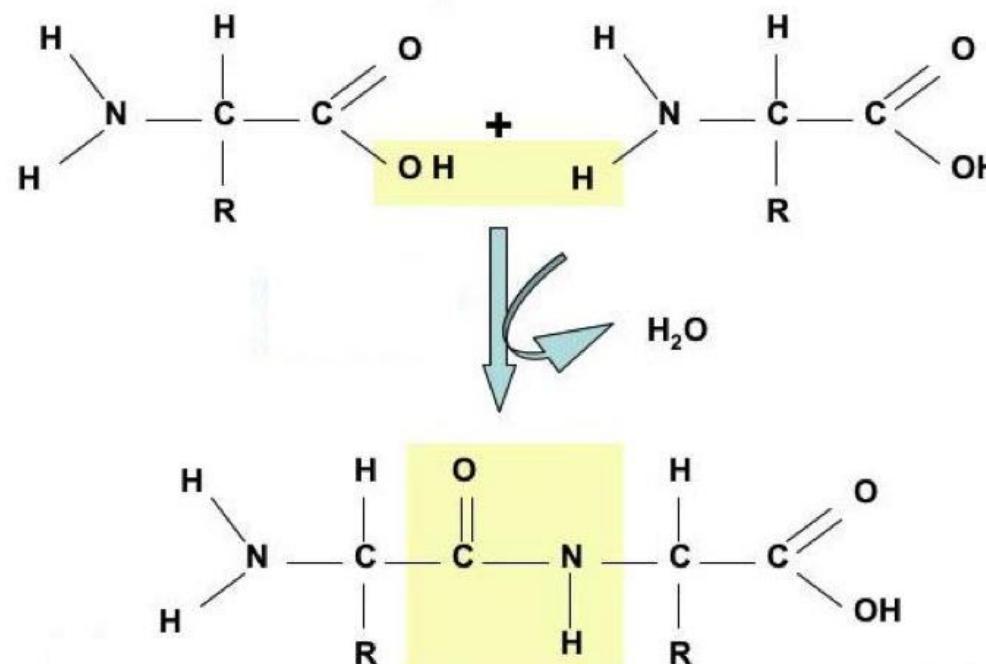
Aspartic Acid  
(Asp / D)



Glutamic Acid  
(Glu / E)

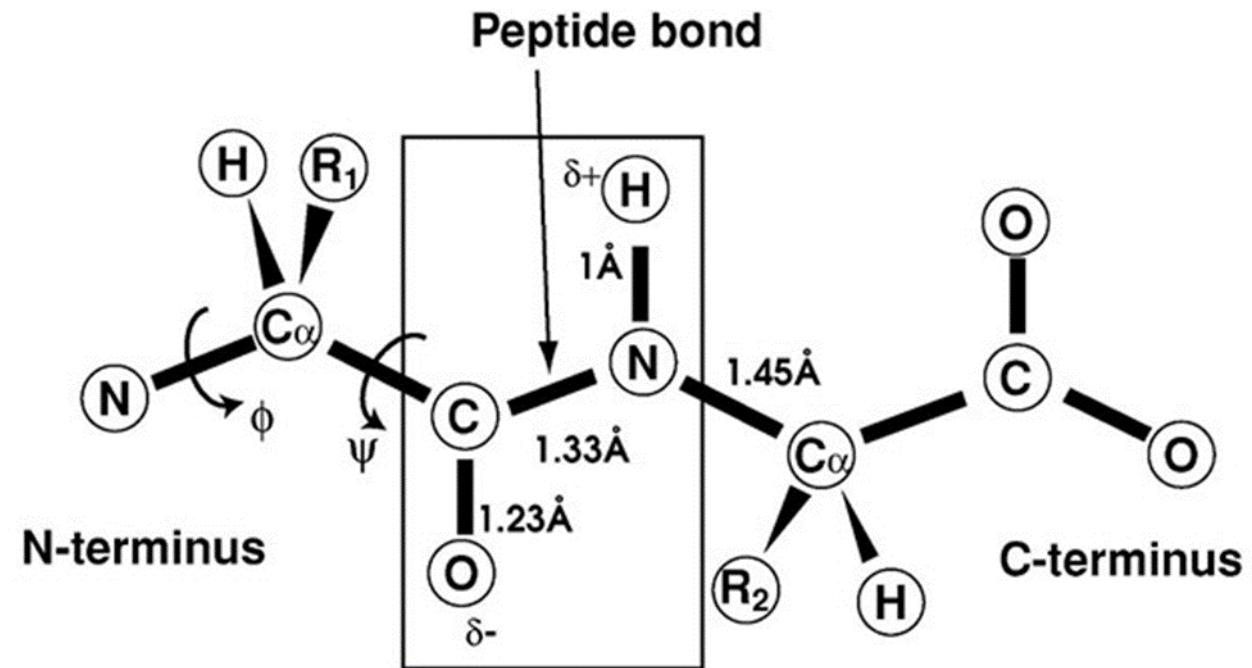
# PEPTID BOND

- A peptide bond is formed between the nitrogen atom of one amino acid and the carbonyl carbon of a second.
- Peptide bonds (covalent) are formed from the amino (N) to the carboxyl (C) terminus by removal of water.
- Peptid bonds can be broken down via hydrolysis reactions, which require water to reverse the process.

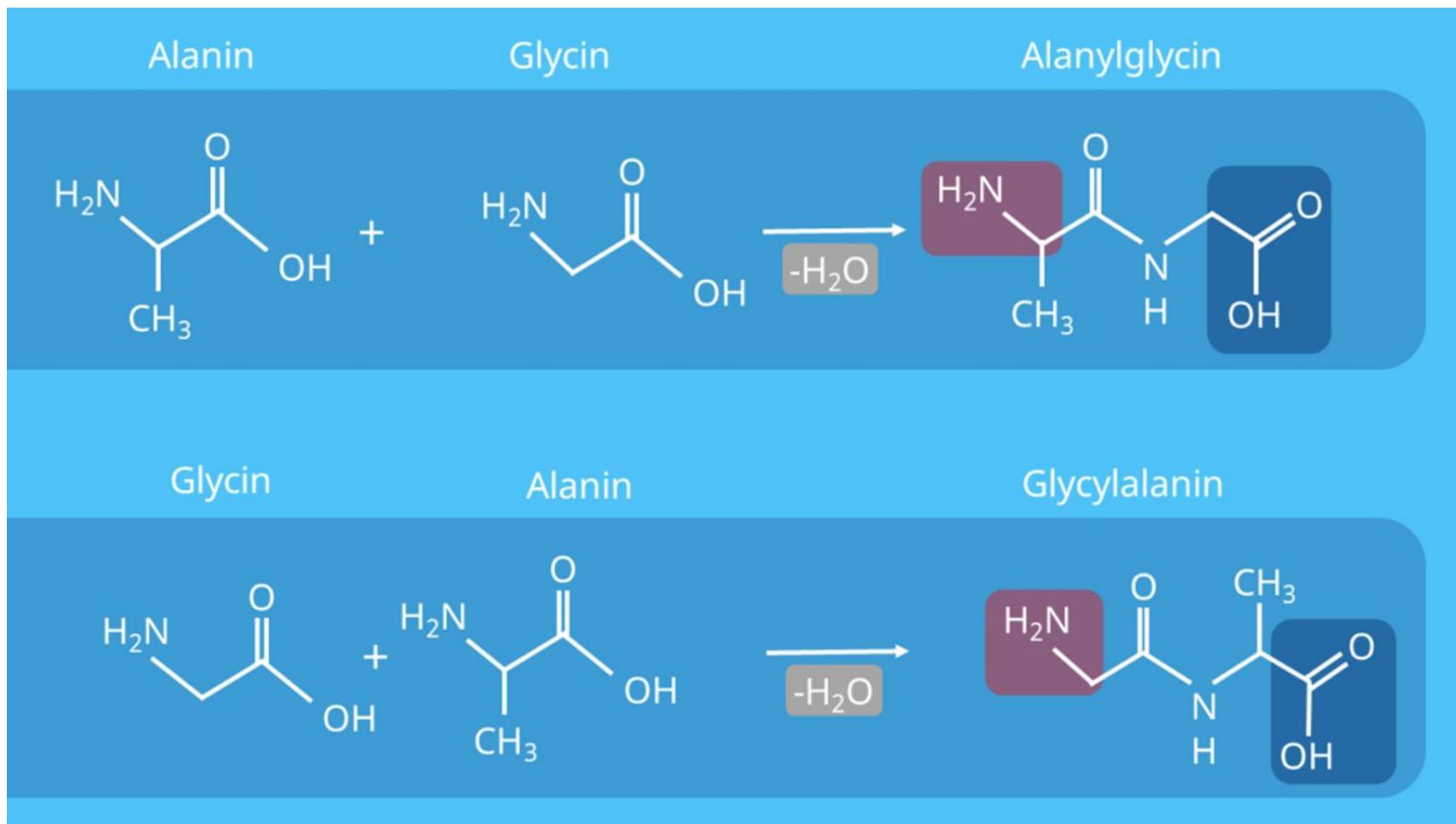


# AMINO ACIDS

- All life is based on bonds between the 20 different amino acids, which all organisms use and modify to their own purpose.
- The number of different combinations is limitless!
- Scientists typically draw and identify proteins starting from the amino, or nitrogen side, going through the carboxyl terminal as a finishing point.

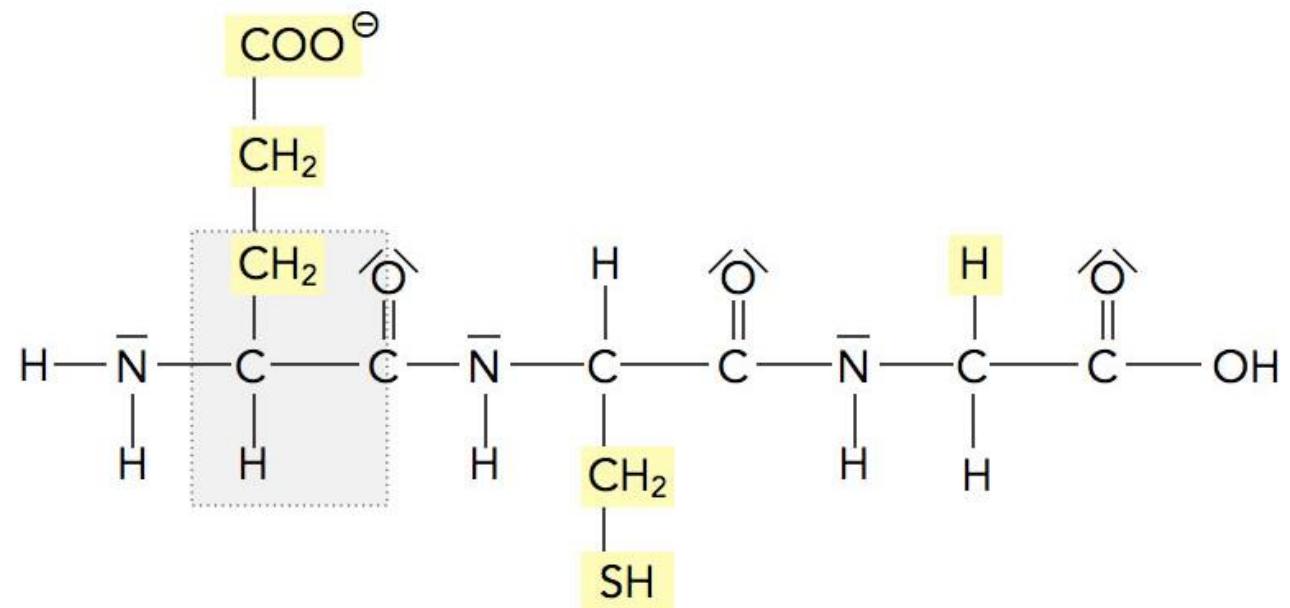


## Dipeptids

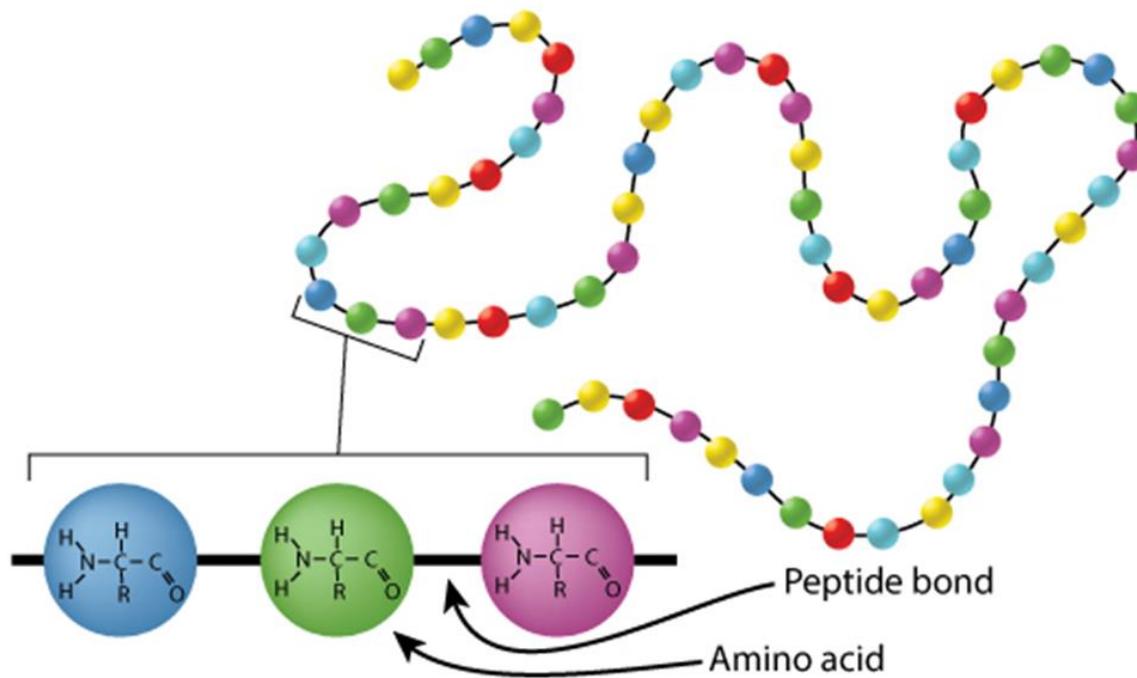


# TRIPEPTIDS

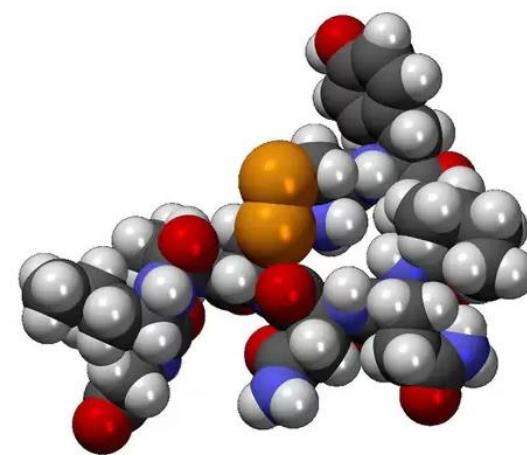
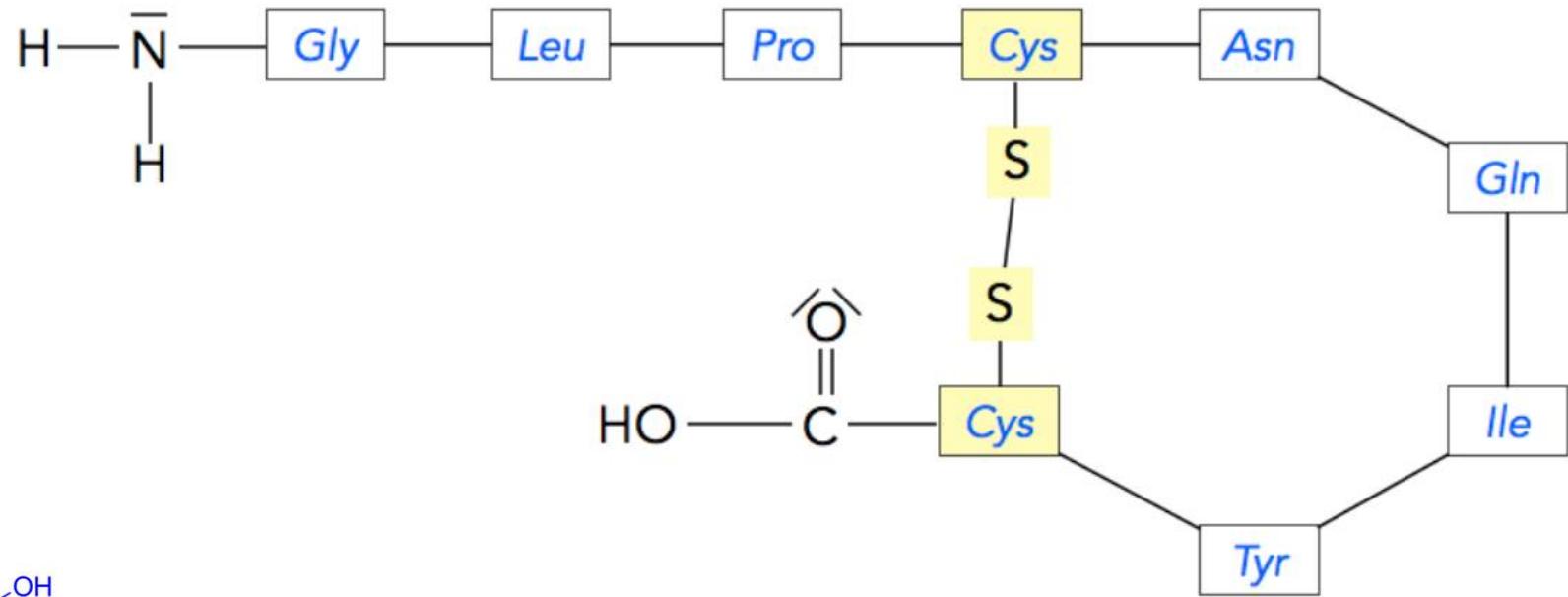
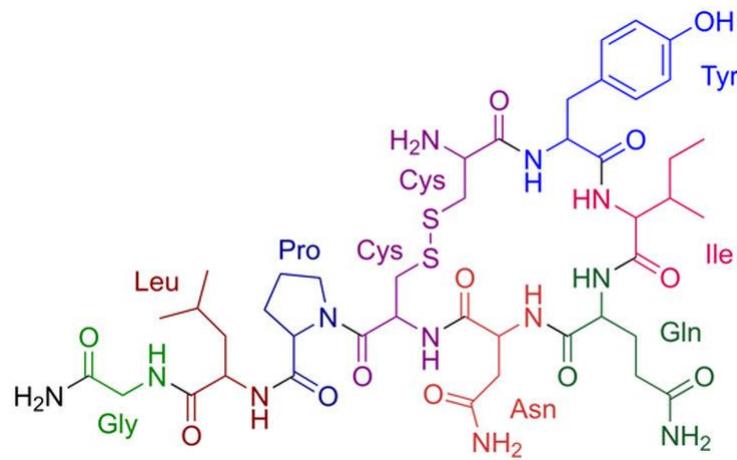
Glutamyl-cysteinyl-glycin

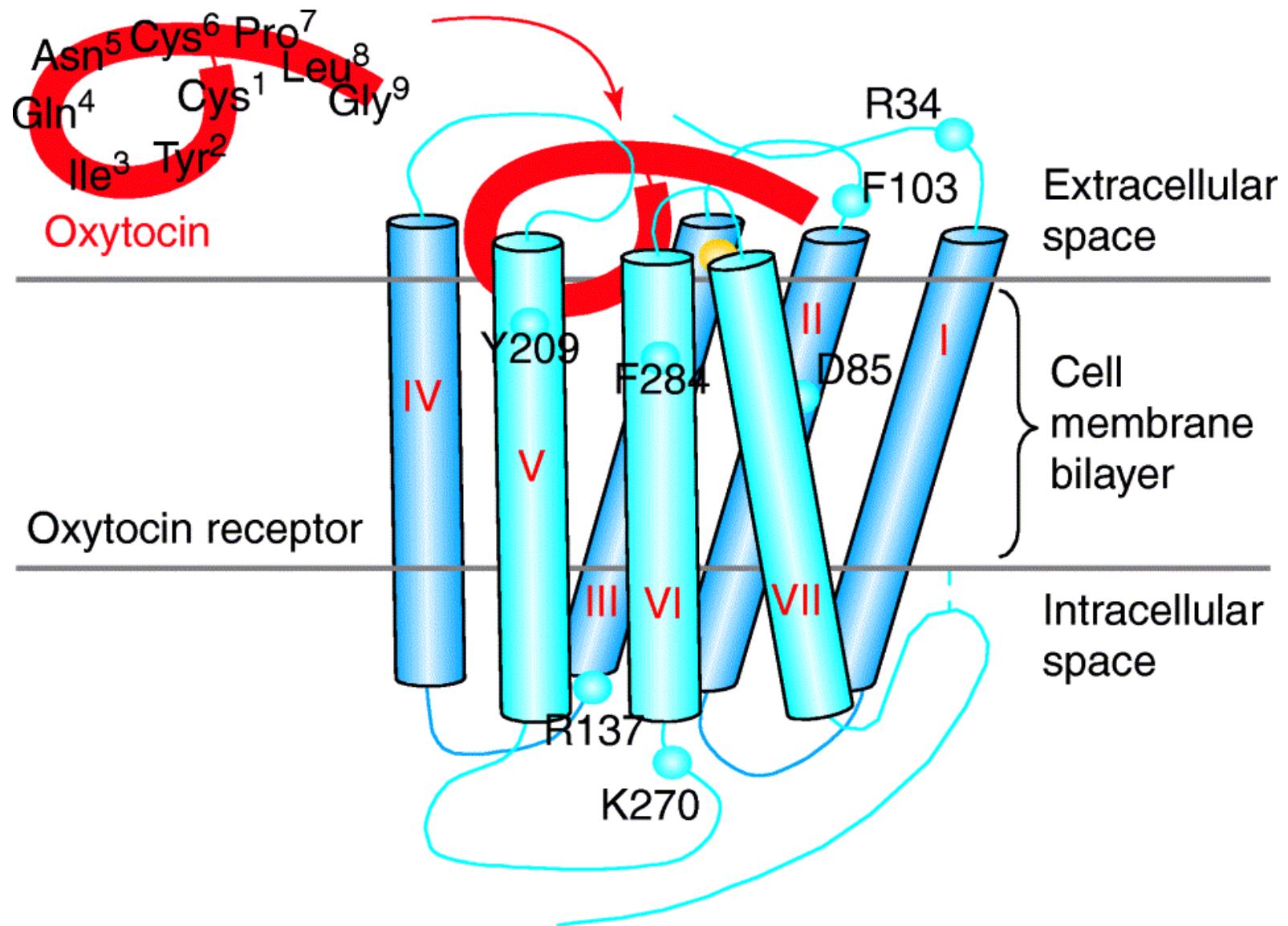


# PROTEINS - PRIMARY STRUCTURE



## OLIGOPEPTIDS (e. g. Oxytocin - a nonapeptid)

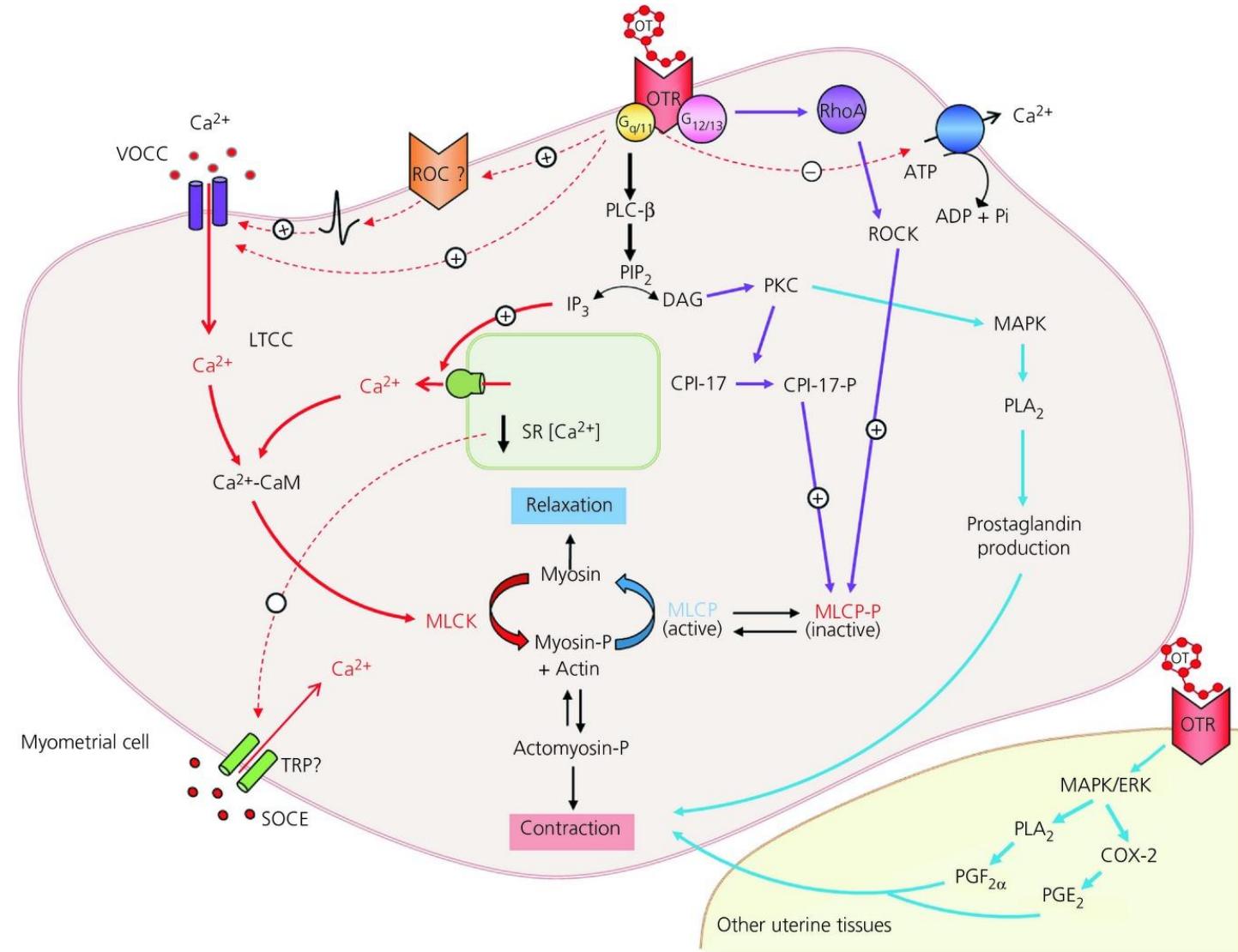




## Mechanisms leading to contraction in myometrium following oxytocin stimulation

OTRs couple to G<sub>q/11</sub> proteins to activate phospholipase C $\beta$  (PLC- $\beta$ ) which controls the hydrolysis of phosphatidylinositol-4, 5-bisphosphate (PIP<sub>2</sub>) into inositol-1, 4, 5-trisphosphate (IP<sub>3</sub>) and diacylglycerol (DAG). These in turn control the mobilisation of Ca from the sarcoplasmic reticulum (SR) and the activation of protein kinase C (PKC) respectively. **The increase in intracellular Ca ([Ca]<sub>i</sub>) brings about contraction via stimulation of Ca-dependent calmodulin and activation of myosin light chain kinase (MLCK). MLCK phosphorylates Ser19 on the regulatory light chains of myosin enabling acto-myosin cross bridge cycling and myometrial contraction.** OT can also raise [Ca]<sub>i</sub> and hence contraction, through Ca entry.

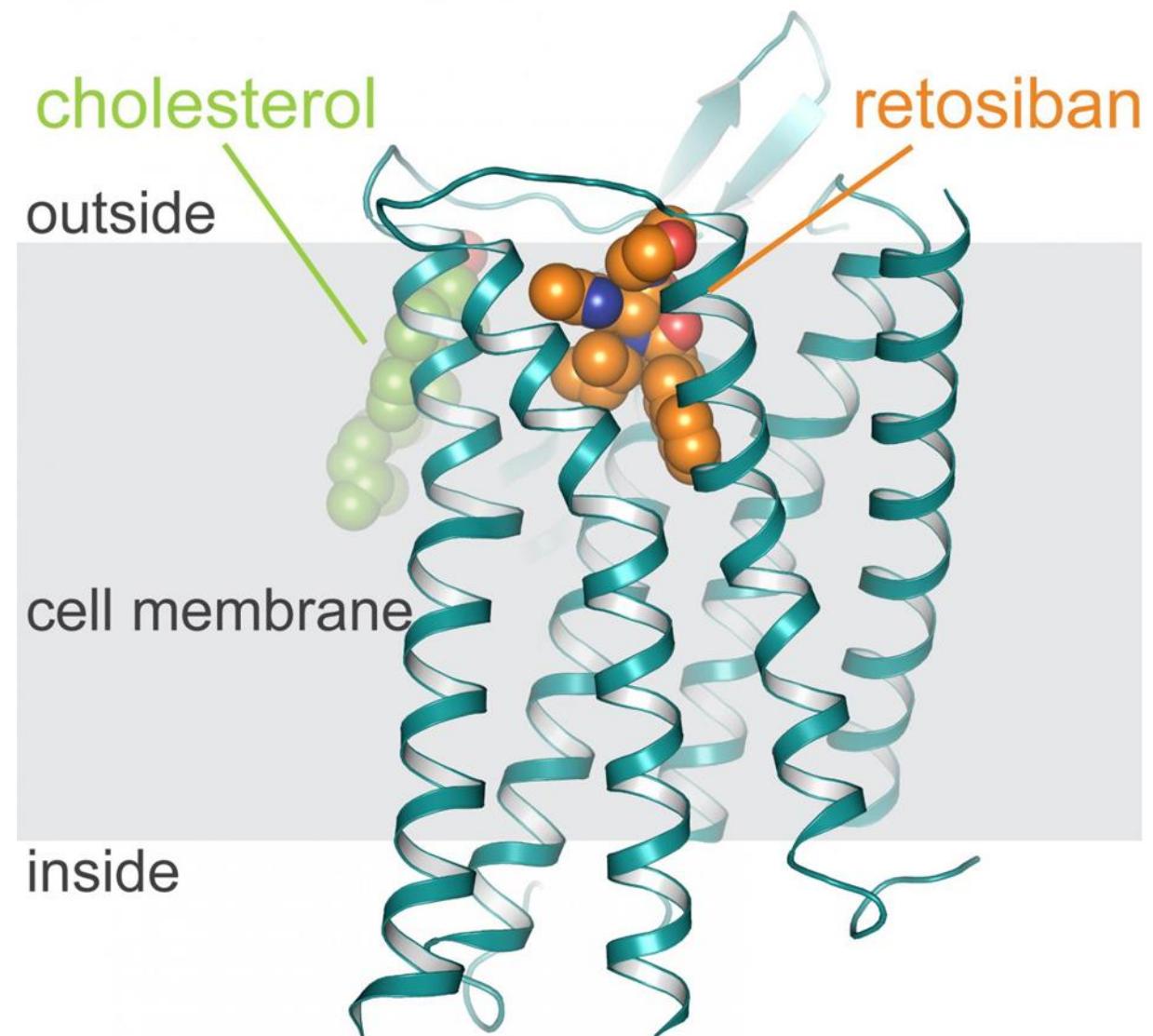
<https://www.sciencedirect.com/science/article/pii/S2468867319301464>



# OXYTOCIN

- Oxytocin is a hormone that is produced in the hypothalamus and released into the bloodstream by the pituitary gland. Its main function is to facilitate childbirth, which is one of the reasons it is called the "love/cuddle hormone."
- Women whose labor is slow to proceed are sometimes given oxytocin to speed the process.
- Retosiban is an oral drug which acts as an oxytocin receptor antagonist. It was developed for the treatment of preterm labour. Retosiban has high affinity for the oxytocin receptor.

## Oxytocin receptor



# PEPTIDES - PROTEINS

Amino acids can be polymerized to form chains:

- if you have two amino acids they will form a dipeptide that has one peptide bond
- if you have three amino acids they will form a tripeptide that has two peptide bond, etc.
- if you have few ( 2-20 amino acids ) they will form oligopeptides
- if you have more amino acids they will form polypeptides
- more than 100 amino acids form proteins.

Most proteins consist of less than 2 000 amino acids, but there are also ones consisting of more than 30 000 (e. g. Titin the largest known protein; its human variant consists of 34 350 amino acids, it is part of human muscles)

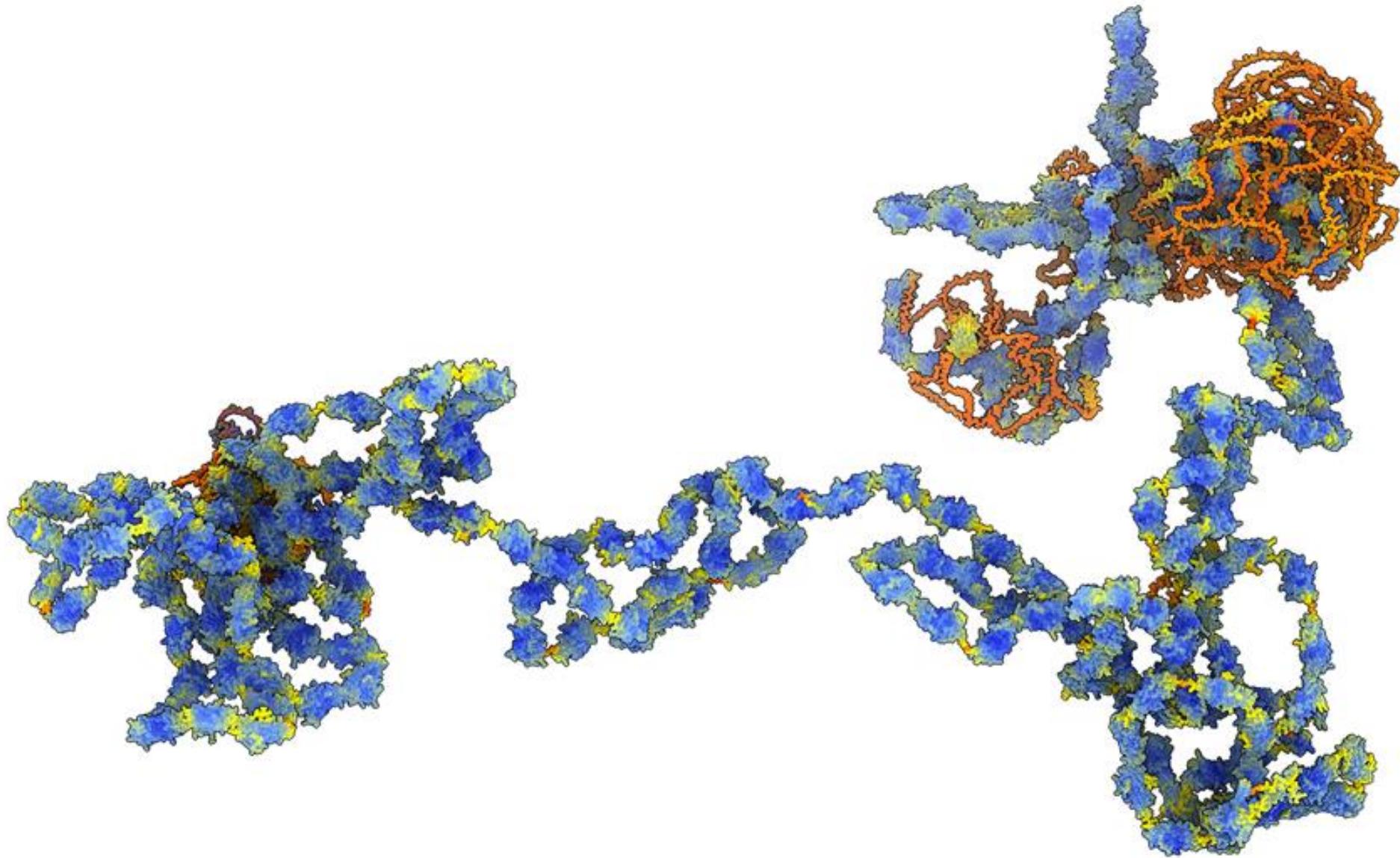
International Union  
of Pure and  
Applied Chemistry  
(IUPAC) is the world  
authority on  
chemical  
nomenclature and  
terminology

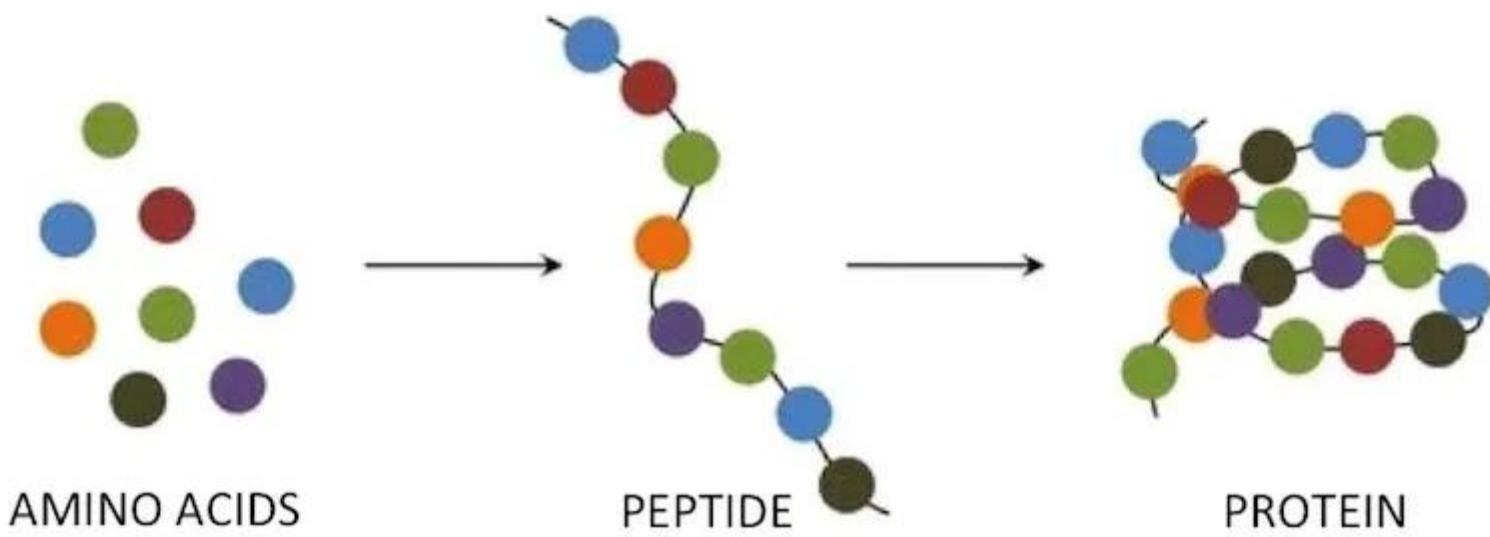
**Titin is the largest known protein.  
(C169723 H270464 N45688 O52243 S912)**

**It has the longest IUPAC name at 189,891 letters. It takes approximately 3.5 hours to read it in its entirety.**

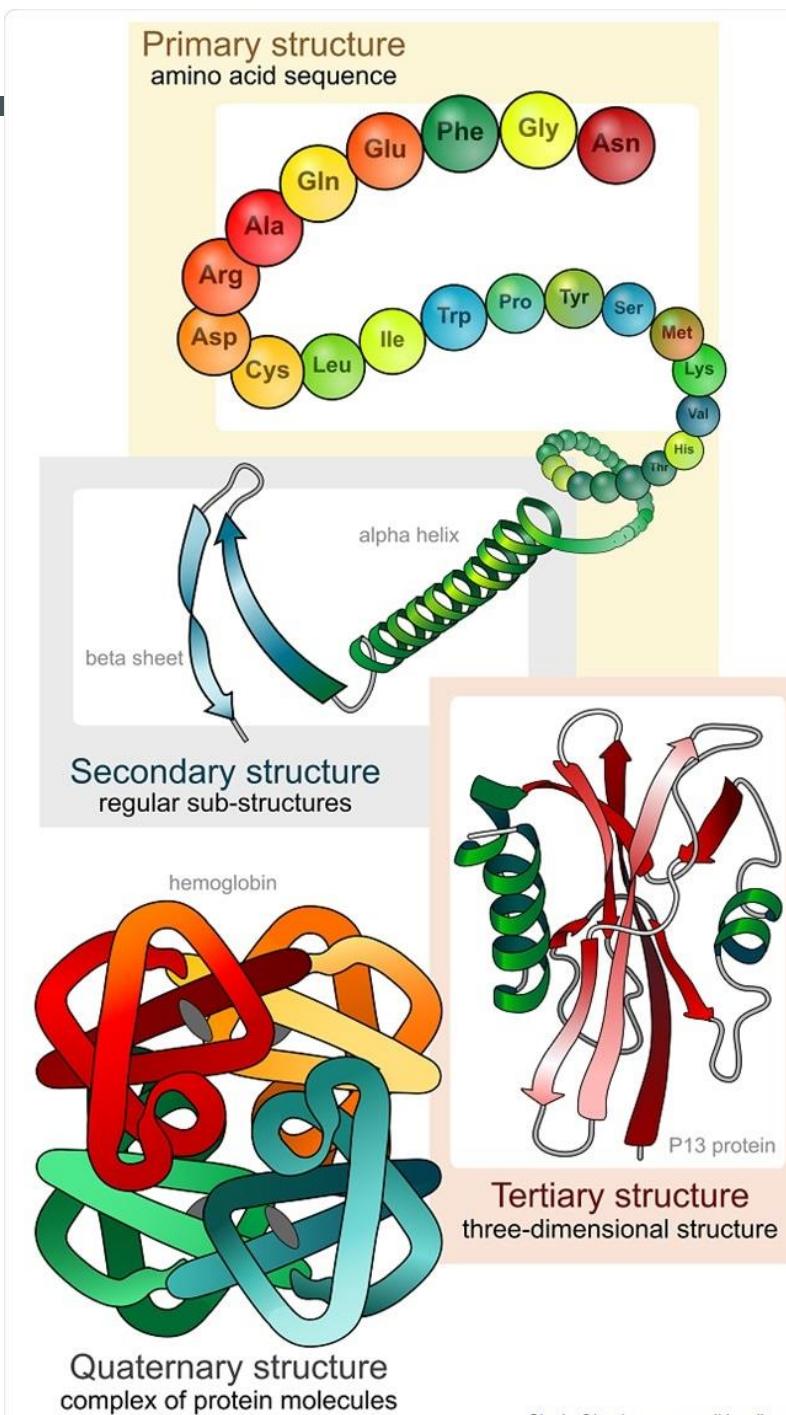
---

# TITIN

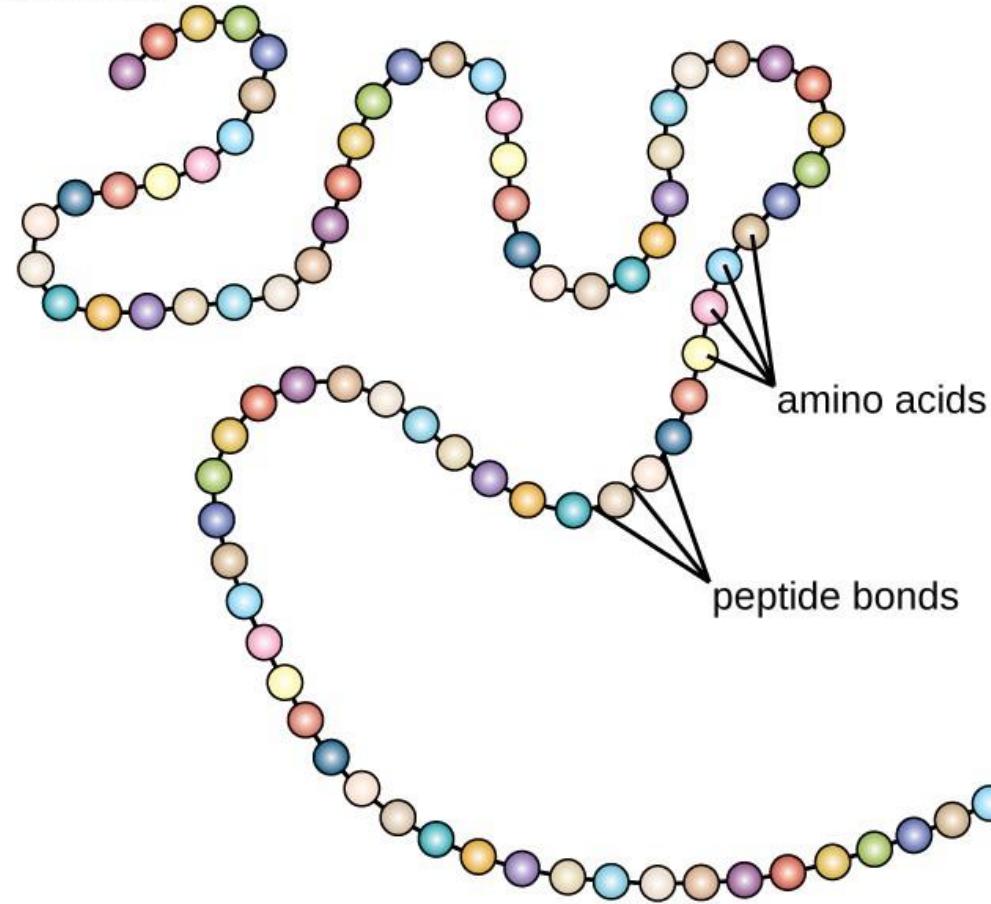




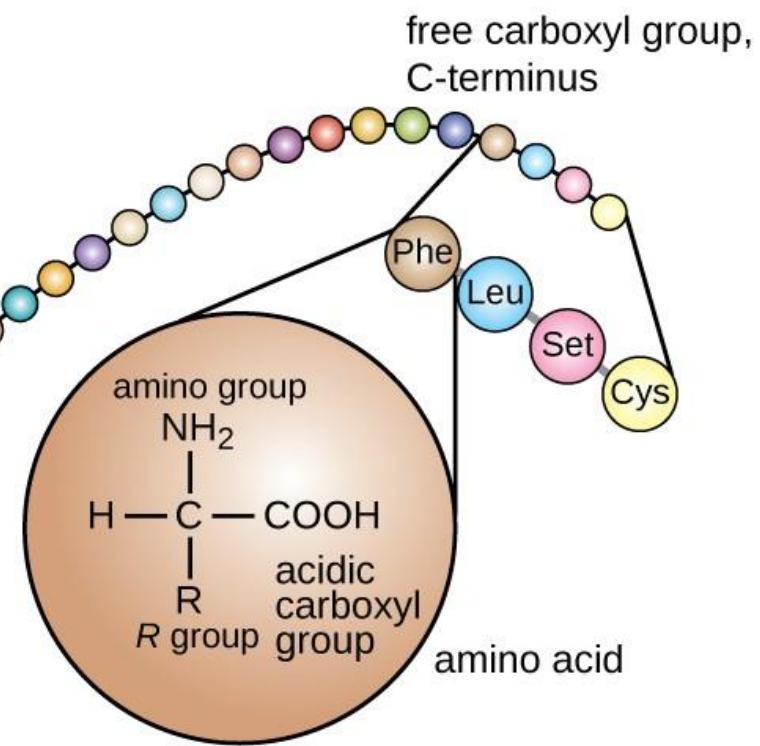
# PROTEINS



free amino group,  
N-terminus



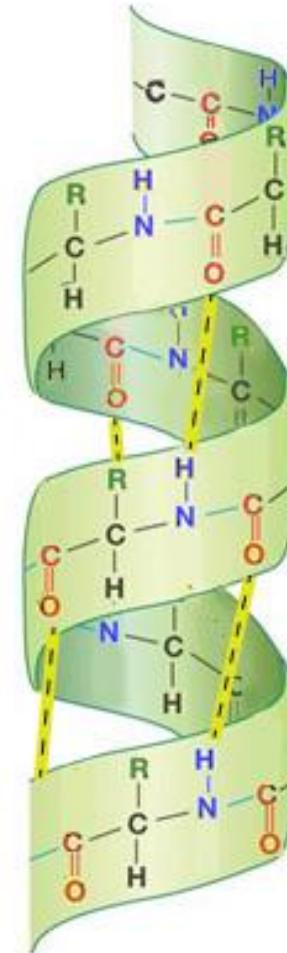
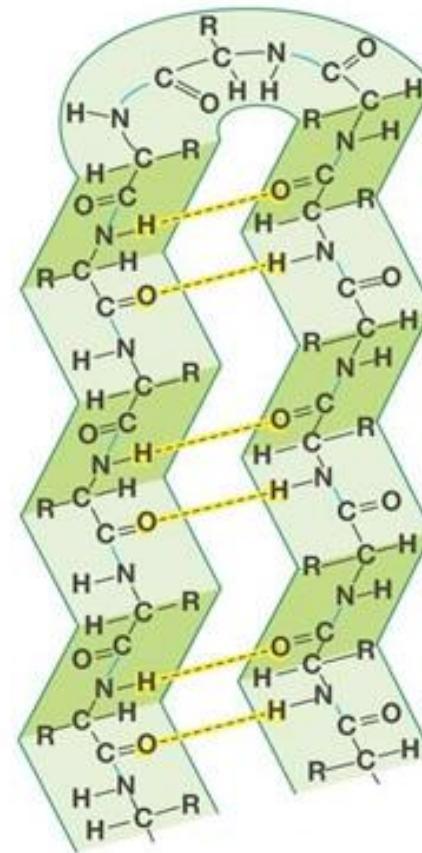
The primary protein structure  
is the chain of amino acids  
that makes up the protein.



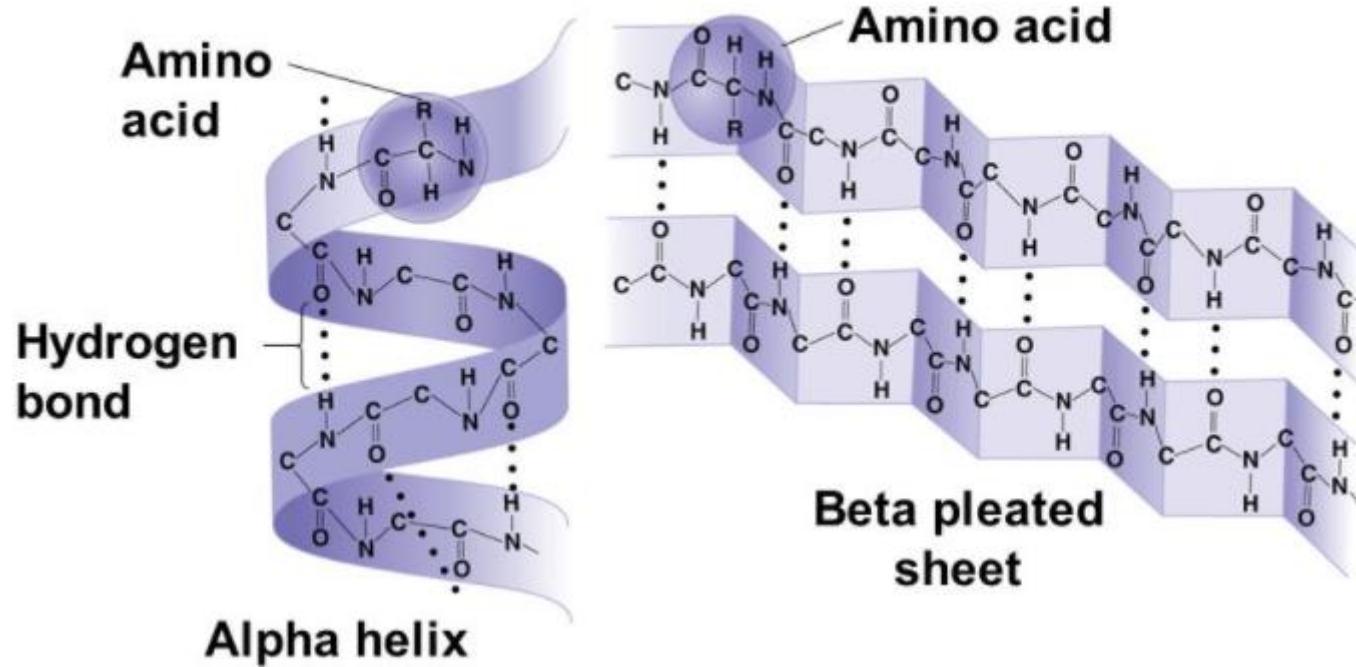
# PROTEINS – SECONDARY STRUCTURE

The secondary structure is caused by hydrogen bonds forming

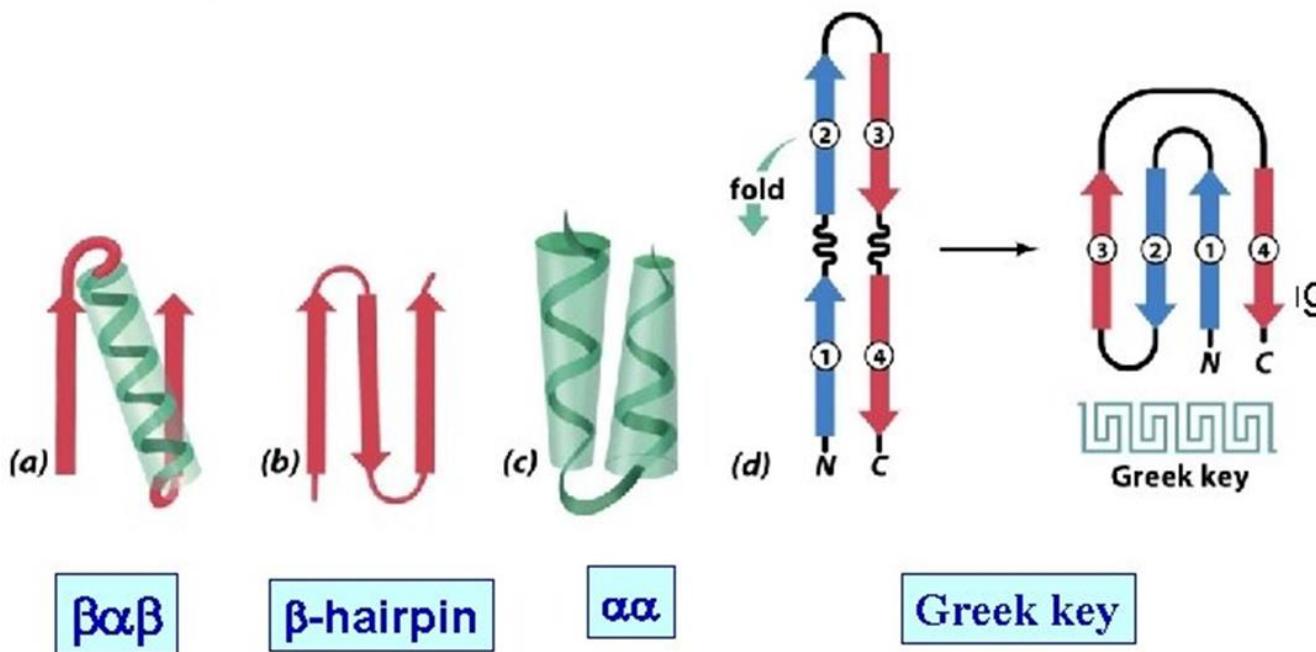
- a helix or
- a pleated sheet



## Secondary structure

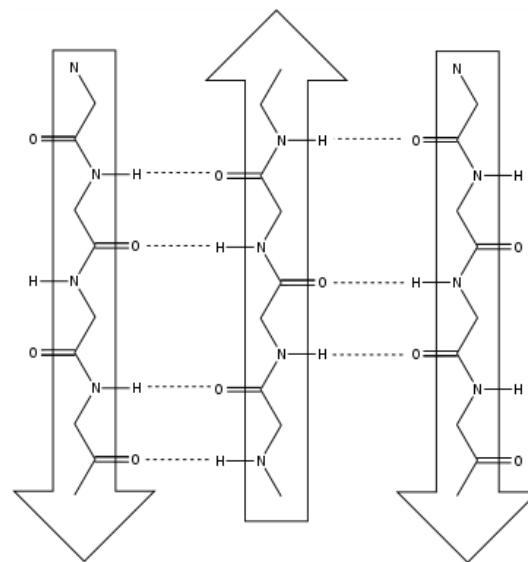


# PROTEINS – SECONDARY STRUCTURE

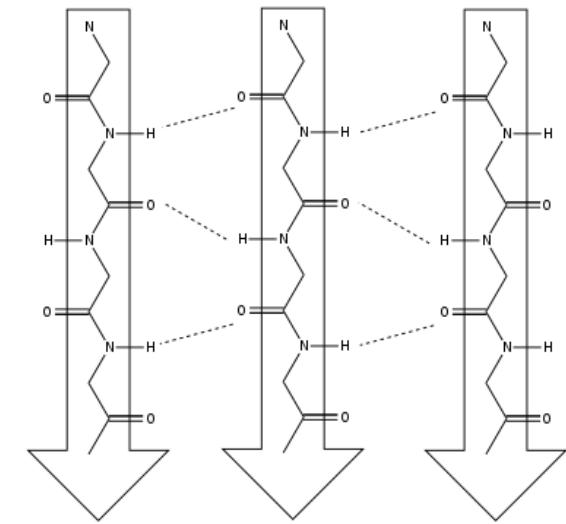


# BETA SHEETS

- In an antiparallel  $\beta$ -sheet, the polypeptide strands are arranged such that a C=O and an NH from adjacent strands face each other, and the H forms an H-bond with the O, with C=O..HN all lying on the same plane.
- In a parallel  $\beta$ -sheet, the strands are literally parallel, and so the H-bond formed has to form a sort of angle, where the C=O and NH no longer lie on the same plane
- antiparallel sheets are more stable than parallel ones** which is consistent with the hydrogen bond geometry:
  - bond length in parallel strands would be longer, thereby weakening it
  - hydrogen bonds are aligned directly opposite each other, making for stronger and more stable bonds
- Most sheets are mixed rather than purely parallel or anti-parallel.

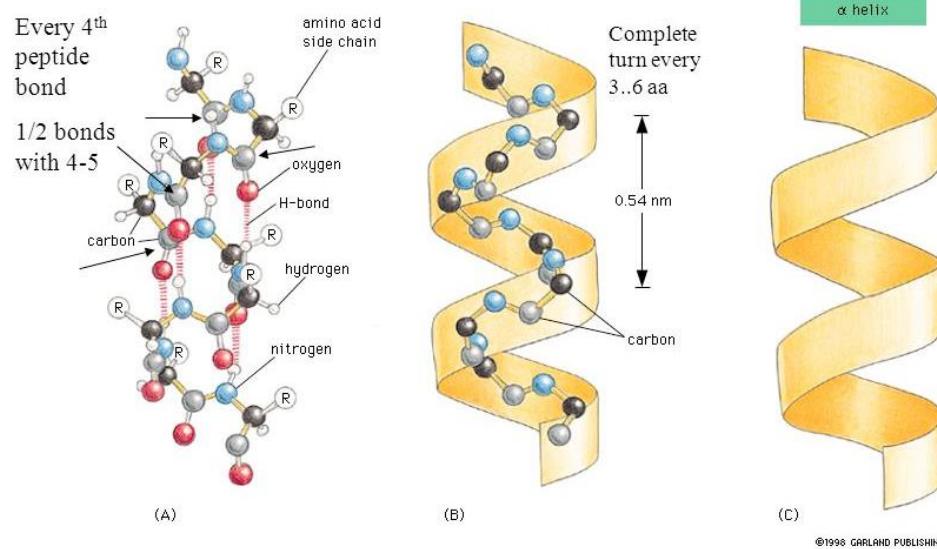


Antiparallel  $\beta$ -sheet

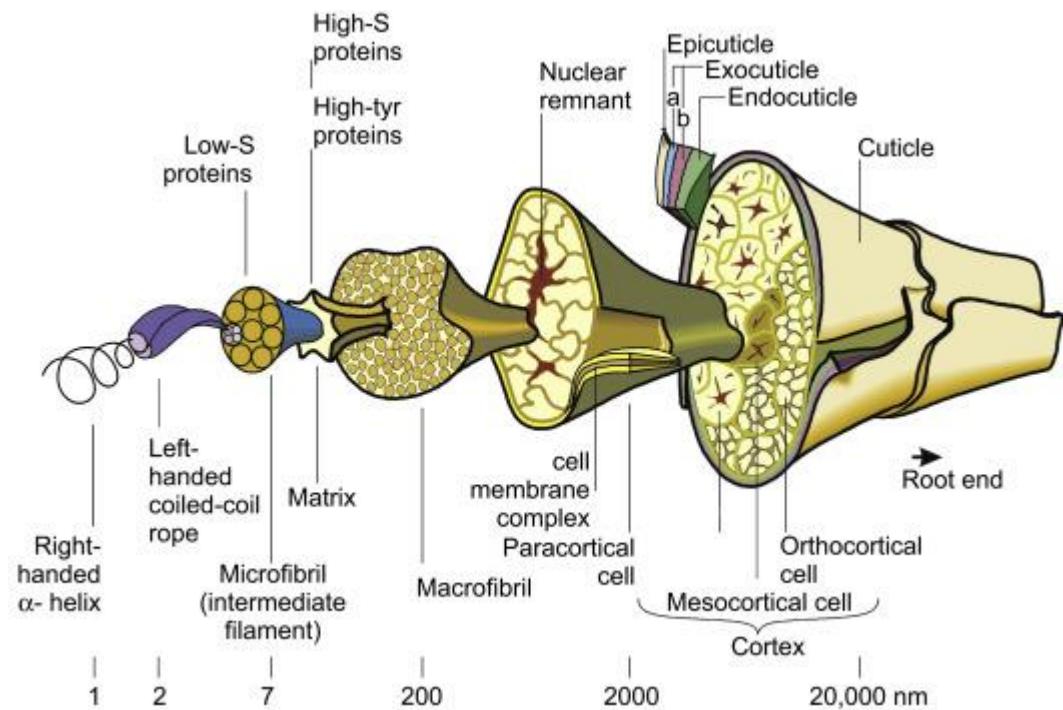


Parallel  $\beta$ -sheet

# ALPHA-KERATIN

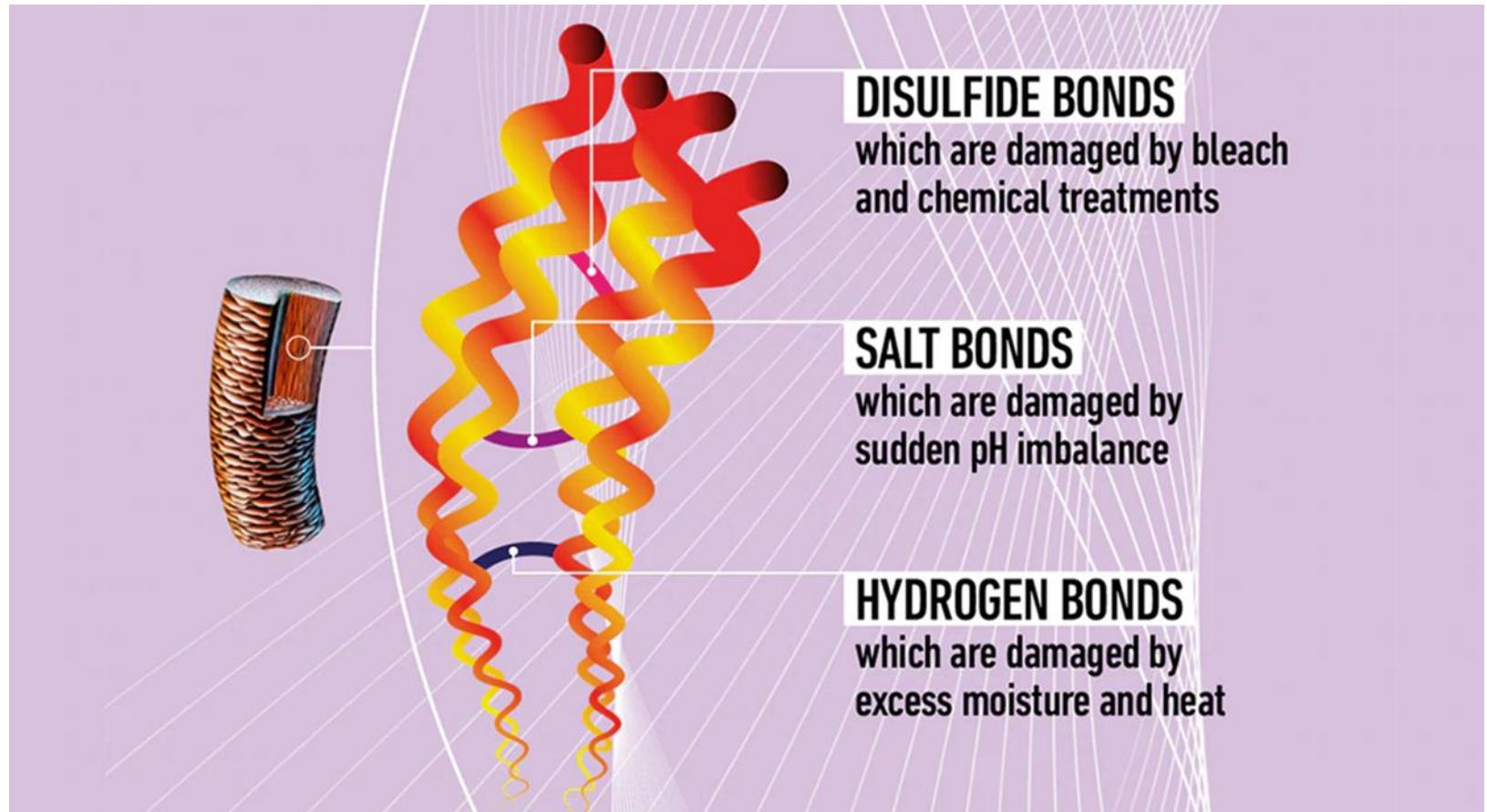


Alpha helix is found in alpha-keratin, abundant in skin, hair, nails



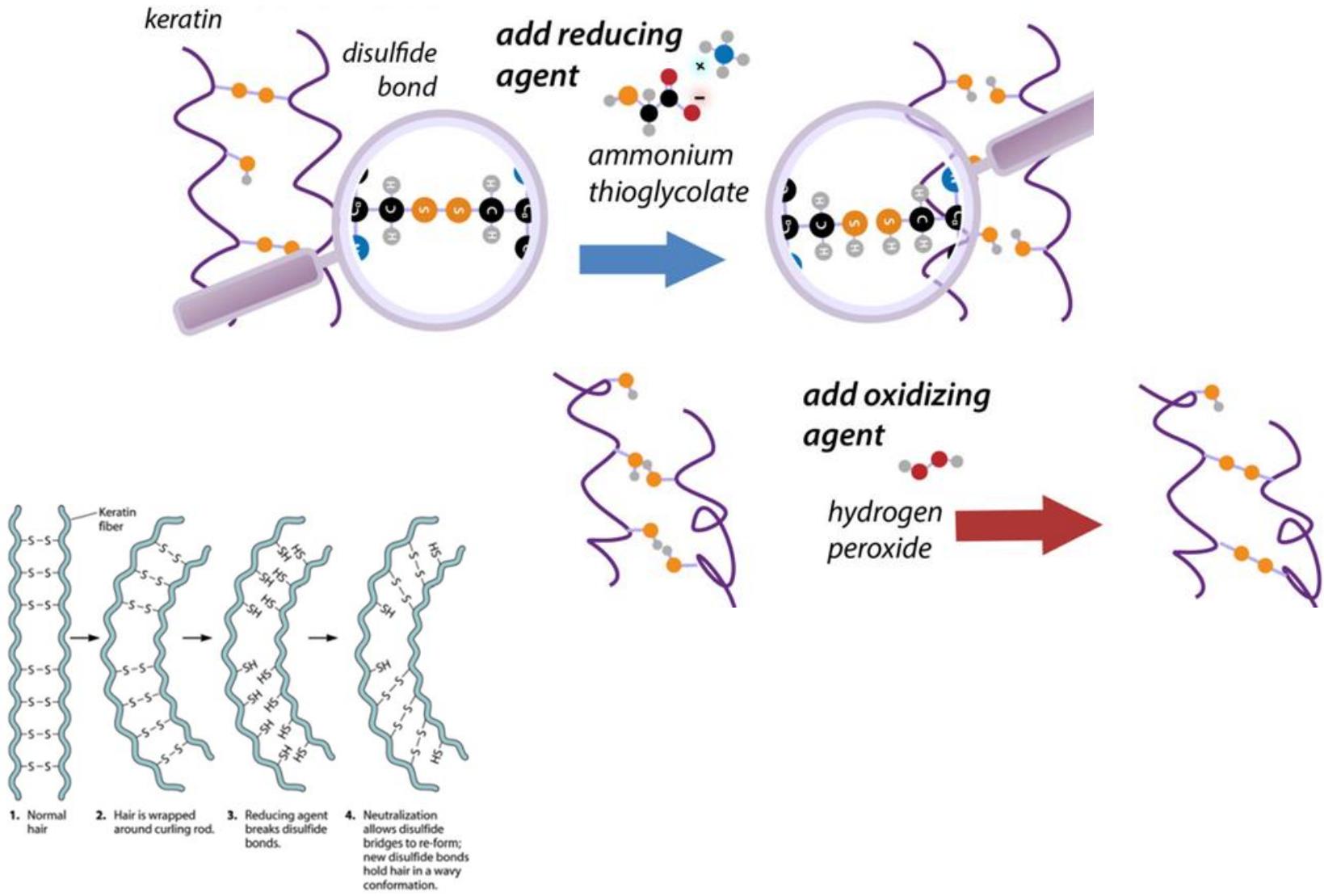
# KERATIN - HAIR

- Once hair has been shampooed, many of the hydrogen bonds are broken, this allows the hair to be stretched around a brush while blow drying.
- Permanent waving: first reduces and then generates new disulfid bonds.



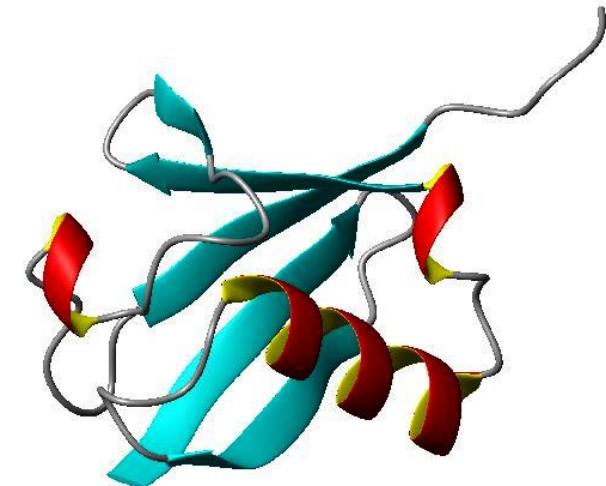
# DISULFIDE BONDS

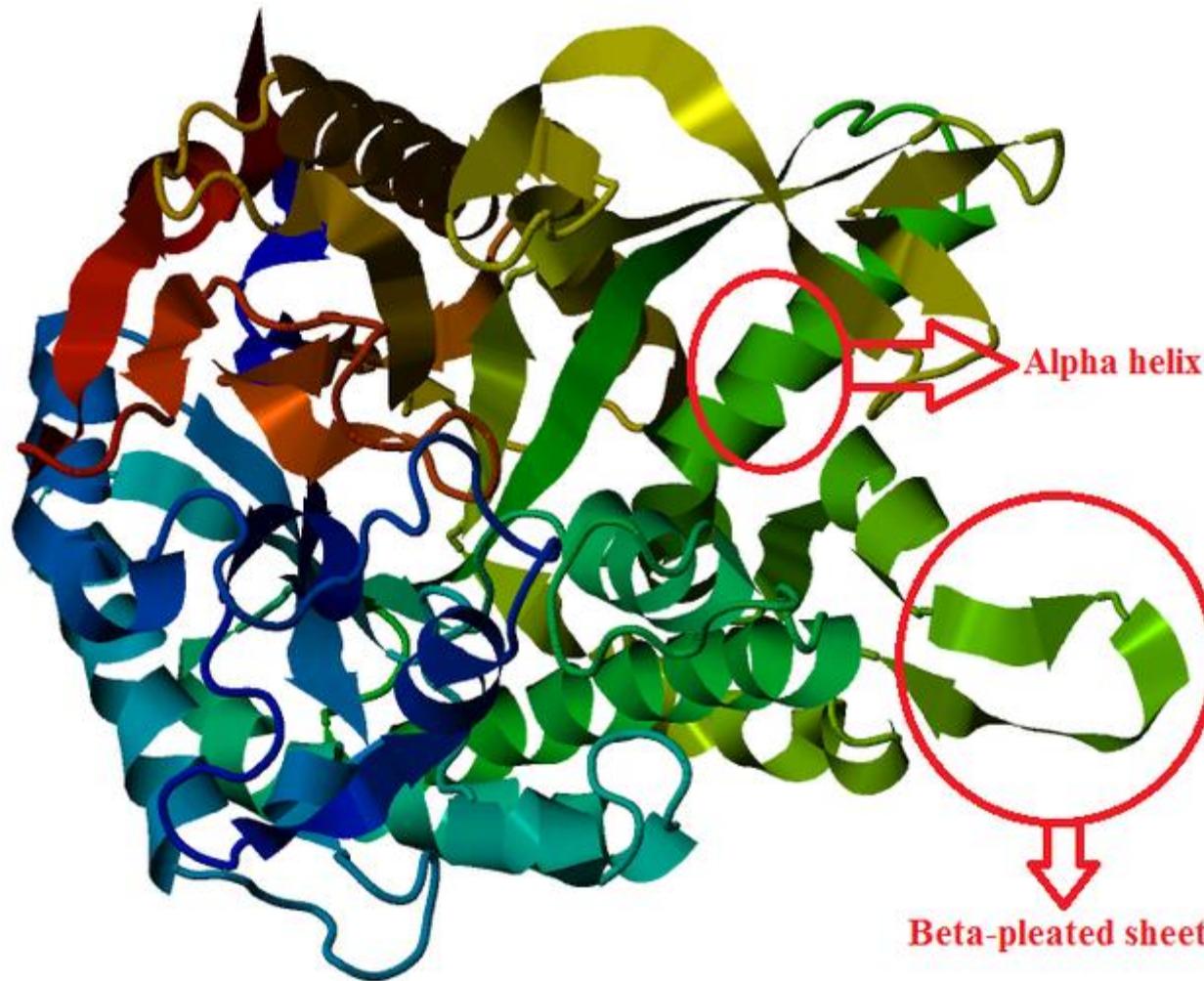
- Cysteine in keratin forms strong crosslinks that hold hair's shape
- Reducing agents break these bonds
- By oxidizing the bonds can be reformed again



# RIBBON DIAGRAMS

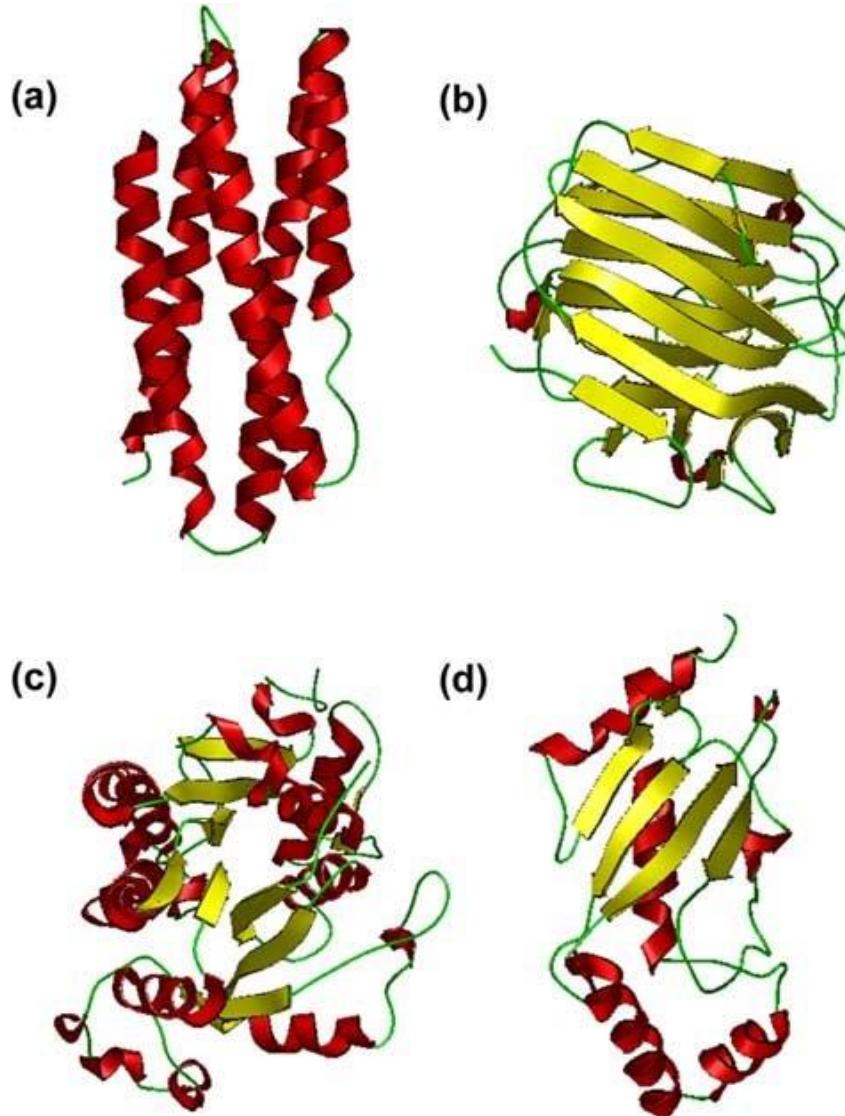
- Ribbon diagrams are 3D schematic representations of protein structure and are one of the most common methods of protein depiction.
- The ribbon shows the overall organization of the protein and serves as a **visual framework** on which to hang details of the full atomic structure.
- Ribbon diagrams are simple yet powerful, expressing the **visual basics** of a molecular structure. This method has successfully portrayed the overall organization of protein structures, reflecting their three-dimensional nature and allowing better understanding of these complex objects.
- $\alpha$ -helices are shown as coiled ribbons,  $\beta$ -strands as arrows



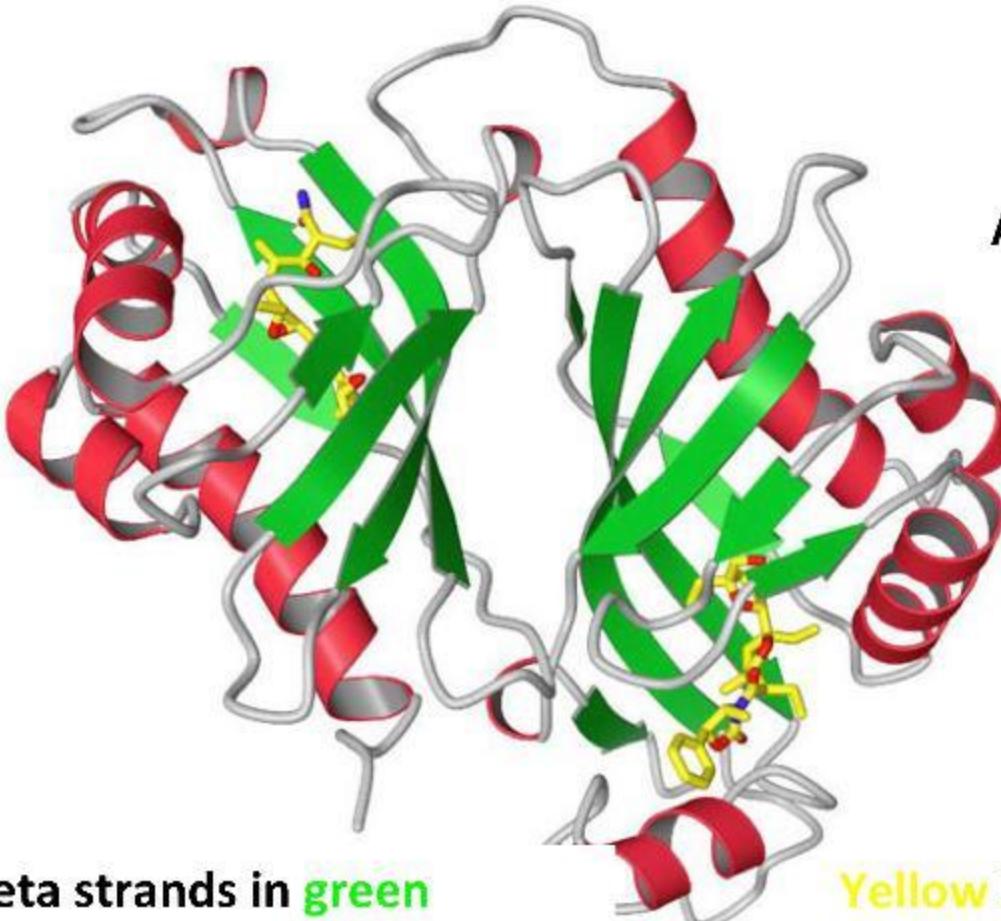


Ribbon drawings to show the four structural classes of proteins:

- (a) all- $\alpha$ ,
- (b) all- $\beta$ ,
- (c)  $\alpha/\beta$ , and
- (d)  $\alpha+\beta$ .



# Ribbon Diagrams



Beta strands in **green**  
(arrow shows direction from N → C)

**Yellow** ball & stick  
indicates bound molecules

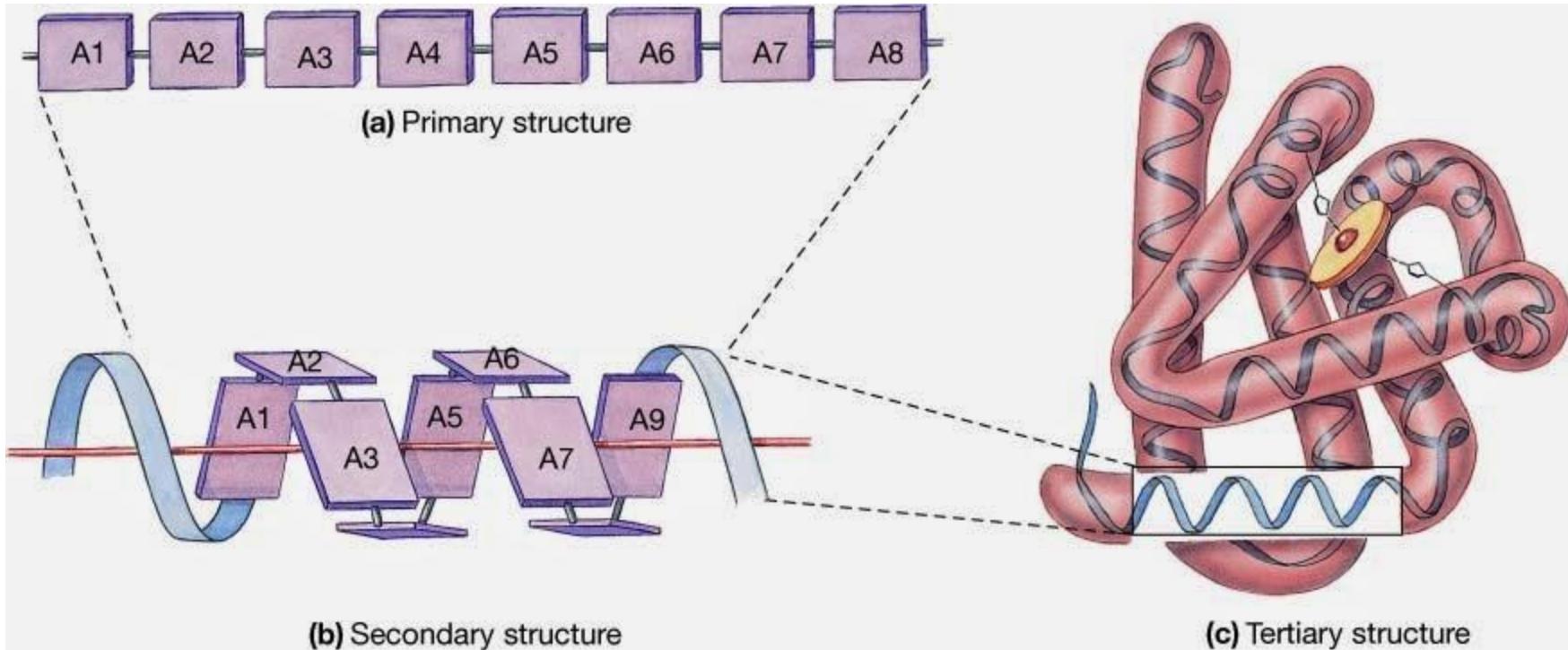
Alpha Helices  
in **red**

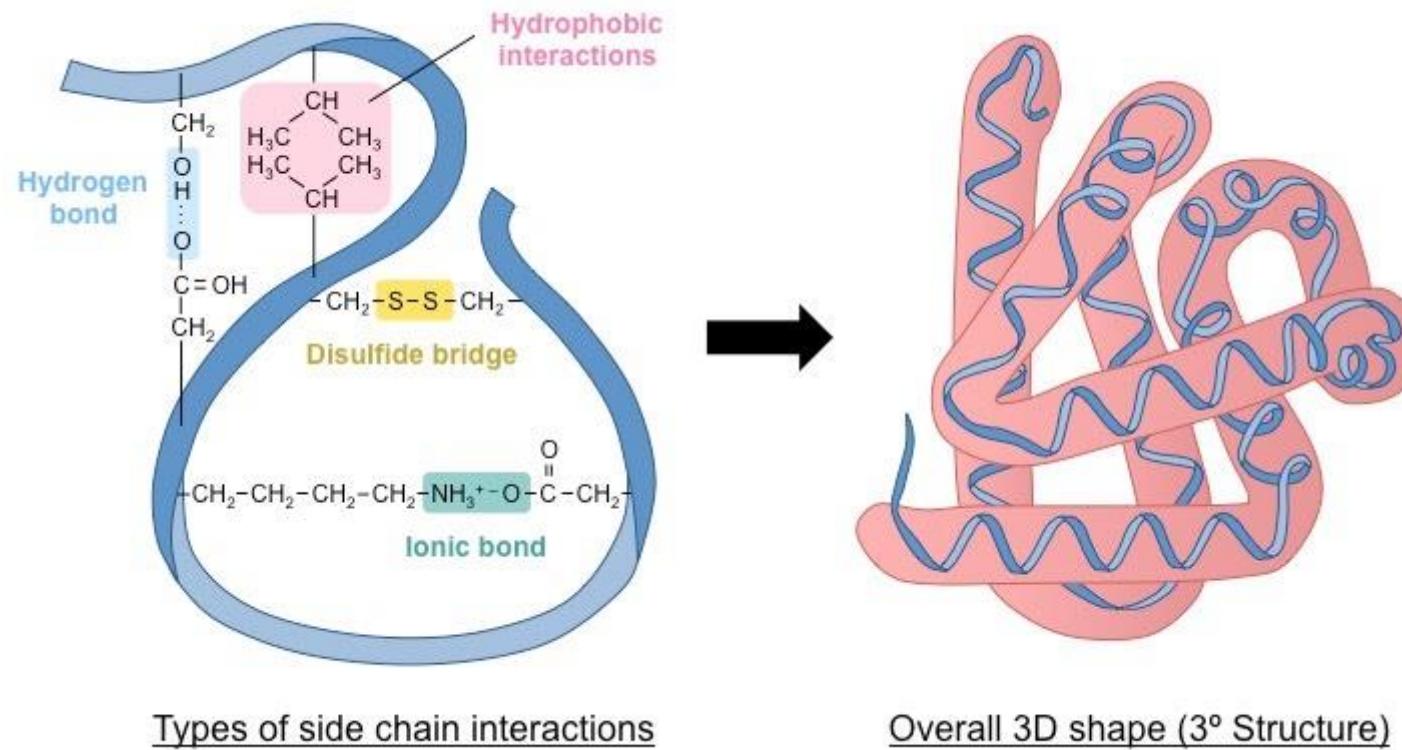
Loops & turns  
in **grey**

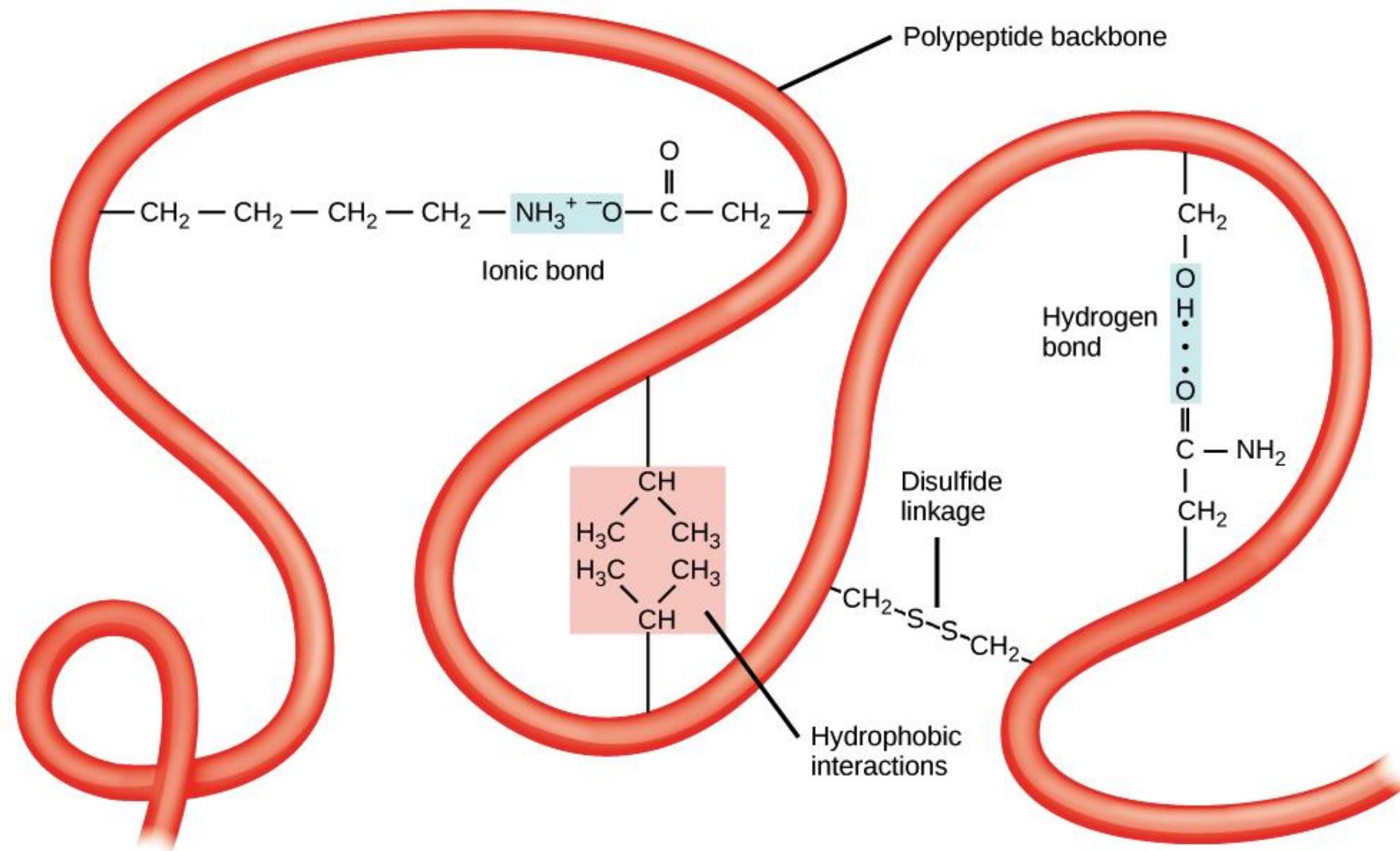
---

## **PROTEINS - TERTIARY STRUCTURE**

- The three dimensional conformation of a polypeptide
- The common features of protein tertiary structure reveal much about the biological functions of the proteins and their evolutionary origins.
- **The function of a protein depends on its tertiary structure.  
If this is disrupted, it loses its activity.**



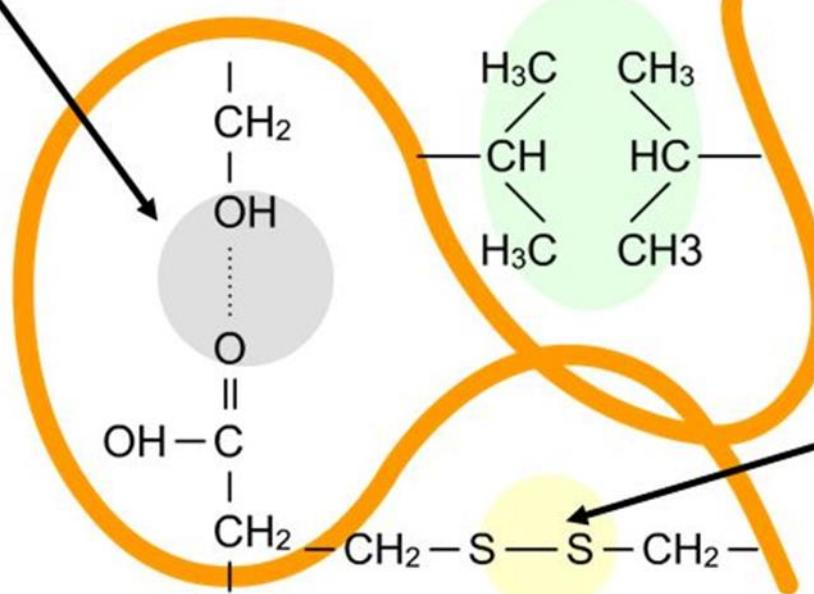




# How is the 3D Shape held together?

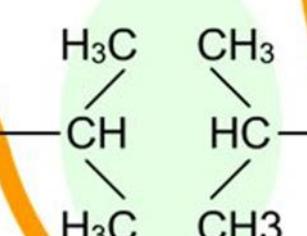
**hydrogen bonds:**

involved in all levels of structure.



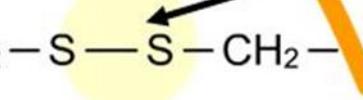
**hydrophobic interactions:**

between non-polar sections of the protein.



**disulfide bonds:** one

of the strongest and most important type of bond in proteins. Occur between two cysteine amino acids.

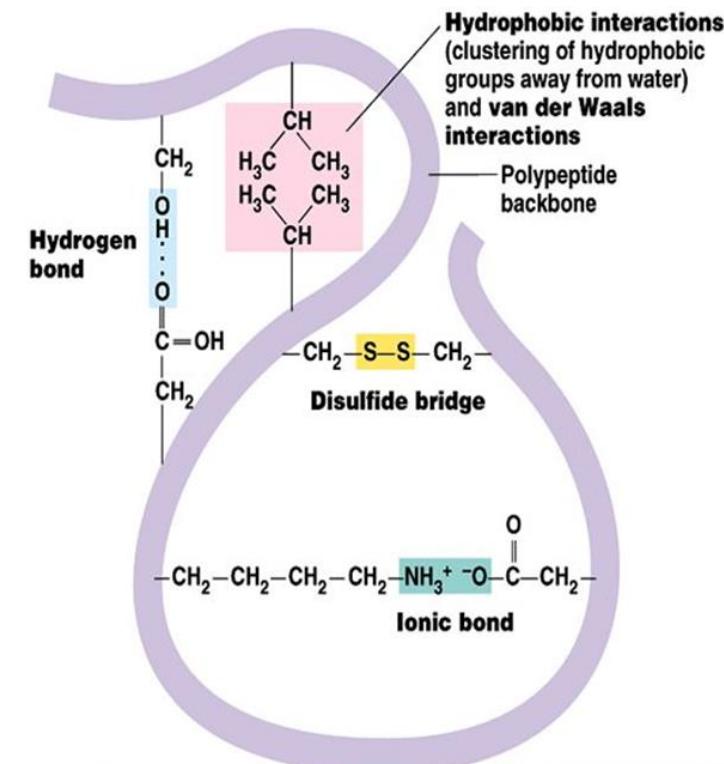


# TERTIARY STRUCTURE

- ▶ is determined by a variety of interactions (bond formation) among R groups and between R groups and the polypeptide backbone.

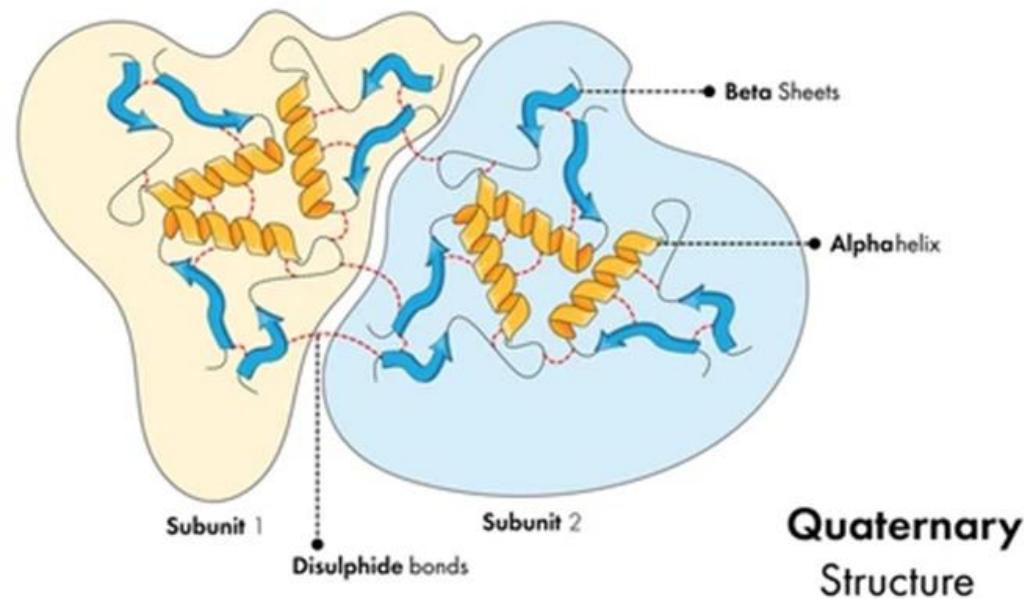
The weak interactions include:

- ▶ Hydrogen bonds among polar side chains
- ▶ Ionic bonds between charged R groups (basic and acidic amino acids)
- ▶ Hydrophobic interactions among hydrophobic (non polar) R groups.
- ▶ Strong covalent bonds include disulfide bridges, that form between the sulfhydryl groups (SH) of cysteine monomers, stabilize the structure.

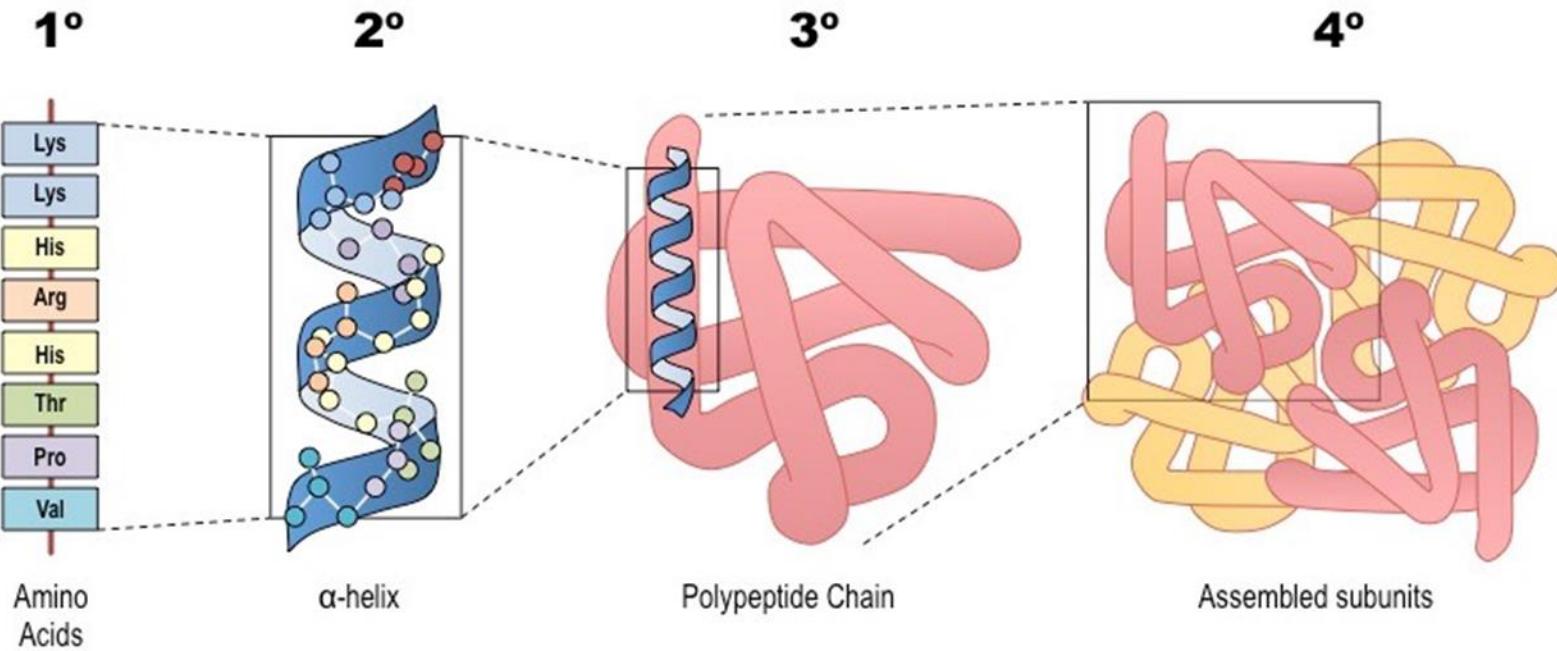


# PROTEINS - QUARTERNARY STRUCTURE

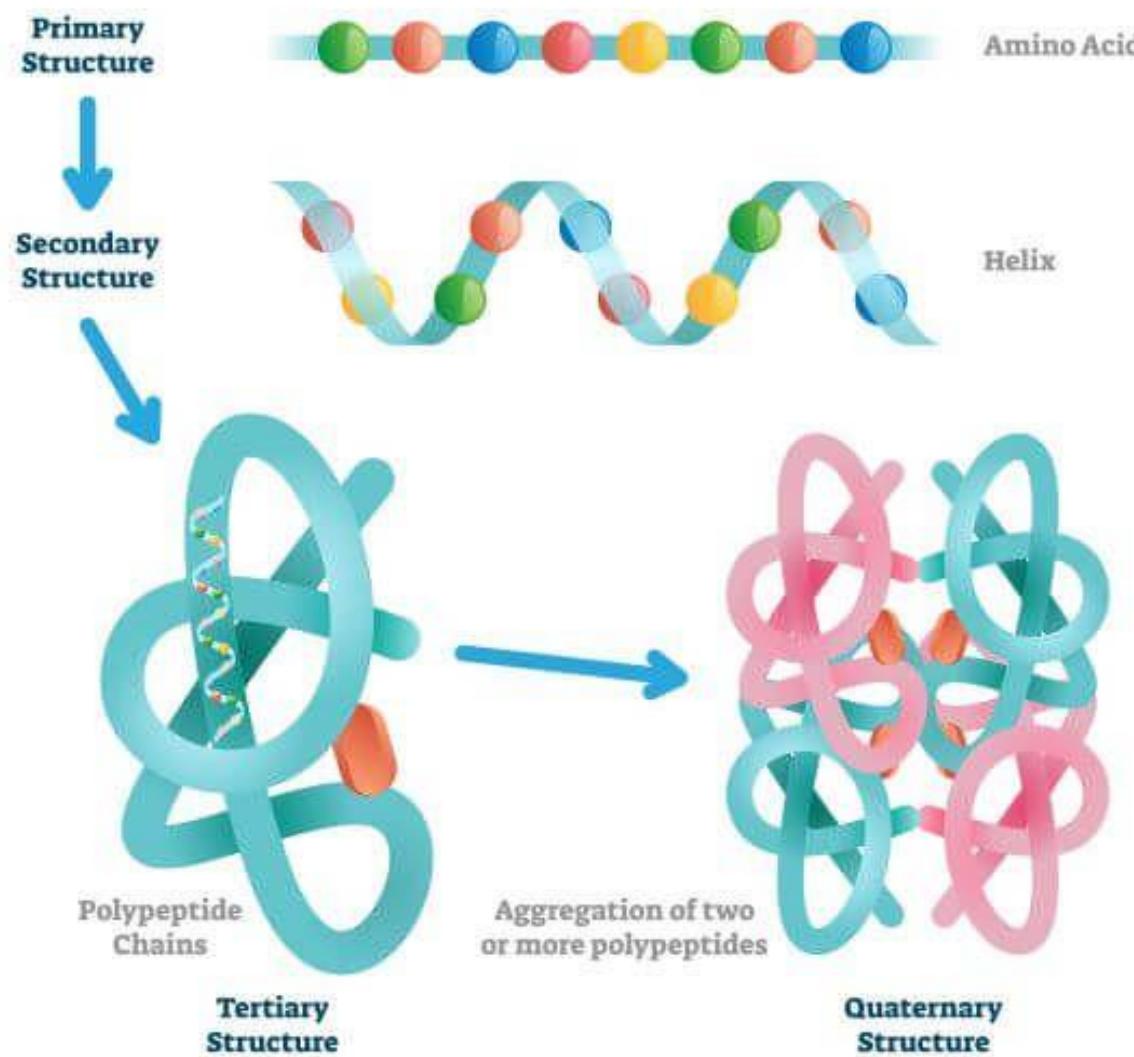
- The arrangement of polypeptide subunits is called quaternary structure.



# PROTEINS - QUARTERNARY STRUCTURE



# PROTEIN STRUCTURE

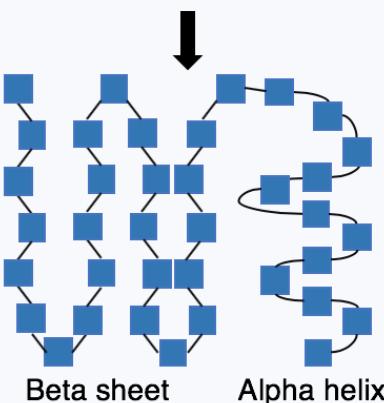


**Primary structure**

Polypeptide chain

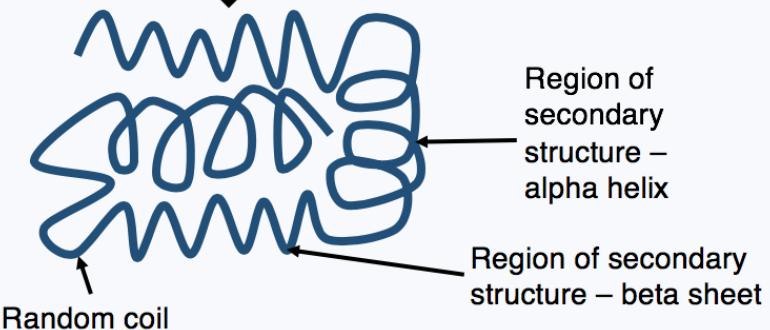
**Secondary structure**

Alpha helix and beta sheet structures produced by hydrogen bonds forming within the polypeptide



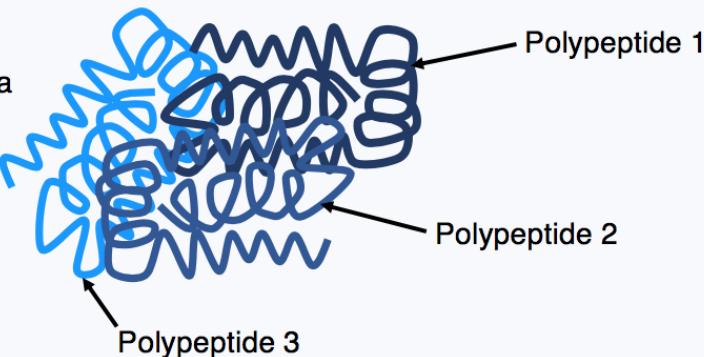
**Tertiary structure**

3D overall fold of the protein containing secondary structures



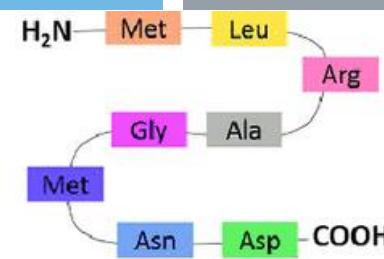
**Quaternary structure**

Multi-subunit complex where each subunit is a distinct polypeptide chain



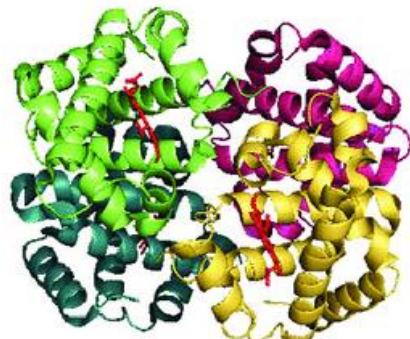
Ala Phe Leu  
Asn Gly Trp Lys  
Ile Cys Asp  
Arg Met

Amino acid



Primary structure

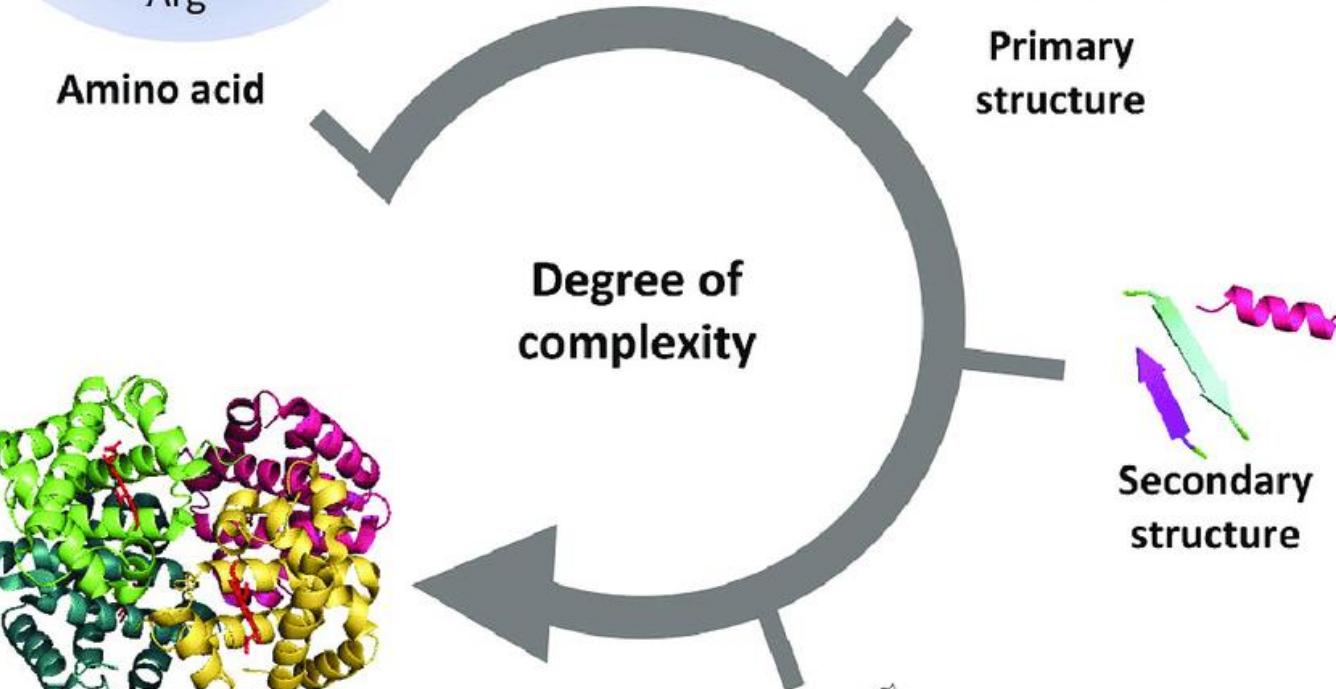
Degree of complexity



Secondary structure



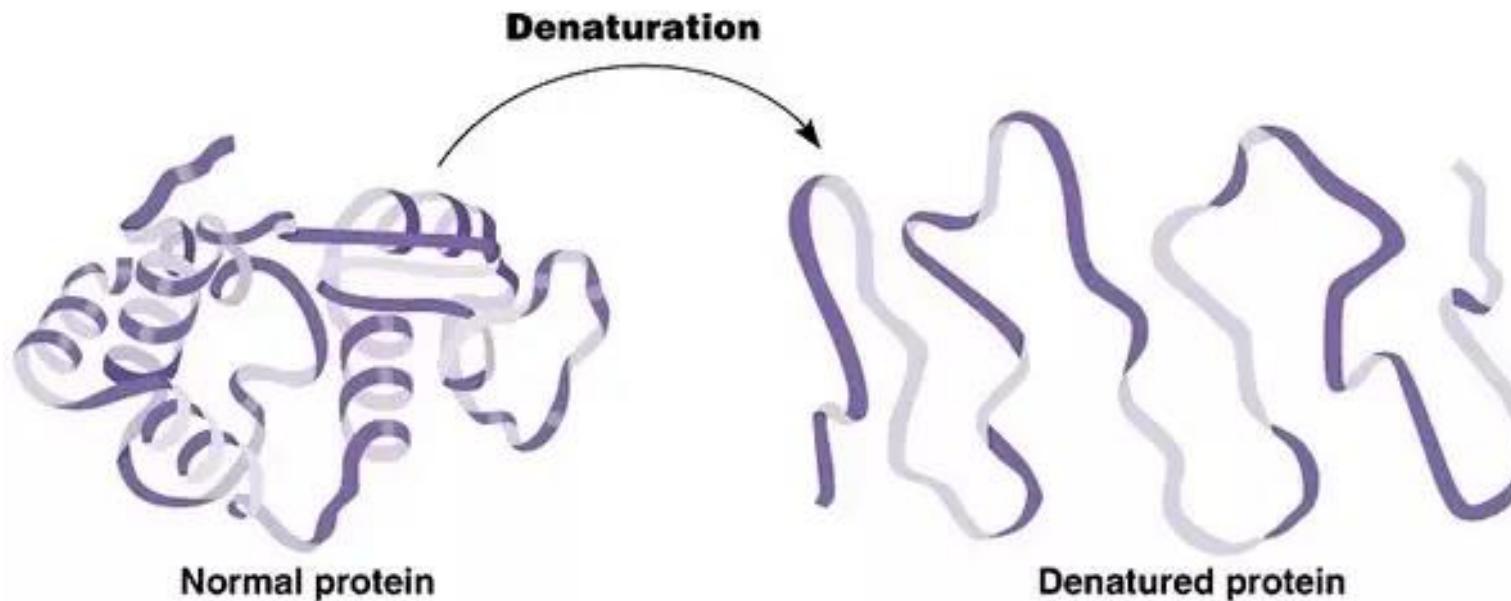
Tertiary structure

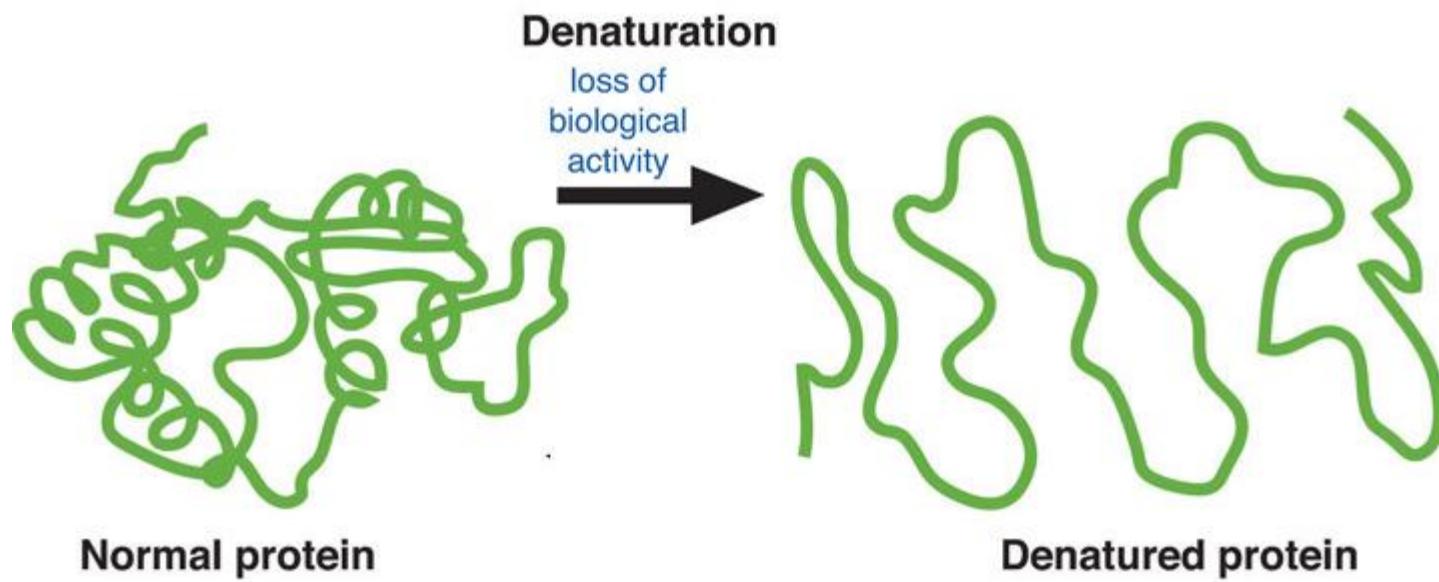


# DENATURATION OF PROTEINS

- the unfolding and disorganization of the protein's secondary and tertiary structures.
- Most proteins, once denatured, remain permanently disordered.
- Denatured proteins are often insoluble
- e.g the protein in the egg “ albumin “ once we expose it to heat will be denatured and become insoluble
- Denaturation agents: strong acids, heat, mechanical mixing, etc.

## DENATURATION – PROTEIN UNFOLDING

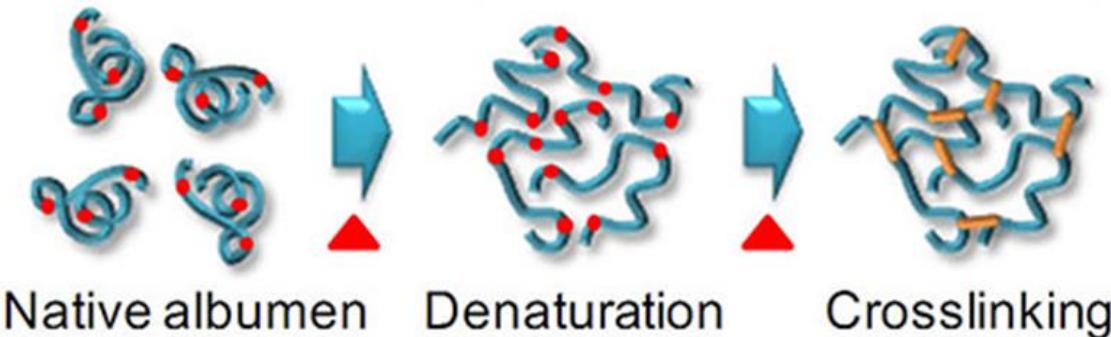




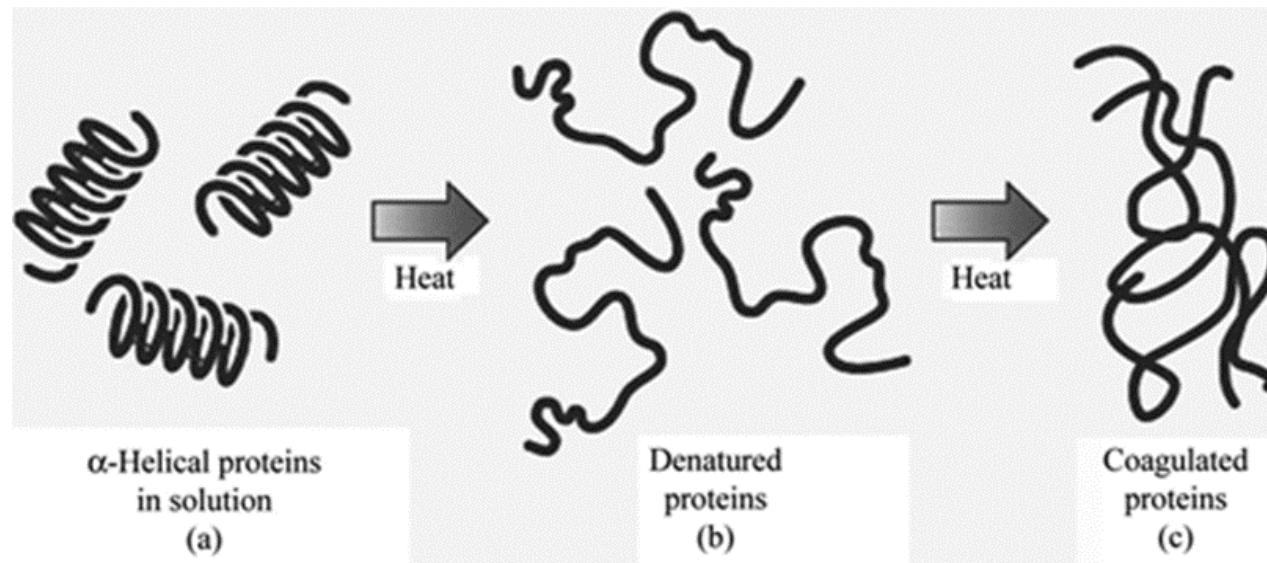
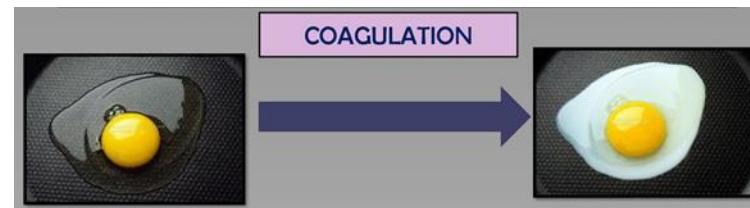
# PROTEIN DENATURATION



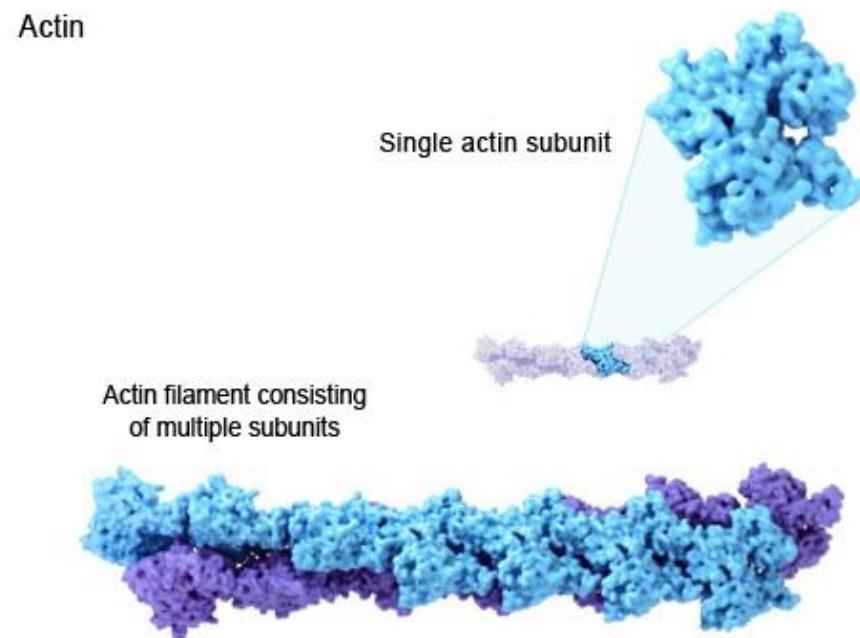
Protein Thermal Irreversible Denaturation



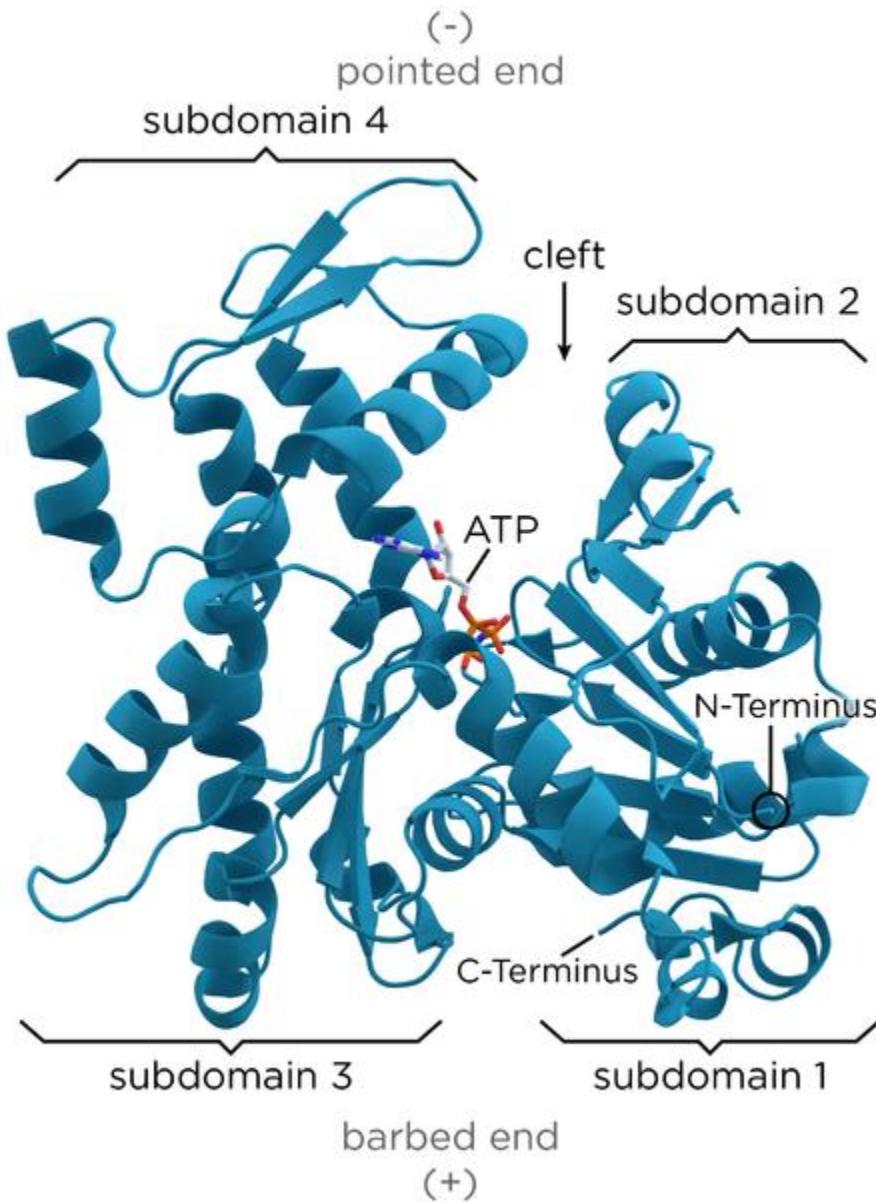
# PROTEIN DENATURATION



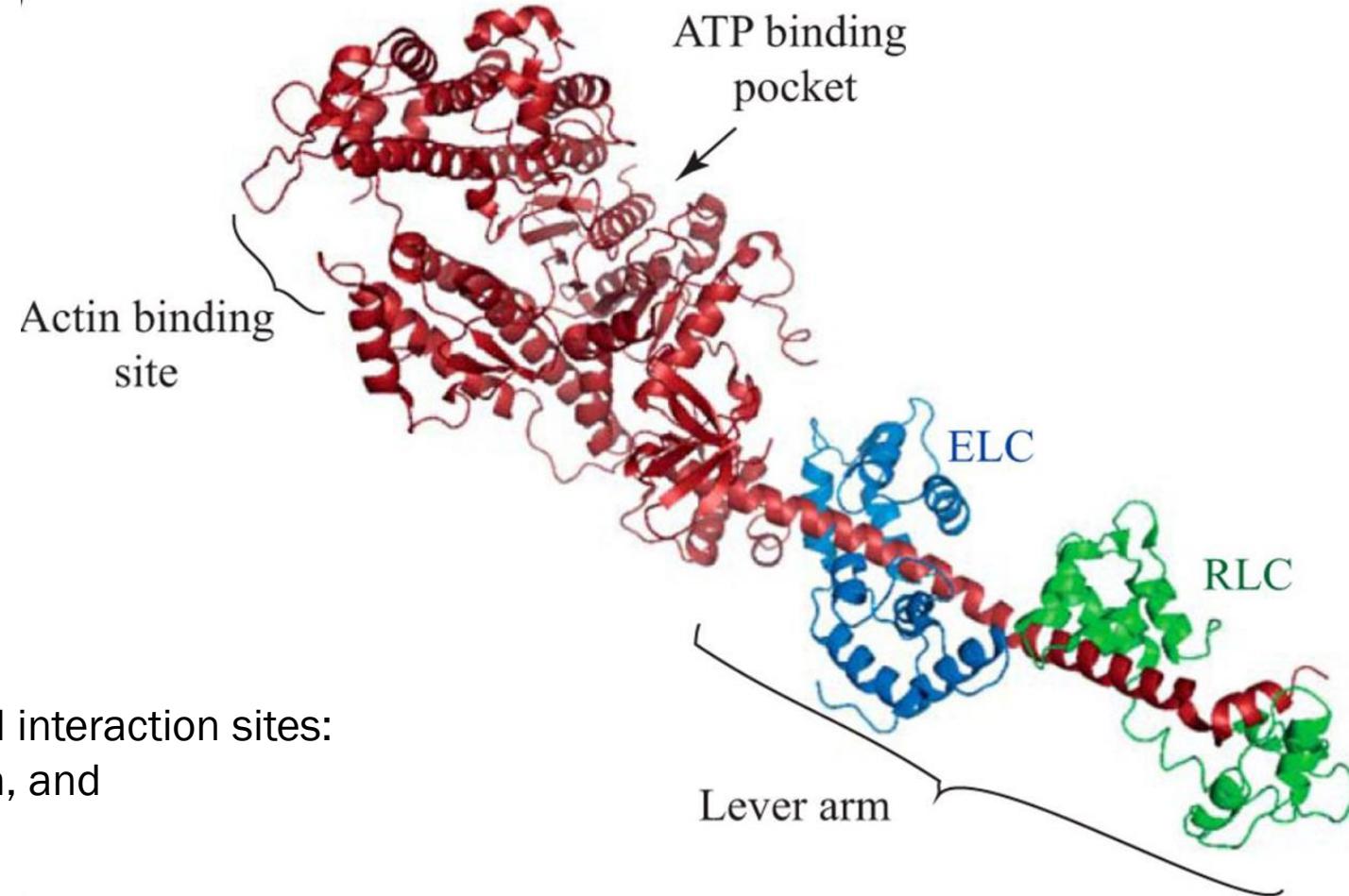
# MUSCLE PROTEINS – ACTIN AND MYOSIN

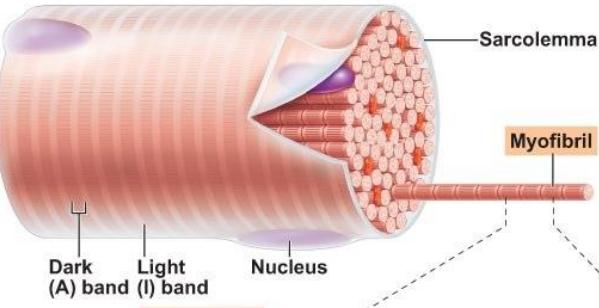


# ACTIN

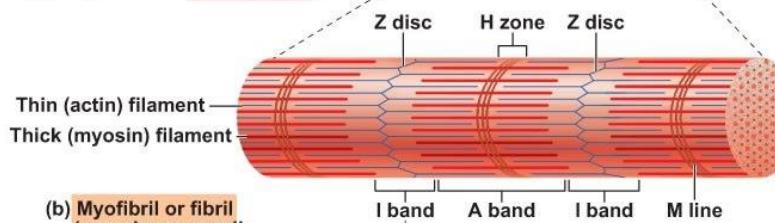


# MYOSIN

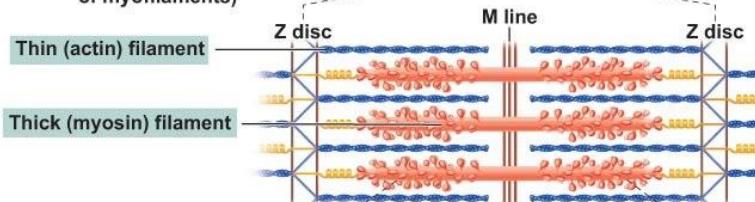




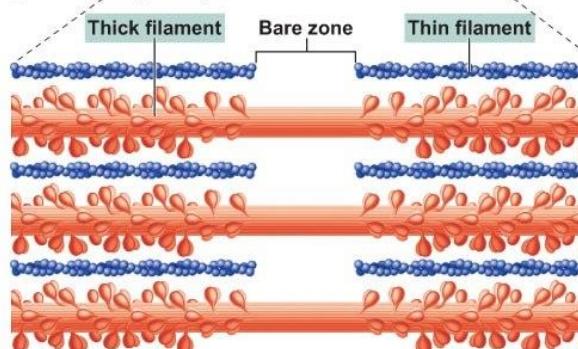
(a) Segment of a muscle fiber (cell)



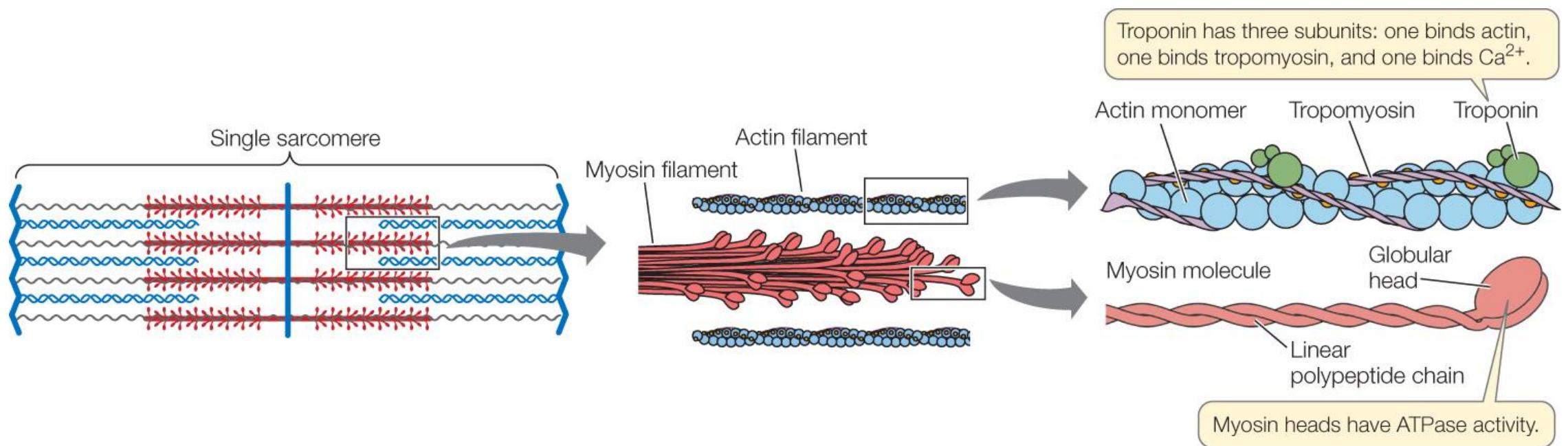
(b) Myofibril or fibril  
(complex organelle composed of bundles of myofilaments)

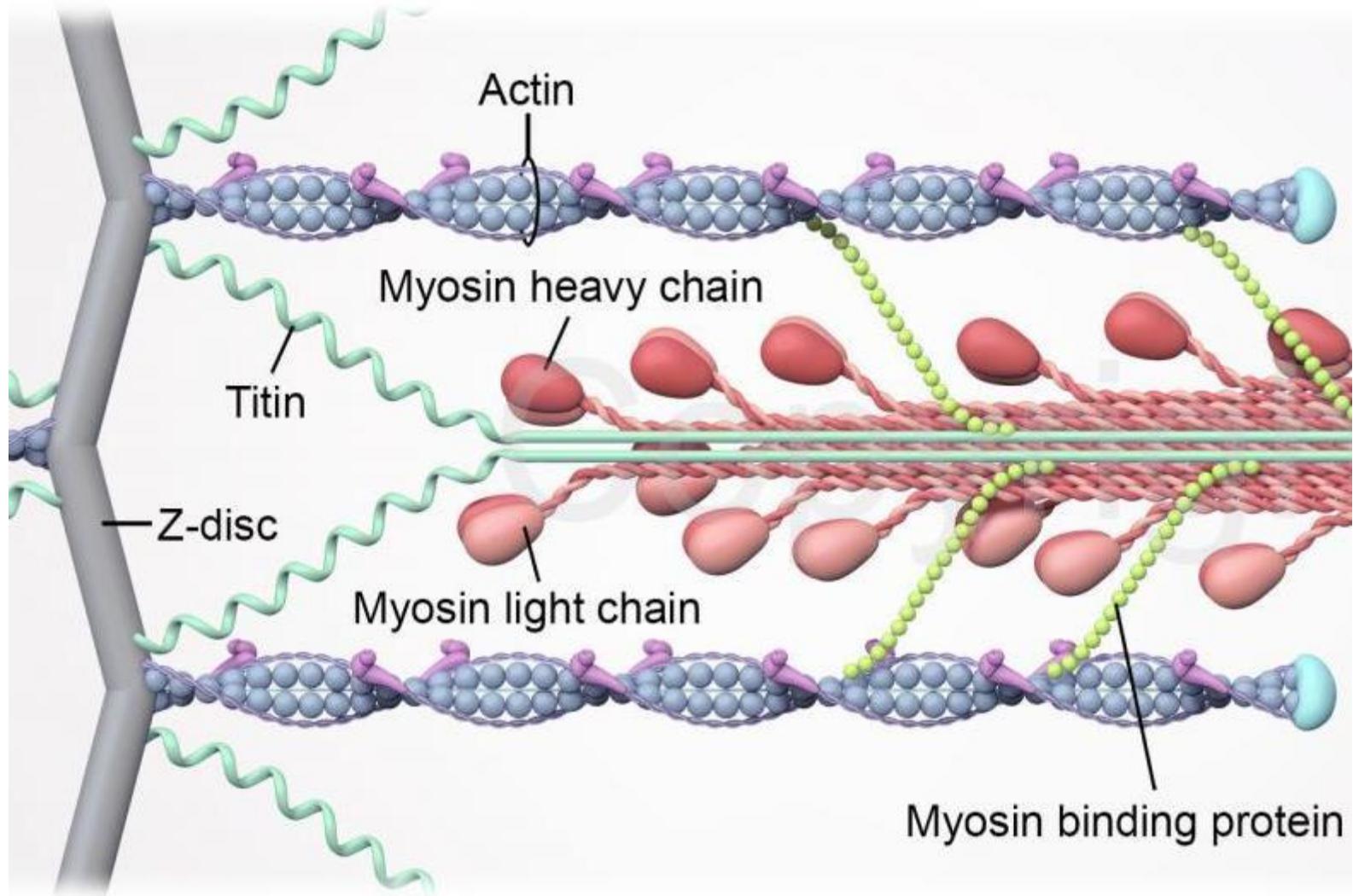


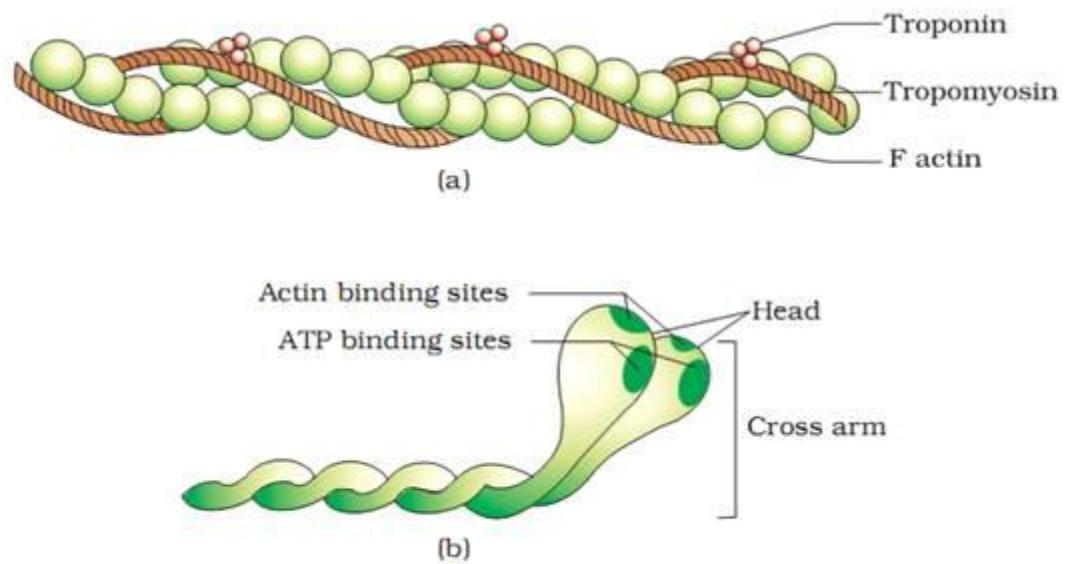
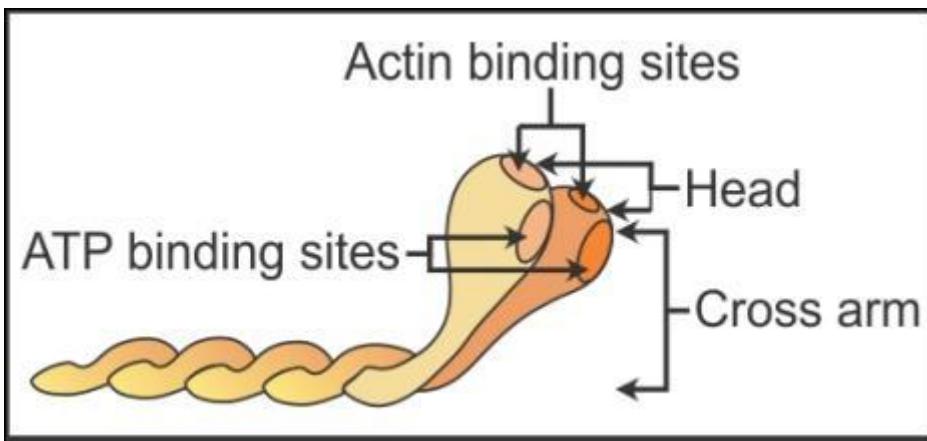
(c) Sarcomere (segment of a myofibril)



(d) Myofilament structure (within one sarcomere)





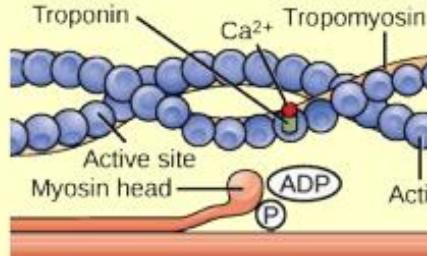


**Figure 20.3** (a) An actin (thin) filament (b) Myosin monomer (Meromyosin)

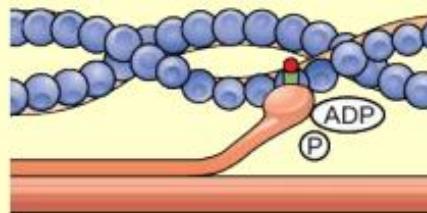
The myosin head is able to bind to the actin because a myosin-binding site is revealed once a calcium molecule binds to the troponin on the actin, causing a conformational change that allows for the myosin-binding site to be revealed and for the myosin head to bind.

Thus, calcium regulates muscle contraction.

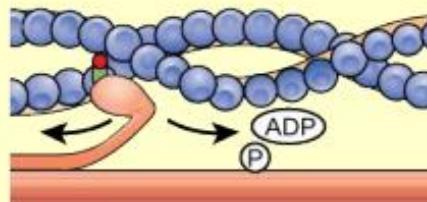
- ① The active site on actin is exposed as  $\text{Ca}^{2+}$  binds troponin.



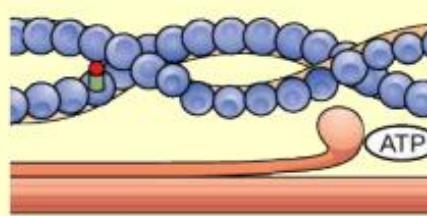
- ② The myosin head forms a cross-bridge with actin.



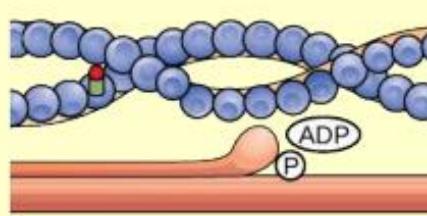
- ③ During the power stroke, the myosin head bends, and ADP and phosphate are released.



- ④ A new molecule of ATP attaches to the myosin head, causing the cross-bridge to detach.



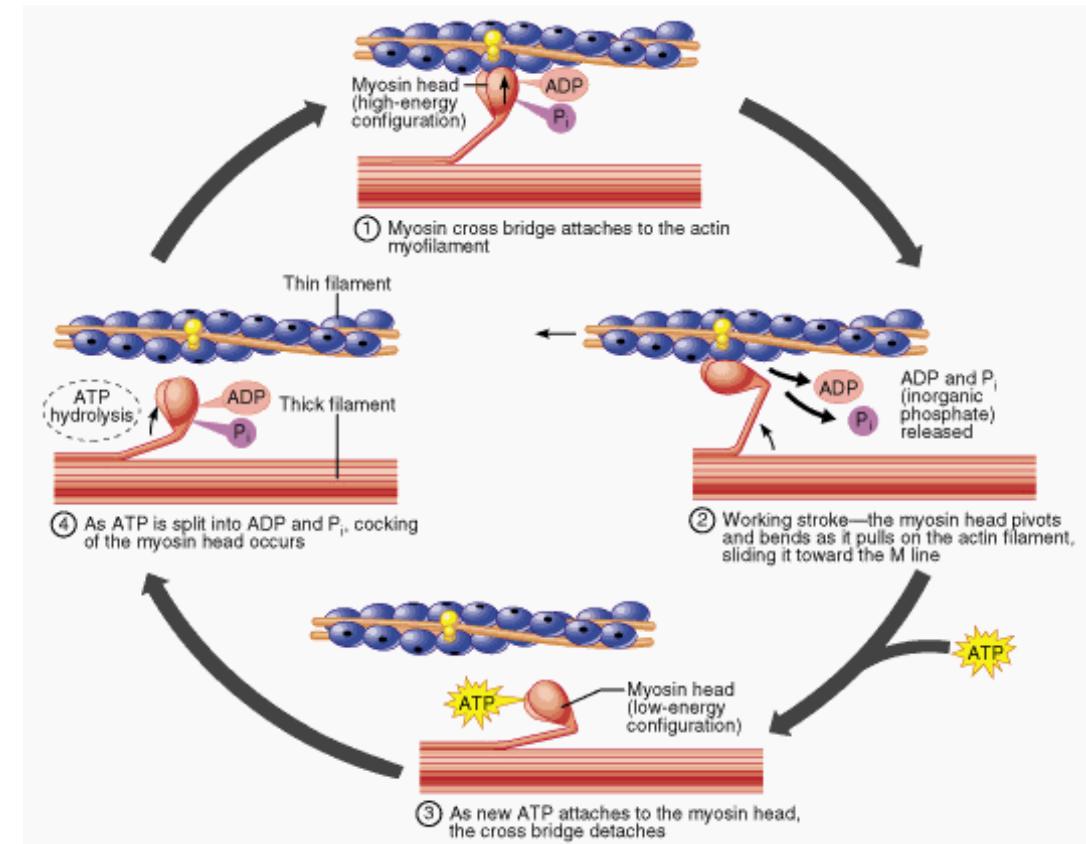
- ⑤ ATP hydrolyzes to ADP and phosphate, which returns the myosin to the "cocked" position.



ATP is necessary to release the myosin from the actin, which allows for the muscles to relax and then be ready to undergo another cycle of contraction and crossbridge forming

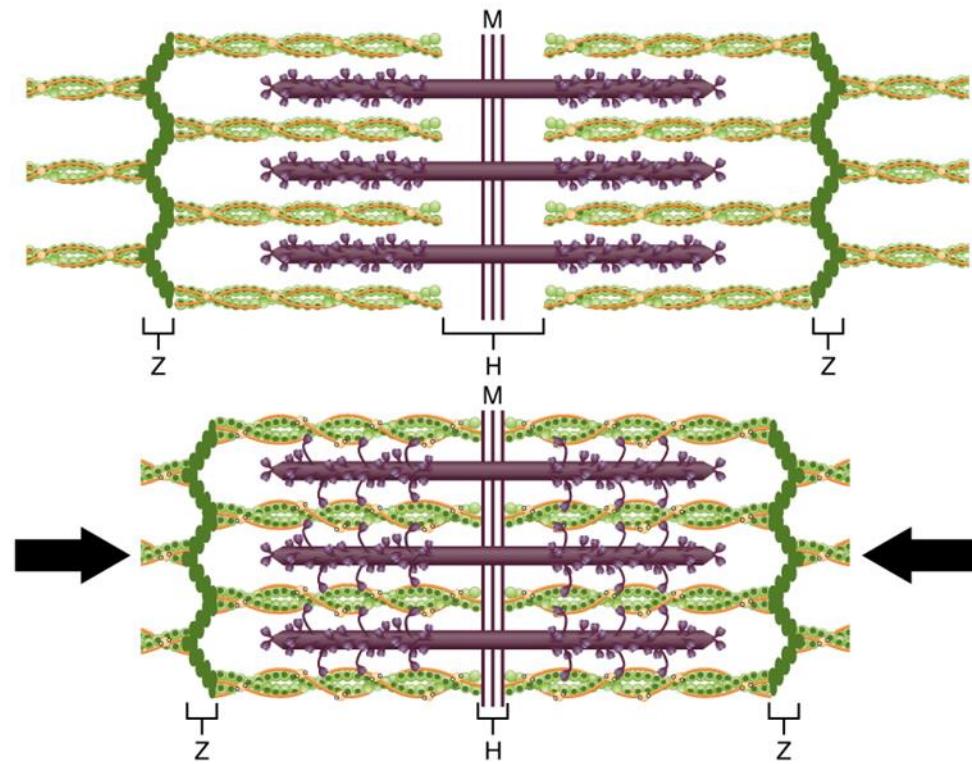
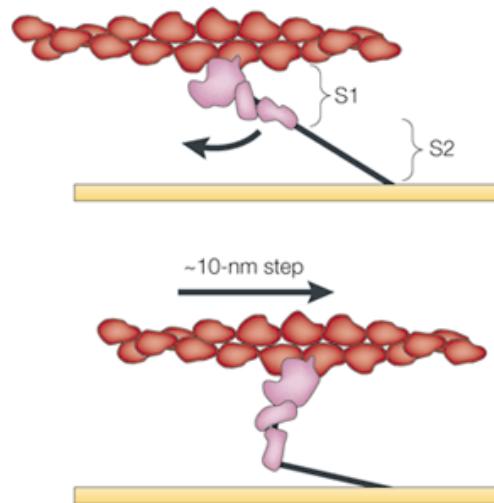
- (1) In the first step of the crossbridge cycle, the myosin head is attached to the actin filament with ADP + Pi separately also attached.
- (2) Once the myosin head binds, it “pivots” and a power-stroke is generated as the myosin causes the actin to shift. This causes for the ADP + Pi to be released.
- (3) A molecule of ATP comes in and binds to the myosin head, which it now able to do because the myosin is no longer bonded to the ADP + Pi molecules. The myosin head detaches from the actin because it is now bonded to the ATP.
- (4) ATP hydrolysis will then cause the myosin head to once again “cock” and be in the appropriate conformation to be ready to bind to “myosin-binding site” on the actin filament once again. As you can see, the ATP is now ADP + Pi.

The myosin head + ADP + Pi once again binds to the actin. The cycle begins.



Once the myosin head binds to the active site of the actin strand, it undergoes hydrolysis of ADP.

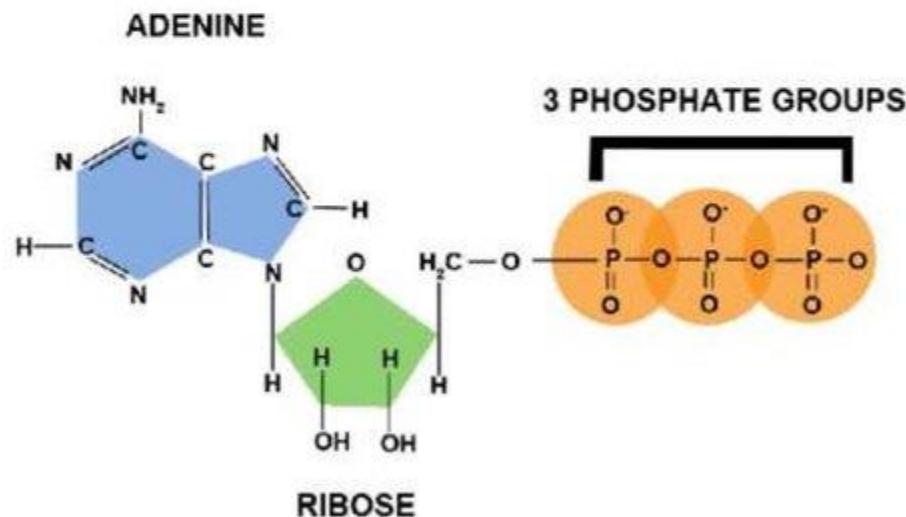
That hydrolysis reaction causes the protein to change its conformation.



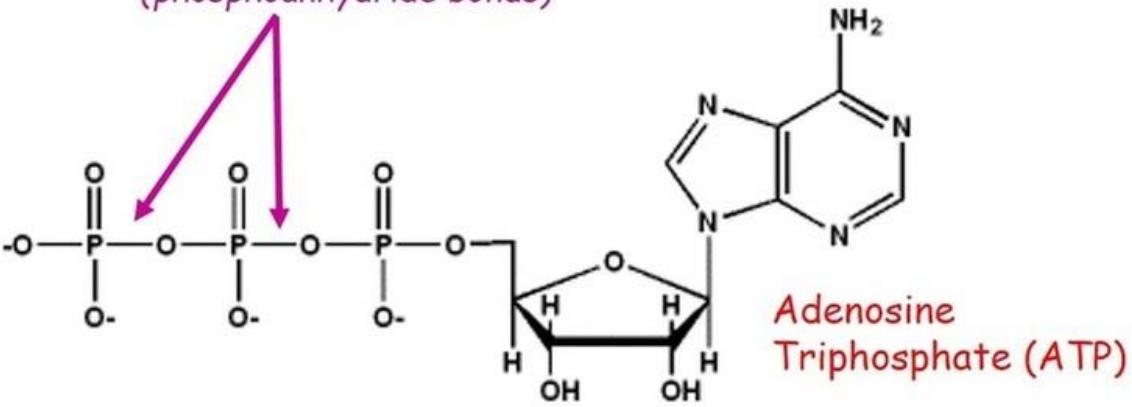
# Adenosine Triphosphate

- Adenosine triphosphate (ATP) is a complex organic chemical that participates in many processes. Found in all forms of life, ATP is often referred to as the "molecular unit of currency" of intracellular energy transfer.

An ATP Molecule

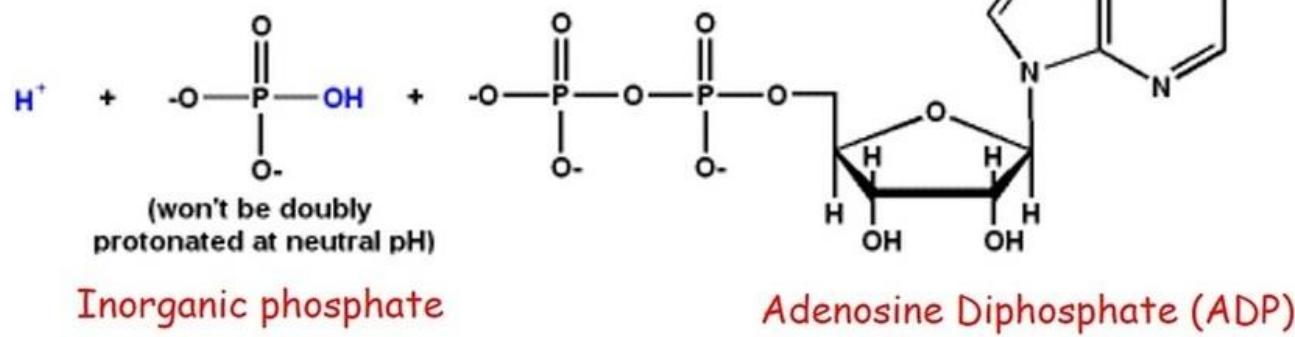


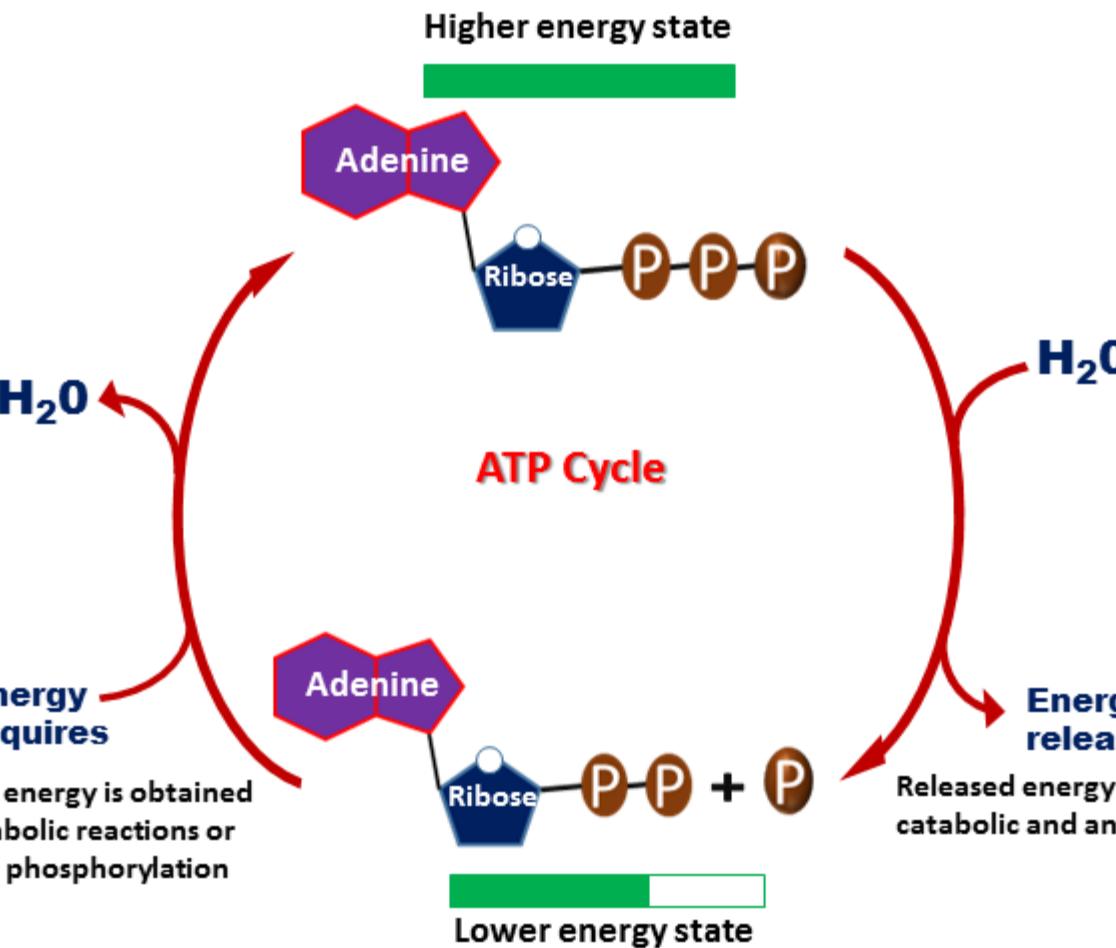
Two high energy bonds  
(phosphoanhydride bonds)



$\text{H}_2\text{O}$

Hydrolysis of the terminal phosphate group by ATPase releases of  $30.5 \text{ kJ/mol}^{-1}$  of free energy

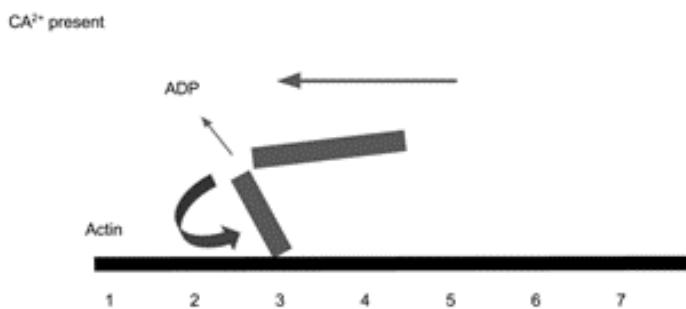
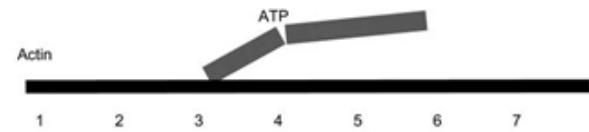
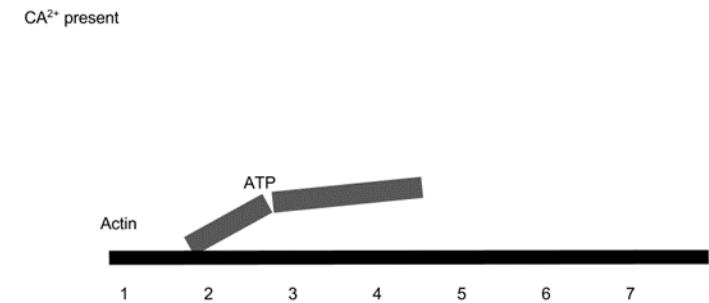
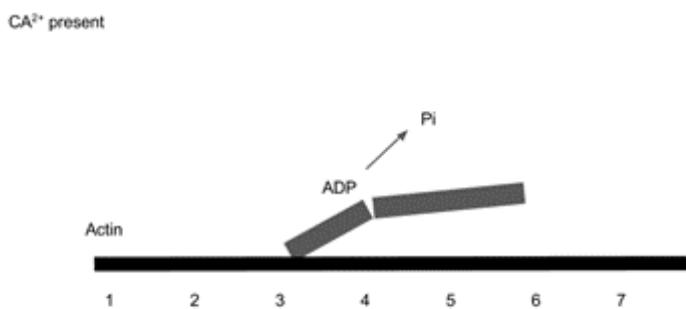
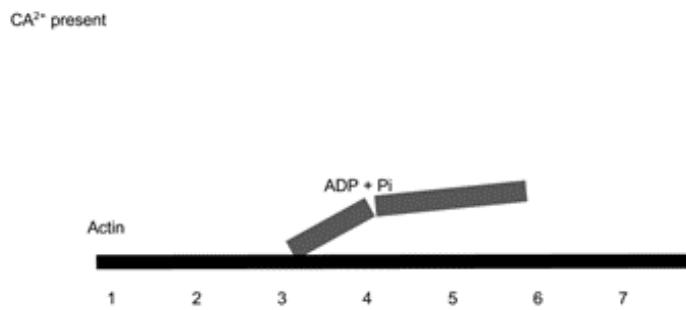
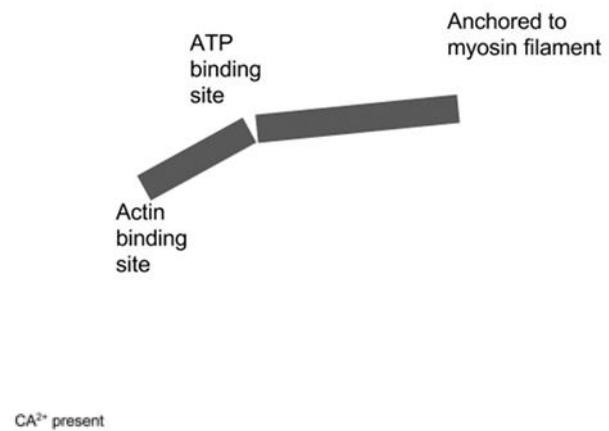


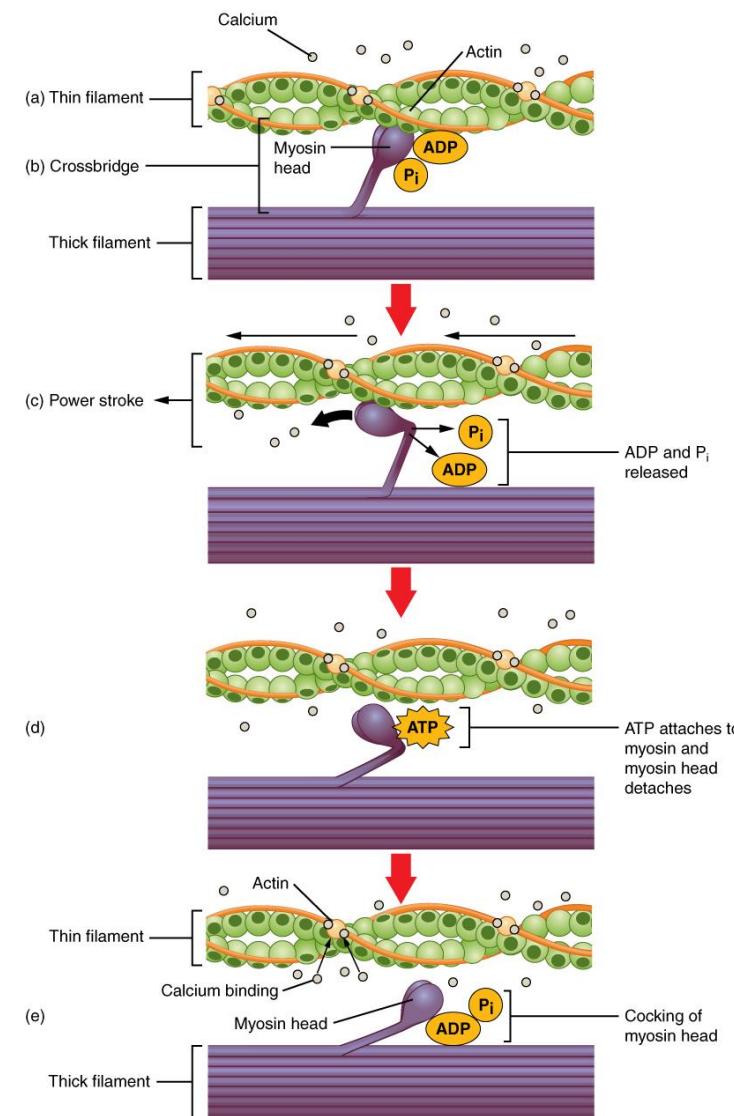
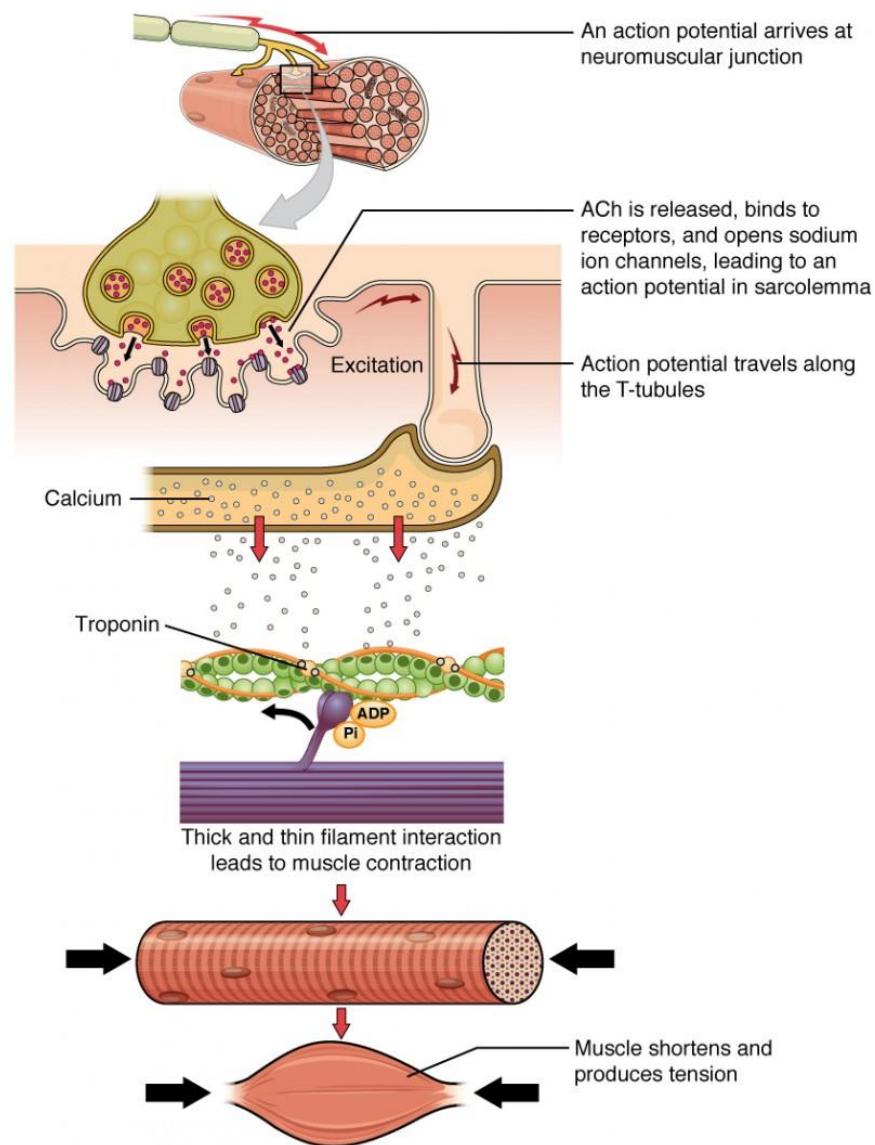


# ATP - HYDROLYSIS

- ATP, or adenosine triphosphate, is a chemical compound that is composed of three phosphate groups all bonded together in series, a ribose group, and an adenine.
- If you remove one of the phosphate groups, then you have adenosine diphosphate ( $ADP + P$ ) and if you were to removed two of the phosphate groups, then you have adenosine monophosphate.
- The **bonds between the phosphate groups** are of very high **energy** and, thus, when the **bonds** in between these groups is **broken**, then a lot of **energy** is then **released**.
- So, in the chemical reaction  $ATP + H_2O \rightarrow ADP + P + H^+$  (also known as ATP Hydrolysis), on the right side of the reaction you also have a lot of energy being released.
- In our muscles, we have hundreds of thousands of muscle fibers, and these fibers are composed of myofibrils, which in turn have structures called sarcomeres. The structure of a sarcomere is shown below:

# ACTIN BINDING



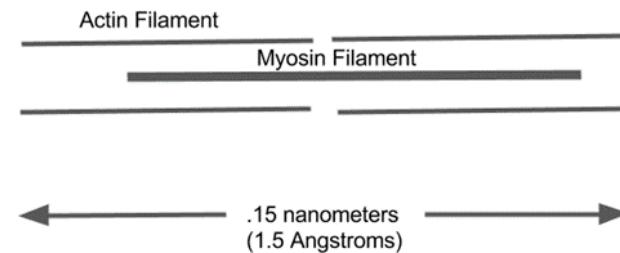
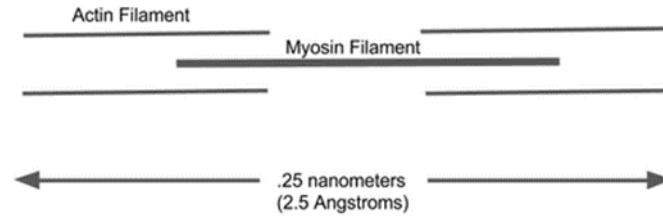


# ACTIN BINDING

- ATP that is bound to the ATP binding site is hydrolyzed, but both the ADP and orthophosphate (Pi) stay bound. There is no change in the shape of the protein complex at this point.
- Orthophosphate is released, leaving only bound ADP. This still doesn't result in a conformational change in the protein complex.
- Once ADP dissociates, the protein complex returns to the conformation state that it is normally in when ATP is not bound. This pulls the anchored end along the actin filament. This is analogous to bending a spring and letting it snap back into its lowest energy position. Thus, the [chemical energy of ATP hydrolysis is transduced into mechanical work](#).
- ATP then attaches to the binding site again, and the actin binding site releases, allowing the complex to return to its original conformation. It then binds to a new site that is closer to the z-disc of the Sarcomere (position 2 instead of position 3, which is where it was before).
- If ATP doesn't bind, myosin won't release, this is what causes rigor mortis in death. Calcium is available but ATP hydrolysis has reached equilibrium and is in too relatively low a concentration to allow the myosin crossbridges to detach. So, we can see here that ATP is necessary for the cycle of muscle contraction to complete - without ATP, the myosin head could not unbind from the actin thread. In rigor mortis, which is a condition that takes place soon after death, the body is no longer making ATP which is why the muscles are incredibly stiff; the myosin heads cannot unbind actin.
- But how does the activity result in contraction? Many of these reactions, in a specific arrangement, and happening in concert results in the transduction of huge amounts of mechanical energy.

# MUSCLE CONTRACTION

- The arrangement depends on the myosin filaments (with many myosin head groups) overlapping with actin filaments that are fixed at their ends.
- The concerted action of the myosin heads pulling in opposite directions against the stationary actin filaments causes the whole apparatus to shorten. Skeletal muscles are comprised of end to end units like this that can shorten by about 30% of the original muscle fiber length. Smooth muscles can shorten by 70% because the filaments are arranged differently.



# QUATERNARY STRUCTURE

- Protein Chains Combine to Make Protein Complexes
- Quaternary structure refers to the way in which the subunits of such proteins are assembled in the finished protein.
- Secondary and tertiary structures are determined by a protein's sequence of amino acids, or primary structure. All proteins have primary, secondary and tertiary structure.
- Some proteins are made up of more than one amino acid chain, giving them a quaternary structure. These multi-chain proteins are held together with the same forces as the tertiary structure of individual protein chains (hydrophobic, hydrophilic, positive/negative and cysteine interactions). Sometimes the various protein chains in a protein complex are identical and other times they are each unique.
- Some proteins contain two or more polypeptide chains that may be structurally identical (homomeric) or totally unrelated (heteromeric).
- Each chain forms a 3D structure called subunit.
- According to the number of subunits : dimeric, trimeric, ... or multimeric.
- Subunits may either function independently of each other, or work cooperatively

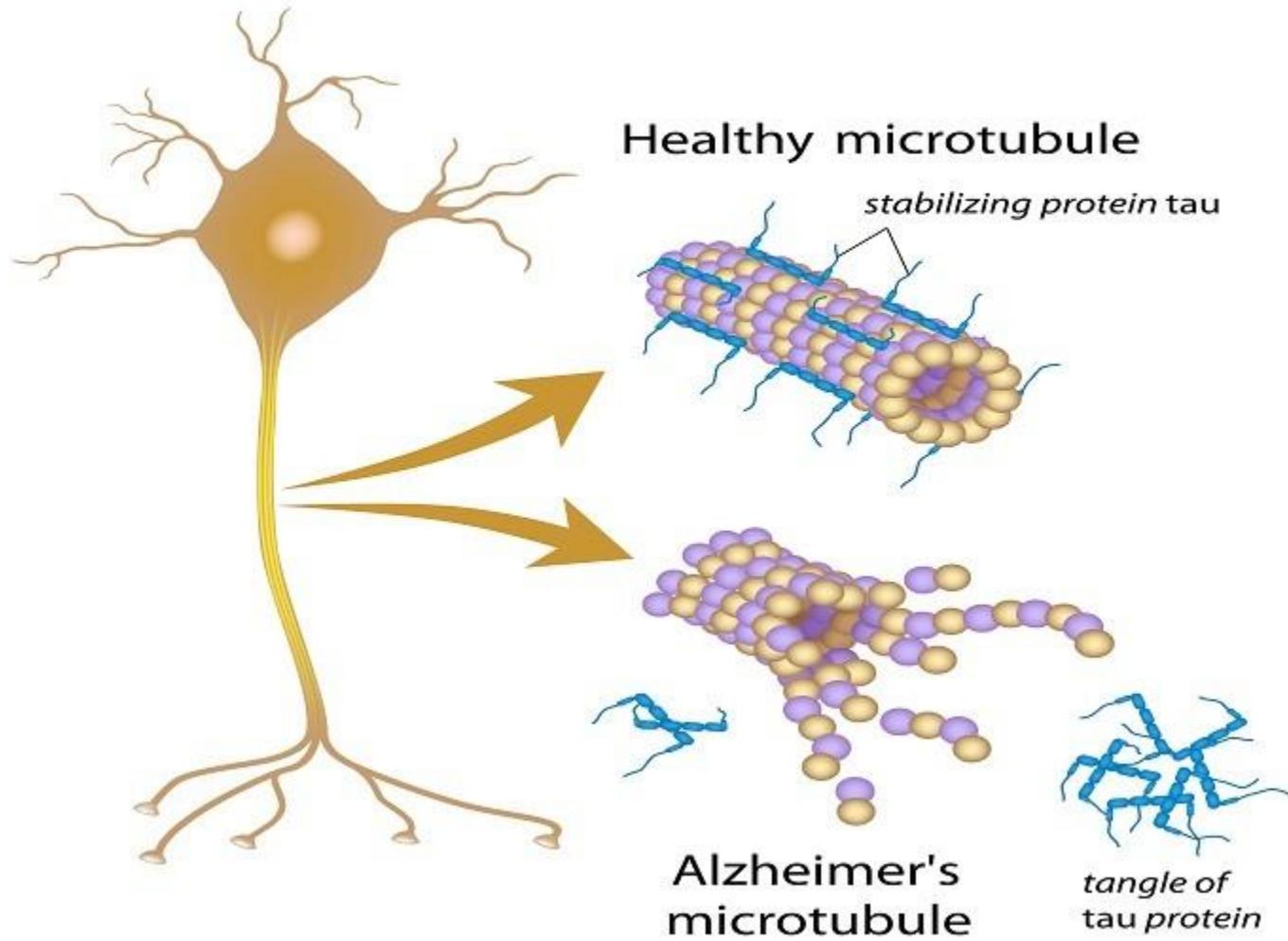
# MISFOLDING

Every protein must fold to achieve its function.

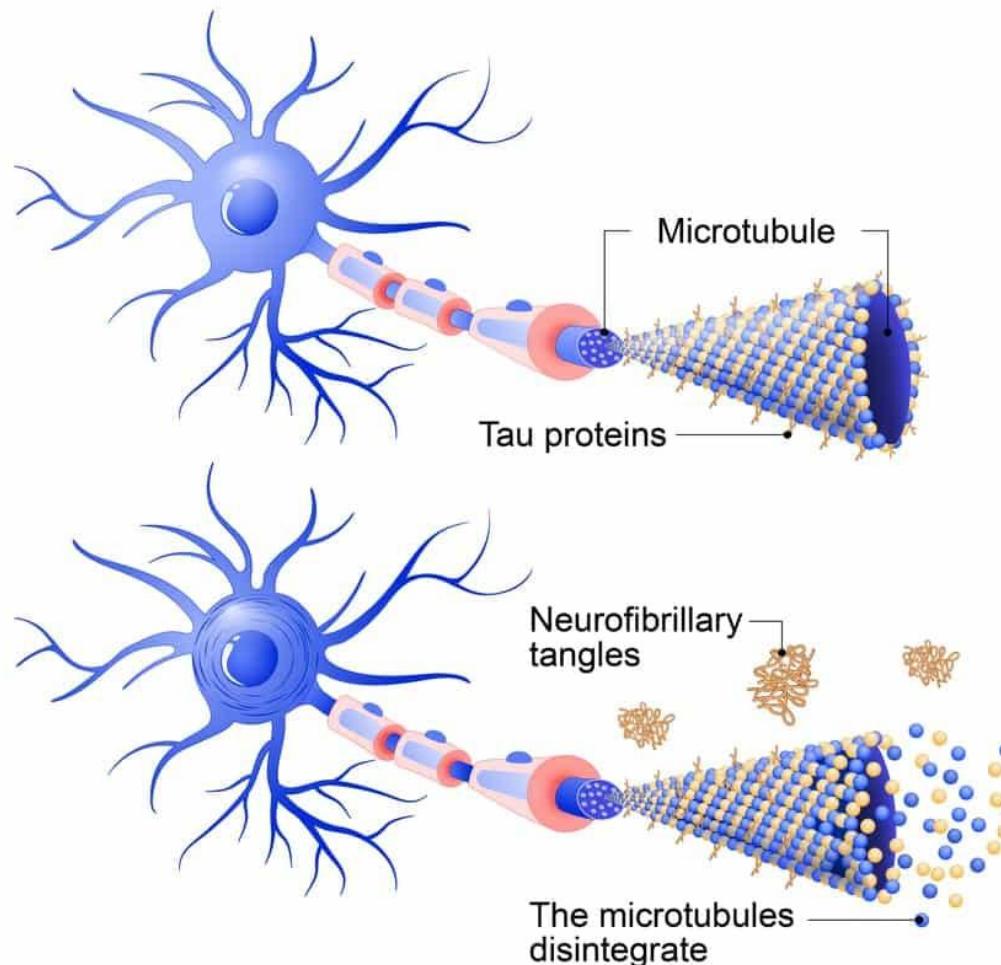
Abnormal folding/misfolding of proteins leads to a number of disorder /diseases in humans

e. g. amyloid protein when misfolded leads to Alzheimer's disease

- $\beta$  amyloid protein is a misfolded protein that aggregates outside neurons, it interferes with the neuron's ability of sending messages. It forms fibrous deposits or plaques in the brains of Alzheimer's patients.



## HEALTHY NEURON

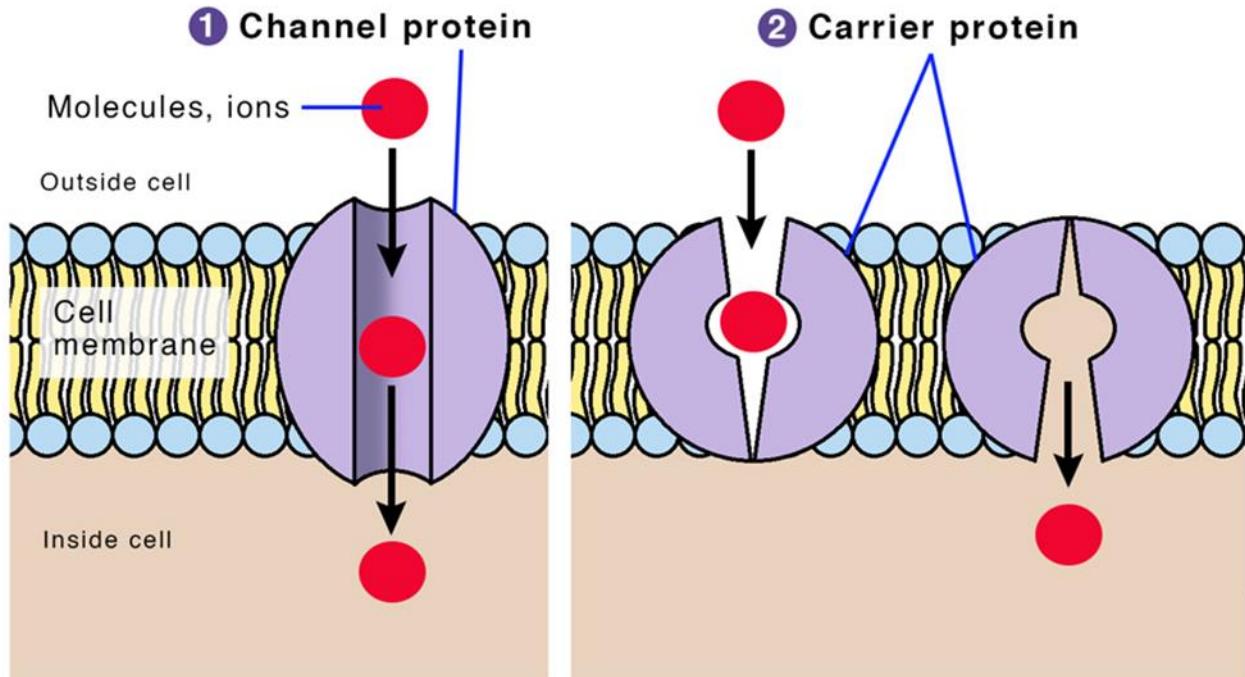


## ALZHEIMER'S DISEASE

# CHANNEL/CARRIER PROTEINS

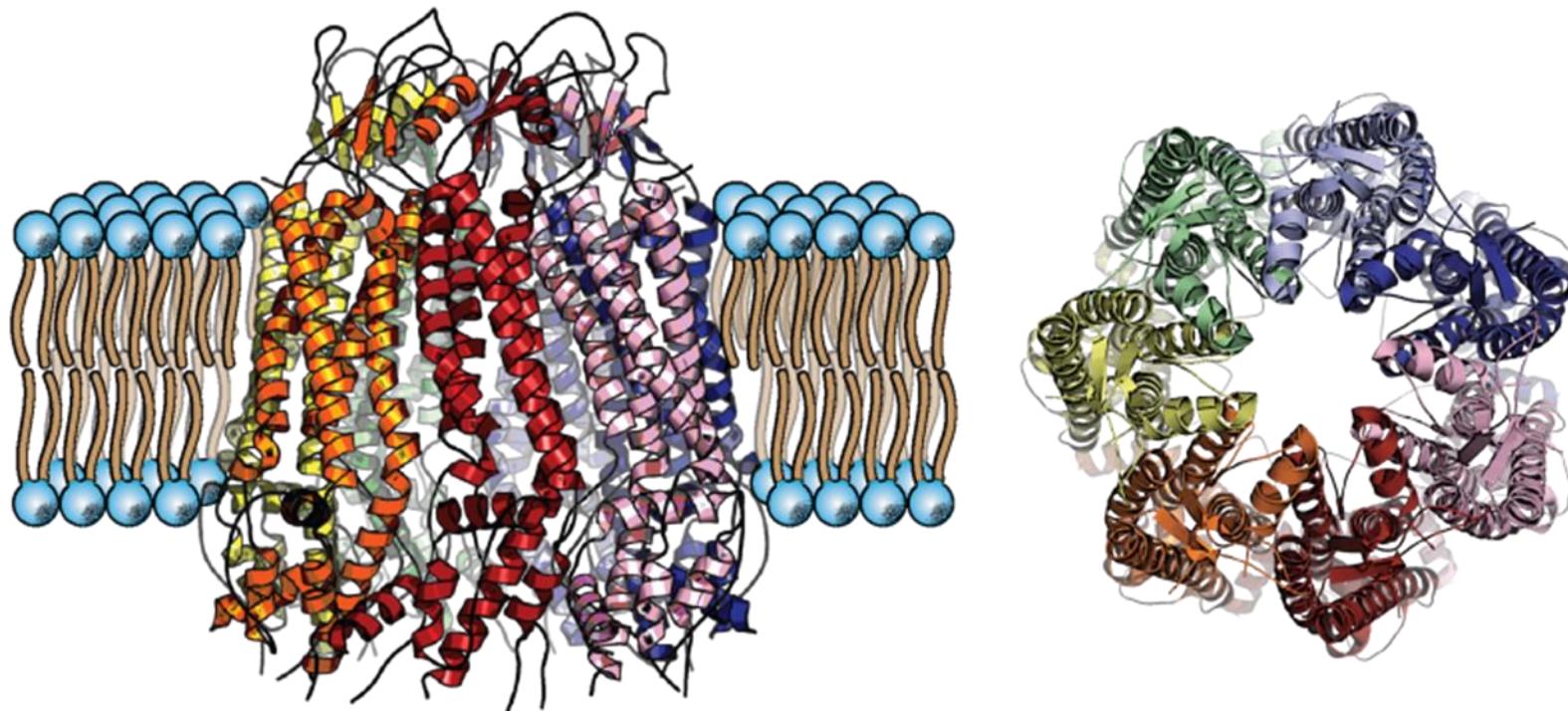
- Each channel or carrier protein acts as a gate which can open and close.
- Each transport protein is specific for the ion or molecule to be transported.

## Types of Transport Proteins



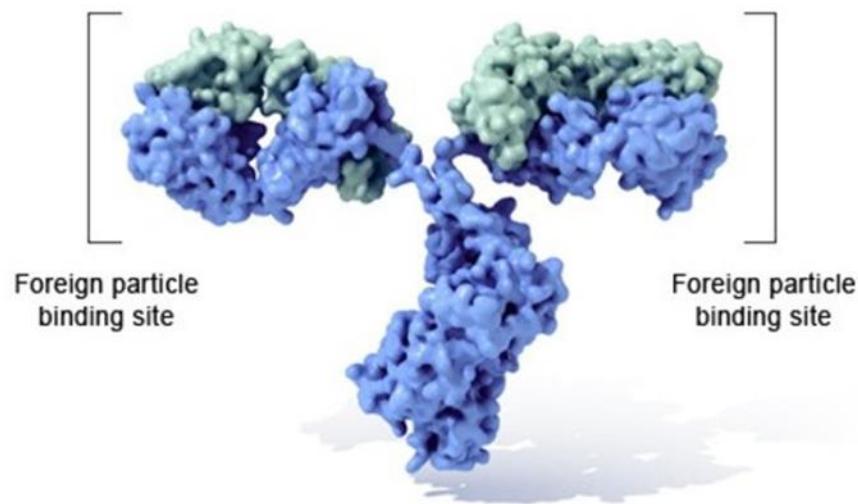
---

# CHANNEL PROTEINS



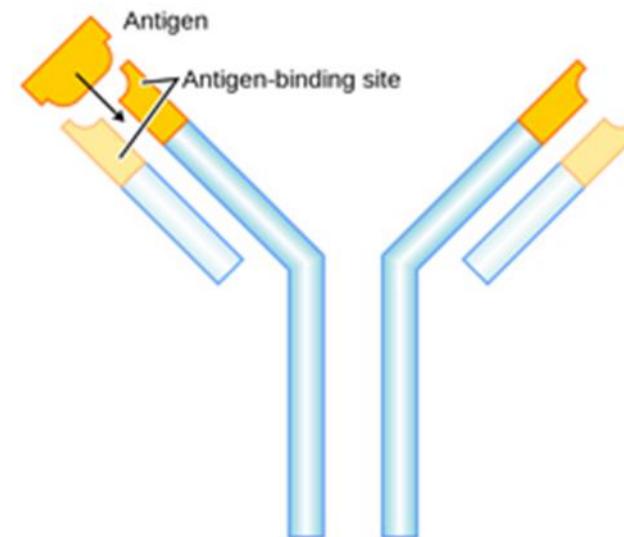
# ANTIBODIES

Immunoglobulin G (IgG)

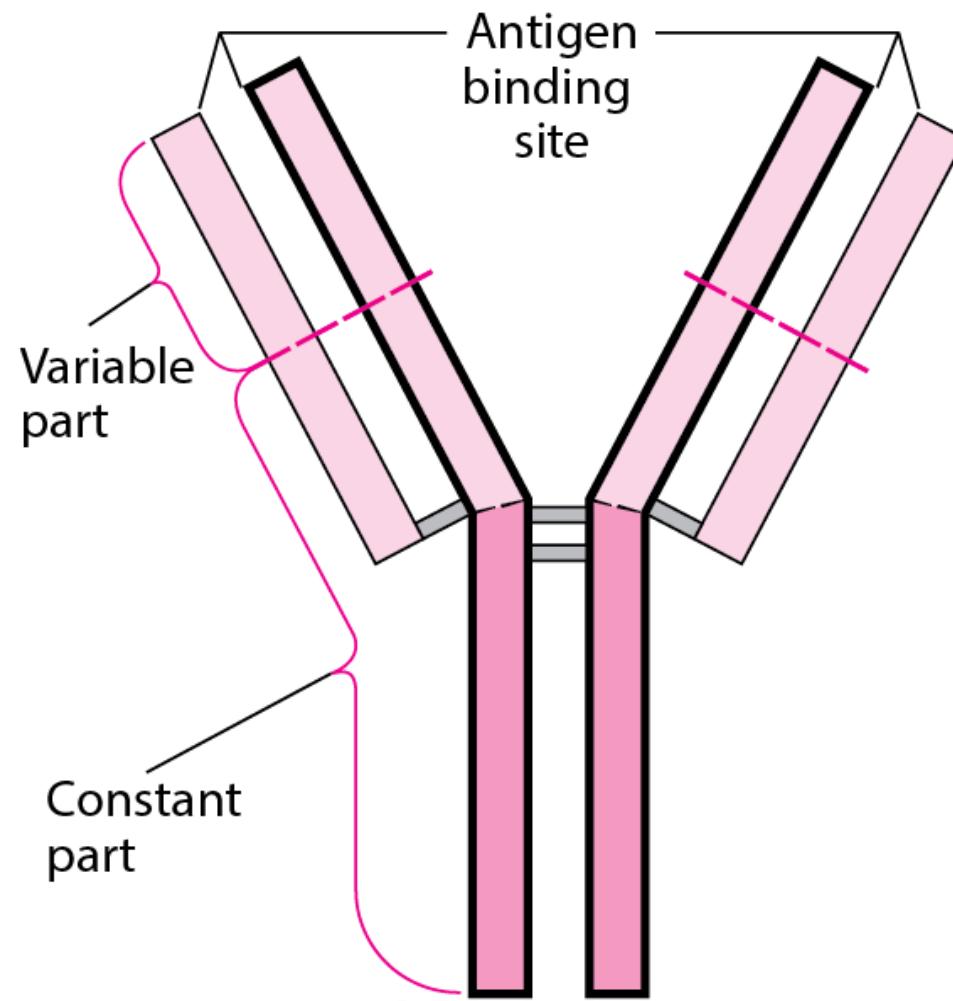


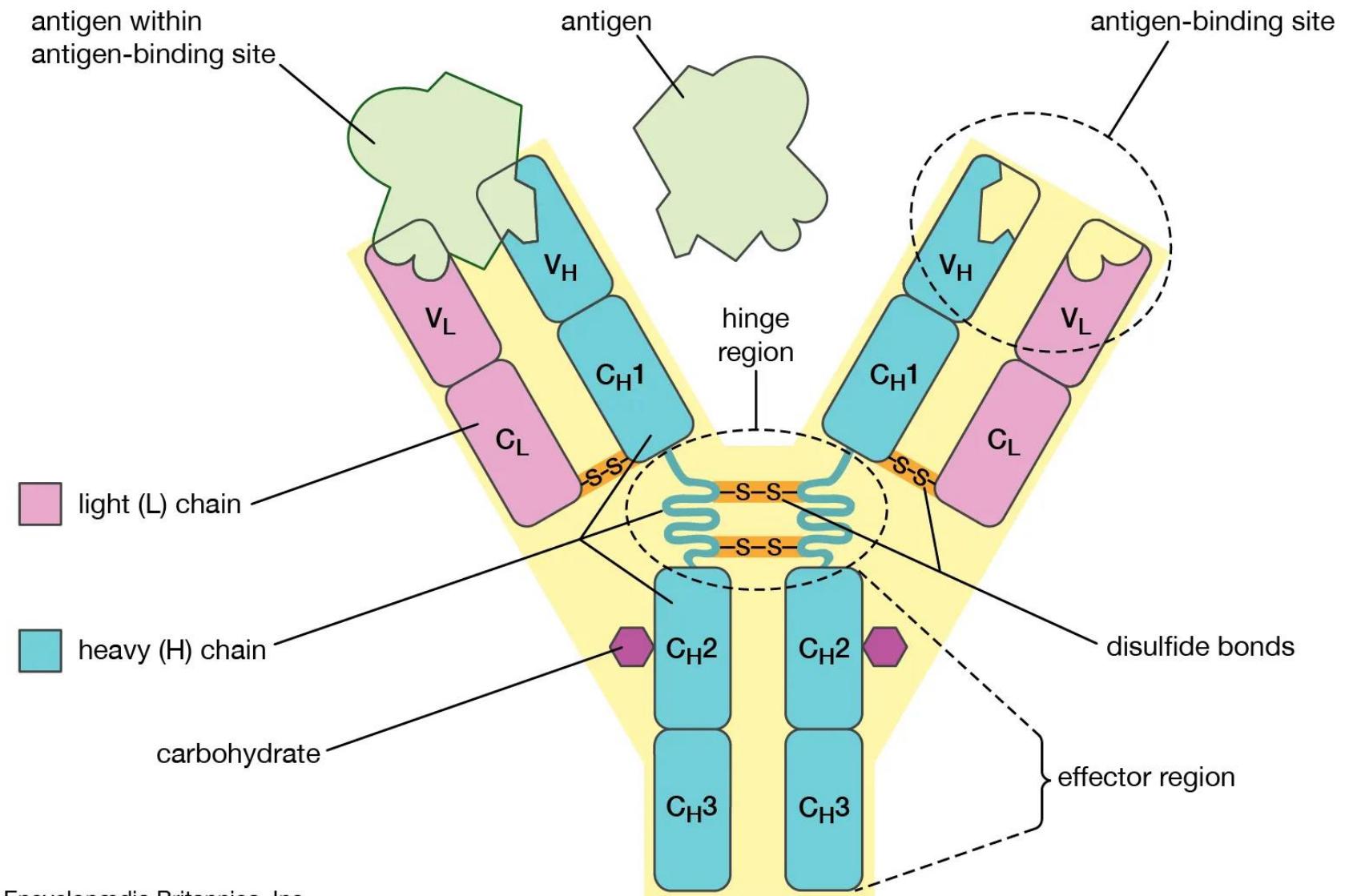
U.S. National Library of Medicine

Antigens

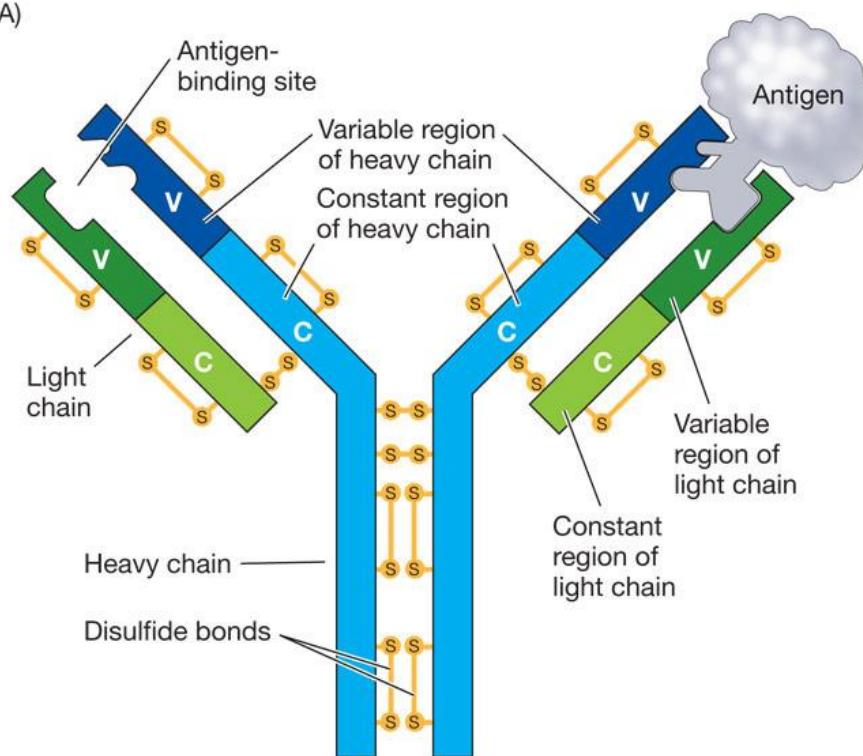


Antibody

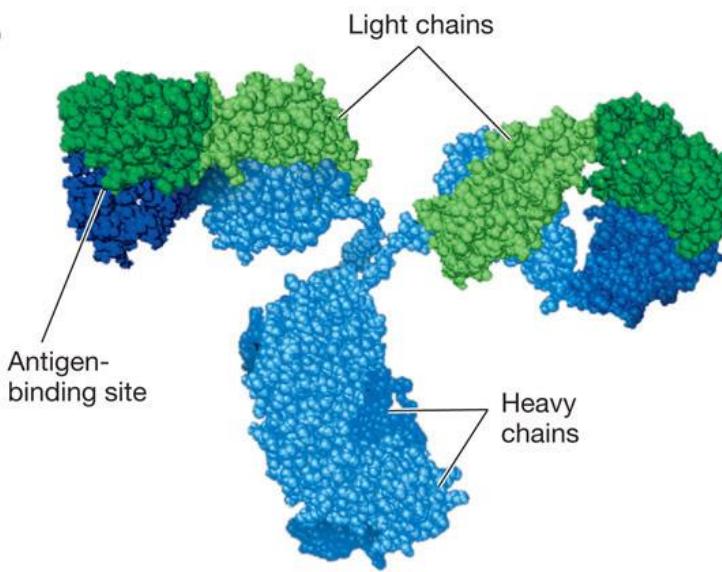


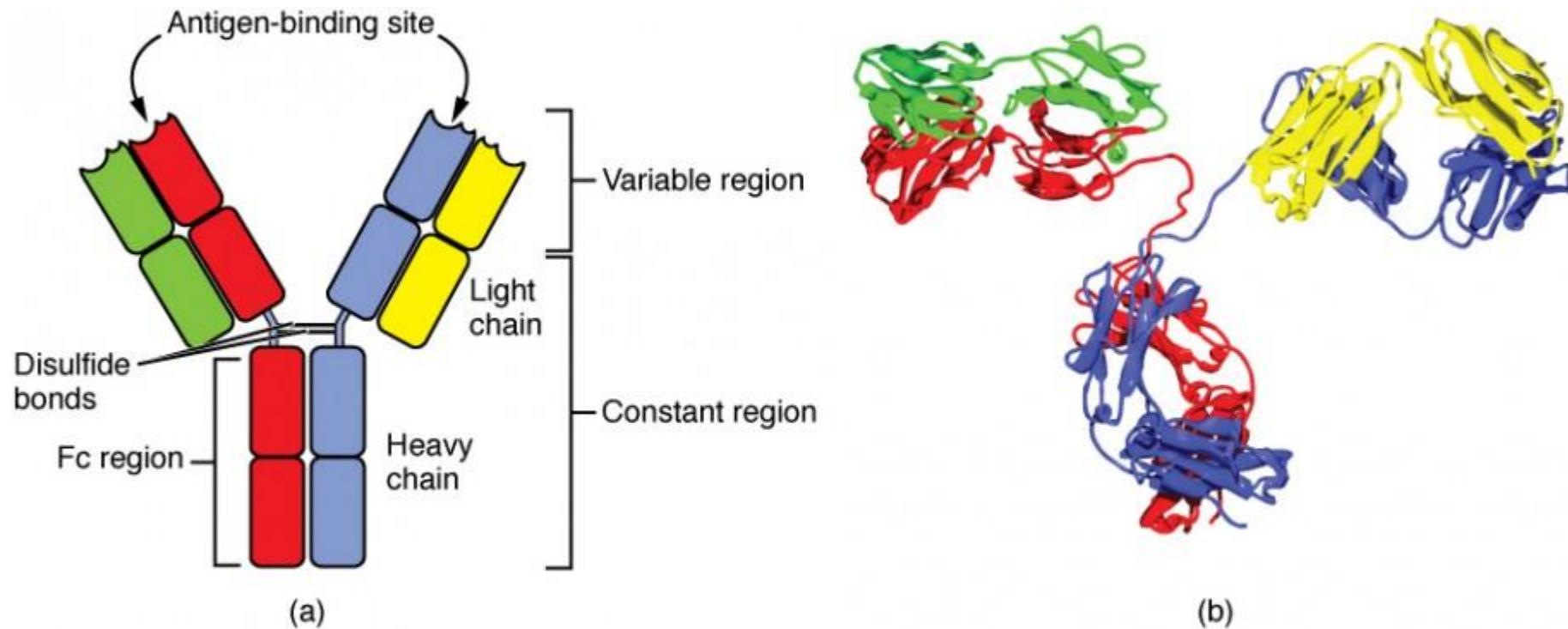


(A)



(B)

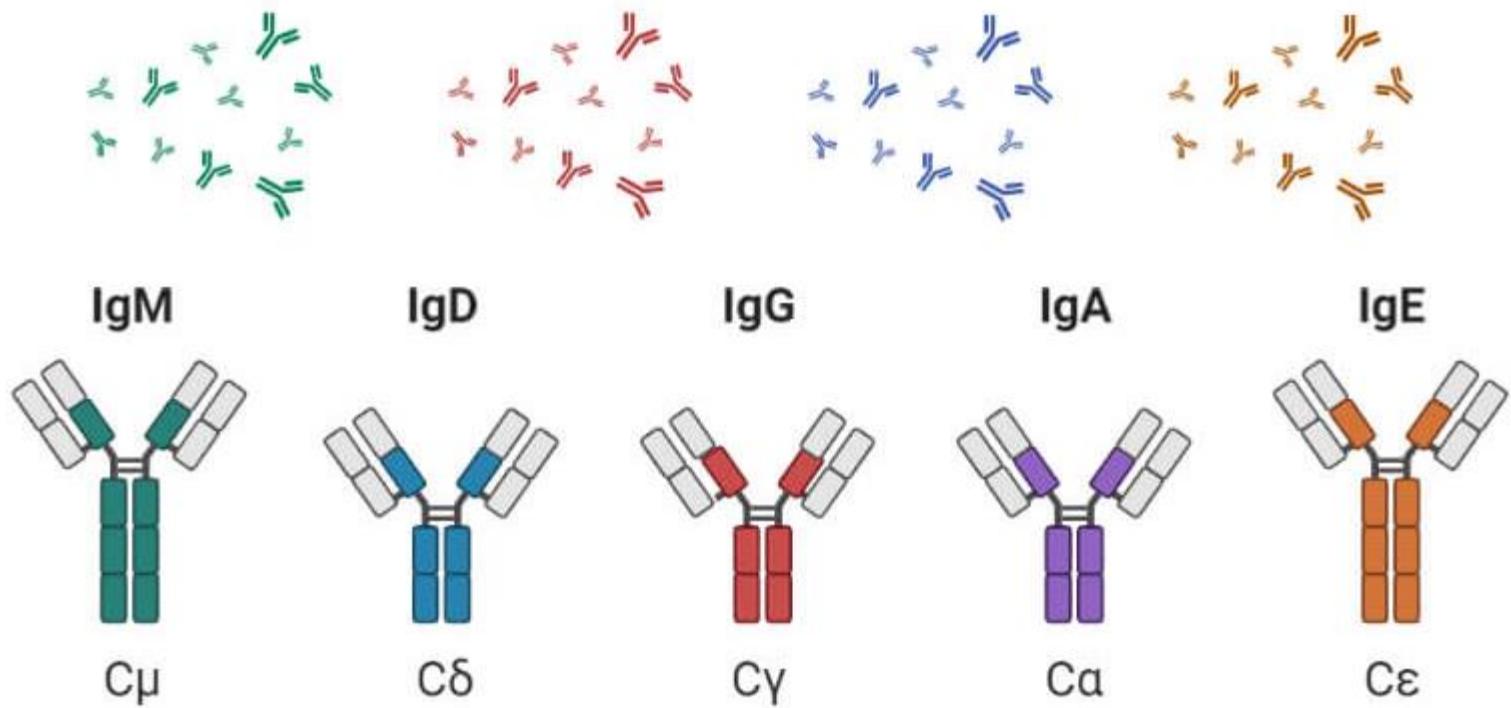


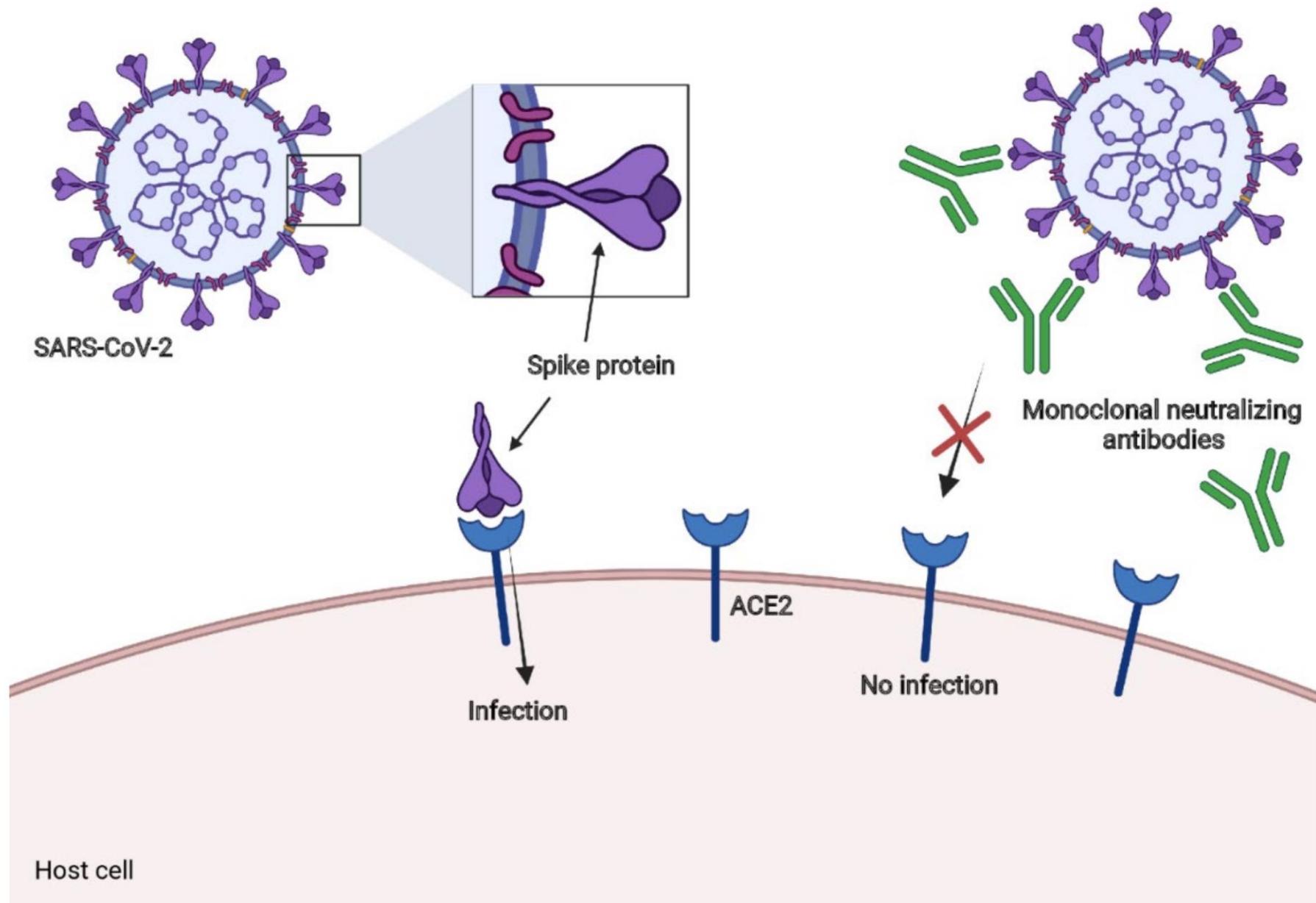


antibody molecule: the two heavy chains are colored red and blue and the two light chains green and yellow

# HUMAN ANTIBODIES

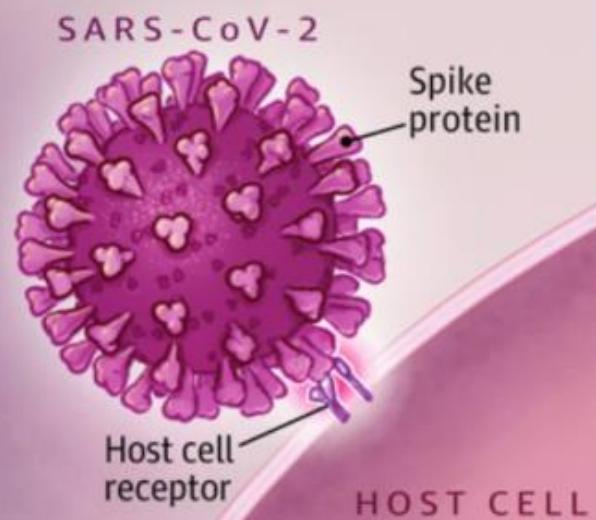
- Human antibodies are classified into five isotypes (IgM, IgD, IgG, IgA, and IgE) according to their heavy (long) chains, which provide each isotype with distinct characteristics and roles.



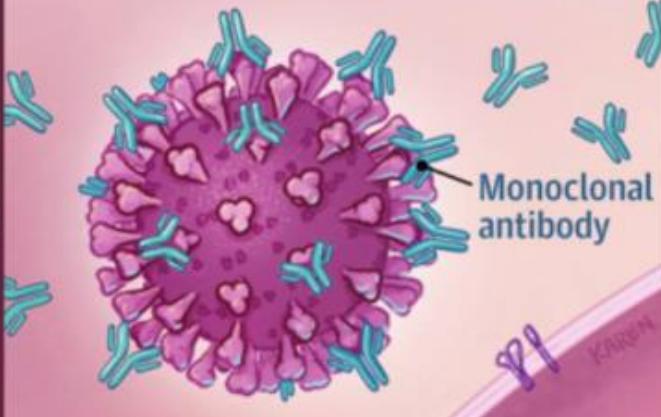


**Monoclonal antibodies are a therapy developed to treat viral infections including COVID-19.**

SARS-CoV-2 uses a spike protein to attach to and enter human cells, which allows it to cause infection.



Monoclonal antibodies bind to the spike protein, prevent the virus from attaching to human cells, and tag it for destruction.



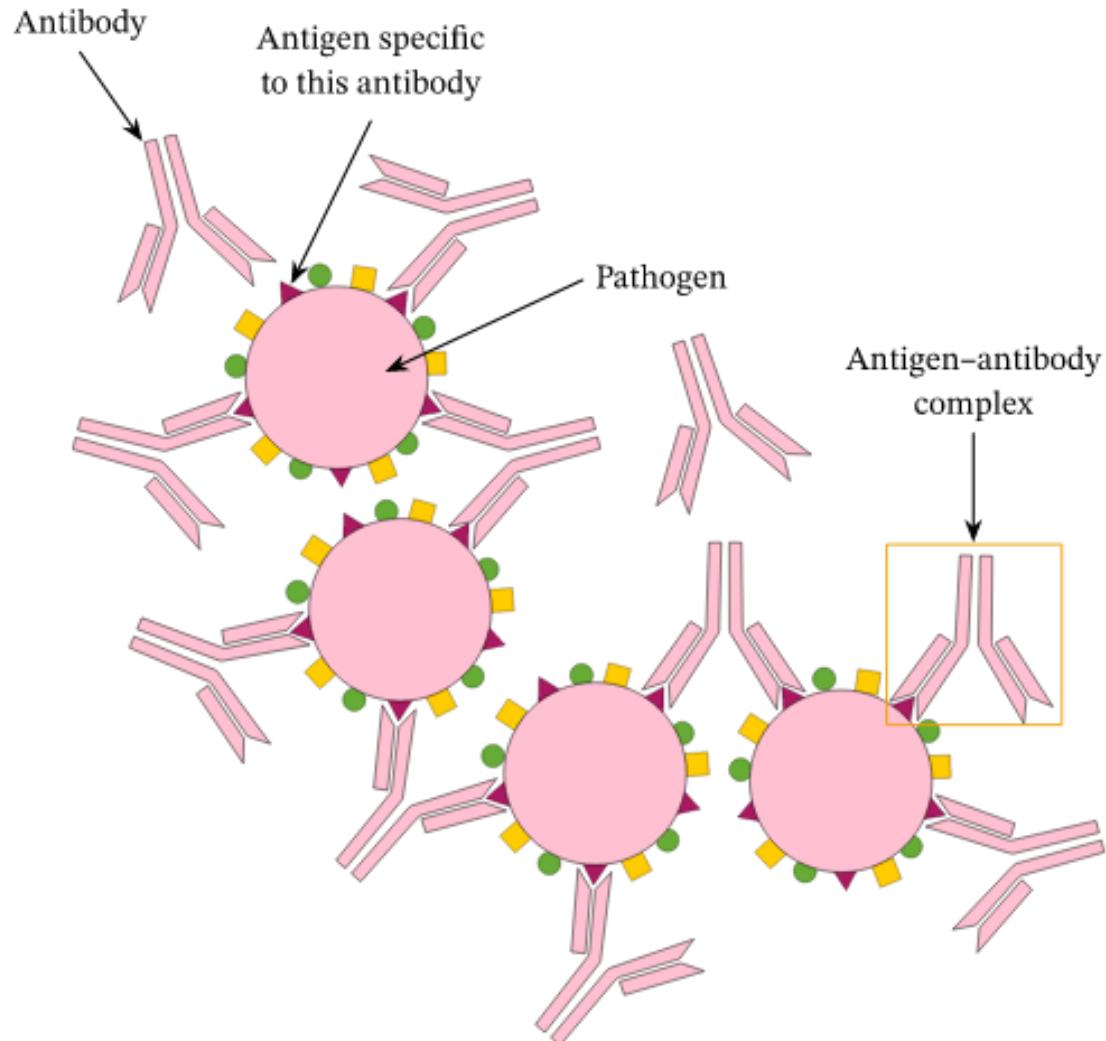
This may prevent development of severe COVID-19.

## Lock and Key Fit

---

- Antigen combines with antibody based on complementary structure. The better the “fit” the more tight the bond between the two.





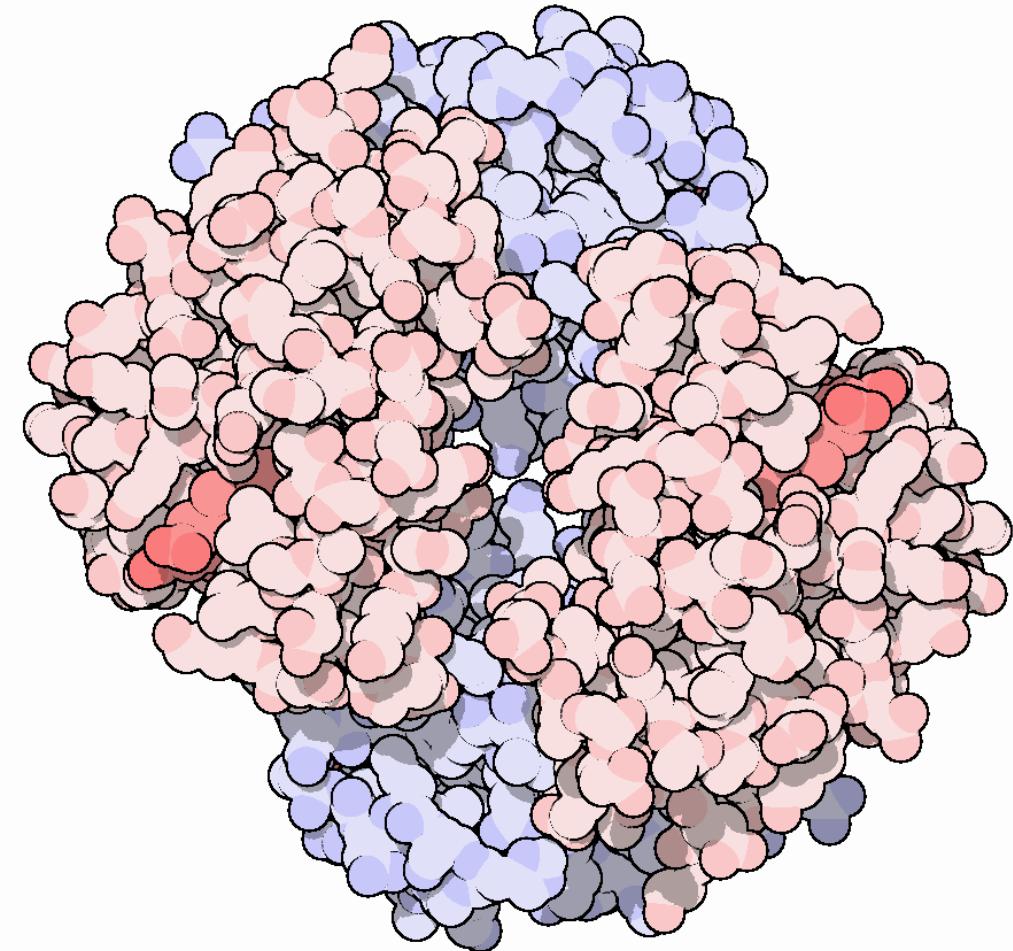
## HEMOGLOBIN, with hemes in red

Hemoglobin uses a change in shape to increase the efficiency of oxygen transport.

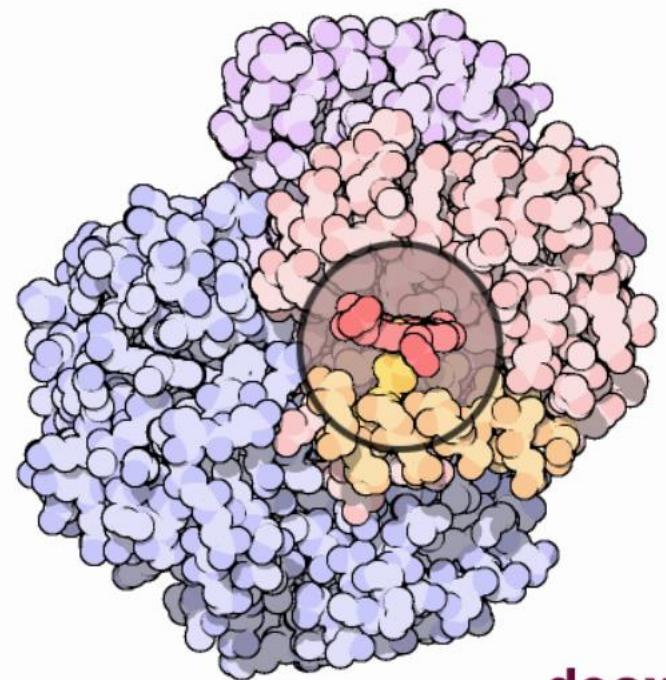
Red Blood, Blue Blood - Ever wondered why blood vessels appear blue?

Oxygenated blood is bright red: when you are cut, the blood you see is brilliant red oxygenated blood. Deoxygenated blood is deep purple.

Hemoglobin is the protein that makes blood red. It is composed of four protein chains, two alpha chains and two beta chains, each with a ring-like heme group containing an iron atom. Oxygen binds reversibly to these iron atoms and is transported through blood.

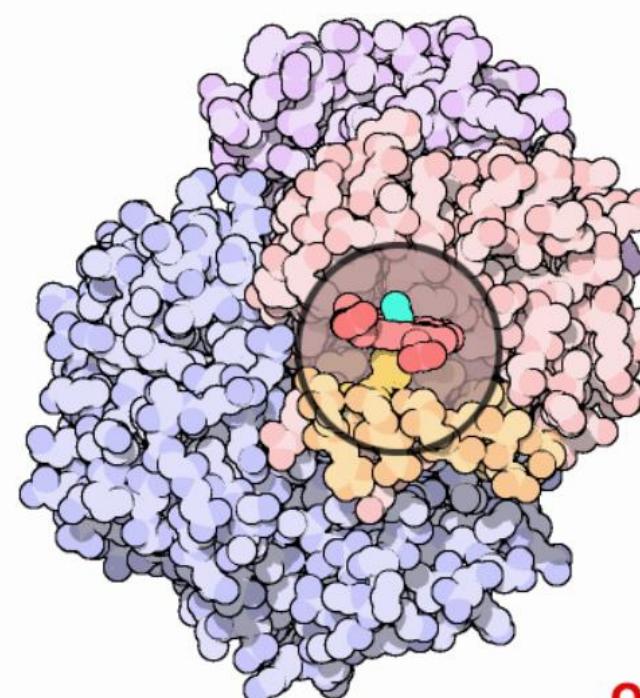


# HEMOGLOBIN



**deoxy**

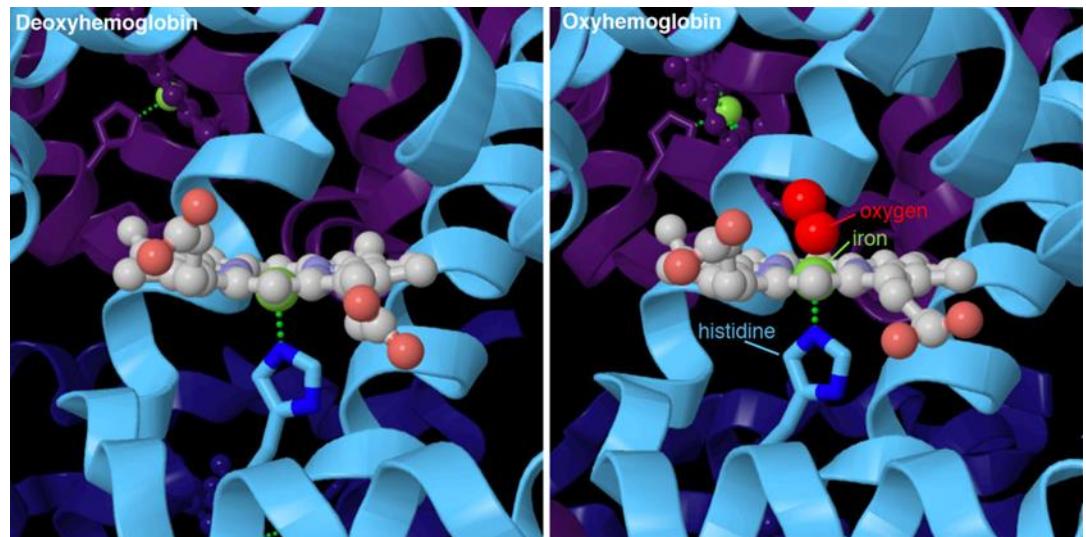
*Allosteric motion of hemoglobin, with an oxygen molecule in turquoise.*



**oxy**

# HEMOGLOBIN

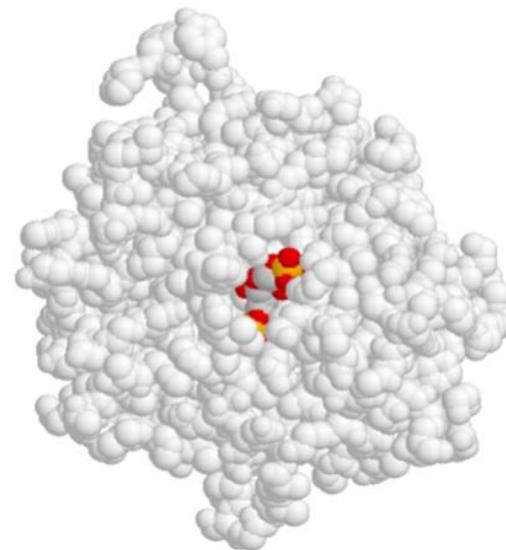
- hemoglobin with no oxygen bound: the heme is seen edge-on with the iron atom colored in green, the key histidines reaching up on the bottom side to bind to the iron atom.
- oxygen bound to the iron, pulling it upwards. This in turn, pulls on the histidine below, which then shifts the location of the entire protein chain. These changes are transmitted throughout the protein, ultimately causing the big shift in shape that changes the binding strength of the neighboring sites.



# ENZYMES

## The active site

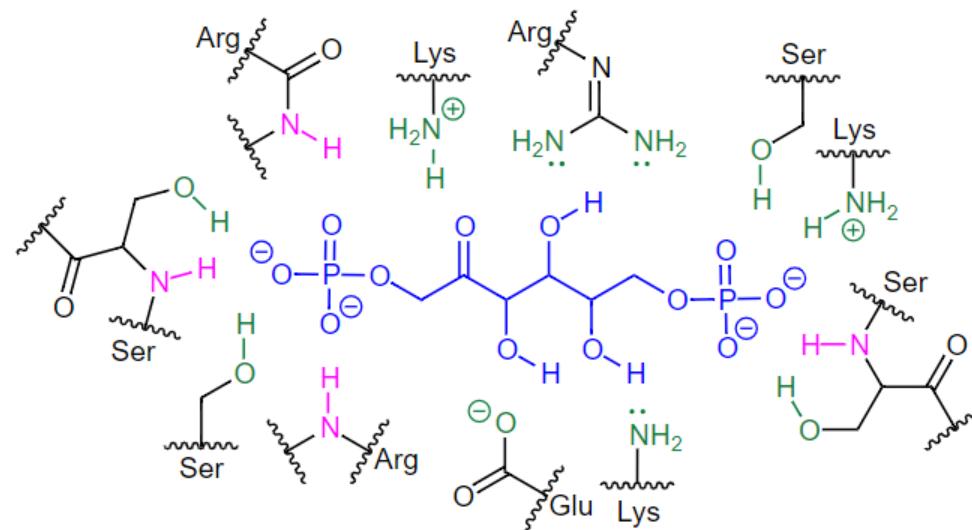
A critical element in the three-dimensional structure of any enzyme is the presence of an '**active site**', which is a pocket, usually located in the interior of the protein, that serves as a docking point for the enzyme's **substrate(s)** ('substrate' is the term that biochemists use for a reactant molecule in an enzyme-catalyzed reaction). It is inside the active site pocket that enzymatic catalysis occurs. Shown below is an image of the glycolytic enzyme fructose-1,6-bisphosphate aldolase, with its substrate bound inside the active site pocket.



When the substrate binds to the active site, a large number of noncovalent interactions form with the amino acid residues that line the active site. The shape of the active site, and the enzyme-substrate interactions that form as a result of substrate binding, are *specific to the substrate-enzyme pair*: the active site has evolved to 'fit' one particular substrate and to catalyze one particular reaction. Other molecules do not fit in this active site nearly so well as fructose 1,6-bisphosphate.



Below is a two-dimensional picture of the substrate (colored blue) surrounded by hydrogen-bonding active site amino acids. Notice that both main chain and side chain groups contribute to hydrogen bonding: in this figure, main chain H-bonding groups are colored purple, and side chain H-bonding groups are colored green.



## Predicting Protein Structure

- It is difficult to predict three dimensional structure from the amino acid sequence
- Comparison to other proteins with known function or structure
  - Easy access of information through public database
    - NCBI (National center for Biotechnology Information) run by National Institutes of Health (NIH)
    - <http://www.ncbi.nlm.nih.gov>
  - Testing structure-function prediction
    - Using molecular biological tools

# PREDICTING

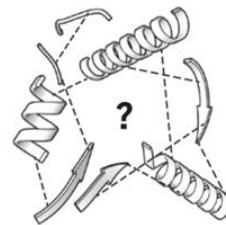
- Predicting 3D structure of protein from its amino acid sequence is one of the most important unsolved problems in biophysics and computational biology.
- A protein's structure determines its function. Experimental protein structure determination is costly, which has driven the search for methods that can **predict protein structure from sequence information**.
- About half of the known proteins are amenable to comparative modeling; that is, an evolutionarily **related protein of known structure can be used as a template for modeling the unknown structure**.
- The fast-paced growth of metagenomics data should enable reliable structure prediction of many protein families.
- Protein data base <https://pdb101.rcsb.org/>

## Structures from sequences

Protein structures are reliably predicted from nothing more than large multiple sequence alignments (13).

### 1 A protein sequence with unknown structure

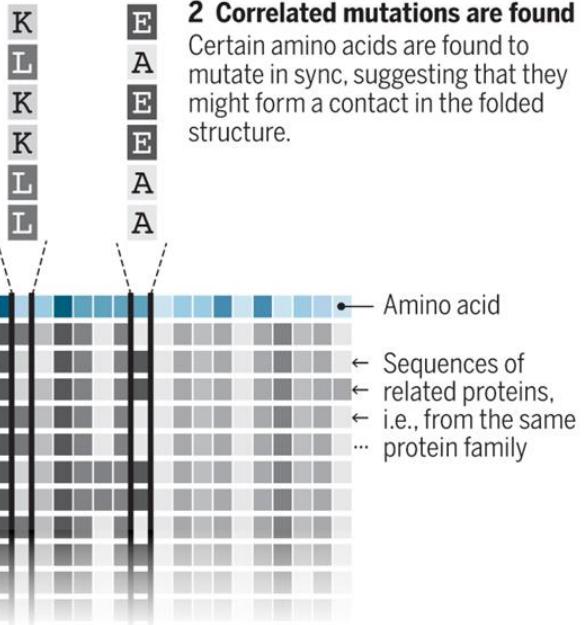
Given a protein sequence (blue) with unknown structure, search databases in order to build huge multiple sequence alignments of the protein's family.



- Protein sequence database (UniRef100)
- Protein sequences from metagenomics

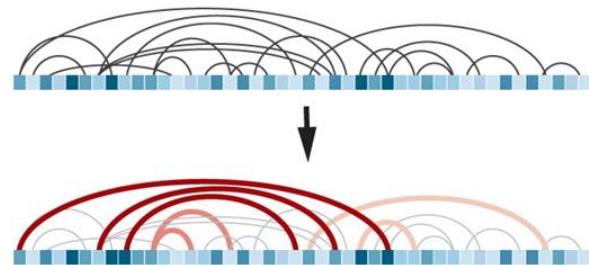
### 2 Correlated mutations are found

Certain amino acids are found to mutate in sync, suggesting that they might form a contact in the folded structure.



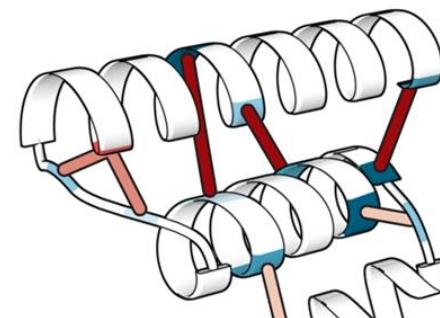
### 3 Find the 3D contacts

Using a statistical method, predict which of the correlations could be due to direct contacts of the amino acids and which ones arise only indirectly from chains of interactions.



### 4 Predict the structure

A 3D structure is predicted de novo, now knowing which residues should be in contact with one another.





## Online Services

- I-TASSER
- C-I-TASSER
- QUARK
- C-QUARK
- LOMETS
- COACH
- COFACTOR
- MetaGO
- MUSTER
- CEthreader
- SEGMER
- FG-MD
- ModRefiner
- REMO
- DEMO
- SPRING
- COTH
- Threpp
- BSpred
- ANGLOR
- EDock
- BSP-SLIM
- SAXSTER



# I-TASSER

Protein Structure & Function Predictions

(The server completed predictions for 675364 proteins submitted by 163571 users from 159 countries)

(The template library was updated on 2022/03/06)

I-TASSER (Iterative Threading ASSEmby Refinement) is a hierarchical approach to protein structure prediction and structure-based function annotation. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template-based fragment assembly simulations. Function insights of the target are then derived by re-threading the 3D models through protein function database BioLiP. I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, CASP13, and CASP14 experiments. It was also ranked the best for function prediction in CASP9. The server is in active development with the goal to provide the most accurate protein structure and function predictions using state-of-the-art algorithms. The server is only for non-commercial use. Please report problems and questions at I-TASSER message board and our developers will study and answer the questions accordingly. ([>> More about the server ...](#))

[Structure models for the SARS-CoV2 Coronavirus genome by C-I-TASSER](#) NEW

[Queue] [Forum] [Download] [Search] [Registration] [Statistics] [Remove] [Potential] [Decoys] [News] [Annotation] [About] [FAQ]

---

I-TASSER On-line Server ([View an example of I-TASSER output](#)):

Copy and paste your sequence within [10, 1500] residues in FASTA format. [Click here for a sample input](#):

Or upload the sequence from your local computer:

Keine ausgewählt

Email: (mandatory, where results will be sent to)

## Protein Engineering

- Manipulation of protein's amino acid sequence to change its function or properties
- Chemical manufacturing
  - Develop enzymes more suitable for industrial applications
  - Increasing enzyme stability
    - e.g. bacteriophage lysozyme: introduce S-S bond to increase heat resistance
    - Proteases in detergent

# PROTEINS - SUMMARY

- Native conformation of the protein is the functional, fully folded protein structure.
- The unique 3D structure of the native conformation is determined by its primary structure, i.e. the amino acid sequence.
- Interactions between the amino acid side chains guide the folding of the polypeptide chain to form secondary, tertiary and sometimes quaternary structures that cooperate in stabilizing the native conformation of the protein.
- Protein denaturation results in unfolding and disorganization of the protein's structure.
- Disease can occur when an apparently normal protein assumes a conformation that is cytotoxic, as in the case of Alzheimer disease.