

# Proteins

## Elements of Biophysics

**Emidio Capriotti**

<http://biofold.org/>

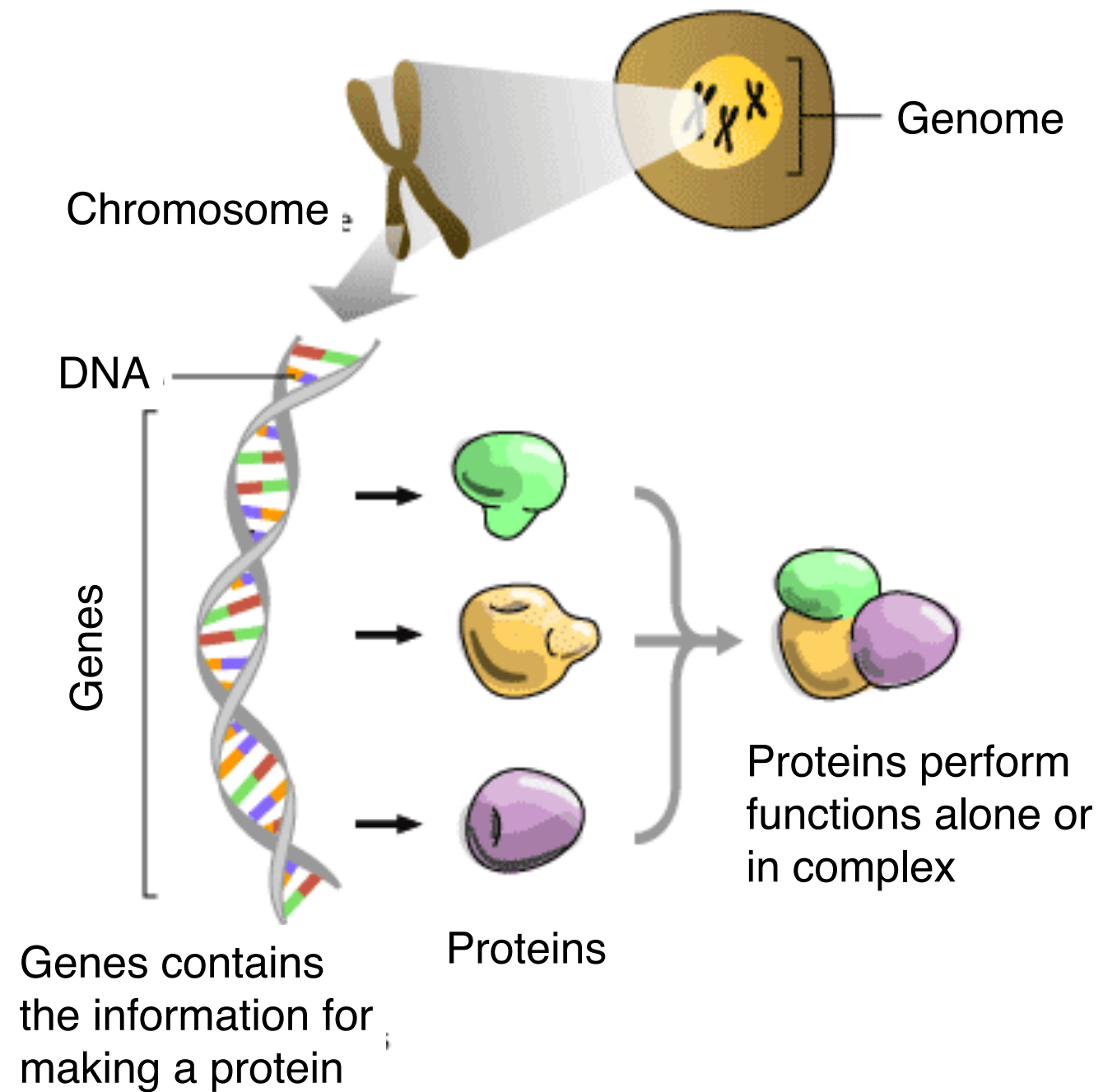


**Biomolecules  
Folding and  
Disease**

Department of Pharmacy and  
Biotechnology (FaBiT)  
University of Bologna

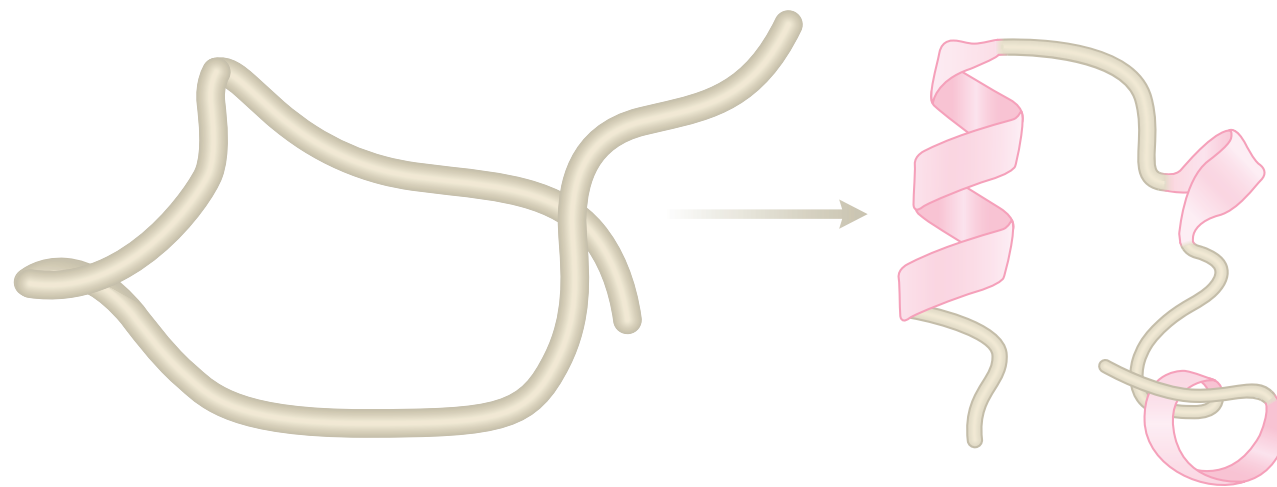


# The Central Dogma



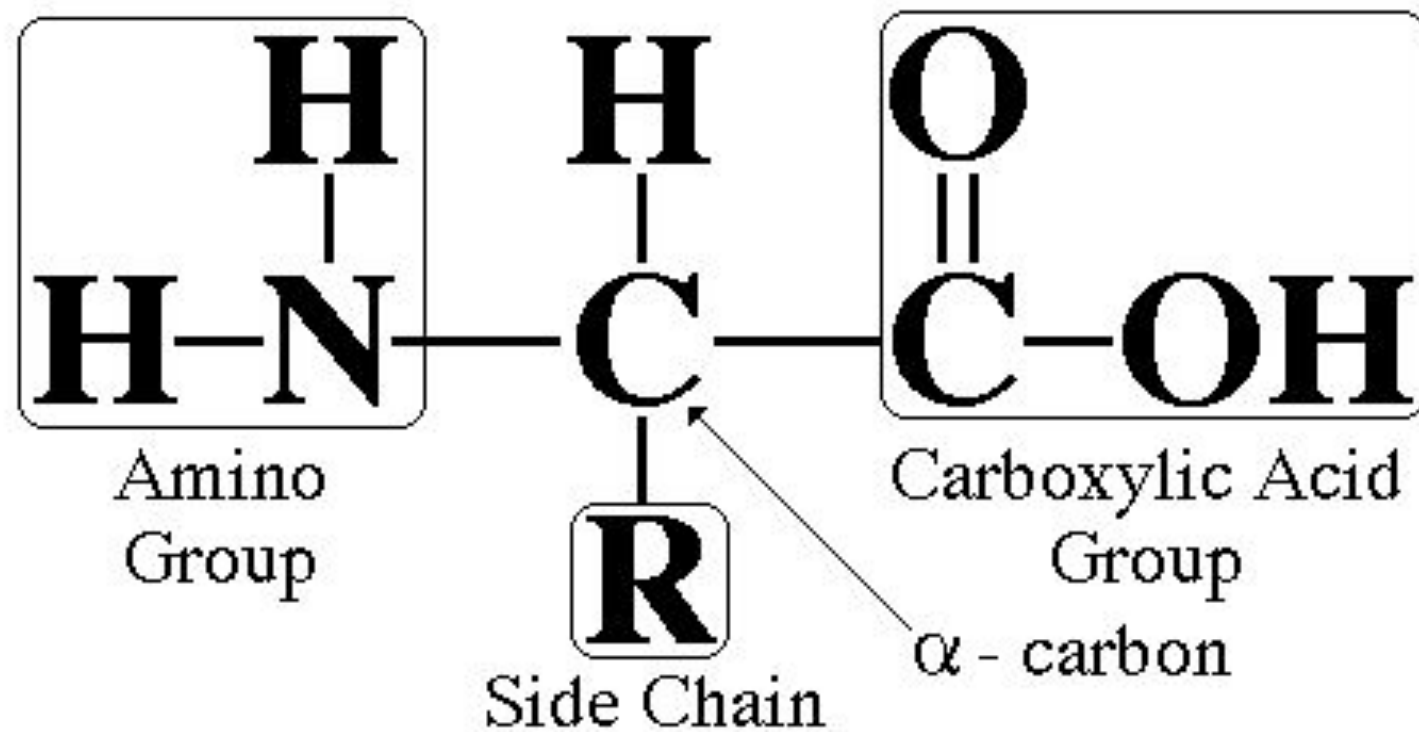
# The Protein

- A polypeptide chain **interacts with the solution** adopting a specific three-dimensional structure with a **hydrophobic interior and hydrophilic exterior**.
- The **folded protein** in the proper conformation can **perform a specific function**. In some cases additional polypeptide chains or cofactors must come together before the protein becomes active.
- The **protein structure** can be understood in terms of a conceptual hierarchy, which is **influenced by different interactions**.



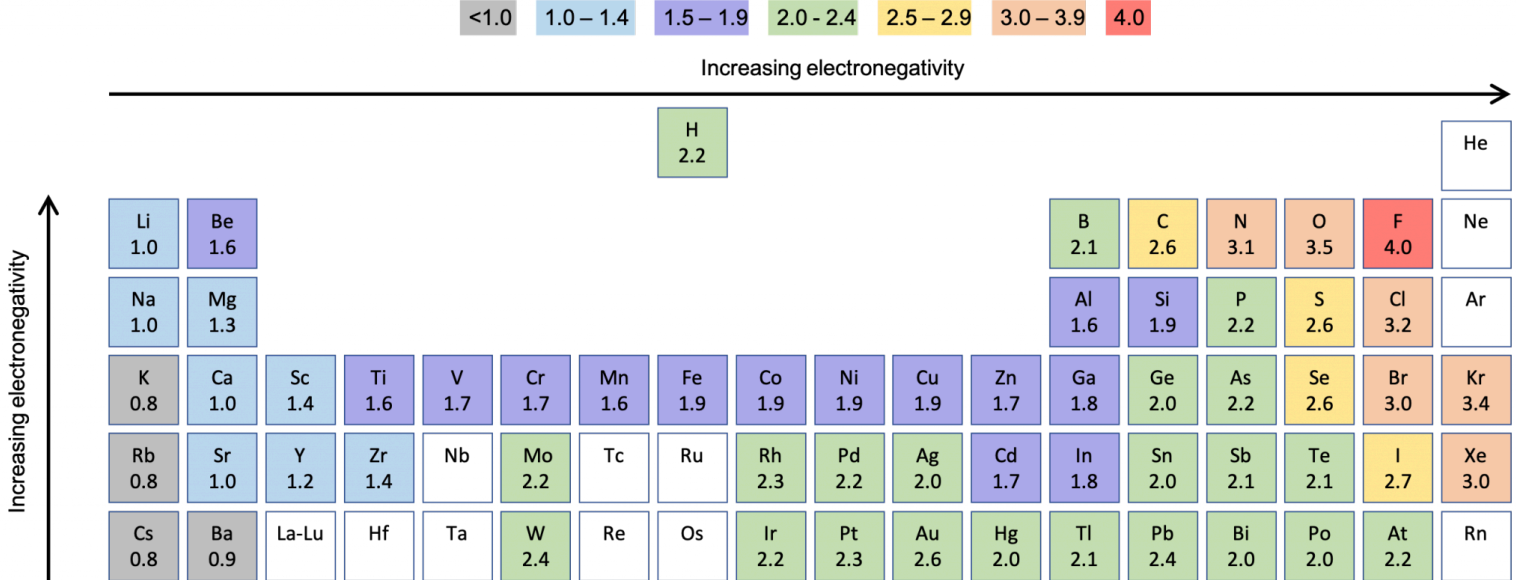
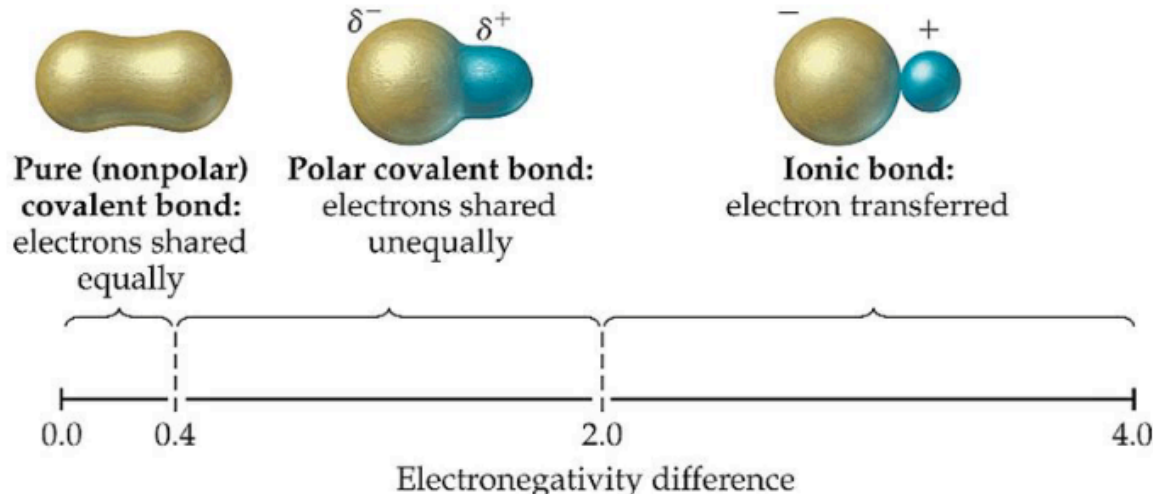
# Amino Acid

The side chain (R) determines the type of the amino acid



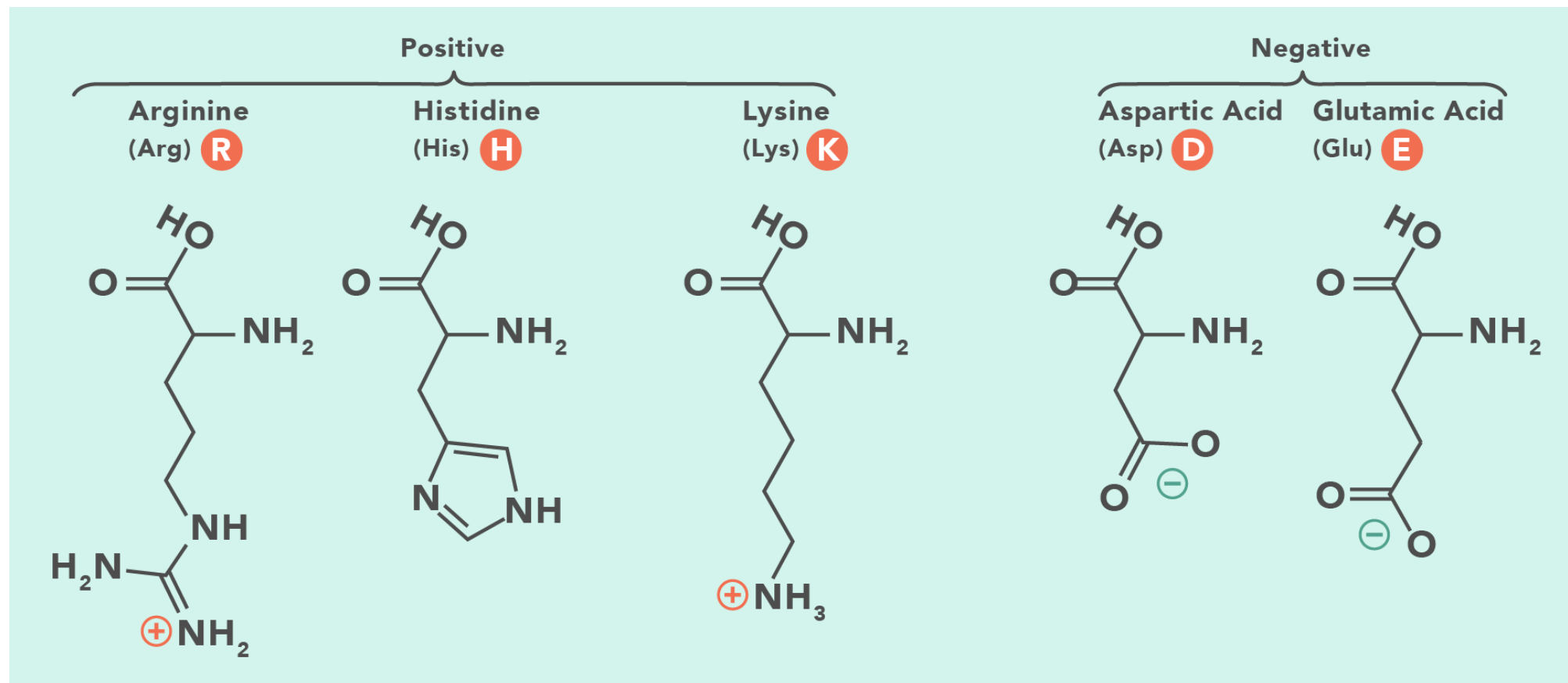
# Electronegativity

Tendency for an atom to attract shared electrons when forming a chemical bond.



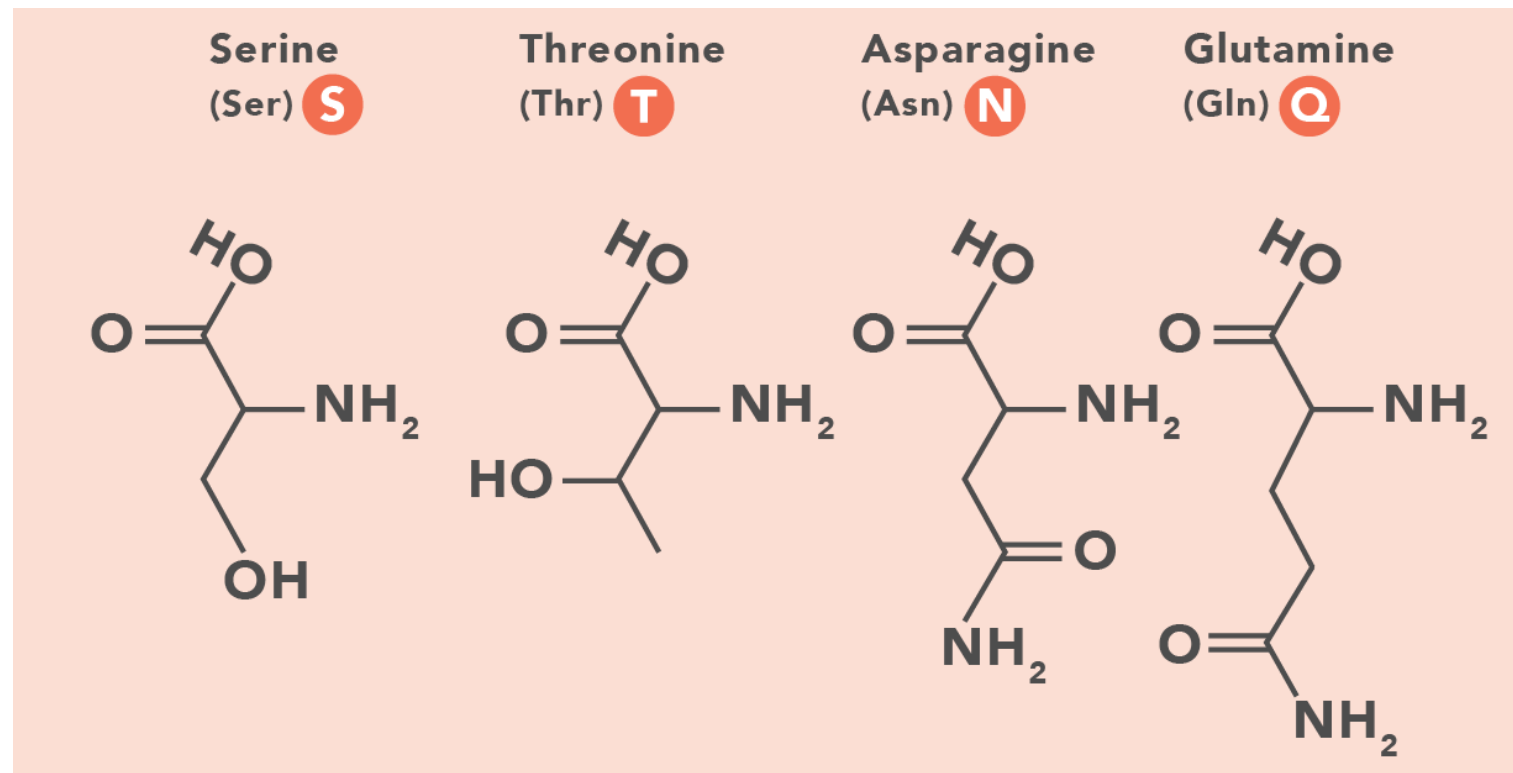
# Charged Amino Acid

The side chain end either with amine (-NH<sub>2</sub>) or carboxylic (-COOH) groups



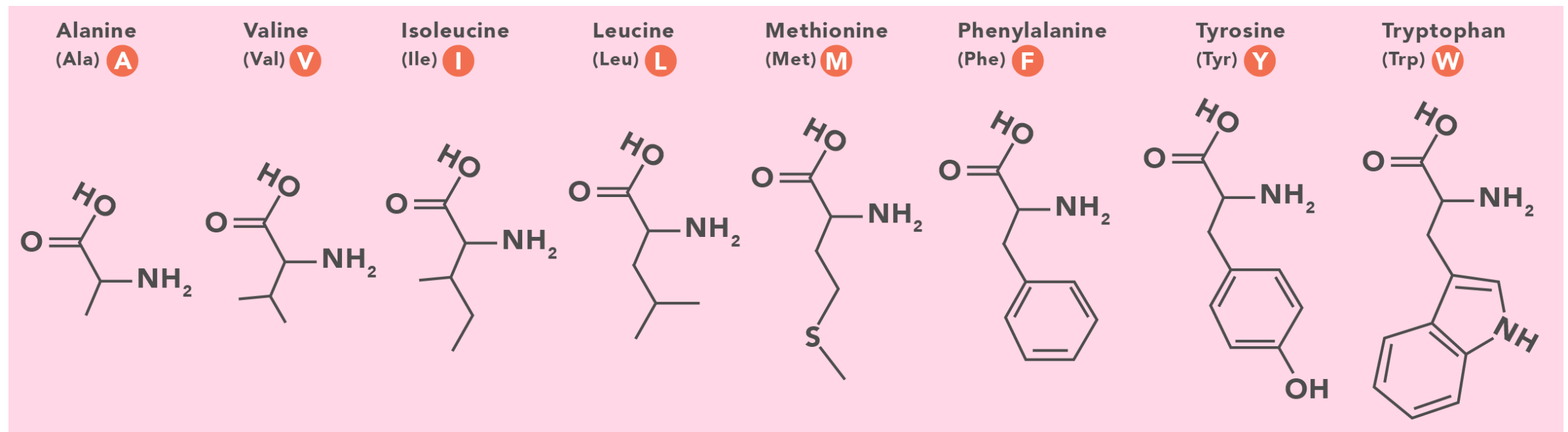
# Polar Amino Acid

The side chain end with partially charged groups



# Hydrophobic Amino Acids

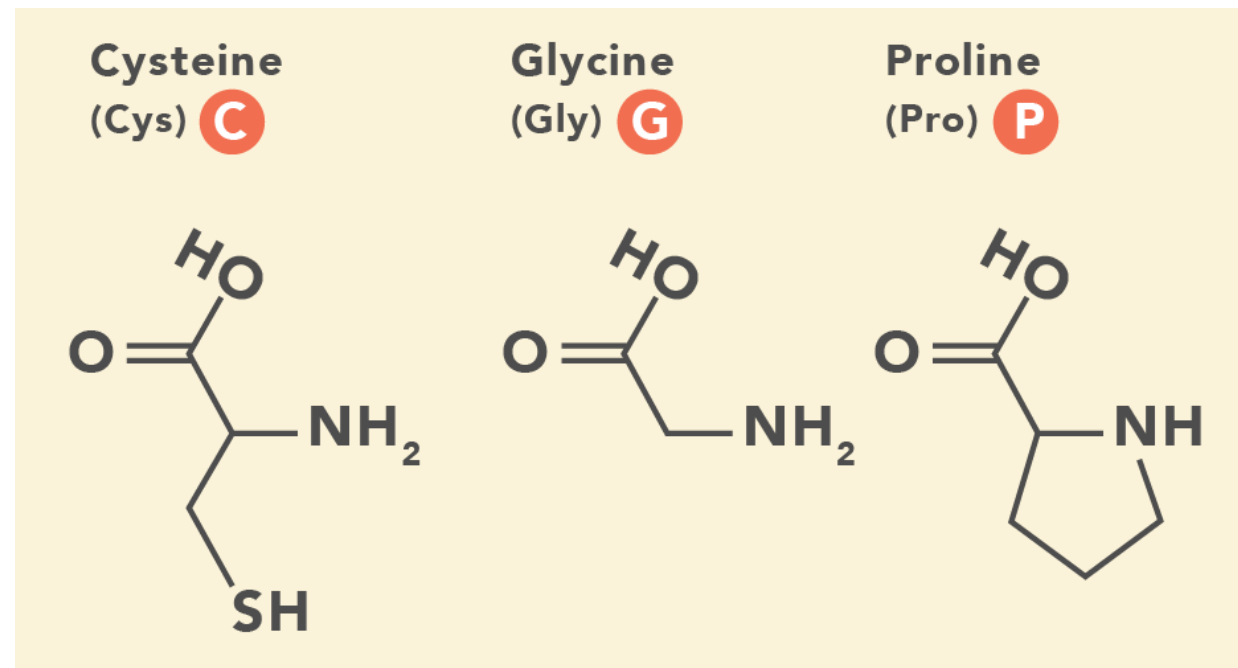
The side chain (R) are alkyl or aromatic groups





# Special Cases

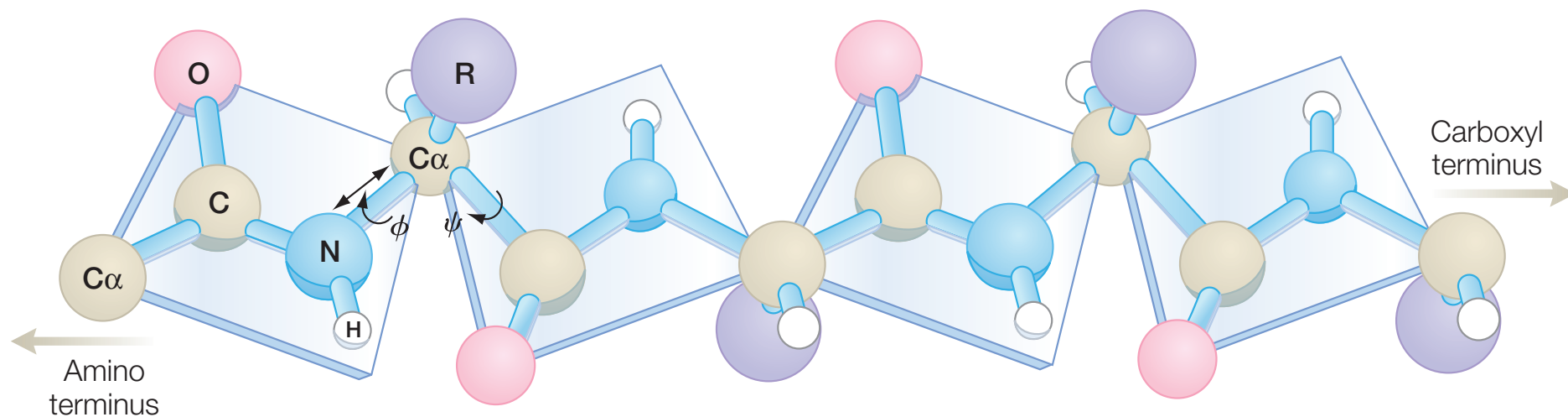
The side chain (R) has particular features



# Polypeptide Chain

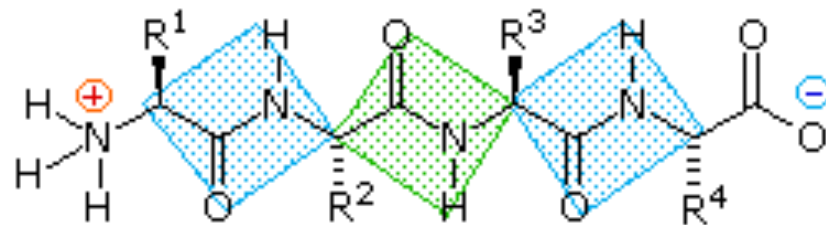
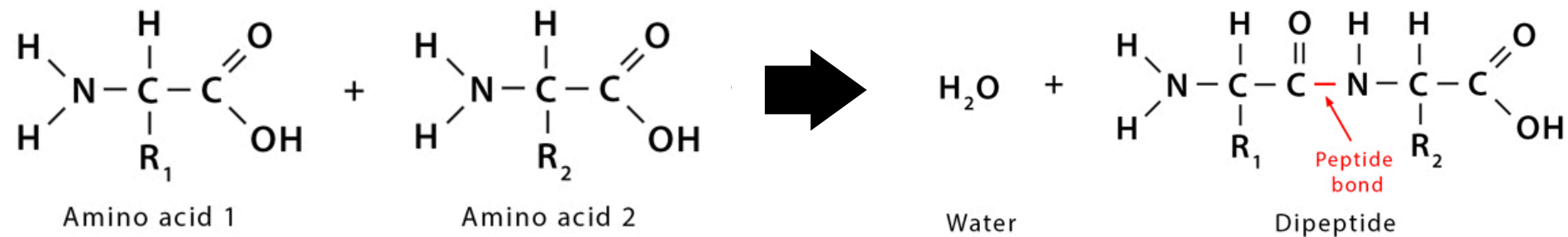
Amino acids form peptide bonds in which the **carboxyl group of one amino acid is joined** to the **amino group of another amino acid**. Many amino acids joined by peptide bonds form a polypeptide chain.

The **protein chain** consists of the regularly repeating main chain or **backbone and the side chains**.

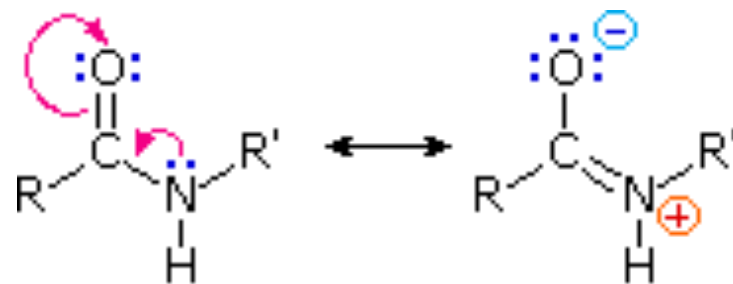


# Peptide Bond

The **peptide unit is always rigid and planar** with the hydrogen of the amino group opposite to the oxygen of the carbonyl group, except for proline.

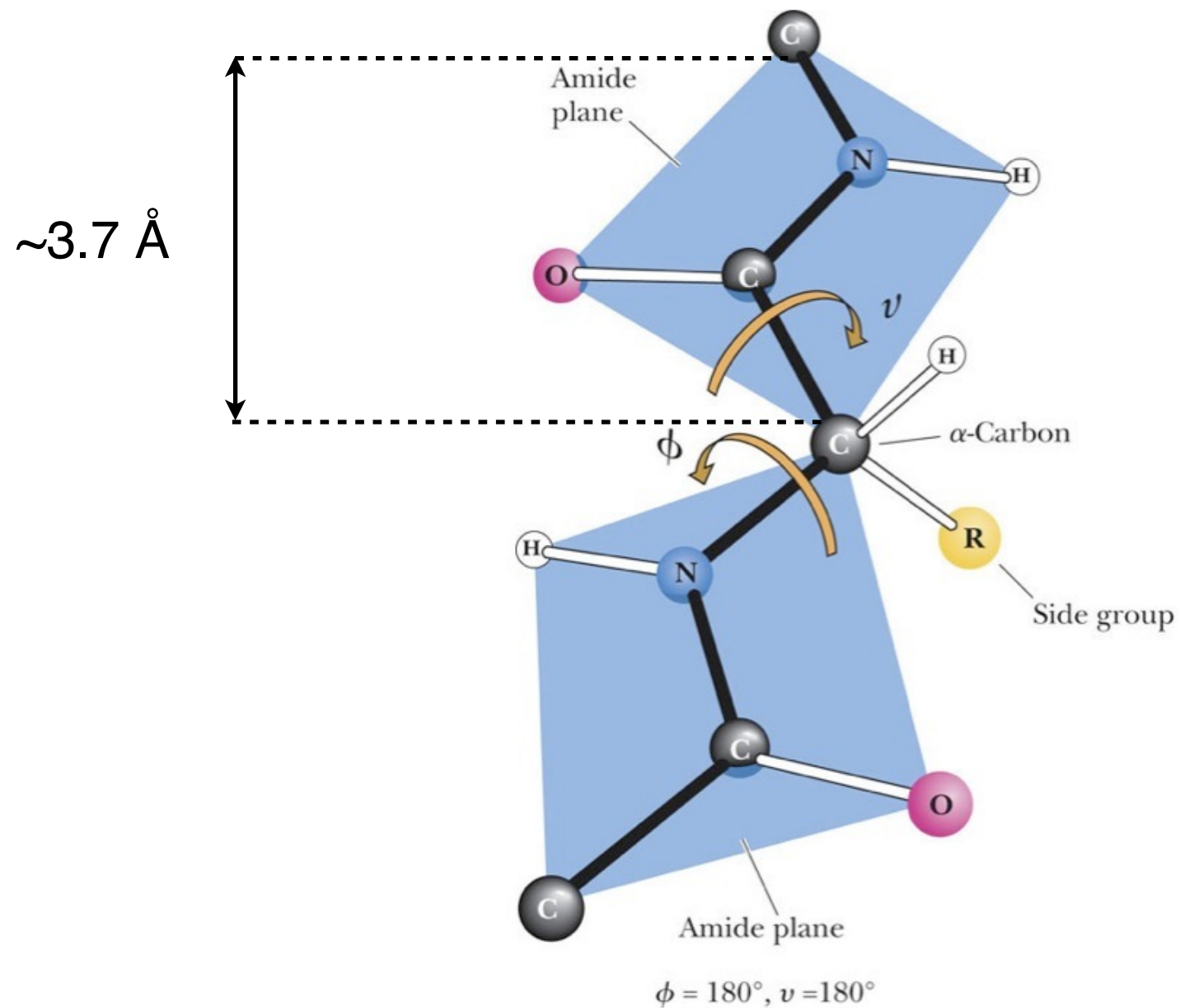


The bond between the **carbonyl carbon atom and the nitrogen atom is not free to rotate.**



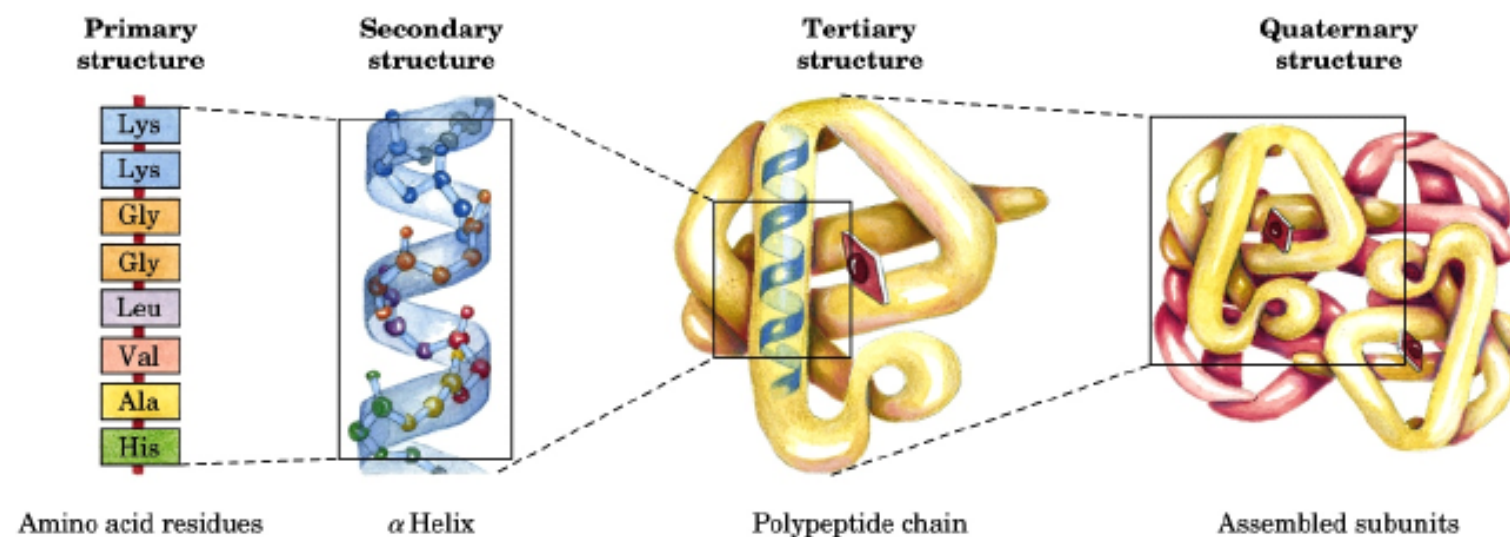
# Torsion Angles

The protein consists of a series of planes that can rotate relative to each other. The bond angles centered around at the  $C_\alpha$  are identified as  $\phi$  and  $\psi$



# Protein Structures

- The **sequence of amino acid residues** linked together by peptide bonds describes the **primary structure** of the protein.
- **Local regions** of the protein **fold into stable arrangements** of amino acid residues that are recurring in proteins ( $\alpha$  helices and  $\beta$  sheets)
- The **overall fold of these secondary-structure** elements describes the tertiary structure of the protein.
- The quaternary structure is the **arrangement of different polypeptides** of the protein.



# Hydrogen Bond

- **Hydrogen bond is an electrostatic interaction** between the partial negative charge (acceptor) and the partial positive charge (donor).
- The strength of the **hydrogen bond** will depend upon **the relative angles and distances**.
- A typical distance is **1.8 Å from the hydrogen to the acceptor**, or about **2.8 Å between the nuclei** of the donor and acceptor.
- Hydrogen bonds involving the main-chain atoms establish the **stability of the secondary structures of proteins**.

# Electrostatic Interactions

The side chains of the amino acid residues lysine, arginine, glutamate, aspartate, and histidine are ionizable. They can form **electrostatic interactions contribute to protein stability and function.**

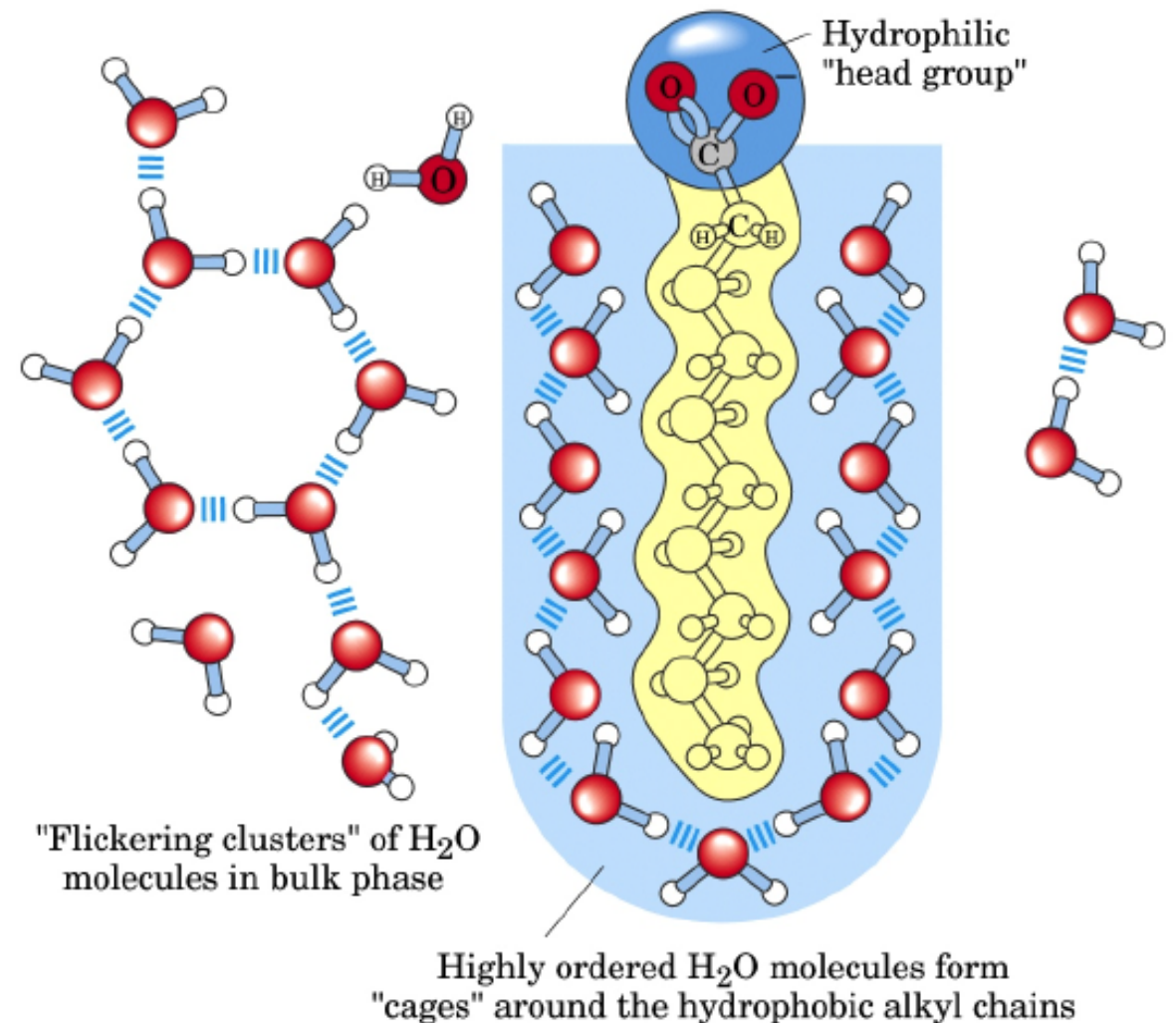
In a typical electrostatic model, the potential between two charges,  $q_1$  and  $q_2$ , that are separated by a distance,  $r$ , is given by:

$$V(r) = \frac{q_1 q_2}{4\pi\epsilon r} = (1389 \text{ kJ mol}^{-1}) \frac{q_1(e)q_2(e)}{\epsilon r(\text{\AA})}$$

where  $\epsilon$  is the **dielectric constant**. The dielectric constant of a vacuum is defined as 1.0 and its value in different solvents **ranges from 80 in a polar solvent such as water to 2 for a nonpolar solvent such as benzene.**

# Hydrophobic effect

- Water molecules form a cage-like structure around the nonpolar molecule.
- The positive  $\Delta H$  is due to the fact that the cage has to be broken to transfer the nonpolar molecule.
- The **positive  $\Delta S$**  is due to the fact that the water molecules are less ordered (an increase in the degree of disorder) when the cage is broken.



$$\Delta G = \Delta H - T\Delta S$$



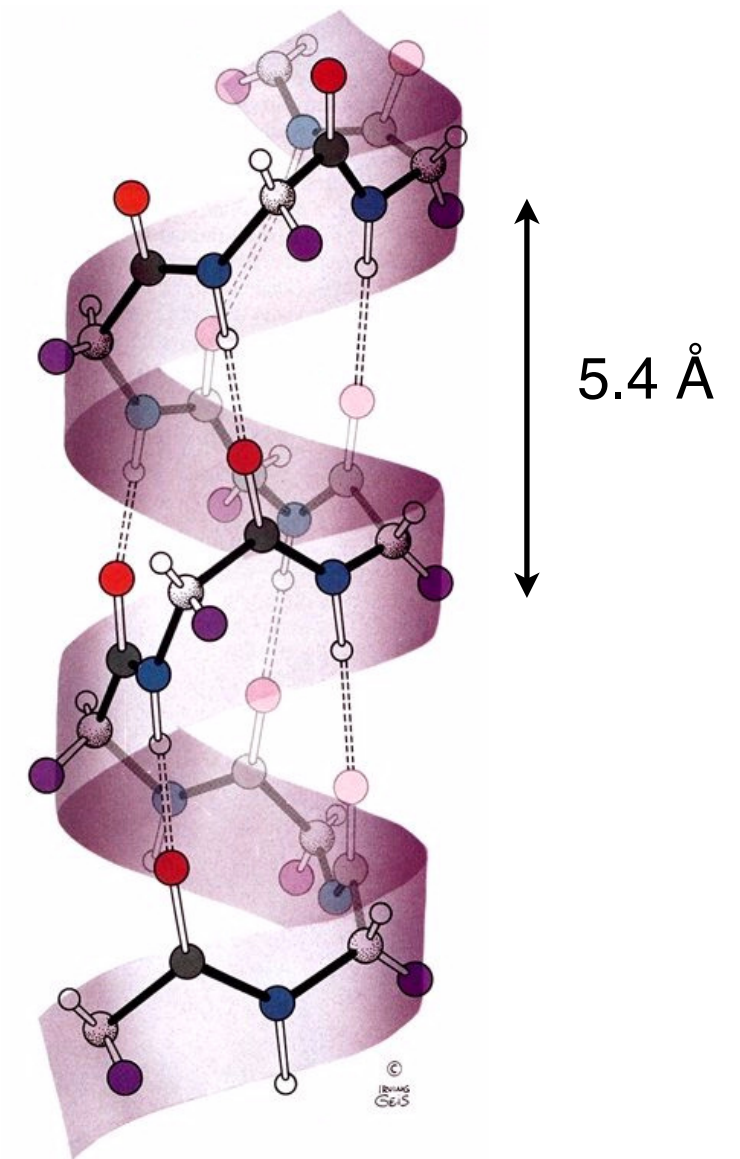
# Folding interactions

Several types of **electrostatic interactions** are **contributing** to the **stability** of the native state but they are **not the driving forces** in the folding process

Type	Examples	Binding energy (kcal/mol)	Change of free energy water to ethanol (kcal/mol)
<b>Electrostatic interaction</b>	Salt bridge	-5	-1
	Dipole-dipole	+0.3	
<b>Hydrogen bond</b>	Water	-4	
	Protein backbone	-3	
<b>Dispersion forces</b>	Aliphatic hydrogen	-0.03	
<b>Hydrophobic forces</b>	Side chain of Phe		-2.4

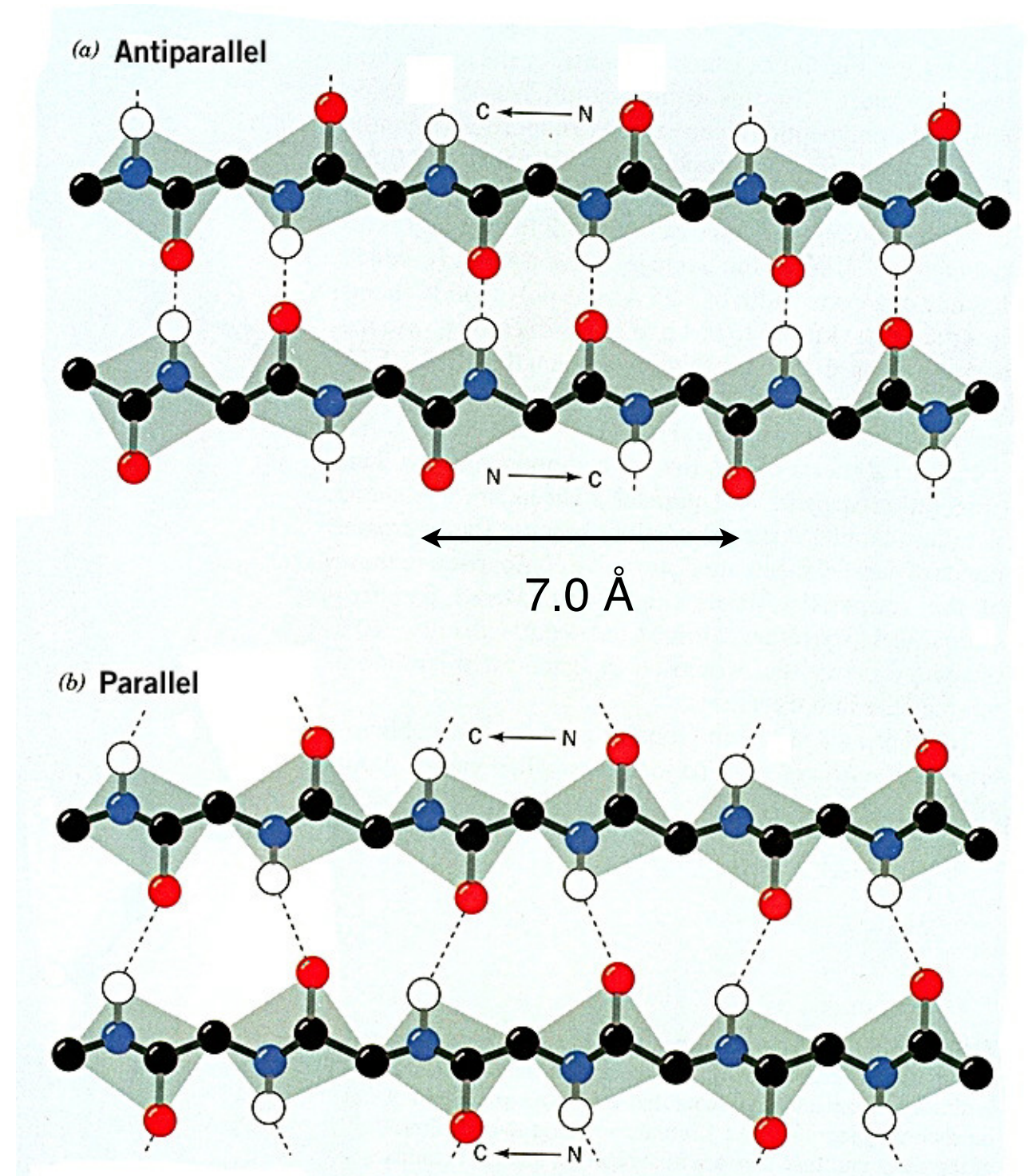
# Secondary structure (I)

- Helices observed in proteins are mostly right-handed.
- Typical  $\phi$ ,  $\psi$  values for residues in  $\alpha$ -helix are around  $-60^\circ$ ;  $-50^\circ$
- Side chains project backward and outward.
- The core of  $\alpha$ -helix is tightly packed.



# Secondary structure (II)

- Typical  $\phi$ ,  $\psi$  values for residues in  $\beta$ -sheet are around  $140^\circ$ ,  $-130^\circ$
- Side chains of neighboring residues project in opposite directions.
- The polypeptide is in a more extended conformation.
- Parallel  $\beta$ -sheets are less stable than anti-parallel  $\beta$ -sheets.





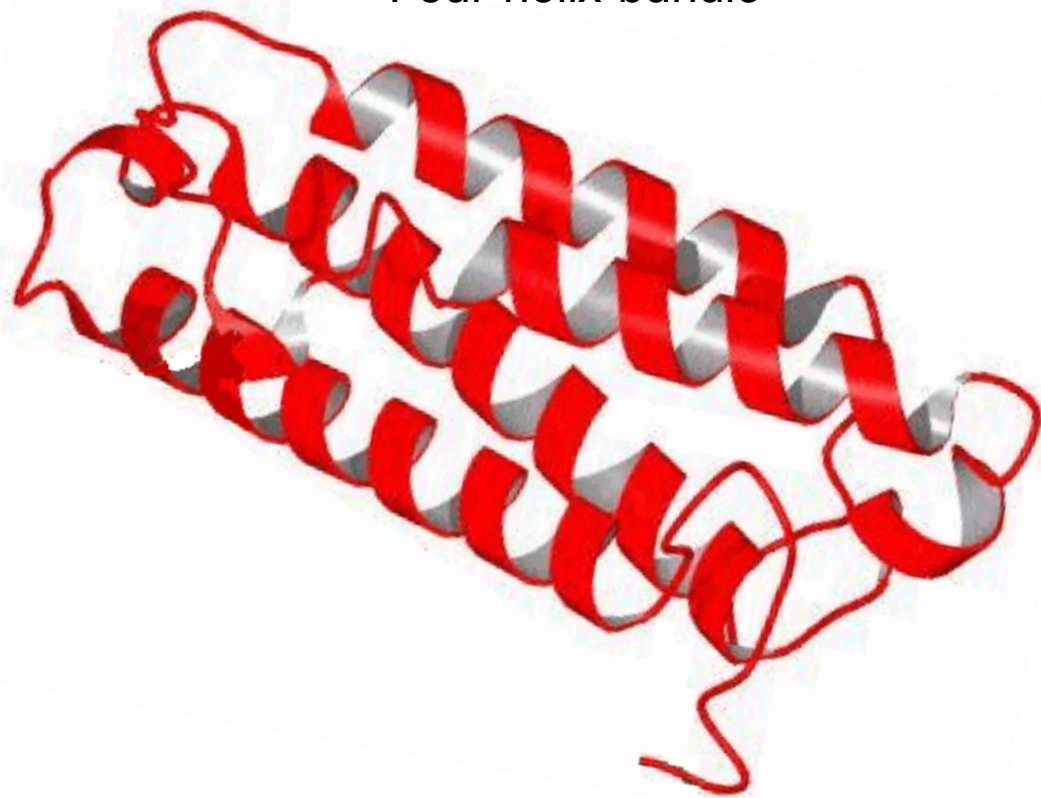
# Quaternary structure

The arrangements of secondary structural elements form the Tertiary Structure of the protein.

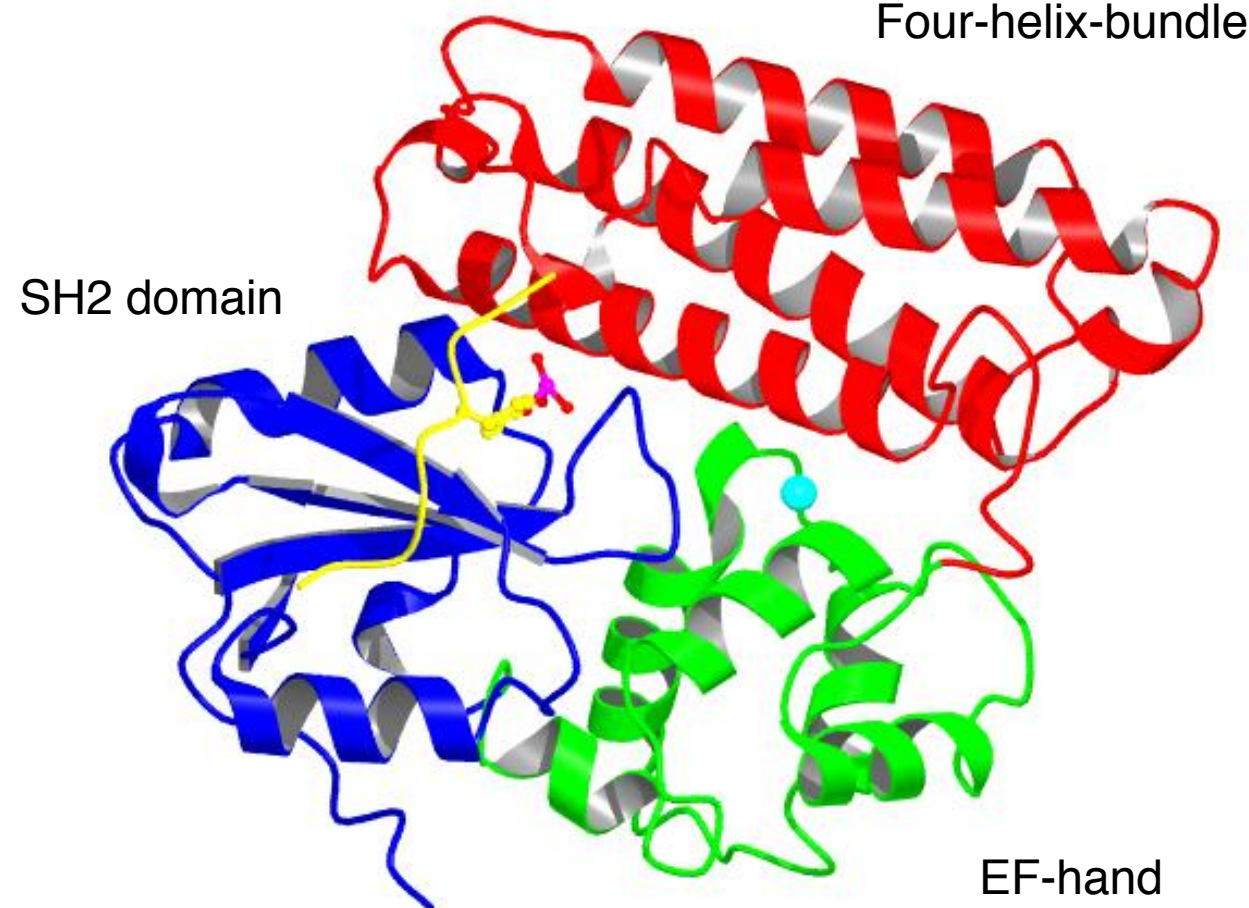
The complex of **two or more protein domains defines the Quaternary Structure**.

In the example Four-helix-bundle, EF-hand and SH2 domains together form an integrated phosphoprotein that functions as a negative regulator of many signaling pathways from receptors at the cell surface.

Four-helix-bundle



Four-helix-bundle



# The Protein Data Bank

The largest repository of macromolecular structures obtained mainly by X-ray crystallography and NMR

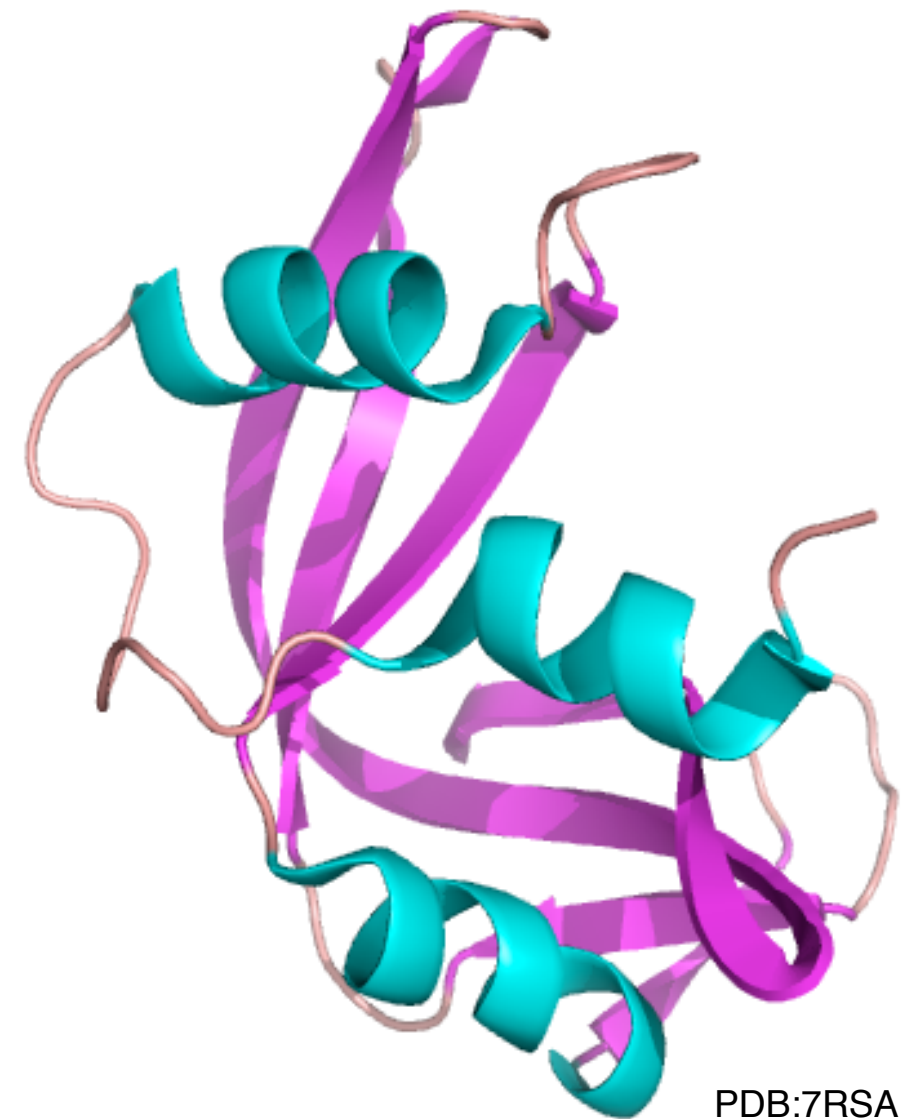
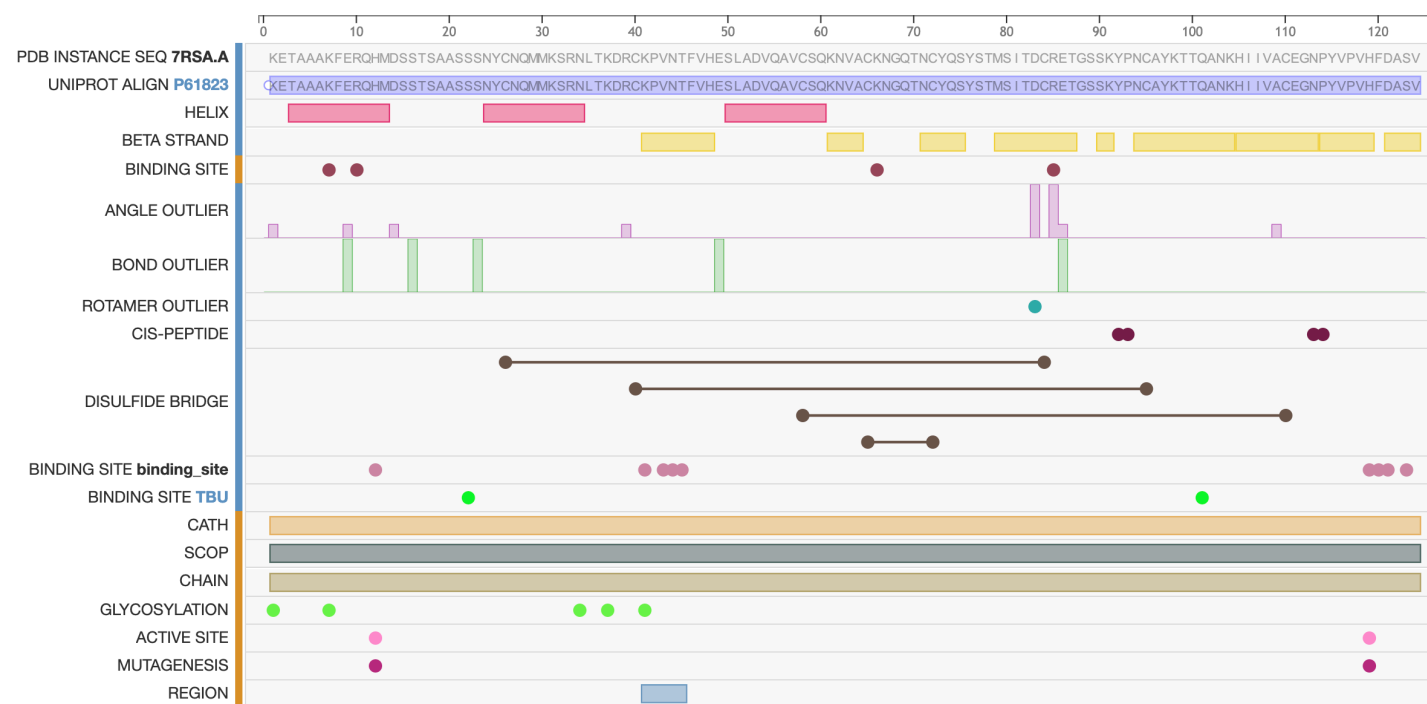
The screenshot shows the RCSB PDB website homepage. At the top, there is a navigation bar with links for Deposit, Search, Visualize, Analyze, Download, Learn, More, Documentation, and Careers. On the right of this bar are buttons for MyPDB and Contact us. Below the navigation bar, the PDB logo is displayed alongside statistics: 196,779 Structures from the PDB and 1,000,361 Computed Structure Models (CSM). A search bar is present with a dropdown menu for '3D Structures' and a search input field containing the placeholder text 'Enter search term(s), Entry ID(s), or sequence'. To the right of the search bar is a toggle for 'Include CSM' and a search icon. Below the search bar are links for 'Advanced Search' and 'Browse Annotations', and a 'Help' link. A secondary navigation bar includes logos for PDB-101, PDB, EMDatabase, Nucleic Acid Database, and wwPDB Foundation, along with social media icons for Facebook, Twitter, YouTube, and LinkedIn. A teal banner below the navigation bar reads 'NEW! Computed Structure Models (CSM)' with a 'Learn more' button. The main content area is divided into three sections. The left section is a dark blue sidebar with a 'Welcome' header and menu items for Deposit, Search, Visualize, Analyze, Download, and Learn. The middle section features a white background with the text: 'RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:'. Below this text are two bullet points: 'Experimentally-determined 3D structures from the Protein Data Bank (PDB) archive' and 'Computed Structure Models (CSM) from AlphaFold DB and ModelArchive'. A paragraph follows: 'These data can be explored in context of external annotations providing a structural view of biology.' At the bottom of this section are two promotional images: one for 'COVID-19 CORONAVIRUS Resources' showing a virus particle, and another for 'Join the RCSB PDB Team' showing hands clasped together. The right section is titled 'October Molecule of the Month' and features a large 3D molecular model of the 'Phytohormone Receptor DWARF14' in blue, orange, and red, with a small inset showing a ball-and-stick model of a ligand.

<http://rcsb.org>

<http://ftp.rcsb.org/pub/pdb/>

# The Bovine Ribonuclease A

Ribonuclease A (RNase A) is a **pancreatic ribonuclease** which belongs to a class of Lyases. This enzyme cleaves at the 3'-side of pyrimidine (uracil or cytosine) phosphate bonds in RNA.

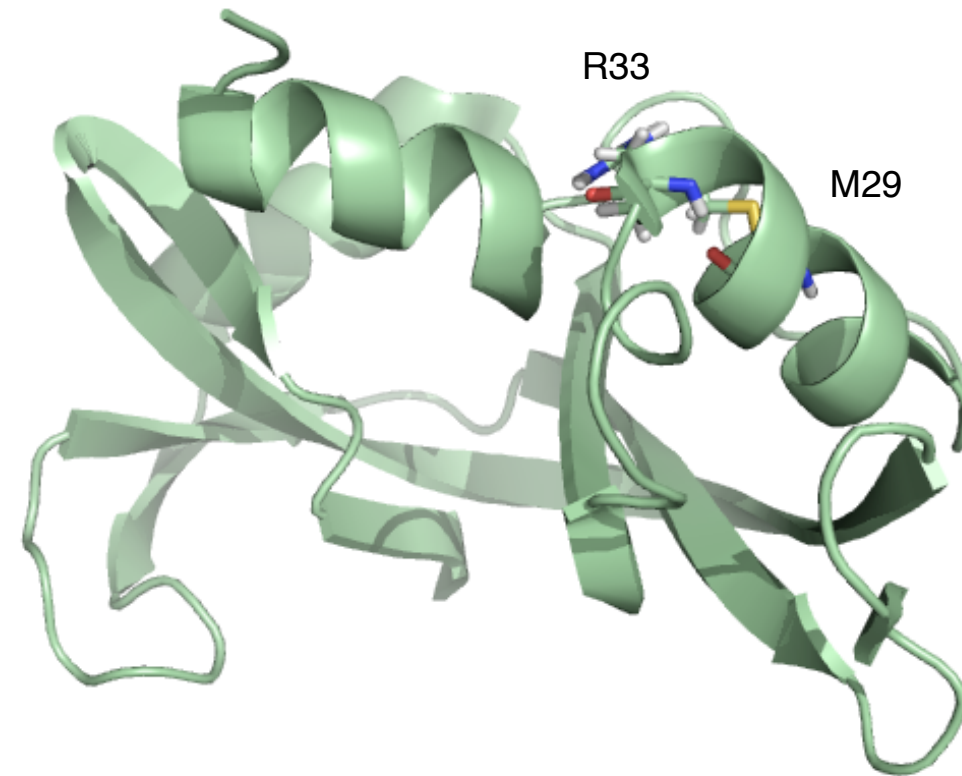
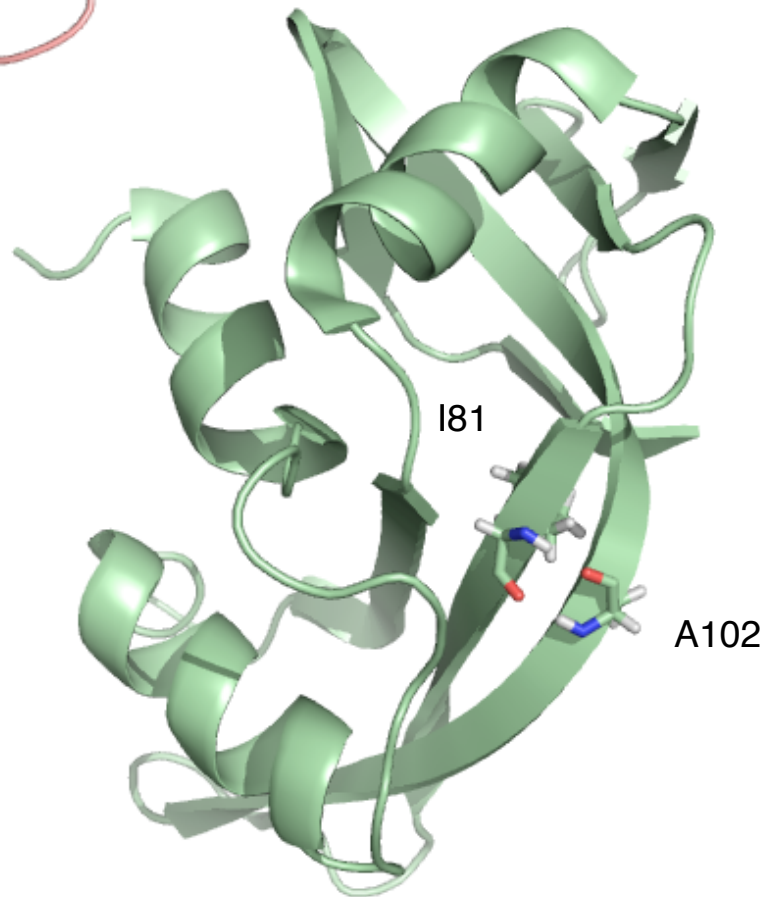
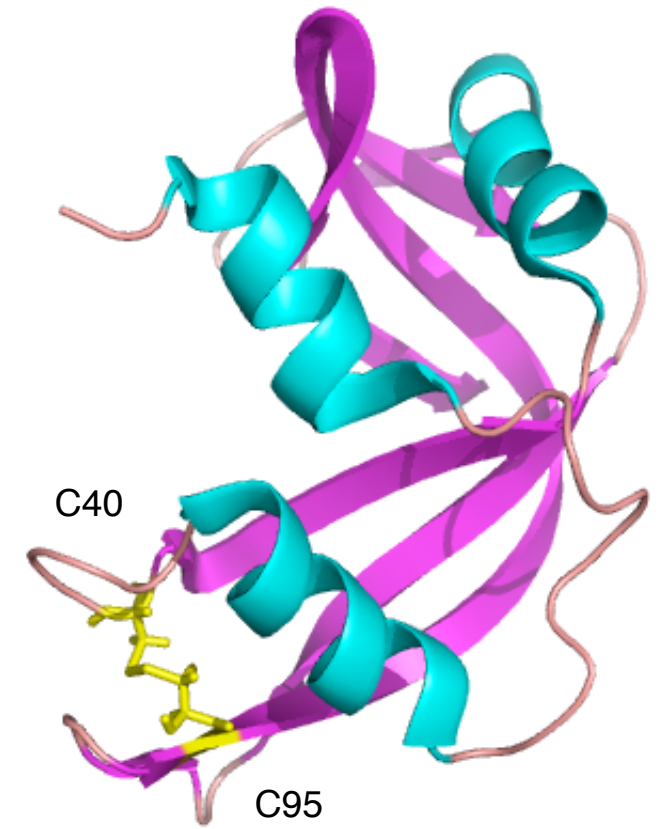
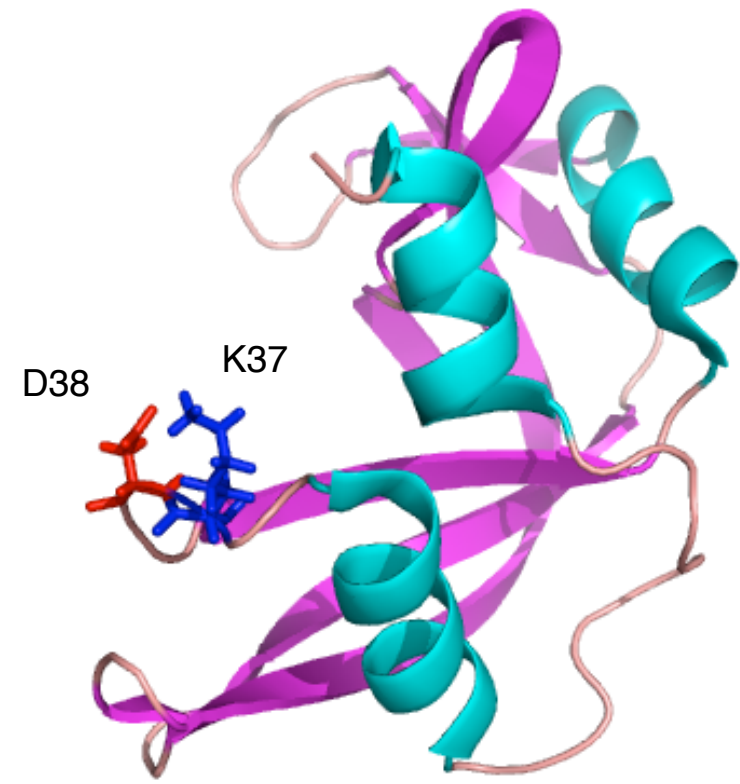


PDB:7RSA



# Bonds and interactions

Examples of salt bridge, disulfide bond and hydrogen bonds in ribonuclease A



# Protein Binding

Protein binding can be formalized with the following reaction:



where  $R$  is the receptor  $L$  the ligand and  $RL$  receptor-ligand complex. This formalism can **apply both to the study of drug-target and protein-protein interactions**. In the first case the ligand will be a small molecule while in the second case the ligand will be a protein. In general the formation constant ( $K_f$ ) is:

$$K_f = \frac{[RL]}{[R][L]}$$

To quantify the affinity of the receptor for ligand, the dissociation constant ( $K_D$ ) is indicated:

$$K_D = \frac{1}{K_f} = \frac{[R][L]}{[RL]}$$

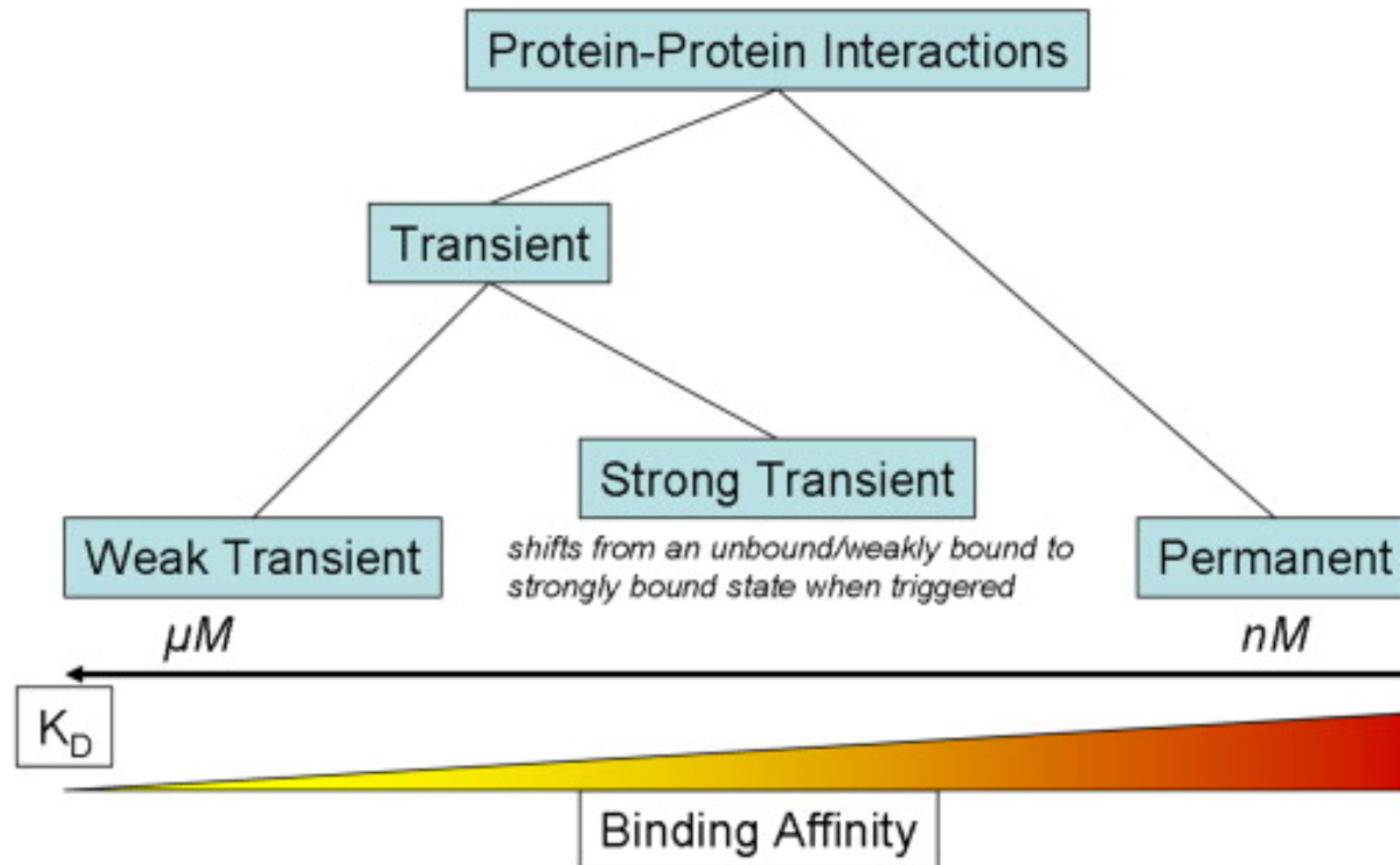
- A small value for  $K_D \Rightarrow$  the equilibrium favors the complex (high affinity)
- A large value for  $K_D \Rightarrow$  the equilibrium favors the separation (low affinity)



# The Molecular Viewpoint

- The affinity of **PPI varies from millimolar to picomolar**, depending on the type of interaction and signaling needed (Chen et al. Protein Sci. 2013)
- Despite affinity varies over a wide range, **proteins maintain a high degree of specificity** for their partners
- Many **proteins exhibit specificity for multiple partners** (Reichmann et al. Curr. Opin. Struct. Biol. 2007).
- The nature of the **interaction surface** determines how proteins interact
- A detailed knowledge of the **interaction surfaces** of proteins and their energetics is necessary to understand the regulatory **mechanisms of biochemical pathways** (especially to modulate or block these pathways for therapeutic purposes)

# Protein-Protein Interactions



**Strong transient:** This category includes interactions that are triggered/stabilised by an effector molecule or conformational change. An example is given by the Ras proteins, which form tight complexes with their partners when GTP-bound and only weak complexes when GDP-bound.

# Surface of Interaction (I)

- The area of PPI interfaces is large (1000 to 4000 Å<sup>2</sup>)
- **Standard-sized** interfaces are 1200 to 2000 Å<sup>2</sup>
- **Short-lived and low-stability complexes**  $\Rightarrow$  smaller interfaces (1150–1200 Å<sup>2</sup>)
- **large surfaces** (2000 to 4600 Å<sup>2</sup>)  $\Rightarrow$ 
  - proteases and particular inhibitors
  - G-proteins and other components of the signal transduction system
- **Protein-small molecule interaction** surfaces have an area of 300 to 1000 Å<sup>2</sup>.

# Surface of Interaction (II)

- Surfaces of PPIs are generally **flat** and lack the grooves and pockets that are present at the surfaces of proteins that bind to small molecules.
- PPI **surfaces are generally hydrophobic** in nature.
- Only certain **hydrophobic spots contribute to the free energy** of binding and help to hold the two proteins together.
- Such regions are called **hot spots**.

# Hot Spots

- **Hot spots** account for less than **50% of the contact area** of PPI
- A region of protein surface is called a hot spot when **replacement of an amino acid** residue by alanine in that spot **lowers the free energy of binding by at least 2 kcal/mol**
- Analysis of the **amino acid composition of hot spots** shows that some residues are found more frequently in hot spots (Tyr, Trp, and Arg)
- The hot spots are surrounded by energetically less important residues that **separate/prevent bulk water from hot spots**

# Analysis of Protein Complex

- identification of **interface residues/hot spots**
- **details** about the interface  
solvent accessible surface area, shape, complementarity between surfaces, residue interface propensities, hydrophobicity, segmentation and secondary structure, and conformational changes on complex formation
- assignment of **protein function**
- recognition of **specific residue motifs**

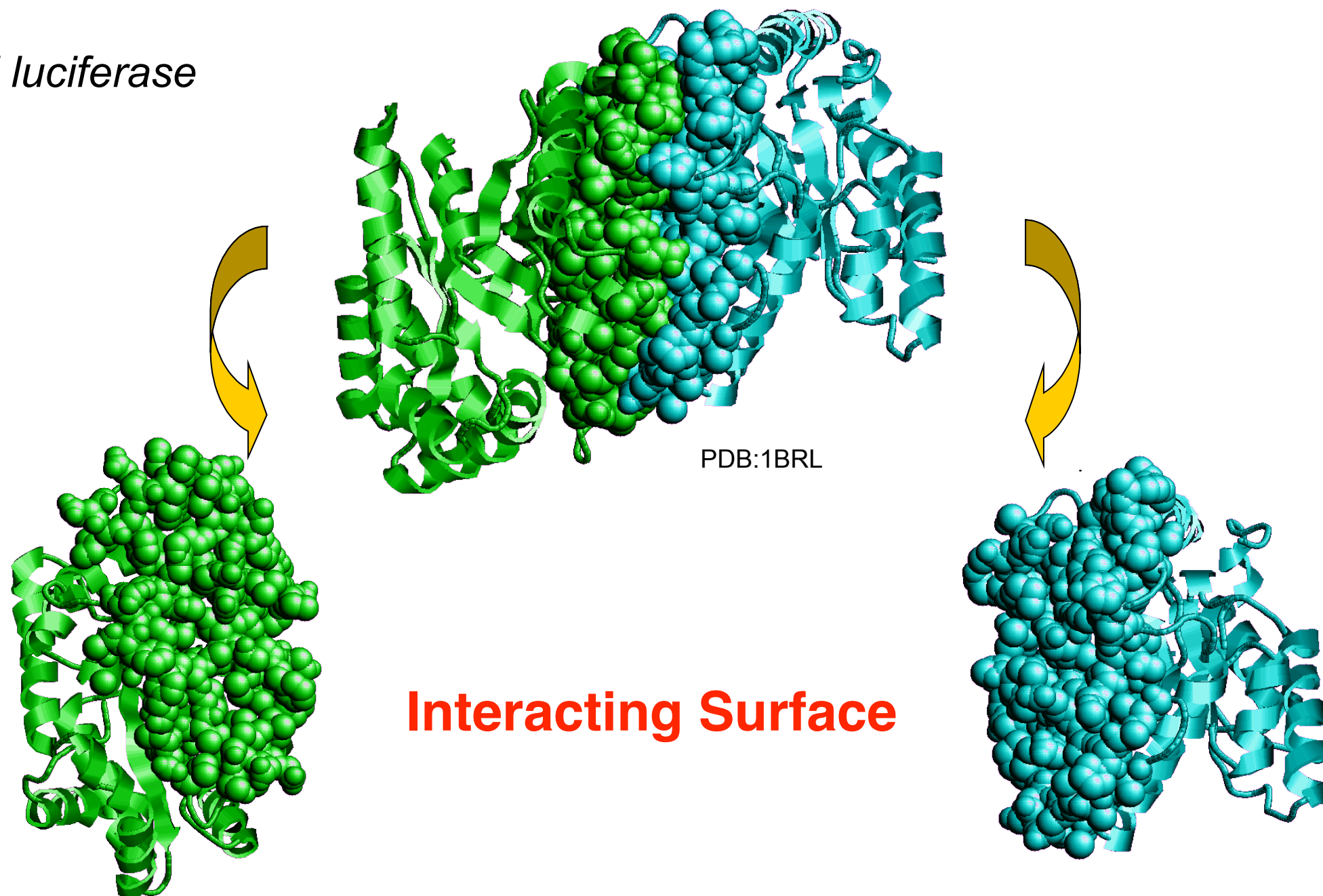
# Structure PPI Data

- The most significant contribution to understanding the PPI surface comes from structural biology via **X-ray crystallography** or **NMR** as well as **mutational studies**
- Prediction of interaction/binding sites
- Prediction of protein-protein complexes

# Interacting surface

Difference in Accessible Surface Area (ASA) between monomers and complex

*Bacterial luciferase*





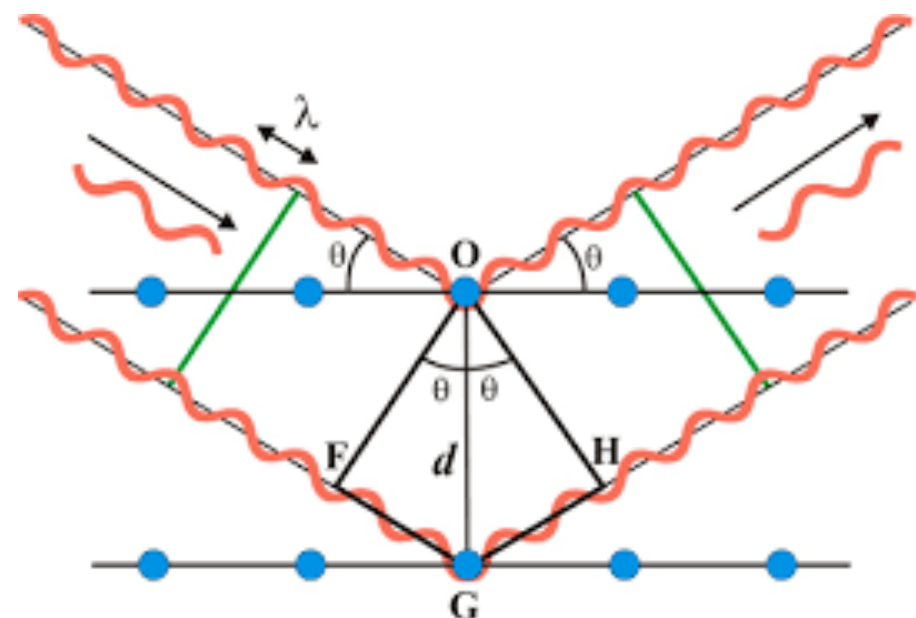
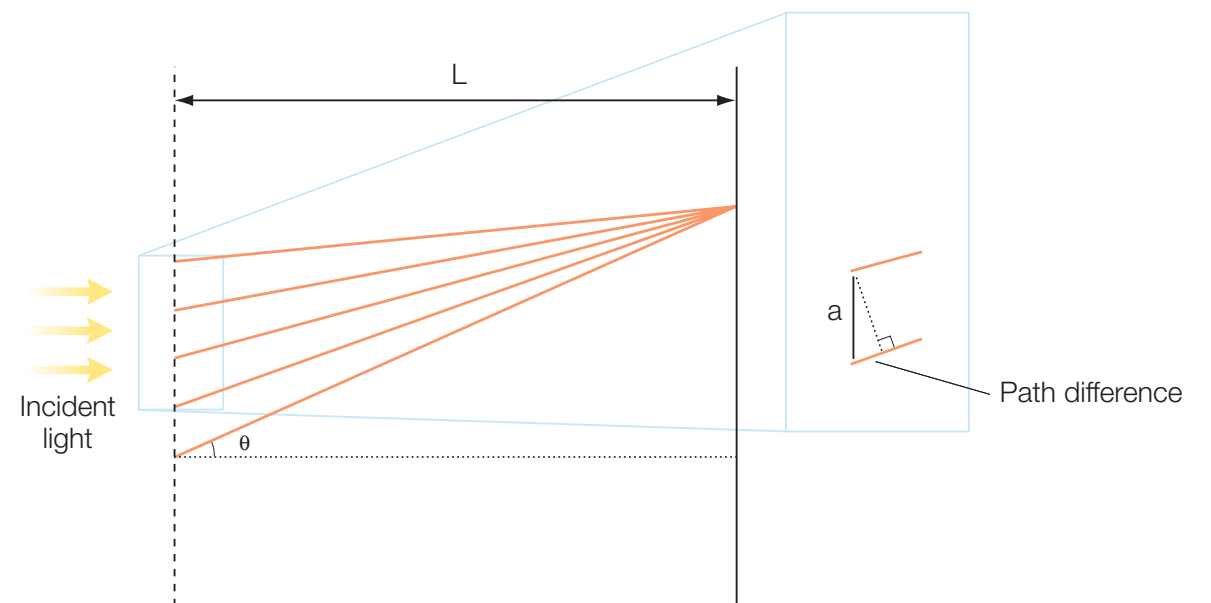
# X-Ray

If the crystallized version of the protein is available the structure can be determined by X-Ray. It is based on the observation that **waves passing through a series of slits generates specific patterns of points.**

Waves combine constructively when the phase **difference is equal to an integer number of wavelengths.**

Considering the reflection of X-Ray the points with **constructive interference** can be calculated using Bragg's Law

$$n\lambda = 2d \sin \theta$$

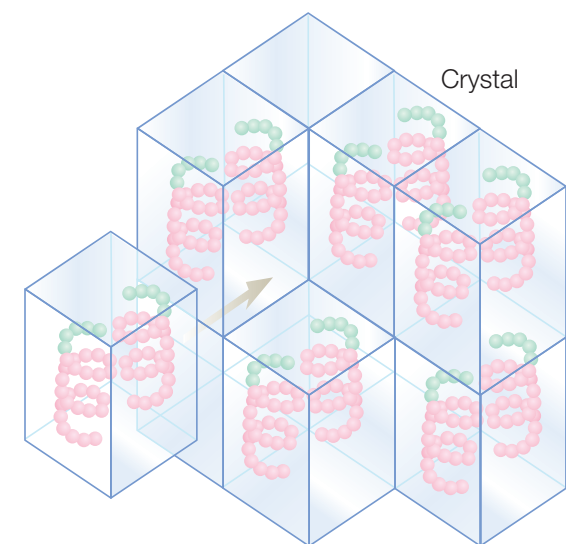
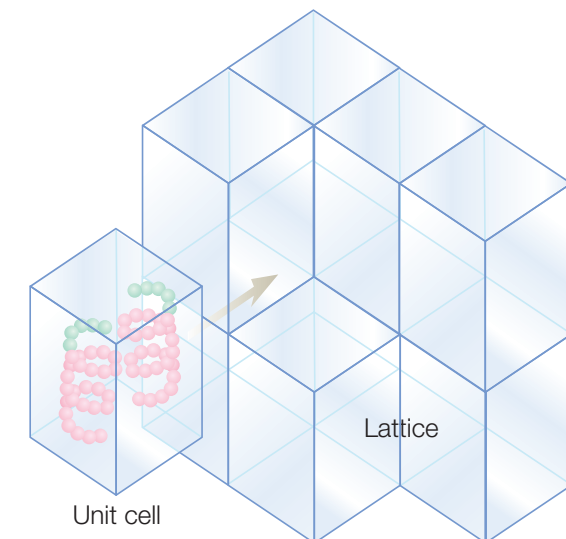


# Protein Crystals

The **formation of protein crystal is a time consuming** process that may requires considerable time, effort, and protein.

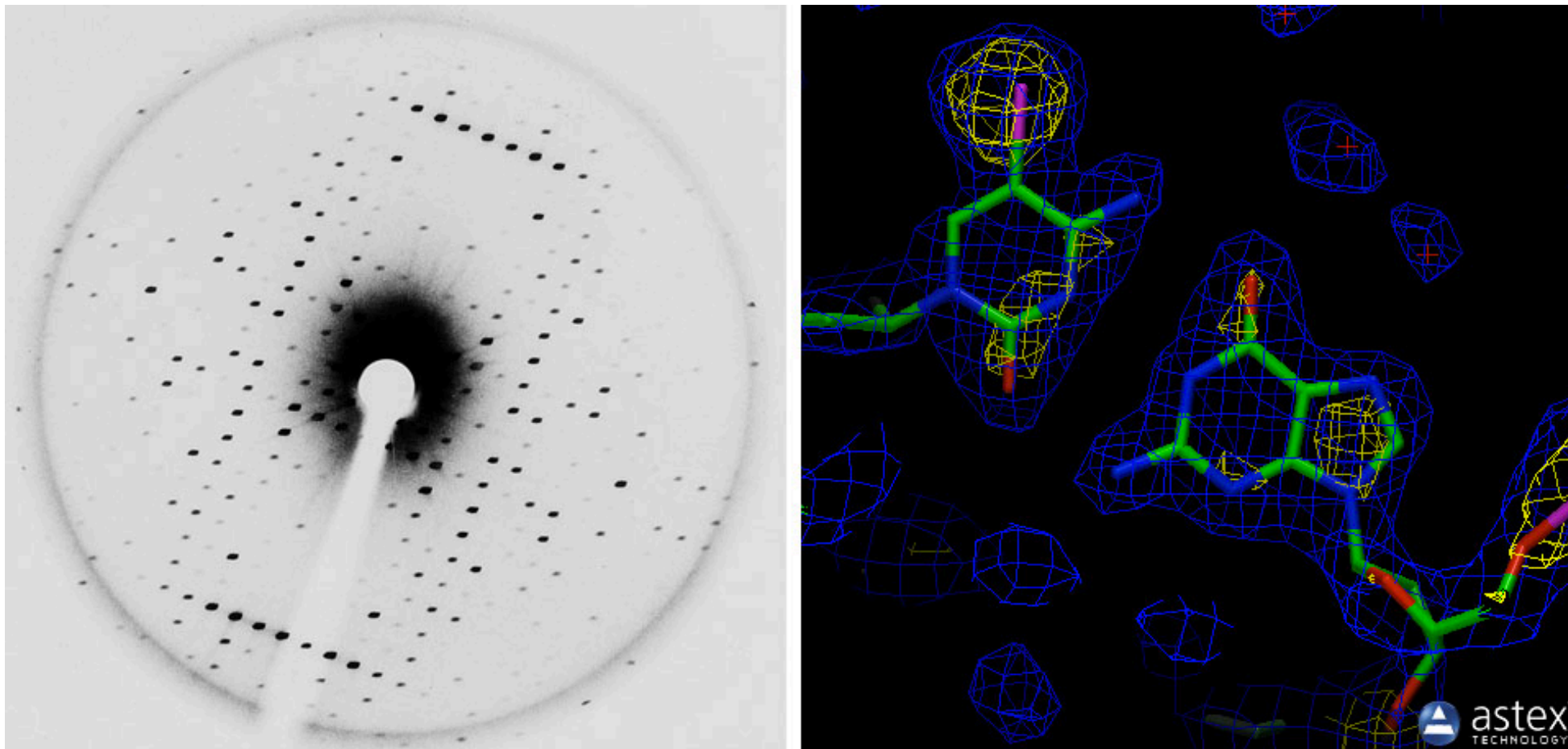
**Molecules can pack into crystal pattern called space group.** For proteins the choice of symmetries is limited because of the intrinsic asymmetry of the protein backbone (65 space groups).

In **absence of gravity the crystals grows assuming a more regular structures** ([video](#)) that allows a better determination of the three-dimensional structure of the protein.



# Electron Density Map

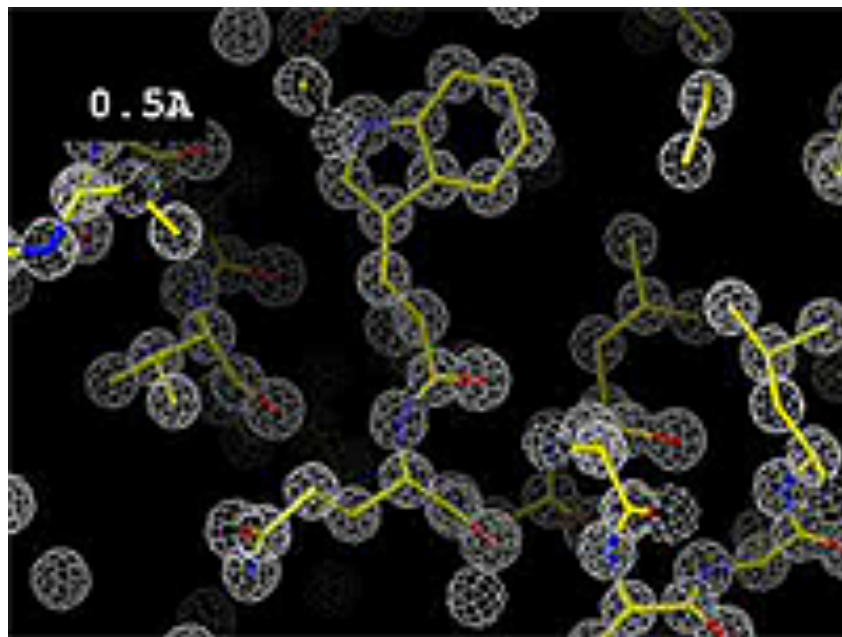
From the the three-dimensional **diffraction patten** (left) in which each spot has a characteristic intensity that is related to the **distribution of electrons in the crystal**. We derived the **electron density map** (right)



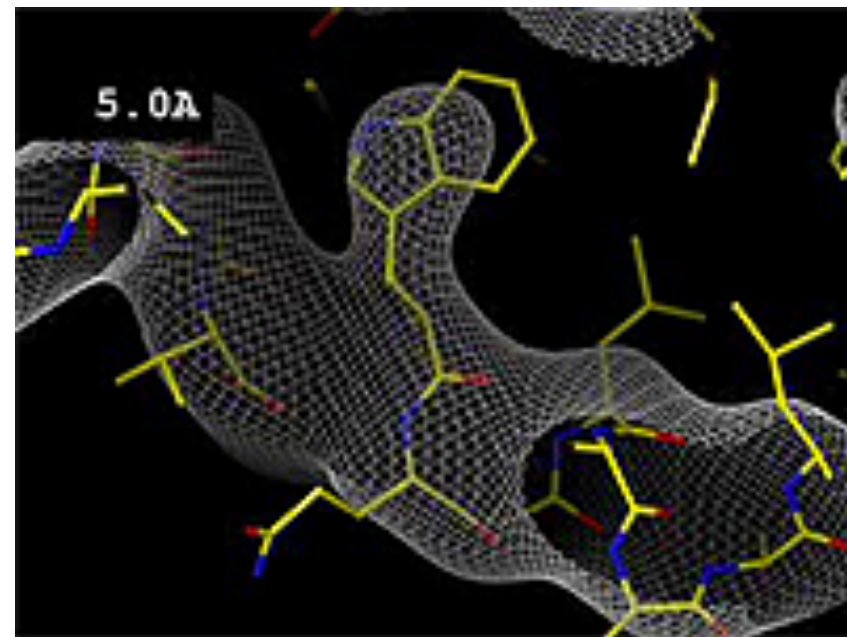
# Resolution

Resolution, in structure determinations, is the distance corresponding to the smallest observable feature. The resolution X-ray protein structures is expressed in Ångström.

High Resolution

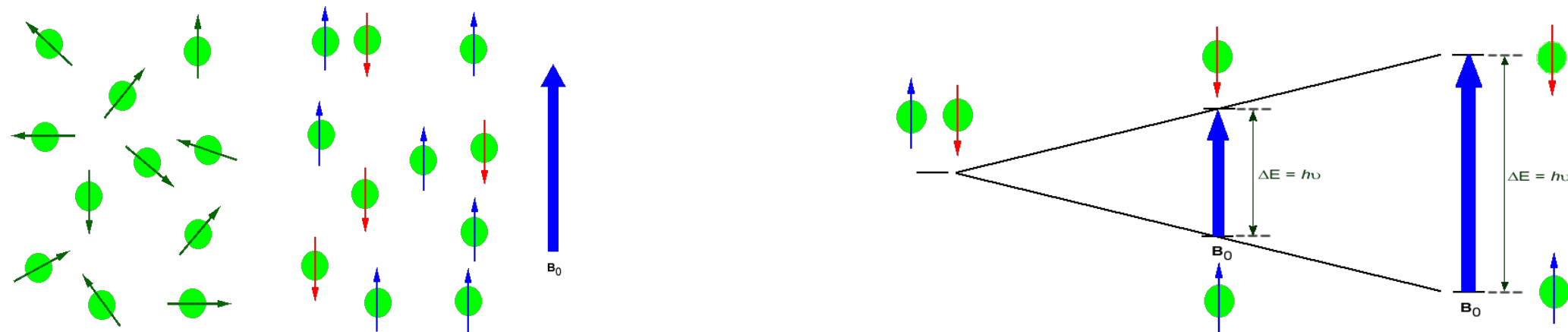


Low Resolution



# NMR Spectroscopy

Atomic nuclei which are positively charged generate an electromagnetic field. The **spin of nuclei like  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  respond to an EM radiation** with specific frequencies aligning their spin.



When the **energy of the EM radiation is equal to the difference between the two spin states**, the radiation is absorbed.

The frequency values of the absorbed EM radiation will vary according to the instrument used. To make this measure independent from the technique, the difference in resonance frequency with respect to a standard reference (tetramethylsilane) is calculated.



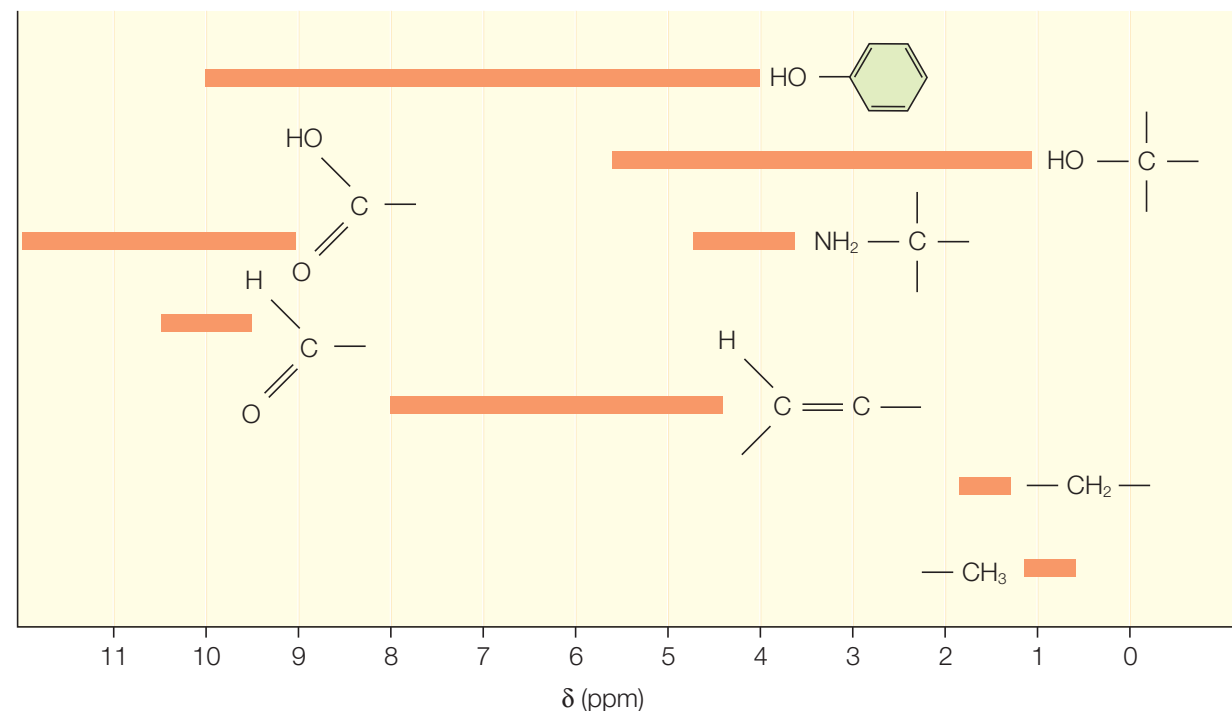
# Chemical Shift

The **chemical shift** is the difference between the frequency of absorbance of the sample and the standard reference.

$$\delta = \frac{\nu_{\text{sample}} - \nu_{\text{ref}}}{\nu_{\text{ref}}},$$

The electrons surrounding the nuclei respond causing an induced magnetic field.

- **Greater density of electrons** corresponds to a larger induced dipole and a **larger shielding**.
- **Electronegative groups** withdraw electron density from the protons, give less shielding, and **hence have larger shifts**.

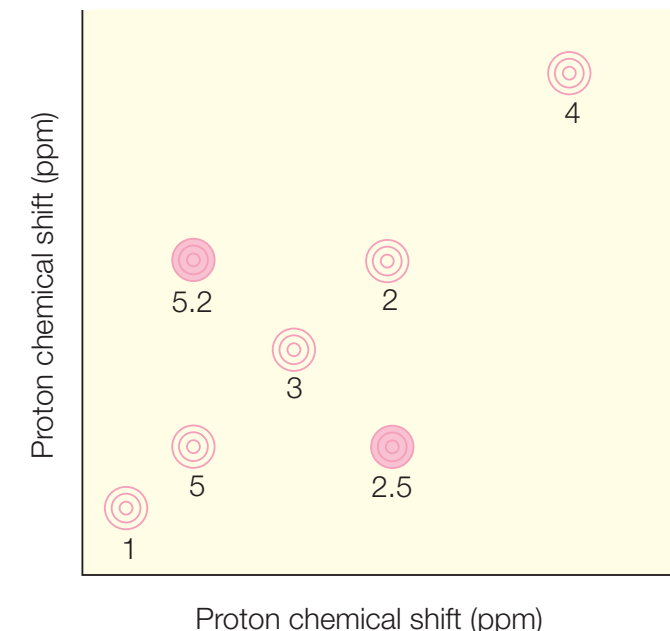
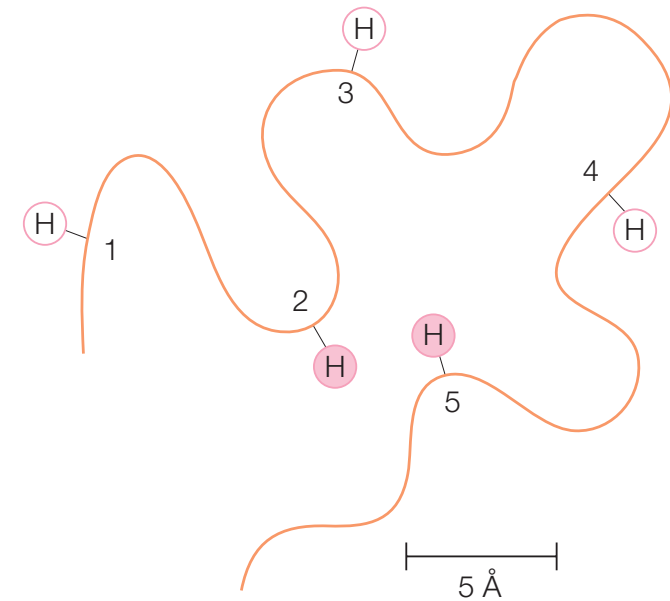


# 2D NMR

Although, one-dimensional NMR is sufficient to observe distinct peaks for the various functional groups, for larger molecules, **many overlapping resonances** can make interpretation of an NMR spectrum difficult.

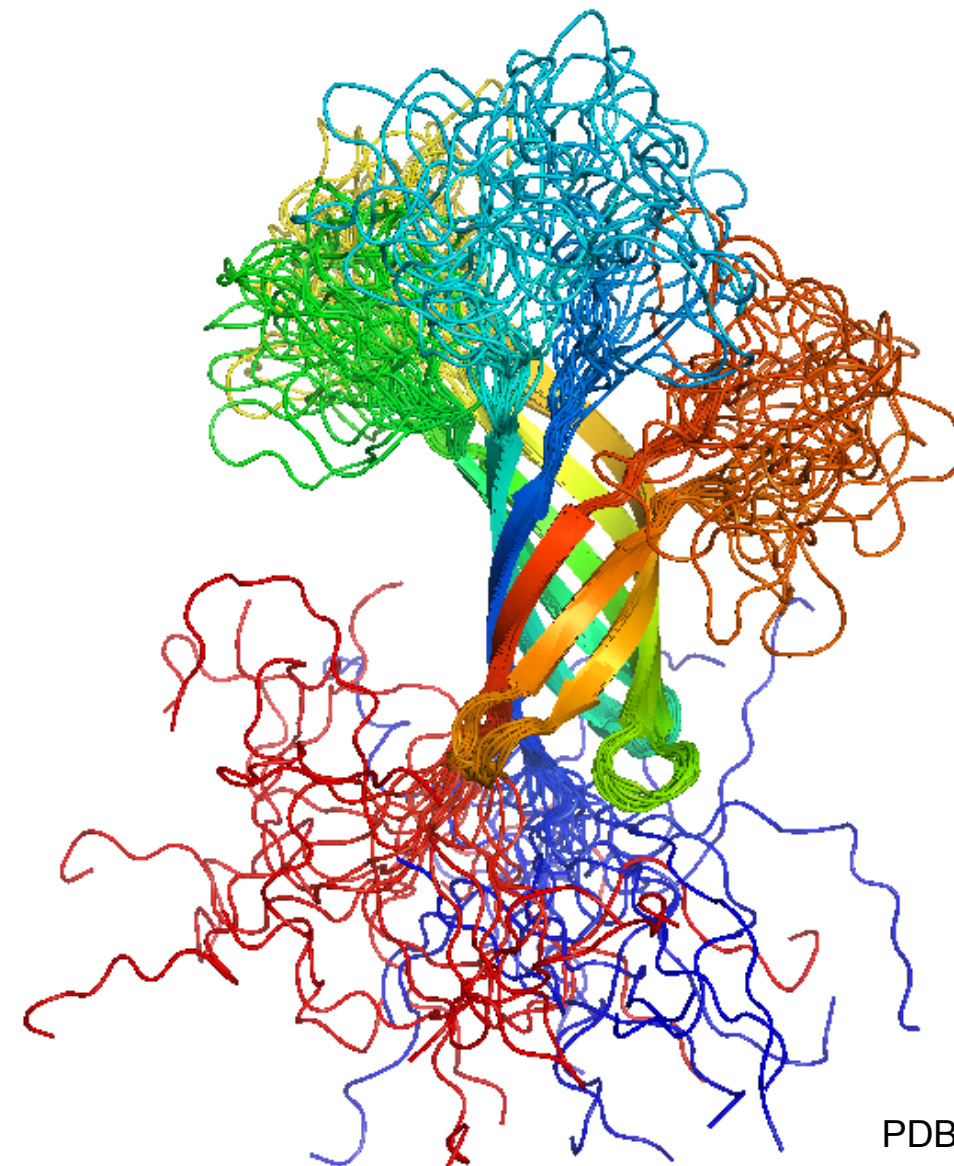
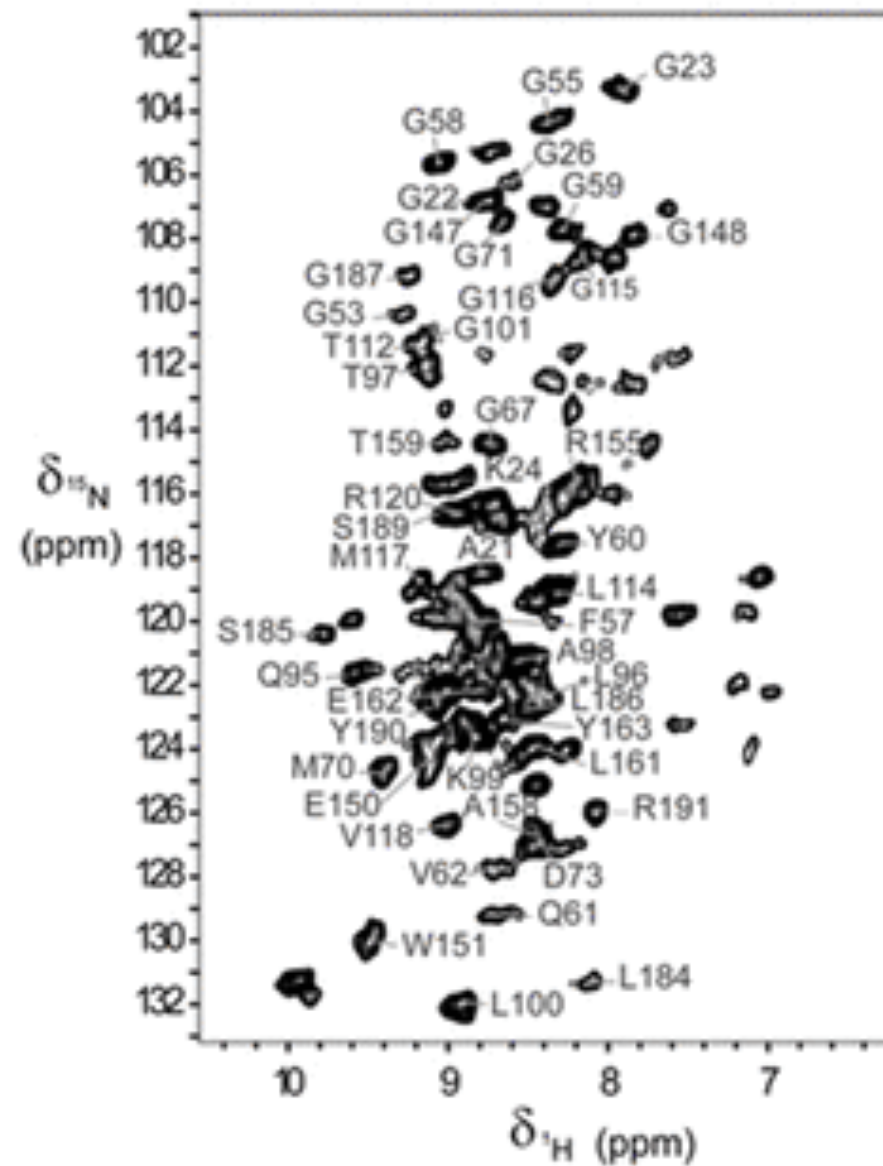
A 2D NMR experiment, however, adds an **additional dimension to the spectra by varying the length of time ( $\tau$ )**. The analysis of the acquired spectrum is useful for **determining the coupling between nuclei** that are connected.

The determination of the protein structure is based on the **Nuclear Overhauser Effect (NOE)** which allow to measure the distance between nuclei that are in close spatial proximity. The NOE arises when the spin relaxation of nuclei A is affected by the nearby nuclei B.



# NMR Spectrum

NMR spectrum is interpreted with tools that transform the experimental data into a protein structure



PDB:2K0L



# Exercise 1

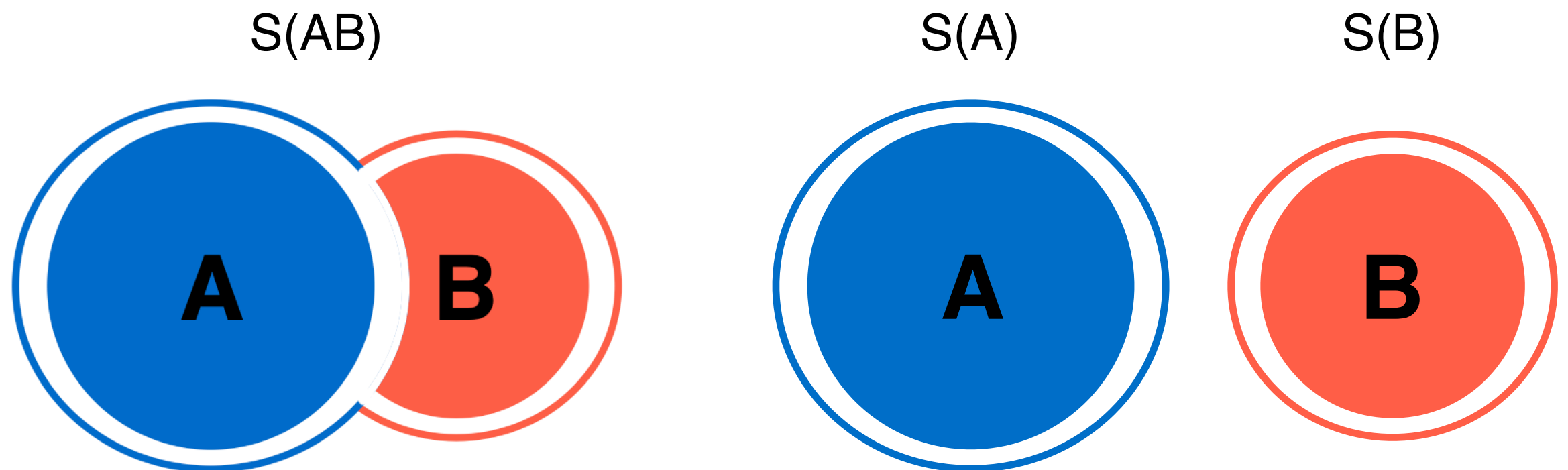
Given the structure of the Bovine Ribonuclease A (PDB: 7RSA) download the structure and measure the distances between the atoms involved in the following interactions:

- LYS37 - ASP38 (salt bridge)
- CYS40 - CYS95 (disulphide bond)
- MET29 - ARG33 (hydrogen bond)
- ILE81 - ALA102 (hydrogen bond)

**Suggestion:** select the coordinate of the atom involved in the interactions reported above.

# Exercise 2

Consider the structure of the Bacterial luciferase (PDB: 1BRL) and determine the size of the surface of interaction between the two monomers A and B. The surface of interaction can be obtained comparing the accessible surface of the complex and the single chains.



**Suggestion:** Use the [dssp-web](http://dssp-web.org/) server to generate the DSSP files of the complex and the monomers. Each file includes the accessible surface of all the residues.