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

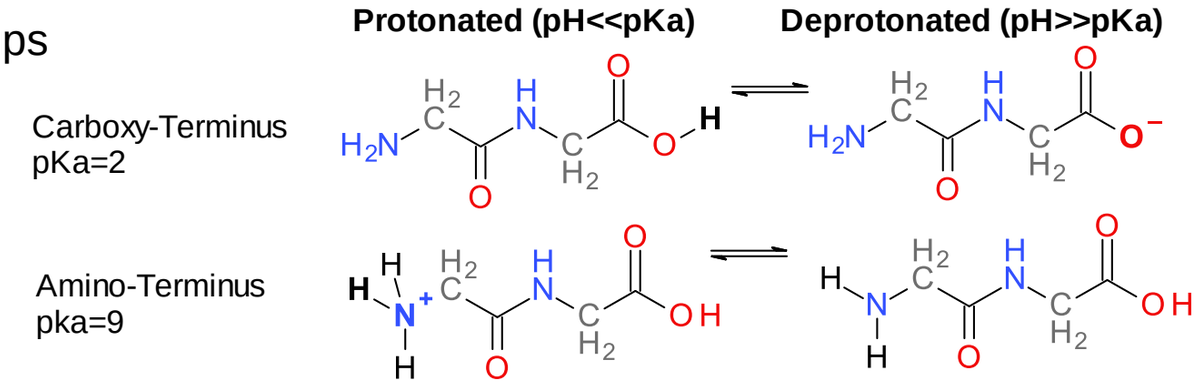
- Affinity chromatography
- Ion exchange
- Size exclusion

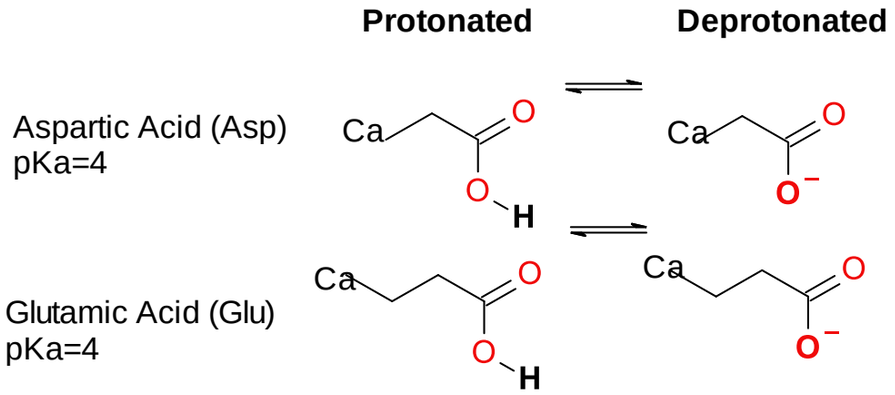
Expression of Recombinant Proteins

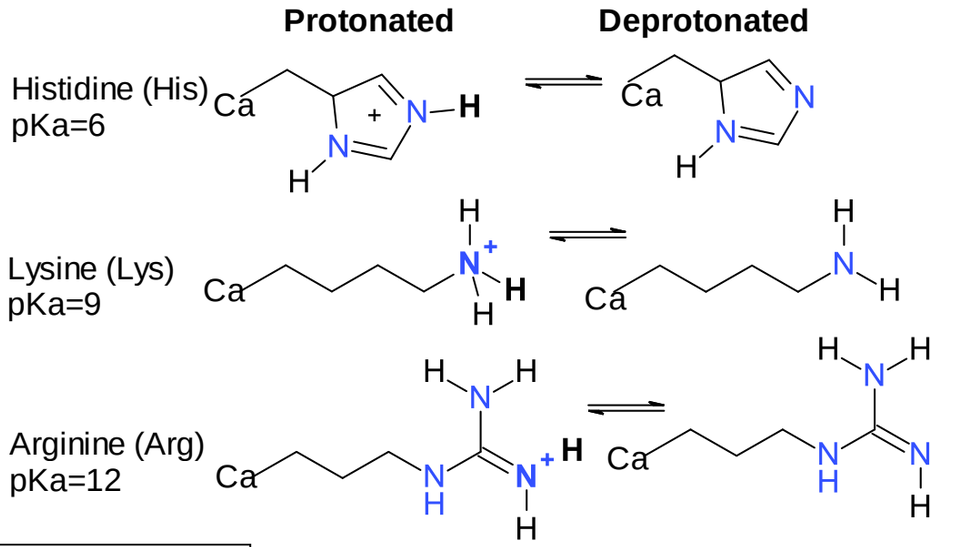
- Transcription and Translation
- E. Coli and T7 RNA Polymerase
- Lac operon as a control system
- pET vectors
- Restriction enzymes & DNA Ligase

Acidic Groups in Amino Acids & Proteins

A. Key Groups

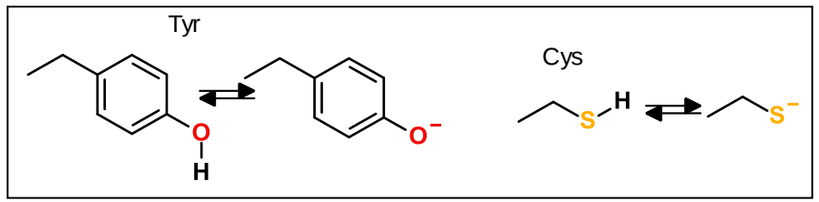






B. Other sidechain ionizations that are usually less important for function:

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



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?

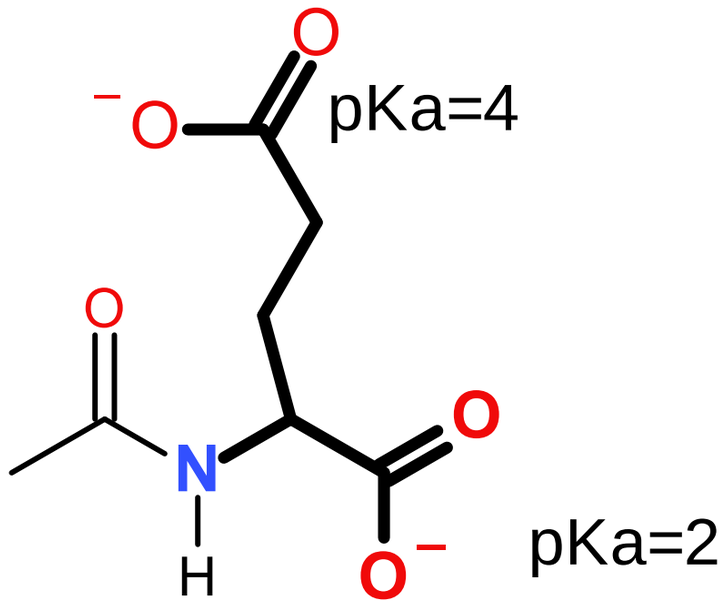
- Chemical Effects
- Environmental Effects

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.

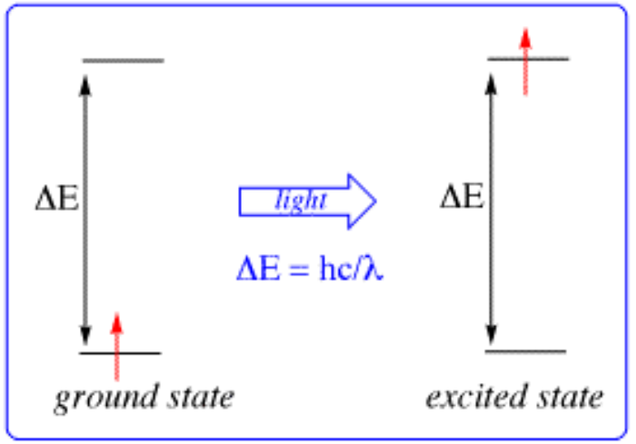
Local charges can shift pKa values.

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

Measuring pKa with NMR - Biochemical Spectroscopy

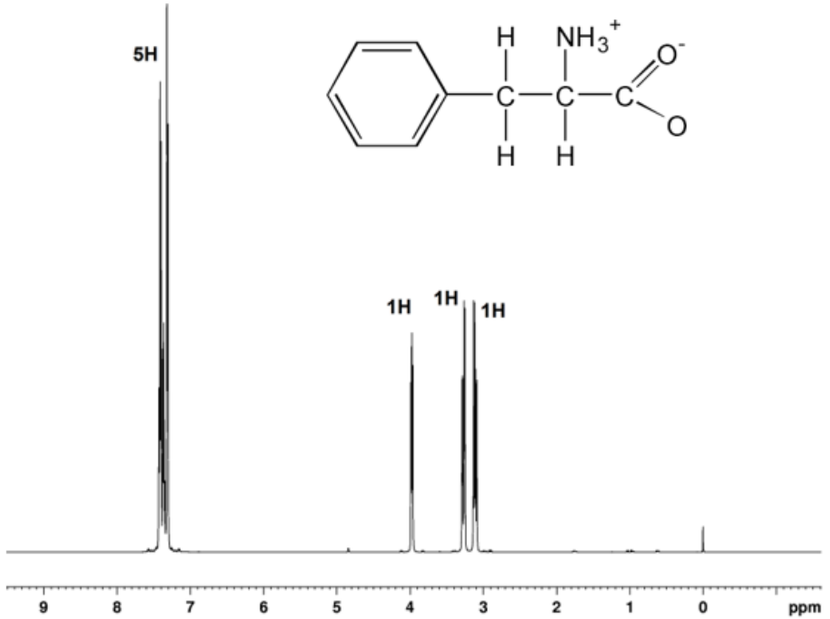
Spectroscopy:

- 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.



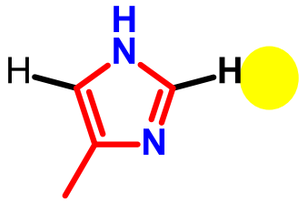
Nuclear Magnetic Resonance (NMR):

- 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).

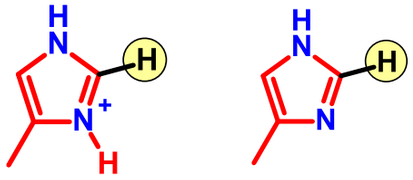


δ_{OBS}

Experimental Measurement of pKa – pH Titrations



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



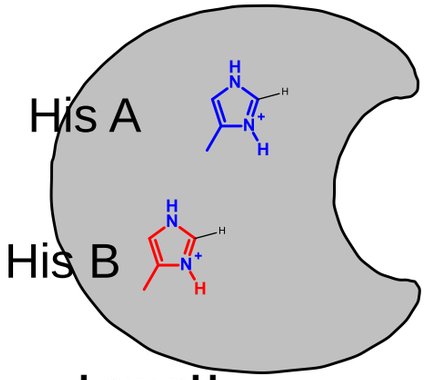
A. Chemical shift depends on the local environment.
B. Chemical shift of highlighted proton changes due to protonation/deprotonation.

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

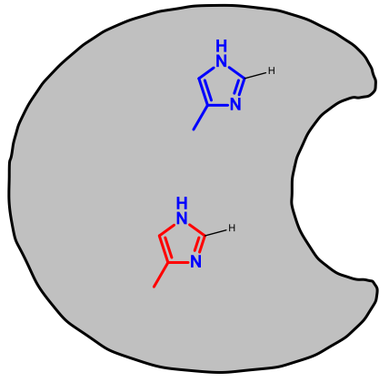
His A – deprotonation increases chemical shift (from 7.4 to 10.0).

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

Example: Protein with two histidines.

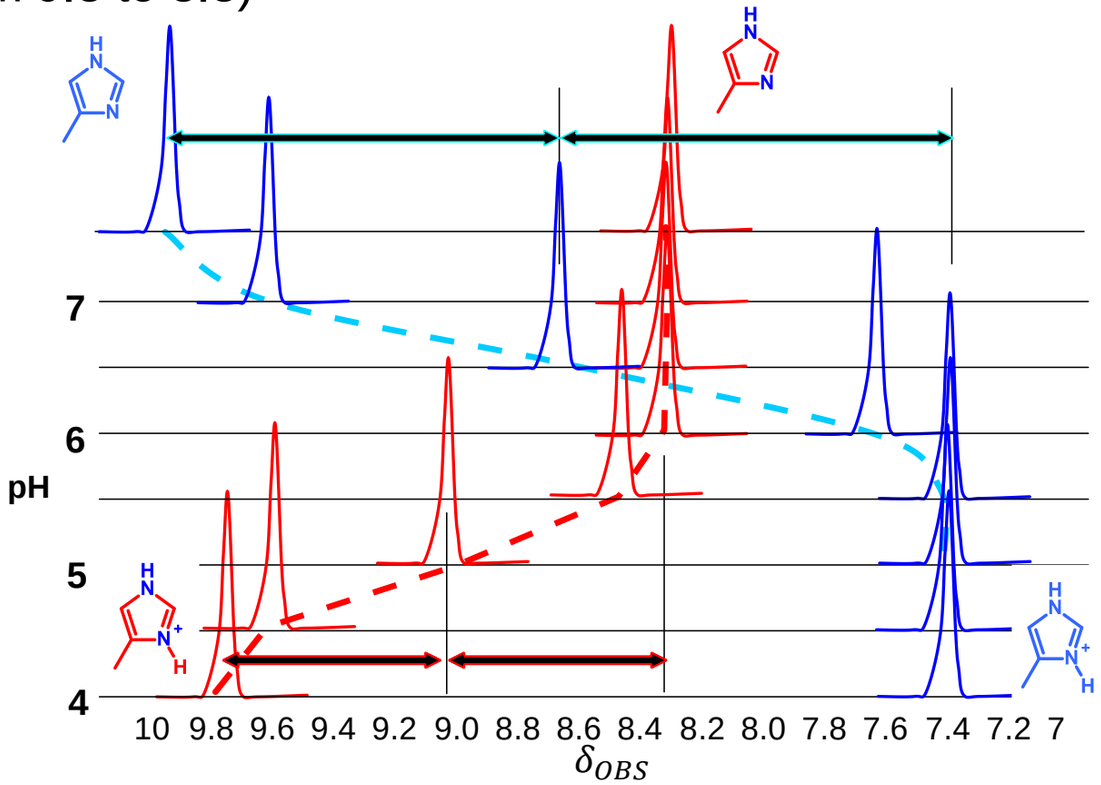


Low pH
Fully protonated



High pH
Fully deprotonated

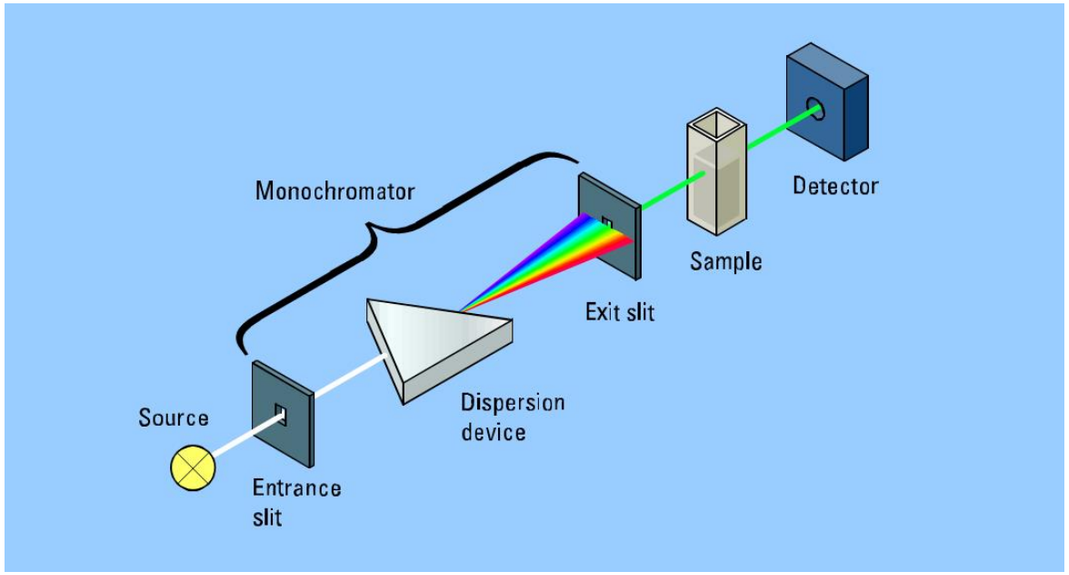
NMR spectra acquired at indicated pH



pKa: pH where peak is ½ way between two extremes

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

Practical Biochemistry - UV Absorption of Proteins to Determine Concentration



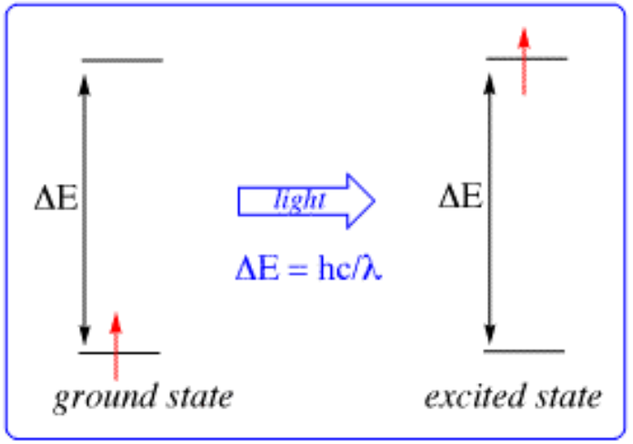
$$A = \log \frac{I_O}{I} = \varepsilon[X]l$$

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).

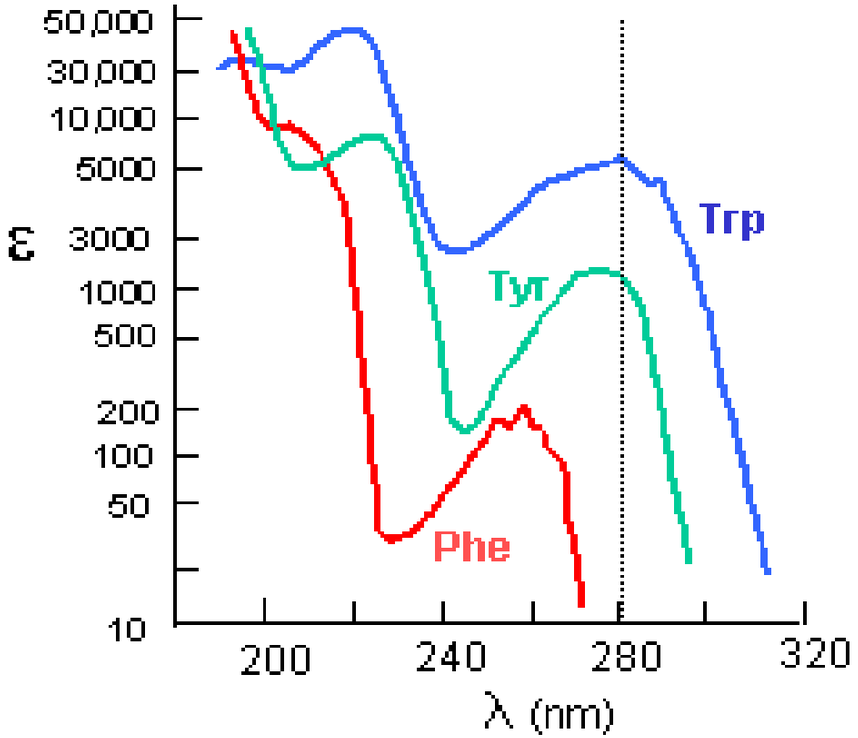
UV-Vis:

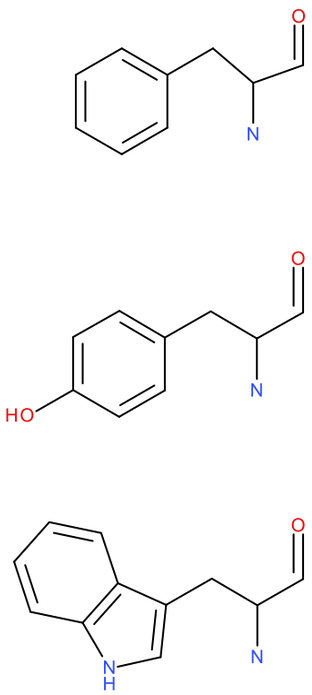
- Light is absorbed by electrons.
- Electrons go from low energy orbital (ground) to higher energy orbital (excited).
- Energy released as heat in non-fluorescent chromophores.
- Absorption energy indicated as wavelength (nm)



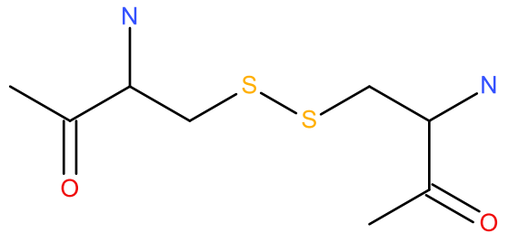
UV Absorption Properties of Proteins

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





Cys in disulfide bonds absorbs weakly at 280 nm, ε = 125/Cys pair



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

Amino acid	λ_{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 $M^{-1}cm^{-1}$

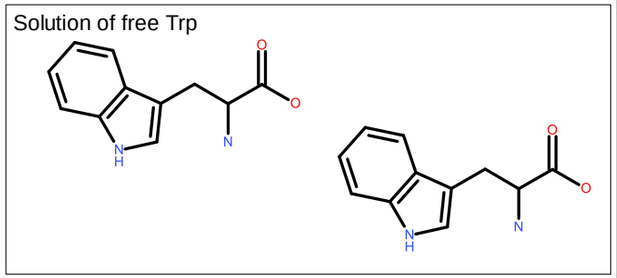
UV Absorption Properties of Proteins

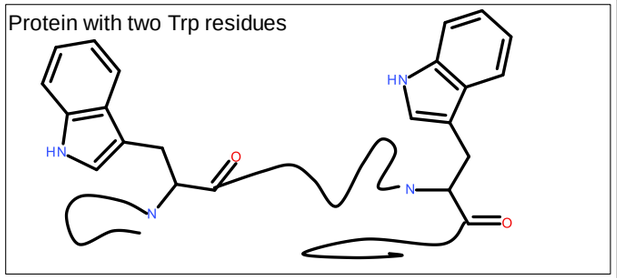
Calculation of molar extinction coefficients:
If a molecule contains a mixture of N different chromophores, the molar extinction coefficient can generally be calculated as the sum of the molar extinction coefficient for each absorbing group in the protein:

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

- 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?





Conc [X]	Abs	ε=A/[X]
200 uM (Conc of Trp)	1.1	5,500 $M^{-1}cm^{-1}$
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

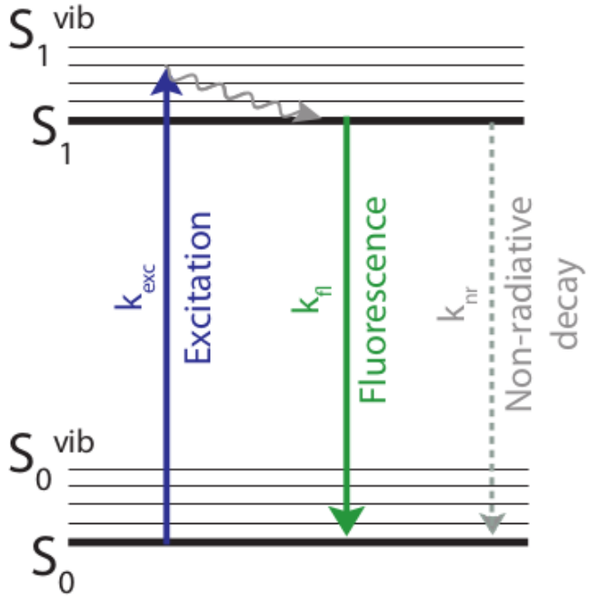
Why do we care? The wavelength and intensity of Trp fluorescence are sensitive to changes in protein structure, providing a convenient way to measure stability of the protein and the binding of molecules to the protein.

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)

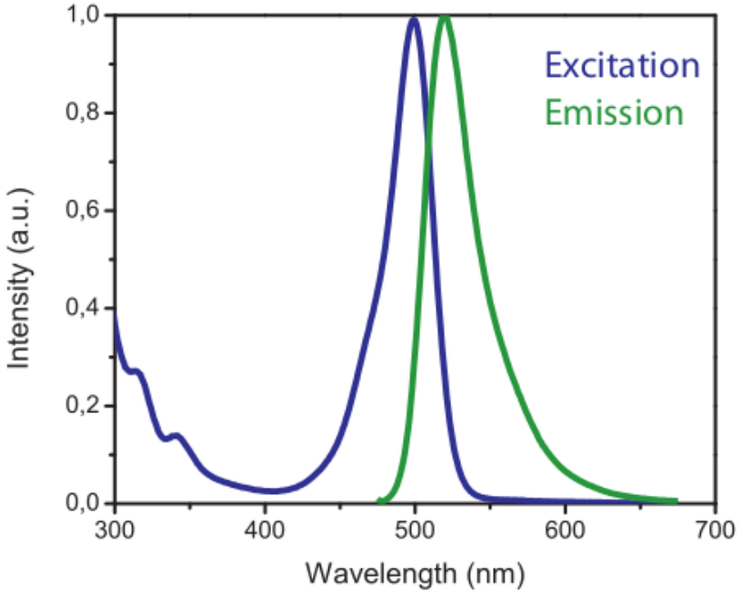
Fluorescence:

- 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.



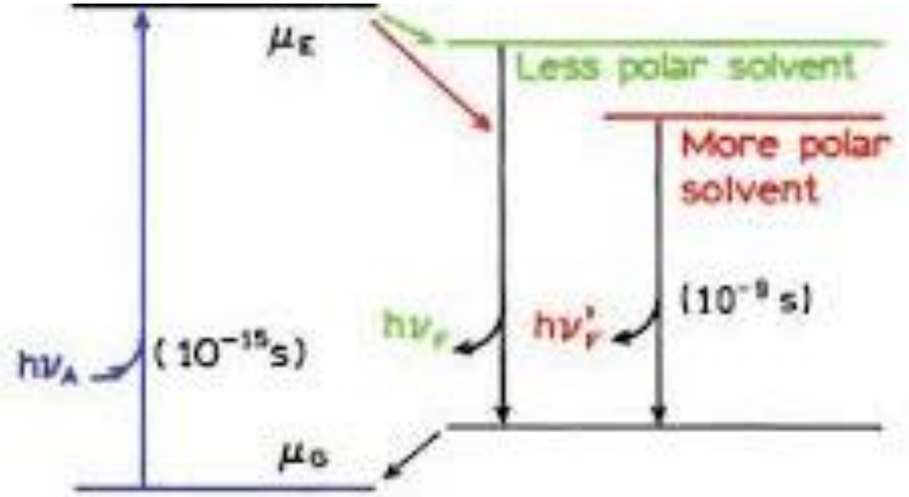
(b)

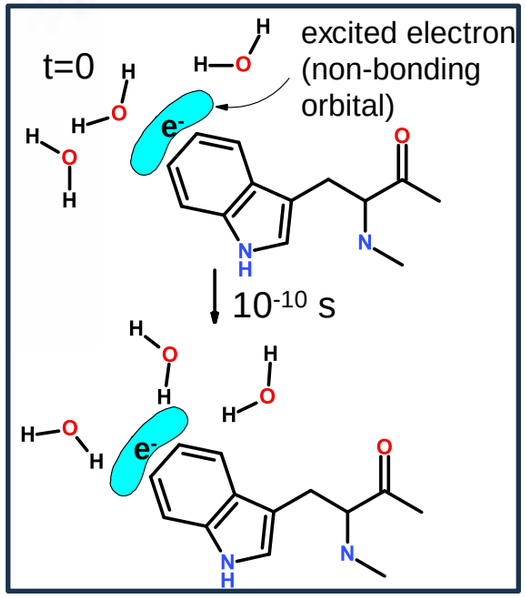
- The emitted photon is always of longer wavelength due to the drop in energy of the excited state.

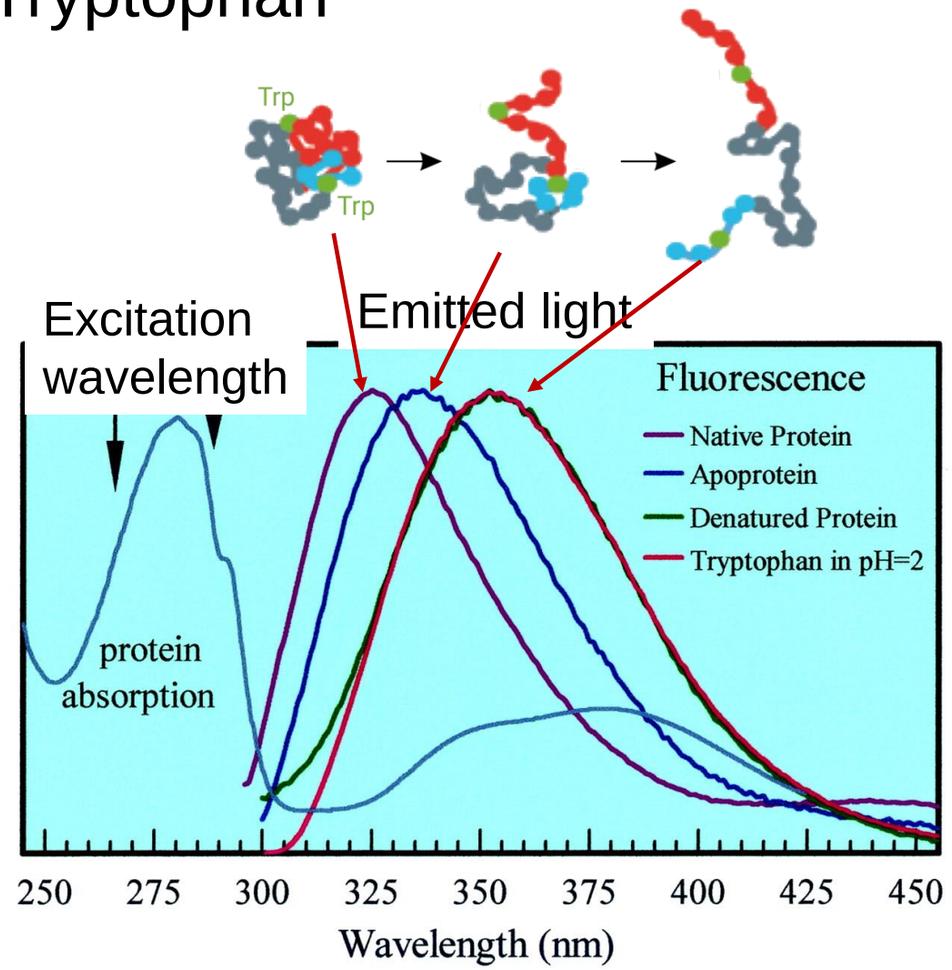
Fluorescence Properties of Tryptophan

- 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.

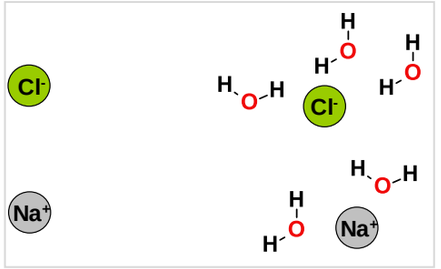
Protein Structure & Stability

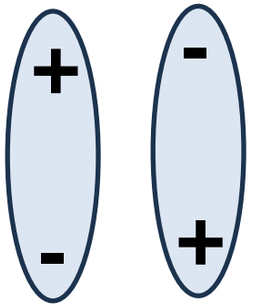
Molecular Forces & Protein Stability

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

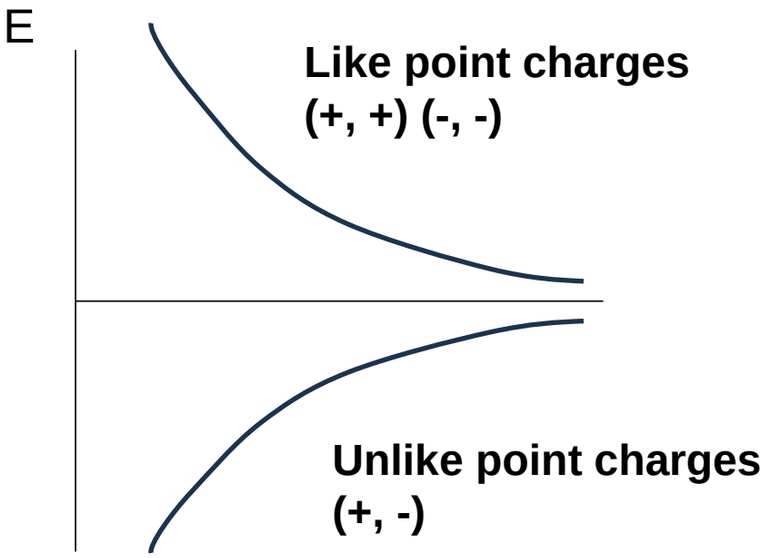
$$E = \frac{q_1 q_2}{Dr}$$

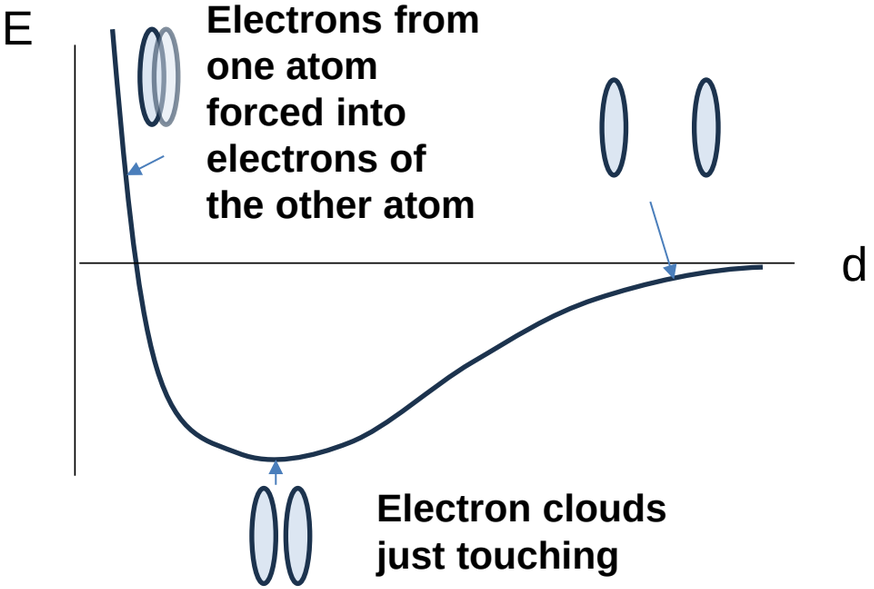
- The energy depends on the charges of the particles (q_1, q_2), distance (r) between the two charges, and the dielectric constant (D) of the media.
- The dielectric constant depends on the solvent. Water has a high dielectric constant because it is a polar solvent.
- The water molecules will orient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





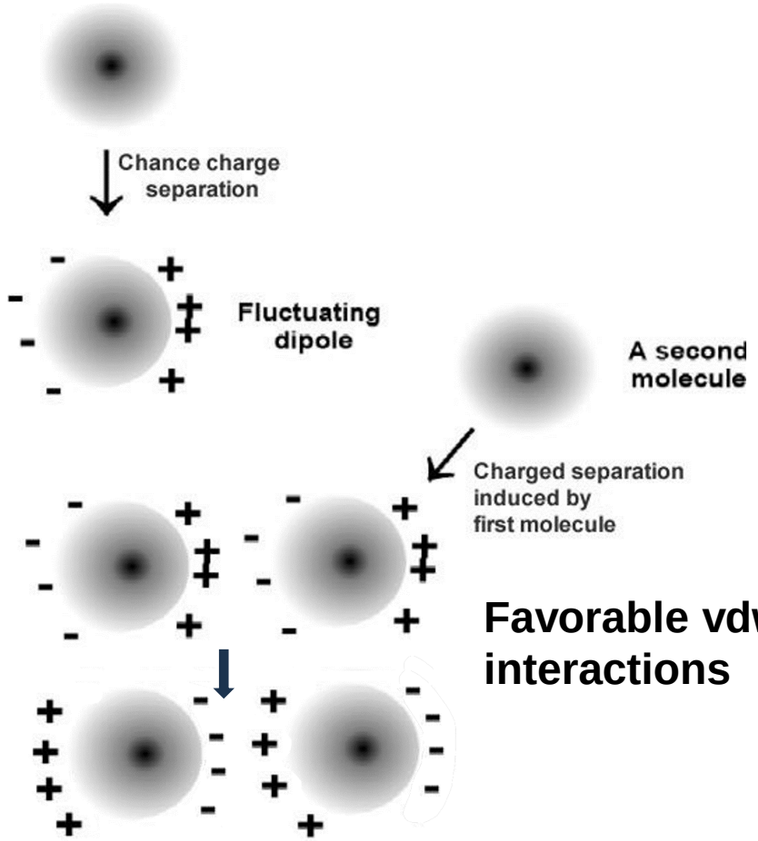
ii) Van der Waals/ Keesom (dipole-dipole) – an electrostatic interaction that involves ***permanent*** partial charges. It does **not** involve **formal charges**.



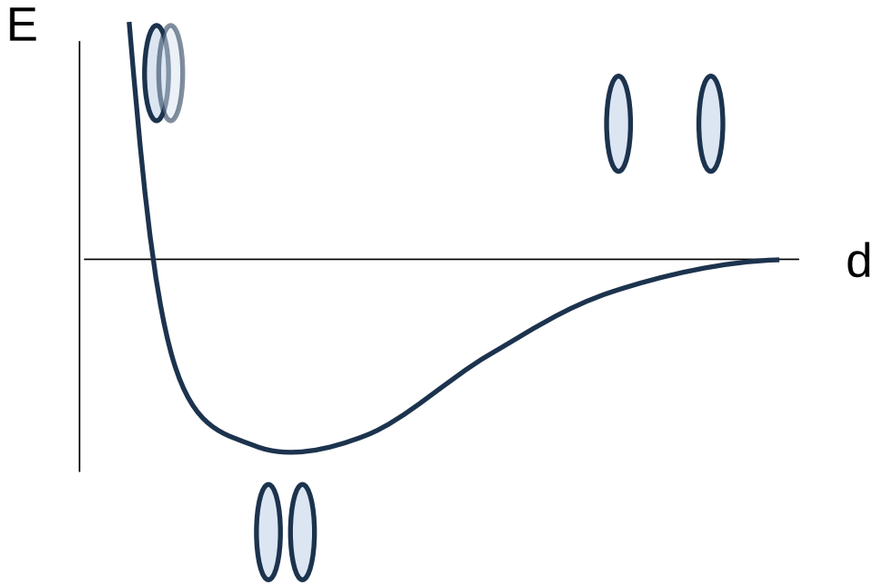


iii) Van der Waals: London Force - Induced dipole

- 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

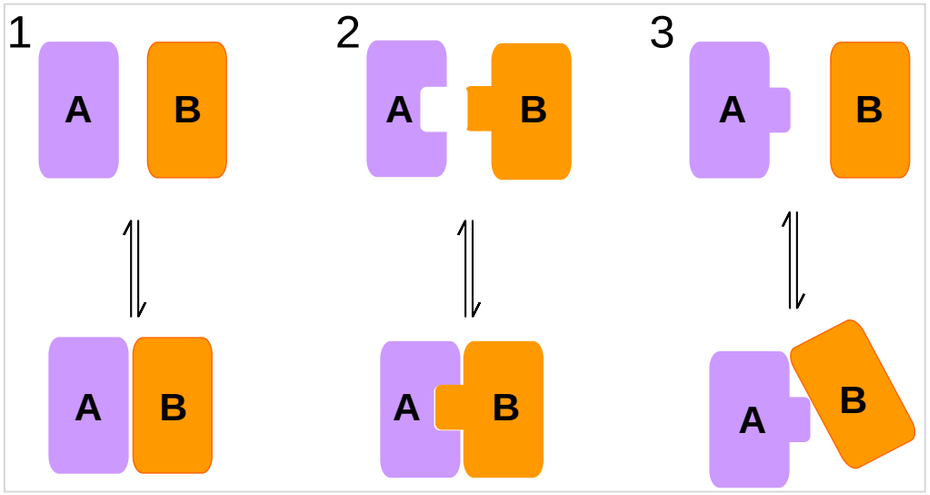


Favorable vdw interactions

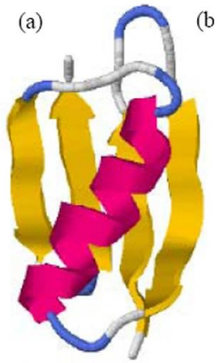


Which of these will have the most favorable vdw interaction:

1 2 3



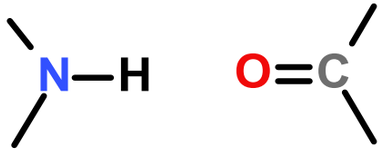
Secondary Structure



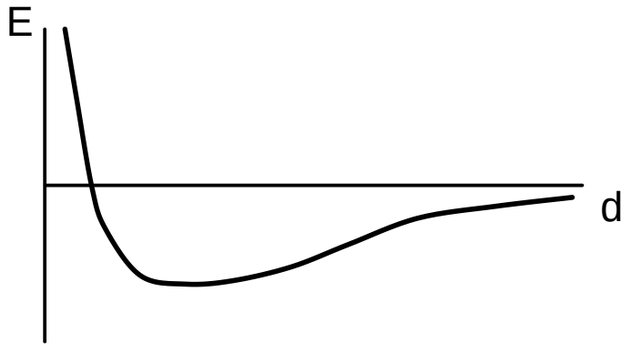
Goals:

- 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)

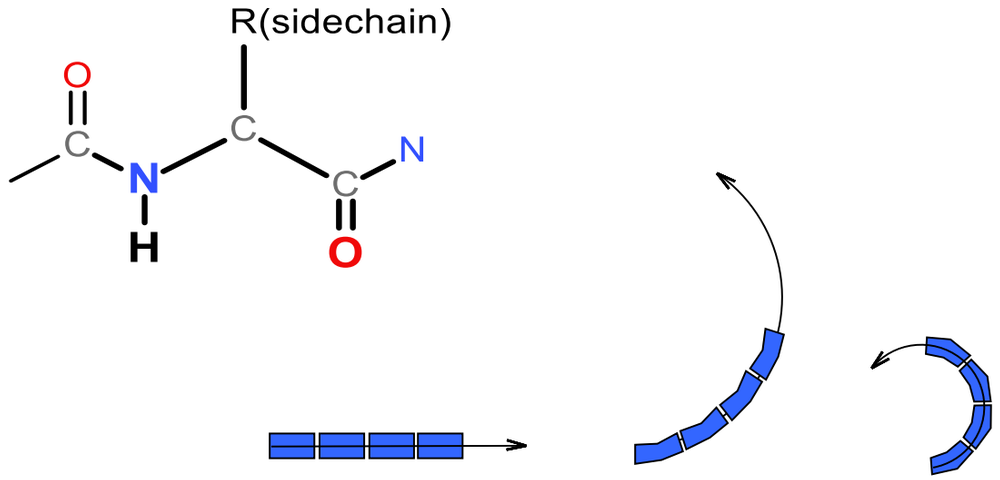


van der Waals (mainchain and sidechain)



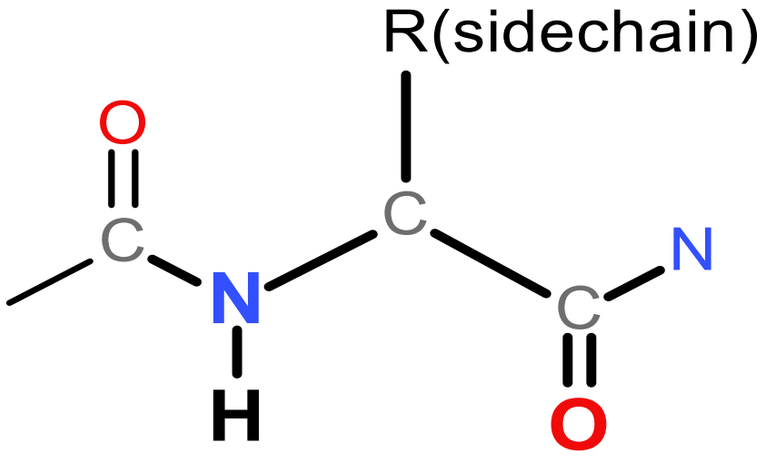
Regular Secondary Structure:

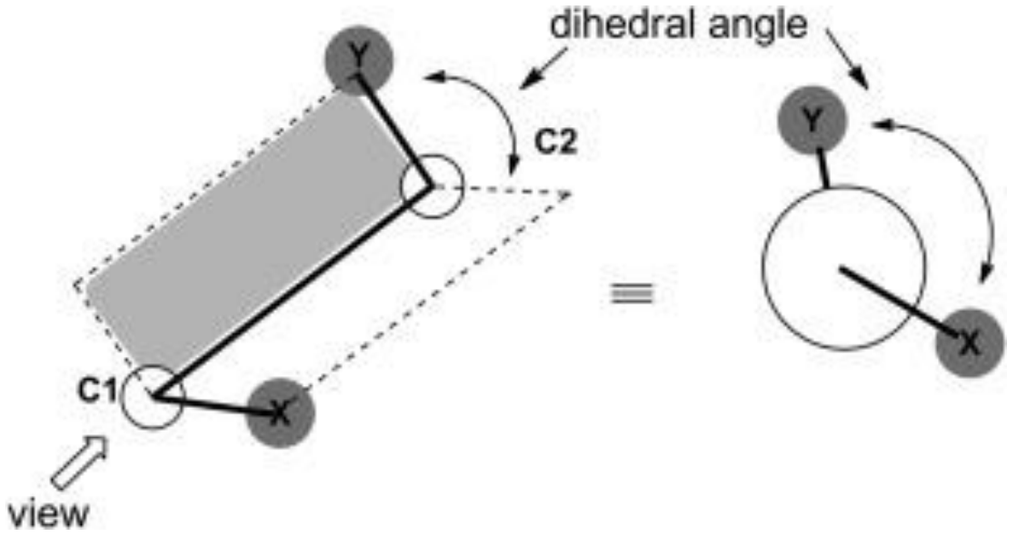
- Local conformation of only the ***mainchain*** atoms.
- Regular secondary structure because each residue within a secondary structural element has the same geometrical shape due to identical rotational angles about bonds in each residue.
- The repeating geometrical properties of the subunits will result 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.



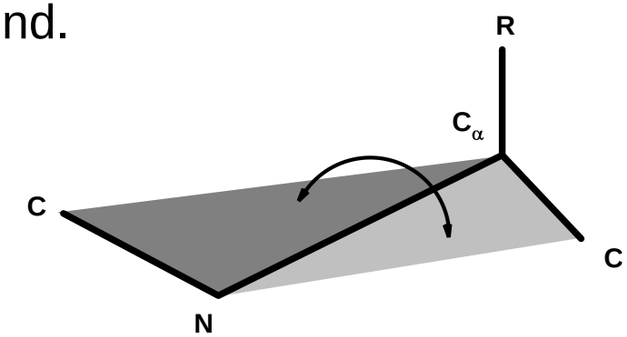
The “shape” of each amino acid depends on the conformation of the bonds within a residue:

- **ω (omega)** - peptide bond
- **Φ (Phi**), the bond between N and C_α
- **Ψ (Psi),** the bond between C_α and C.



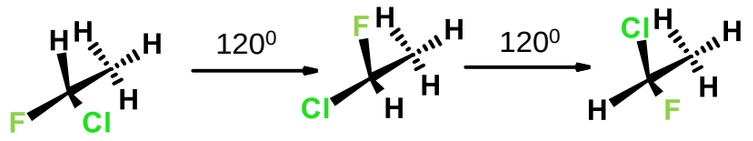


The angle of rotation around a bond is specified as the **dihedral angle** between two planes that share the bond.

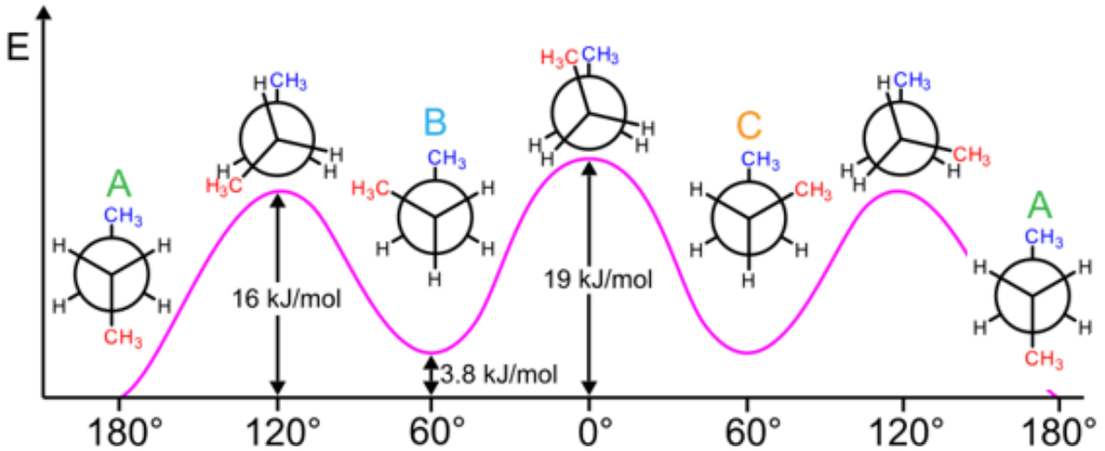


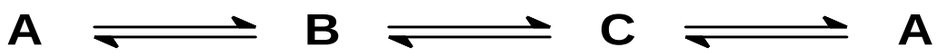
The two planes that are used to specify the phi angle are illustrated above.

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.

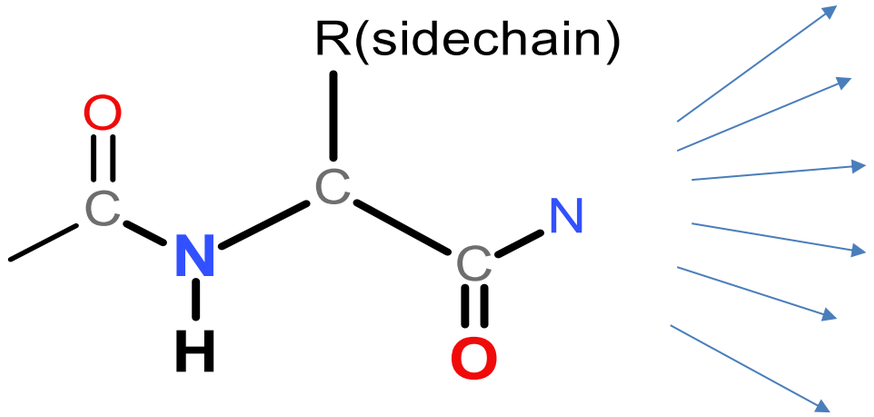


Bond rotation about the C-C bond in butane indicates relatively small energies between each stable state, so interconversion between rotational conformations will easily occur and a molecule will sample all possible low energy states.





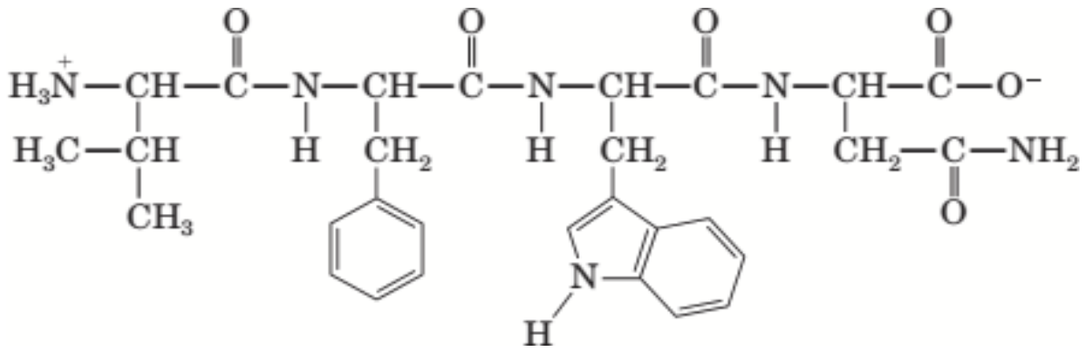
***Reflection**: How many different conformations could the backbone atoms of one residue in a protein take, assuming free rotation around **each of the** three bonds (C-N, N-C_α, C_α-C)?*



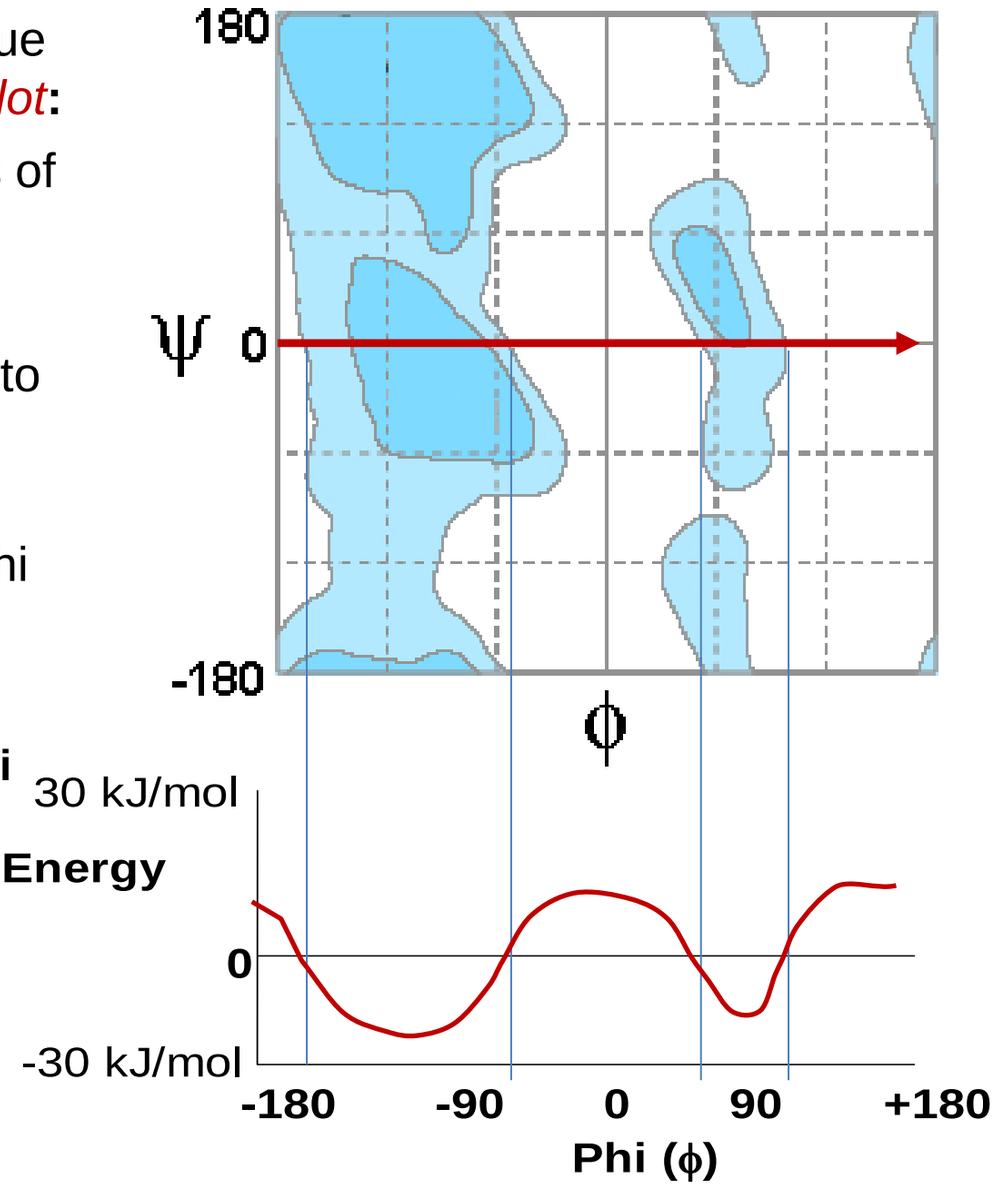
Ramachandran Plots – Representing the phi and psi angles of each residue

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 to close together.



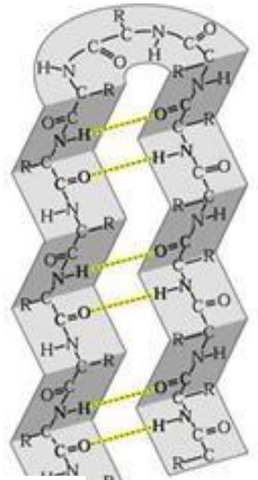
Plot of Energy versus Phi for Psi =0.

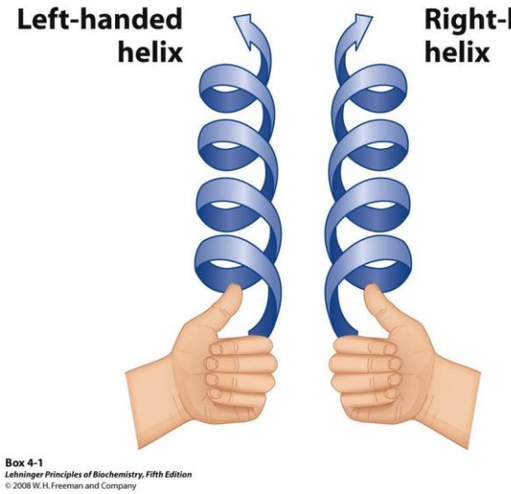


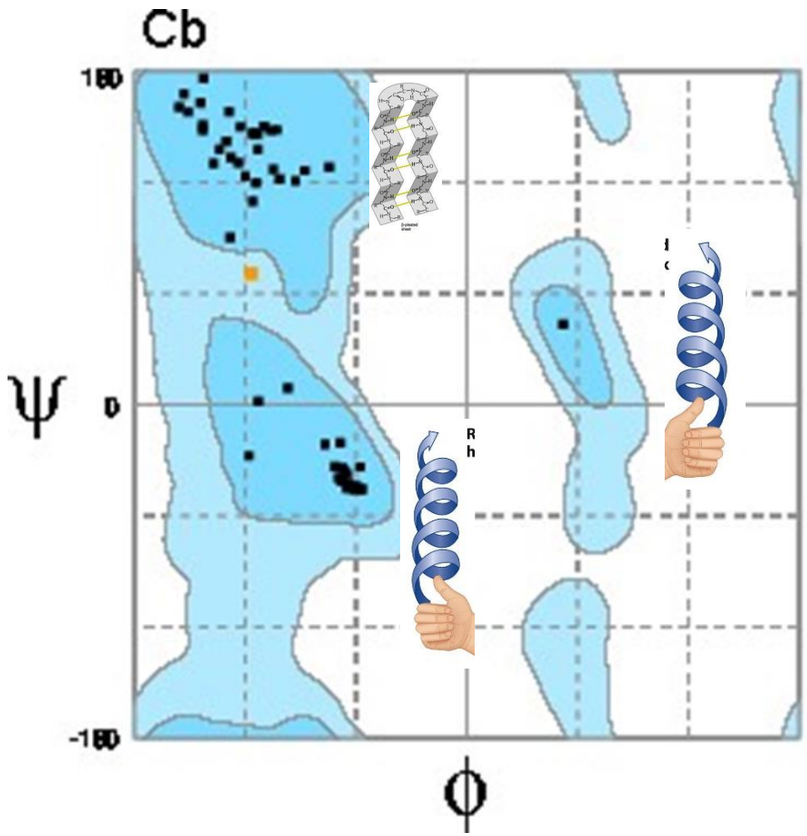
How many Secondary Structures are Found in Nature?

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.







Expectations – Ramachandran plot

- How it is constructed
- Location of different secondary structures
- Origin of colored and white regions (vdw)

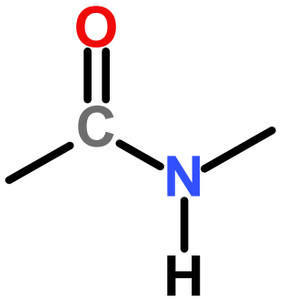
Why are there only three possible "shapes"?
Free rotation about all three bonds should give us 27 possible shapes!

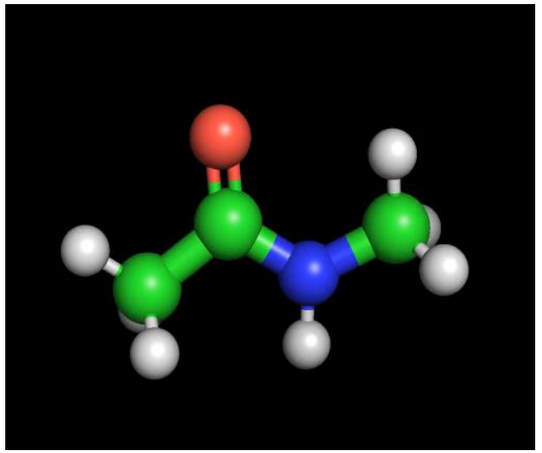
What limits the numer of conformations to three?

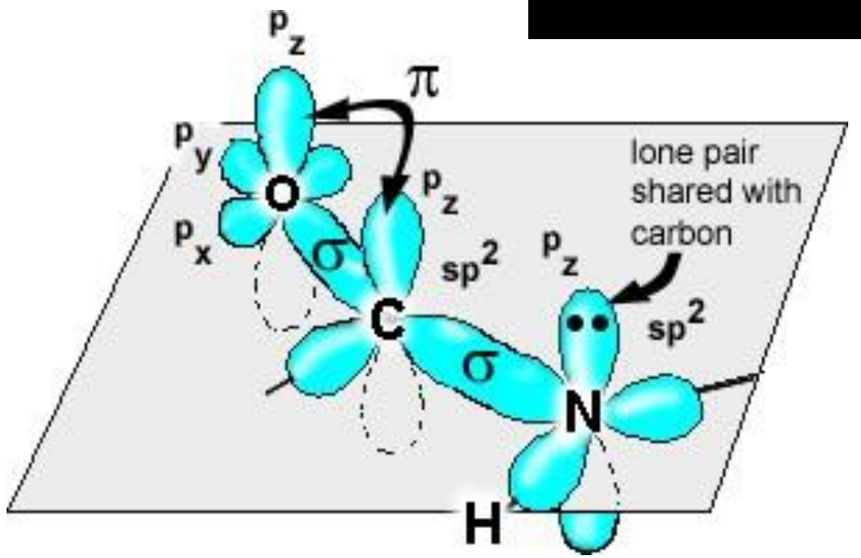
Conformational Freedom of Proteins – One bond at a time

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".



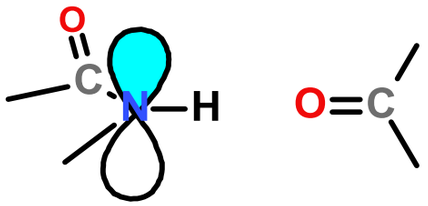


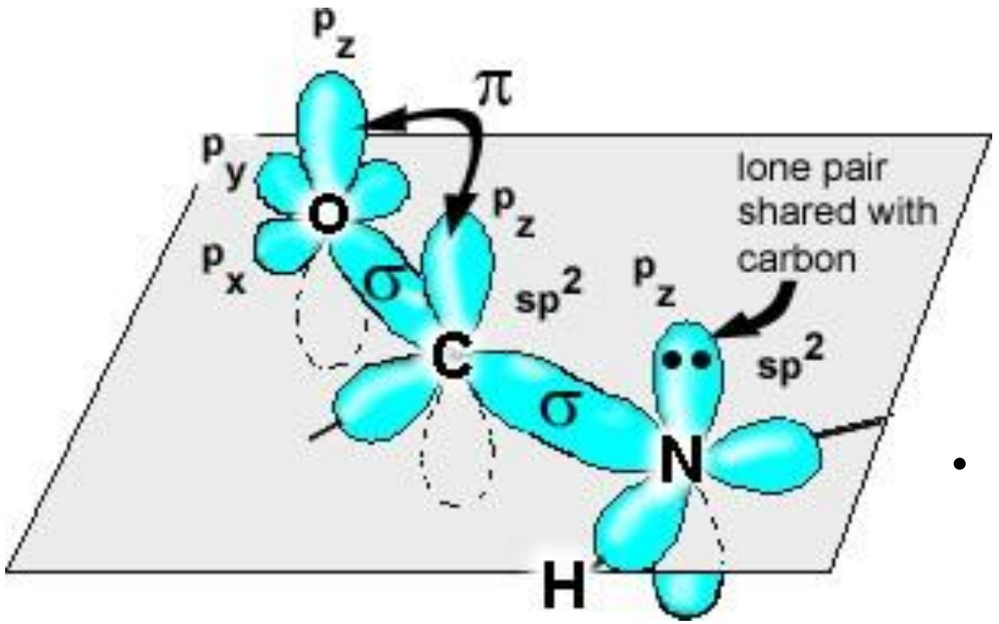


- **The N-H group** within the peptide bond can ***only*** act as a hydrogen bond donor. The partial negative charge on the nitrogen is delocalized over the entire conjugated system so it is not energetically favorable for it to accept a hydrogen bond (a similar argument applies to the NH group on Trp sidechain, and the amide group found on Asn and Gln sidechains).

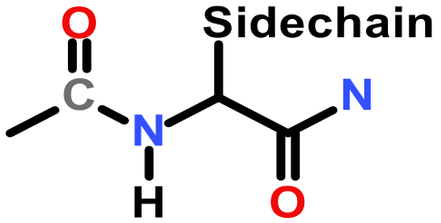
Expectations – Peptide bond

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

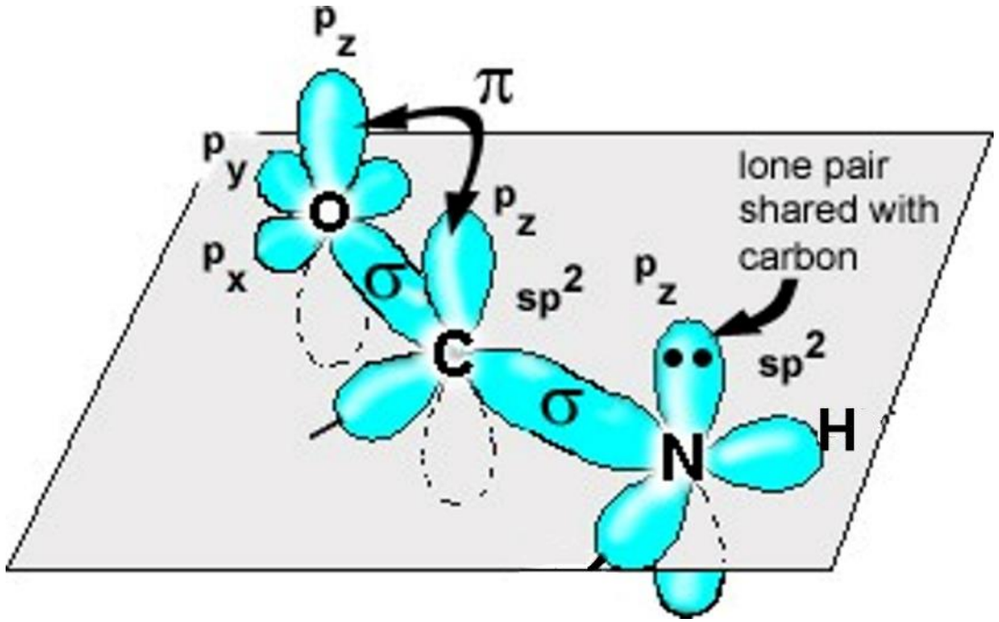




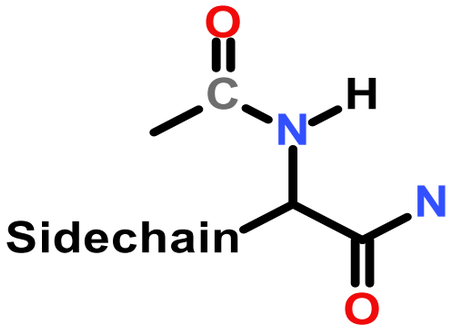
Trans



- 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.

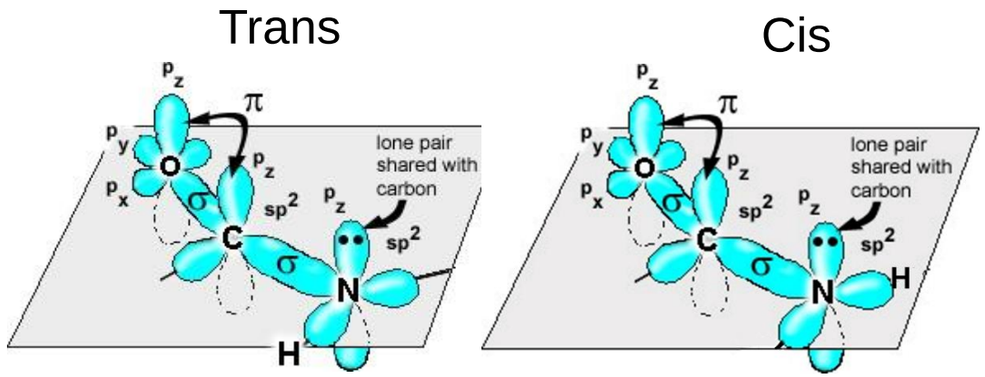


Cis



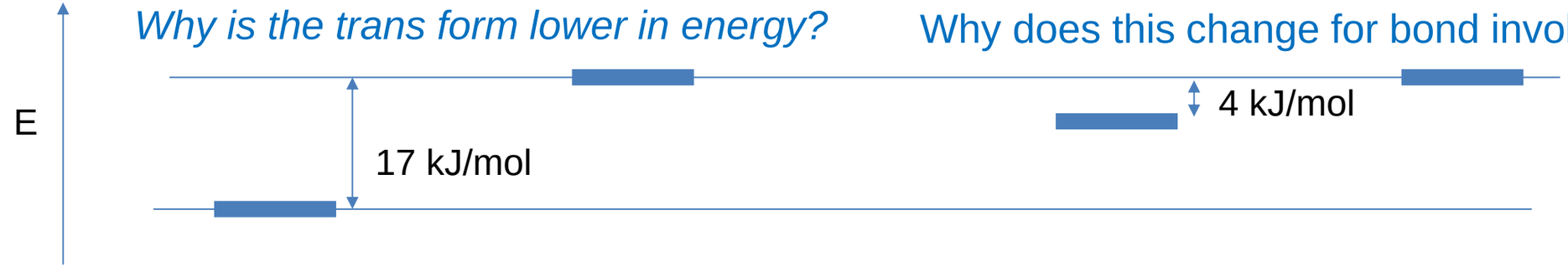
Cis and trans Peptide Bonds:

- Two possible orientations of the peptide bond allow overlap of the p_z orbital between the C and N.
- They are related by a 180º 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

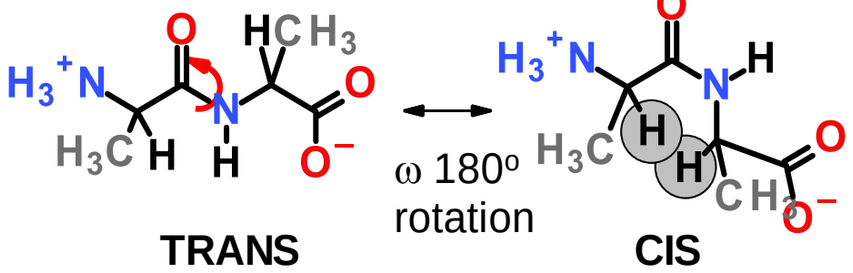


Why is the trans form lower in energy?

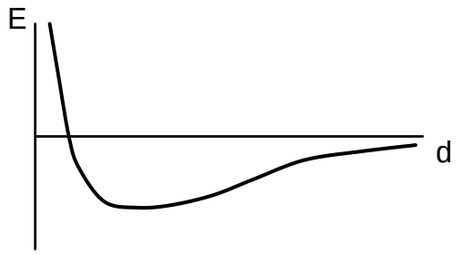
Why does this change for bond involving proline?



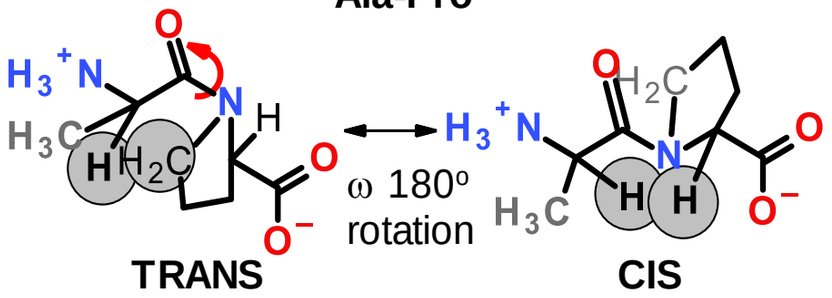
Ala-Ala



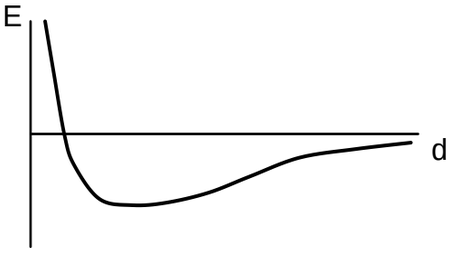
TRANS **CIS**



Ala-Pro

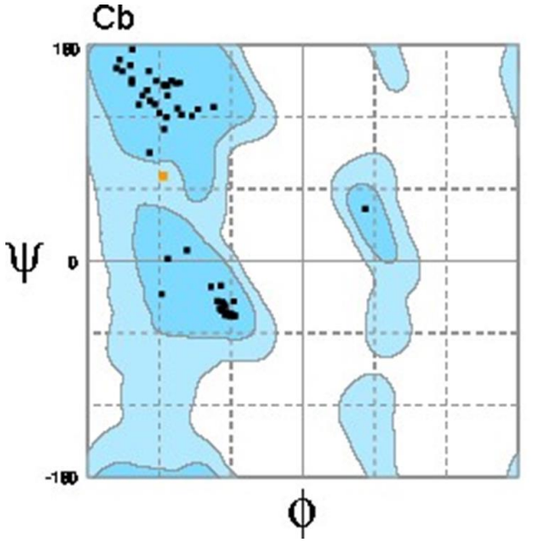


TRANS **CIS**

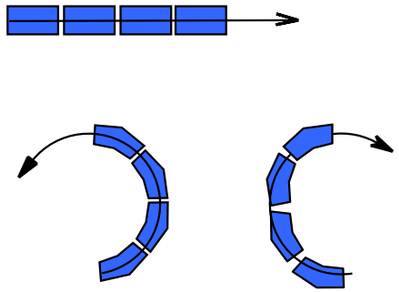


N - C_α & C_α - C Bonds:

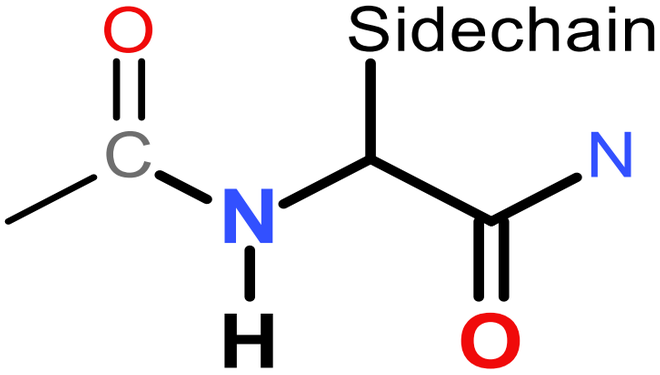
- There is free rotation about *both* of these bonds.
- Since both the phi and psi bonds are free to rotate, a total of 9 (3x3) different stable pairs of phi and psi angles should be observed.
- However, there are only three pairs, each of which correspond to a secondary structure:



Φ = -120°, and Ψ = 125°. extended, β-structure
Φ = -60°, and Ψ = -45° α-helix, right handed
Φ = +60°, and Ψ = +45°. α-helix, left handed



- Ψ and Φ 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).

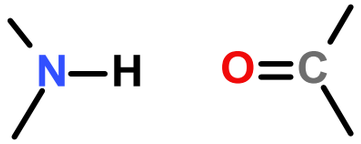
***Reflection**:*

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

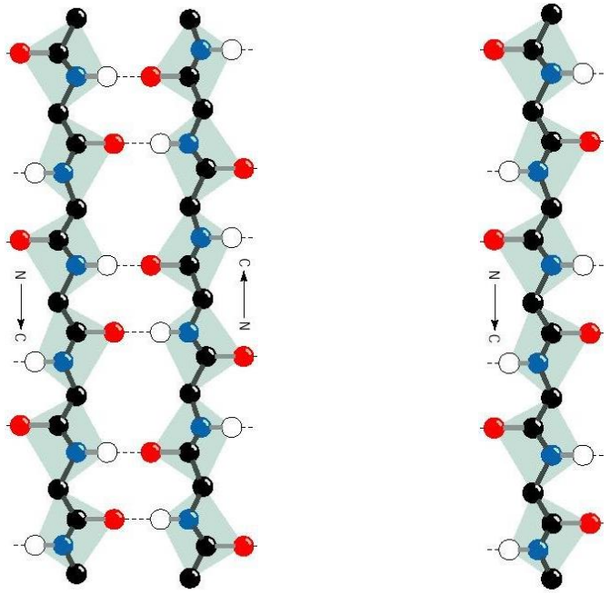
What forces are stabilizing, and which are destabilizing?

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

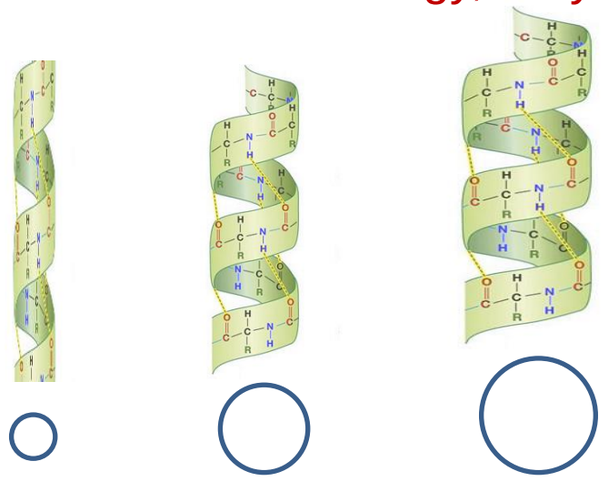
H-bonds. Not all phi & psi angles will allow ***mainchain*** H-bond donors and acceptors to align properly, this excludes many of the possible phi and psi angles.

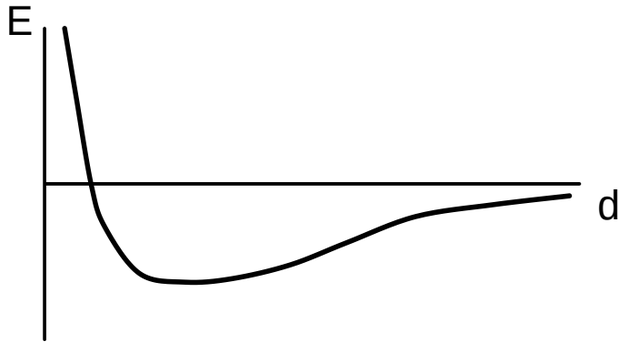


B-strands allows H-bonding



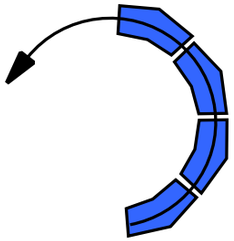
Many helices allow the formation of H-bonds, but only one is lowest in energy, Why?





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

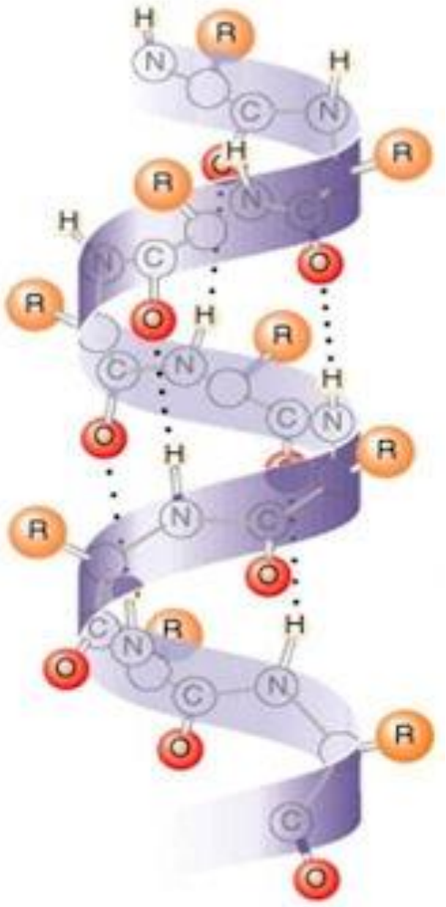
α-Helix Structures (Φ = -60°, Ψ = -45°)

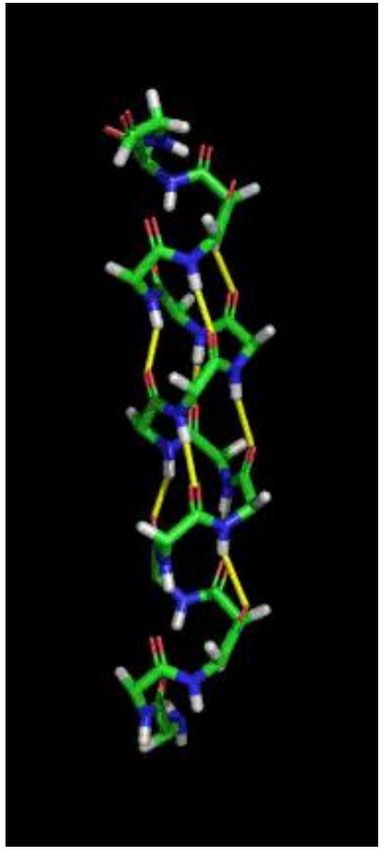


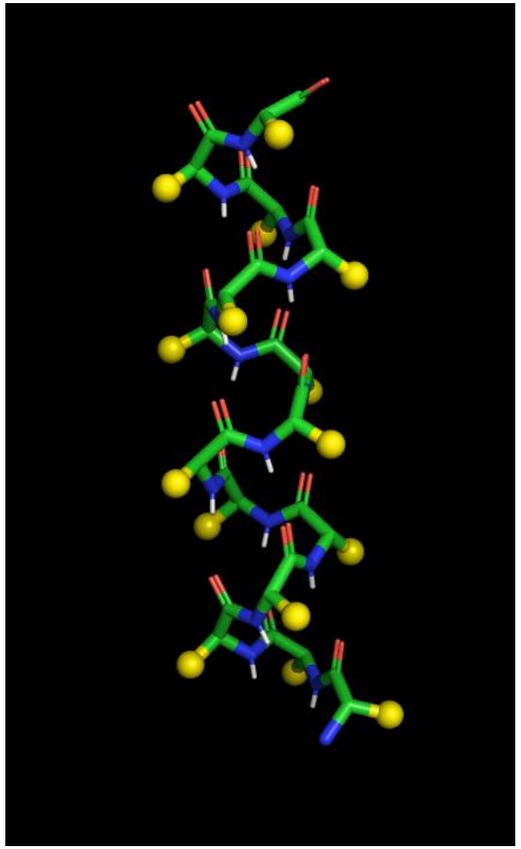
_____ residues/turn

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°)

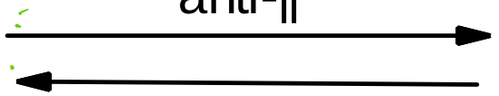
- H-bonds perpendicular to strands.
- Strands can be parallel or anti-parallel. This refers to direction of strand, from N to C, the strands always align adjacent to each other, forming H-bonds.
- Sidechains:

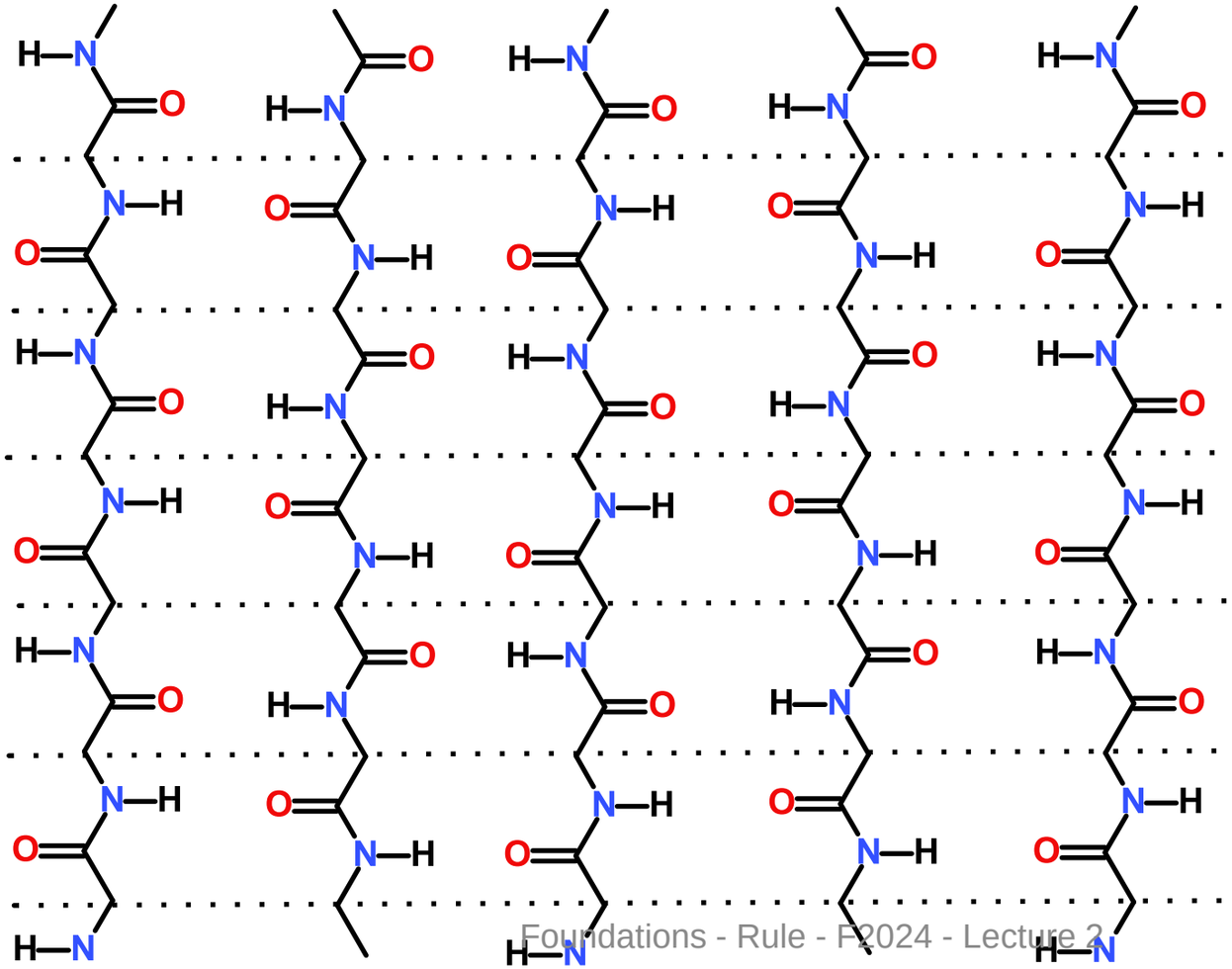
✓ alternate up and down on any ***one strand,*** i.e. above and below the sheet

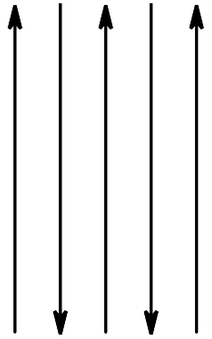
✓ sidechains on ***adjacent strands are pointing in the same direction.***

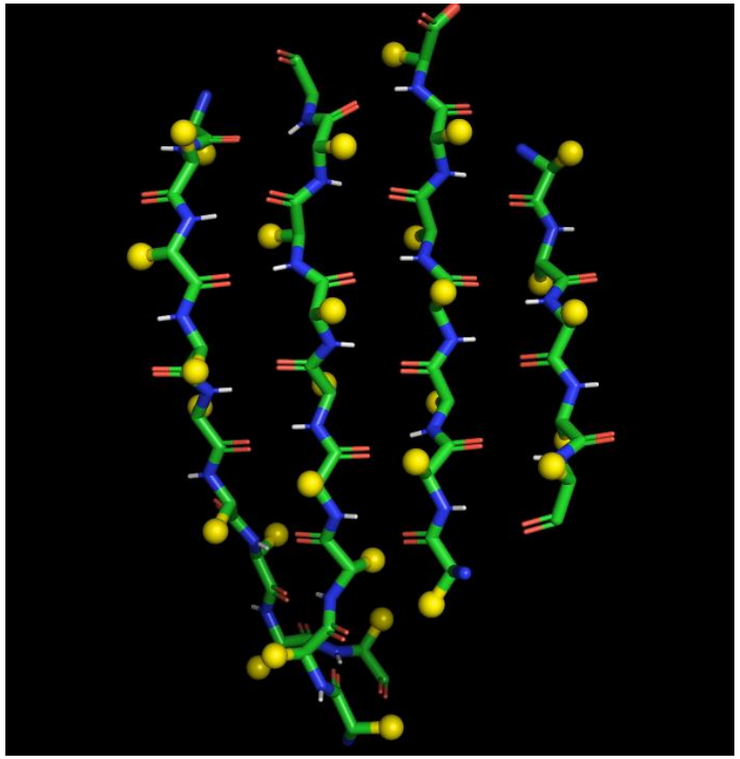
parallel

anti-‖



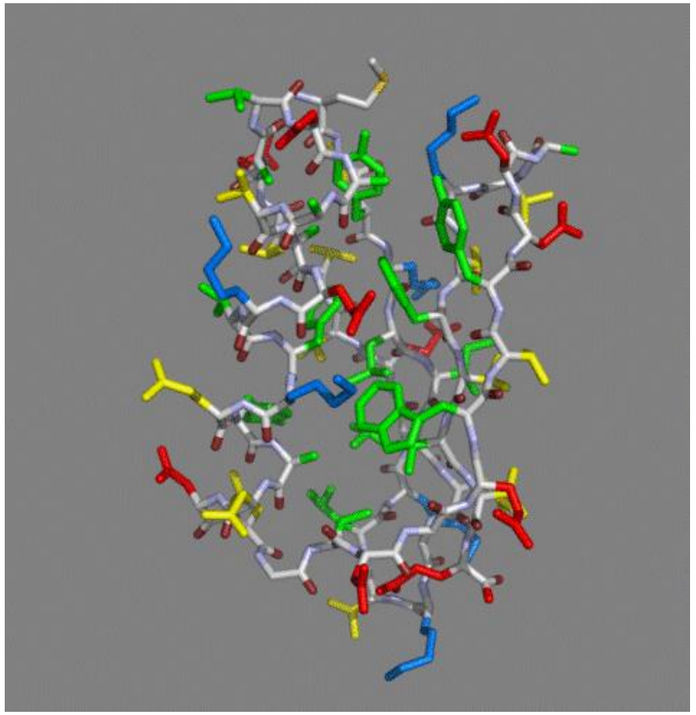






Tertiary Structure = Complete Structure of One Chain

Location of Residues in Globular Proteins

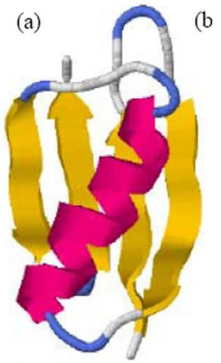


Red - amino acids with neg. sidechains

Blue - amino acids with pos. sidechains

Yellow – amino acids with polar sidechains.

Green - amino acids with hydrophobic side chains

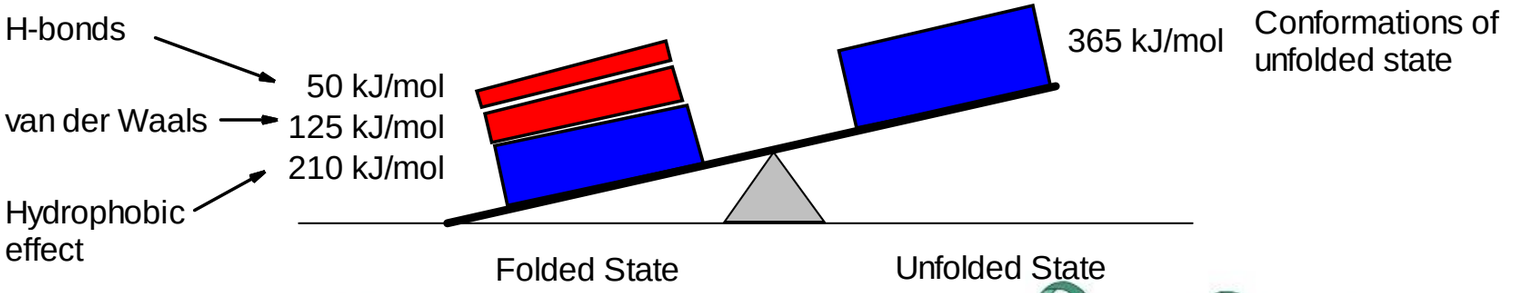


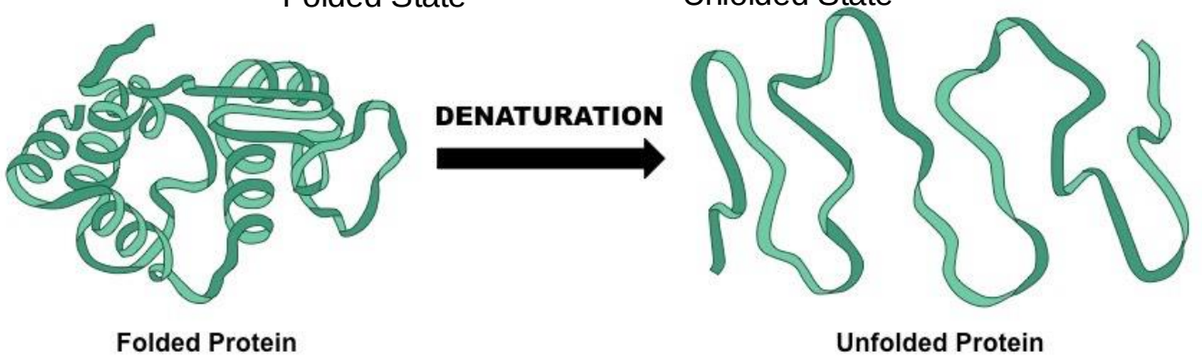
Amino Acid		
	Inside	Surface
Charged		
Polar		
Non-polar		

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.

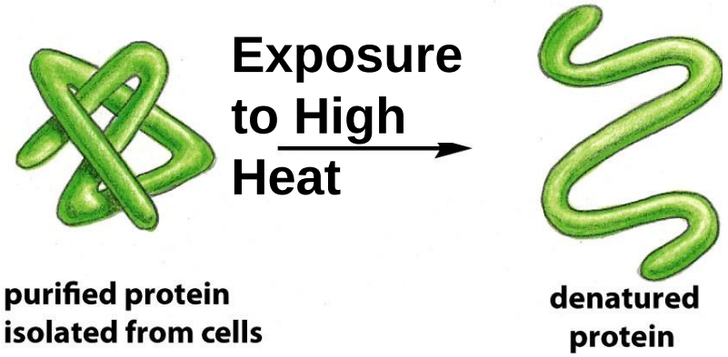
ΔS^o : Change in disorder of the system. ***positive*** entropy change is ***favorable*** since the ***disorder*** in the system is ***increased***.

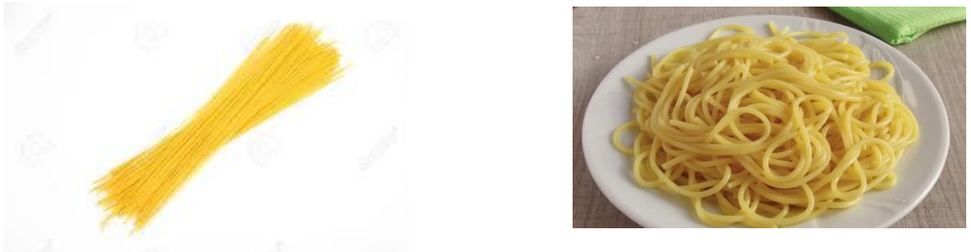
ΔS^o is positive for unfolding, favoring the unfolded state.

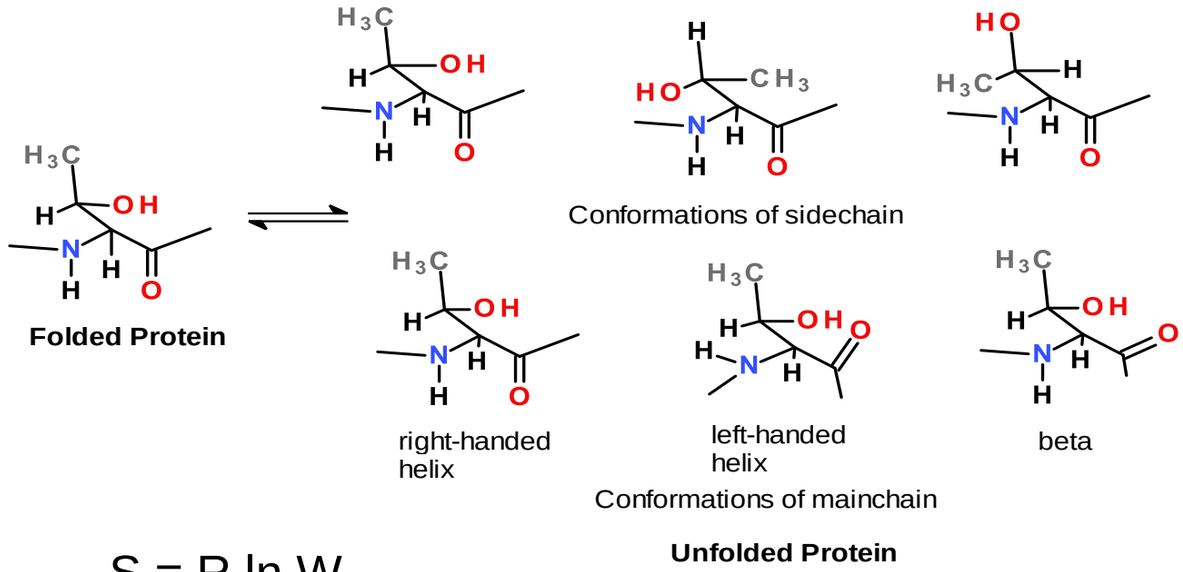
ΔH^o : Enthalpy – A change in the electronic configuration of the system that either releases *heat* ($\Delta H^o < 0$) or absorbs *heat* ($\Delta H^o > 0$). **Release of heat is favorable**.

ΔH^o is positive for unfolding, heat needs to be added to unfold proteins

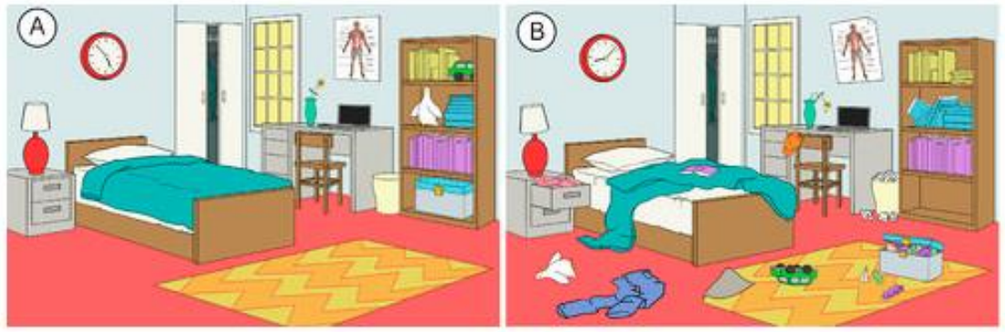
Unfolded Polypeptides Are Flexible – High Entropy (Disorder) Stabilizes the Unfolded State







Energy and Entropy (disorder)



$S = R \ln W$

For an N-residue Protein:

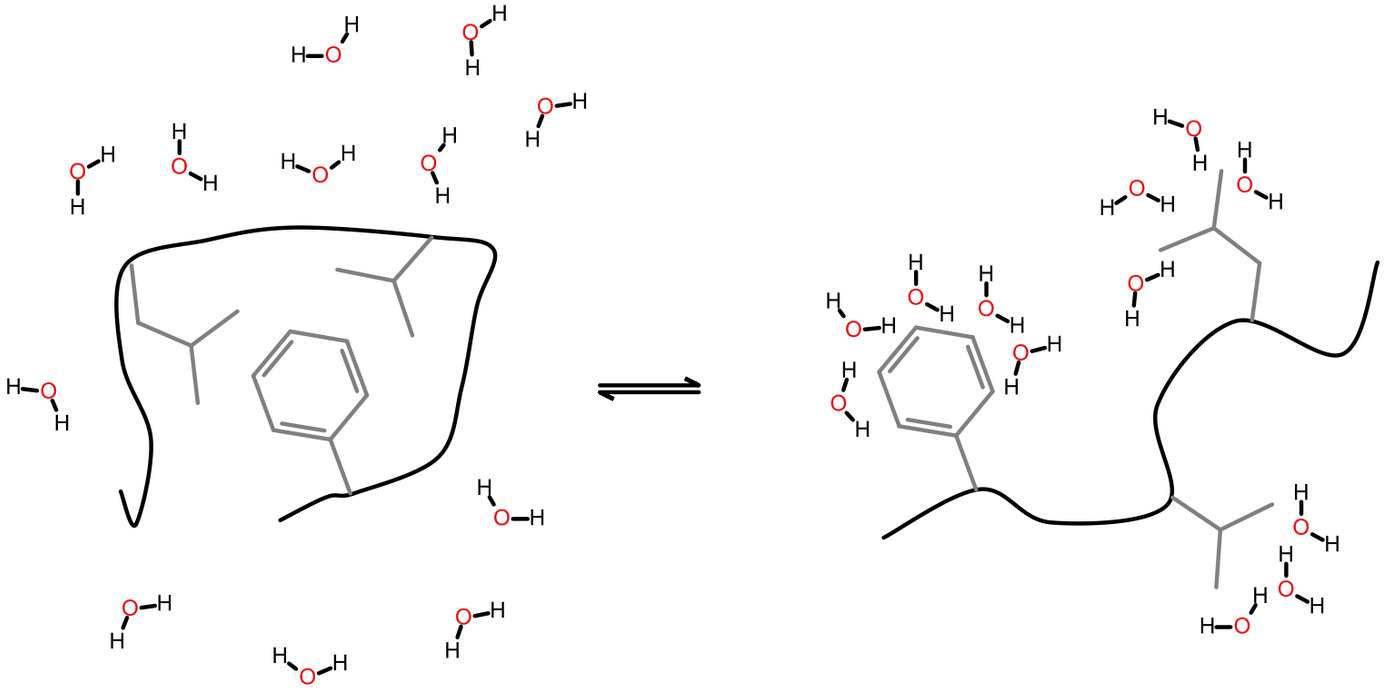
$S_N = R \ln 1 = 0$

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

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

$= 355$ kJ/mol

ΔS^o - 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.



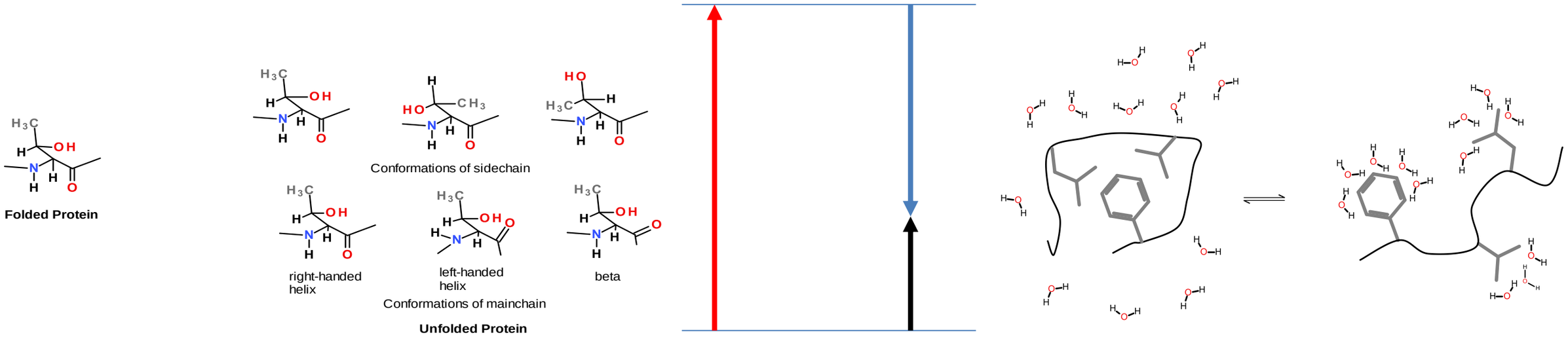
Dissolved butane, ordering hydrogen bonded waters

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, ΔS^{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:

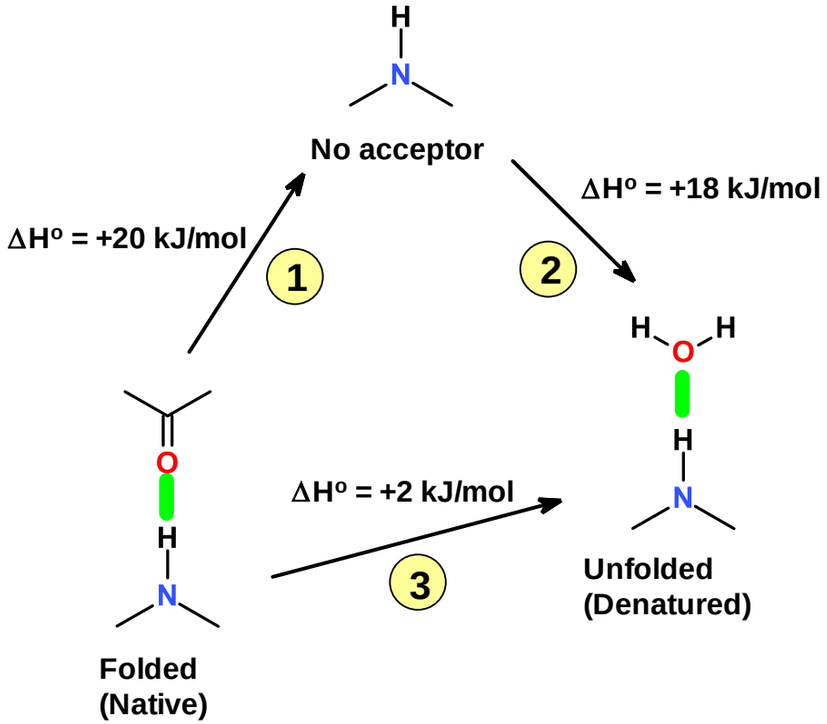


$$\Delta\overrightarrow{S_{OBS}} = \overrightarrow{\Delta S_{Chain}} + \overrightarrow{\Delta S_{Solvent}}$$

Enthalpy (ΔH°) is the amount of heat generated/consumed by the reaction when 1 mole of reactants are converted to one mole of products.

ΔH° - Hydrogen bonds:

- ΔH° 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.



Hypothetical reaction:

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

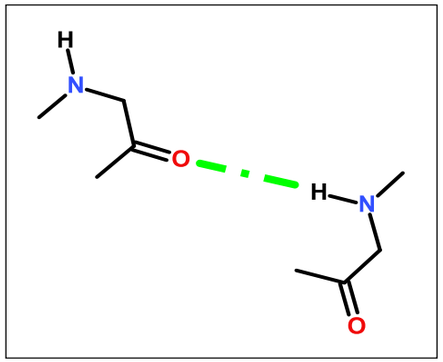
2 – Reforming hydrogen bond with water. Energy released is 18 kJ/mol. Weaker due to less favorable geometry.

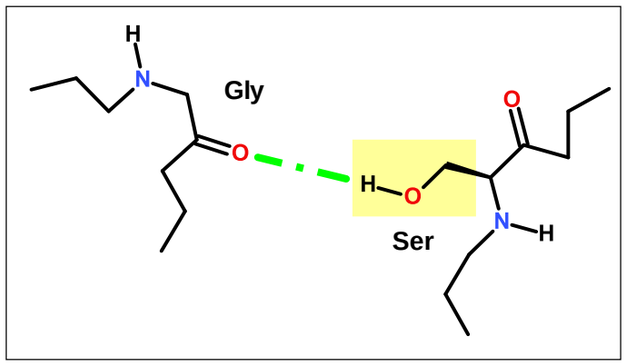
3- Overall change in enthalpy is the sum of 1 + 2

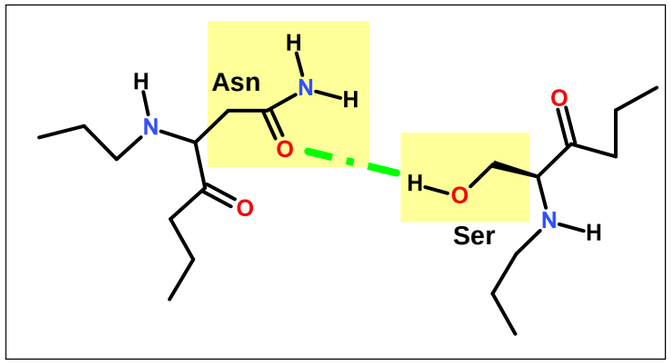
H-bond in folded protein is ~2 kJ/mol ***more*** stable than H-bond to water.

Tertiary structure stabilized by:

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



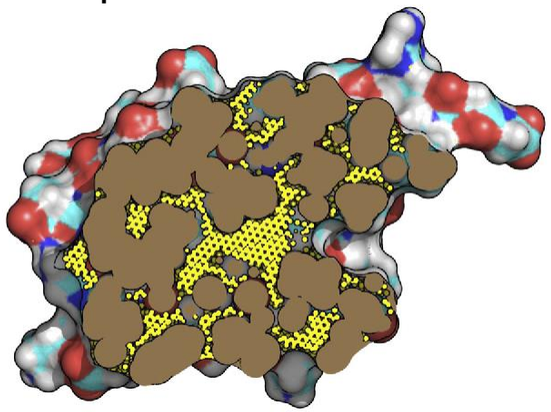




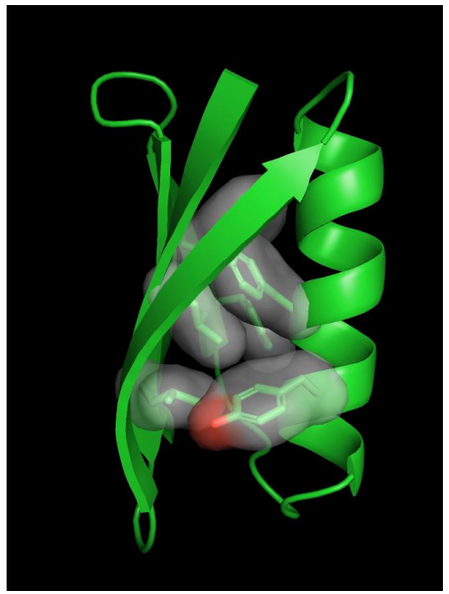
ΔH^o - Van der Waals (London) Forces. This is **un**favorable for **un**folding. Van der Waals interactions are more stable in the native form of the protein.

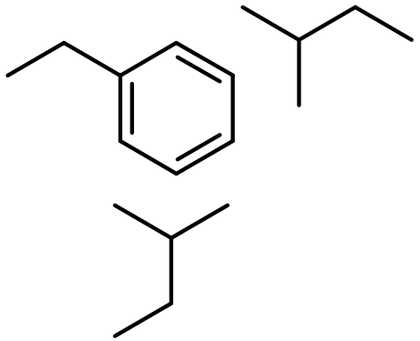
*What enhances the higher stability of van der Waals interactions in **folded proteins**?*

Optimal packing of sidechains in core of protein

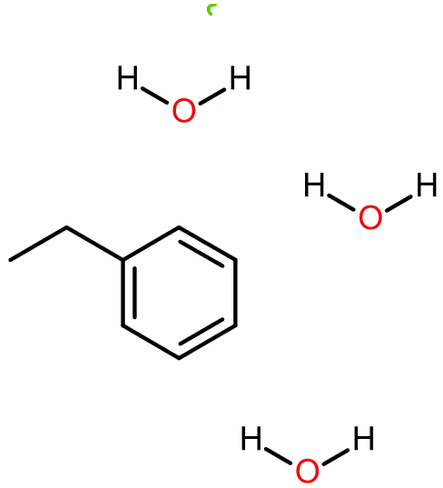


Cross section of a folded protein – knobs fit into holes

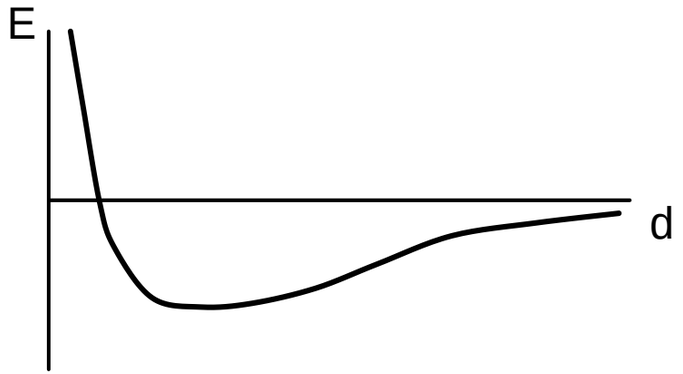




Native

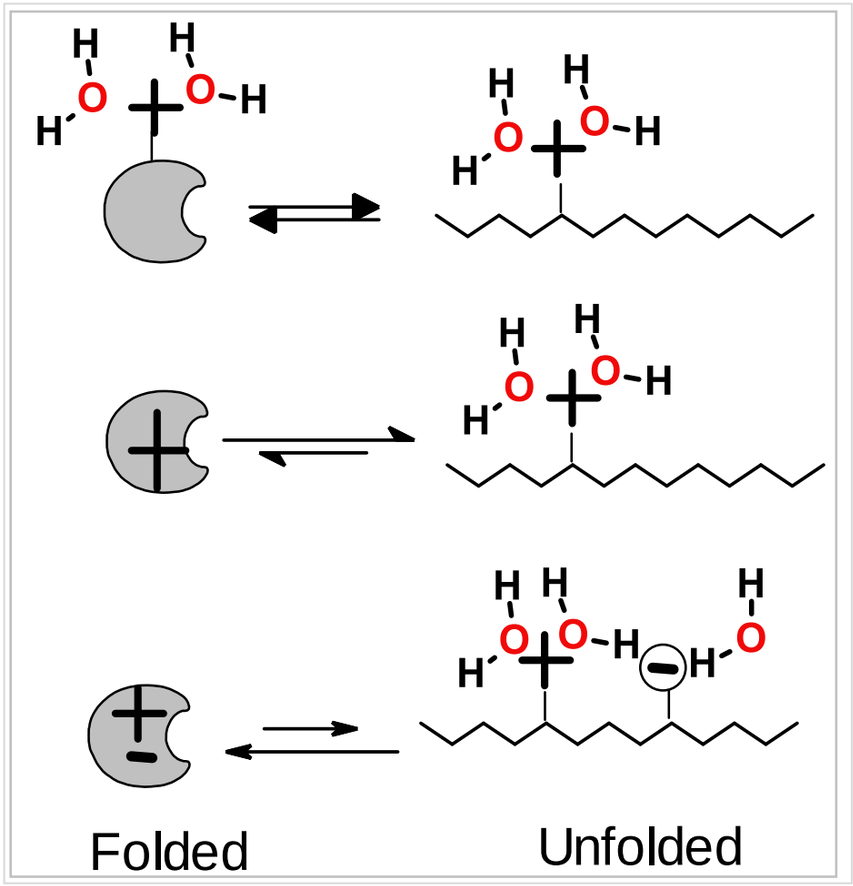


Unfolded

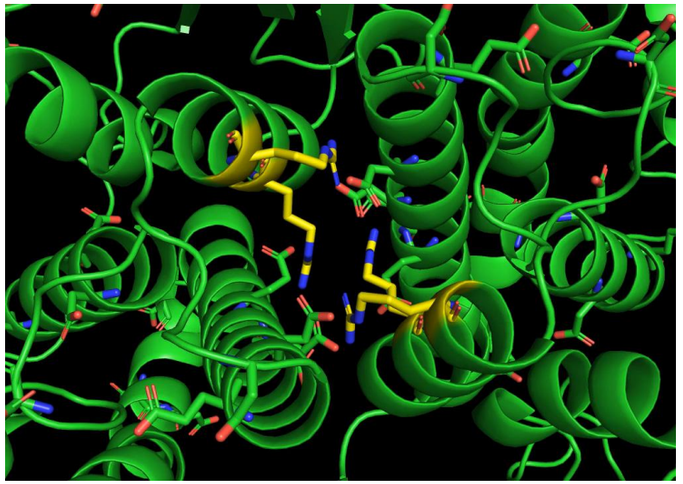


ΔH^o - **Electrostatic forces**:

- **Surface charges:** Although these forces can contribute to ΔH^o 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 $E \propto q_1 q_2 / Dr$).
  - $D_{water} = 80$
  - $D_{interior} \sim 5$



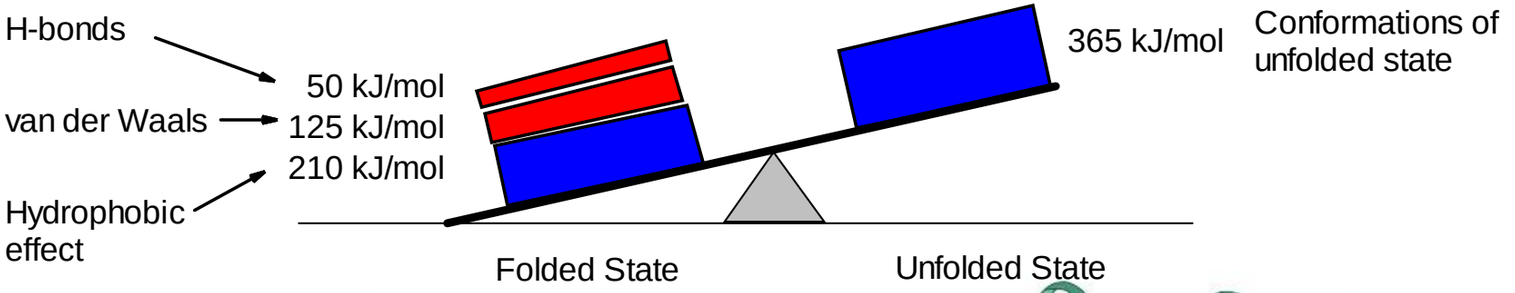
Buried charges stabilize some protein-protein interfaces.
In this homodimeric enzyme two Arg residues from each subunit interact via vdw and interact with Asp residues by charge-charge interactions.

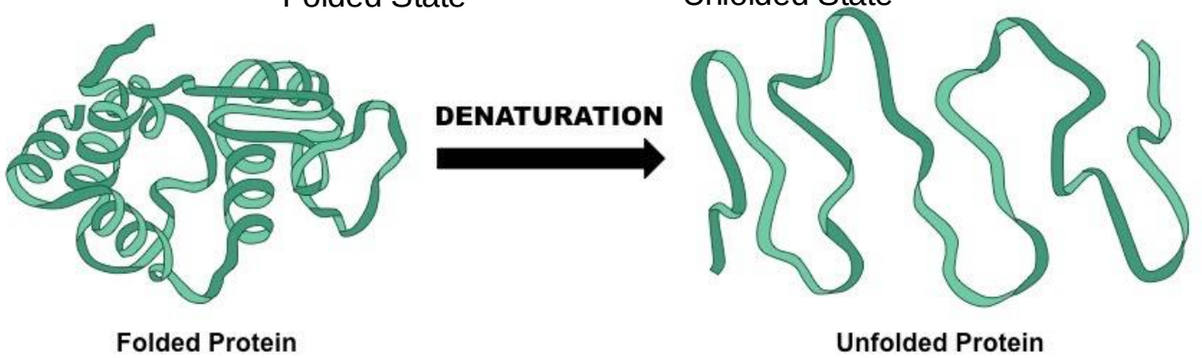


Tertiary Structure – Stability

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.

ΔS^o : Change in disorder of the system. ***positive*** entropy change is ***favorable*** since the ***disorder*** in the system is ***increased***.

ΔS^o is positive for unfolding, favoring the unfolded state.

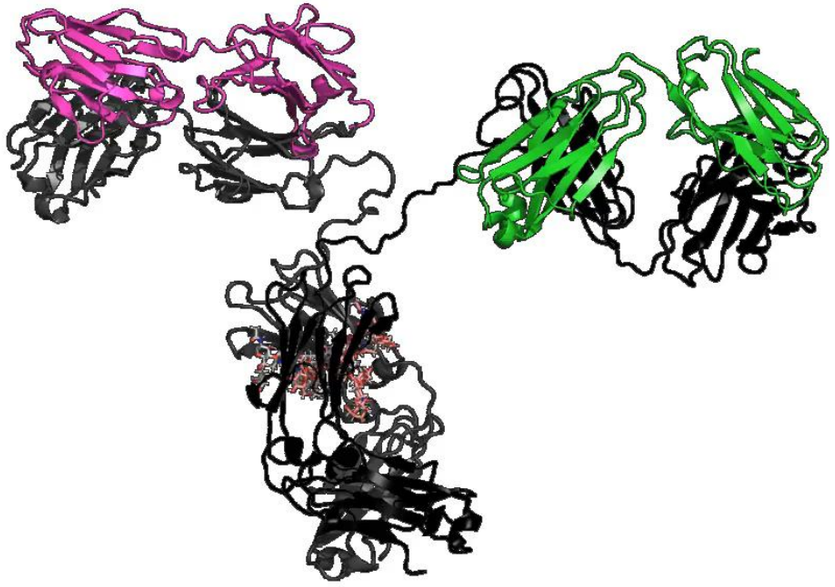
ΔH^o : Enthalpy – A change in the electronic configuration of the system that either releases *heat* ($\Delta H^o < 0$) or absorbs *heat* ($\Delta H^o > 0$). **Release of heat is favorable**.

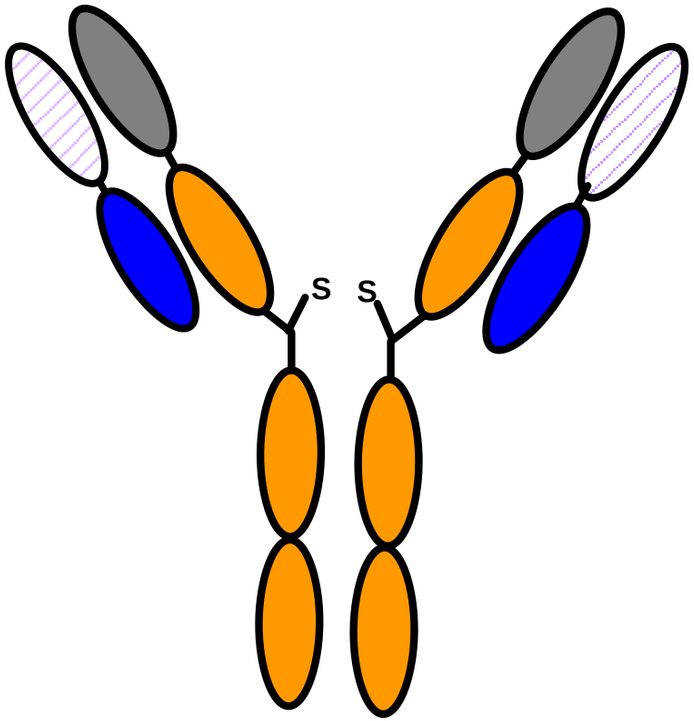
ΔH^o is positive for unfolding, heat needs to be added to unfold proteins

Quaternary Structure

- Combinations of polypeptide subunits (combinations of tertiary structures).
- Chains associate by ***non-covalent interactions*** between R groups on the different chains.
- *Additional* stabilization can be provided by disulfide bonds, usually in extra-cellular proteins.
- Proteins can be a dimer, trimer, tetramer, pentamer, etc.

- If the chains are the same, called homo_____. If chains are different, hetero_____





Quaternary structure of Antibody

- two light chains are identical
- two heavy chains are identical

- heterotetramer held together by non-covalent forces & disulfide bonds

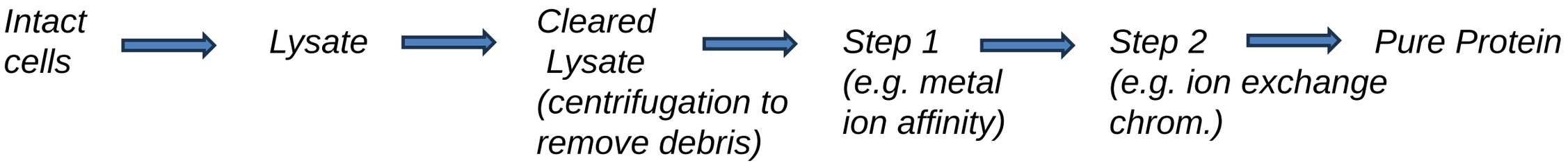
- Antibodies bind two antigens at the top of the Y.

Protein Purification.

Overall Expectations:

1. Describe overall process of column chromatography
2. Understand separation process of:
    a. Size exclusion chromatography (gel filtration) (***Native*** MW)
    b. Cation exchange (+ charged proteins)
    c. Anion exchange (- charged proteins)
    d. Affinity chromatography (ligand, antibody)
    e. Tags (his-tag), genetically encoded.
3. Understand how different components of SDS-PAGE work:
    - Electrophoretic migration: v = q/m
    - Role of SDS: uniform charge to mass ratio -> separation by size only.

Multi-step Purification:



Typical Purification Methods & Their Prevalence

1. Selective precipitation of proteins by ammonium sulfate:
   - Advantage was being able to handle very large volumes of lysate from natural sources
   - Disadvantage is relatively poor resolution (separation of similar proteins from another)
   - Seldom used in the era of recombinant DNA expression of target protein

2. Column Chromatography – separation by size, charge, tags, selective affinity
   - Very scalable (mg to gms)
   - High capacity
   - High resolution (easier to separate similar proteins)

Types of Column Chromatography & Their Frequency of Use

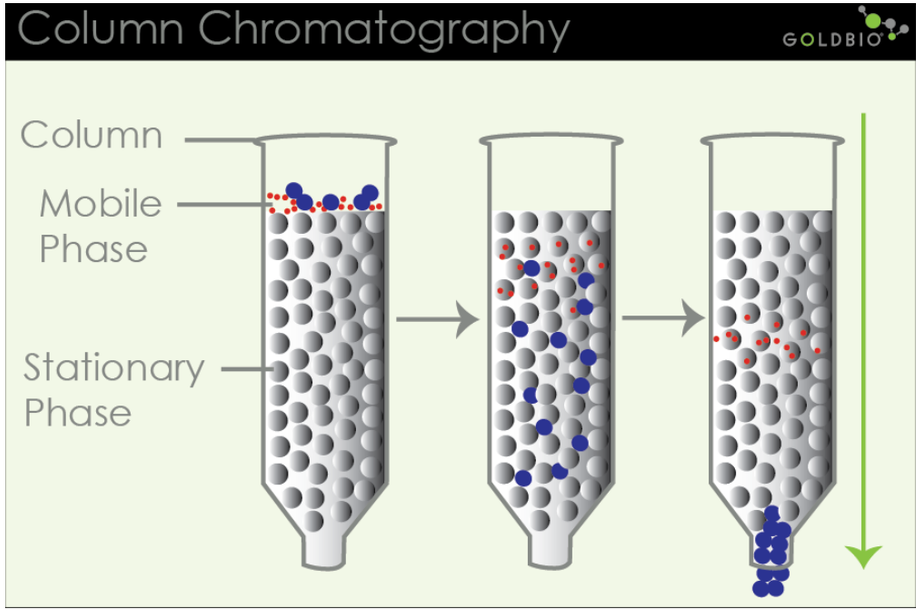
Method	Separation by	Frequency of Use
Metal-ion affinity	Tagged with6 His residues = HisTag	Extremely common (95%)
Maltose and glutathione (GSH) affinity chromatography	Tagged with maltose binding protein or glutathione transferase	Somewhat common (4%)
Ion-exchange chromatography	Charge	Very common 2nd step (70%)
Size exclusion chromatography	Size	Somewhat common 2nd step (30%)
Antibody affinity chromatography	Binding to specific antibody	Relatively rare
Ligand based affinity chromatography	Binding to specific ligand	Targeted to particular classes of enzymes
Hydrophobic affinity chromatography	Polarity of protein surface	Rare

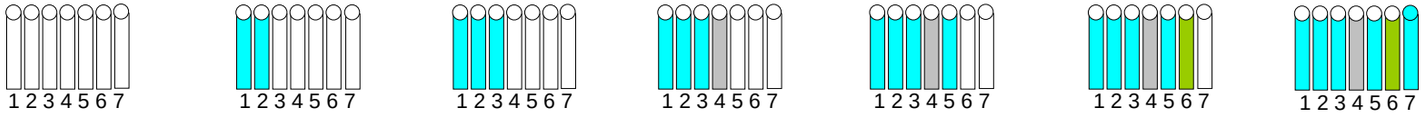
Column Chromatography

- Chromatography is performed in long glass tubes filled with a matrix or resin (particle size similar to fine sand).
- The liquid is a buffered salt solution to keep the proteins in their native (folded) form.
- The mixture of proteins is added to the top of this column and buffer is allowed to flow through the column.
- As the buffer flows through the column the mixture of proteins is drawn down through the column and interacts with the resin. The actual mode of separation depends on the nature of the resin.
- The protein ***elutes*** from the column at a specific volume (V_e) at a specific time.
- The elution volume is divided into separate **fractions** (e.g. separate test-tubes)
- Usually, several different chromatographic steps are performed with different resins, sequentially, during a purification scheme.

Different resins allow separation by:

- Affinity (Metal ions for HisTag)
- Charge (ion exchange)
- Size (size exclusion/gel filtration)





His-Tags – The most common Affinity Tag

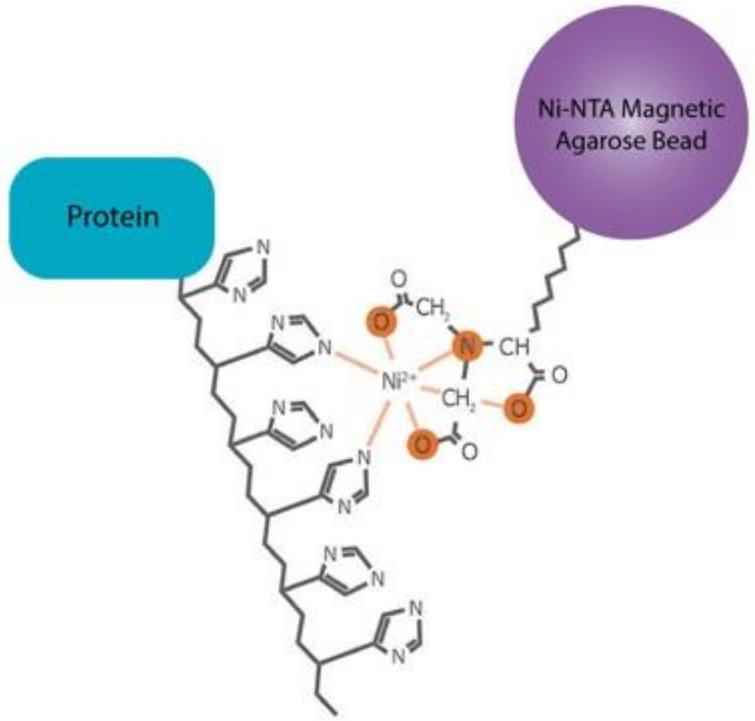
- A His-tag is the addition of 6-10 Histidine residues to either the N-terminus or the C-terminus of the target protein.
- The tag has to be added by changing the DNA sequence of the target protein.
- The His residues will cause the tagged protein to stick to immobilized Ni ions on a resin bead.
- The impurities can be washed away, and the tagged protein released by adding a high concentration of imidazole. The imidazole will compete for the His and bind to the Ni ions.

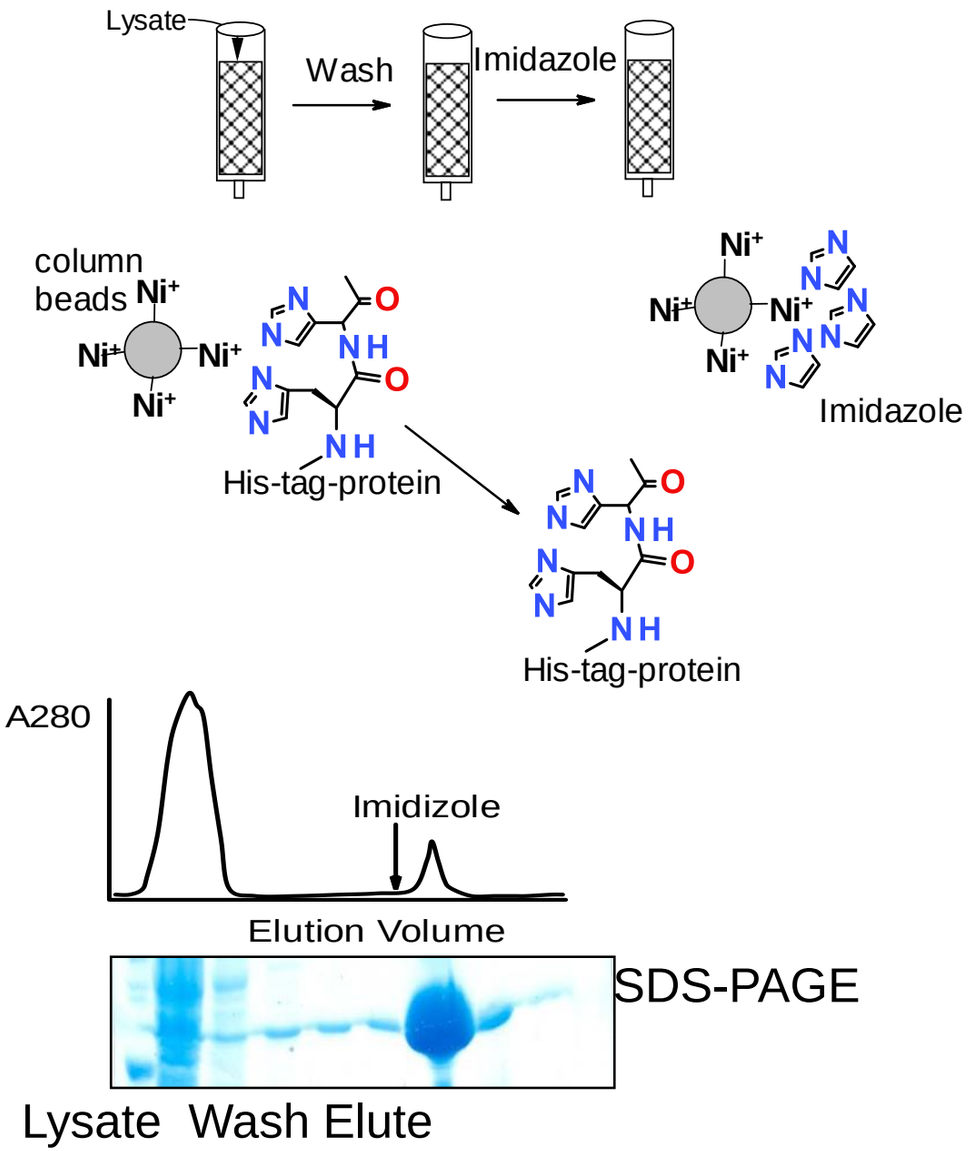
Original DNA/protein sequence:

```
-ATGCCCGGC--
 MetProGly--
```

Modified DNA/tagged protein sequence:

-ATG**CACCACCACCACCACCAC**CCCGGC--
 Met**HisHisHisHisHisHis**ProGly--





Ion Exchange - Effect of pH on Net Charge of Proteins

pI (Isoelectic pH)- pH at which proteins have no *net* charge:

- Proteins become positively charged for pH < pI
- Proteins have a negative charged if pH > pI.

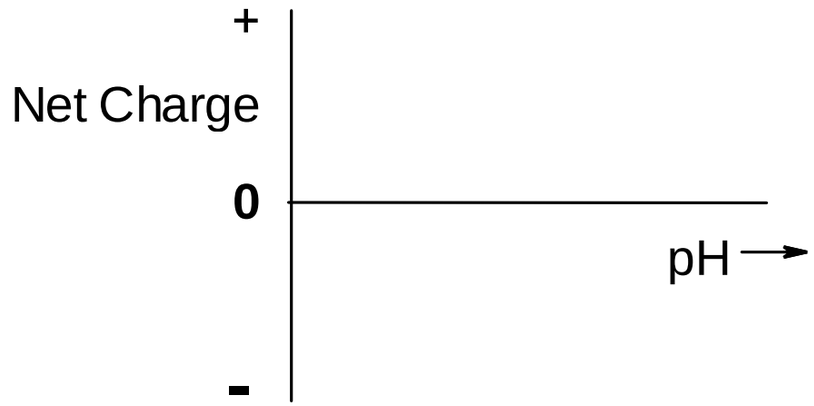
Therefore, the charge on a protein can be changed by altering the pH, increasing or decreasing the binding to ion exchange resins.

The pI can be predicted from the amino acid sequence using the same charge calculation that we used earlier:

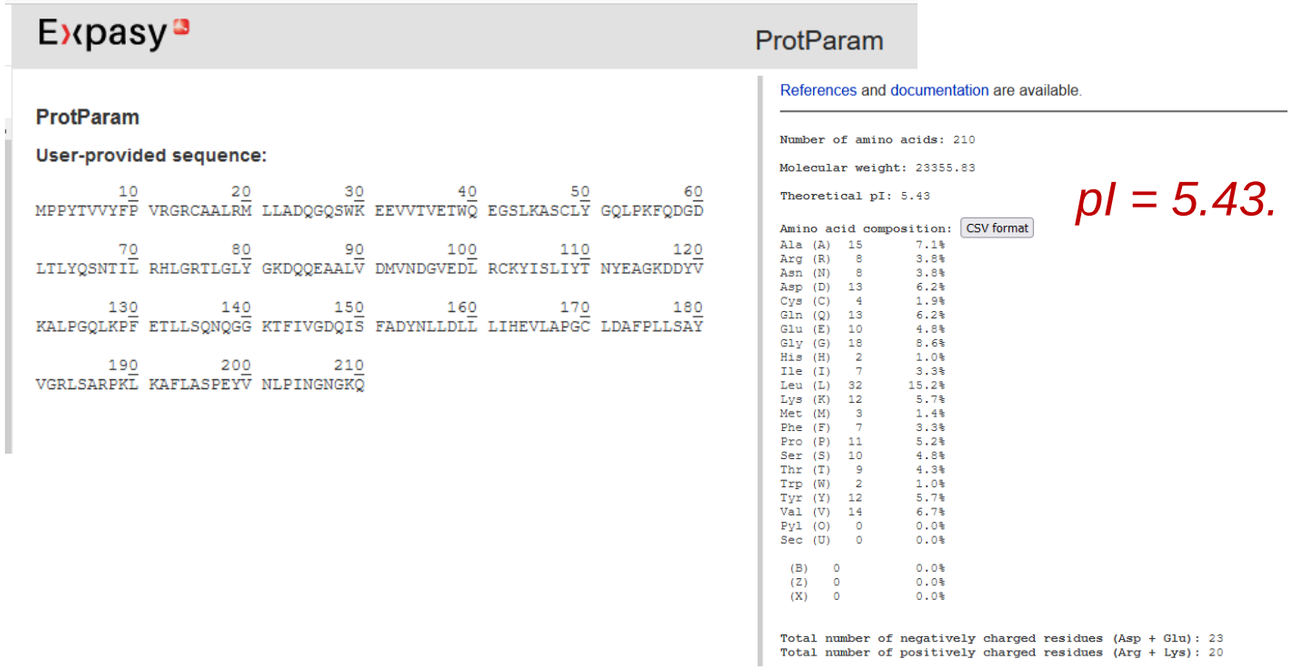
$$\langle q \rangle = \sum_{i=1}^{n} f_A \times q_A + f_{HA} \times q_{HA}$$

and finding the pH value where the charge is zero.

This calculation can be done by web-servers.



https://web.expasy.org/protparam/

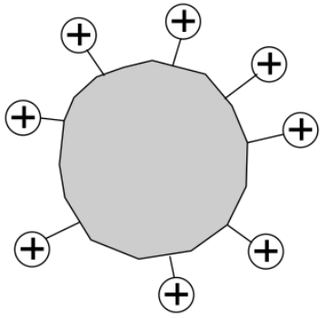
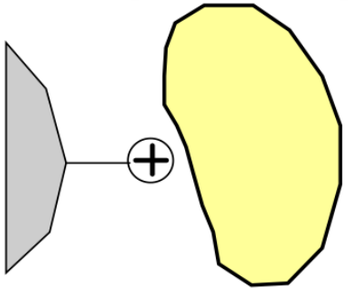
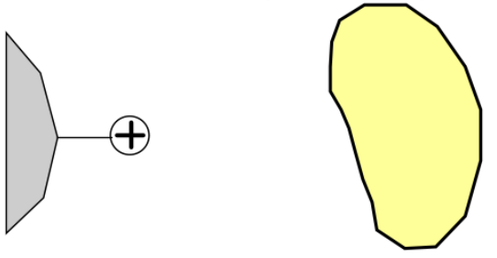
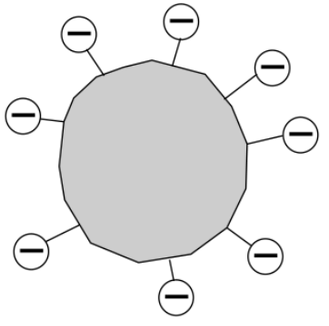
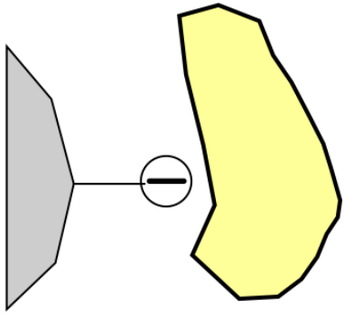
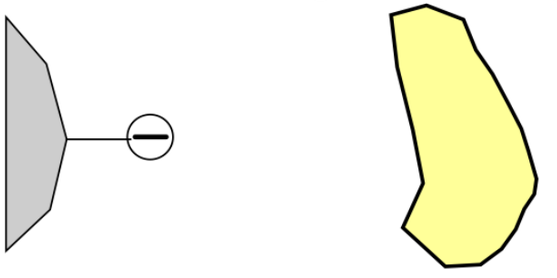


pI = 5.43.

1. *What is the charge on this protein at pH=5.43?*
2. *What is the sign of the charge on this protein at pH=7?*

   negative or positive

Types of Resins – Ion Exchange

Type	Resin	Principal of Separation	How to Elute the Protein
Anion Exchange *Separation by charge.*	• Beads with a **positive** charge 	Protein stick to resin because of: • Overall **negative** charge (*anions*) • Proteins have patches of negative charge 	• Wash, bound proteins **may** elute. • Increase salt concentration to weaken electrostatic interaction. • Change of pH to pH < pI (protein becomes positively charged) 
Cation Exchange *Separation by charge.*	• Beads with a **negative** charge 	Protein stick to resin because of: • Overall **positive** charge (*cations*) • Proteins have patches of positive charge 	• Wash, bound proteins **may** elute. • Increase salt concentration to weaken electrostatic interactions. • Change of pH to pH > pI (protein becomes negatively charged) 

Ion Exchange - Column Chromatography

Cation exchange resin, binds + charge ions.

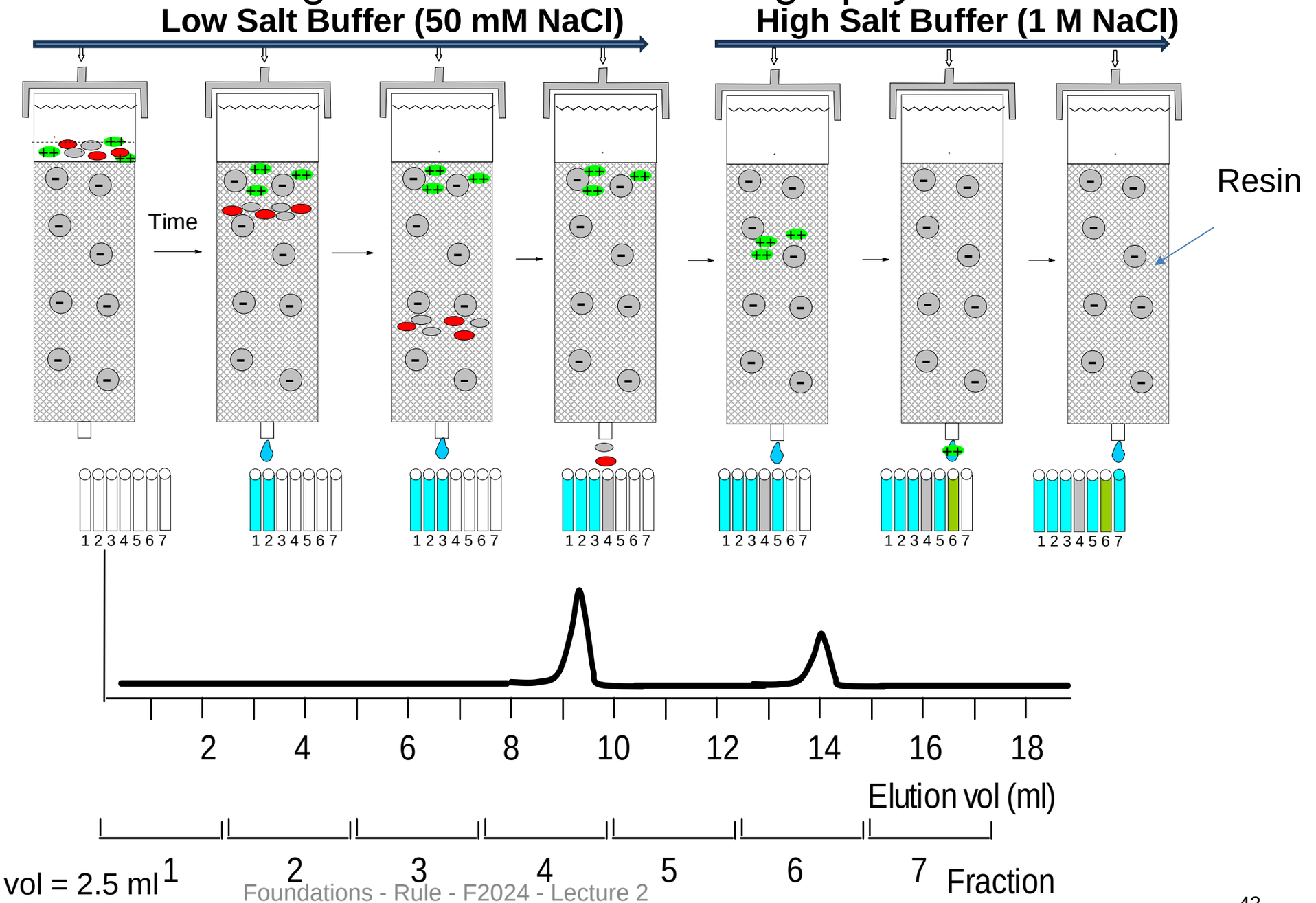
Mixture contains:

A. + Charge (green)
B. Neutral (gray)
C. - Charge (red)

Elution Volume: Volume that has dripped out of the column.

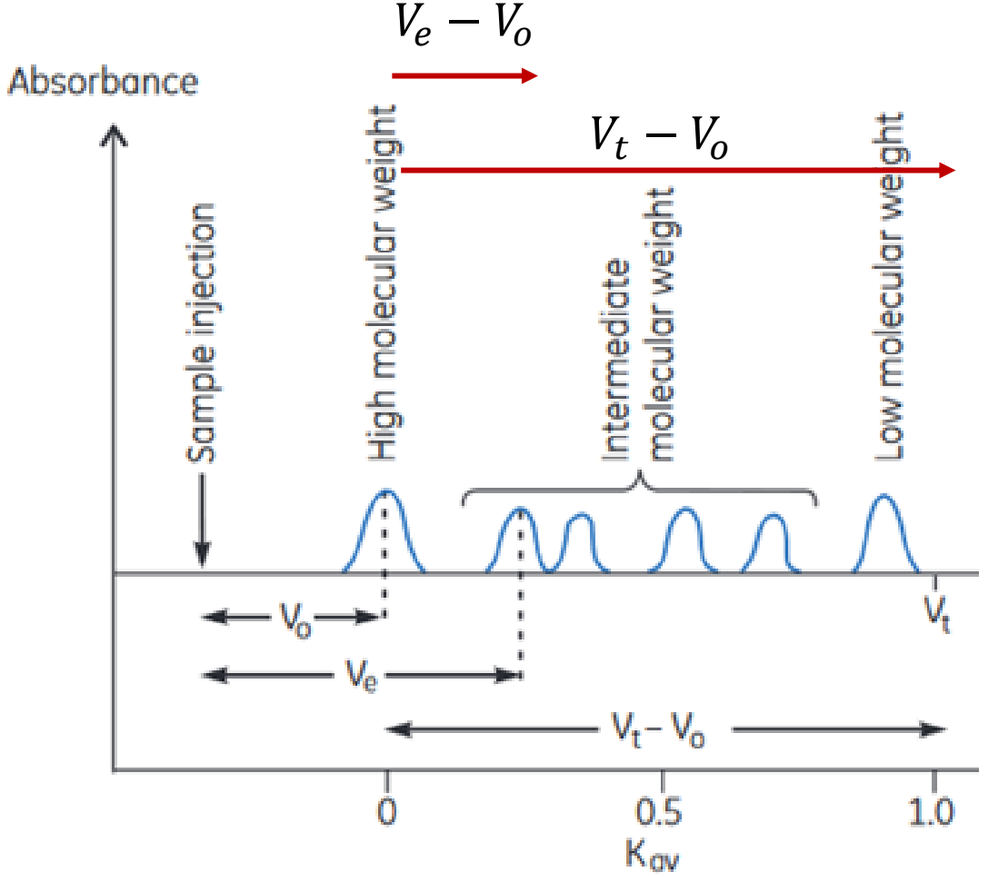
Fraction: Collection of liquid that drips out the bottom of the column – i.e. the total amount of liquid is *fractionated* into separate fractions.

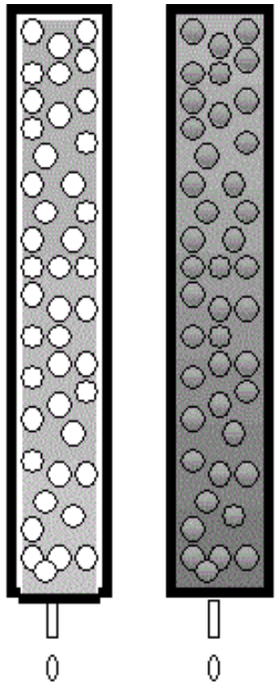
Fractions that contain the protein of interest are pooled (combined) before next step.



Separation by Size - Size Exclusion (aka Gel Filtration)

Size Exclusion (gel filtration) *Separation by size*		• Proteins don't "stick". • Small proteins enter the interior of the beads, and therefore take longer to wash off of the column	• Simply washing the column with buffer will eventually wash the proteins out of the column. • Smaller proteins elute last.





Fractional Elution $K_{AV} = \frac{(V_e - V_o)}{(V_t - V_0)}$

V_e = volume that peak elutes (exits column)

V_t = total column volume, ***accessible to small proteins.***

V_o = void volume, the volume ***outside*** the beads, this amount of liquid has to come through the column before any proteins can elute - ***volume accessible to large proteins***. (Typically ~30% of the total column volume)

V_t - V_0 = volume available for separation

The volume available depends on the MW:
Huge protein – V_o is the available volume, $V_e = V_o$
Tiny protein – V_t is the available volume, $V_e = V_t$

Evaluating Final Purity

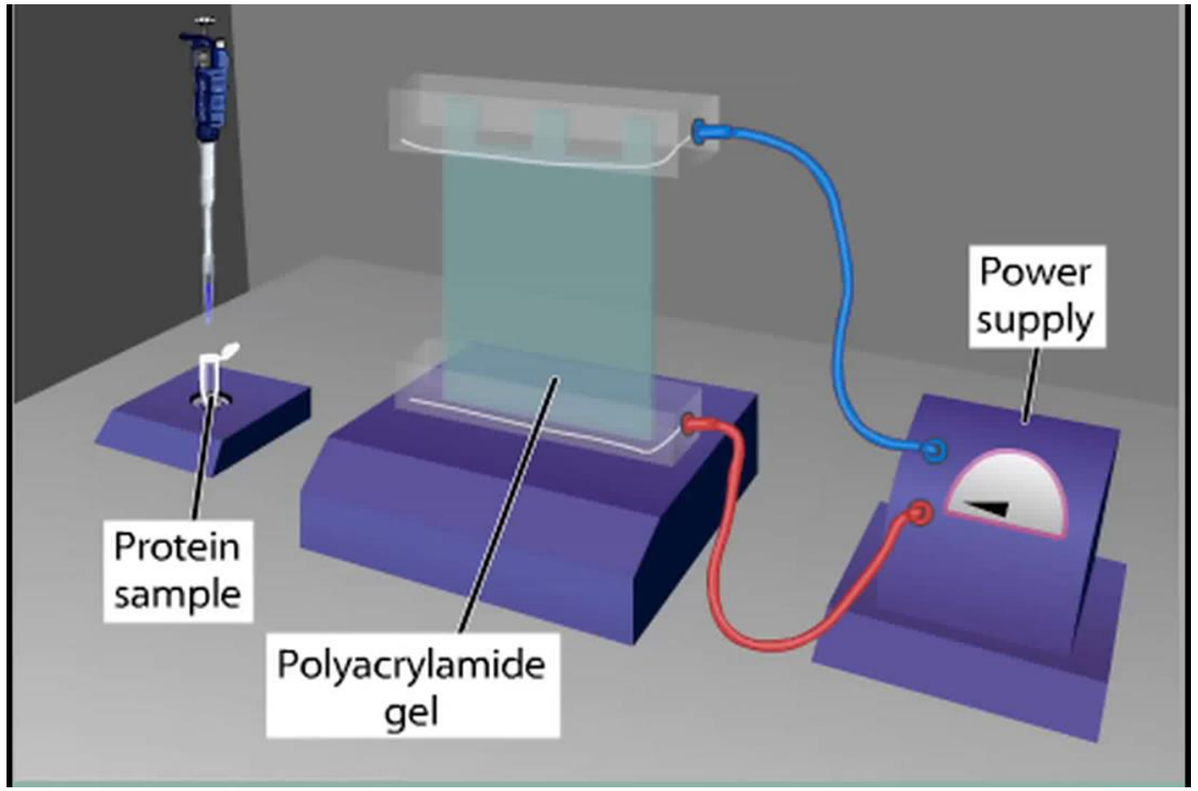
After the protein is purified, its purity can be monitored by:

a) SDS-page gel electrophoresis (most common)
b) Mass spectrometry (sometimes)
c) Amino terminal sequencing (rare)
d) Isoelectric focusing (separation by pI) (rare)

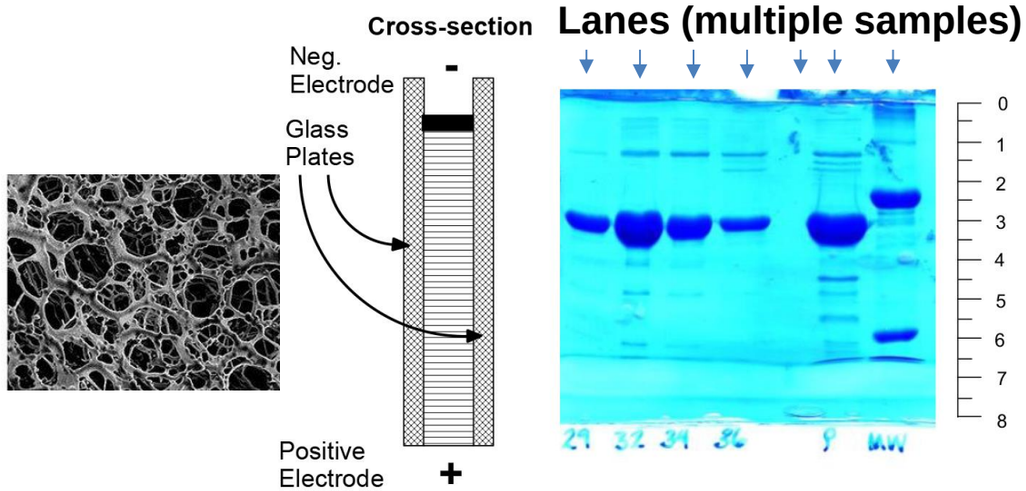
SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Electrophoresis = movement of charged molecules in an electric field: *velocity* $\propto$ *(q/M) x V* (V, volts)

Denatured proteins are pulled through a polymer gel by electrophoresis, separating them solely by size (ideally).



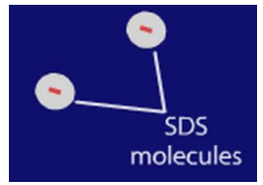
Top View of Gel

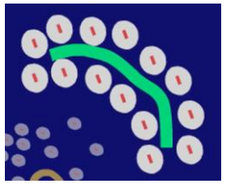


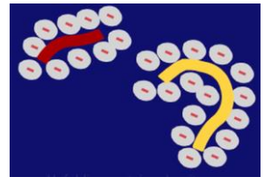
Lanes (multiple samples)

Based on the video:

1. *What happened to all proteins when interacting with SDS?*
2. *What happened to the protein with two chains, after interacting with SDS?*





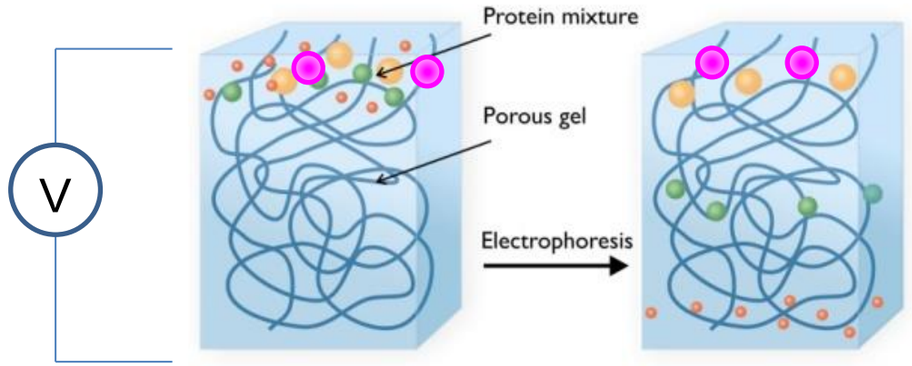


SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

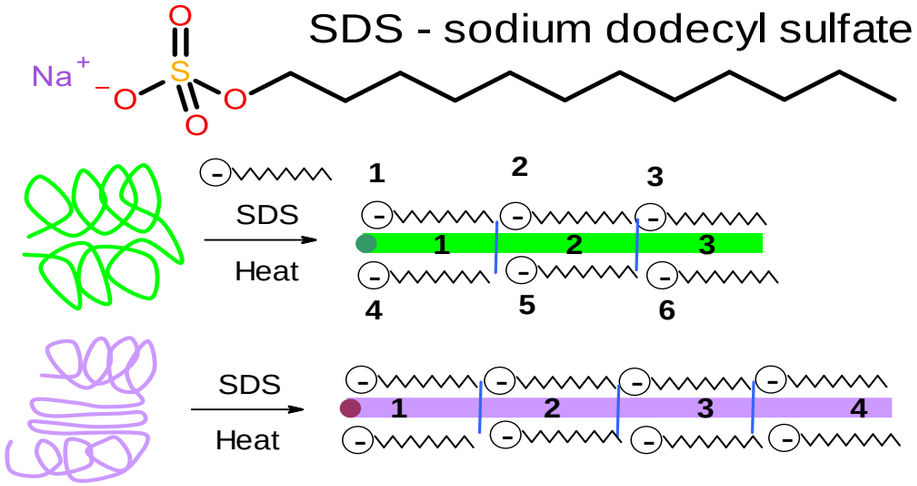
SDS – denatures proteins, giving them:

- a negative charge
- a uniform charge to mass ratio

- All proteins migrate in the same direction.
- Separate only by size as they pass through the gel (Smaller proteins move faster)

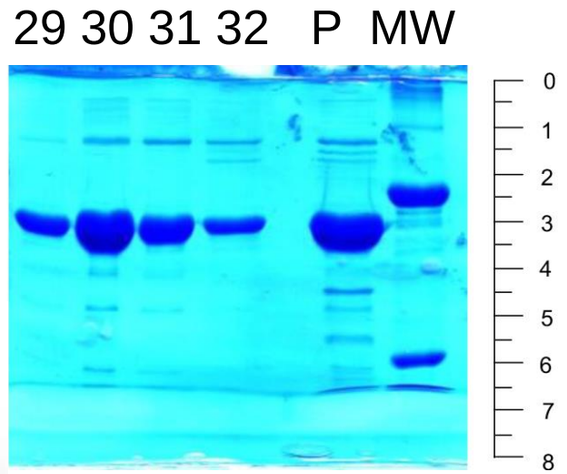


Yellow, green, orange + SDS = - charge
Magenta + SDS = - charge



Charge (q)	Mass (M)	v = q/M
6	3	2 = 6/3

Gels are stained with a stain that is specific for protein. Most commonly used stain is Coomassie blue. Gel shows four fractions (#29, #30, #31, #32) obtained from chromatography, an empty lane, pooled fractions (P), molecular weight standards (MW).



Measuring Denatured Molecular Weight With SDS-PAGE.

A. Molecular Weight Determination – distance migrated ~ ∝ log(MW).

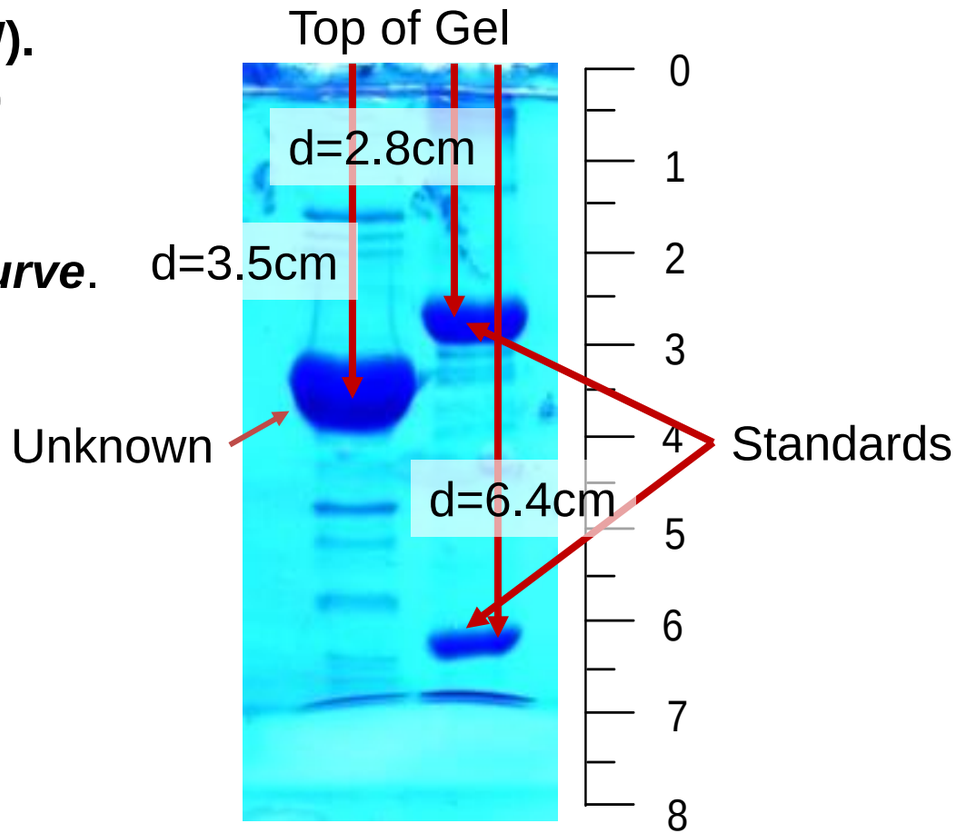
i) Load molecular weight standards on one lane (right lane in above gel)

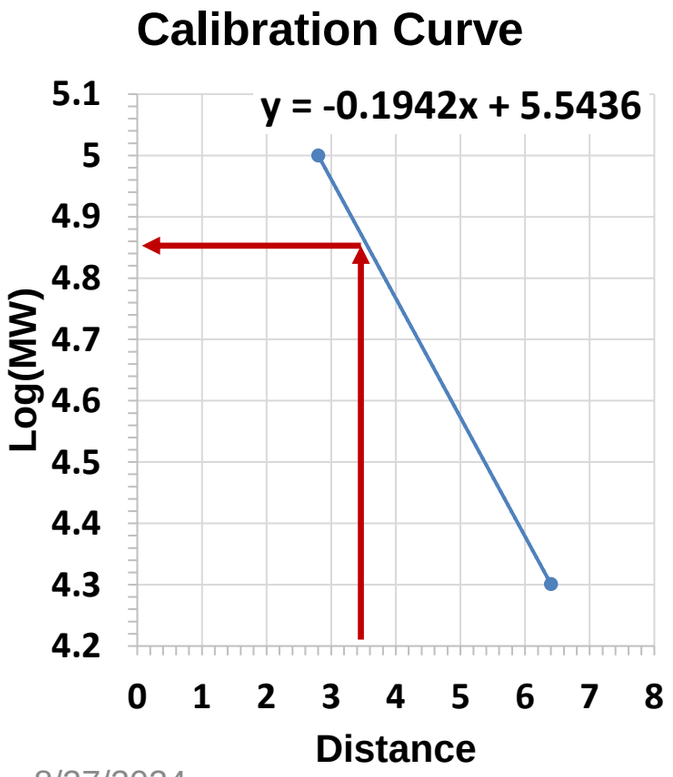
ii) load unknown(s) in other lane (left lane), apply voltage.

iii) Measure the distance migrated of each band.

iv) Plot log(MW) versus *d* for standards, generating a ***calibration curve***.
This is fit to a polynomial function e.g. $log(MW) = \alpha d^2 + \beta d + \gamma$.

v) Use calibration curve to get logMW of unknown.





Obtaining MW

1. Analytical Solution (better)

Log(MW) = -0.1942 x d + 5.5436

d = 3.5 for unknown

Log(MW) = 4.86

MW = $10^{4.86}$ = ~ 73,100 Da

2. Graphical solution

i. Sketch calibration curve using known MW (standards)

ii. Find intersection of line for d=3.5

iii. Read log(MW) from y-axis

Standards

S1: MW 100,000, distance = 2.8 cm

S2: MW 20,000, distance =6.4

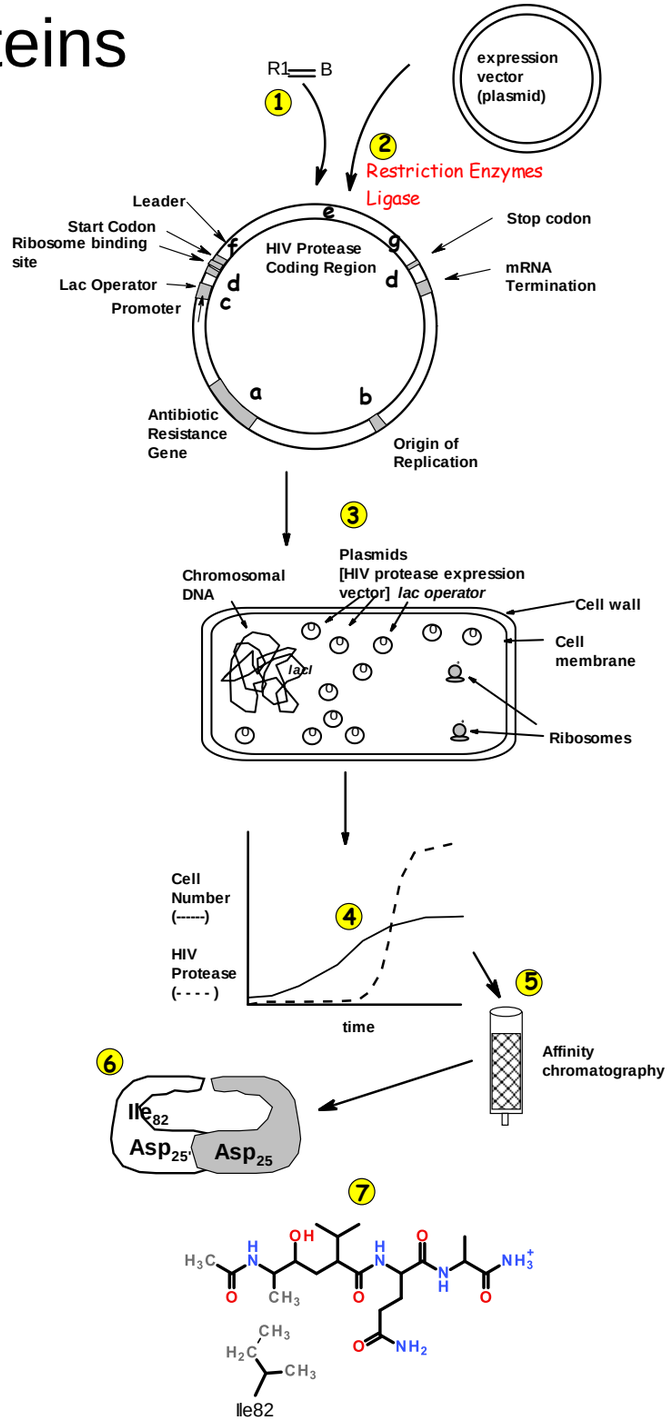
Unknown: Distance =3.5

Production of Recombinant Proteins

The Goal: To produce any desired protein in in E. coli (a widely used bacteria) using recombinant DNA methods.

The overall procedure is as follows:

1) Obtain the codons for the desired protein:
   - PCR from native source
   - Chemical synthesis of gene.
2) Insert the codons into a bacterial expression vector (a specialized form of a DNA plasmid) using restriction endonucleases and DNA ligase.
3) Transformation of the bacteria with the plasmid.
4) Growth of the transformed bacteria, production of the recombinant protein.
5) Purification of the recombinant protein.
6) Structure determination
7) Biochemical studies, e.g. rational drug design, enzyme mechanism



Properties of Expression Vectors

Common Required Features:

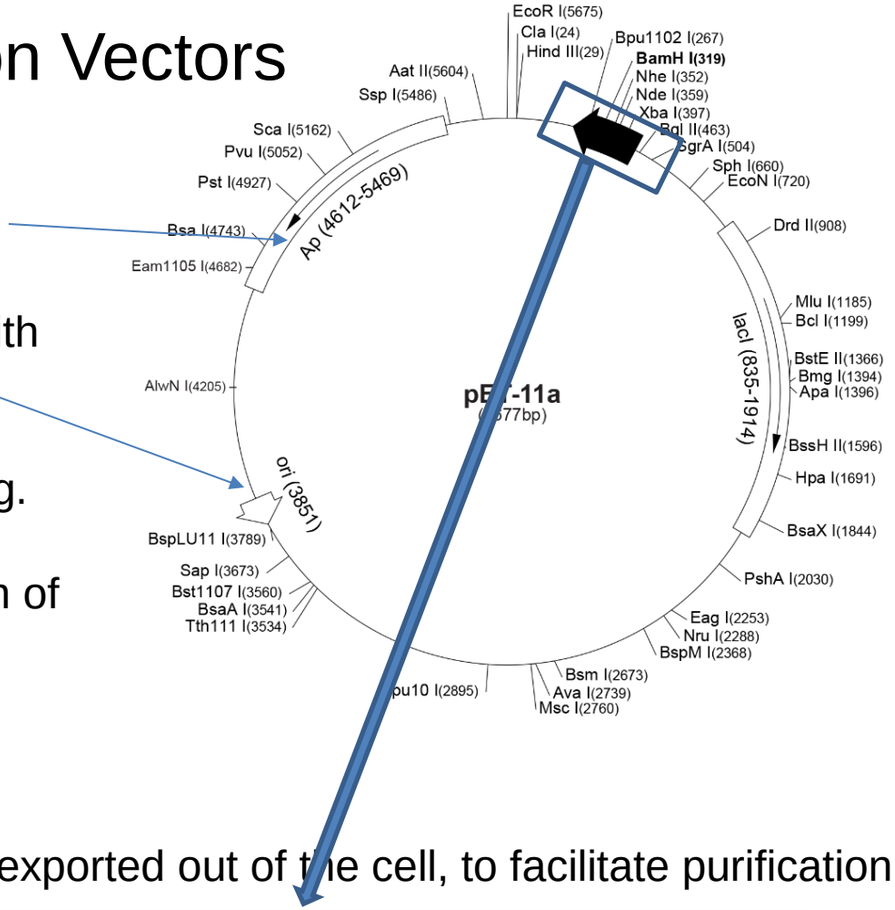
a) Provide antibiotic resistance to the host bacteria, such that only bacteria containing the plasmid will grow in the presence of the antibiotic.

b) An origin of DNA replication so that the plasmid will be replicated with the bacterial DNA

c) DNA sequences that cause the production of mRNA, copying the information in the DNA to mRNA, including a regulated promotor (e.g. lac operator).

d) Sequences in the resultant mRNA that start and stop the production of the recombinant protein.

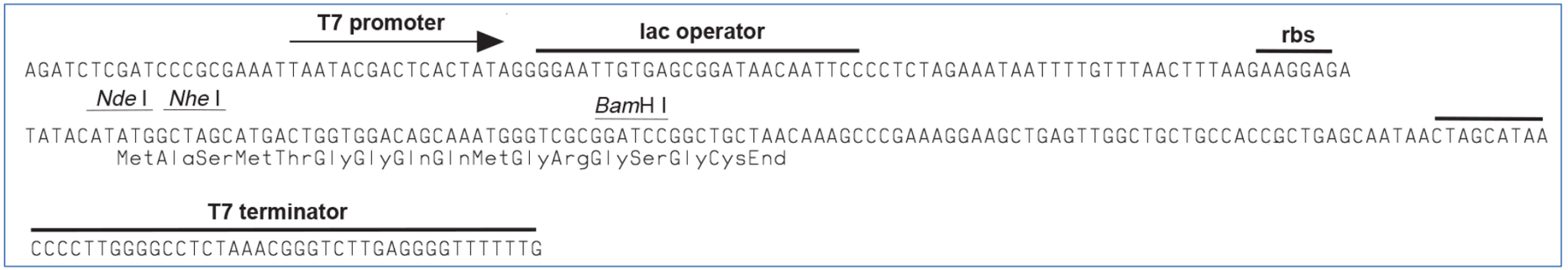
e) Coding region for protein to be expressed.



Optional features on some vectors:

a) His6 tag for affinity purification on a Ni column.

b) Sequences (pelB leader) that cause the recombinant protein to be exported out of the cell, to facilitate purification.

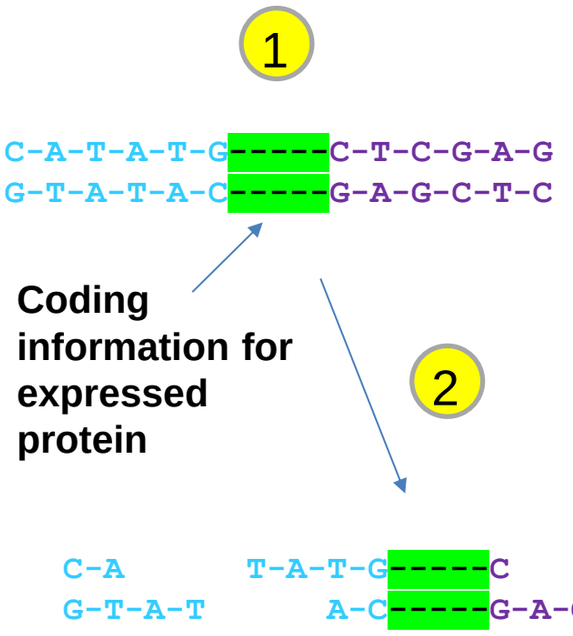


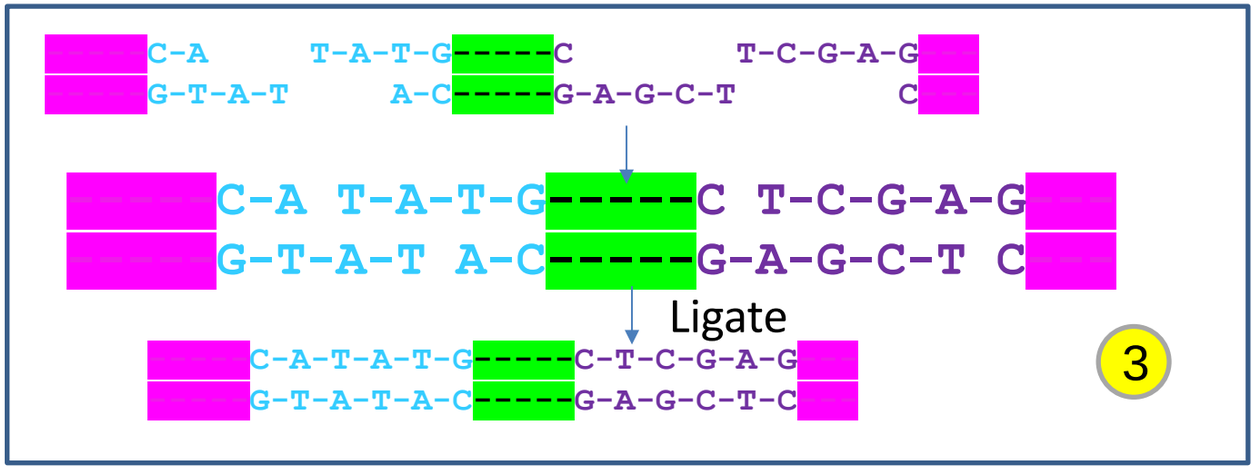
Gene for protein to be expressed inserted by restriction digest of insert & plasmid, mix & ligate.

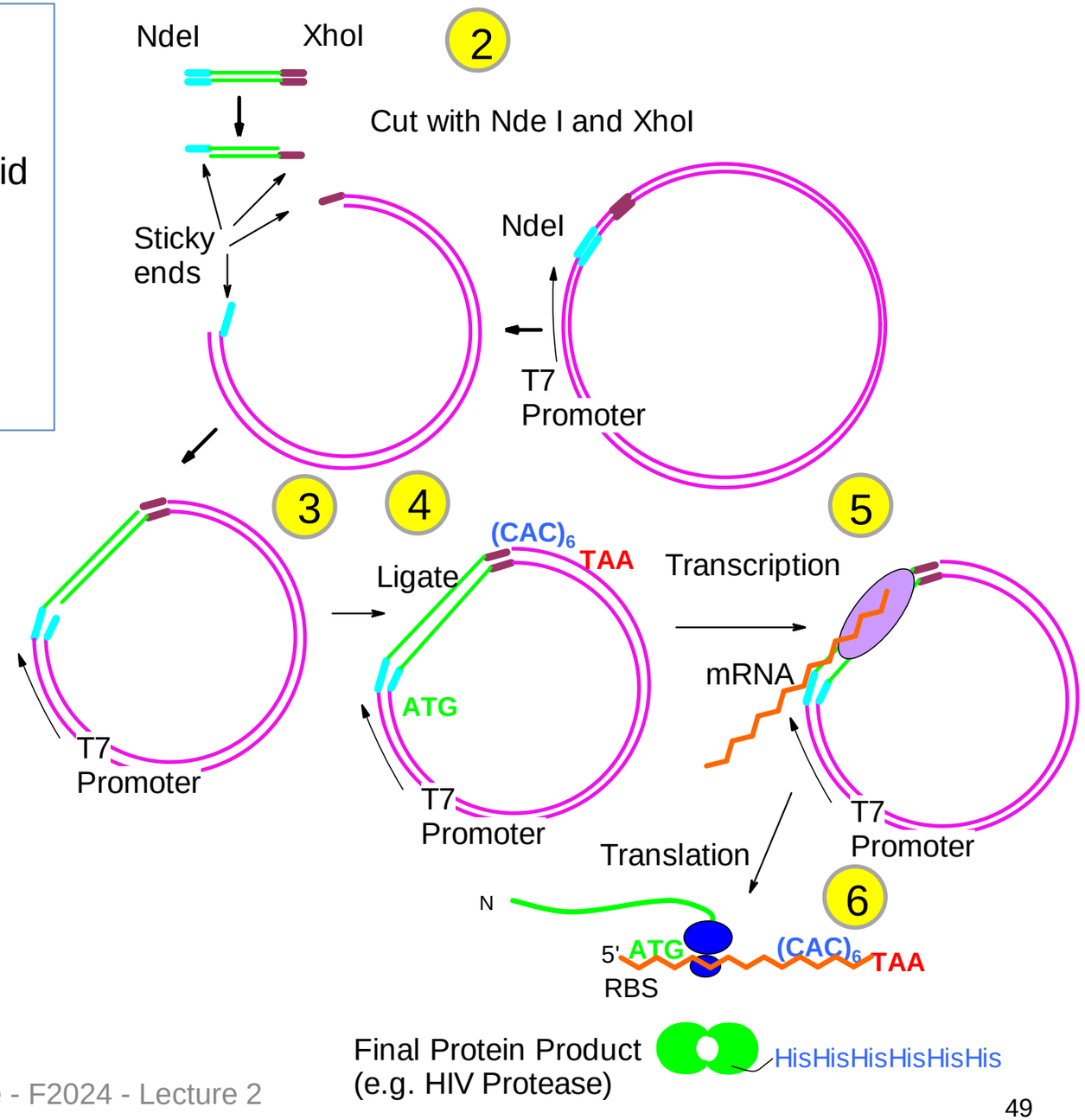
Insertion of DNA into Expression Plasmid

STEPS:

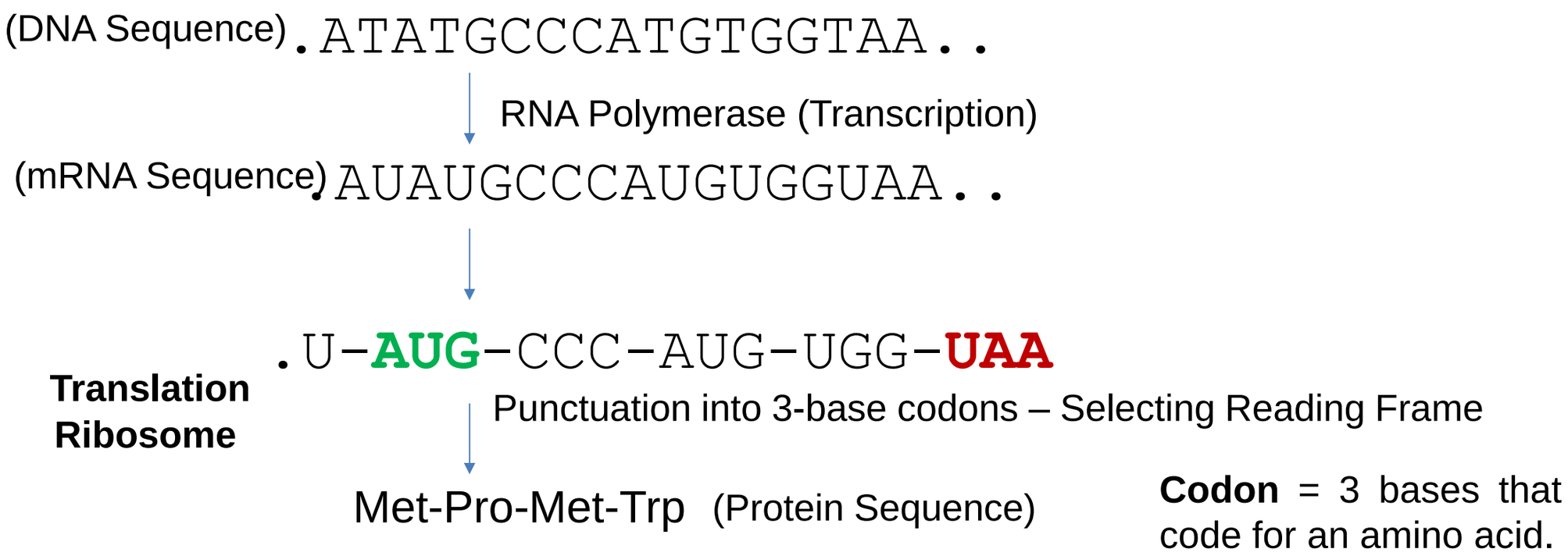
1. Obtain DNA segment that codes for desired protein
2. Cut DNA segment & plasmid with restriction enzymes
3. Ligate (join DNA together)
4. Transform
5. Transcription
6. Translation







Information Transfer In Biology



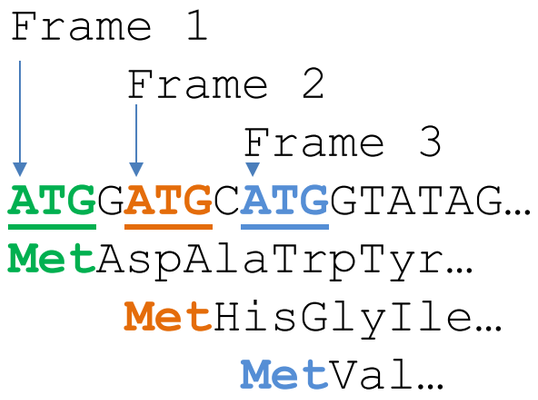
Codon = 3 bases that code for an amino acid.

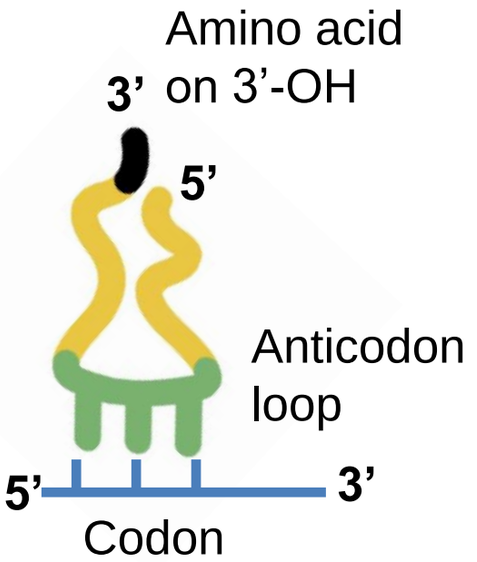
DNA/RNA sequence to Amino Acid Sequence – Codon Table

5' Base	Middle Base				3'
	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Term	Term	A
	Leu	Ser	Term	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

The codon TGG = ____?

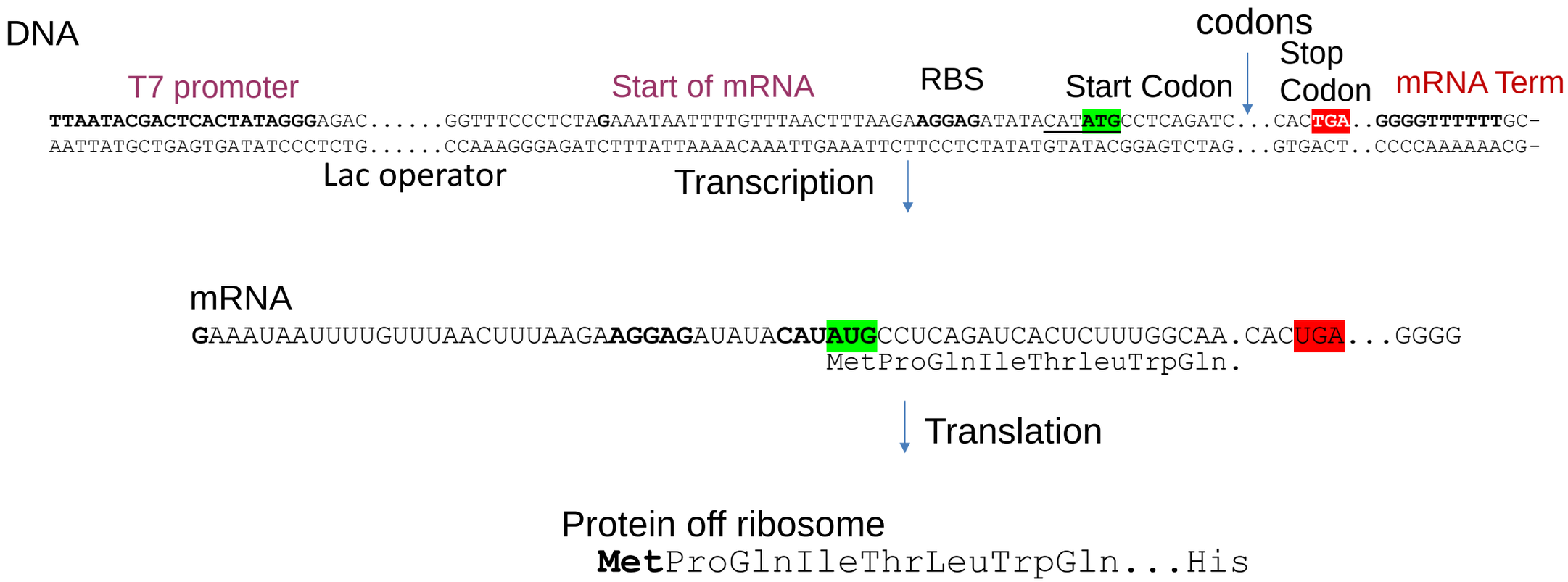
Start codon selection sets the correct reading frame





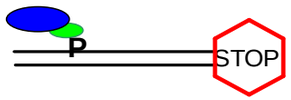
Expectations:
- Concept of reading frame
- Converting codons to AA

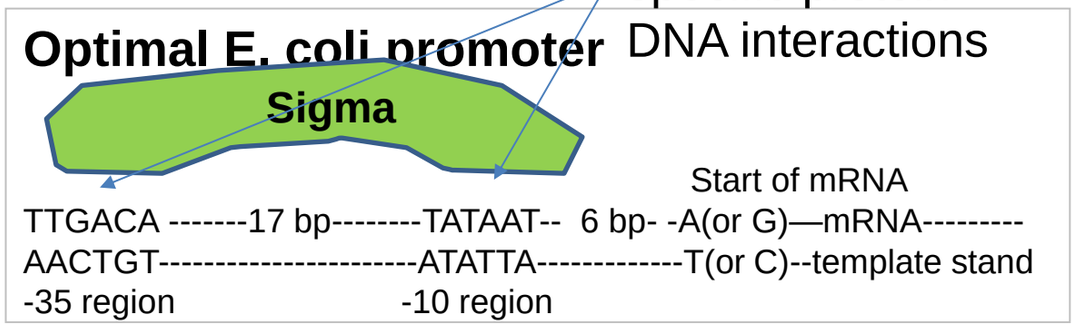
mRNA Synthesis (Transcription) – Important Signals



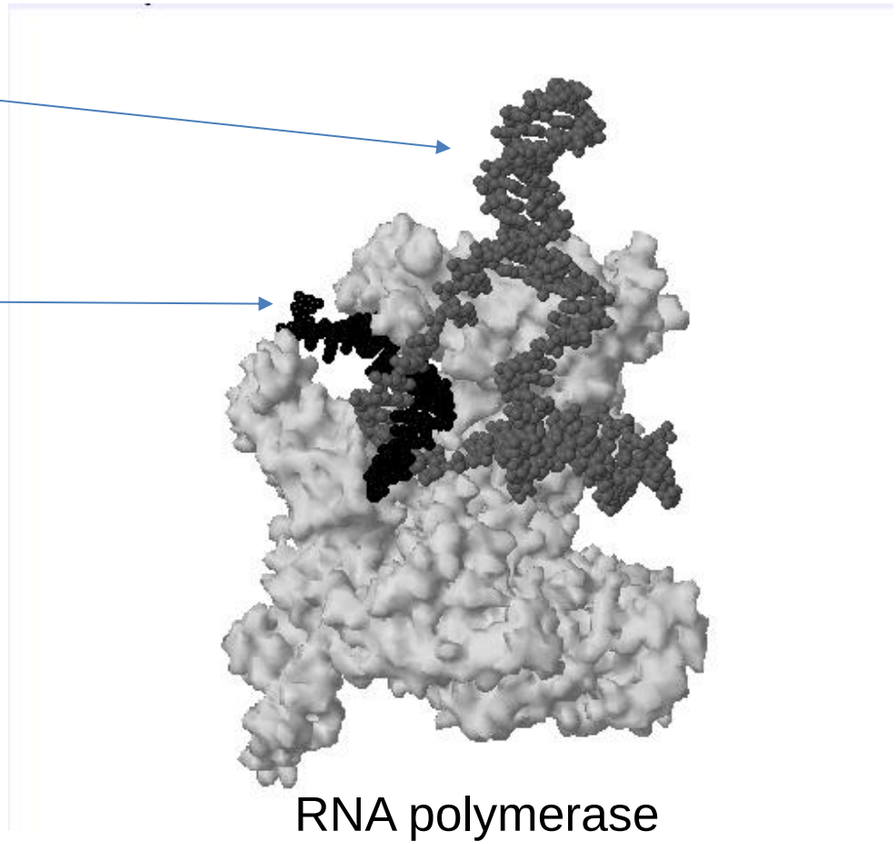
E. Coli RNA Polymerase:

- Holoenzyme: $\sigma + \alpha_2\beta\beta'$
- Binds to promoter (P) sequence in a base specific manner **via σ subunit**.
- Core polymerase = $\alpha_2\beta\beta'$ – sufficient for elongation (sigma factor falls off core enzyme)





-35 = 35 bases from start site



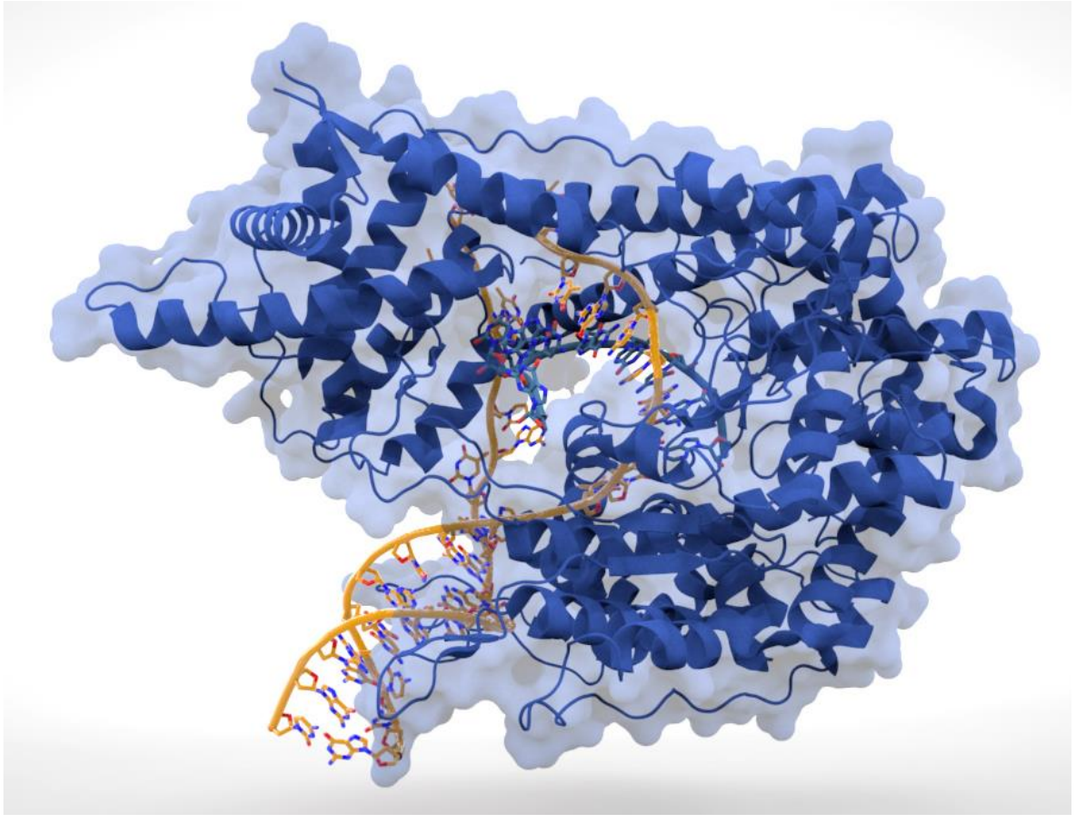
T7 RNA Polymerase

- Isolated from a bacterial virus called T7
- A single chain polymerase
- Recognizes a different promoter sequence – will only transcribe genes with a T7 promoter – selective synthesis of one gene (the one on the plasmid)

T7 is a more robust RNA polymerase than E. coli, seldom stalls during transcription – gives more mRNA and therefore more protein. It is typically used when expressing recombinant proteins from plasmids.

T7 Promoter: TAATACGACTCACTATAGGGAGA

E. Coli Prom: TTGACA -------17 bp----------TATAAT-



Protein Synthesis

(DNA Sequence) .ATATGCCCATGTGGTAA..

RNA Polymerase (Transcription)

(mRNA Sequence) .AUAUGCCCAUGUGGUAA..

Translation
Ribosome

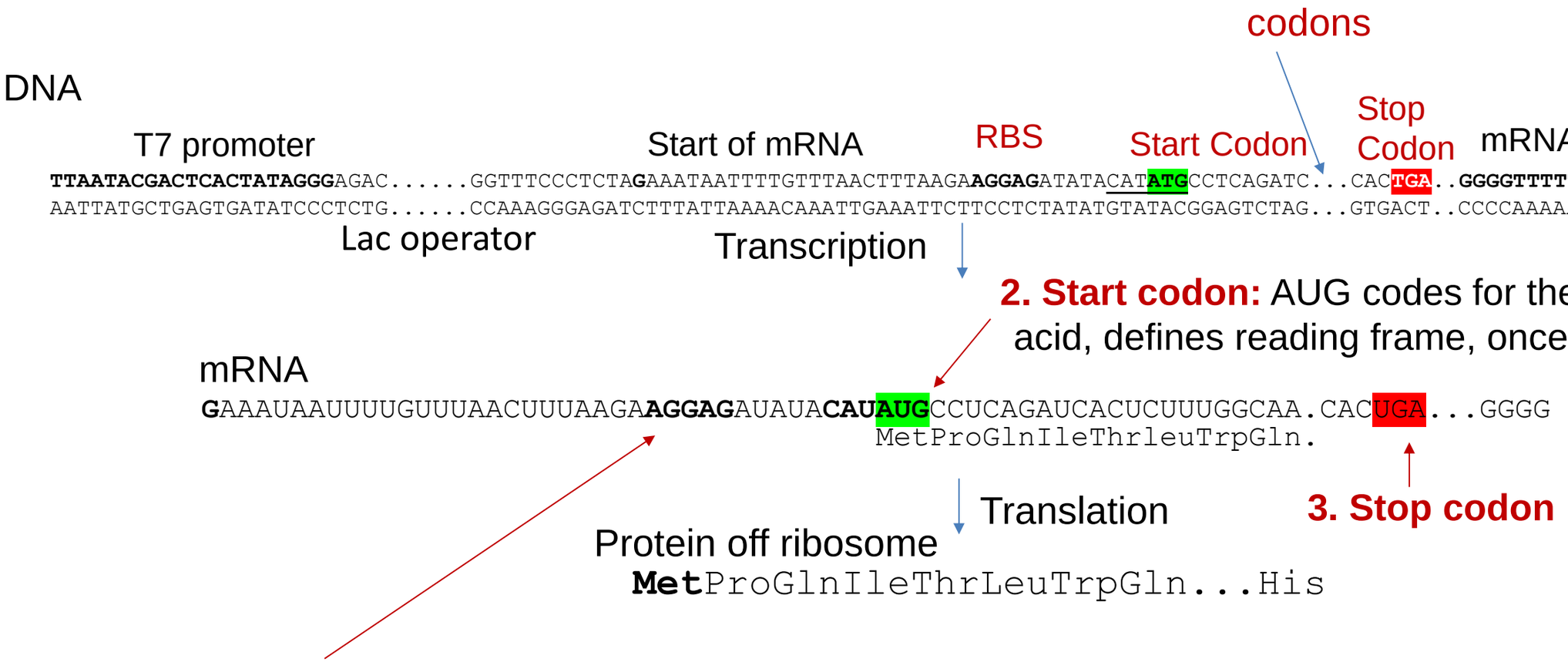
.U–**AUG**–CCC–AUG–UGG–**UAA**

Punctuation into 3-base codons – Selecting Reading Frame

Met-Pro-Met-Trp (Protein Sequence)

Codon = 3 bases that code for an amino acid.

Protein Synthesis – Important Signals



1. Ribosome Binding Site (RBS): (Shine-Dalgarno [SD] sequence)[Prok only]

- *Positions mRNA on the ribosome so that the correct start codon is used.*
- The optimal spacing between the SD sequence and the AUG is 6-9 bases.

Regulation of Transcription by Lac Repressor

The constitutive expression of high levels of almost any protein is toxic to the bacteria:

- Protein may be toxic (e.g protease)
- Production completes for valuable cellular resources.

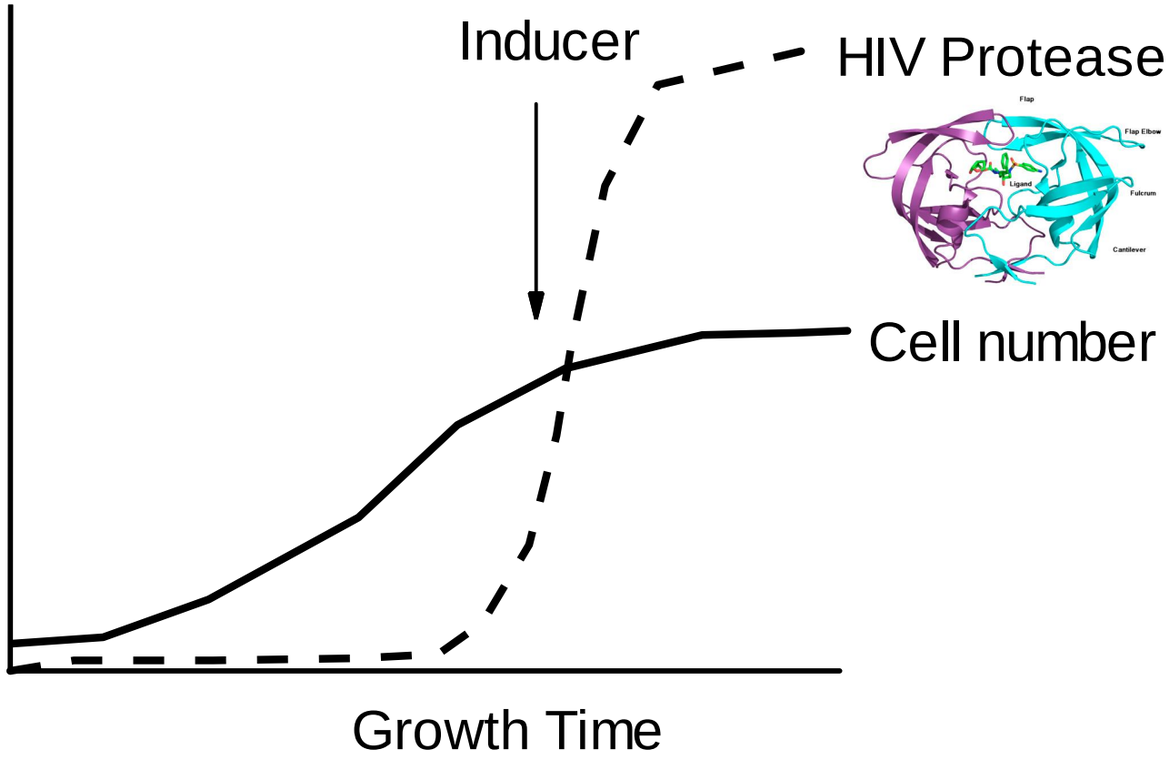
Cell death can occur due to metabolic stress.

An ON/OFF switch is essential.

Expression of the recombinant protein occurs when desired, i.e. when sufficient cell growth has been attained.

The most common ON/Off switch is the **lac repressor-operator** system.

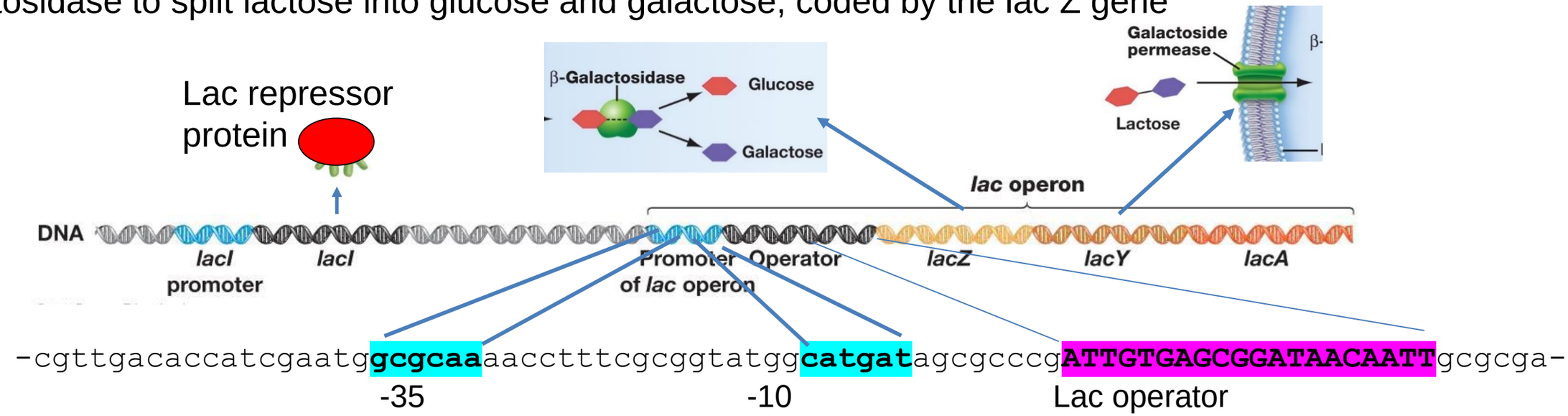
The addition of an **inducer** (IPTG) will start the production of the protein on the plasmid



Transcriptional Regulation: The *lac* Operon – A Natural ON/OFF Switch

Lactose metabolism in bacteria requires 2 proteins – produced from the lac operon (operon is a cluster of protein coding sequences that are produced from a single mRNA, generated from a single promoter)

- Galactoside permease to allow the lactose to enter the cell, coded by the lac Y gene
- Galactosidase to split lactose into glucose and galactose, coded by the lac Z gene

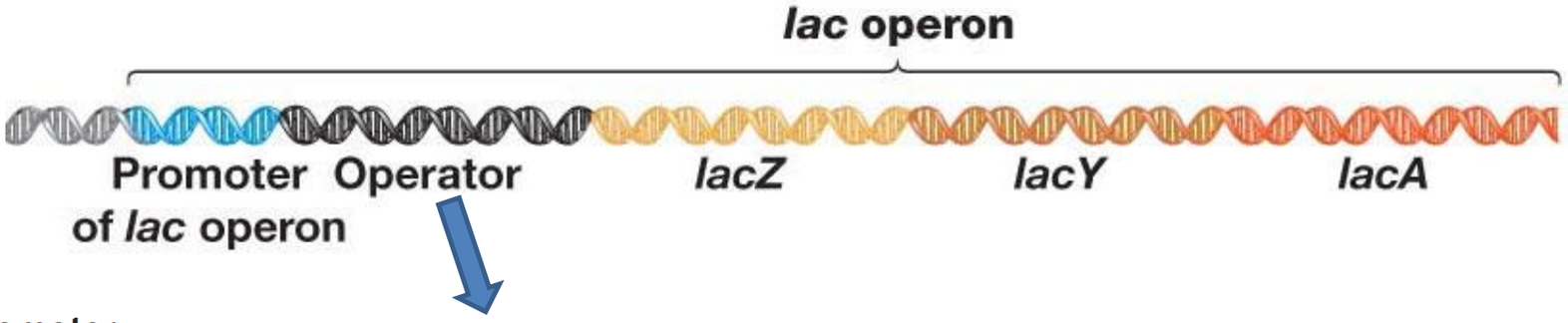


- The production of these proteins is regulated by the **lac repressor**.
- The repressor is a protein that binds to a specific DNA sequence (**lac operator**) and prevents transcription of lacZ and lac Y by blocking E. coli RNA polymerase from moving off the promoter.
- The lac repressor is produced from a separate **lac I gene (not part of the operon)**, which has its own promoter.
- The lac repressor is made all of the time.

Why? The costly synthesis of enzymes that are required for lactose metabolism are only made when lactose is present in the growth media.

How do we use this control machinery?

Place the lac operator sequence between the T7 promotor and the gene we want to control.



Bgl II — T7 promoter — lac operator — *Xba* I — rbs

```
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA
```

Nde I — HIV Protease

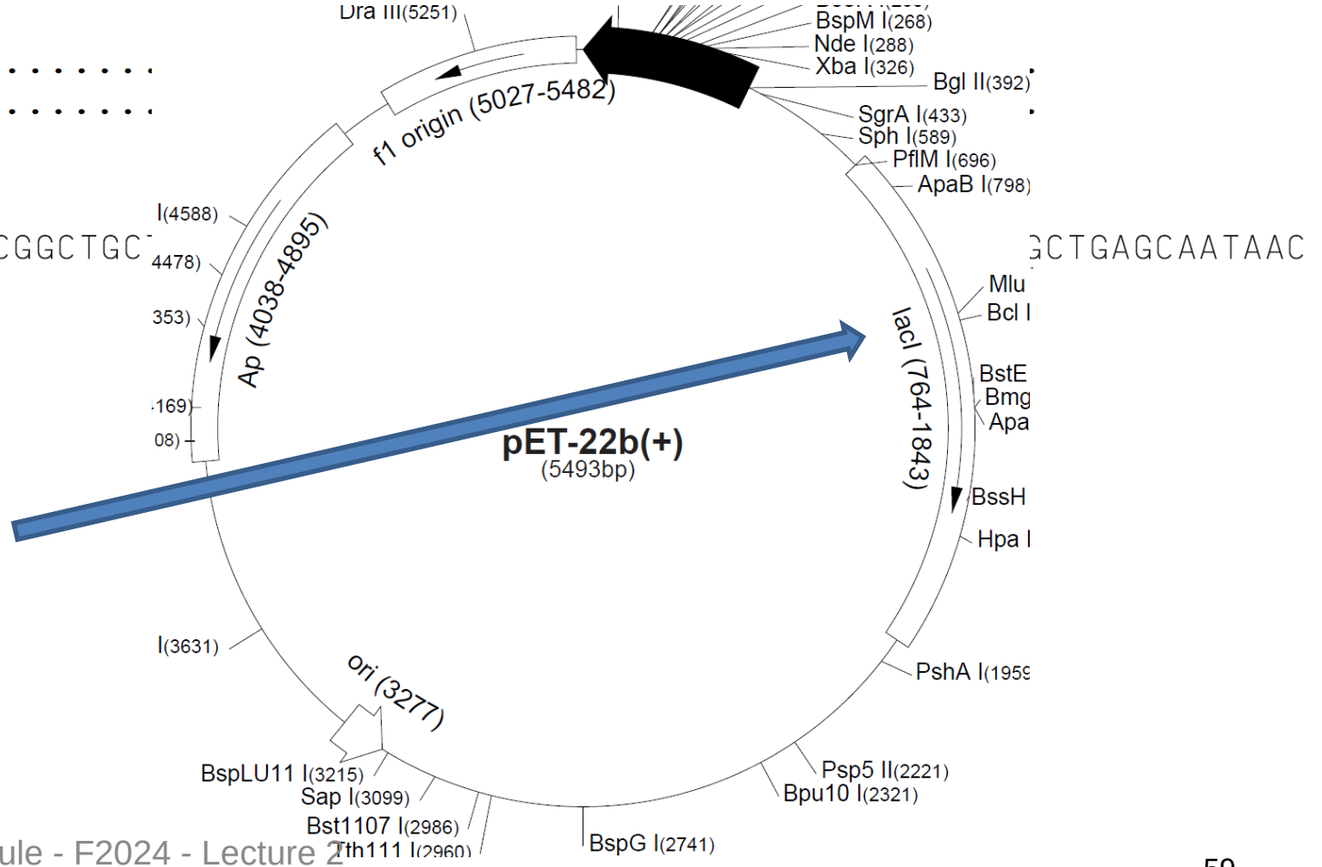
```
TATACATATG CCTCAGATCACTCTTTGGCAA.....................
      Met ProGlnIleThrLeuTrpGln.....................
```

Xho I — His•Tag

```
......TACTTTAAATTTC CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGC
..........LeuAsnPhe LeuGluHisHisHisHisHisHisEnd
```

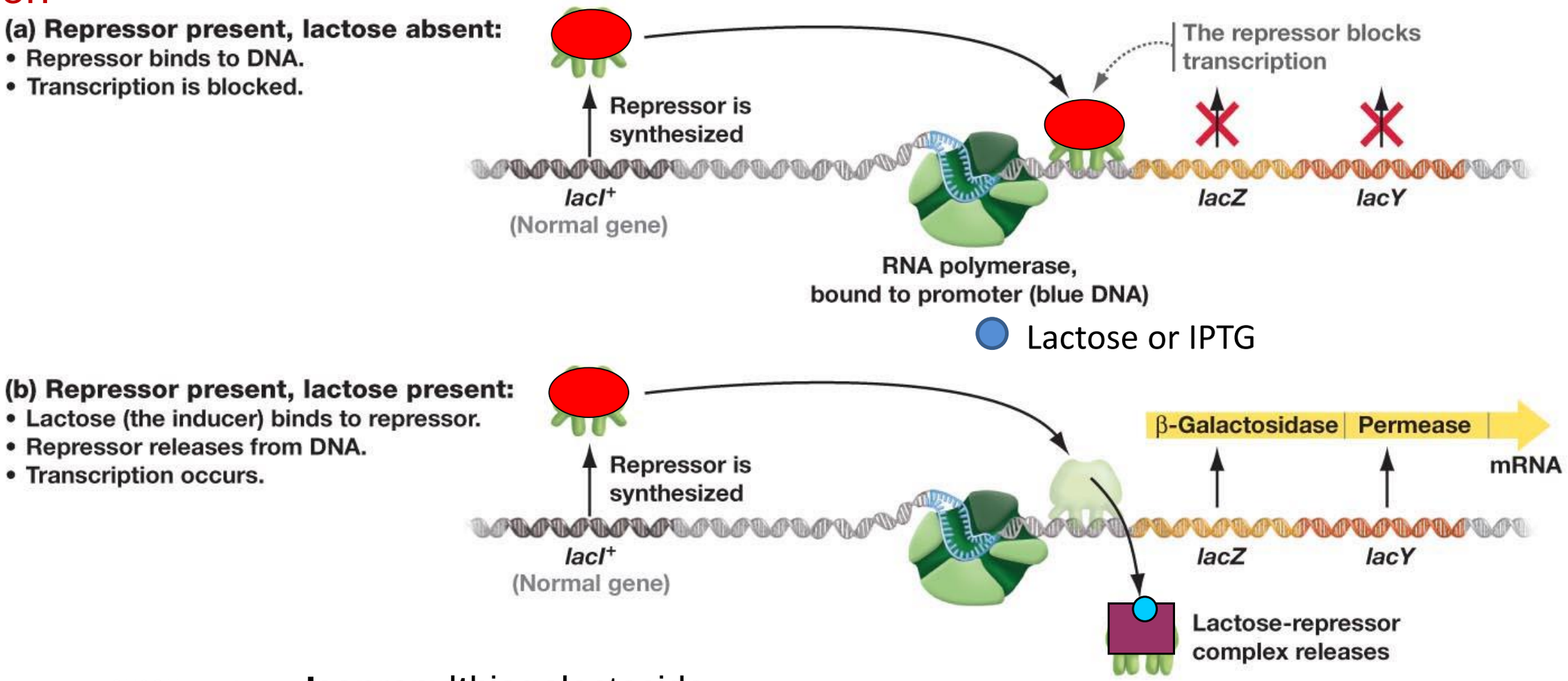
T7 terminator

```
TAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
```

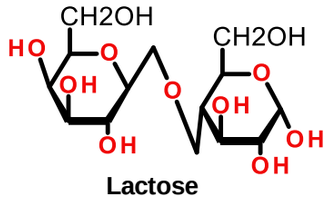


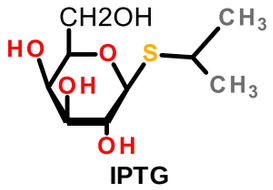
Since there are multiple copies of the plasmid, and not enough lac repressor to go around, the lac I gene is also placed on the plasmid.

The *lac* Operon



Lactose is the Inducer in nature





IPTG

Isopropyl**thio**-**g**alactoside (IPTG) is used in the lab.

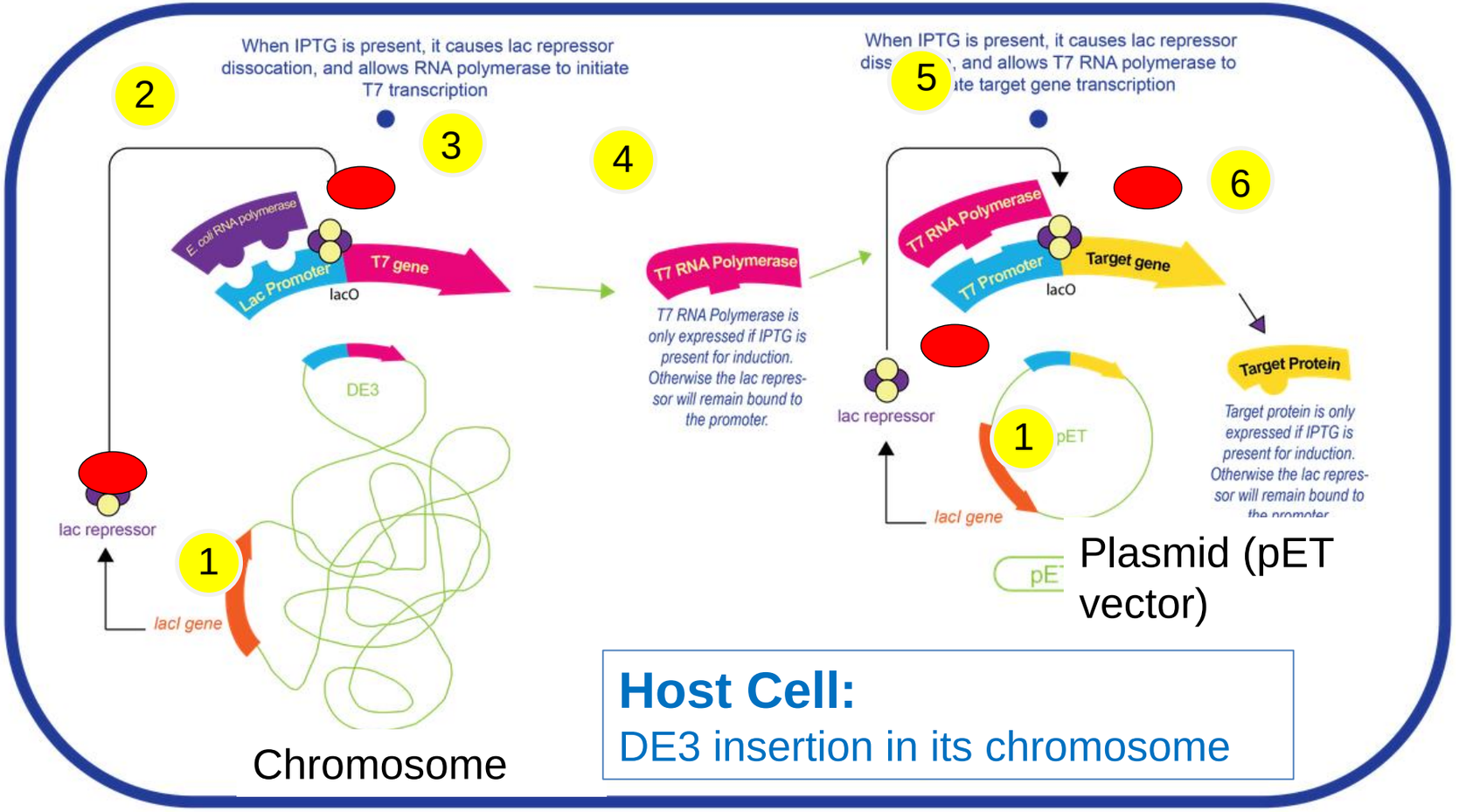
It is a non-hydrolysable inducer that will not be degraded by the cell.

IPTG (and lactose) induce an allosteric change in the lac repressor protein.
R (high DNA affinity) → T (low DNA affinity)

Expression of Recombinant Proteins utilizing the T7 Promotor/Lac Operon

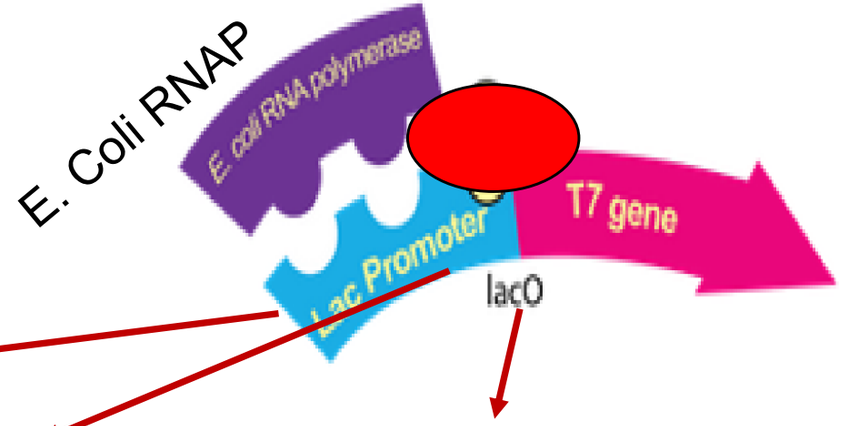
- **T7 RNA polymerase gene** – on the **chromosome** and under control of lac promoter and repressor.
- **T7 promoter** – on **plasmid**, upstream from the gene to be expressed. Usually regulated by lac (shown here).

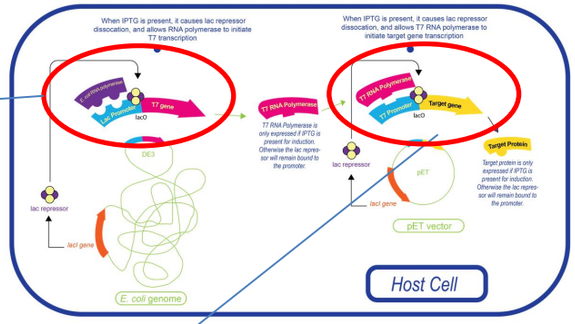
1. The lac repressor is always produced from the lacI gene (chromosome) and from the lacI gene on the plasmid.
2. The lac repressor is initially bound to operator sites, on the chromosome (and the plasmid if regulated)
3. Addition of IPTG causes lac repressor to leave lac operator, upstream from the T7 gene on the chromosome.
4. E. coli RNA polymerase now makes T7 RNA polymerase, using a standard E. coli promoter.
5. T7 RNA polymerase binds to the T7 promoter on the plasmid.



6. T7 RNA polymerase transcribes gene on plasmid, using the T7 promoter. HIV protease mRNA is produced, which is then used by the ribosome to make the mutant HIV protease.

DE3 Sequence Element – Chromosomal: Synthesis of T7 RNA polymerase under lac control

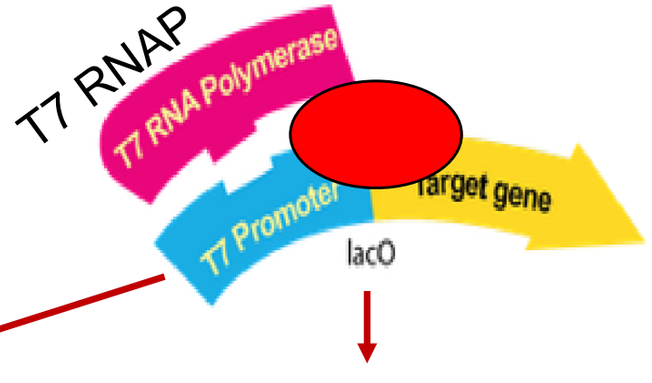




AAA**TTGACA**xxxxxxxxxxxxxxxx**TATAAT**GG**GGAATTGTGAGCGGATAACAATTCC**….**AAGGAG**ATATACAT**ATG**..T7 RNA Poly…

|-35 -E.coli Lac Promoter -10 | |-------Lac operator----------| RBS

T7 – Plasmid: Synthesis of the target protein by T7 RNA polymerase. Usually under lac control

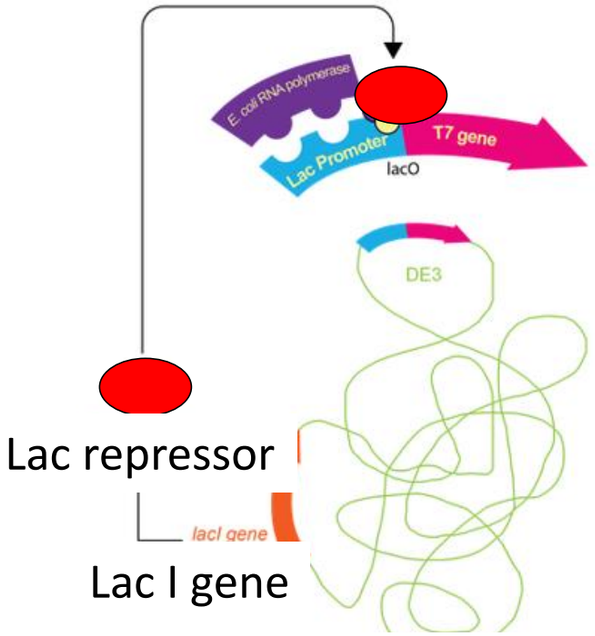


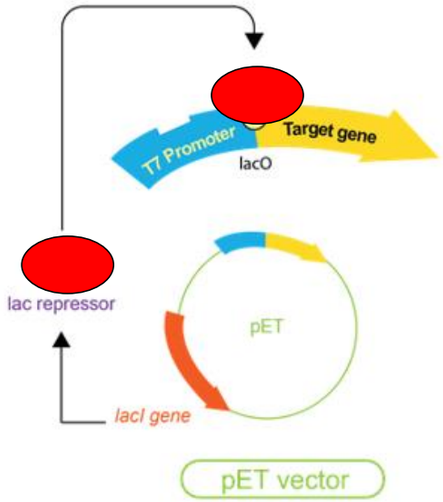
AAA**TTAATACGACTCACTATA**GG**GGAATTGTGAGCGGATAACAATTCC**….**AAGGAG**ATATACAT**ATG**..HIV Prot…

|---T7 promoter--------| |-------Lac operator----------| RBS

No Inducer

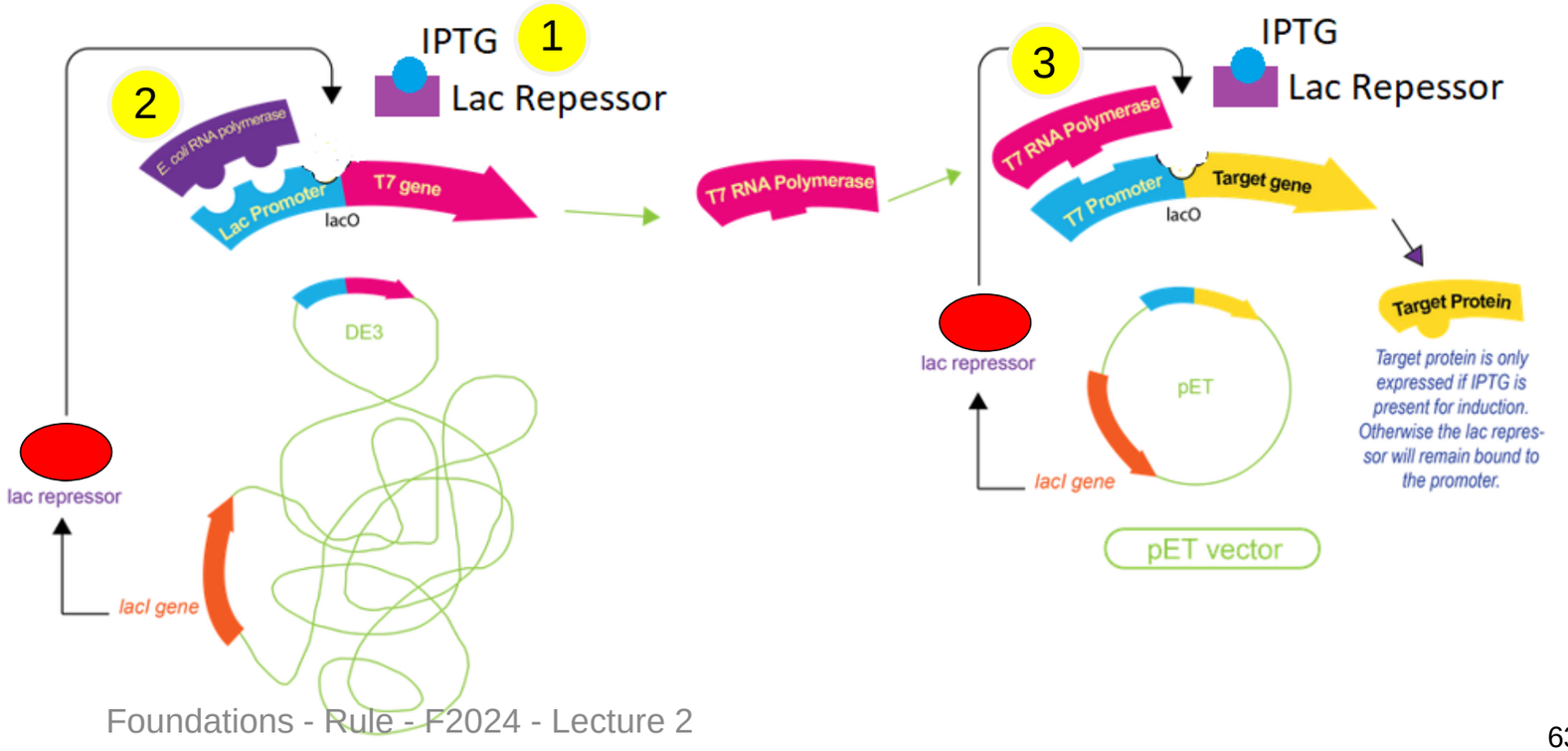
- T7 RNA Pol is not made – its transcription is blocked by lac repressor
- No expression of gene on the plasmid since its promotor is a T7 promoter ***and*** there is no polymerase.





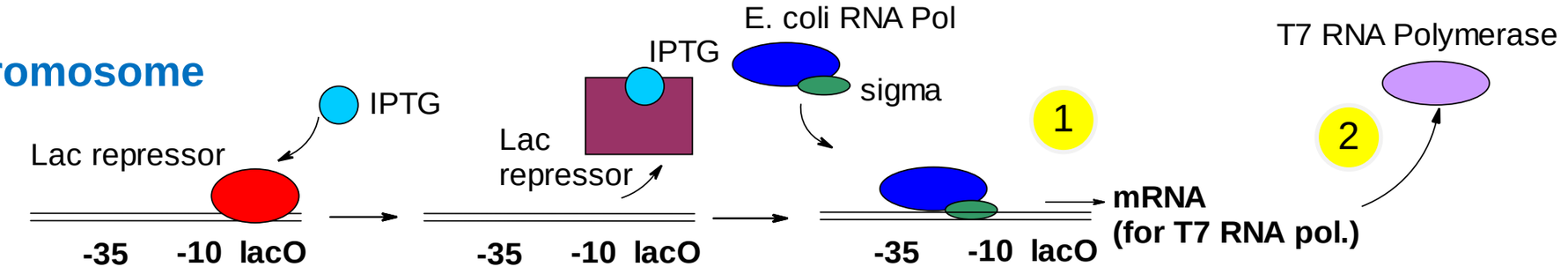
With Inducer (IPTG or Lactose)

1. Lac repressor binds IPTG, dissociates from its operator, on *both* the chromosome & the plasmid.
2. E. Coli RNA polymerase makes mRNA for T7 RNA polymerase, producing T7 RNAP
3. T7 RNAP will generate a mRNA of the gene on the plasmid, producing the target protein.



Regulation of T7 Transcription by Lac Repressor

Events on Chromosome



Addition of IPTG causes:

1. Production of mRNA for T7 RNA polymerase *from the chromosome.*
2. Translation of mRNA to produce T7 RNA polymerase.
3. T7 RNA polymerase produces mRNA *from gene on the plasmid.*
4. mRNA is translated to final target protein.

Events on Plasmid