# Foundations in Biomedical Sciences – Lecture 2

Gordon Rule
Department of Biological Sciences
Carnegie Mellon University

rule@andrew.cmu.edu

Amino Acids
- pKa, chemical and environmental effects
- Measuring pKa
- UV-Vis absorption
- Fluorescence properties of Trptophan

Protein Structure and stability
- Primary, secondary, tertiary structure
- Standard energy
- Entropic effects
- Enthalpic effects

Protein Purification Methods
- Affinity chromatography
- Ion exchange
- Size exclusion

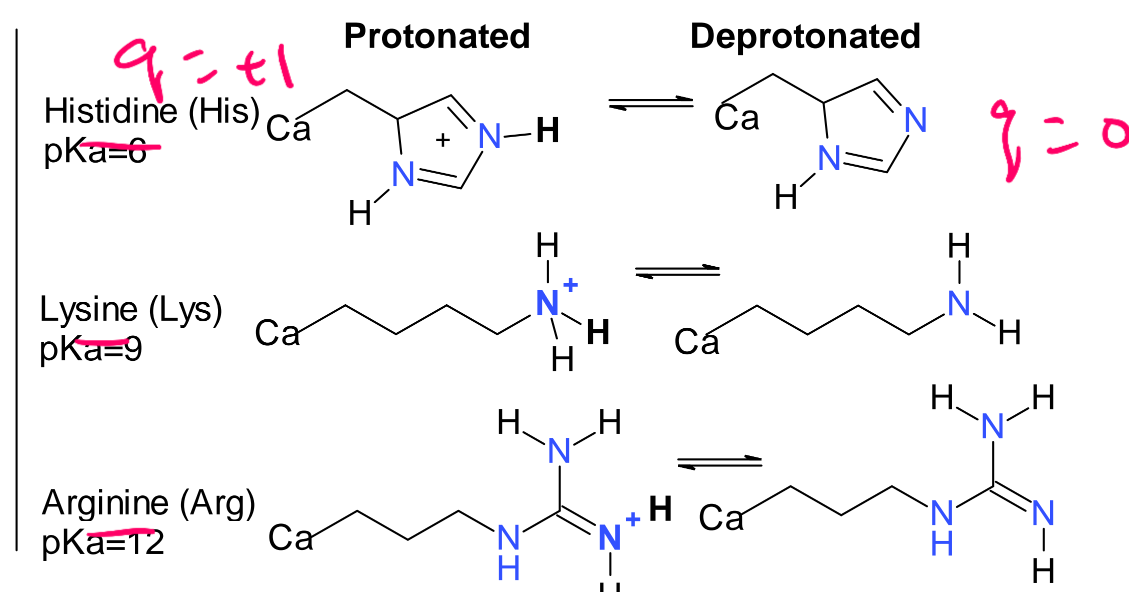
Expression of Recombinant Proteins
- Transcription and Translation
- E. Coli and T7 RNA Polymerase
- Lac operon as a control system
- pET vectors
- Restriction enzymes & DNA Ligase

# Acidic Groups in Amino Acids & Proteins

## A. Key Groups

| | Protonated (pH<<pKa) | Deprotonated (pH>>pKa) |
|---|---|---|
| Carboxy-Terminus pKa=2 | | |
| Amino-Terminus pka=9 | | |

| | Protonated | Deprotonated |
|---|---|---|
| Aspartic Acid (Asp) pKa=4 | q = 0 | q = -1 |
| Glutamic Acid (Glu) pKa=4 | | |




B. Other sidechain ionizations that are usually less important for function:

- Tyr-OH $pK_a$=10
- Cys-SH, $pK_a$=8.

Tyr

Cys

C. Other sidechain that do not ionize over a pH range of 0-10:

- Amide (Asn, Gln)
- Alcohol (Ser, Thr)
- Indole (Trp)

What Determines pKa values?

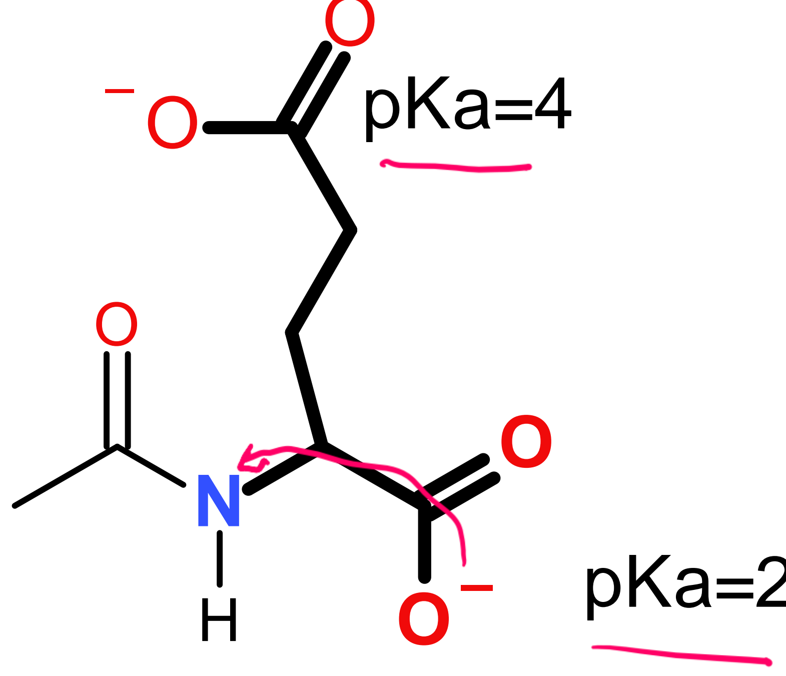
- Chemical Effects
- Environmental Effects

**Chemical Effects**

*Consider a terminal glutamic acid in a protein.*

*Why do the pKa values of the two carboxylates differ? Explain why one is 2 and the other is 4.*

Full negative charge on the oxygen is unfavorable.




*What will the electronegative nitrogen do to the negative charge?*

**Environmental Effects on pKa Values:**

In addition to chemical bonding effects, the pKa values of groups can change due to environment.

i.e. a protein with multiple His residues will likely show different pKa values for each His residue.

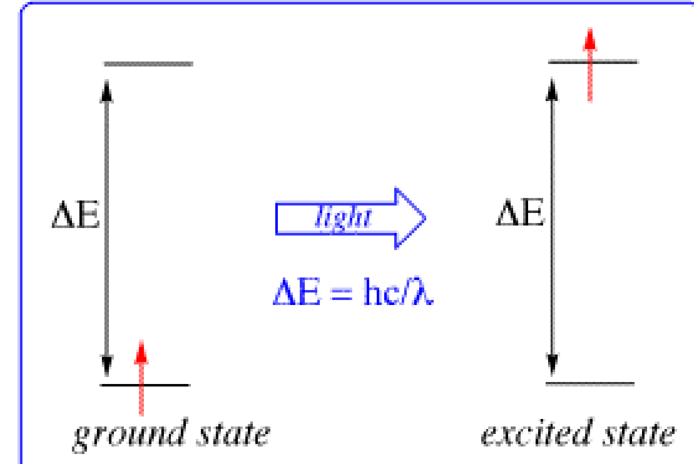
Local charges can shift pKa values.

*How do we measure the pKa of individual residues in a protein?*

# Measuring pKa with NMR - Biochemical Spectroscopy

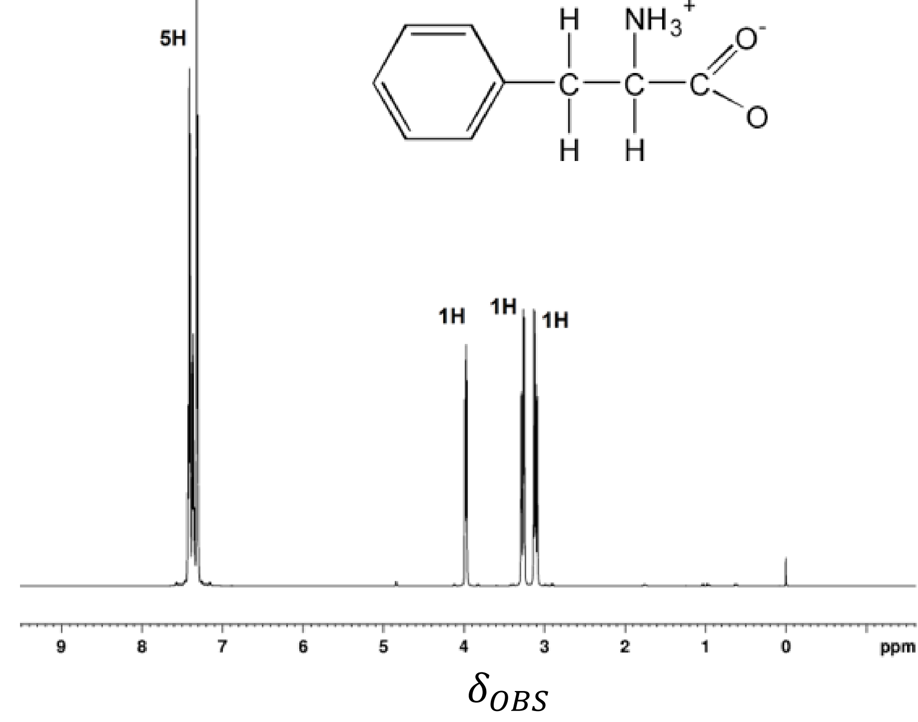
**Spectroscopy:**

- System exists in a ground and higher energy excited state.
- Light of a specific wavelength is absorbed.
- Energy of light corresponds to energy difference between ground and excited state
- Absorption spectrum is the amount (intensity) of energy absorbed versus wavelength.




**Nuclear Magnetic Resonance (NMR):**

- Light is absorbed by nuclear spins.
- Spins go from one direction to the other (spin up/spin down)
- Absorption energy is indicated by chemical shift (δ).



# Experimental Measurement of pKa – pH Titrations

$$\delta_{OBS} = f_{HA} \times \delta_{HA} + f_A \times \delta_A$$

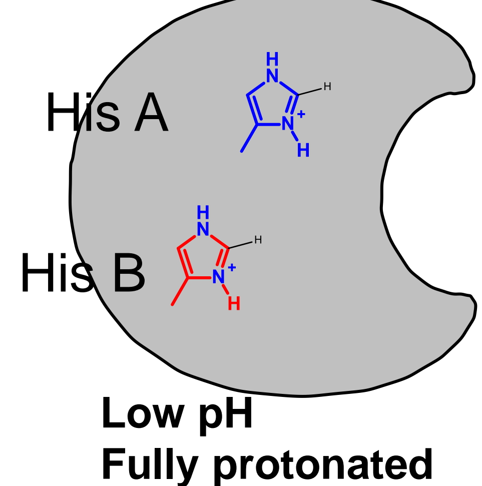
**A.** Chemical shift depends on the local environment.
**B.** Chemical shift of highlighted proton changes due to protonation/deprotonation.

**C.** Observed shift is weighted average and can be used to determine fraction protonated and deprotonated.

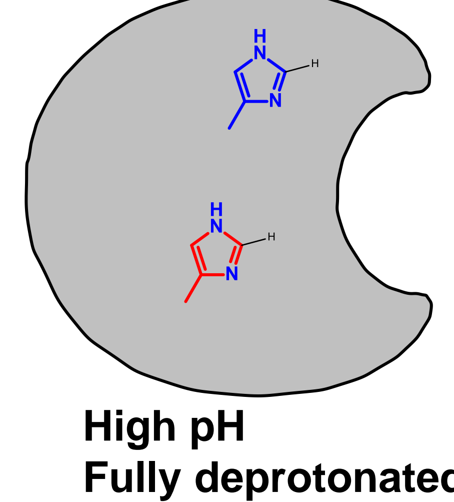
His A – deprotonation increases chemical shift (from 7.4 to 10.0). pKa = 6.5

His B – deprotonation decreases chemical shift (from 9.8 to 8.3) pKa = 5.0

**Example:** Protein with two histidines.

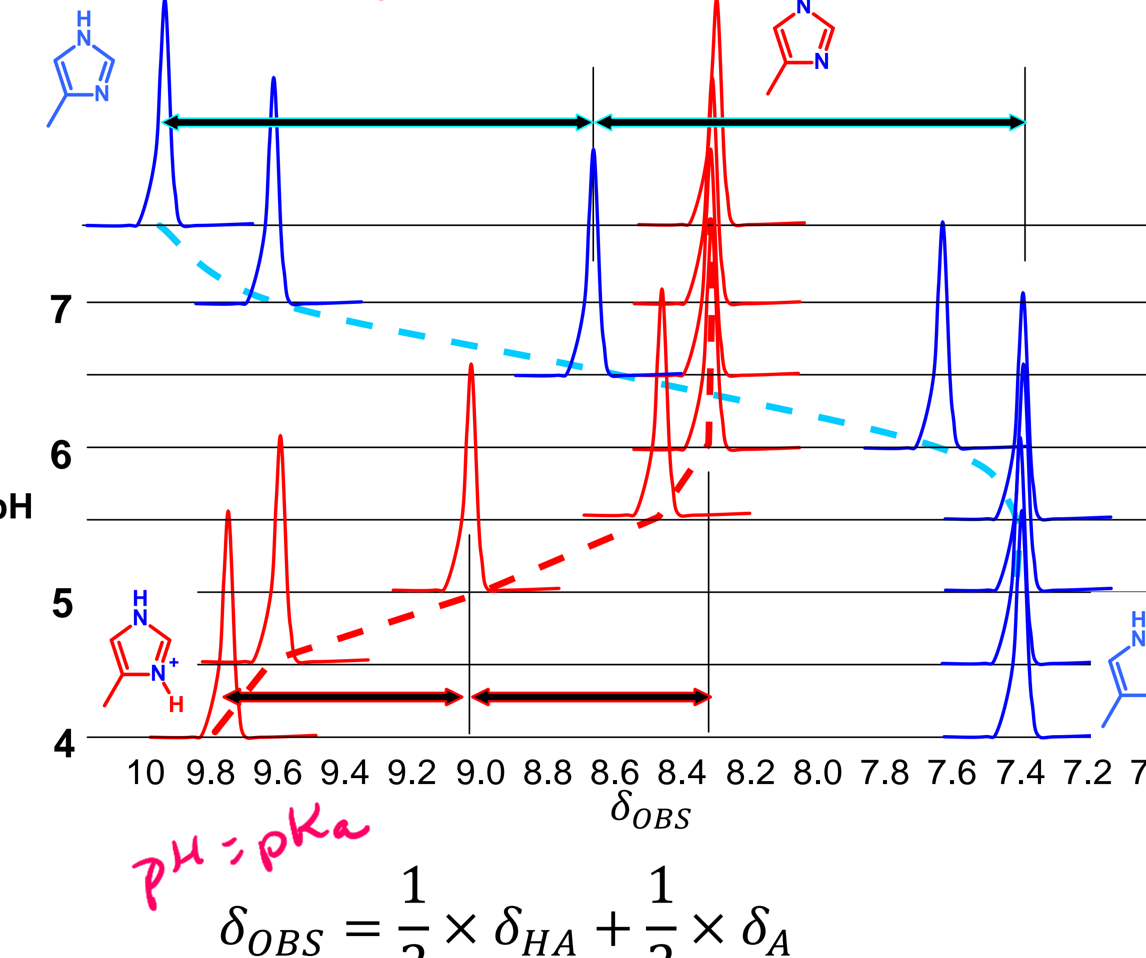



**Low pH**
**Fully protonated**



**High pH**
**Fully deprotonated**

**NMR spectra acquired at indicated pH**

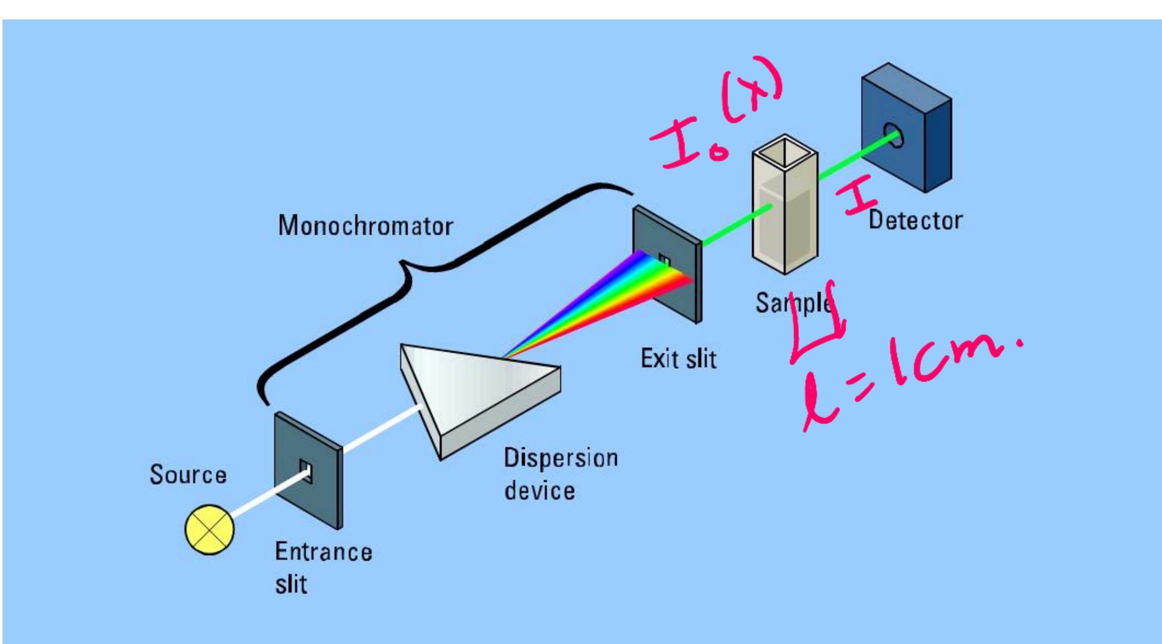



pKa: pH where peak is ½ way between two extremes

pH = pKa

$$\delta_{OBS} = \frac{1}{2} \times \delta_{HA} + \frac{1}{2} \times \delta_A$$

# Practical Biochemistry - UV Absorption of Proteins to Determine Concentration



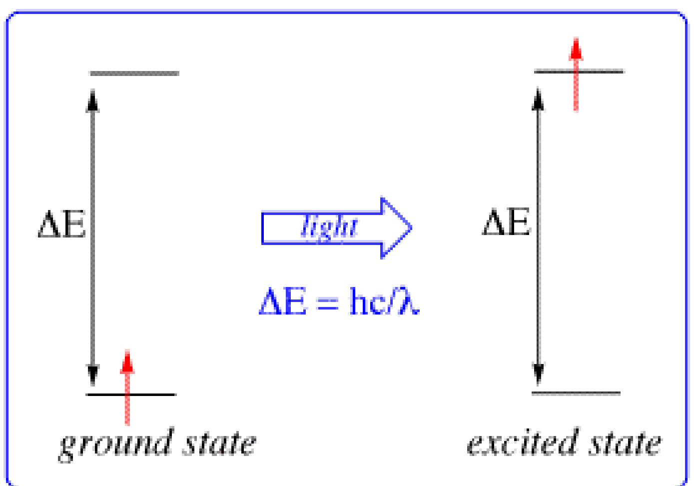

$$A = \log \frac{I_0}{I} = \varepsilon [X] l$$

where

- A is the absorbance of the sample;
- $I_0$ is the intensity of the incident light;
- I is the intensity of the light that leaves the sample.
- ε is the molar extinction coefficient at a specific wavelength
- [X] is the concentration of the absorbing species
- l is the path length (usually 1 cm).

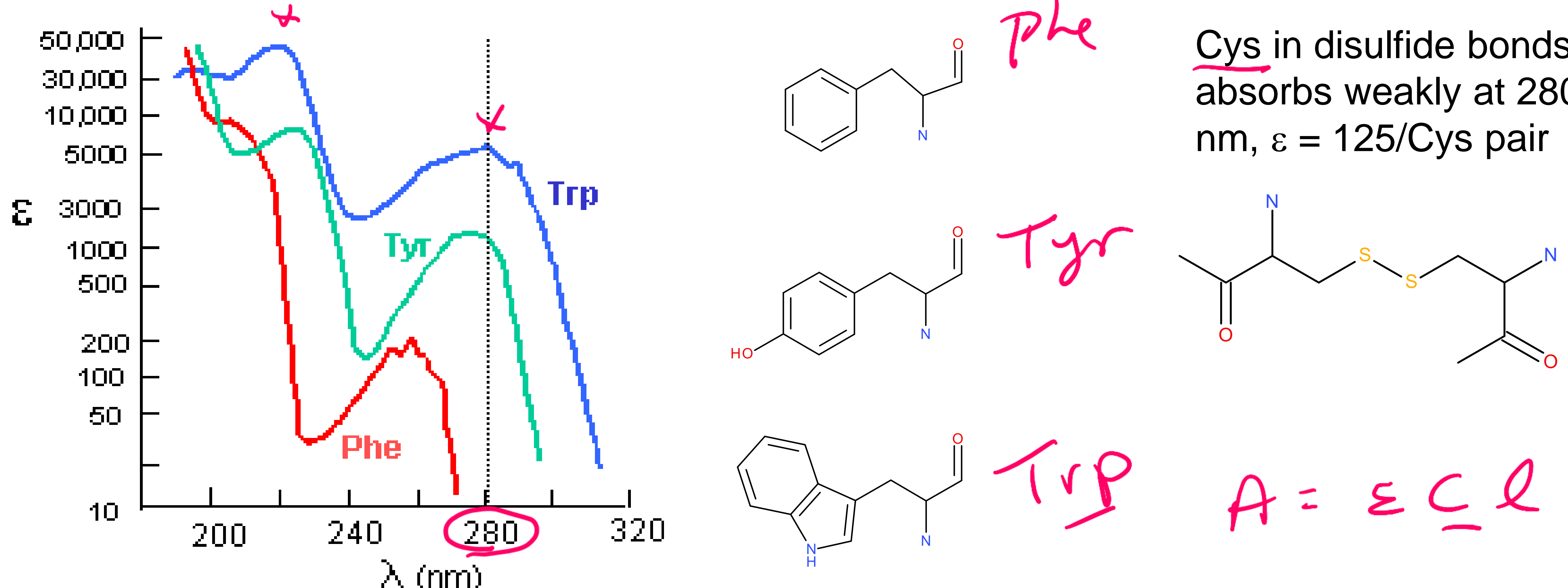
**UV-Vis:**

- Light is absorbed by electrons.
- Electrons go from low energy orbital (ground) to higher energy orbital (excited).
- Energy released as heat in non-fluorescent chromophores.
- Absorption energy indicated as wavelength (nm)



# UV Absorption Properties of Proteins

Three aromatic amino acids absorb light in the ultraviolet range (UV).




The extinction coefficients (or molar absorption coefficients) of these amino acids are:

| **Amino acid** | **$\lambda_{MAX}$** | **Extinction Coefficient, ε, (at 280 nM)** |
|---|---|---|
| Trp | 280 nm | 5,500 $M^{-1}cm^{-1}$ |
| Tyr | 274 nm | 1,490 $M^{-1}cm^{-1}$ |
| Phe | 260 nm | ~0 $M^{-1}cm^{-1}$ |

# UV Absorption Properties of Proteins

**Calculation of molar extinction coefficients:**
If a molecule contains a mixture of N different chromophores, the molar extinction coefficient can generally be calculated as the sum of the molar extinction coefficient for each absorbing group in the protein:

$$\varepsilon_{Protein} = \sum_{i=1}^{N} \varepsilon_i$$

- Therefore, the molar extinction coefficient for a protein can be calculated from its amino acid composition.

**Example**:
A protein has two Tryptophan (Trp) residues and one Tyrosine (Tyr) and one Phe, what is its extinction coefficient at 280 nm?

$\lambda = 280$

$\varepsilon_{protein} = 2 \times 5{,}500 + 1 \times 1490$

| | Conc [X] | Abs | ε=A/[X] |
|---|---|---|---|
| Solution of free Trp | 200 uM (Conc of Trp) | 1.1 | 5,500 $M^{-1}cm^{-1}$ |
| Protein with two Trp residues | 100 uM (Conc of protein) | 1.1 | 11,000 $M^{-1}cm^{-1}$ |

| Amino acid | Extinction Coefficient, ε, (280 nm) |
|---|---|
| Trp | 5,500 $M^{-1}cm^{-1}$ |
| Tyr | 1,490 $M^{-1}cm^{-1}$ |
| Cys-Cys | 125 $M^{-1}cm^{-1}$ |
| Phe | ~0 $M^{-1}cm^{-1}$ |

Calculation of pI and other properties are done via web servers: https://web.expasy.org/protparam/

# Fluorescence Properties of Tryptophan

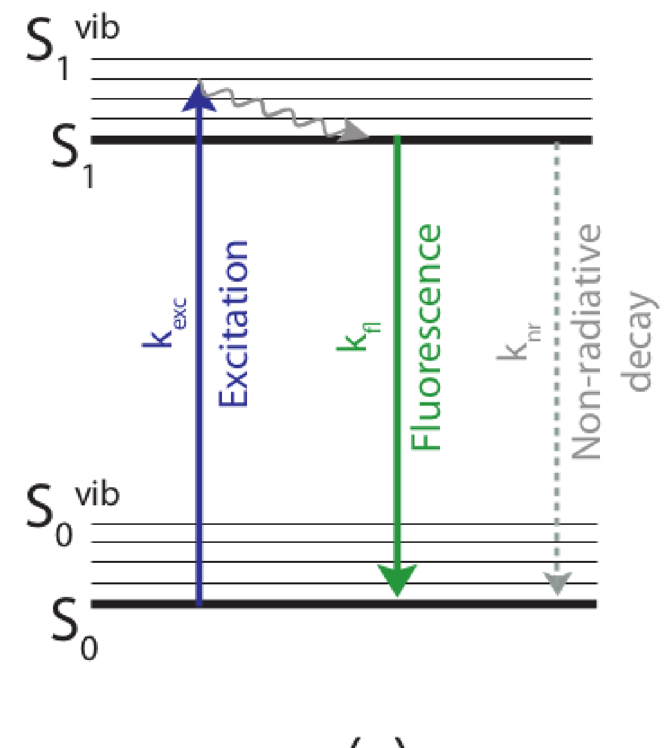
*Why do we care? The wavelength and intensity of Trp fluorescence are sensitive to changes in protein structure, providing a convenient way to measure stability of the protein and the binding of molecules to the protein.*

**Absorbance:**

- Absorbed photon moves electron from the low energy ground state to an empty orbital.
- Energy of excited state drops due to a number of reasons, including a response to the environment.
- When electron goes back to the ground state its *energy is released as heat* (non-radiative decay)

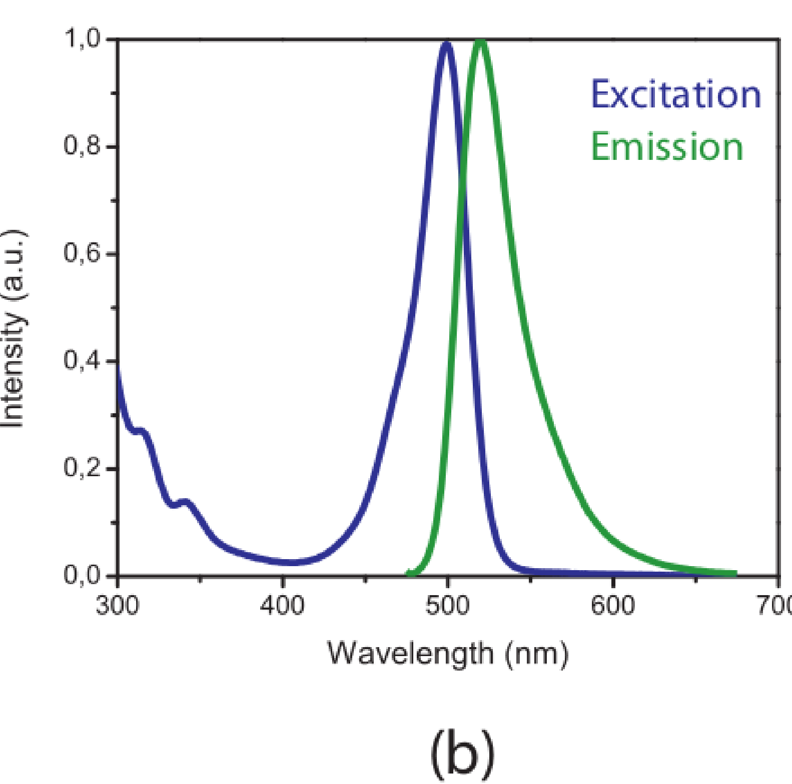
**Fluorescence**:

- Absorbed photon moves electron from the low energy ground state to an empty orbital.
- Energy of excited state drops due to a number of reasons, including a response to the environment.
- When electron goes back to the ground state its *energy is released as a photon* (radiative decay)




(a)

- Intensity of fluorescence depends on relative rates of fluorescence and Non-radiative decay.



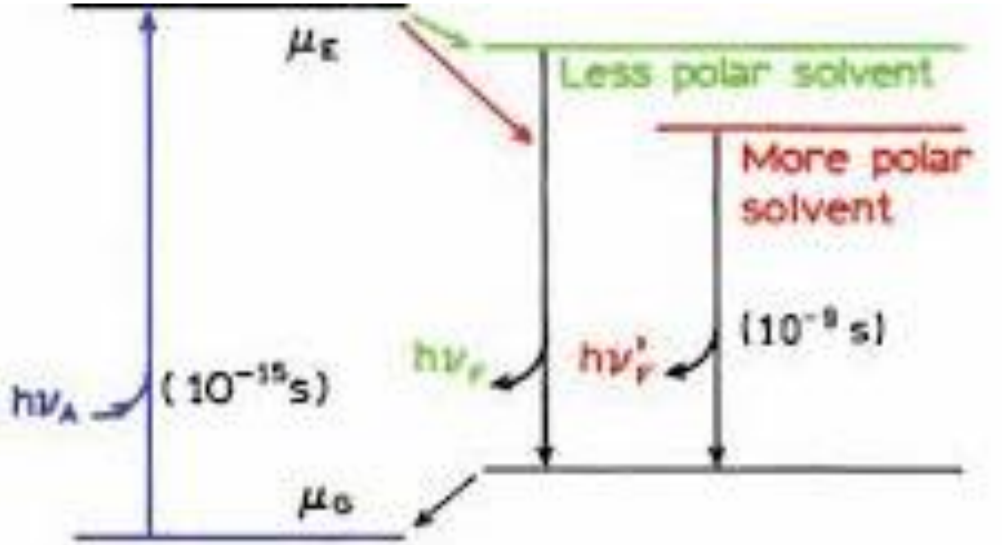

(b)

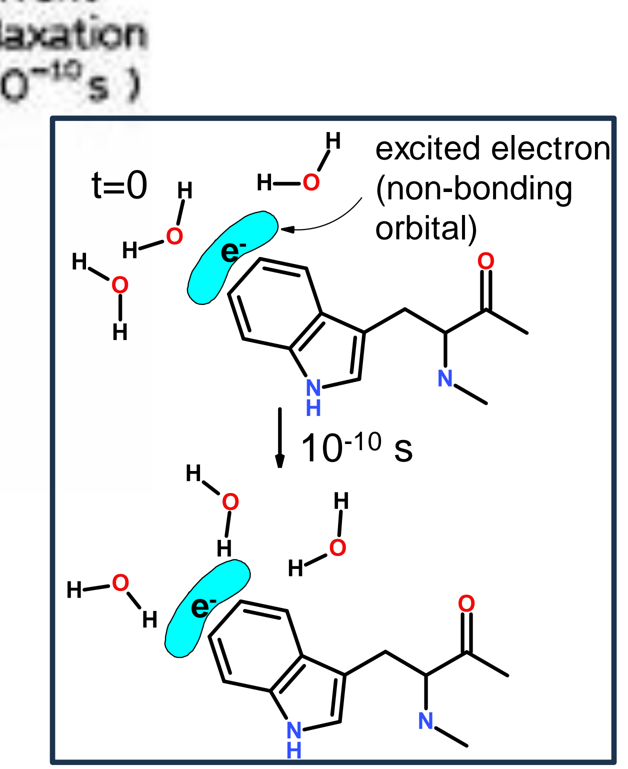
- The emitted photon is always of longer wavelength due to the drop in energy of the excited state.

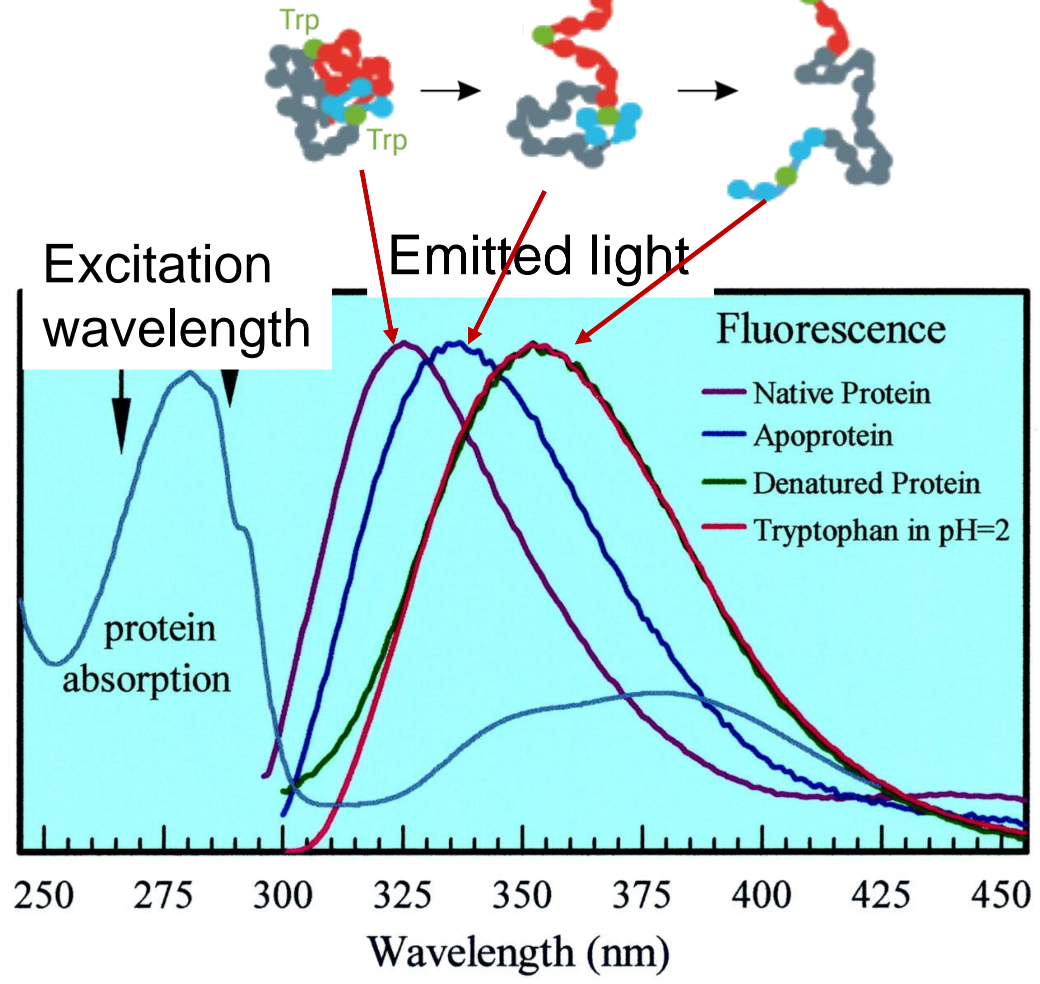
# Fluorescence Properties of Tryptophan

- Tryptophan is strongly fluorescent, absorbing at 280 nm and emitting at ~350 nm.
- Tyr and Phe have weaker fluorescence and are of little practical use due to the low intensity of fluorescence.

- The Trp emission wavelength depends on the solvent polarity:
  - Buried Trp emits at ~320 nm (non-polar environment), excited state is not stabilized by water dipoles.
  - Solvent exposed Trp emits at ~360 nm (polar), lower energy due to stabilization of excited state by water.




- The above curves are scaled to be the same intensity. Unfolding can cause either a decrease or increase in fluorescence intensity, depending on the environment of the Trp in the folded state.

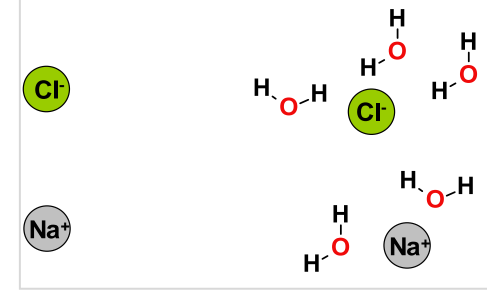
# Protein Structure & Stability

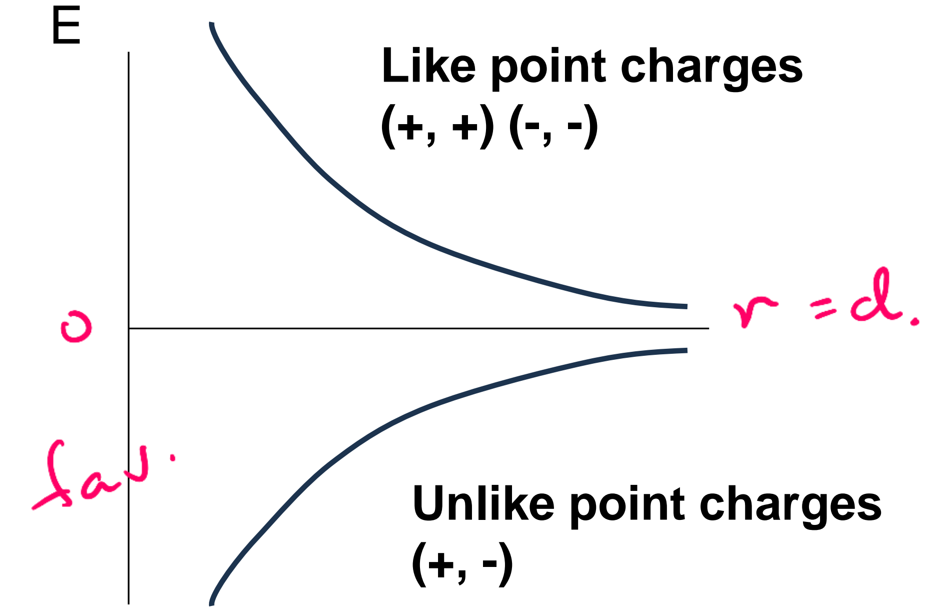
# Molecular Forces & Protein Stability

**i) Electrostatics:** The interaction energy between two charged particles is:

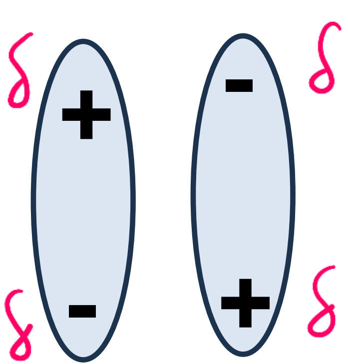
$$E = \frac{q_1 q_2}{Dr}$$

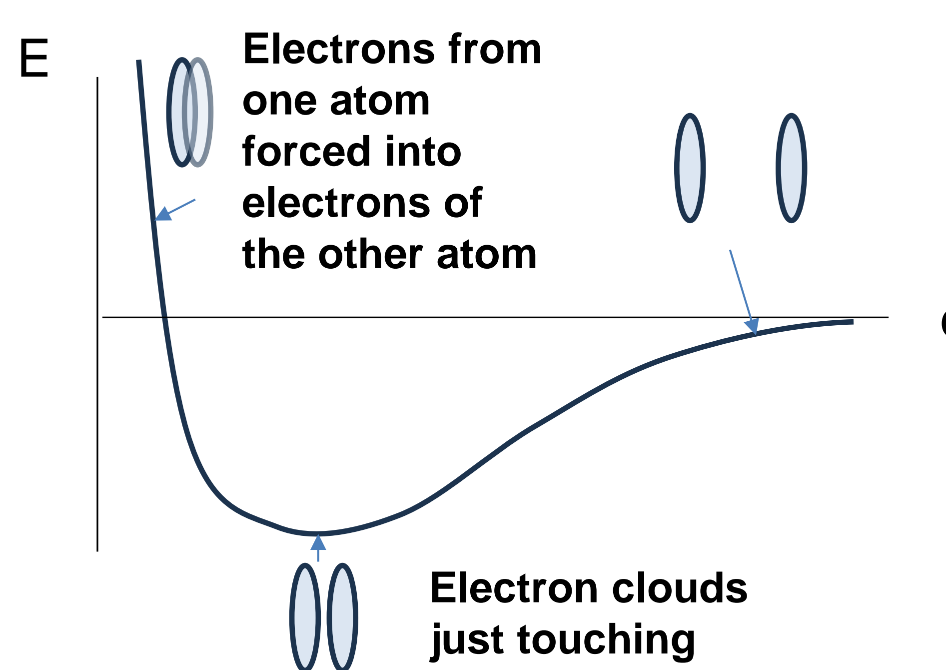
- The energy depends on the charges of the particles ($q_1$, $q_2$), distance (r) between the two charges, and the dielectric constant (D) of the media.
- The dielectric constant depends on the solvent. Water has a high dielectric constant because it is a polar solvent.
- The water molecules will orient themselves around the ions, reducing the effective charge on the ion. The high dielectric constant of water reduces electrostatic interactions to ~10 kJ/mol



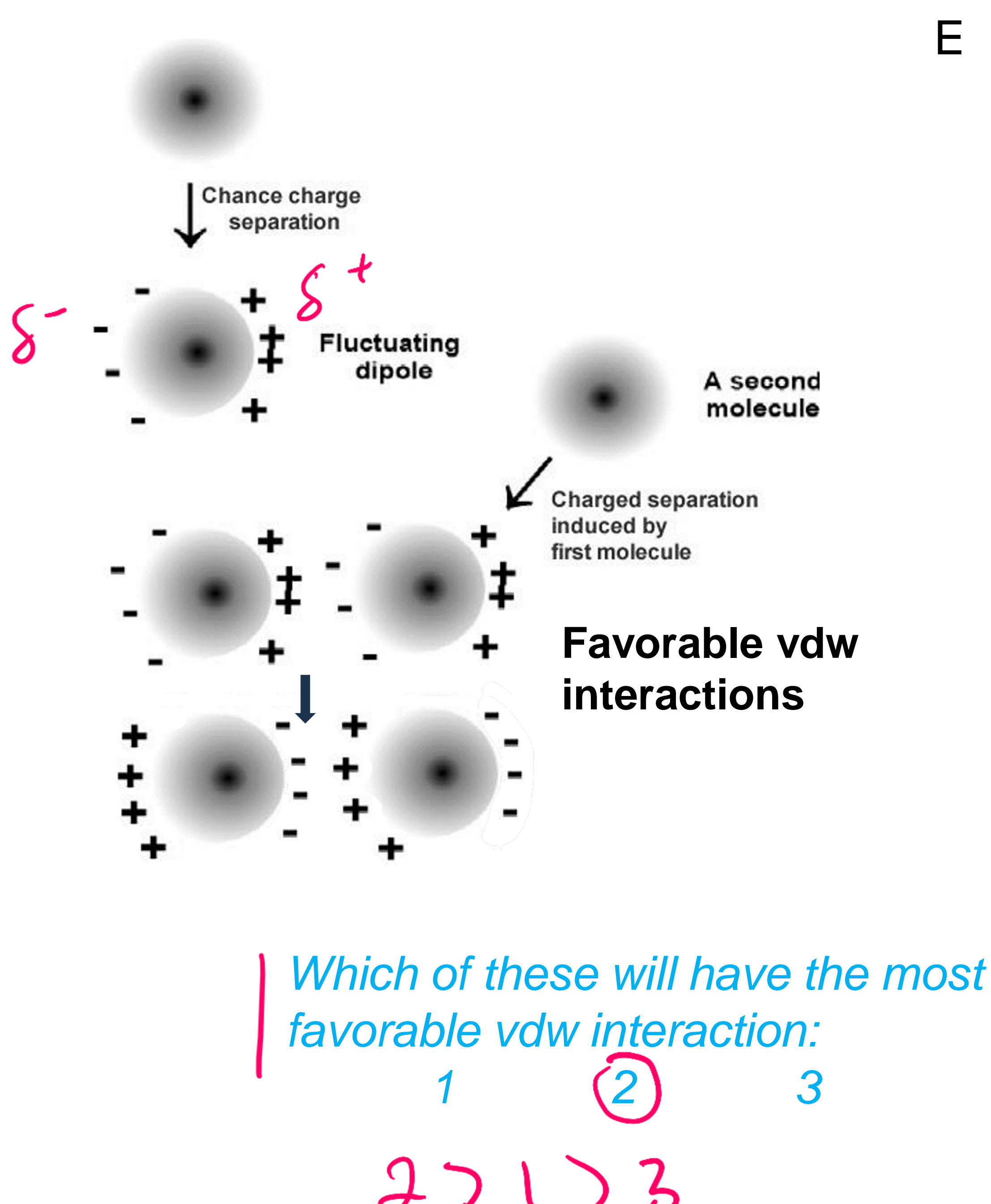

**ii) Van der Waals/ Keesom** (dipole-dipole) – an electrostatic interaction that involves ***permanent*** partial charges. It does **not** involve **formal charges**.

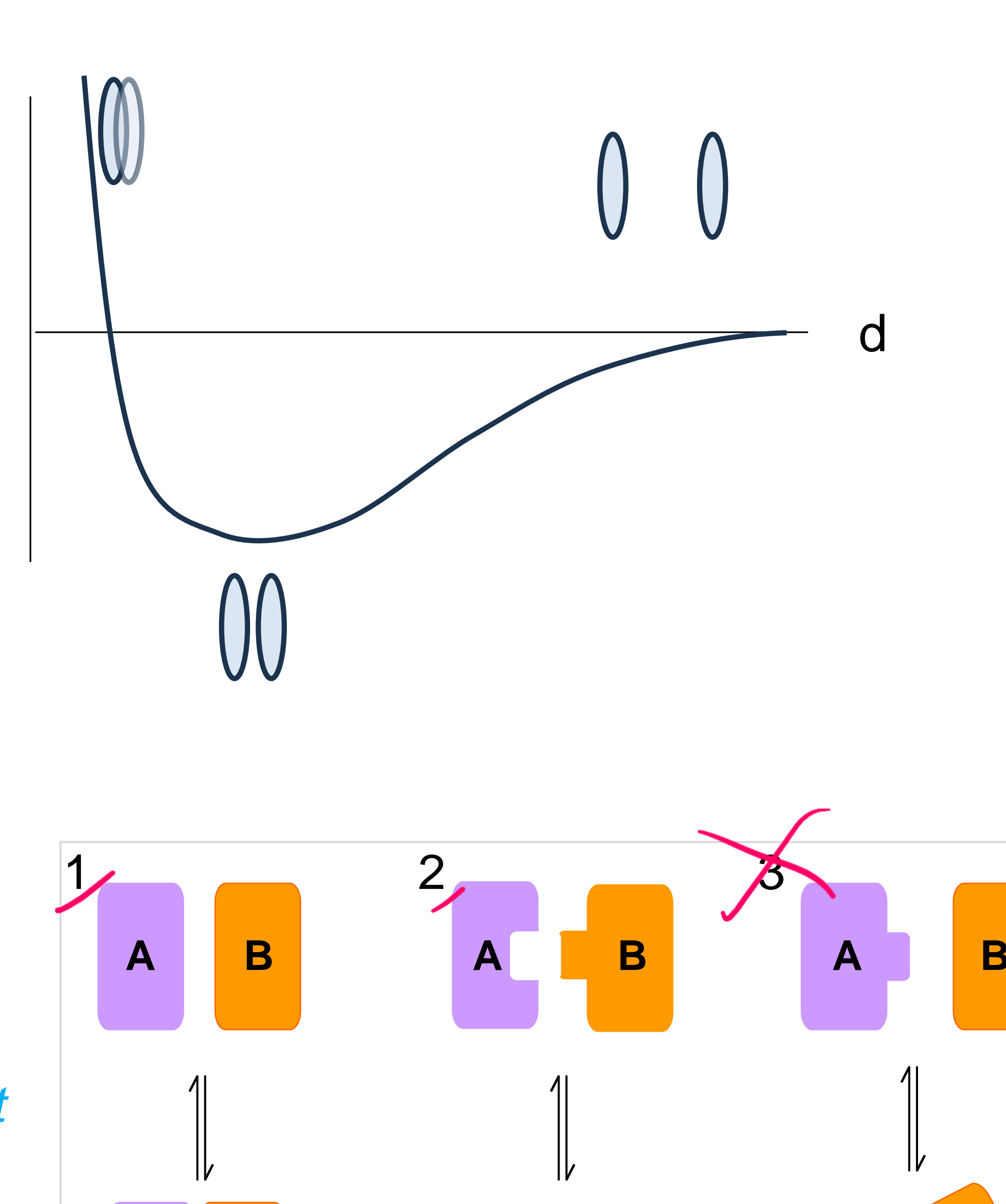


# iii) Van der Waals: London Force - Induced dipole

- Fluctuation of electron density leads to charge distribution on a neutral atom.
- The charge imbalance polarizes near-by atom
- An attractive electrostatic force is generated.
- Charge reversal quickly occurs, but response of other atoms maintains attractive interaction



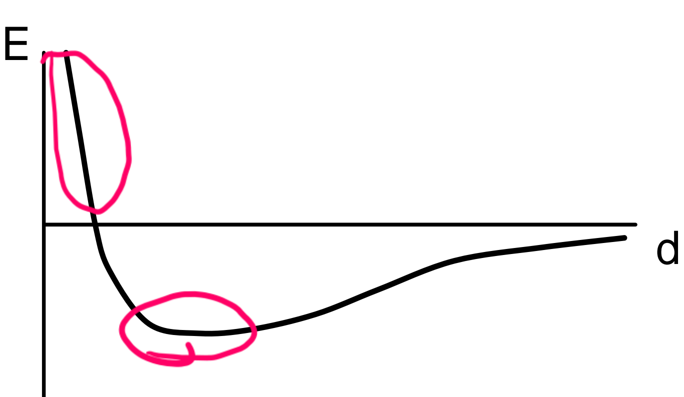



# Secondary Structure

**Goals:**

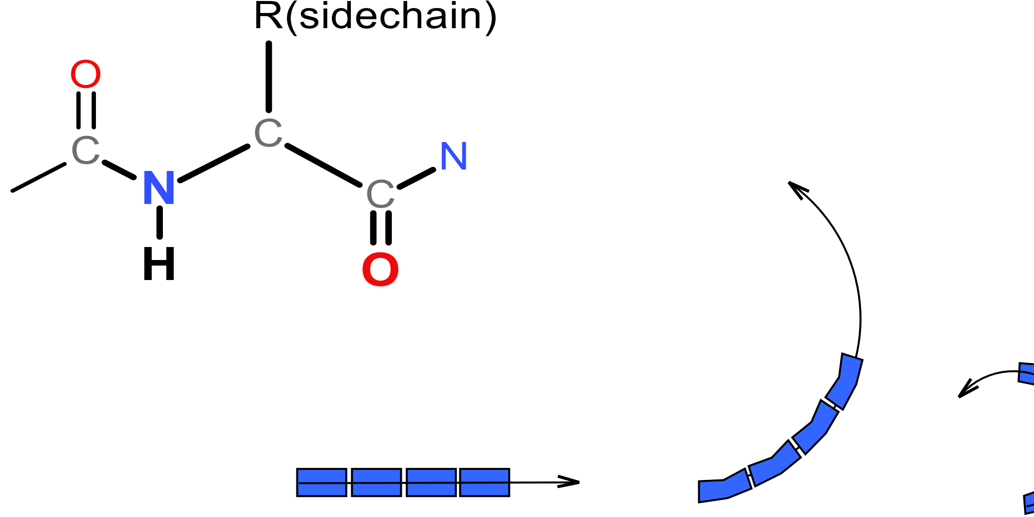
- Why are peptide bonds planer and usually *trans*.
- Why H-bonding and van der Waals restrict secondary structures.
- Generation and interpretation of Ramachandran plots

H-bonding (mainchain atoms)



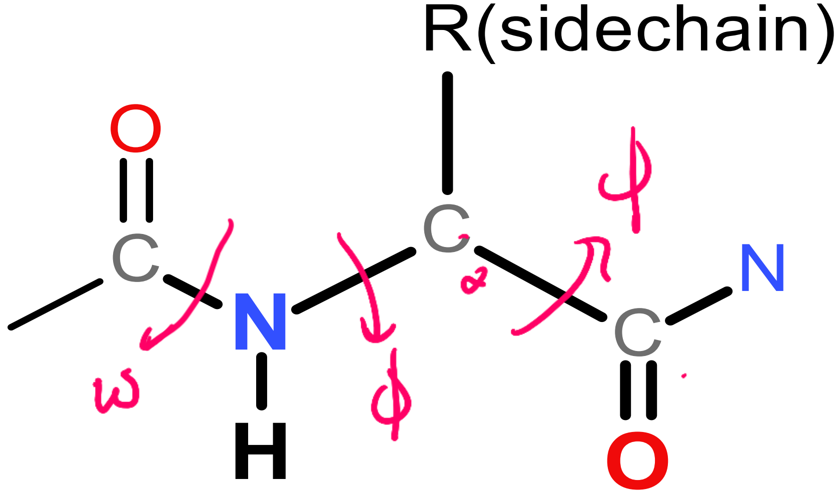
**Regular Secondary Structure:**

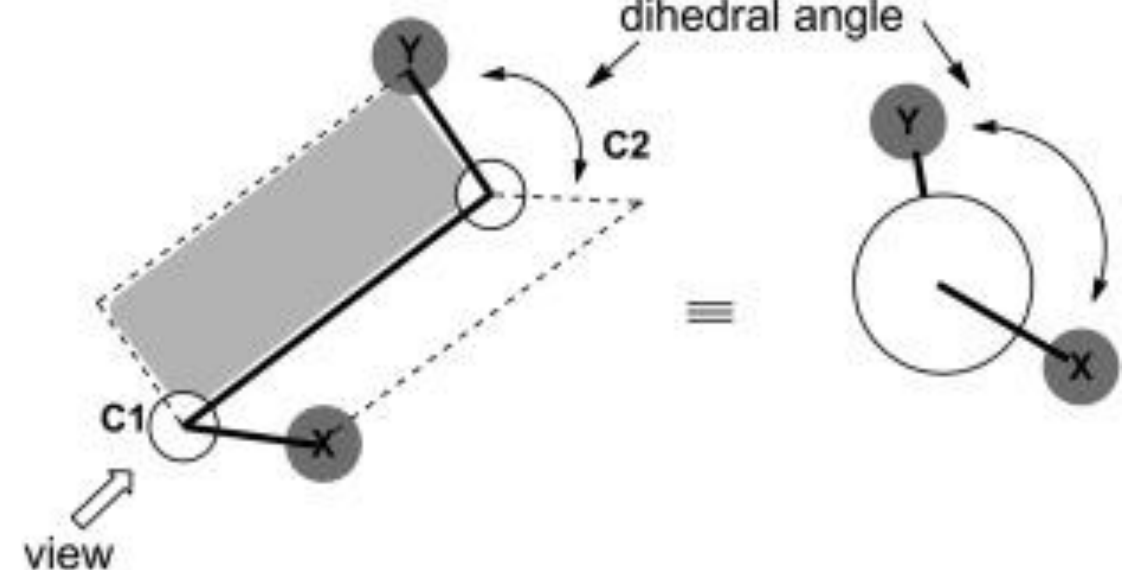
- Local conformation of only the ***mainchain*** atoms.
- Regular secondary structure because each residue within a secondary structural element has the same geometrical shape due to identical rotational angles about bonds in each residue.
- The repeating geometrical properties of the subunits will result in a three-dimensional shape if the units are laid end to end:
  - Rectangular blocks will generate a linear shape.
  - Curved blocks will generate curved shapes, these will be helices in three dimensions.



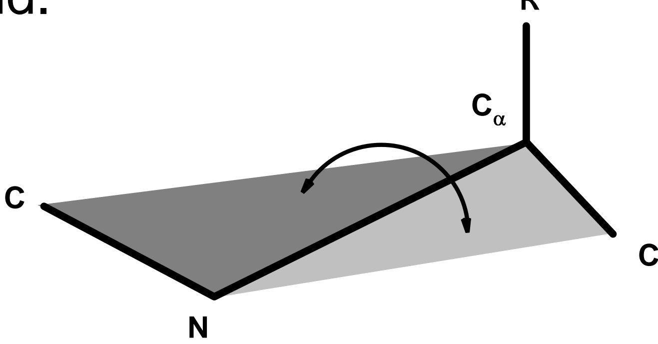
The "shape" of each amino acid depends on the conformation of the bonds within a residue:

- **ω (omega)** - peptide bond
- **Φ (Phi)**, the bond between N and $C_\alpha$
- **Ψ (Psi),** the bond between $C_\alpha$ and C.




The angle of rotation around a bond is specified as the **dihedral angle** between two planes that share the bond.

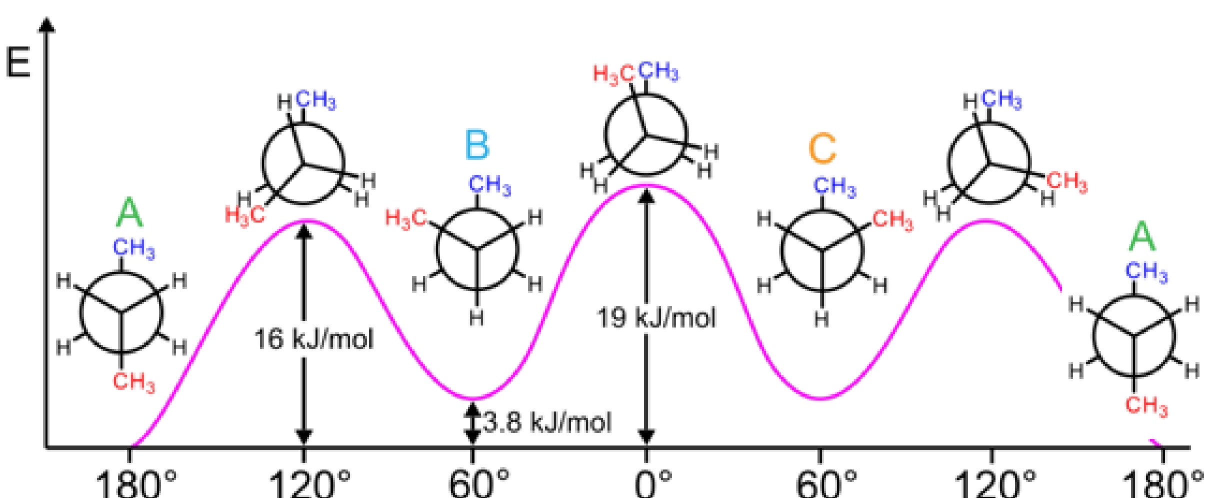



The two planes that are used to specify the phi angle are illustrated above.

In the case of a simple molecule, such as a 1-chloro-1-fluoroethane three different conformations are generally considered stable for free rotation about a single bond.

120$^0$ 120$^0$

Bond rotation about the C-C bond in butane indicates relatively small energies between each stable state, so interconversion between rotational conformations will easily occur and a molecule will sample all possible low energy states.




**A** ⇌ **B** ⇌ **C** ⇌ **A**

***Reflection***: *How many different conformations could the backbone atoms of one residue in a protein take, assuming free rotation around* ***each of the*** *three bonds (C-N, N-$C_\alpha$, $C_\alpha$-C)?*

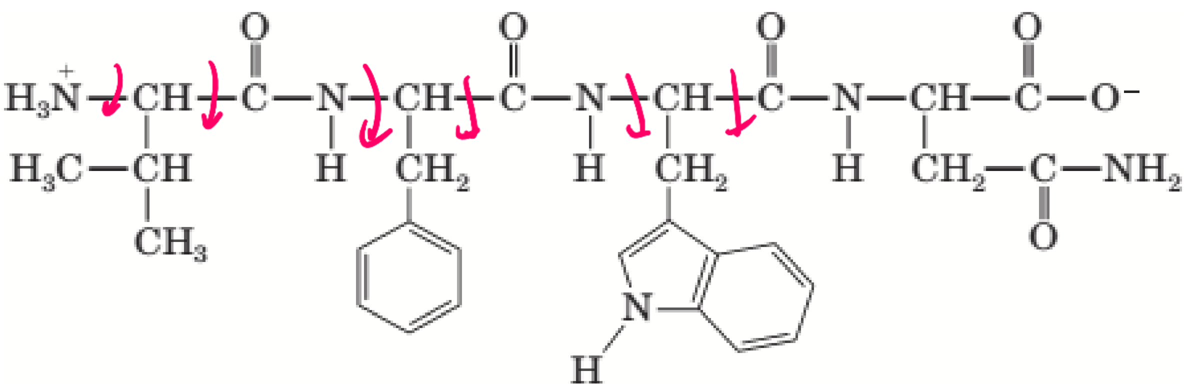
R(sidechain)

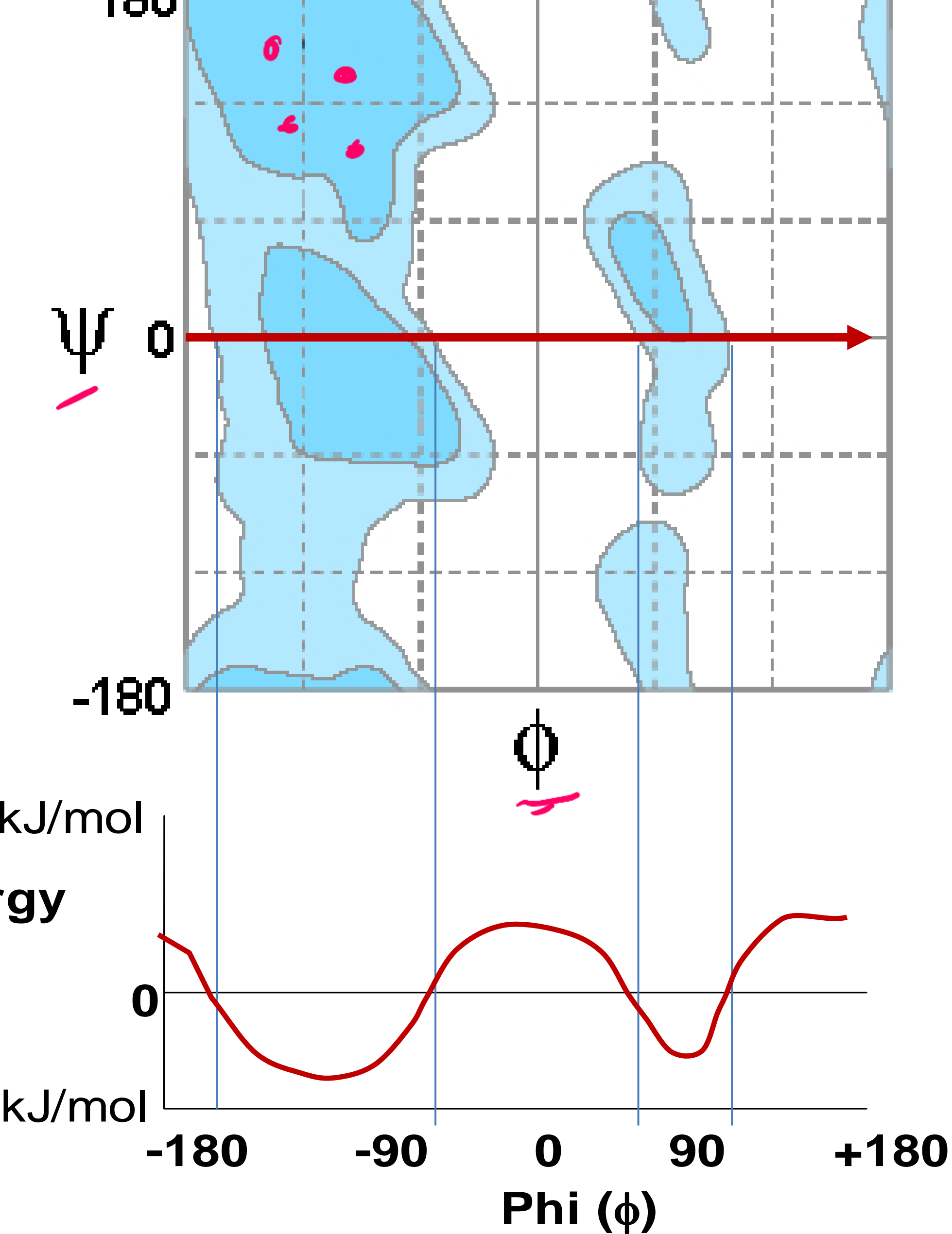
27!

# Ramachandran Plots – Representing the phi and psi angles of each residue

The conformations that are experimentally observed in proteins can be visualized by plotting the Φ (Phi) and Ψ (Psi) angles for each residue in a two-dimensional plot. Such a plot is called a *Ramachandran Plot*:

- The horizontal and vertical axes represent the phi and psi angles of each residue.
- A single point in the plot represents ***one*** residue in the protein.
- Colored regions represent regions of low (favorable) energy due to van der Waals. These colored areas are for residues with a β carbon.
- The white areas correspond to unfavorable van der Waals, the phi and psi angles would push atoms too close together.





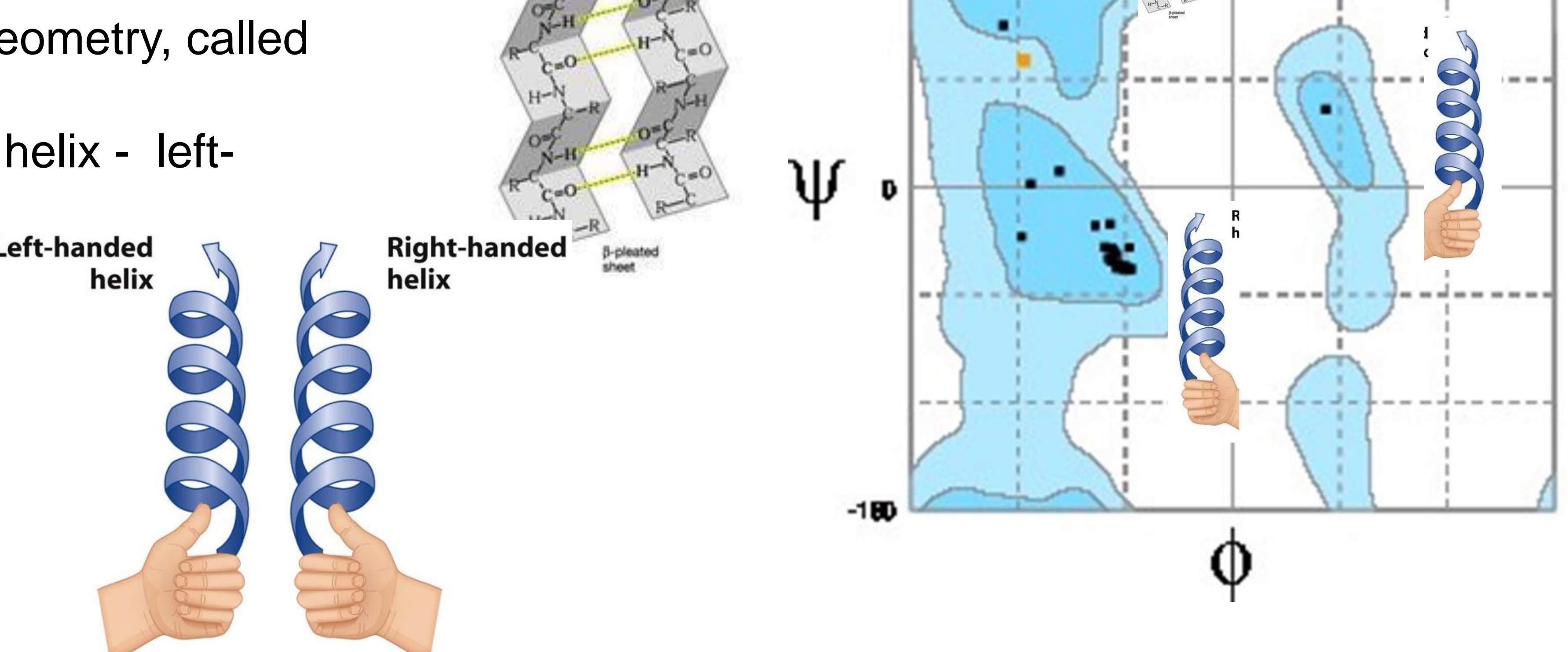

**Plot of Energy versus Phi for Psi =0.**

**Expectations – Ramachandran plot**
- How it is constructed
- Location of different secondary structures
- Origin of colored and white regions (vdw)

# How many Secondary Structures are Found in Nature?

The Ramachandran plot of proteins shows ***only*** three possible low energy "shapes":

- Linear strands, called β-strands, which usually form a multi-stranded sheet.
- A right-handed helical geometry, called an α-helix.
- The mirror image of the helix - left-handed, relatively rare.



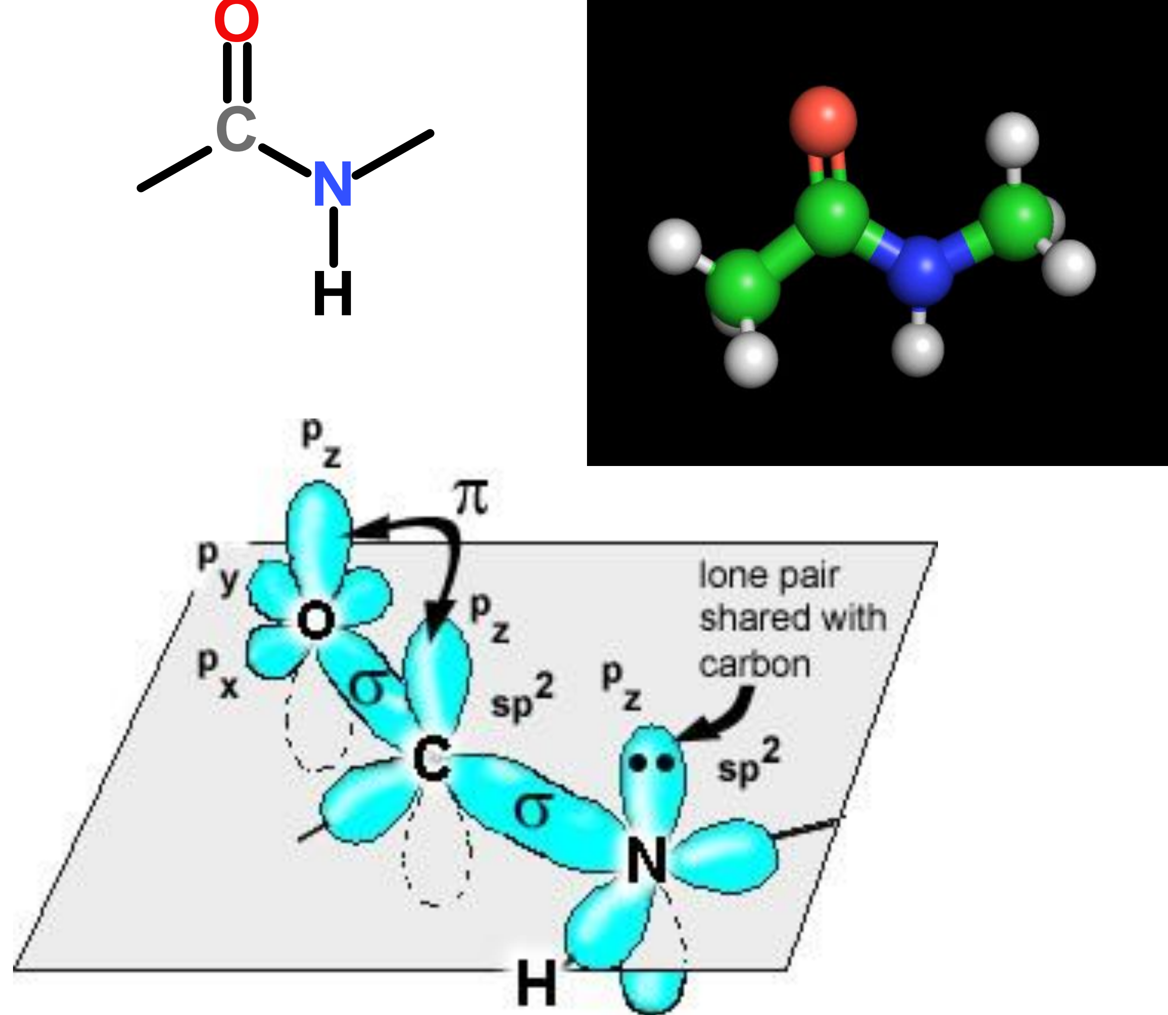

*Why are there only three possible "shapes"? Free rotation about all three bonds should give us 27 possible shapes!*

*What limits the numer of conformations to three?*

# Conformational Freedom of Proteins – One bond at a time

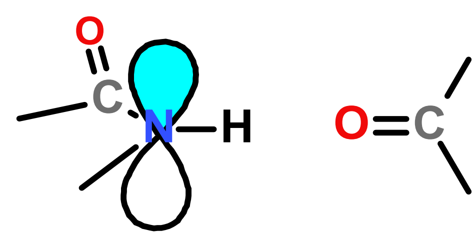
**$C_{i-1}$ - N (Peptide bond):**

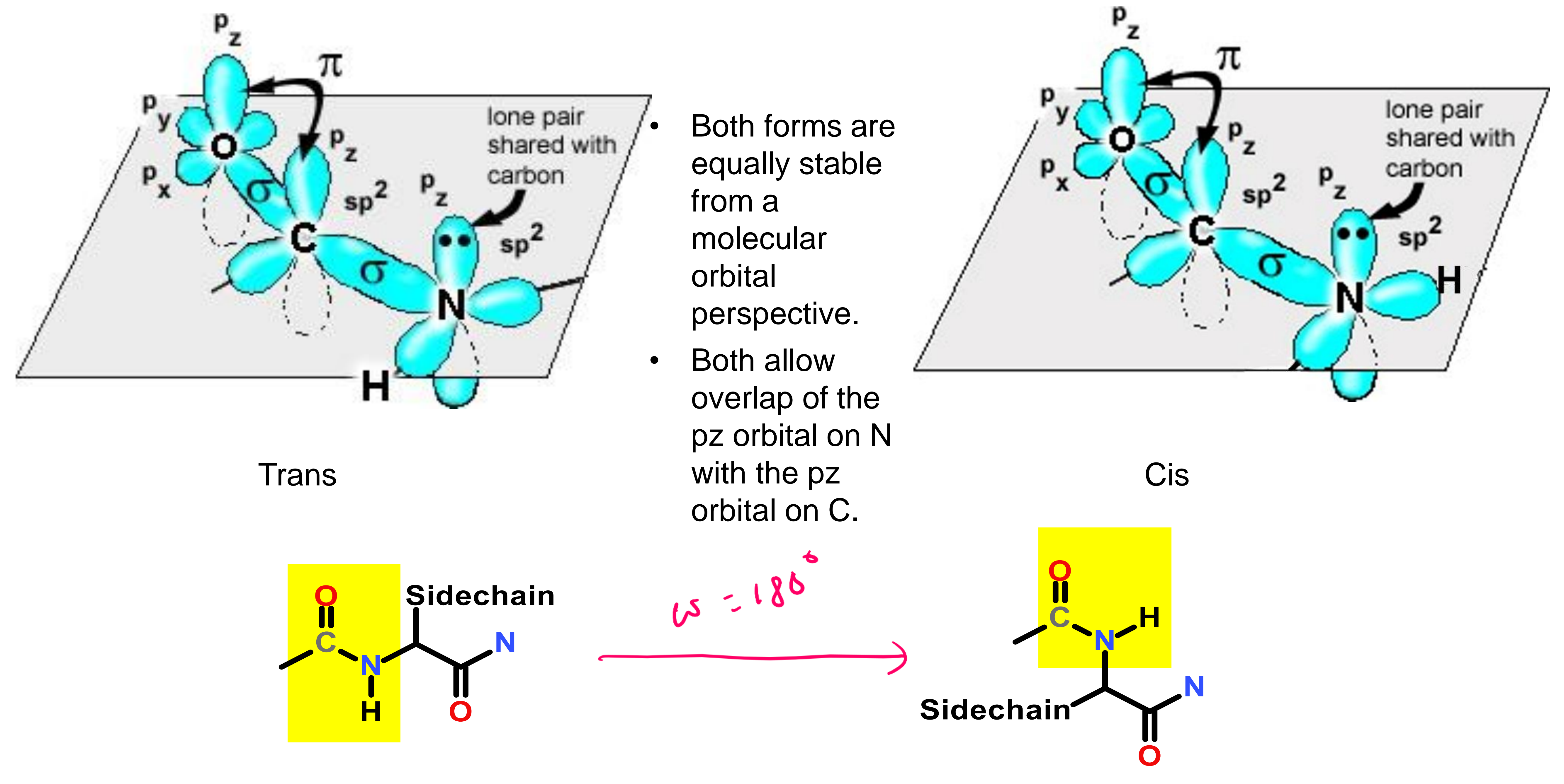
- The four atoms that make up this bond are planar due to the hybridization properties of the carbonyl carbon and the nitrogen (both $sp_2$).
- Free rotation about the bond is **not** possible since the $p_z$ orbitals of oxygen, carbon, and nitrogen form a delocalized system. Rotation about the peptide bond would break the interaction between the $p_z$ orbital of the nitrogen and carbon atoms and is therefore unfavorable. The peptide bond is said to be a "partial double bond".
- **The N-H group** within the peptide bond can ***only*** act as a hydrogen bond donor. The partial negative charge on the nitrogen is delocalized over the entire conjugated system so it is not energetically favorable for it to accept a hydrogen bond (a similar argument applies to the NH group on Trp sidechain, and the amide group found on Asn and Gln sidechains).




*Expectations – Peptide bond*

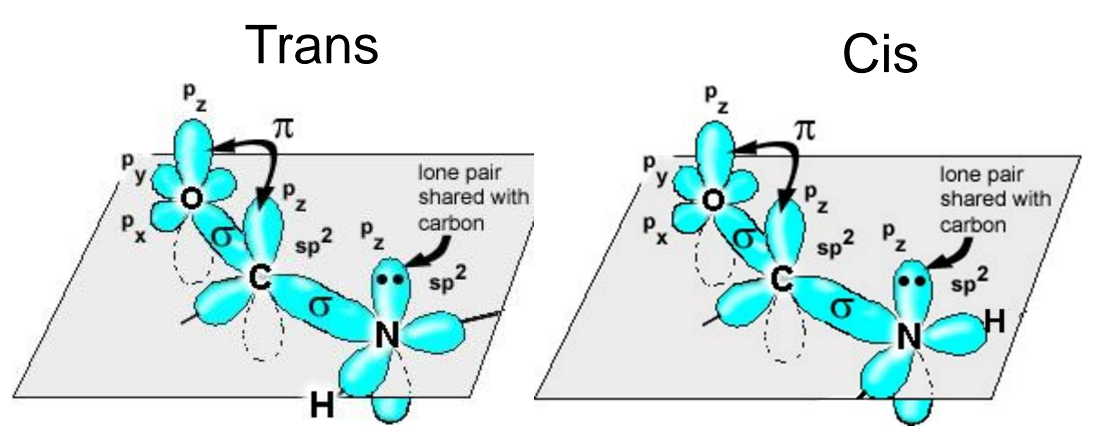
- Why the peptide bond is planer
- Why trans is favorable over cis.




p z
π
p y
O
p x
σ
p z
C
sp 2
p z
lone pair
shared with
carbon
sp 2
σ
N
H
Trans
Both forms are equally stable from a molecular orbital perspective.
Both allow overlap of the pz orbital on N with the pz orbital on C.
Cis
O
C
N
H
Sidechain
N
O
ω = 180°

**Cis and trans Peptide Bonds**:

- Two possible orientations of the peptide bond allow overlap of the $p_z$ orbital between the C and N.
- They are related by a 180° flip of the peptide bond, giving two possible conformations: ***trans or cis***.
- The trans form is seen to be 1000x more stable than the cis form.
- For peptide bonds involving proline, the trans form is 5 x more stable




*Why is the trans form lower in energy?*

*Why does this change for bond involving proline?*

E

17 kJ/mol

4 kJ/mol

**Ala-Ala**

TRANS

ω 180° rotation

CIS

**Ala-Pro**

TRANS

ω 180° rotation

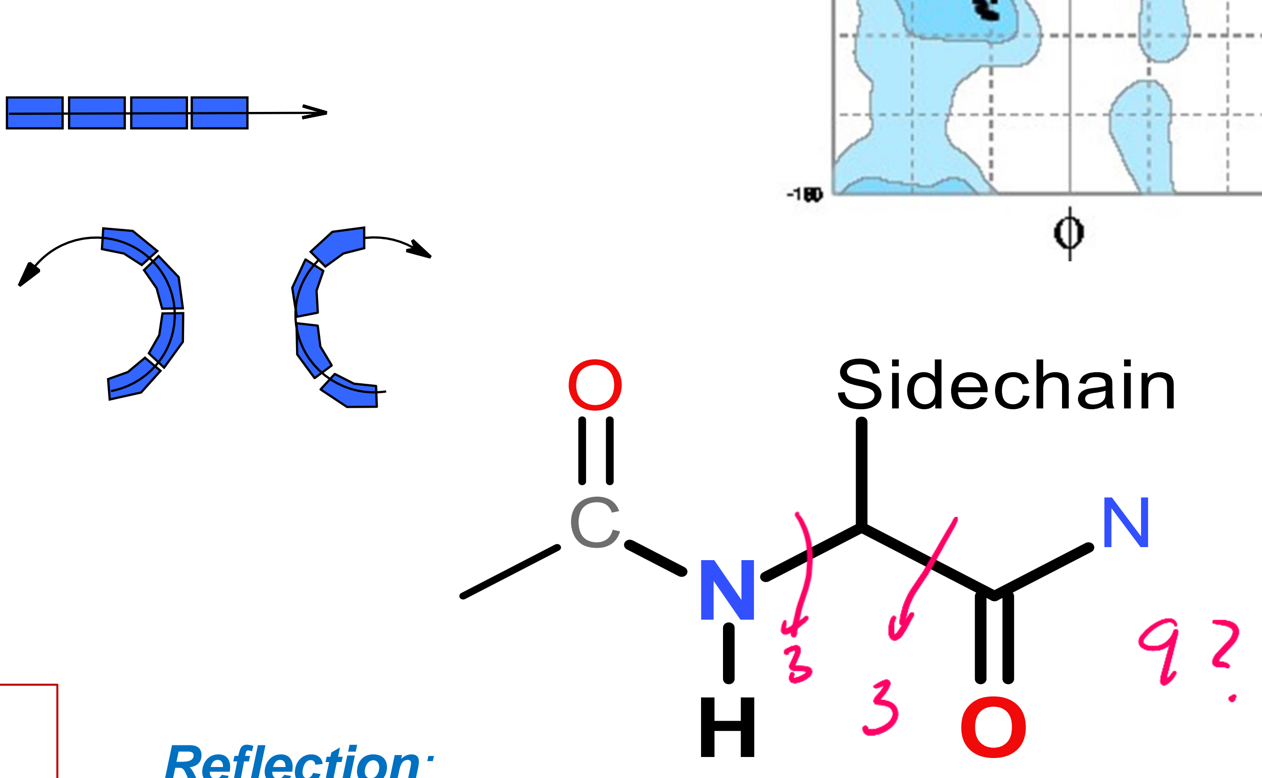
CIS

**N - $C_\alpha$ & $C_\alpha$ - C Bonds:**

- There is free rotation about *both* of these bonds.
- Since both the phi and psi bonds are free to rotate, a total of 9 (3x3) different stable pairs of phi and psi angles should be observed.
- However, there are only three pairs, each of which correspond to a secondary structure:

**$\Phi$ = -120°, and $\Psi$ = 125°. extended, β-structure**
**$\Phi$ = -60°, and $\Psi$ = -45° α-helix, right handed**
**$\Phi$ = +60°, and $\Psi$ = +45°. α-helix, left handed**

- $\Psi$ and $\Phi$ torsional angles are the *same* for *each* residue within the element of secondary structure.
- In these structures, each peptide bond is rigid and planar and in the *trans* conformation.




Expectations – Secondary Structure Overview

- You should know there are three secondary structures.
- The right-handed helix and the beta are more common
- That each residue in the secondary structure has the same phi and psi angles (regular sec. structure).

***Reflection:***

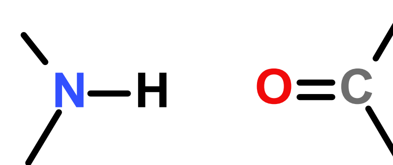
*Why are there only three – what limits forming the other six possibilities?*

*What forces are stabilizing, and which are destabilizing?*

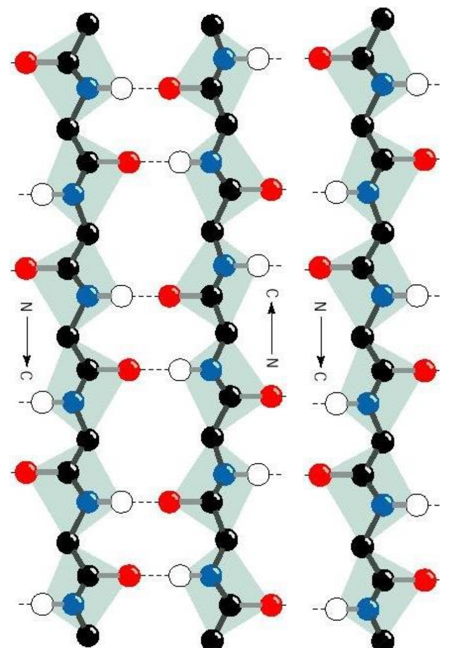
***Reflection***: *Why are there only three structures– what limits forming the other six? What forces are stabilizing, and which are destabilizing?*

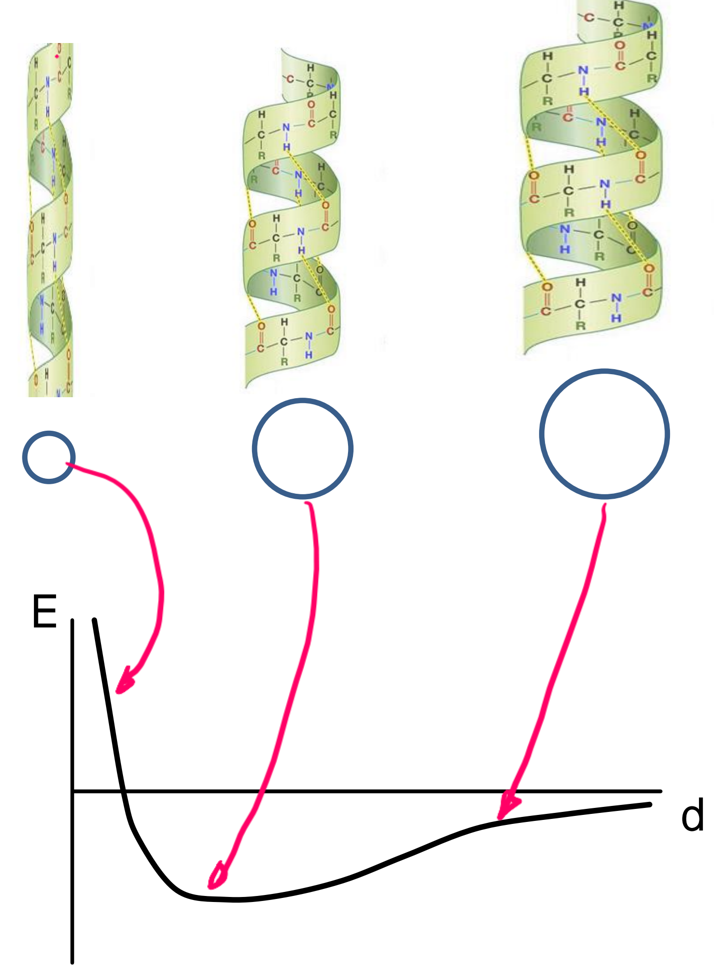
Many helices allow the formation of H-bonds, but only one is lowest in energy, Why?

**H-bonds.** Not all phi & psi angles will allow ***mainchain*** H-bond donors and acceptors to align properly, this excludes many of the possible phi and psi angles.



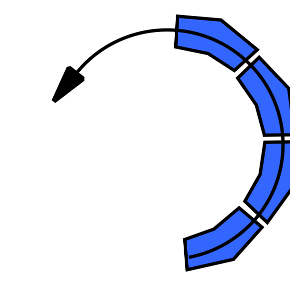

B-strands allows H-bonding






**van der Waals** makes one helix more favorable than others – ideal packing of ***mainchain*** atoms.

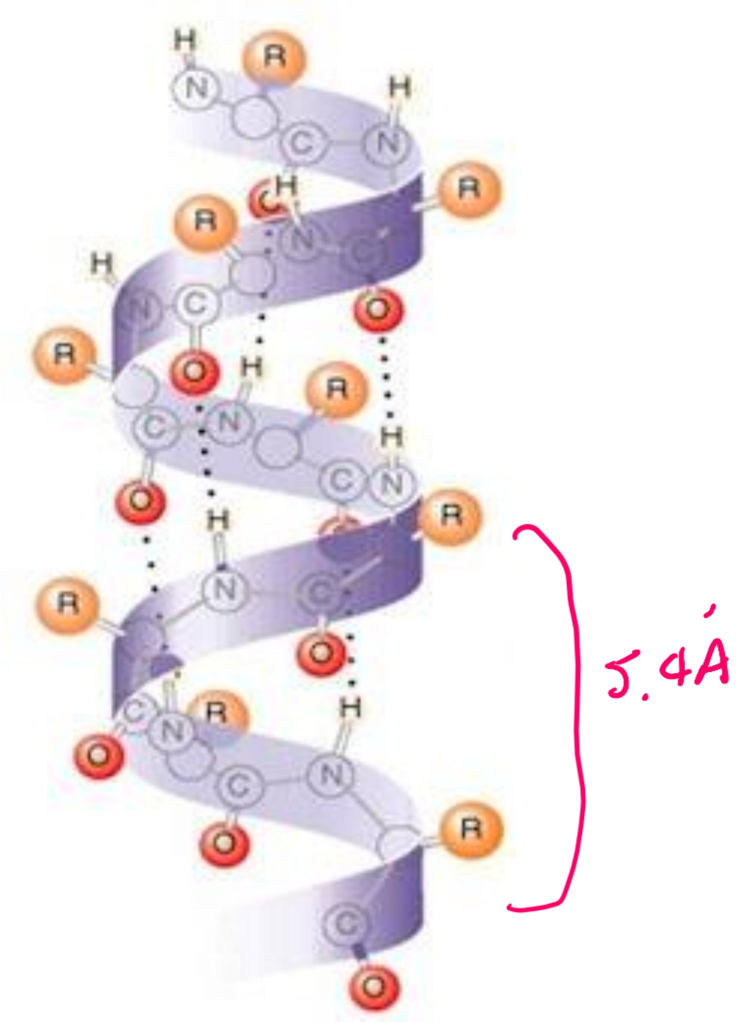
# $\alpha$-Helix Structures ($\Phi$ = -60°, $\Psi$ = -45°)

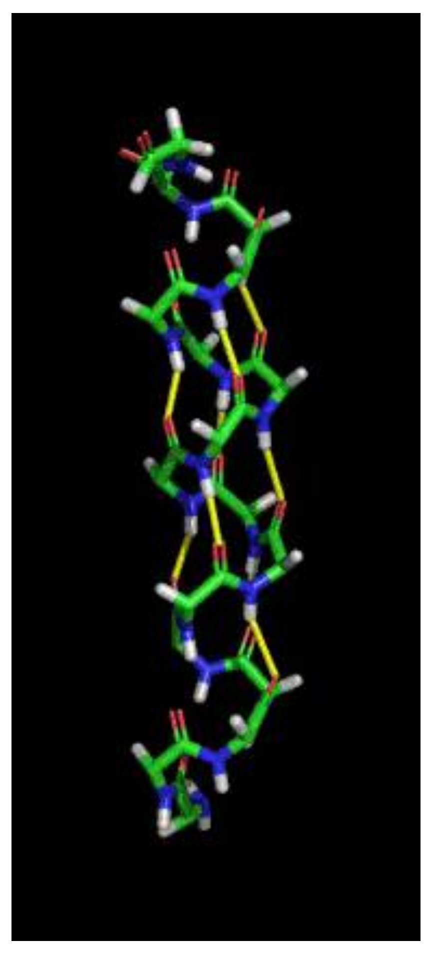


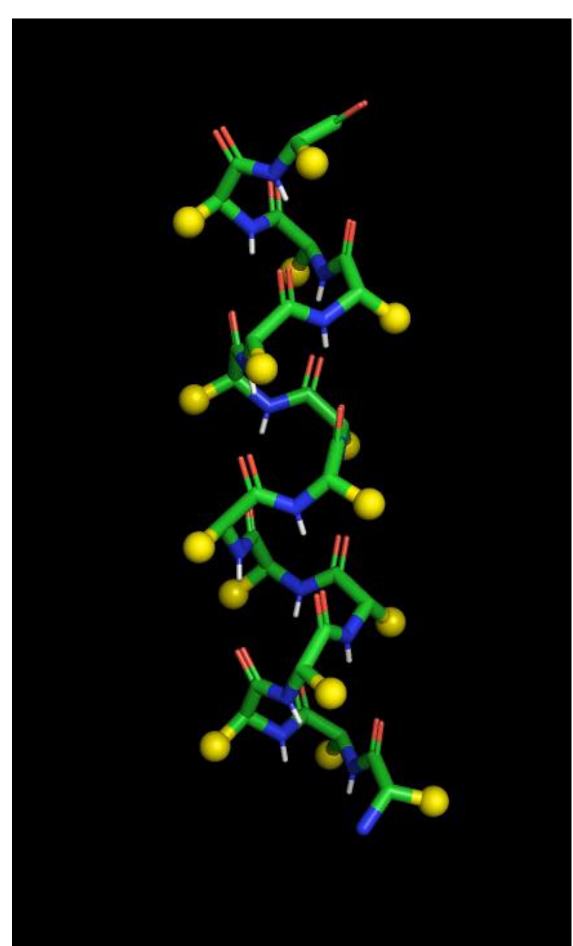
3.6 residues/turn

pitch = 5.4 Å/turn

- H-bonds || to helix axis.
- There is a repeating hydrogen bonding pattern within the helix.
- Sidechains point outwards, and slightly down towards the N-terminal of the helix.
- Right-handed



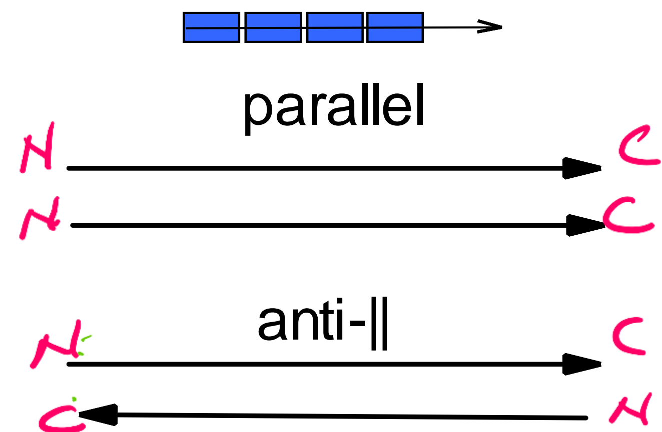


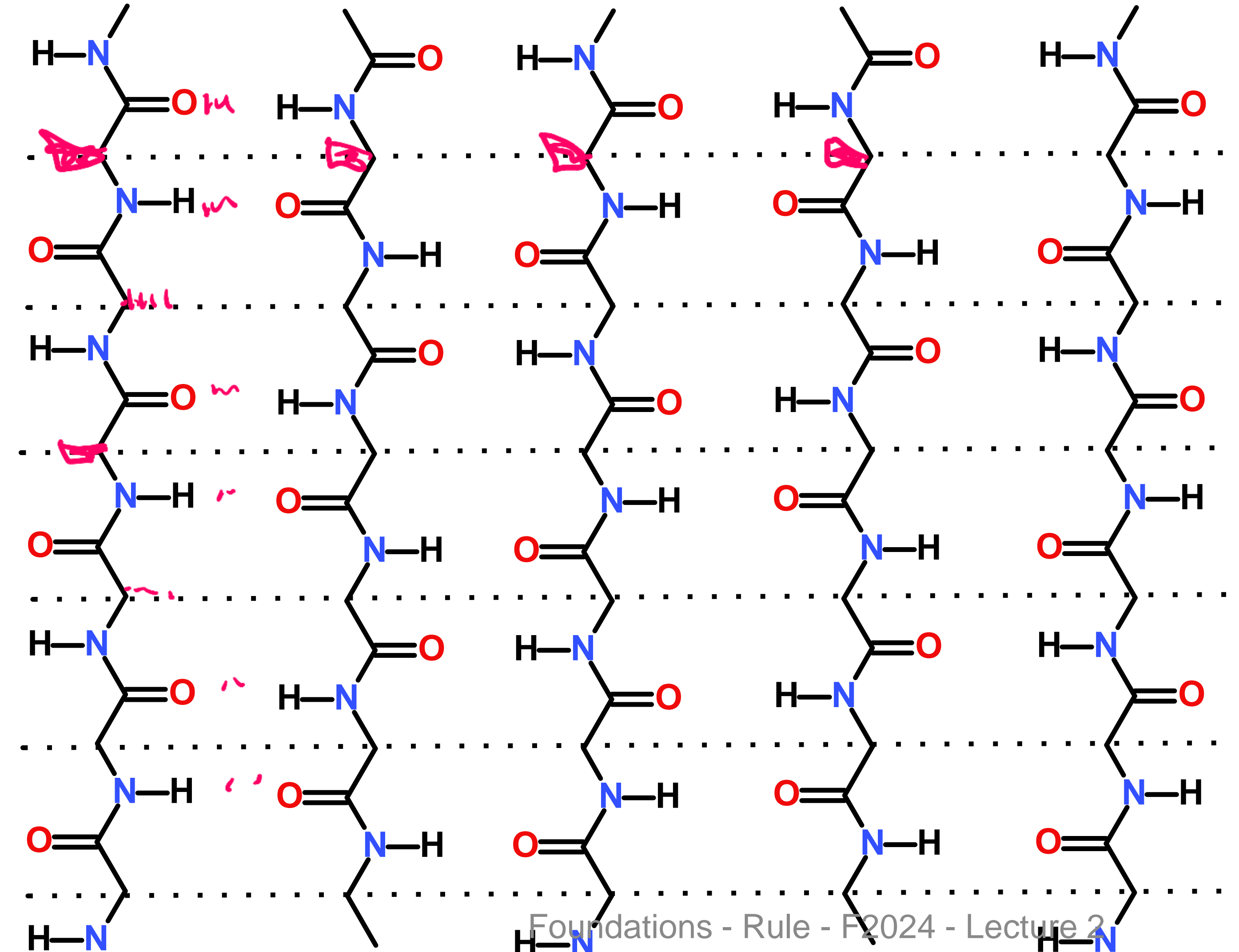
# Beta Structures (Φ = -120°, Ψ = 125°)

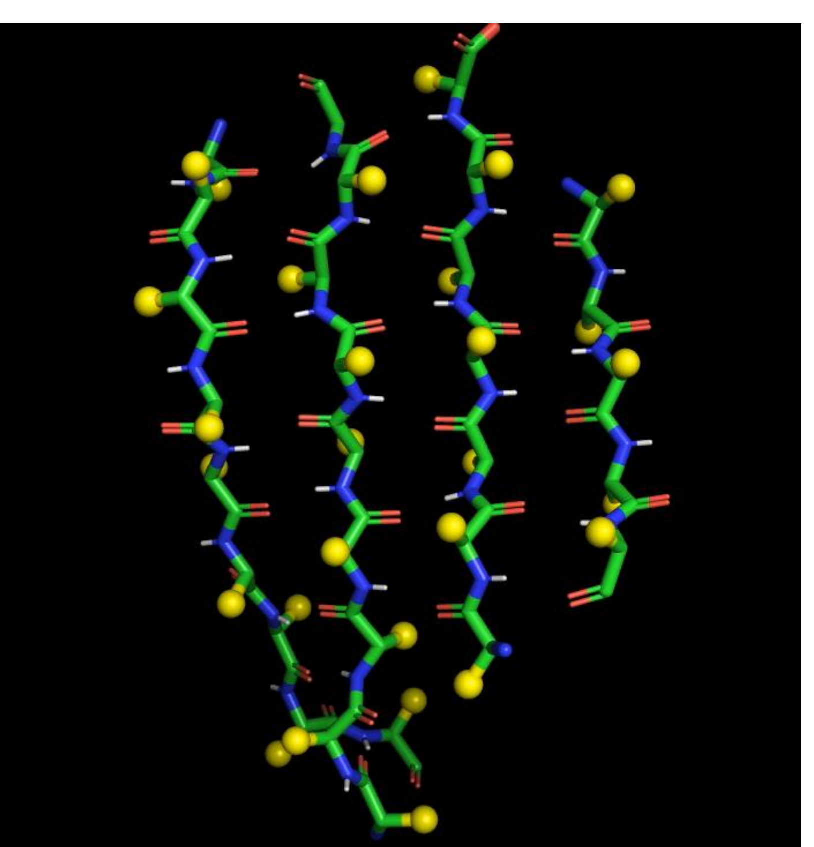
- H-bonds perpendicular to strands.
- Strands can be parallel or anti-parallel. This refers to direction of strand, from N to C, the strands always align adjacent to each other, forming H-bonds.
- Sidechains:

✓ alternate up and down on any ***one strand,*** i.e. above and below the sheet

✓ sidechains on ***adjacent strands are pointing in the same direction.***

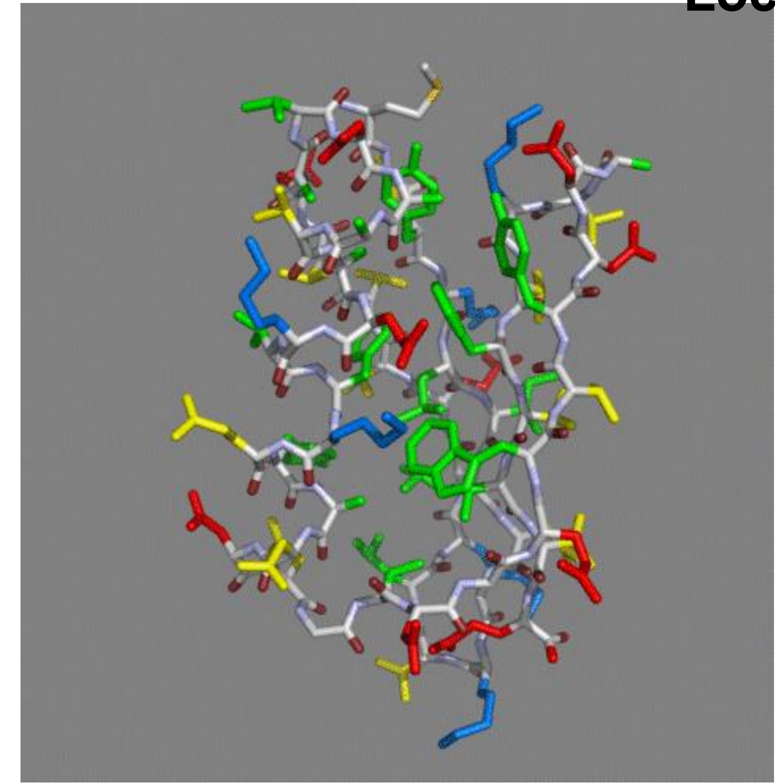




# Tertiary Structure = Complete Structure of One Chain

## Location of Residues in Globular Proteins

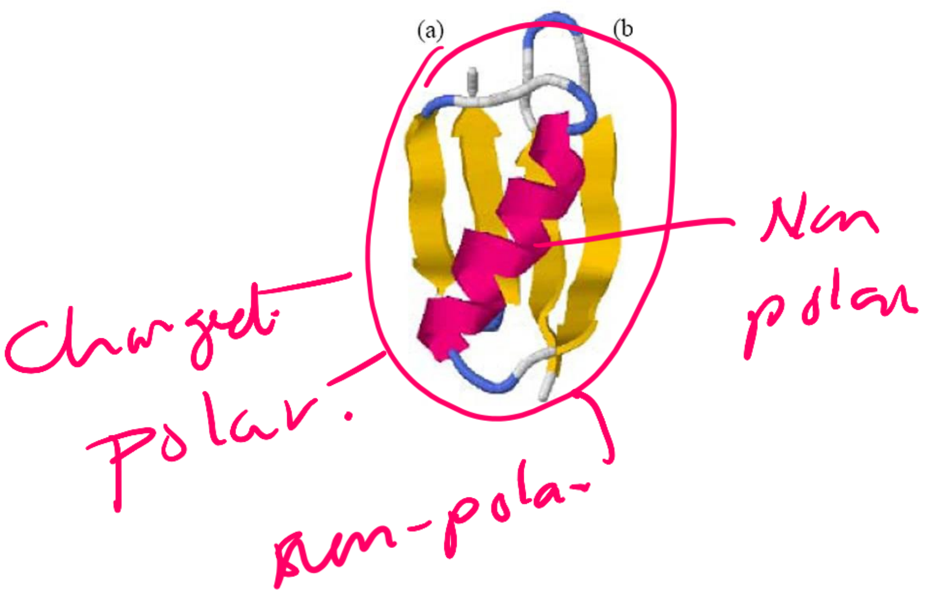


Red - amino acids with neg. sidechains

Blue - amino acids with pos. sidechains

Yellow – amino acids with polar sidechains.

Green - amino acids with hydrophobic side chains

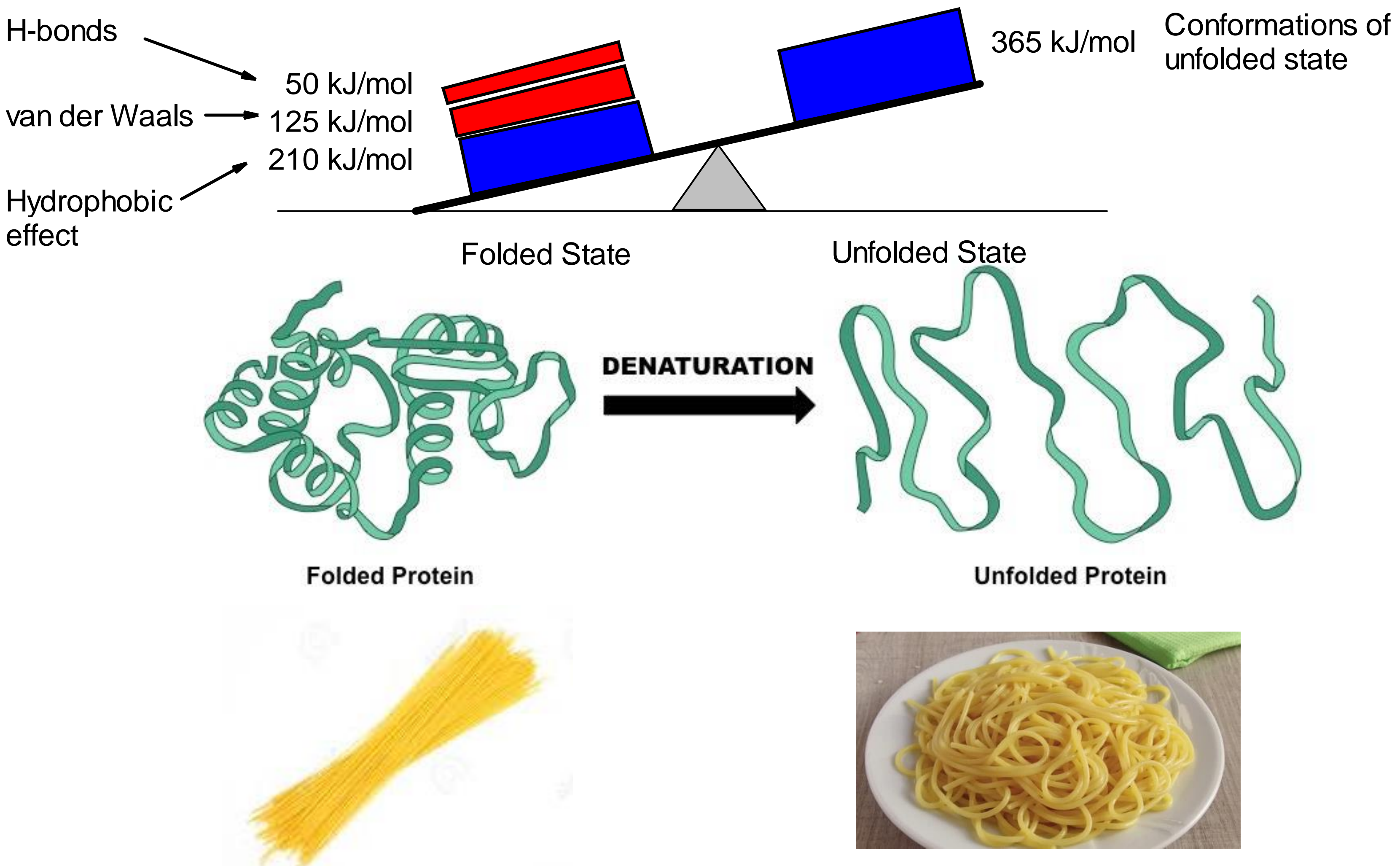



| Amino Acid | | |
|---|---|---|
| | Inside | Surface |
| Charged | | |
| Polar | | |
| Non-polar | | |

# Tertiary Structure – Stability

**Goals:**

- Relate molecular interactions to stability of proteins.
- Categorize enthalpic and entropic effects.
- Understand marginal stability of folded proteins.




Position of equilibrium depends on the energy difference between the two states:

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

Higher energy states are unfavorable.

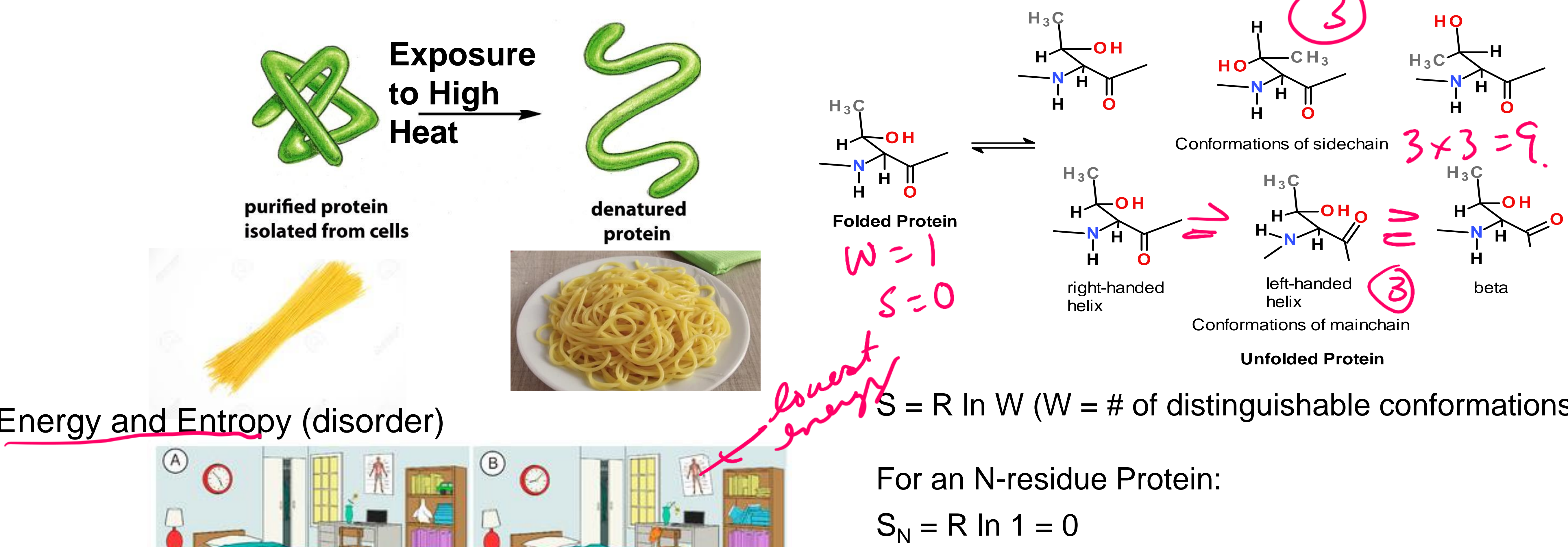
$\Delta S^o$ : Change in disorder of the system. ***positive*** entropy change is ***favorable*** since the ***disorder*** in the system is ***increased***.

$\Delta S^o$ is positive for unfolding, favoring the unfolded state.

$\Delta H^o$ : Enthalpy – A change in the electronic configuration of the system that either releases *heat* ($\Delta H^o < 0$) or absorbs *heat* ($\Delta H^o > 0$). **Release of heat is favorable**.

$\Delta H^o$ is positive for unfolding, heat needs to be added to unfold proteins

# Unfolded Polypeptides Are Flexible – High Entropy (Disorder) Stabilizes the Unfolded State




$S = R \ln W$ (W = # of distinguishable conformations)

For an N-residue Protein:

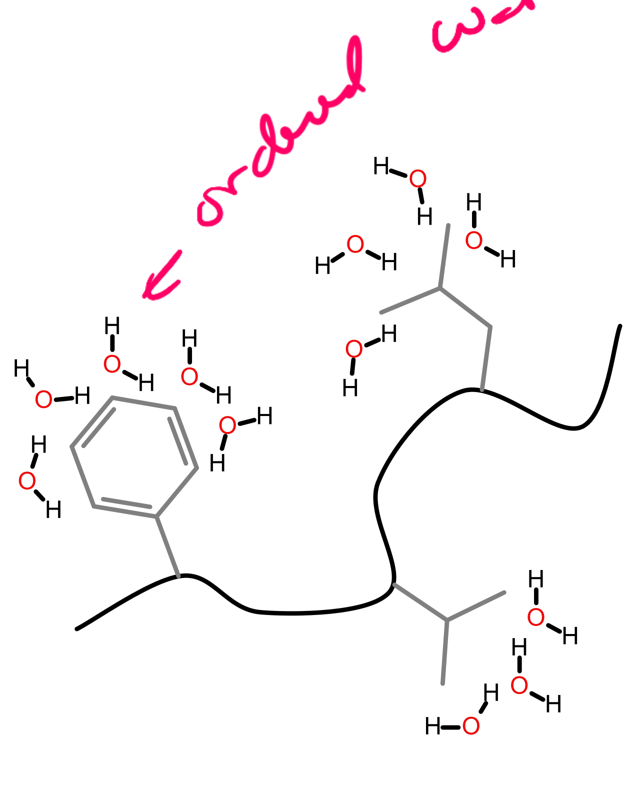
$S_N = R \ln 1 = 0$

$S_U = R \ln (3x3)^N$ = Large and positive

65 residues: $-T\Delta S = 300 \times 8.3 \times \ln (9)^{65}$

$= 355$ kJ/mol

**ΔS° - Hydrophobic effect - Entropy Changes of the Solvent:** The hydrophobic effect is due to the entropy of the *water* in the system. When a non-polar side chain is exposed to water it orders, or decreases the entropy, of the water molecules. However, when the non-polar residue becomes buried in the non-polar center of the protein it releases all of the water which coated it. The released water can now freely diffuse in the solvent, resulting in an increase in entropy of the water, thus non-polar groups are "forced" into the non-polar core of the protein.



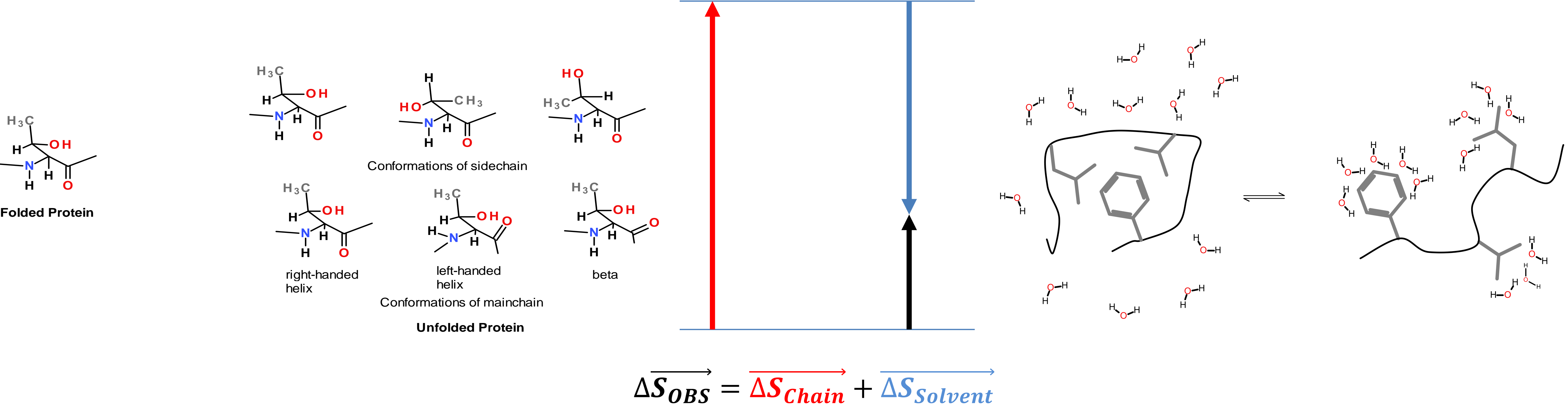
**Dissolved butane, ordering hydrogen bonded waters**

The larger the sidechain, the larger the hydrophobic effect.

## Overall entropy change:

$$\Delta S^o_{OBS} = \Delta S^o_{Chain} + \Delta S^o_{Solvent}$$

Note that these are opposite in sign, for N→U, $\Delta S^o_{Chain}$ is large and positive (favorable), while $\Delta S^o_{Solvent}$ is large and negative (unfavorable for unfolding), overall the entropy of unfolding is positive, favoring the unfolded form of proteins. The entropy changes can be represented as a vector diagram:




$$\overrightarrow{\Delta S_{OBS}} = \overrightarrow{\Delta S_{Chain}} + \overrightarrow{\Delta S_{Solvent}}$$

**Enthalpy ($\Delta H^o$)** is the amount of heat generated/consumed by the reaction when 1 mole of reactants are converted to one mole of products.

## $\Delta H^o$ - Hydrogen bonds:

- $\Delta H^o$ associated with hydrogen bonding is *favorable* for folding. Hydrogen bonds are more stable in the native form of the protein by about 1-5 kJ/mol.
- Hydrogen bonds that are broken during folding and then **not** reformed in the folded state cost about 20 kJ/mol.

No acceptor

$\Delta H^o$ = +20 kJ/mol

1

$\Delta H^o$ = +18 kJ/mol

2

$\Delta H^o$ = +2 kJ/mol

3

Folded (Native)

Unfolded (Denatured)

## Hypothetical reaction:

1 – Breaking of hydrogen bond in a protein, without reforming H-bond, cost 20 kJ/mol.

2 – Reforming hydrogen bond with water. Energy released is 18 kJ/mol. Weaker due to less favorable geometry.

3- Overall change in enthalpy is the sum of 1 + 2

H-bond in folded protein is ~2 kJ/mol ***more*** stable than H-bond to water.

## Tertiary structure stabilized by:

- Mainchain-mainchain H-bonds in secondary structure
- Mainchain-sidechain H-bonds
- Sidechain-sidechain H-bonds

Gly

Ser

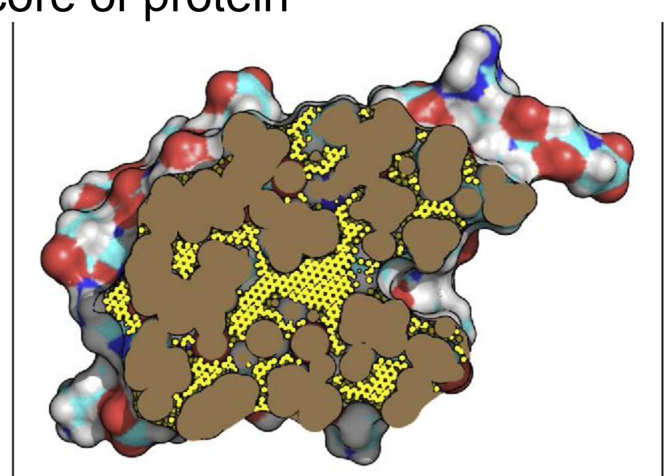
Asn

Ser

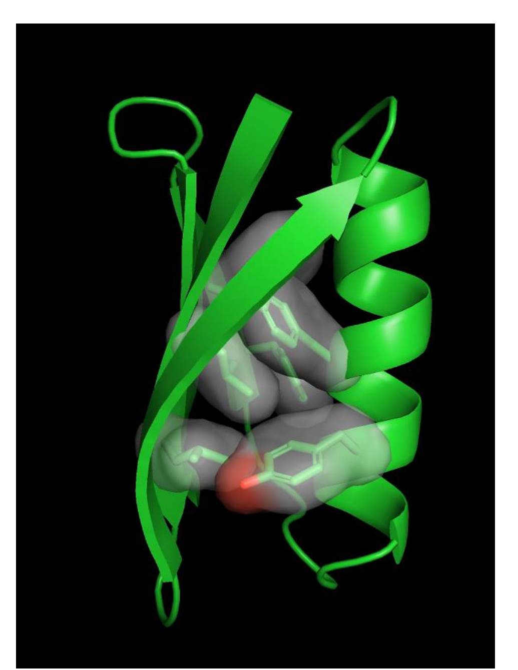
**ΔH° - Van der Waals (London) Forces.** This is **un**favorable for **un**folding. Van der Waals interactions are more stable in the native form of the protein.

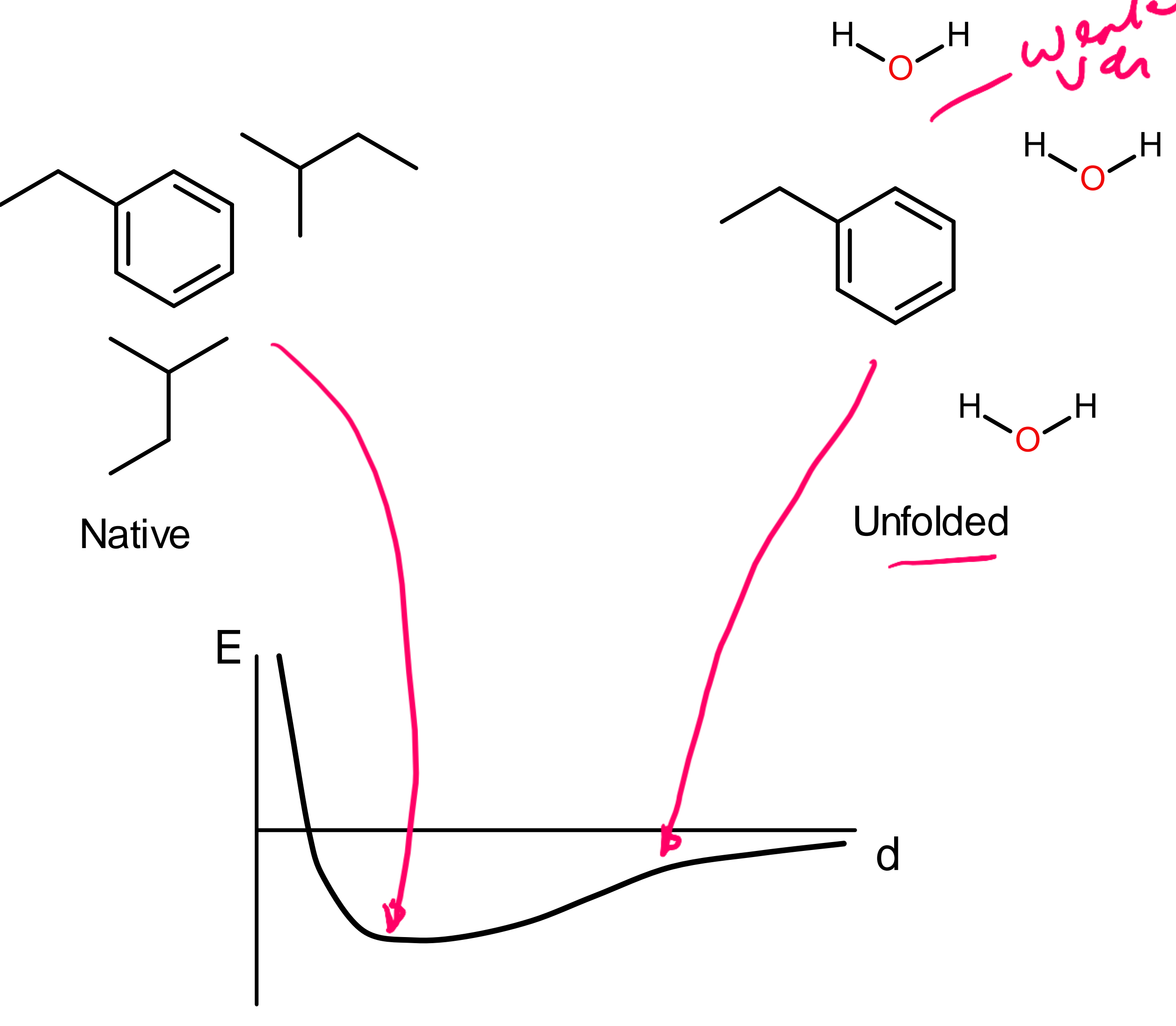
*What enhances the higher stability of van der Waals interactions in **folded proteins**?*

Optimal packing of sidechains in core of protein



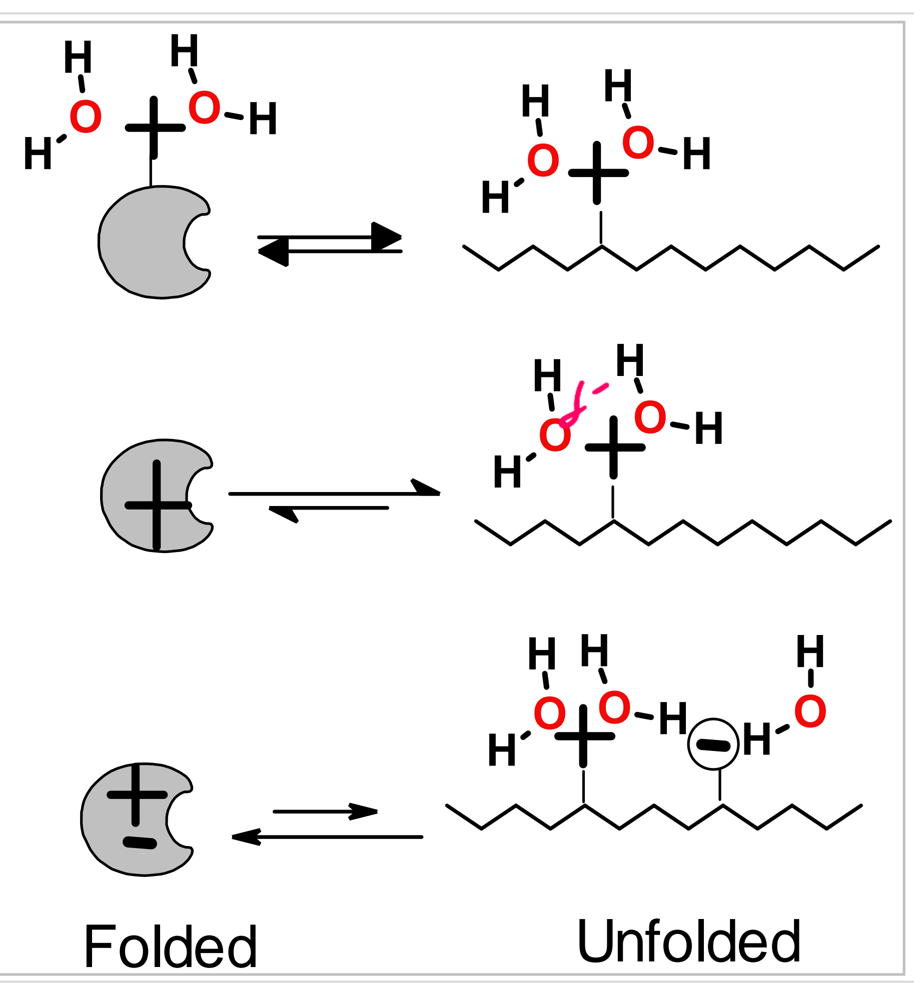
Cross section of a folded protein – knobs fit into holes



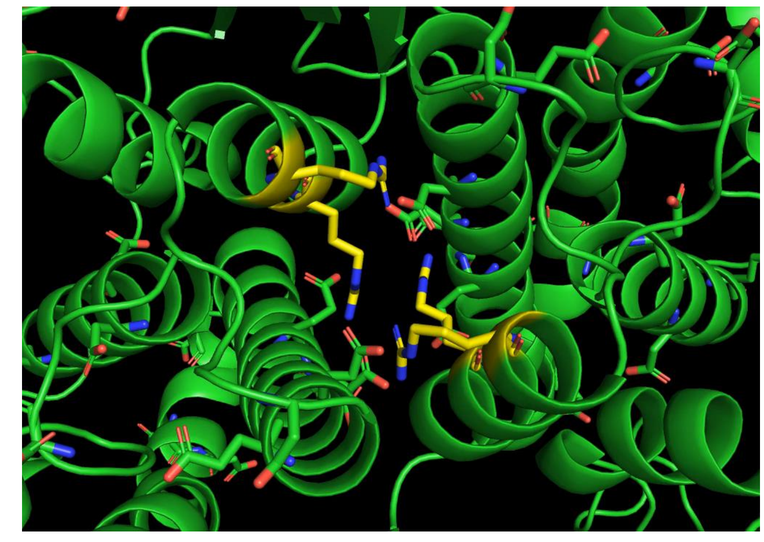


## ΔH° - Electrostatic forces:

- **Surface charges:** Although these forces can contribute to ΔH° changes in many biochemical interactions, they are generally not important for protein folding because the charged residues remain on the surface and therefore interact with water equally well in both the native and the denatured state.
- **Buried charges:** The energetic cost of burying a single charge in the core of a protein is extremely high, largely due to desolvation of the ion during the folding process, therefore single buried charges in proteins are very rare.
- In some proteins *charge-pairs* are buried. These are very stabilizing because the loss of energy due to desolvation is regained by favorable electrostatic interactions in a low dielectric media (recall $E \propto q_1q_2/Dr$).
    - $D_{water} = 80$
    - $D_{interior} \sim 5$



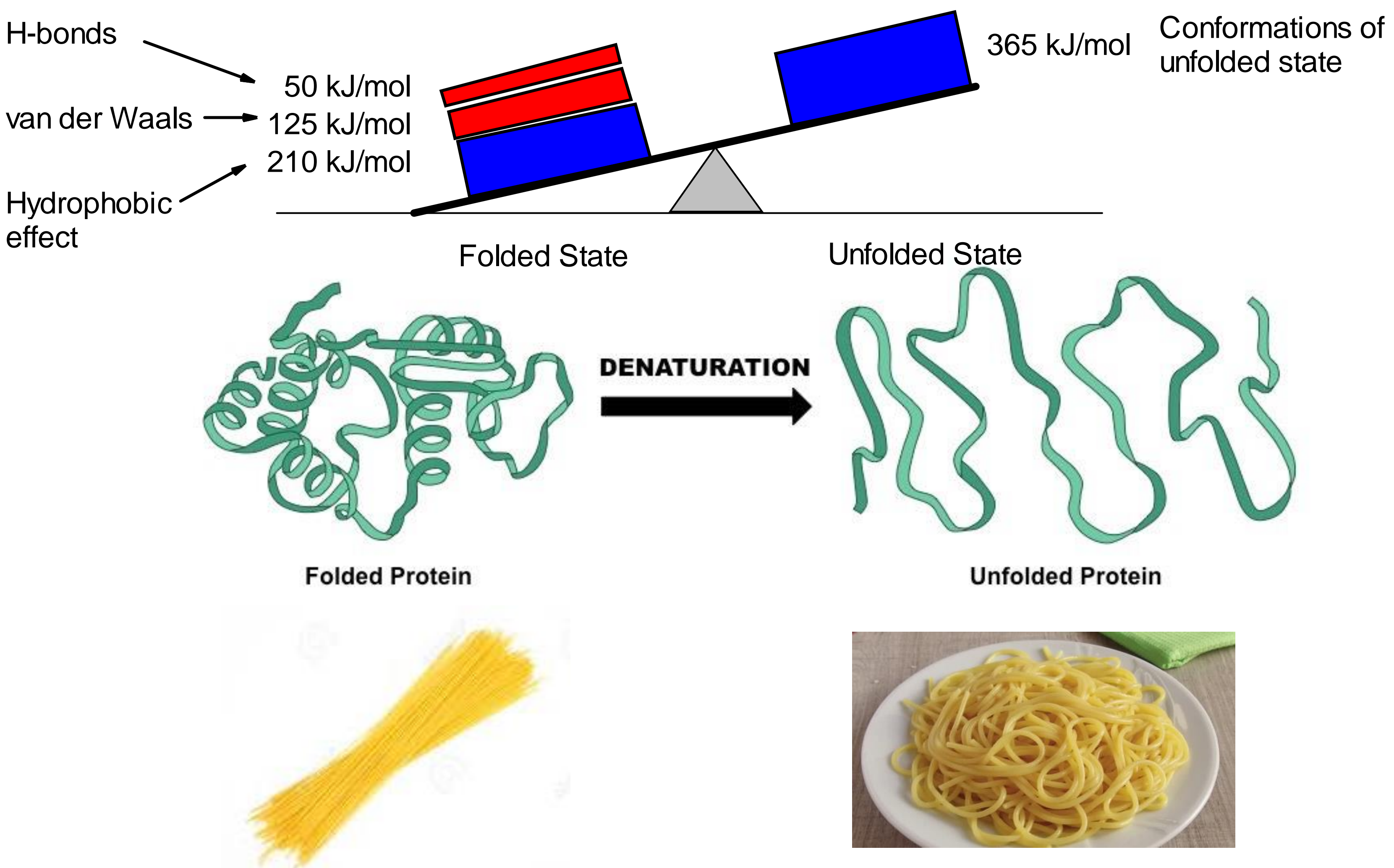

**Buried charges stabilize some protein-protein interfaces.**
In this homodimeric enzyme two Arg residues from each subunit interact via vdw and interact with Asp residues by charge-charge interactions.



# Tertiary Structure – Stability

**Goals:**

- Relate molecular interactions to stability of proteins.
- Categorize enthalpic and entropic effects.
- Understand marginal stability of folded proteins.




Position of equilibrium depends on the energy difference between the two states:

$\Delta G^o = \Delta H^o - T\,\Delta S^o$

Higher energy states are unfavorable.

$\Delta S^o$ : Change in disorder of the system. ***positive*** entropy change is ***favorable*** since the ***disorder*** in the system is ***increased***.

$\Delta S^o$ is positive for unfolding, favoring the unfolded state.

$\Delta H^o$ : Enthalpy – A change in the electronic configuration of the system that either releases *heat* ($\Delta H^o$ <0) or absorbs *heat* ($\Delta H^o$ >0). **Release of heat is favorable**.

$\Delta H^o$ is positive for unfolding, heat needs to be added to unfold proteins