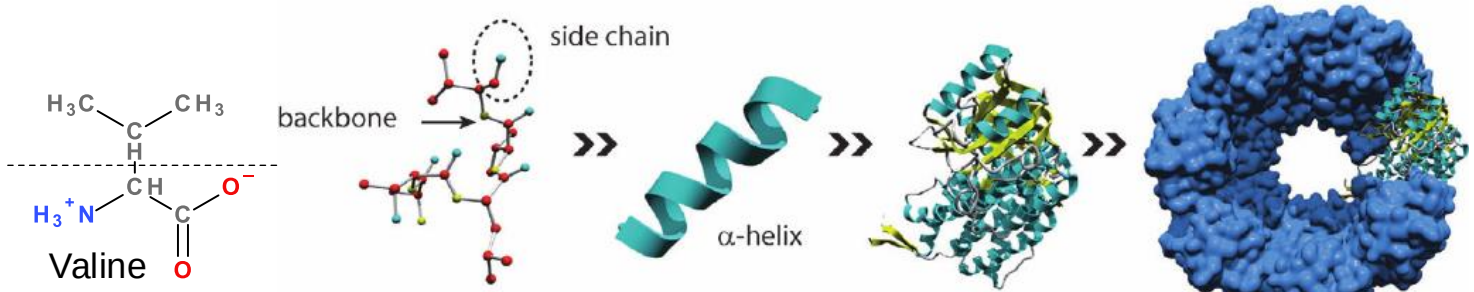
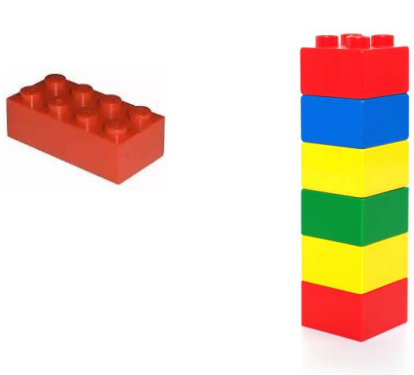
Lecture 2:
Biology Fundamentals

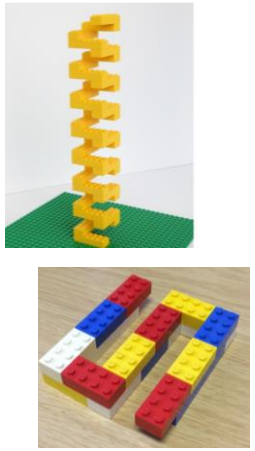
- Review of Protein Structure and Stability
- Ligand Binding
- Proteins as enzymes
- Carbohydrates
- Nucleic acid structure
- Introduction to central dogma
- DNA Sequencing
- Polymerase Chain Reaction (PCR)

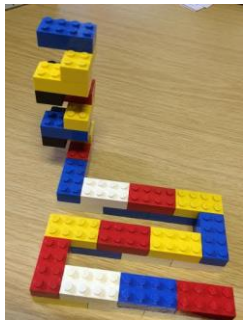
Structural Hierarchy of Proteins

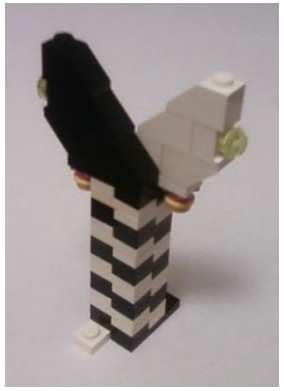
- Primary - sequence of amino acids, no 3D structural information
- Secondary - local structural elements, only mainchain atoms involved
- Tertiary - 3D position of ***all*** atoms, functional form of many proteins.
- Quaternary - multiple chains – multiple chains often required for function.







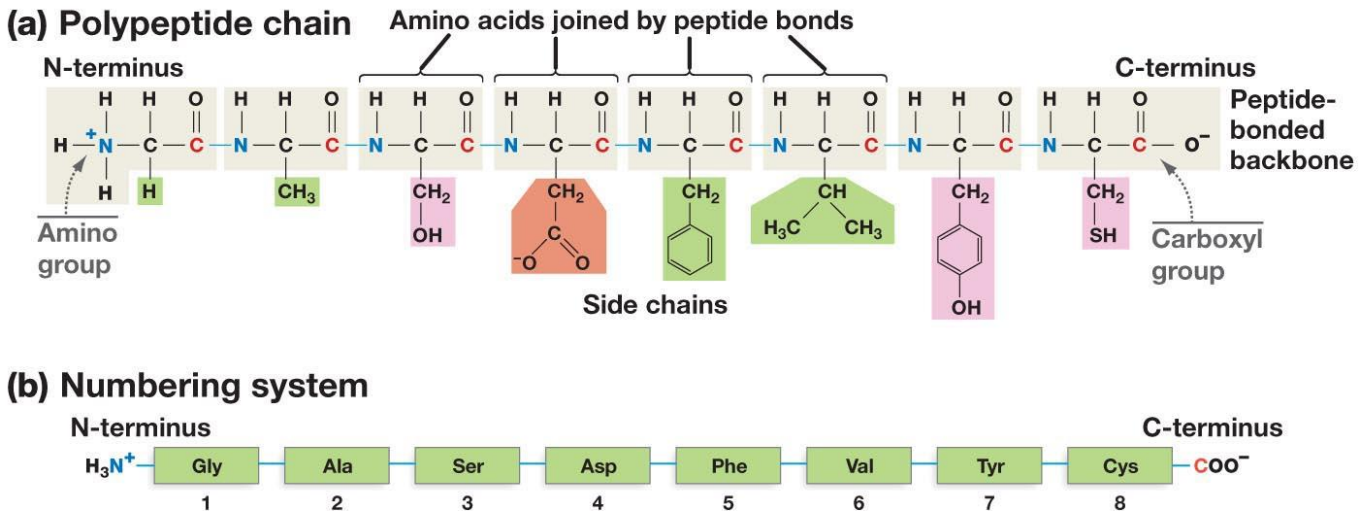




(White and black represent two different chains)

Primary Structure = Protein Sequence

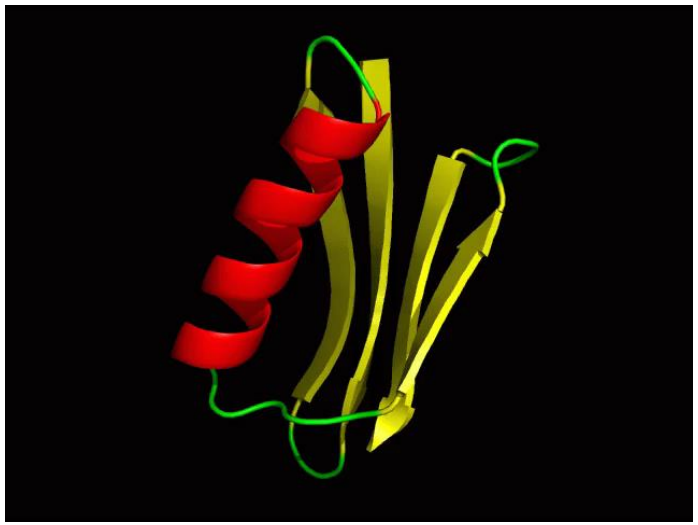
Order of amino acids, from the amino terminus to the carboxy terminus, e.g. Gly-Ala-Ser-Asp…..

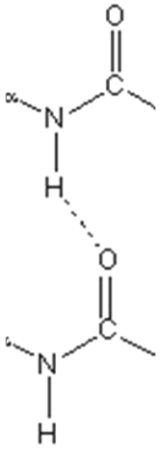


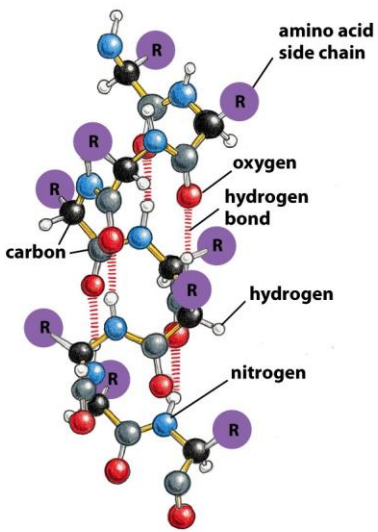
- Within the polypeptide, the "backbone" has three key characteristics:
  1. **Directionality**
     - Free amino group, on the left, is called the **N-terminus.**
     - Free carboxyl group, on the right, is called the **C-terminus.**
  2. **R-group orientation**
     - Side chains can interact with each other or water.
  3. **Flexibility**
     - Single bonds on either side of the peptide bond can rotate, making the entire structure flexible.

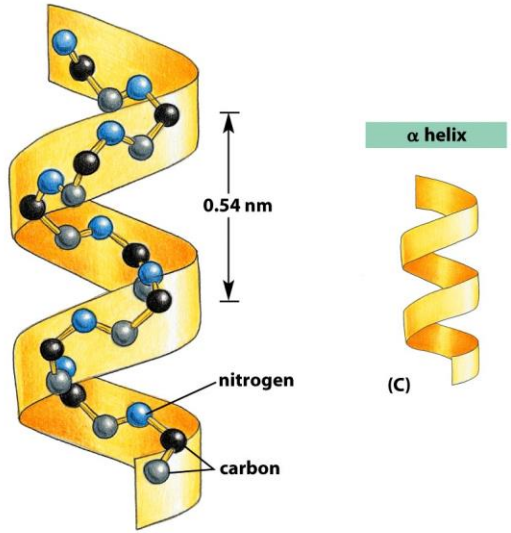
Secondary Structure

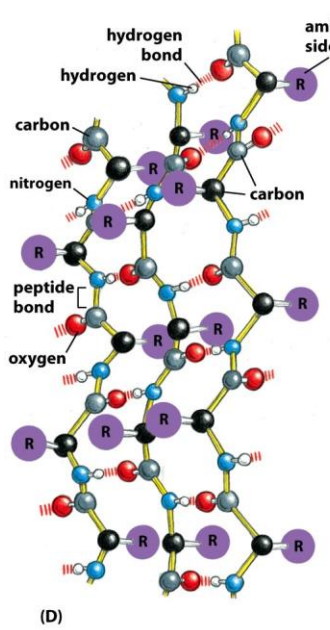
"Building blocks of proteins"

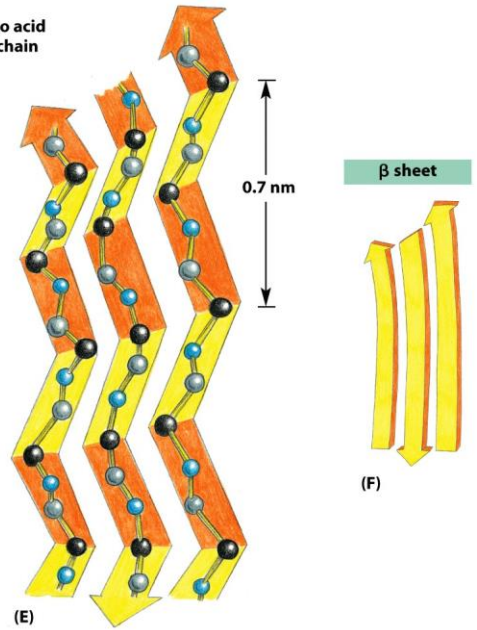




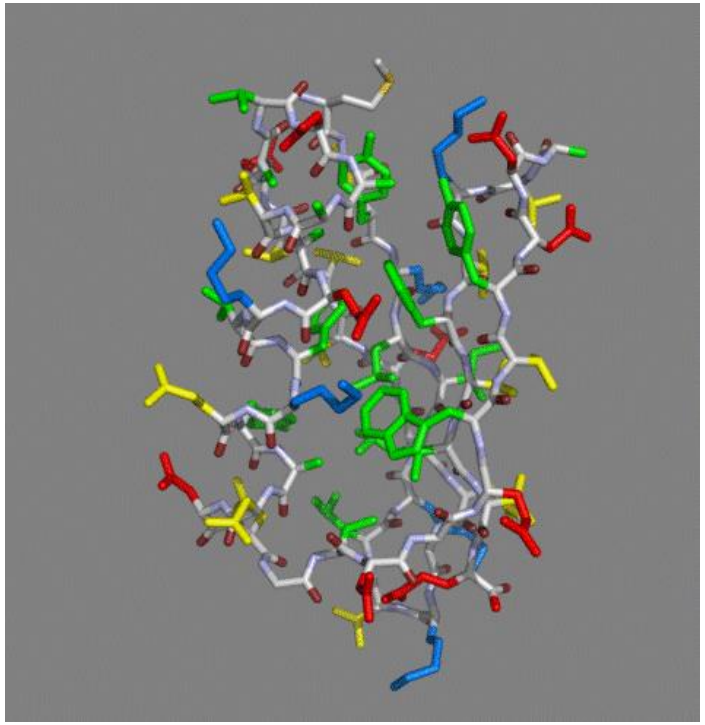








Location of Residues in Globular Proteins

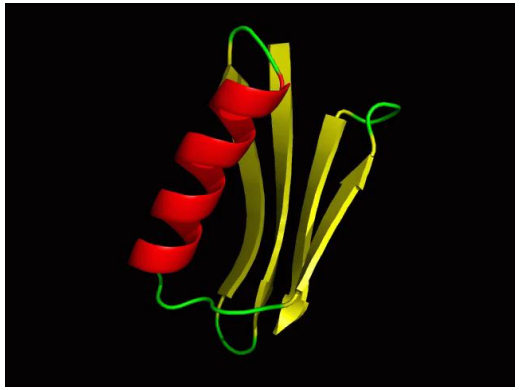


Red - amino acids with neg. sidechains

Blue - amino acids with pos. sidechains

Yellow – amino acids with polar sidechains.

Green - amino acids with hydrophobic side chains



Amino Acid	Inside (core)	Surface
Charged		
Polar		
Non-polar		

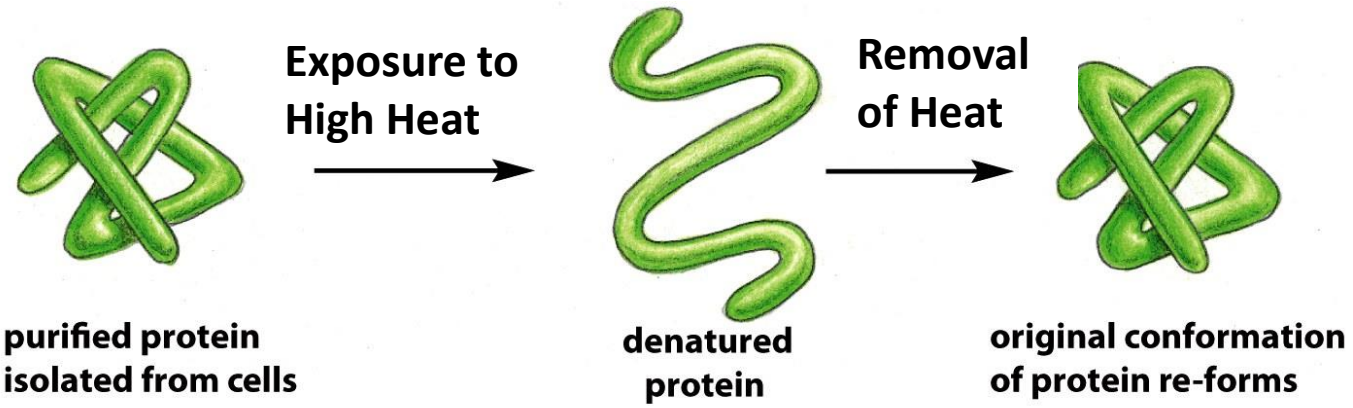
Protein Stability

H-bonds
van der Waals
Hydrophobic effect

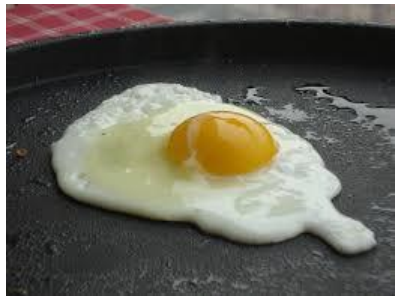
Chain disorder

Protein Denaturation

Native Unfolded



- Often, unfolded protein aggregate, which prevents refolding.



Hydrogen Bonding Stabilizes the Folded Form

General Pattern:

X-H Y

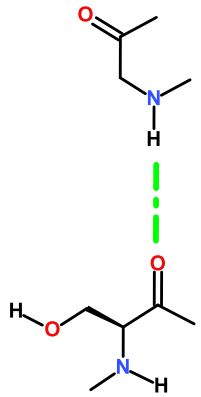
Possible Donors:

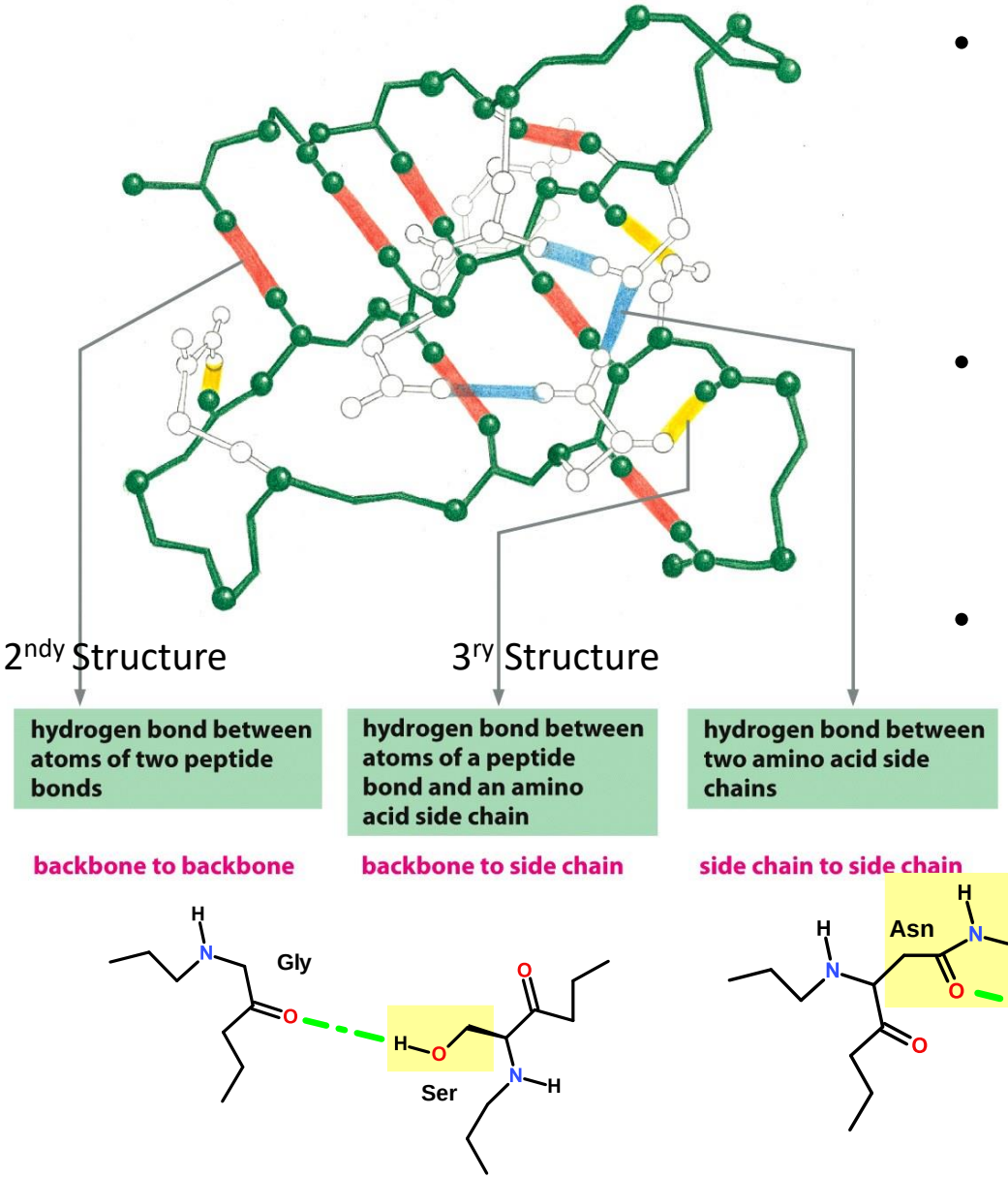
N-H or O-H

Possible Acceptors

C=O or -O-H or

N in amino group (Lysine)

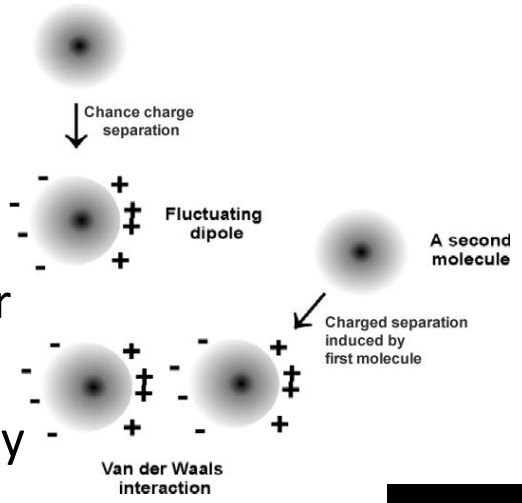


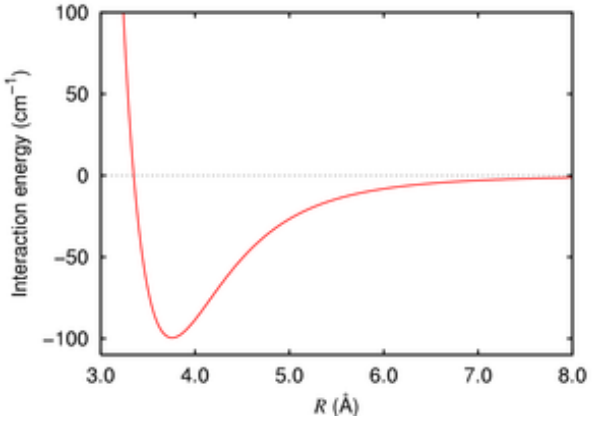


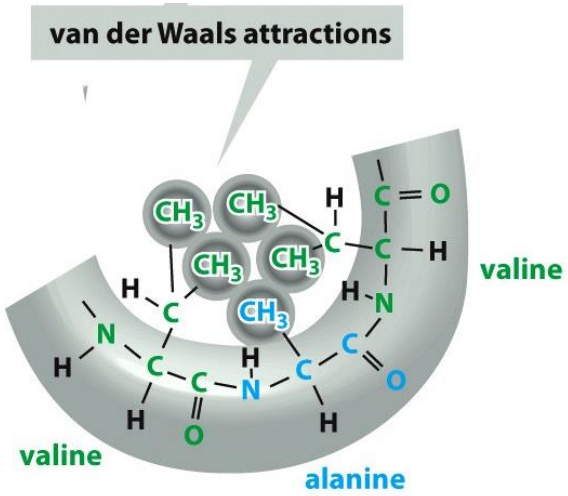
- **Hydrogen bonds** form between hydrogen atoms and the carbonyl group in the peptide-bonded backbone = secondary structure
- Hydrogen bonds are also found between hydrogen and electronegative atoms on side chains (sidechain-sidechain)
- Sidechains can form hydrogen bonds to the mainchain too.

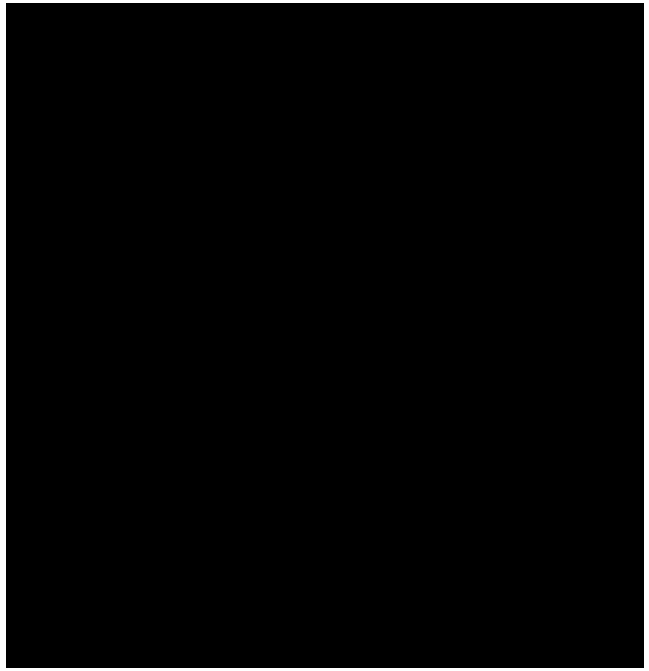
Van der Waals Interactions Stabilize the Folded State

- VdW are weak electrostatic interactions between side chains due to temporary (fluctuating) charges.
- Attractive from long distance
- Distance at lowest energy is at the van der Waals radii of the atoms.
- Optimized in the core of folded proteins by "knobs fitting into holes"





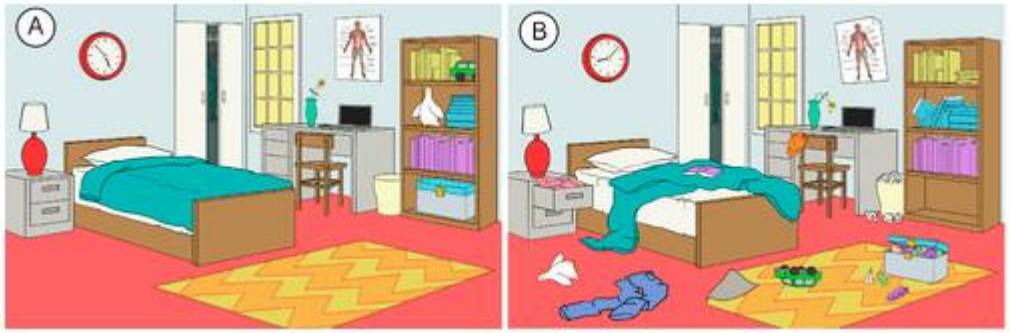


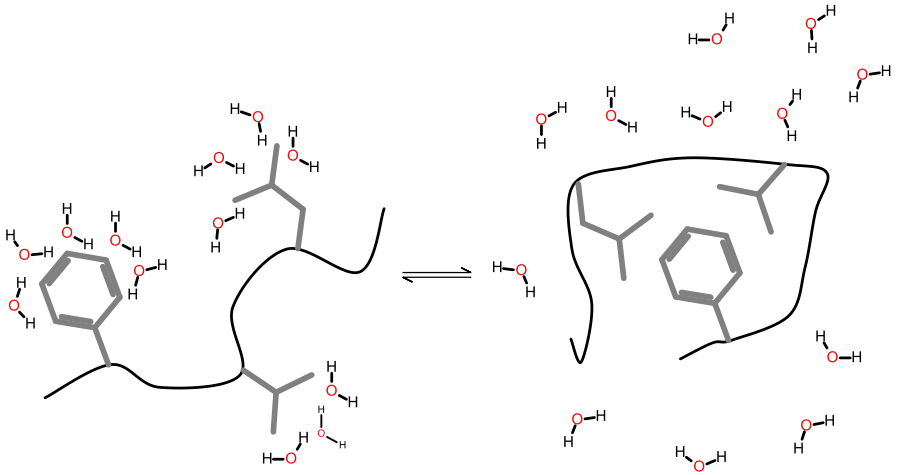


Hydrophobic Interactions are Critical for Stabilizing the Folded Structure

Ordered water hydrating a non-polar group

Energy and Entropy





Hydrophobic interactions within a folded protein increase stability of the folded protein by releasing the ordered water that surrounded exposed non-polar groups in the unfolded protein. *Folding increases the entropy of the water – favorable.*

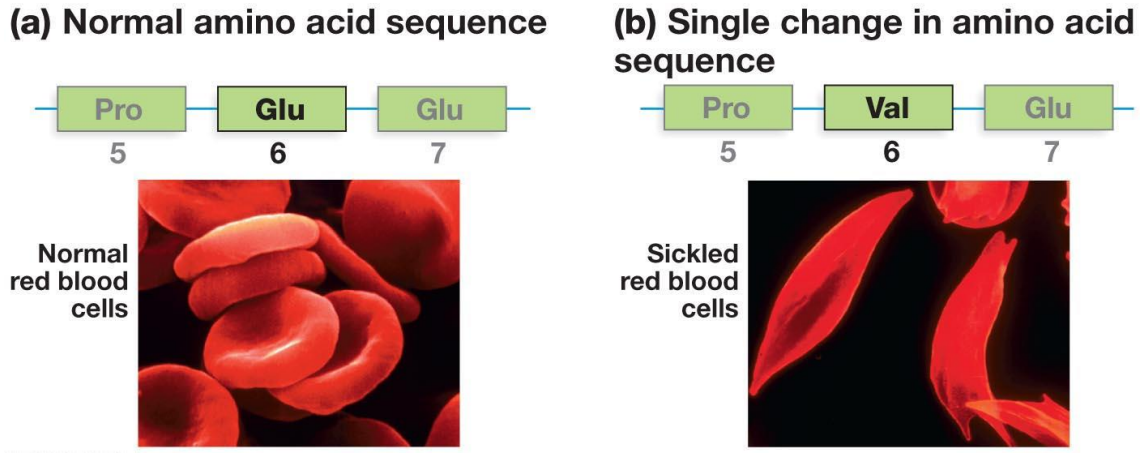
TABLE 3.1 **How Amino Acids Interact with Water**

20 amino acids are ranked according to how likely they are to interact with water. Color codes are based on Figure 3.3.

Isoleucine	Highly hydrophobic
Valine	
Leucine	
Phenylalanine	
Methionine	Moderately hydrophobic
Alanine	
Glycine	
Cysteine	
Tryptophan	
Tyrosine	Mildly hydrophobic
Proline	
Threonine	
Serine	
Histidine	Mildly hydrophilic
Glutamate	
Asparagine	
Glutamine	
Aspartate	
Lysine	
Arginine	Highly hydrophilic

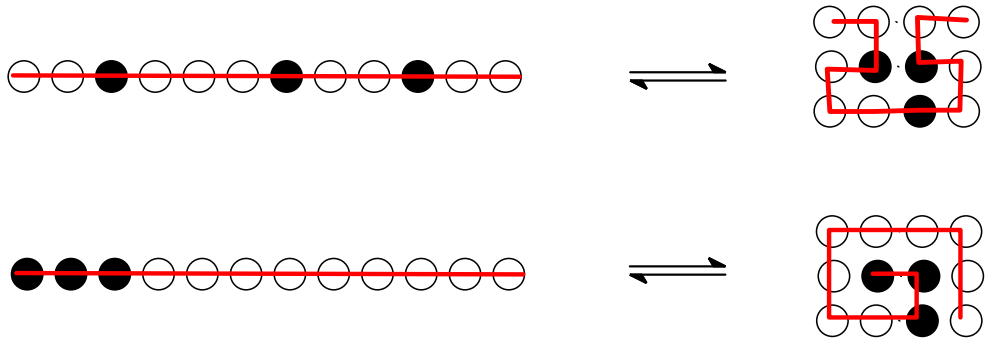
Fold Depends on Amino Acid Sequence

Effect of mutations on protein folding – sickle cell anemia

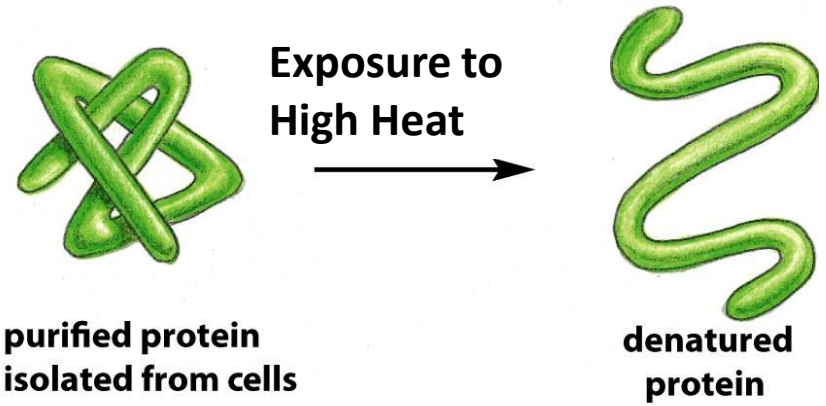


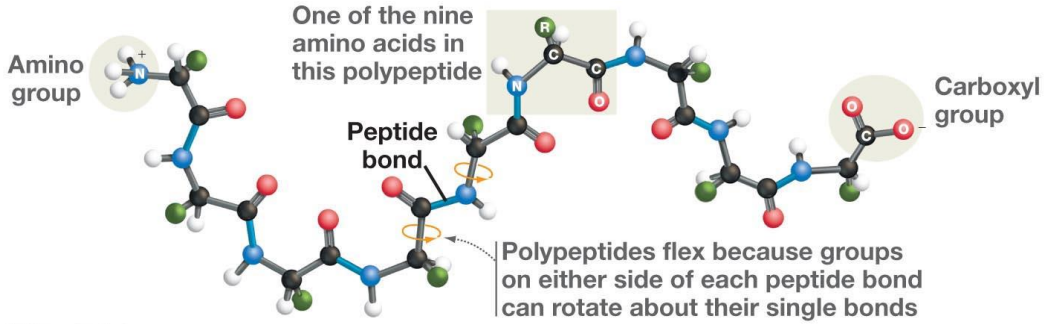
A single change in the amino acid sequence can change the function of a protein, and often affecting how it folds.

The *position* of non-polar residues (filled circles) mostly affects the final fold:

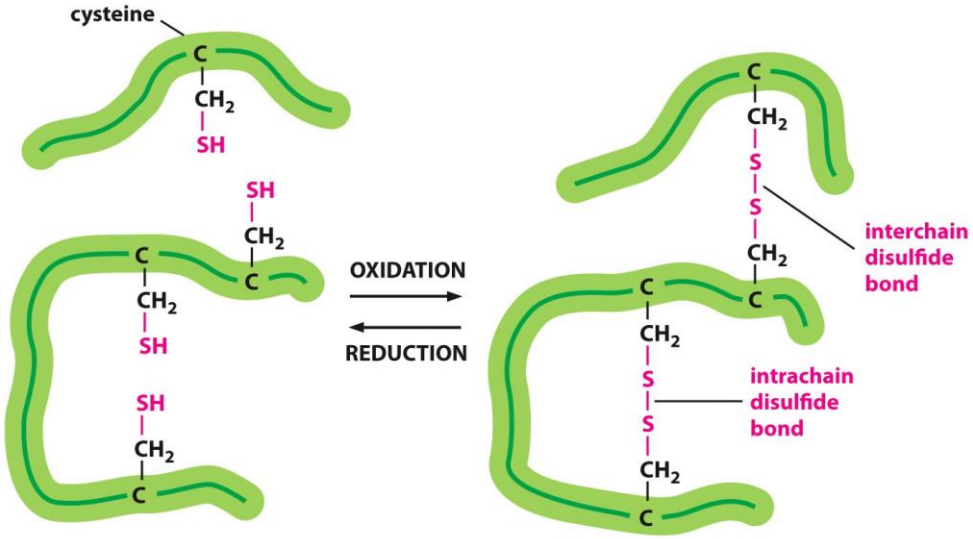


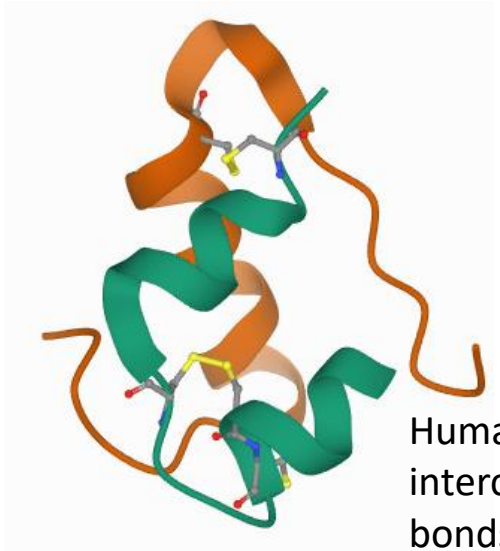
Unfolded Polypeptides Are Flexible – High Entropy stabilizes the Unfolded state



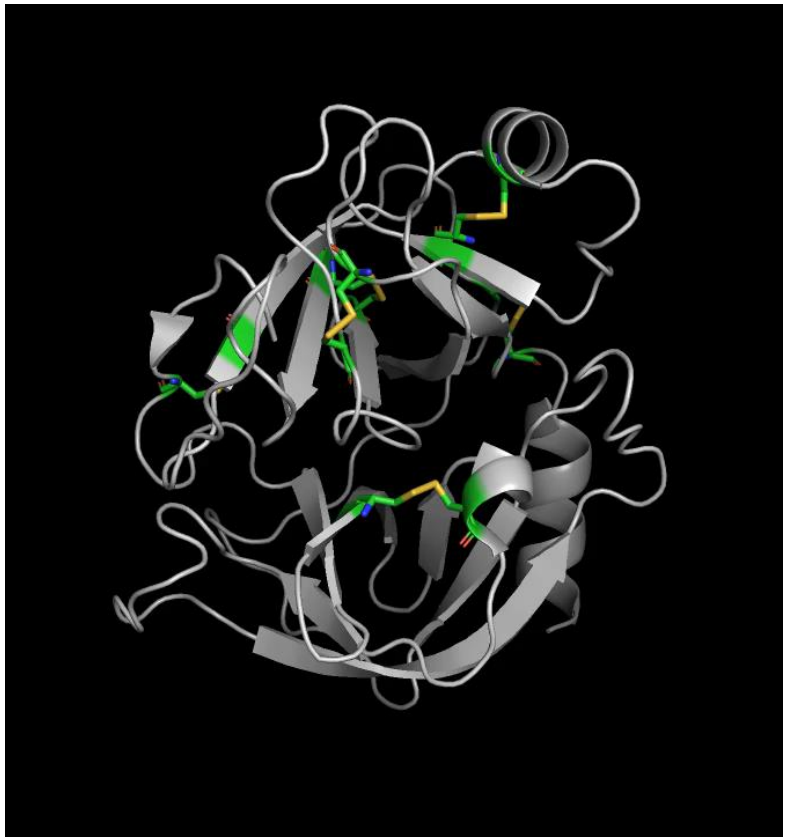


Disulfide Bonds Stabilize Some Proteins Outside the Cell (and body)





Human Insulin – interchain disulfide bonds



Trypsin – a digestive enzyme produced in the pancreas, exported to the small intestine – disulfide bonds within a single chain.

Summary - Interactions that Stabilize Folded Proteins.

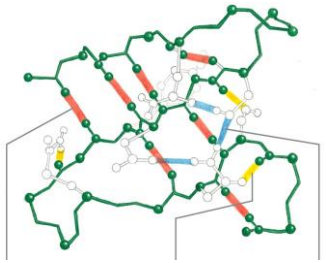
- **Hydrogen bonds** form between hydrogen atoms (NH) and the carbonyl group in the peptide backbone (mainchain), and between and donors and acceptors on sidechains. *Mainchain-mainchain H-bonds are responsible for secondary structures.*
- **Hydrophobic interactions** within a protein increase stability of the folded state by *increasing entropy due to the release of water that was ordered by the exposed non-polar groups in the unfolded protein.*
- **van der Waals interactions** are *optimized in the well packed core of the protein*.

- **Covalent disulfide bonds** form between sulfur-containing cysteine residues stabilizing them (usually only exported, secreted proteins).

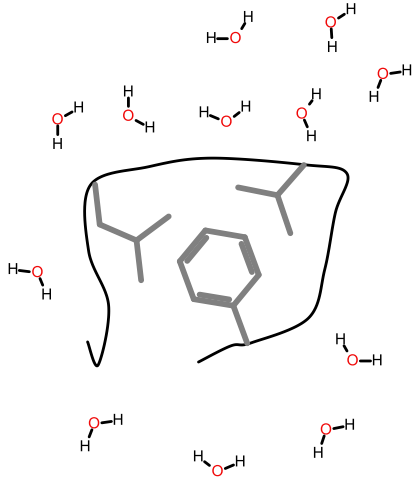
H-bonds
van der Waals
Hydrophobic effect

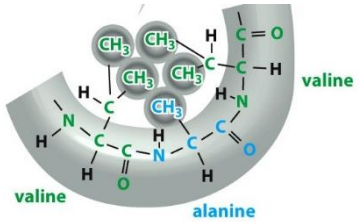
Chain disorder

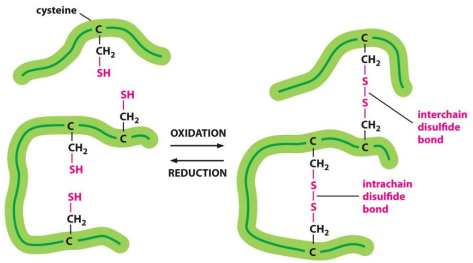
Native



Unfolded

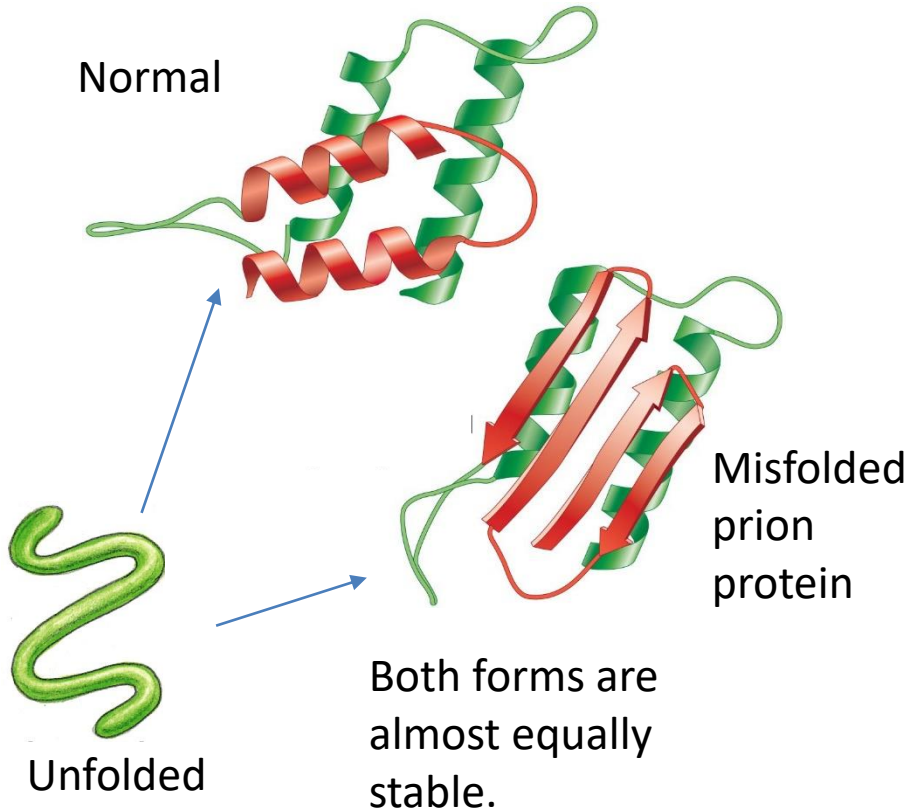






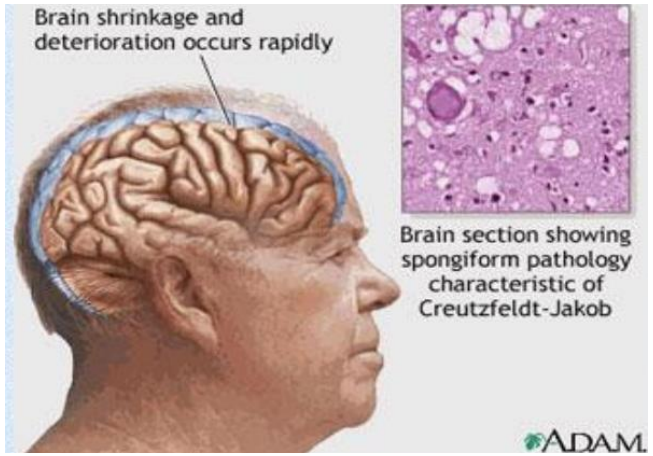
What happens when proteins don't fold properly?

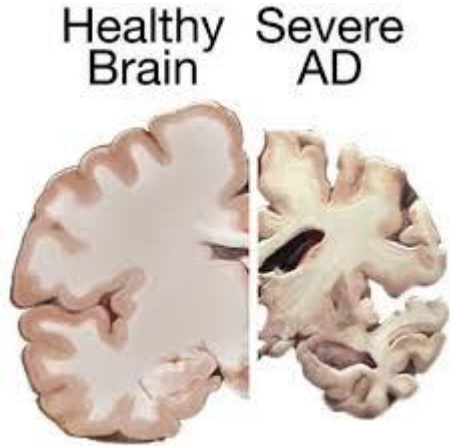
Prions are improperly folded proteins that cause neurodegenerative diseases



Both forms are almost equally stable.

Why does this occur?





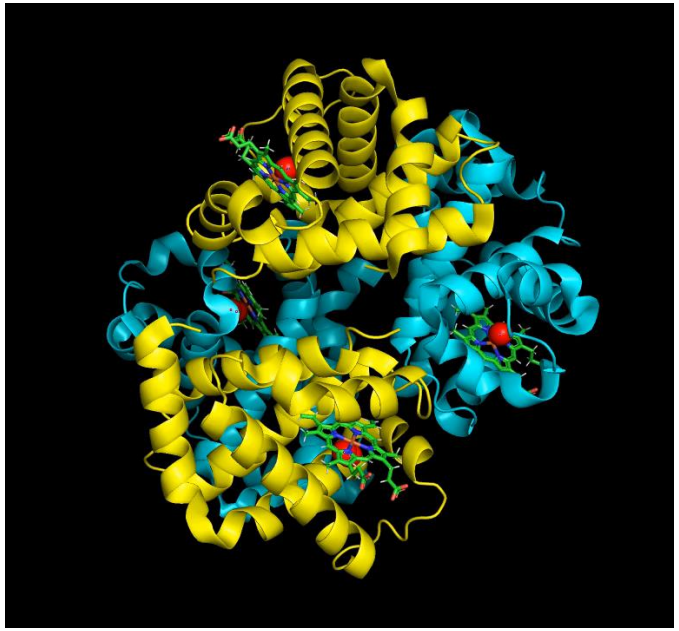
What is the effect on the brain?

Unfolded protein response (UPR):
The presence of unfolded proteins can trigger the UPR, which can turn off protein synthesis in the cell, leading to cell death.

Why do the brain cells die?

Quaternary Structure

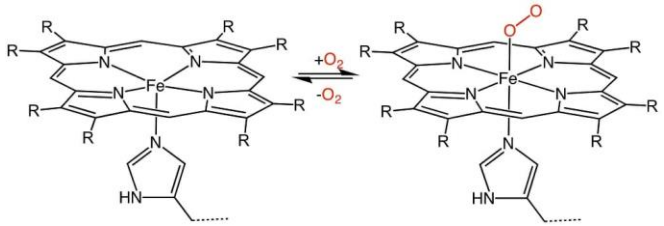
- Combinations of polypeptide subunits (combinations of tertiary structures).
- May be held together by covalent bonds (disulfide), but usually non-covalent interactions between R groups on the different chains.
- Proteins can be a dimer, a tetramer, etc.
- If the chains are the same, called homo_____. If chains are different, hetero_____



Quaternary structure of hemoglobin (oxygen transport protein):

- two α chains
- two β chains

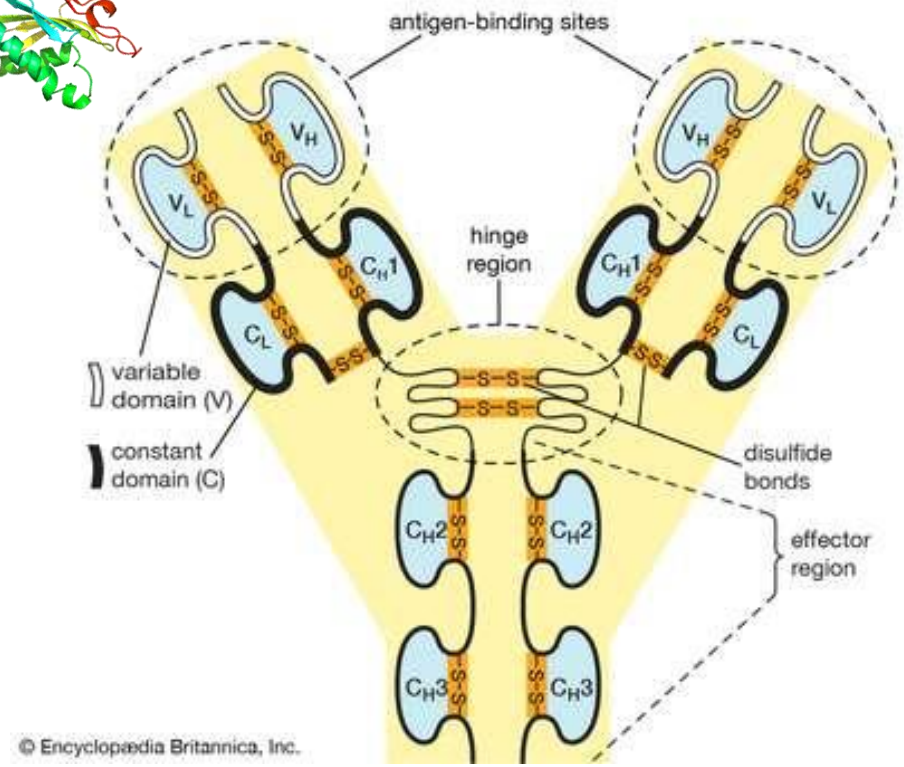
Oxygen is carried on Fe^{2+} within heme groups:

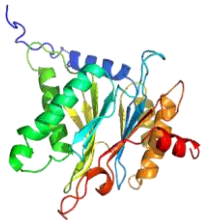


Antibodies – Produced by the Adaptive Immune system to Fight Pathogens.

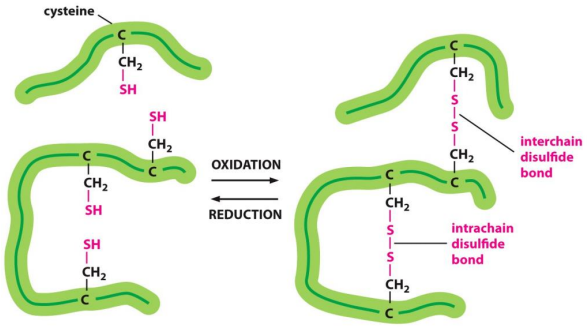
Antigen

Antigen = ligand that binds to an antibody.





Antigen



Properties of Antibodies:

- 4 chains – two identical light (200 aa), two identical heavy (400 aa).
- Bind two identical antigens (pathogens, toxins)
- Chains crosslinked with disulfide bonds, increasing stability.

Protein Structure - Summary and Expectations

Primary Structure:

- Can you describe the mechanism of peptide bond formation
- Can you draw structure of peptides.
- Can you identify amino terminus and give the sequence of amino acids, N -> C

Secondary structure:

- Identify helical and sheet secondary structures,
- know that they are stabilized by **mainchain** hydrogen bonds between N-H and O=C.
- Location of H-bonds and sidechains

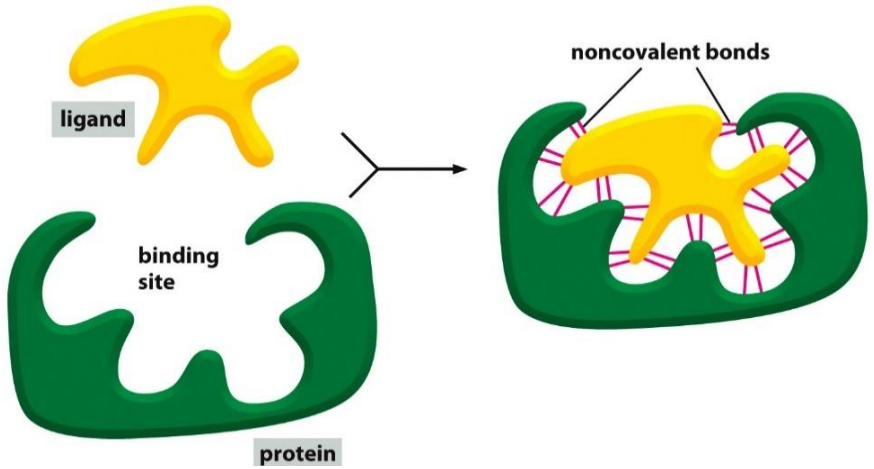
Tertiary Structure:

- Can you describe and identify role of the following in stabilizing the folded state.
  - H-bonds,
  - van der Waals,
  - hydrophobic effect
- Can you predict, based on sidechain, which amino acids are found in the core of a protein, and which are found on the surface.

Quaternary Structure:

- Multiple chains, stabilized by non-covalent and covalent (disulfide bonds) interactions.
- What is the quaternary structure of an antibody?

Ligand Binding: Most Proteins Bind to Other Molecules in Biological Interactions:



Ligand: Something that binds to a protein, usually small molecules (e.g. cyclicAMP, cAMP).

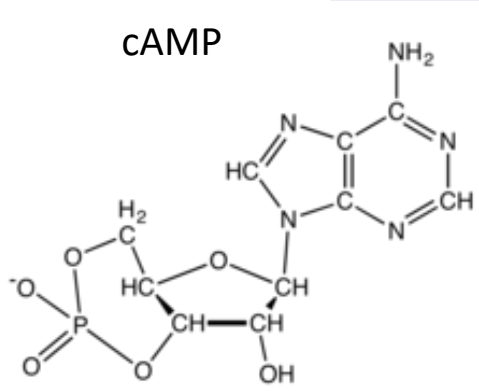
Binding site allow a protein to interact with specific **ligands**

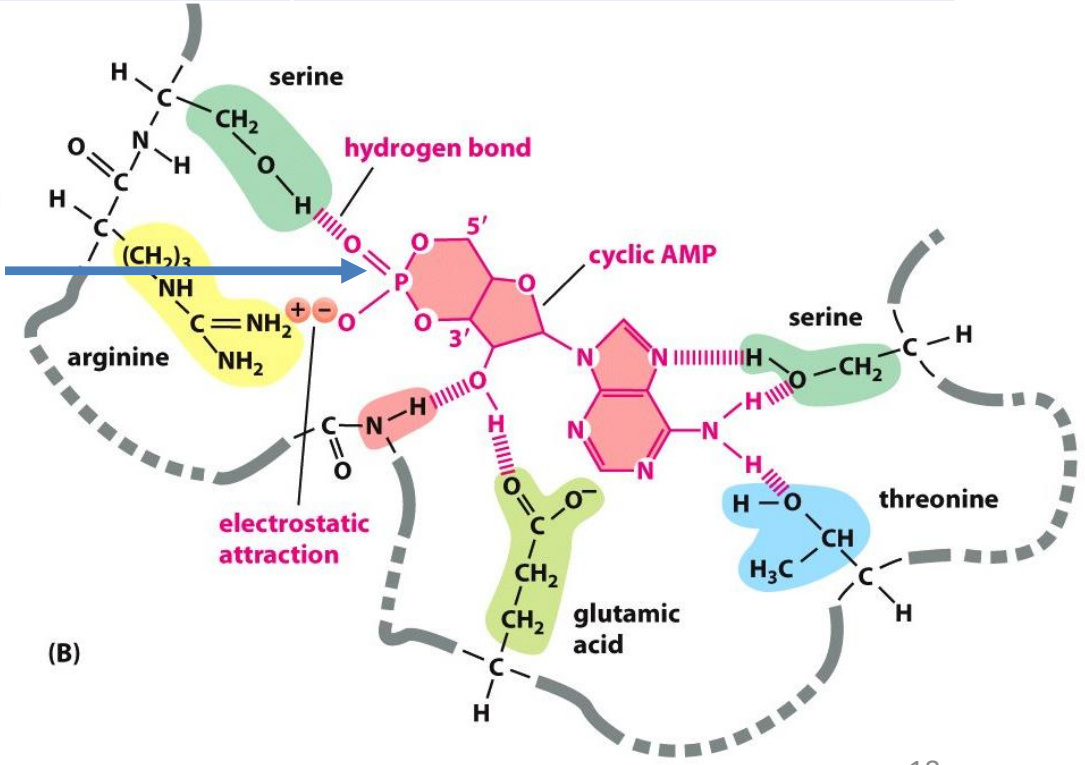
Binding site is generated by the **folded** form of the protein.

The bound ligand can be stabilized by any and all of:

Interaction	Which stabilize cAMP Binding?
van der Waals	
H-Bonding	
Electrostatic	
Hydrophobic effect	

cAMP





Ligand Binding & Saturation:

Define fraction saturated: $Y = \frac{[ML]}{[M] + [ML]}$

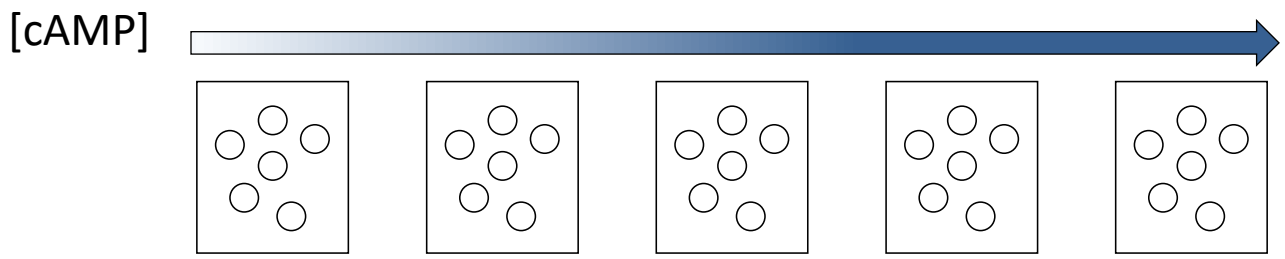
[M] = free macromolecule (e.g. antibody with no antigen).

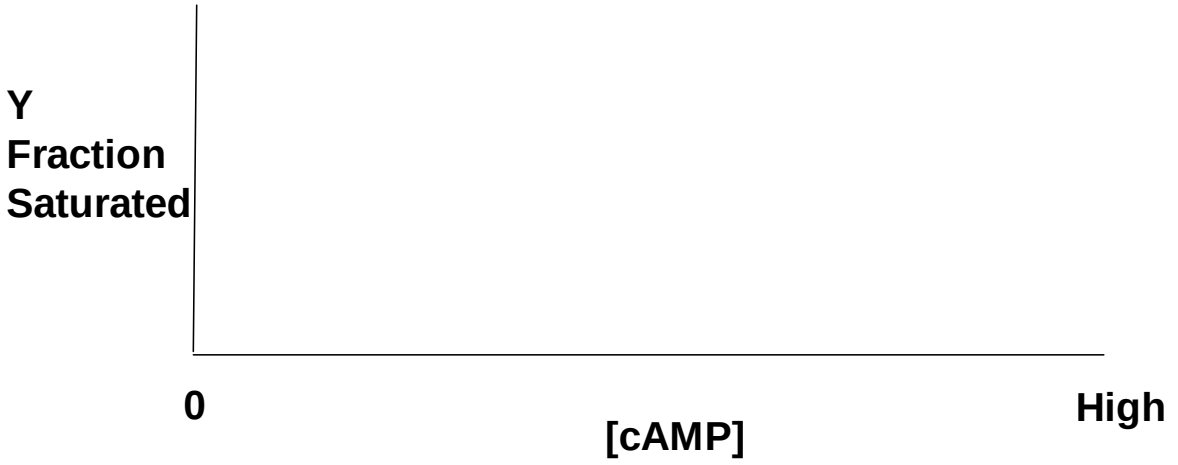
[ML] = macromolecule with ligand bound (e.g. antibody with antigen bound).

The boxes with circles represent proteins with no cAMP bound, each box (left to right) is at a higher [cAMP]. Filled circles indicate bound ligand.

1. How will the number of filled circles depend on the cAMP concentration?

2. Plot the location on the fraction saturated curve for each box.





Key Points:

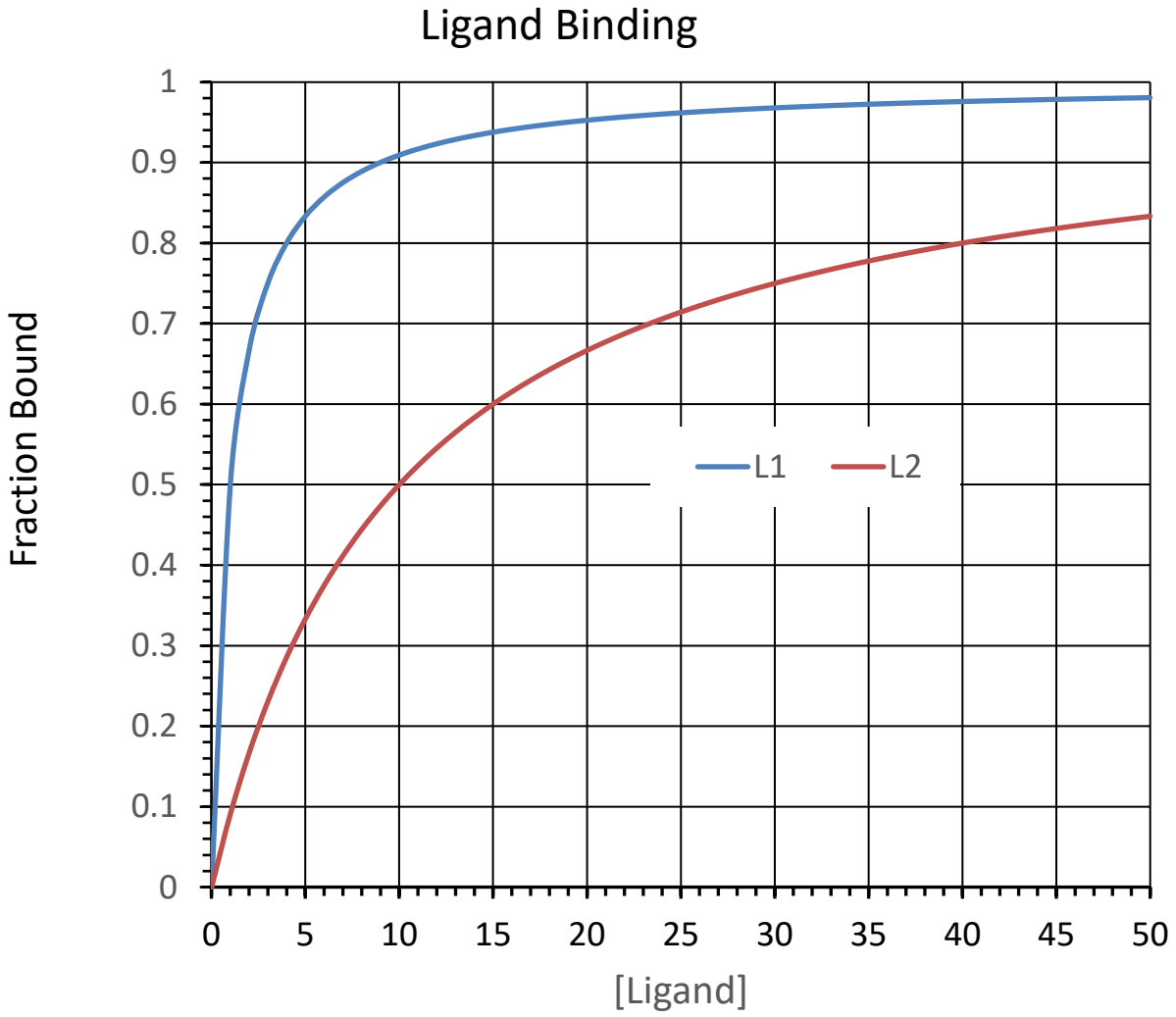
1. The binding sites saturate, when all are full no more ligand can bind.
2. There is a ligand concentration, [L], where ½ the sites are full. This [L] is K_D.
3. K_D is the equilibrium constant for ligand dissociation:

$$K_{Eq} = \frac{[products]}{[reactants]}$$

$$(ML) \rightleftharpoons (M) + (L)$$

$$K_D = \frac{[M][L]}{[ML]}$$

Using K_D to Compare Ligand Binding



The binding of two different molecules to the same protein was measured and the data is shown on the right. L1 is cAMP, L2 is similar to cAMP

Which ligand has a K_D of 1? L1 or L2?

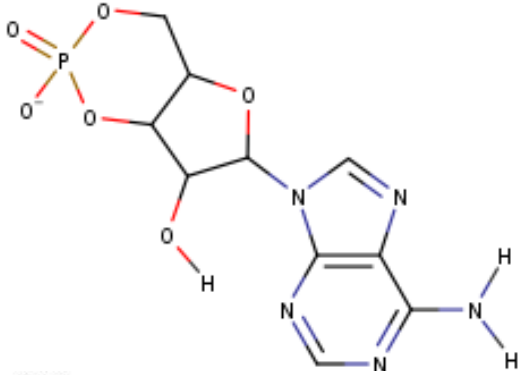
Which ligand has a K_D of 10? L1 or L2?

Which ligand binds more tightly to the protein (higher affinity)? L1 or L2?

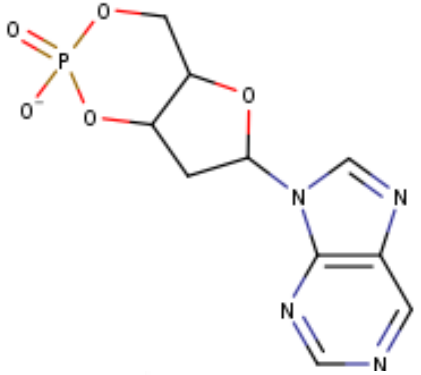
Why does L1 bind more tightly (higher affinity)?

1. What are the chemical differences between L1 and L2 (Upper diagram)

Ligand 1 (cAMP)

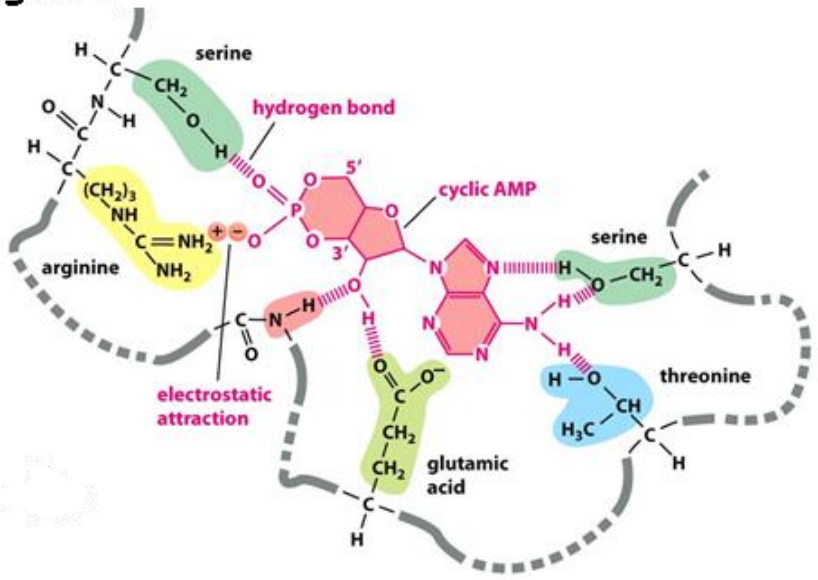


Ligand 2

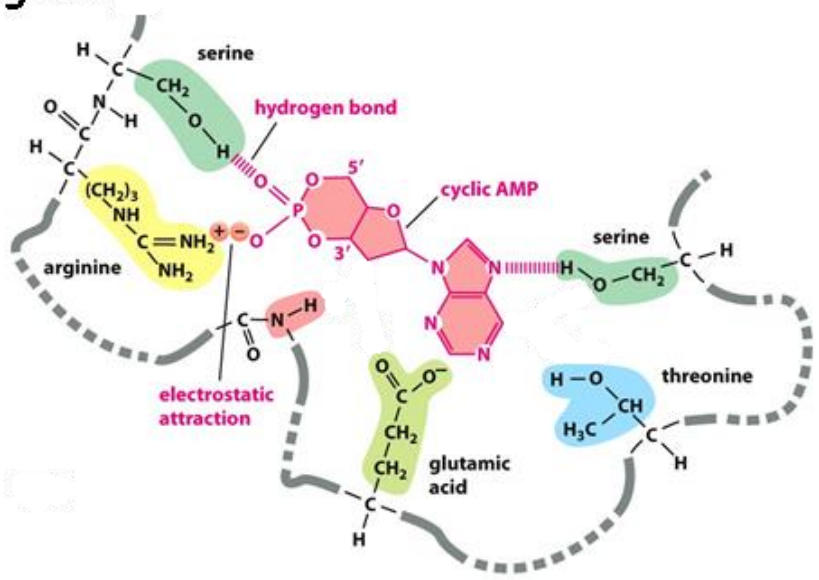


2. How do these differences affect the interactions with the protein (lower diagram)?

Ligand 1



Ligand 2



3. How do the differences affect K_D?

Key Points:

Binding:

Folded proteins have ***binding sites*** that recognize other molecules (***ligands***) using ***any and all*** of the following:

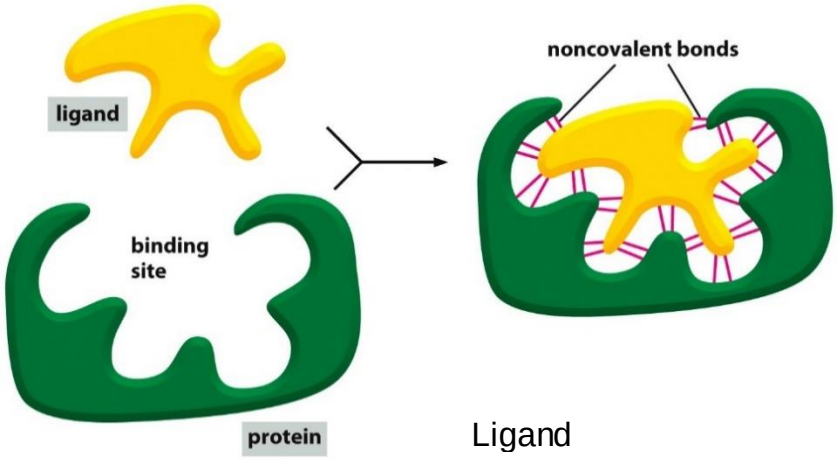
- H-bonds,
- van der Waals,
- Electrostatic,
- Non-polar interactions (hydrophobic)

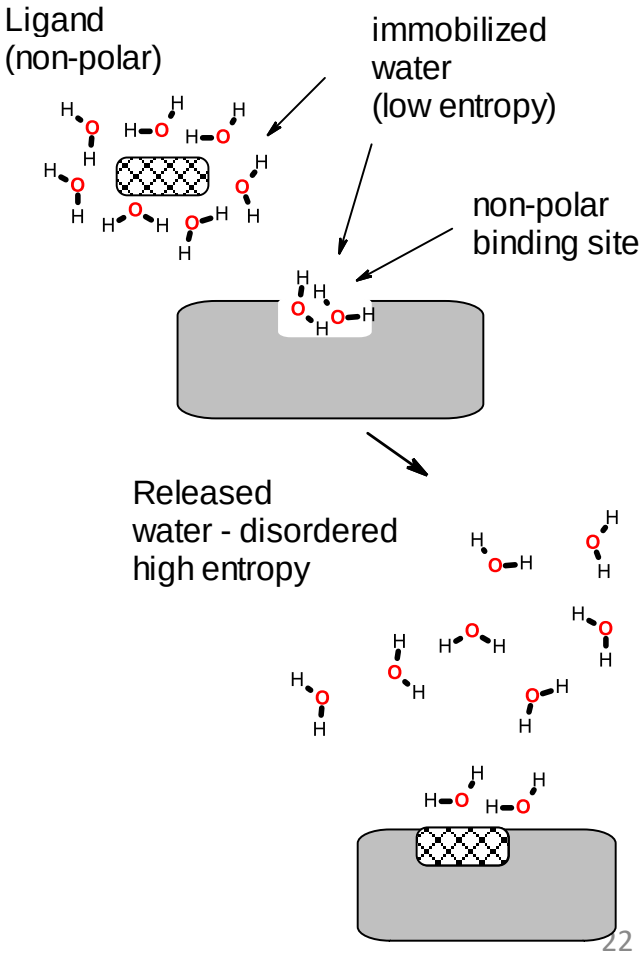
Binding is **reversible**

Binding is **saturable**

Binding ½ point (Y=0.5) occurs at K_D

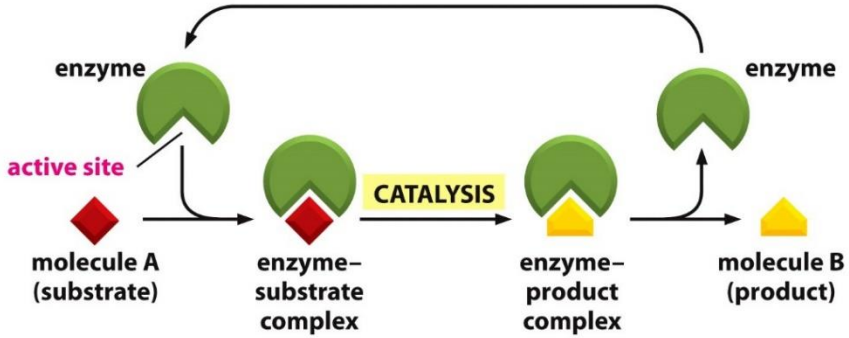
The higher the affinity (strength of interaction), the lower the K_D

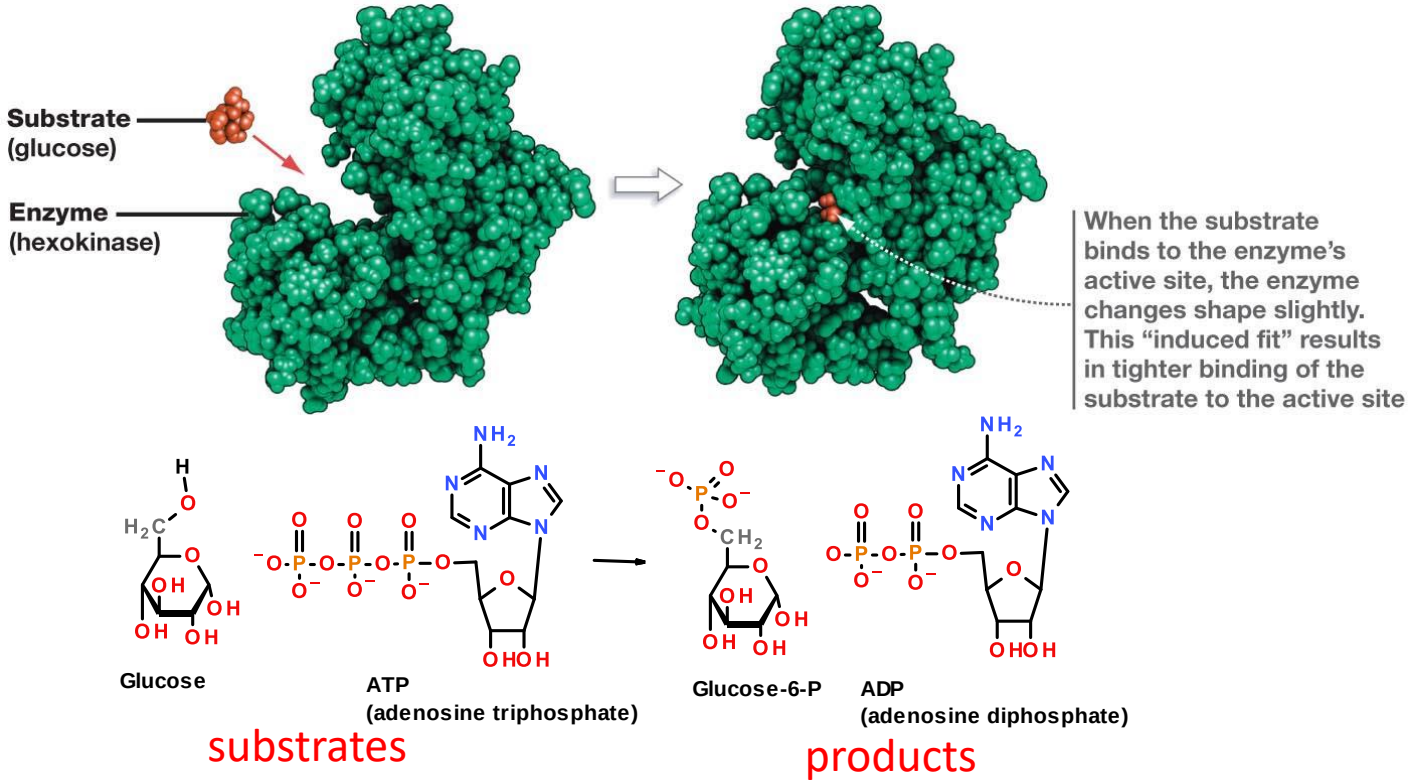




Enzymes

- **Enzymes** are protein or RNA catalysts. They increase the rate of the reaction.
- They bind "substrates" and convert them to "products". Usually, the substrate undergoes a chemical reaction and is changed in its structure.
- Most biological chemical reactions occur at meaningful rates only in the presence of an enzyme.
- Substrates bind specifically to the enzyme's **active site,** interacting with amino acid side chains (or RNA bases). Usually a single enzyme binds one substrate.
- The chemical change caused by the enzyme is catalyzed by additional functional groups in the active site.
- Many enzymes undergo a conformational change when the substrates are bound to the active site; this change is called an **induced fit**.





Enzyme – Chemical Diversity

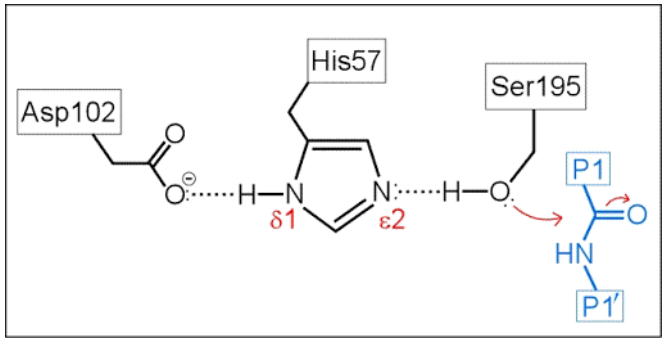
TABLE 4–1 SOME COMMON FUNCTIONAL CLASSES OF ENZYMES

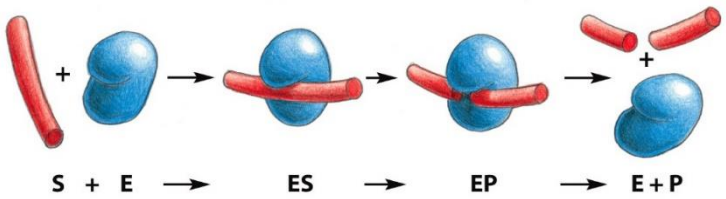
ENZYME CLASS	BIOCHEMICAL FUNCTION
Hydrolase	General term for enzymes that catalyze a hydrolytic cleavage reaction.
Nuclease	Breaks down nucleic acids by hydrolyzing bonds between nucleotides.
Protease	Breaks down proteins by hydrolyzing peptide bonds between amino acids.
Synthase	General name used for enzymes that synthesize molecules in anabolic reactions by condensing two molecules together.
Isomerase	Catalyzes the rearrangement of bonds within a single molecule.
Polymerase	Catalyzes polymerization reactions such as the synthesis of DNA and RNA.
Kinase	Catalyzes the addition of phosphate groups to molecules. Protein kinases are an important group of kinases that attach phosphate groups to proteins.
Phosphatase	Catalyzes the hydrolytic removal of a phosphate group from a molecule.
Oxido-reductase	General name for enzymes that catalyze reactions in which one molecule is oxidized while the other is reduced. Enzymes of this type are often called oxidases, reductases, or dehydrogenases.
ATPase	Hydrolyzes ATP. Many proteins with a wide range of roles have an energy-harnessing ATPase activity as part of their function, including motor proteins such as myosin and membrane transport proteins such as the sodium–potassium pump.

- Most enzyme names end in “-ase”
- Usually named by their substrates and the reactions they catalyse, i.e. glucose kinase

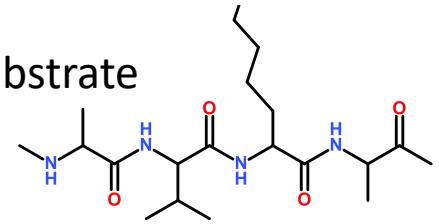
Example of Active Site Functional Groups:
Catalytic triad (Asp, His, Ser) in Protease Trypsin cleaves after Lys Residues

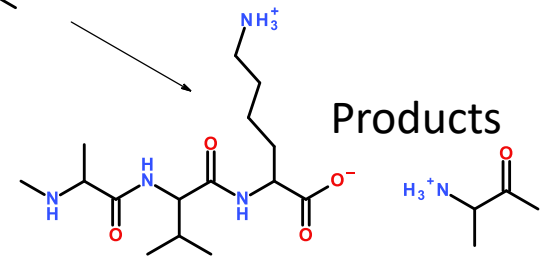
Catalytic triad

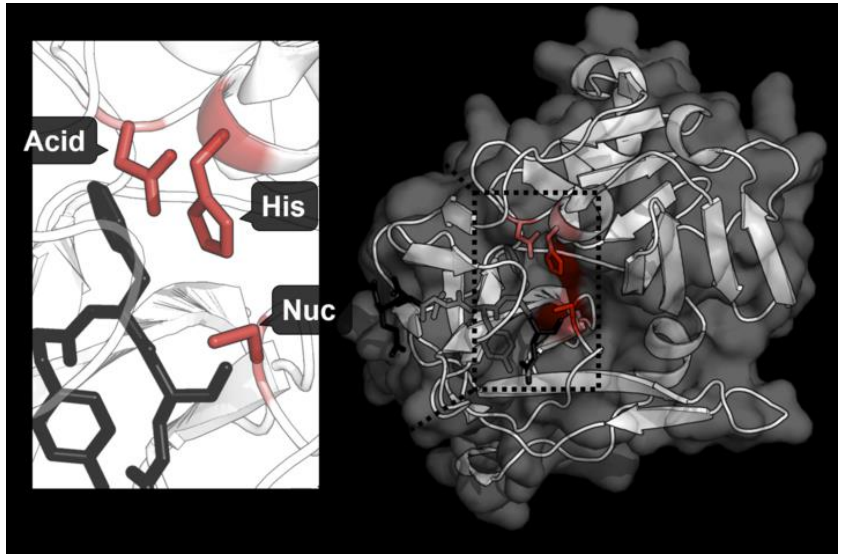




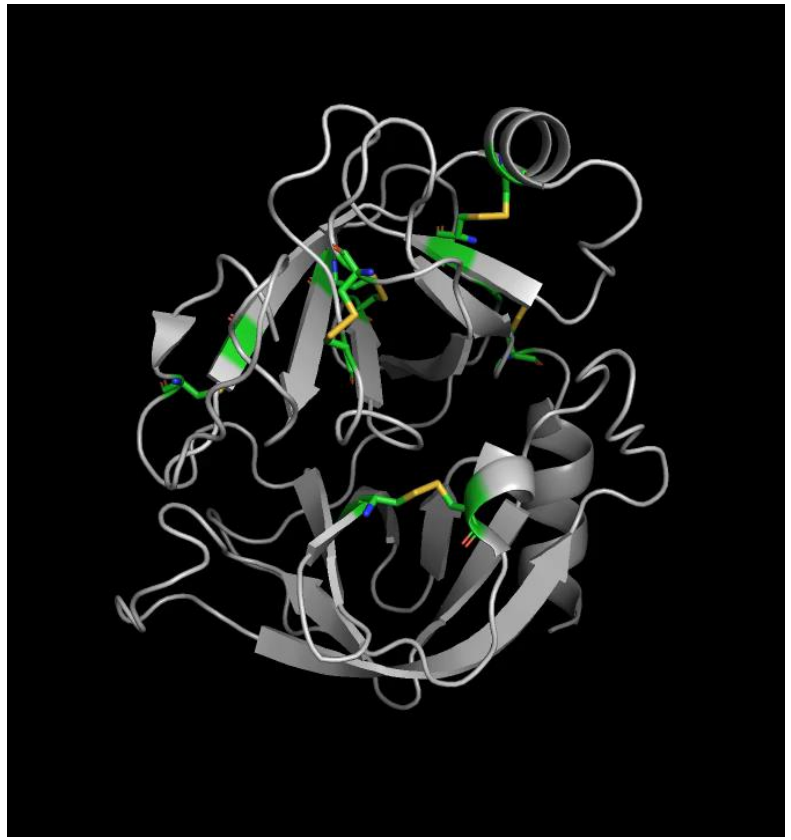
Substrate







https://shirleychemproject.weebly.com/



Disulfide bonds in trypsin

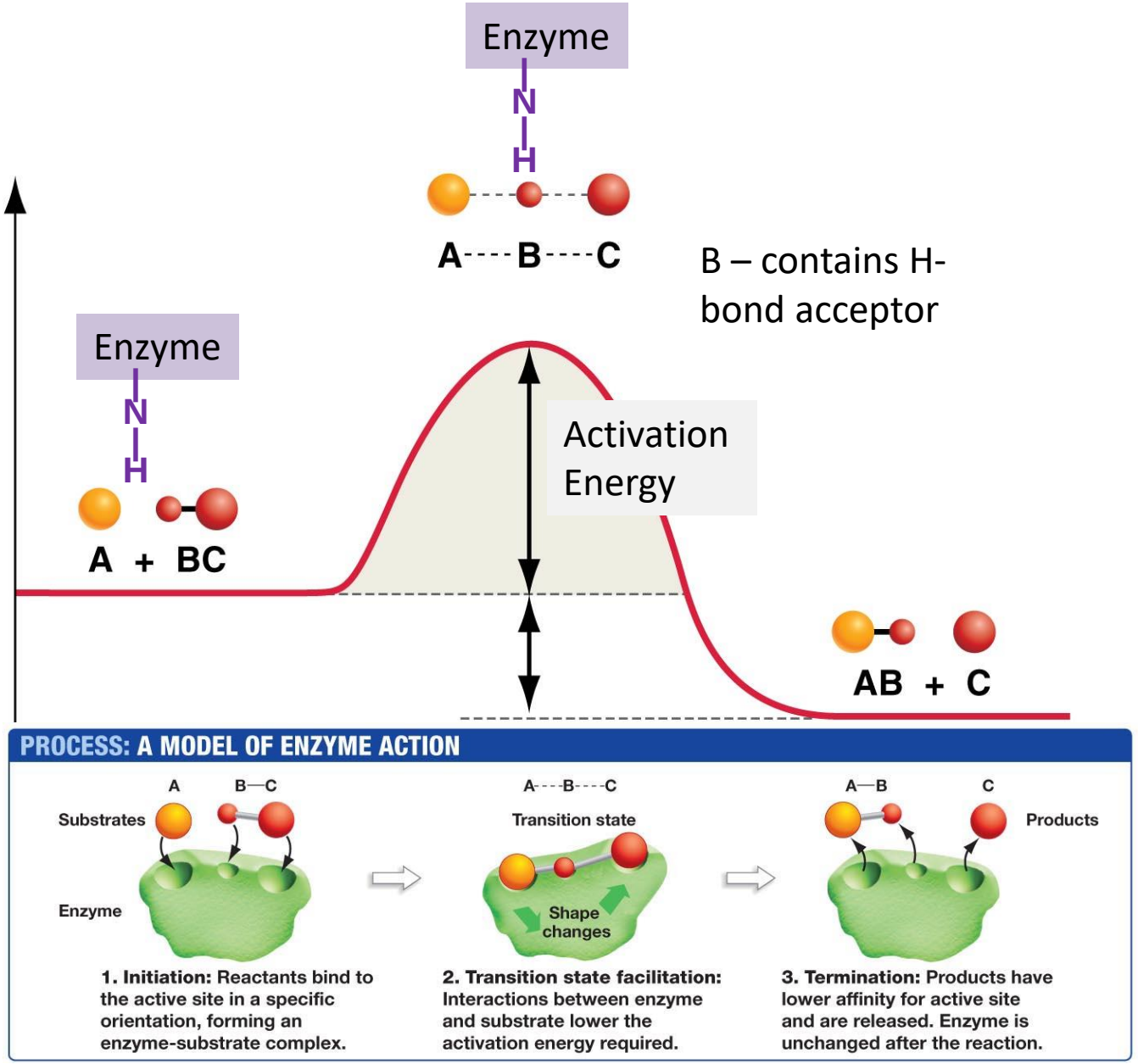
How Do Enzymes Increase Rates?

- **Transition state** = high energy intermediate that occurs during the reaction.
- Energy barrier is called the activation energy.
- Rate of product formation depends on the concentration of the transition state.

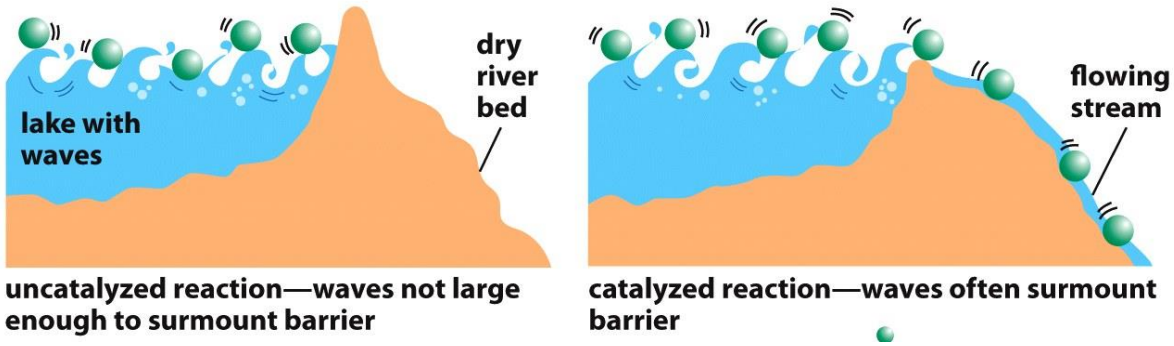
Low [X] = Slow reaction

Higher [EX] = Faster reaction

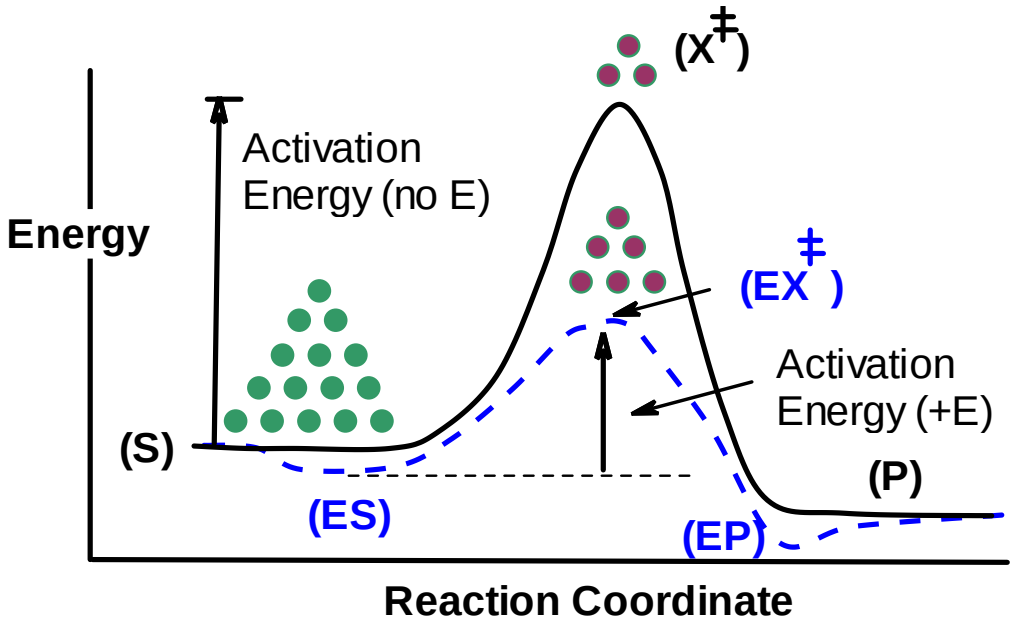
- Interactions between the enzyme and the substrate stabilize the **transition state** (X) and lower the activation energy required for the reaction to proceed.
- Stabilization can include:
  - Pre- alignment of key groups in the active site, reducing entropy cost of organizing groups.
  - Direct interactions with the transition state (see diagram)



A model of transition state stabilization.



Lower energy of transition state allows more substrates to reach transition state due to their thermal energy.



[S] = 15

[X] = 3
[EX] = 6

How much faster will the rate be when the enzyme is present?

Key Points:

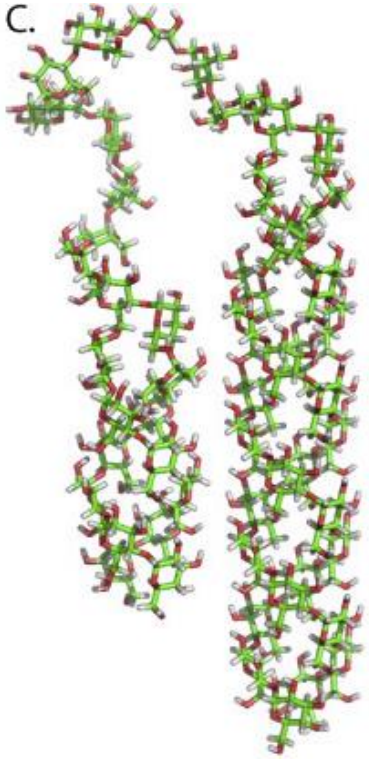
Enzymes:

Enzymes bind substrates (S), forming (ES) complex in active site, converting to P, releasing P.

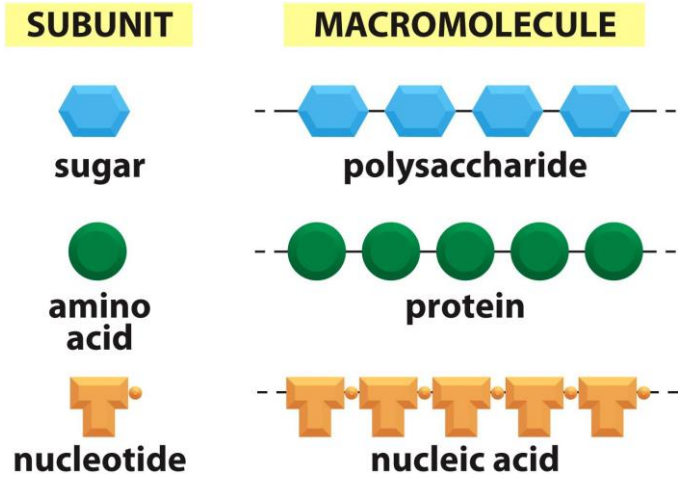
Rate enhancement since the transition state complex (EX) forms more readily with enzymes due to:

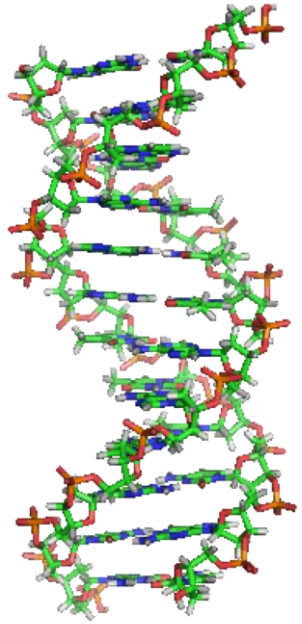
- Bringing substrates and functional groups on the enzyme together by binding (less entropy change)
- Directly lowering energy of transition state (X) through favorable interactions that are unique to the transition state, such as forming unique hydrogen bonds.

Carbohydrates



Polysaccharide



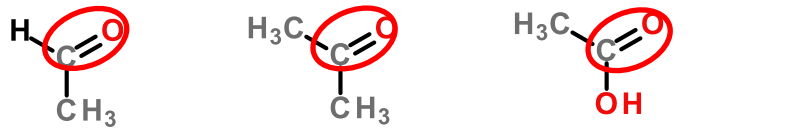


DNA (Nucleic Acid)

Carbohydrates

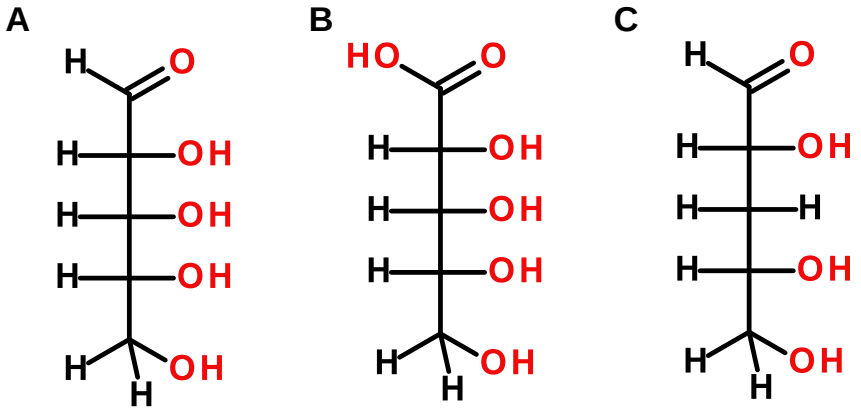
- Monosaccharides (one sugar),
- oligosaccharides (few sugars)
- polysaccharides (many sugars)
- Chemical formula is $(CH_2O)_n$ (e.g. hydrated carbon)
- They are molecules with:
  - one aldehyde or ketone group, on 1st or 2nd carbon
  - -OH group on *all* other carbons, leading to a chiral carbon for most carbons.

Functional groups:



aldehyde ketone carboxylic acid

Carbonyl group –> C=O

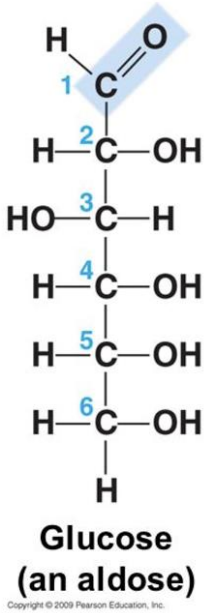


Only one of these is a carbohydrate, which one?

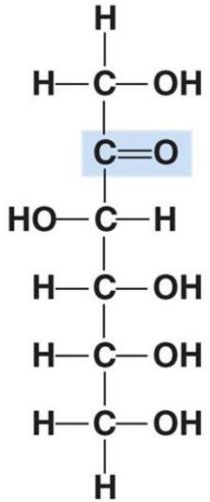
3 ways simple sugars (monosaccharides) differ from each other

1. Location of the carbonyl group
2. Number of carbons
3. Spatial arrangement of atoms (the position of the OH groups)

Carbonyl group is located on C_1



Glucose (an aldose)

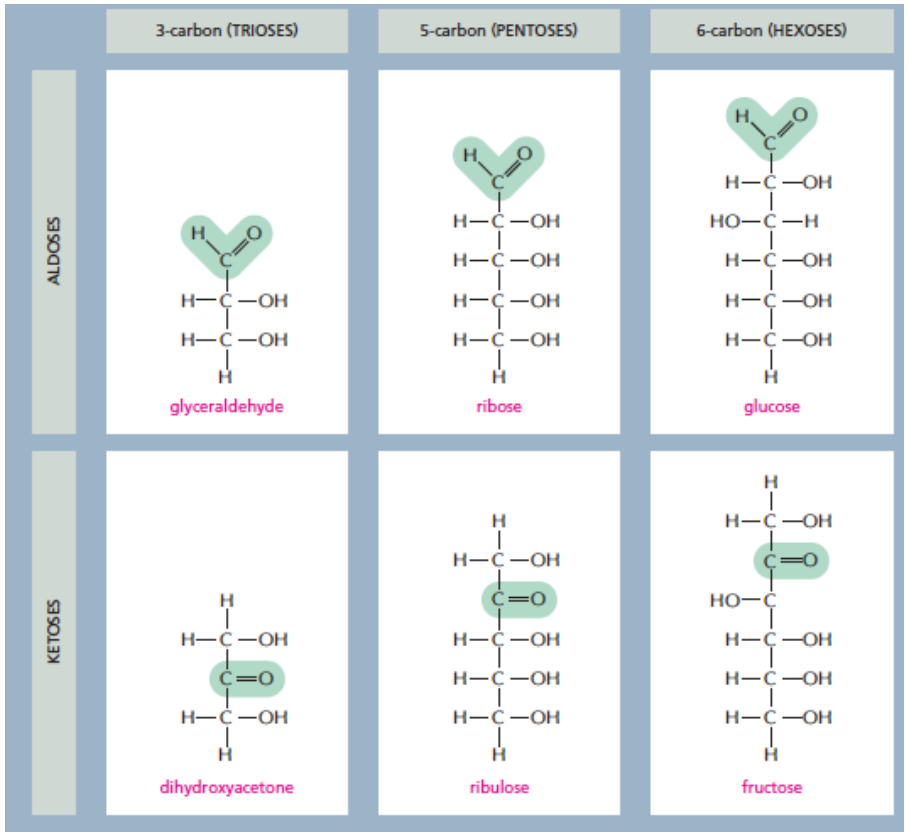


Fructose (a ketose)

Numbering carbons:
Carbon 1 is at the end closest to the C=O group.

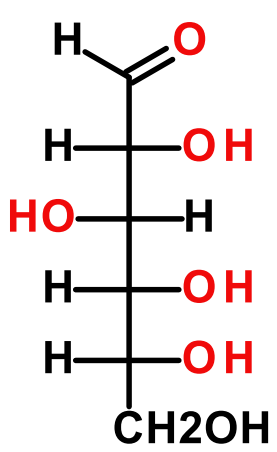
What carbon is the carbonyl?

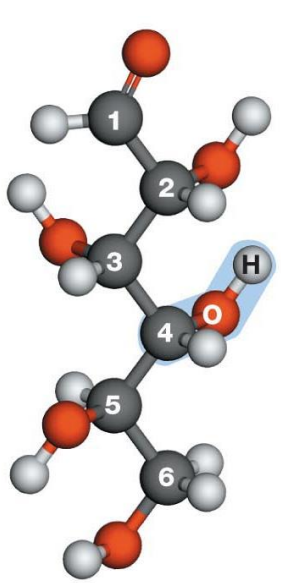
1. Location of the carbonyl group
2. Number of carbons
3. Spatial arrangement of atoms (the position of the OH groups)



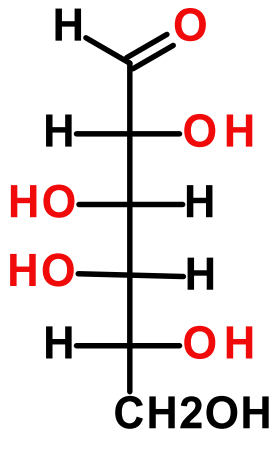
3 ways simple sugars (monosaccharides) differ from each other

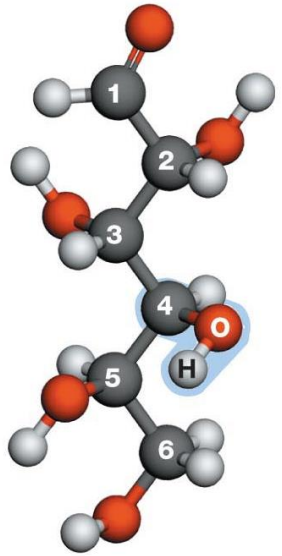
1. Location of the carbonyl group
2. Number of carbons
3. **Spatial arrangement of atoms (the position of the OH groups)**





Glucose





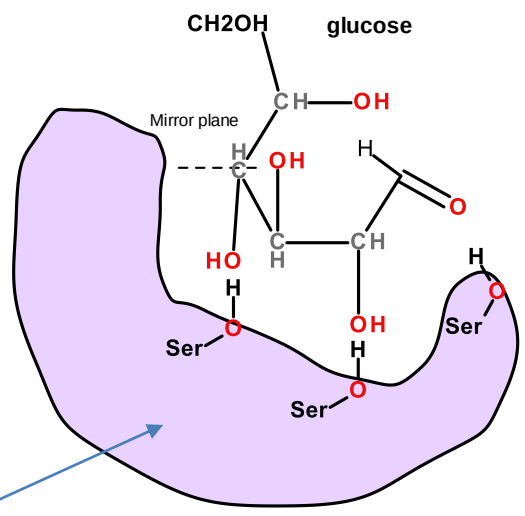
Galactose

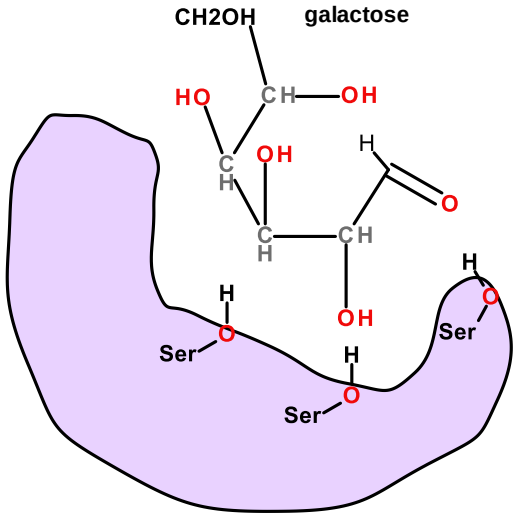
Both have the same chemical formula $C_6H_{12}O_6$.
Both are aldose sugars with 6 carbons.
Yet their functions are different.

- Glucose can be used for energy immediately.
- Galactose has to be converted to glucose before it can be used for energy.

They have different interactions with enzymes due to the different chirality at carbon 4.

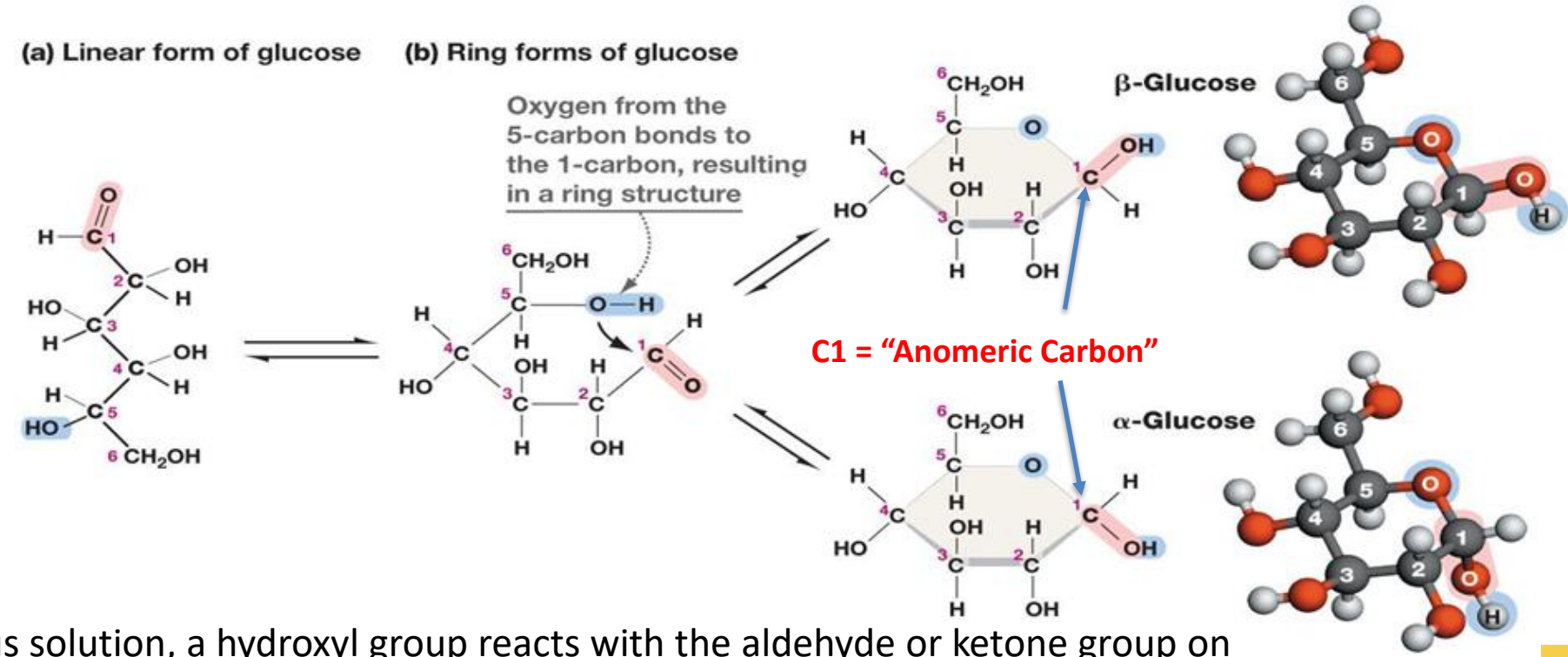
- OH is down in glucose
- OH is up in galactose





Enzyme specific for a-glucose

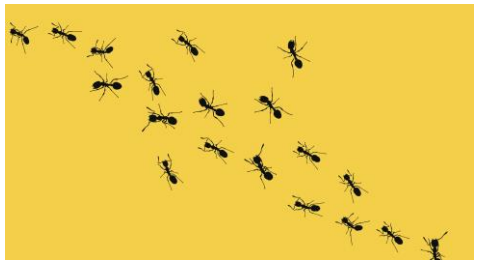
Ring formation in monosaccharides:



- In aqueous solution, a hydroxyl group reacts with the aldehyde or ketone group on the same molecule, closing the molecule into a ring, with a bridging oxygen
- Stable ring sizes are 5 atoms or 6 ato
- It is usually the 2nd to last -OH group, i.e. C5 in glucose, C4 in ribose.
- No atoms are lost or gained in this reaction.
- The carbonyl carbon becomes chiral, and is called the ***anomeric carbon*****.**
- The rings with different chirality at C1 are different:

  α (new OH is down), β (new OH is up)

  "(ants are down, birds are up)"



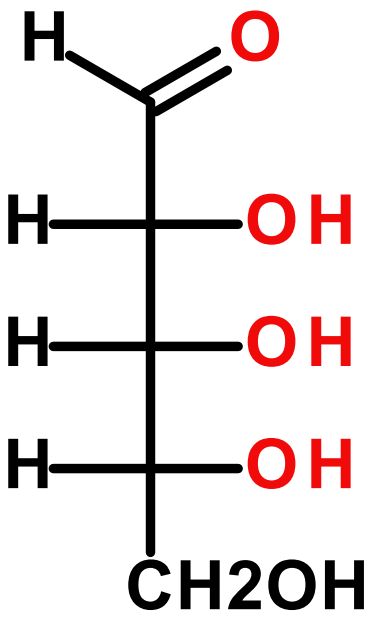
Example Problem:

The linear form of ribose, a 5 carbon aldose is shown on the right. This sugar is found in RNA (ribonucleic acid).

1. Number the carbons.

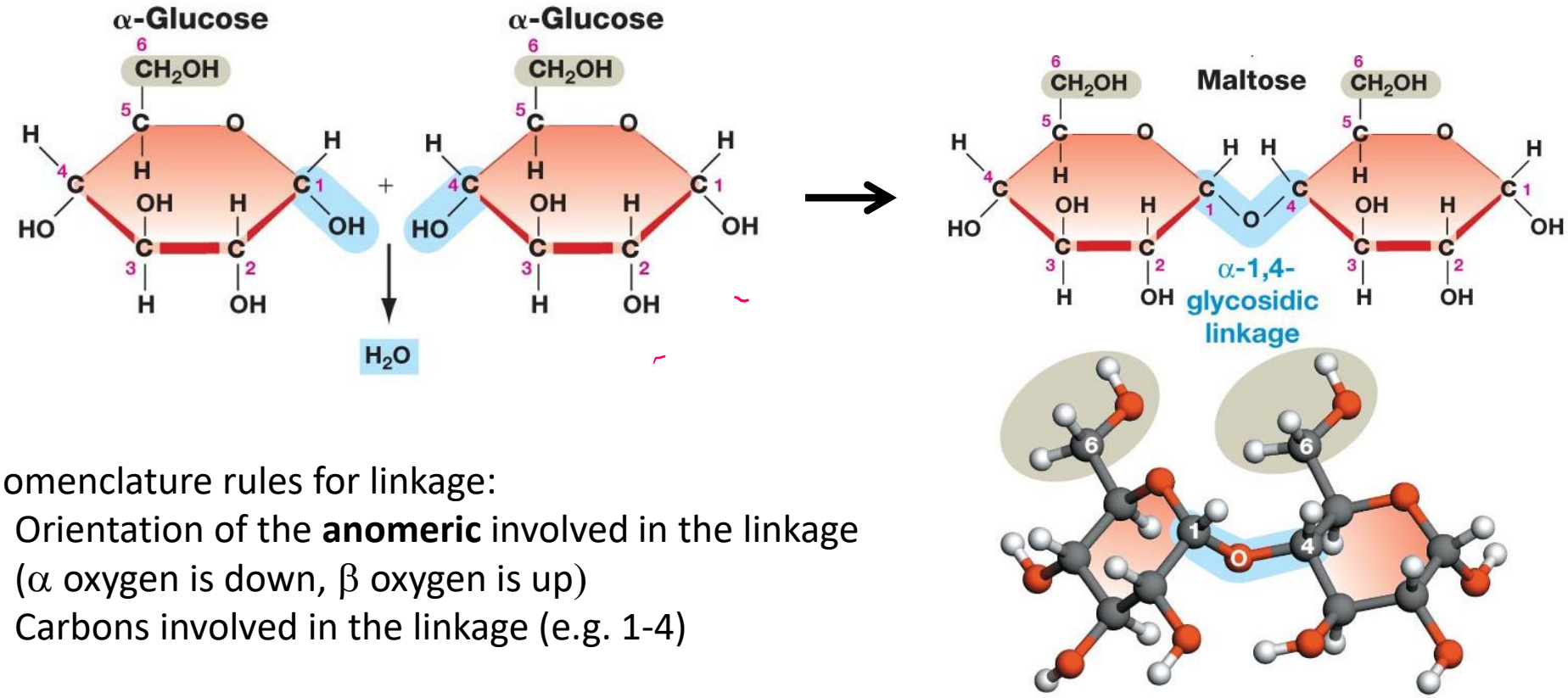
2. Which carbons are chiral? Mark them with a *.

3. Draw the cyclic form of α-ribose



Glycosidic Linkages - Disaccharides

- **Glycosidic** bond formed between **any** -OH of one sugar and the anomeric carbon of another (e.g. at least one anomeric is involved.
- Water released (dehydration reaction).

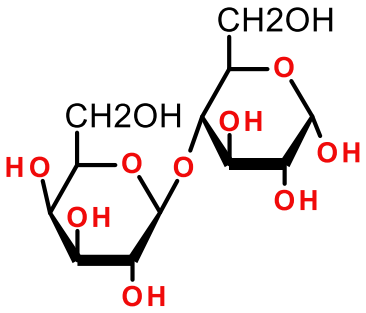


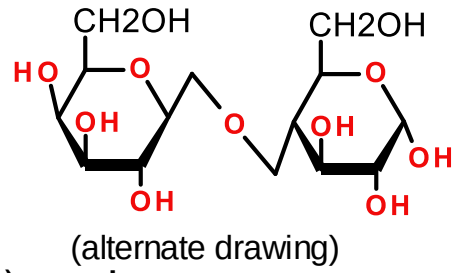
Nomenclature rules for linkage:

- Orientation of the **anomeric** involved in the linkage (α oxygen is down, β oxygen is up)
- Carbons involved in the linkage (e.g. 1-4)

Disaccharides

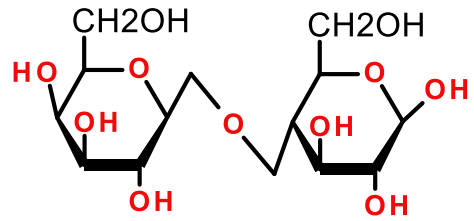
Lactose (milk sugar)





(alternate drawing)

β-galactopyranosyl-(1→4)-α-glucopyranose

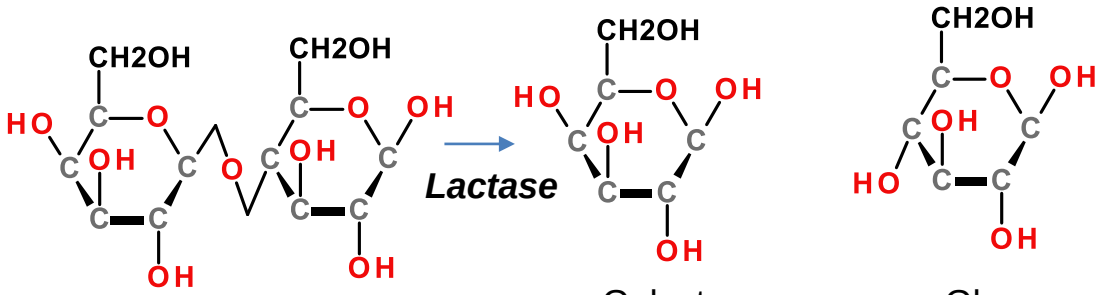


β-galactopyranosyl-(1→4)-β-glucopyranose

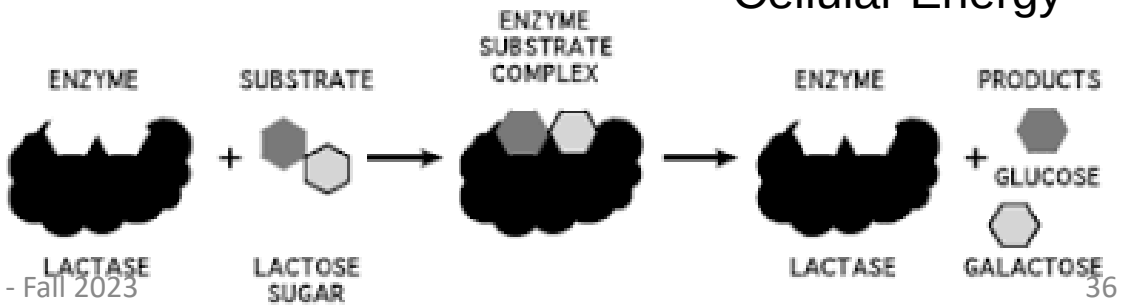
Lactose is the major sugar in mammalian milk.

- Infants produce the enzyme ***lactase*** to hydrolyze the disaccharide to monosaccharides.
- Some adults have low levels of lactase (genetically programmed). This leads to *lactose intolerance*. The ingested lactose is not absorbed in the small intestine, but instead is fermented by bacteria in the large intestine, producing uncomfortable volumes of CO_2 gas.
- It is possible to purchase dairy products that have been treated with lactase, reducing the effects of lactose intolerance.

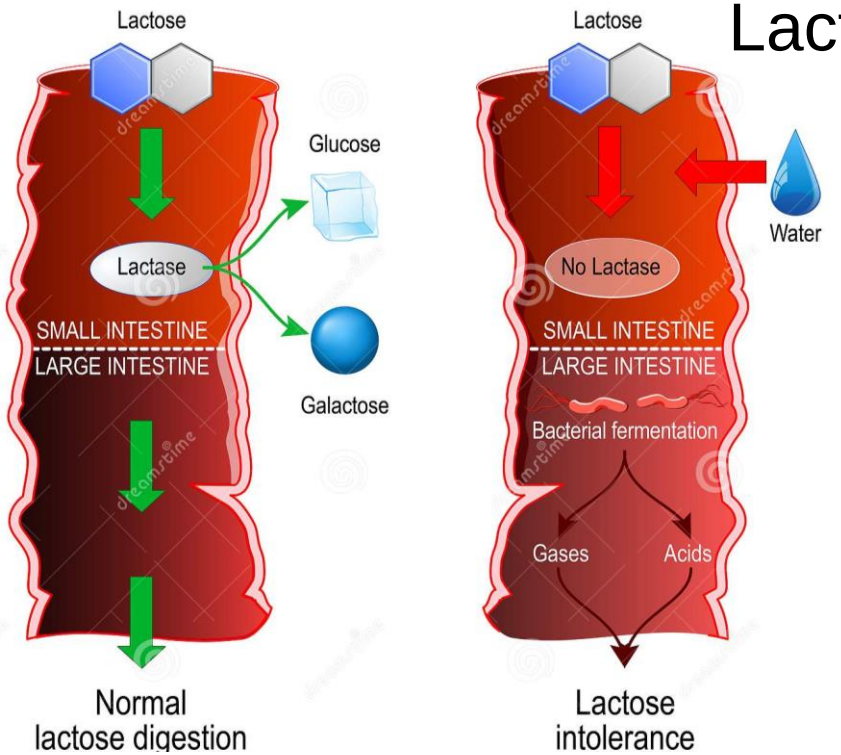
Metabolism of Lactose



Cellular Energy



Lactose Intolerance



Most individuals with lactose maldigestion can tolerate up to 12g of lactose as a single dose with no, or minor, symptoms
The European Food Safety Authority (EFSA)

Whole/low-fat Milk
200ml
9-10g lactose

Cheddar Cheese
25g
0.03g lactose

Whole Milk Yogurt
125g
6g lactose

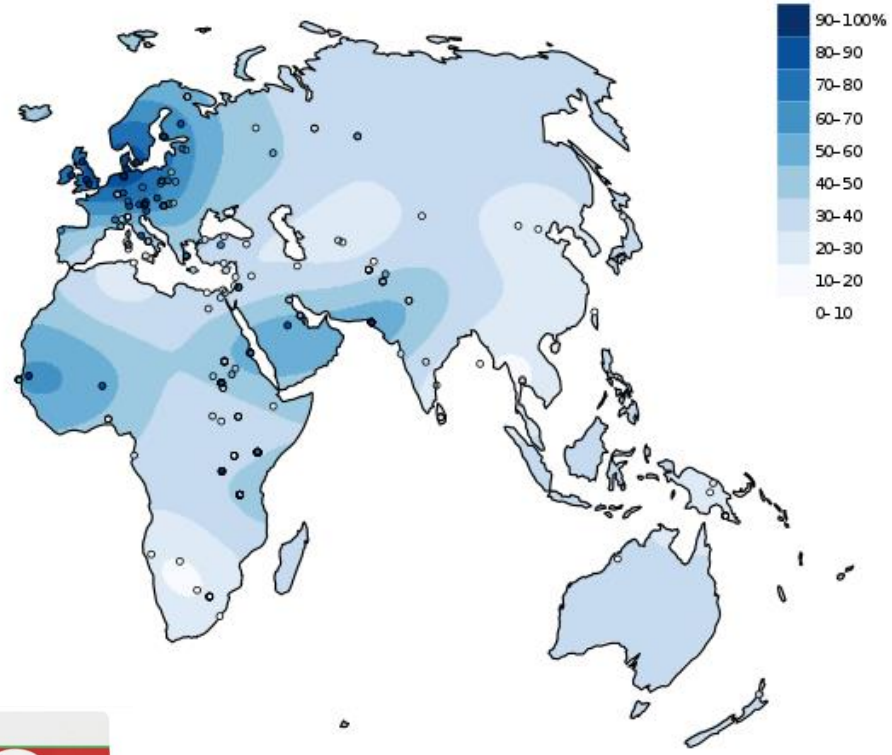
Lactose-free Milk
200ml
0g lactose

Lactase Persistence

Which region can individuals consume lactose (High lactase persistence)?

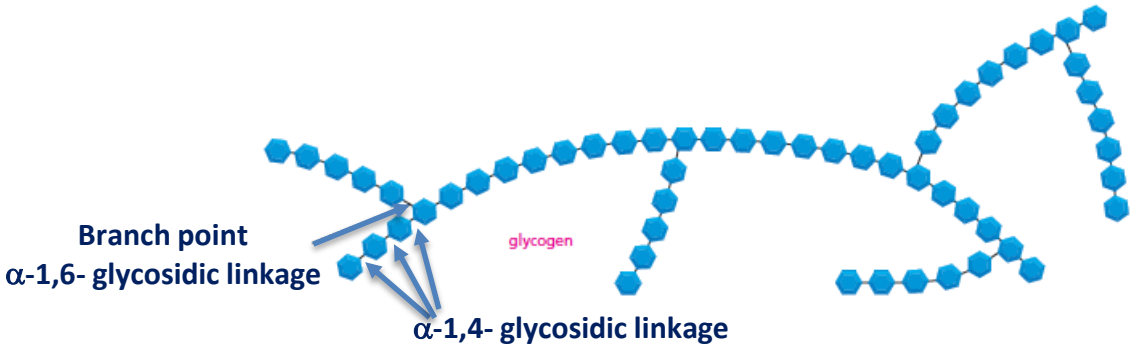
A) UK

B) South Africa



Polysaccharides as Energy Storage – Glycogen Storage Disease

Glycogen and is made entirely of glucose units and is used for glucose storage.

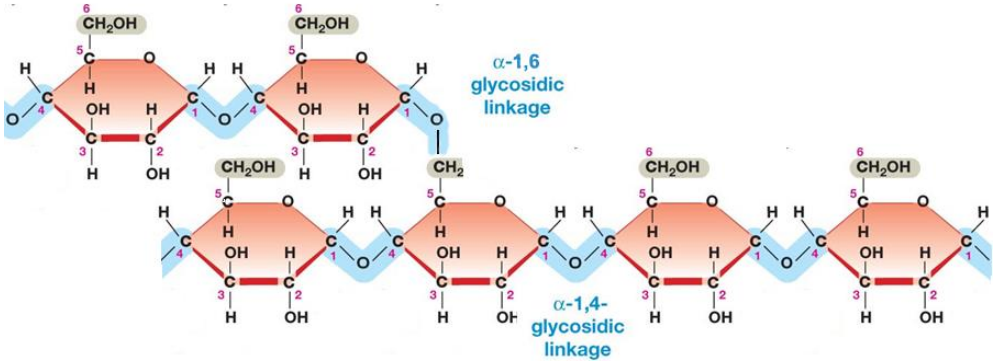


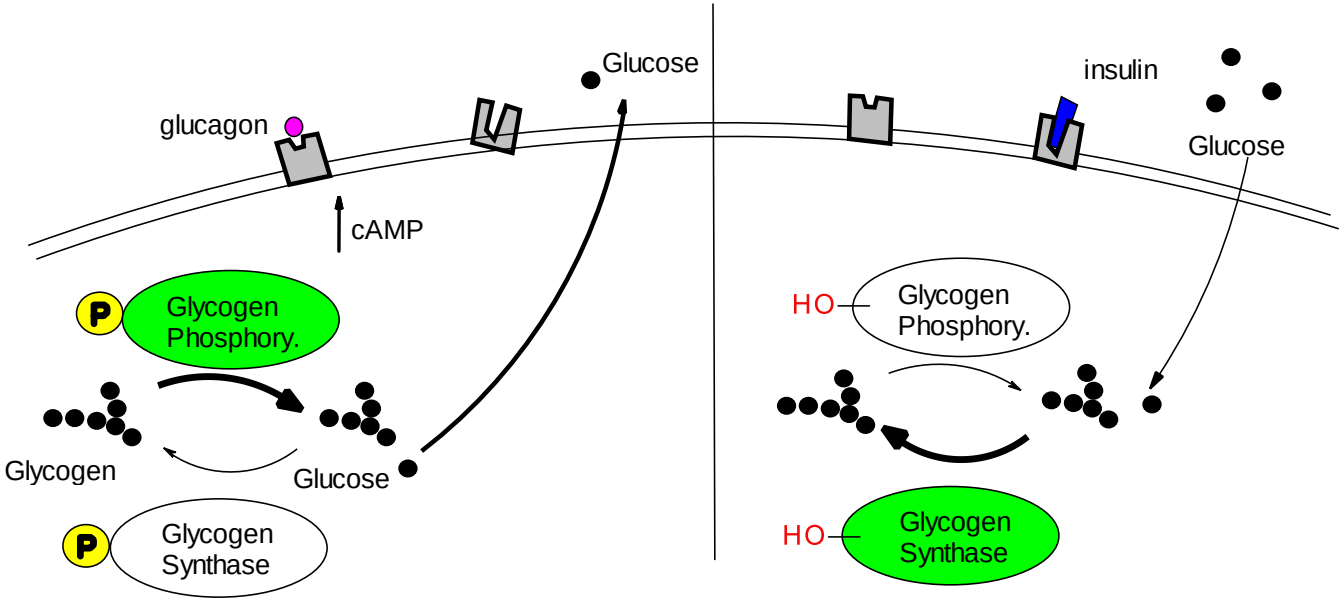
Glycogen Levels are regulated by hormones secreted due to blood glucose levels.

- Glucagon – low blood sugar
- Insulin – high blood sugar

Two enzymes degrade or synthesize glycogen

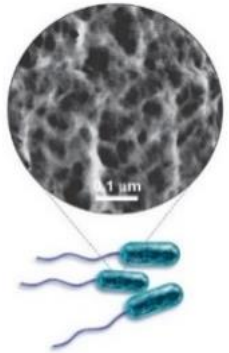
- Glycogen phosphorylase – releases glucose from glycogen
- Glycogen synthase – stores glucose in glycogen

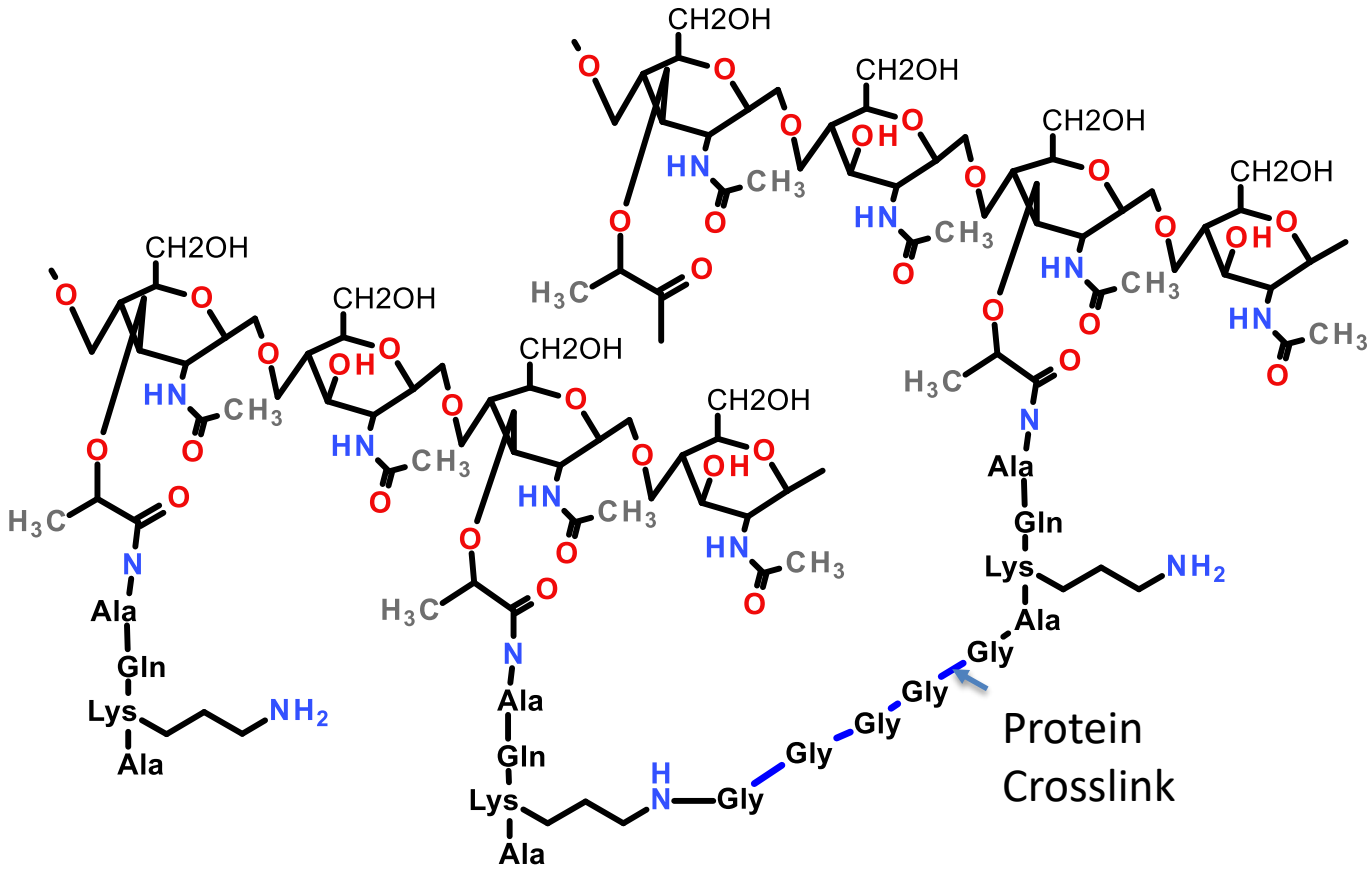




Polysaccharides as Structural Molecules

Peptidoglycan (protein + sugar) in bacterial cell wall





Peptidoglycan (Bacterial Cell Wall)

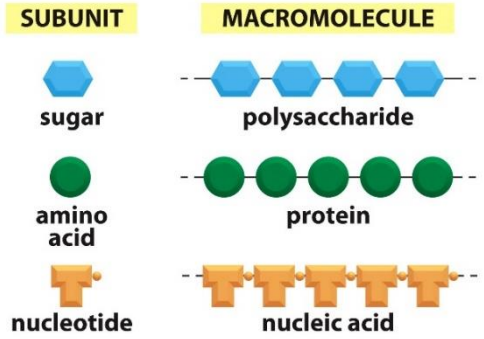
Many antibiotics interfere with cell wall synthesis (e.g. penicillin)

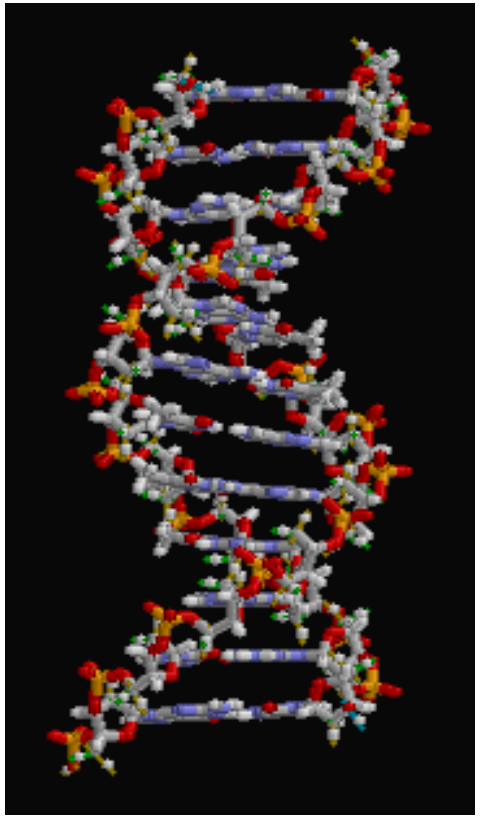
Summary and Expectations for Carbohydrates

Key Points:

- General structure of monosaccharides - be able to distinguish between aldose and ketose (and identify compounds that are not sugars).
- Know how to number carbons on aldoses and ketoses
- Be able to describe the linkage between two monosaccharides (configuration at the anomeric carbon, atoms linked)
- Be able to describe the linkage between glucose molecules in:
  - Glycogen (glucose storage)
- Be able to describe the overall structure of the peptidoglycan in bacterial cell walls.

Nucleic Acids & Central Dogma

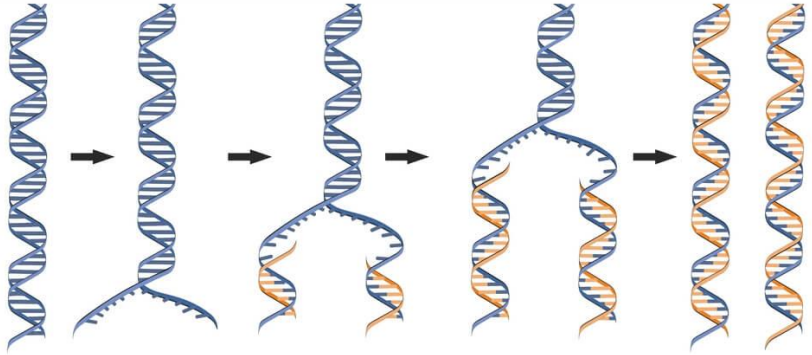




Double stranded DNA

Introduction to Central Dogma

Genome: Entire DNA content of an organism, contains all of the instructions for life. Single circular molecule in Proks, multiple linear molecules (chromosomes) in Euks. The genome is *replicated* when cells divide.

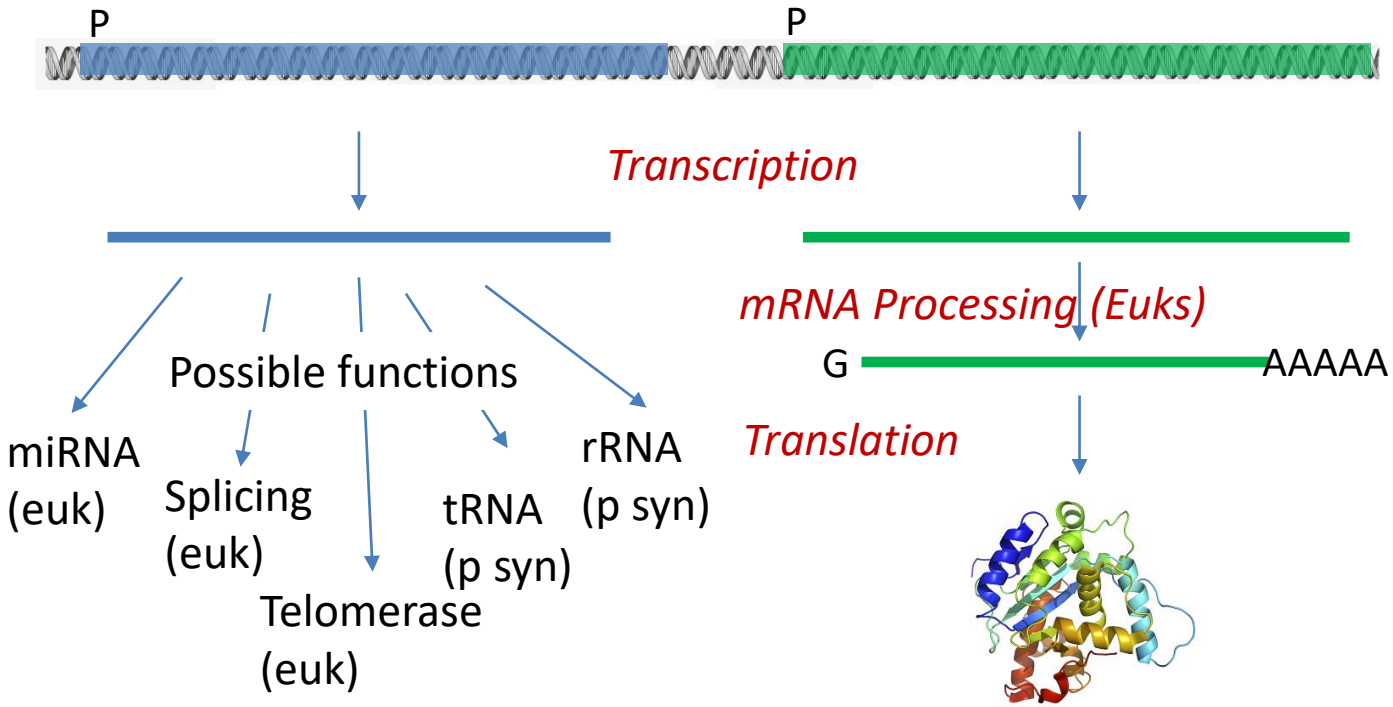


Gene – a segment of DNA that is converted (*transcribed*) to RNA. A *promoter* (P) sequence on the DNA is the minimal requirement for the production of RNA.

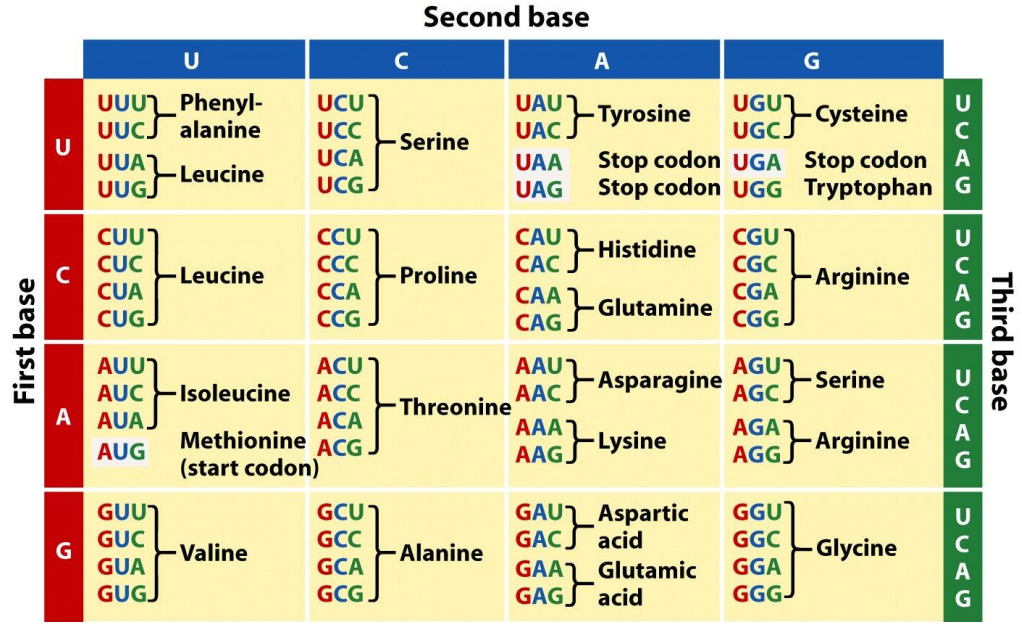
RNA molecules are often processed in **Eukaryotic cells** before they are functional

Many RNAs are functional on their own

mRNA are *translated* to a protein.



The Genetic Code



...ATATGCCCATGTGGTAA..

(DNA Sequence)

...AUAUGCCCAUGUGGUAA..

(mRNA Sequence)

..U–AUG–CCC–AUG–UGG–UAA

(Punctuated RNA sequence – how the ribosome interprets the sequence)

(Protein Sequence)

_____-_____-_____-_____

Note:

- Each codon codes for one amino acid.
- Many amino acids are coded by more than one codon.
- Most organisms use the same codon table – some codons have different meanings in some organisms.

Special Codons:

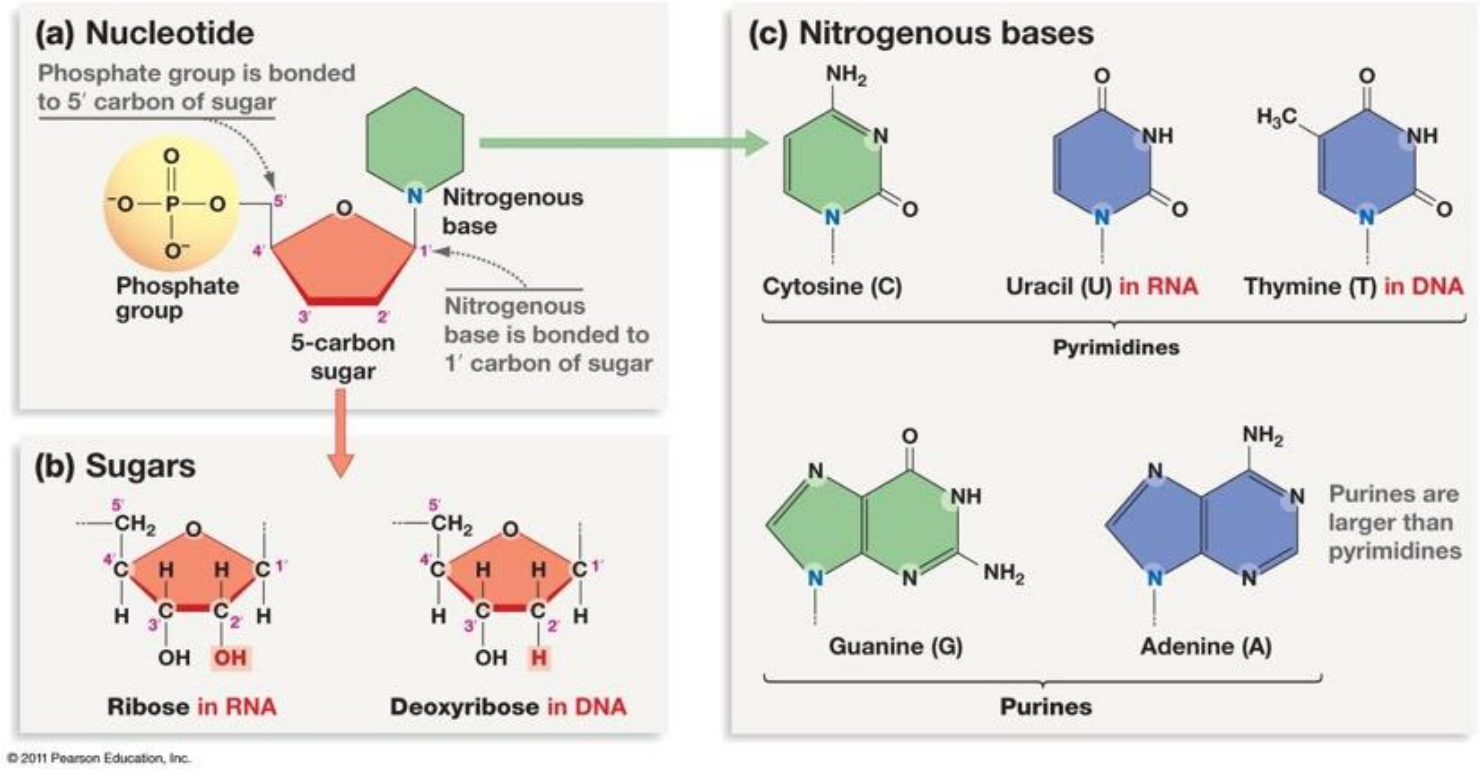
AUG = Is used to begin almost all proteins that are synthesized on the ribosome, it also codes for methionine when found internally.

UAA, UAG, UGA = stop codons (do not code for any amino acids), terminate synthesis

What is a Nucleic Acid?

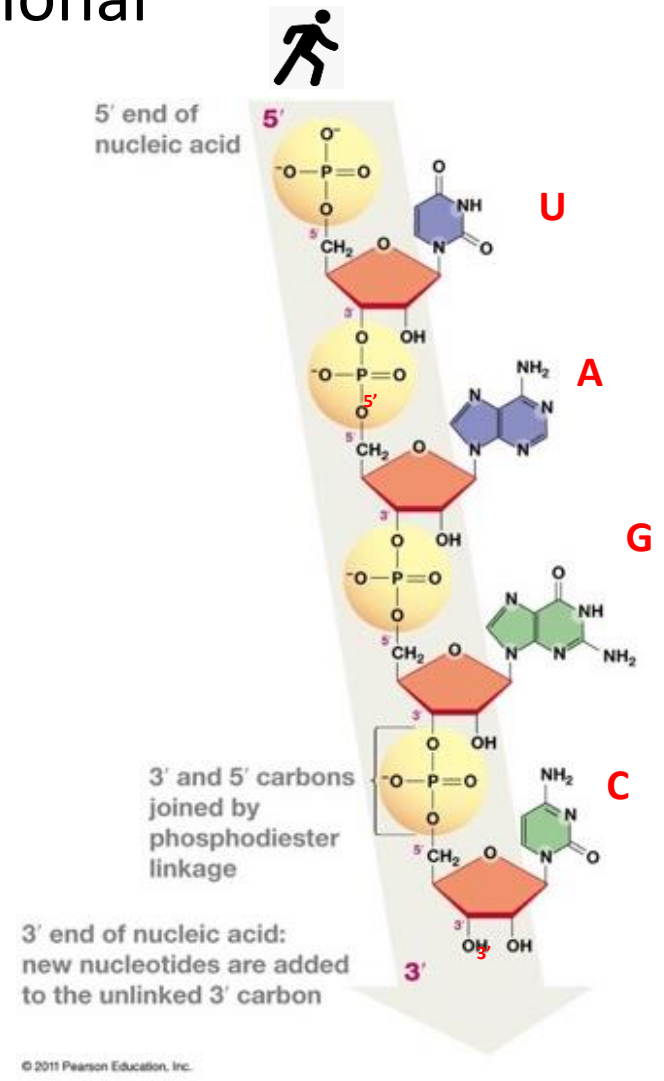
- Nucleic acids (DNA & RNA) are made up of monomers called "Nucleotides"
- Nucleotide = Nitrogenous base + five carbon sugar + phosphate
  - DNA: 2'-deoxyribose sugar + A, T, G or C
  - RNA: ribose sugar + A, U, G, or C
  - Phosphates give DNA and RNA molecules overall negative charge
  - Pyrimidines = one aromatic ring Purine = two aromatic rings

Note the carbon numbering on the ribose is 1', 2' etc.



DNA (& RNA) are directional

What are the two different ways we could write the sequence of this nucleic acid?

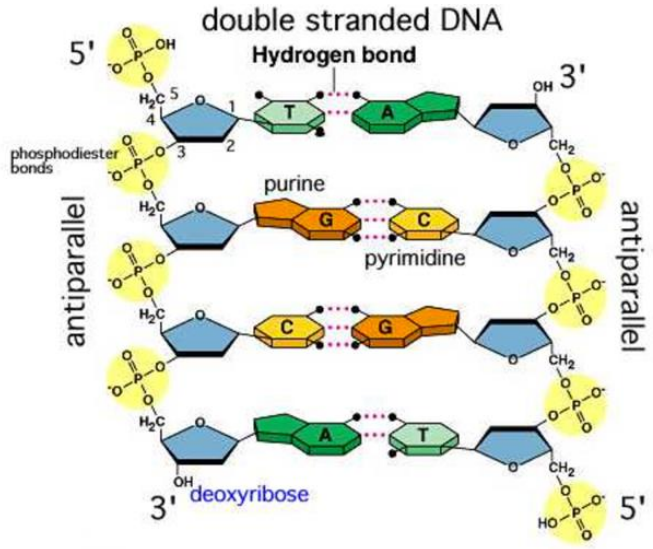


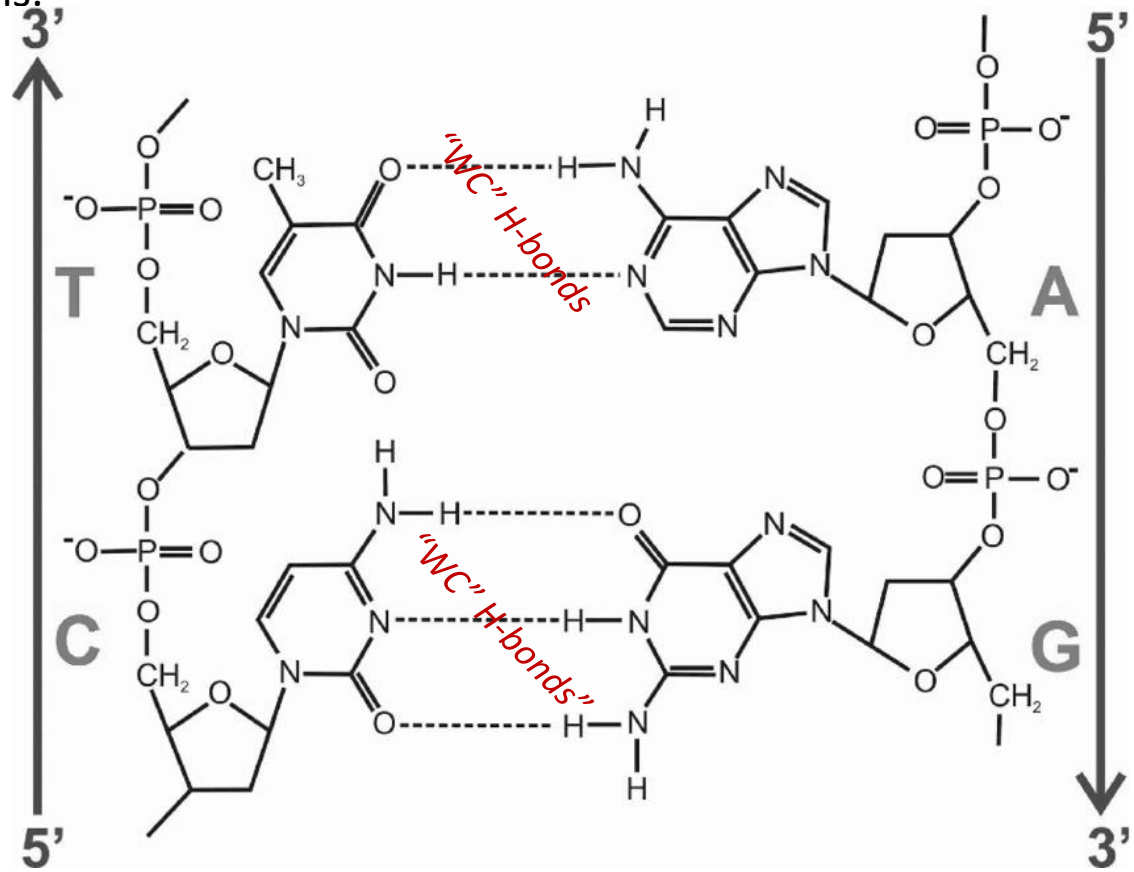
Double Stranded DNA structure

Complementary base pairing: Hydrogen bonds form between bases, thus linking the 2 stands with weak non-covalent interactions.

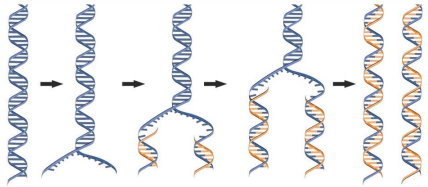
DNA twisted into double helix

- Strands anti-parallel
- Sugar-phosphate backbone outside
- Nucleotide bases project inward.
- Basepairs are stacked on each other.
- Uniform width
- H-bonds between bases:
  - A=T (two hbonds)
  - G Ξ C (three hbonds)



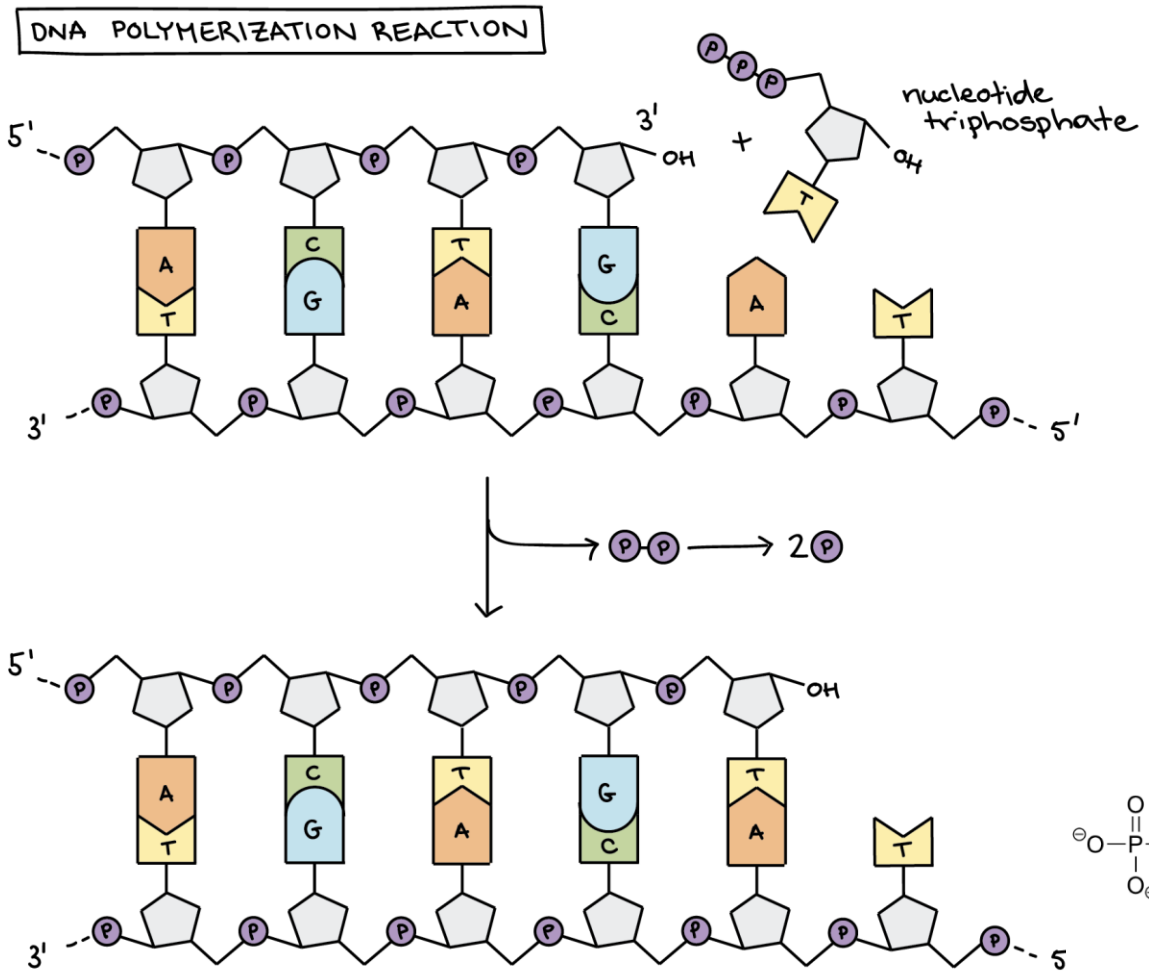


How to indicate the sequence of dsDNA?

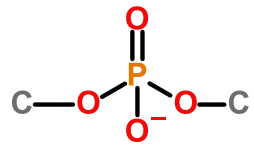


DNA Polymerase – Fundamental Activity.

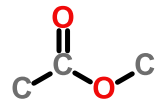
5’ to 3’ polymerization



1. *Where on the deoxyribose is the new base added?*
2. *What determines which base is to be added?*

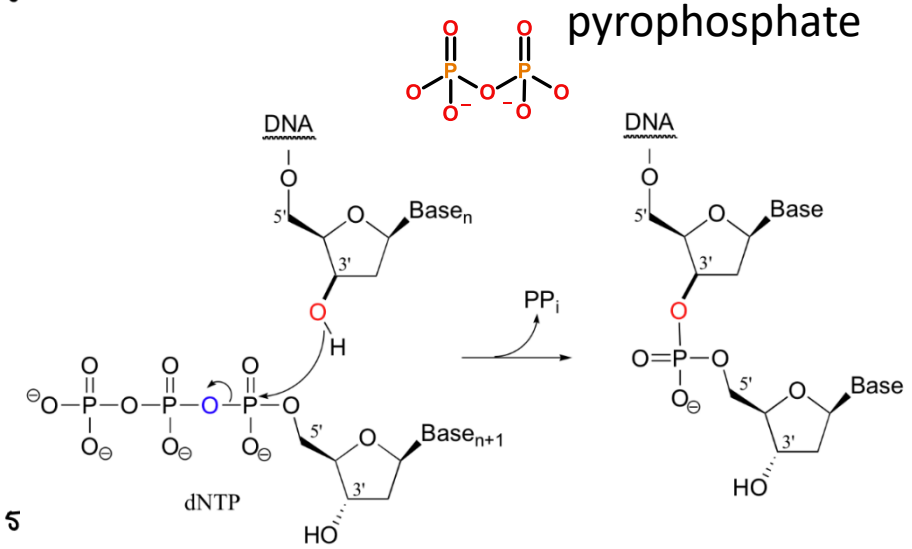


Phosphodiester linkage



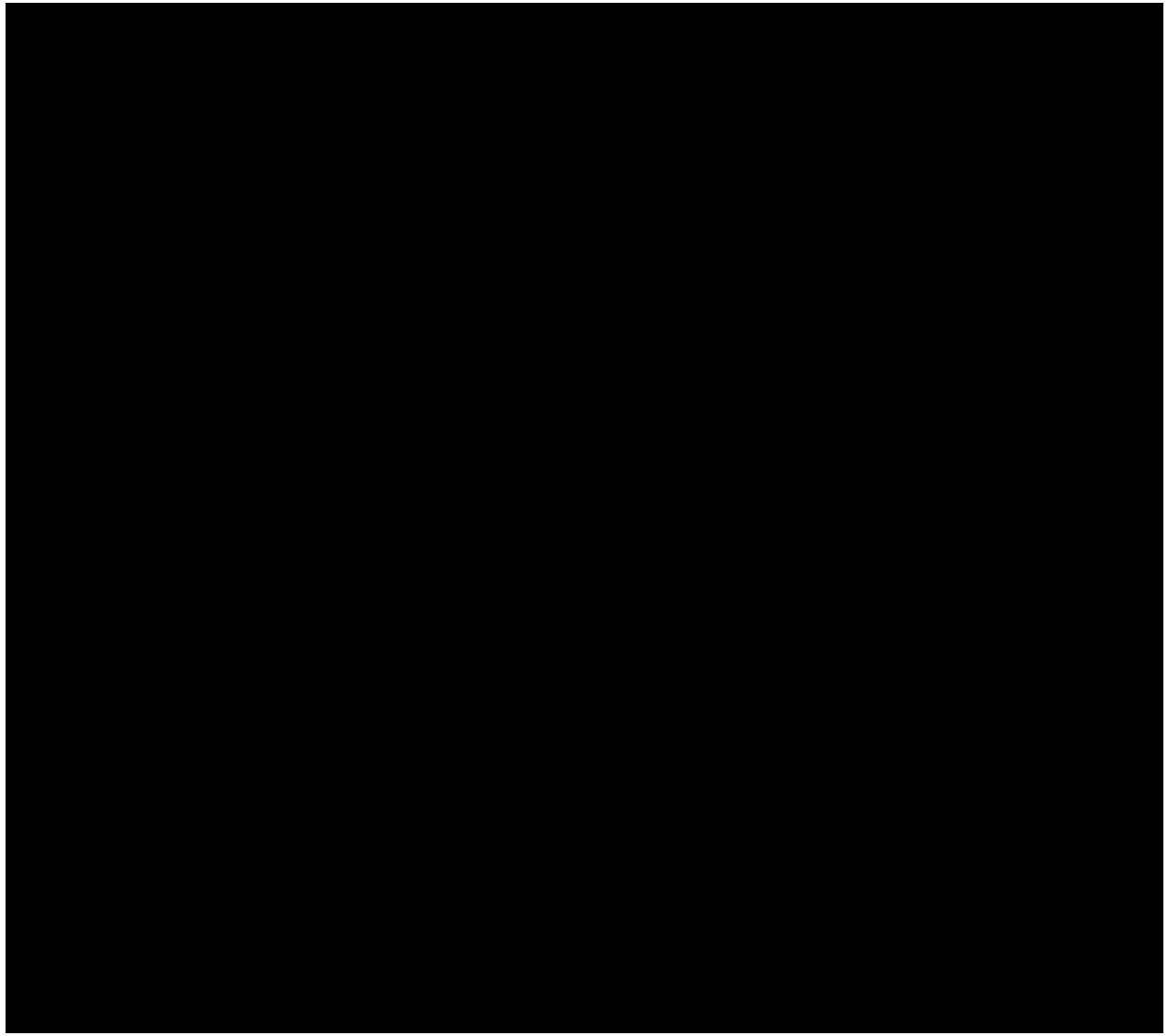
ester linkage

pyrophosphate



DNA Polymerase – Fundamental Activity.

- Synthesize new polymers of DNA.
- Require a short region of double stranded DNA to start synthesis – primer-template junction.
  - Primer can be a short DNA or RNA oligonucleotide (oligo) that is complementary to the DNA template.
  - RNA primers are used in DNA replication in the cell
  - DNA primers are used in other biotechnology applications (PCR, DNA Sequencing)
- Require single stranded template to provide information on which base to add.
- Add new dNTPs to 3'-OH of the primer, elongating in the **5' to 3'** direction.
- Elongation will go to the end of the template.



(1:48 syn starts)

DNA Polymerase – Fundamental Activity.

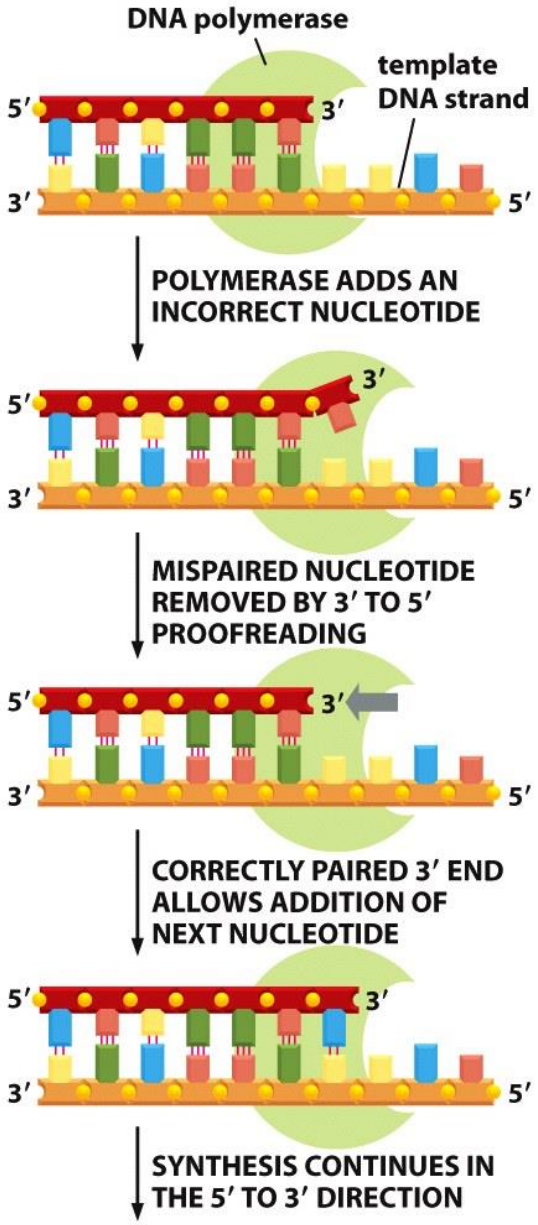
```
5′ A-T-C-A

3′A-T-G-C-C-G-T-A-G-T-C-G-T-A-C-A-G-T-A-C-G-T-G-C-A
  1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
                    1                   2
```

1. *Where (what position) will this primer (ATCA) anneal?*
2. *What base will be added first?*
3. *What is the last base added?*

DNA Polymerase – Error Correction – 3’ Exonuclease

- Incorrectly incorporated bases are removed by a 3’ exonuclease activity.
- Most DNA polymerases have this activity.
- The polymerase used by the HIV virus has no proofreading activity
- The polymerase used by Covid-19 has limited proofreading activity.

Reflection: What are the consequences of poor error correction in HIV and Covid viruses?

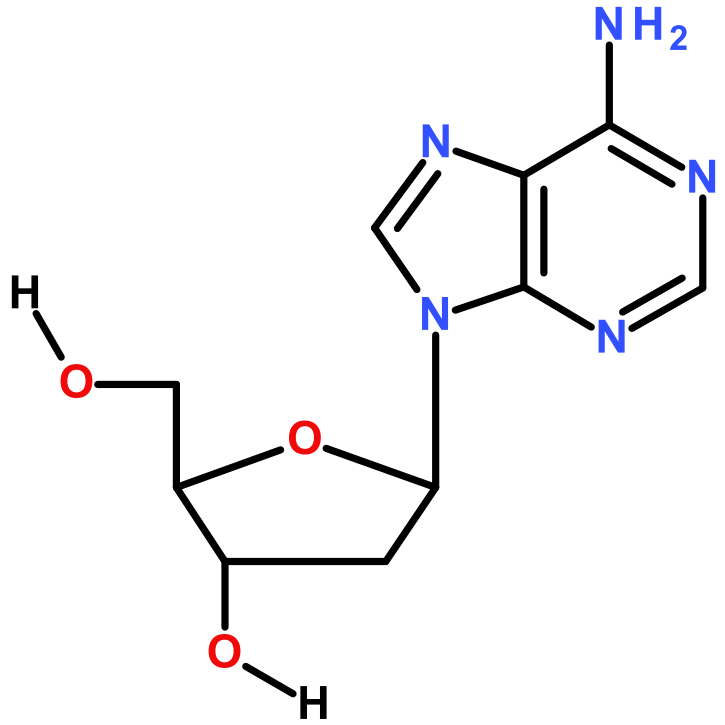


Polymerase Expectations:

1. Identify where primer anneals to the template.
2. Predict order of base addition.
3. Explain the mechanism of dNTP addition by polymerases (addition of dNTP to 3’OH, release of P-P)
4. Explain how polymerases correct errors (3’ -> 5’ exonuclease)

Exercise

A cell acquires deoxyadenosine from the environment.



A. Indicate the (three) steps that have to occur before this base can get incorporated into DNA

B. Indicate the ***two*** steps that will result in this base becoming part of a DNA strand, after the events in part A.

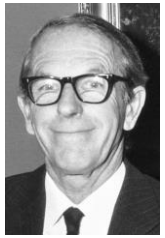
C. What will happen to DNA synthesis if the base is missing the 3'-OH?

DNA Sequencing – Sanger (dideoxy) Sequencing

DNA Sequencing - Determining the Order of Bases in the DNA.

Sanger Sequencing:

- Second method to generate long (~1000 base) sequence information (an earlier chemical method developed by Gilbert proved to be impractical for most laboratories (hydrazine = rocket fuel was required)
- Sanger was awarded his 2nd Nobel prize for this work in 1980, shared with Gilbert.



Determine the position of all four bases in a DNA strand = Sequence (video)

Sanger Sequencing:

Sequenced region (~1000 bases)

5′ ----C-T-T-C-A-G-C-T-T-A-G-T-A-A-T-C-C-G-G-T-A-C-G-T-G-C-A----

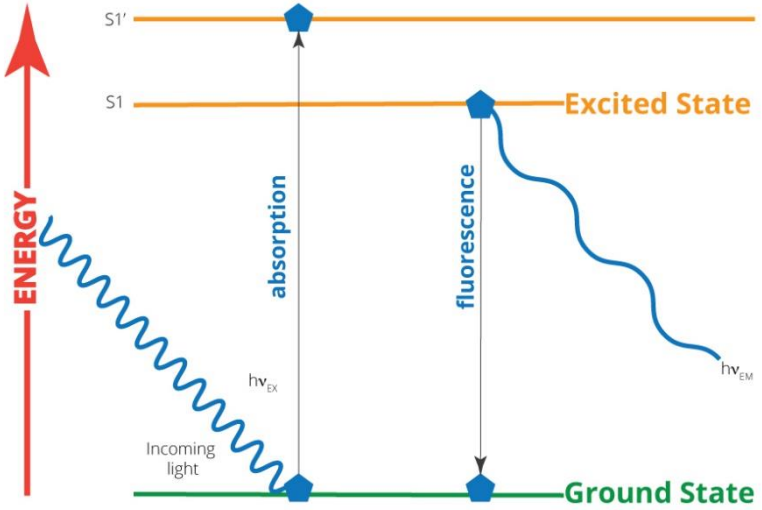
Template 3′ ----G-A-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C-C-A-T-G-C-A-C-G-T----

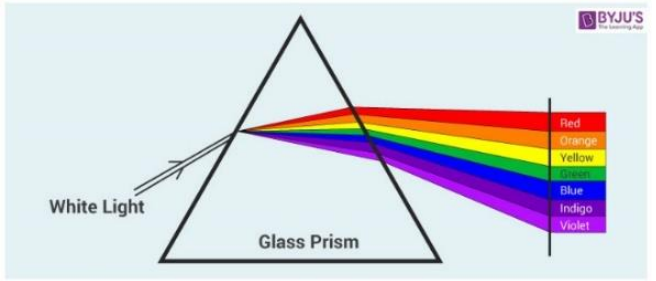
Primer 5′**C-T-T-A**^{OH}

Template 3′ ----G-A-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C-C-A-T-G-C-A-C-G-T----

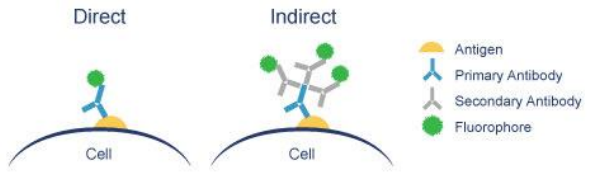
What is fluorescence?

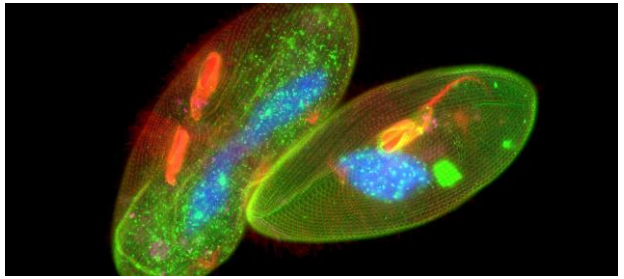
- When molecules absorb light an electron goes from a lower shell to a higher shell. This is where the energy from the light goes.
- In most molecules the electron goes back down to its original shell with the release of heat.
- Fluorescent molecules emit the energy as light of a longer wavelength (different color).
- The color that is emitted depends on the molecule.





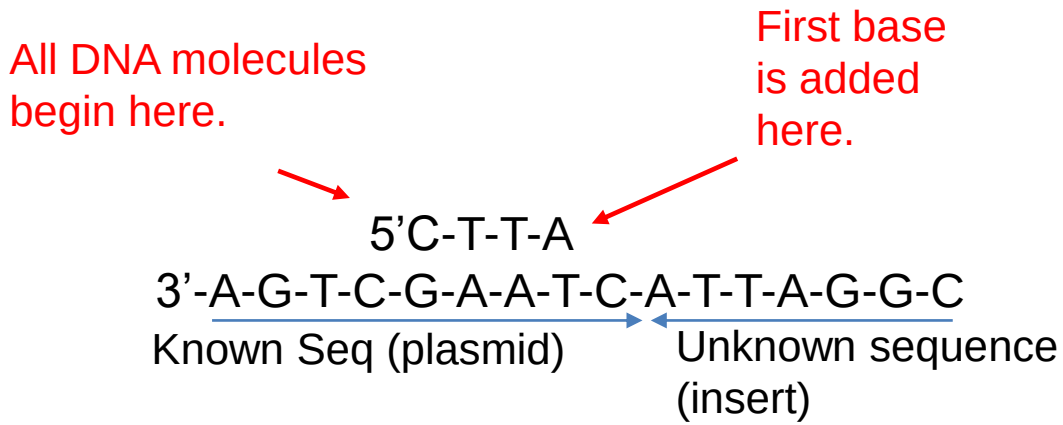
Fluorescently tagged antibodies can be used to stain components of cell with fluorophores.





DNA Sequencing - Determining the Order of Bases Added by DNA Polymerase

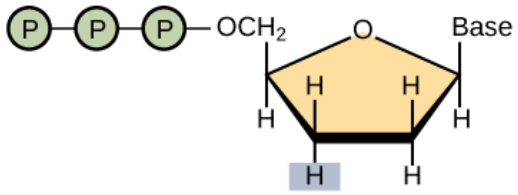
1. Start sequencing at known location with primer that anneals at a ***unique*** location on the plasmid, “upstream” from the region to be sequenced.



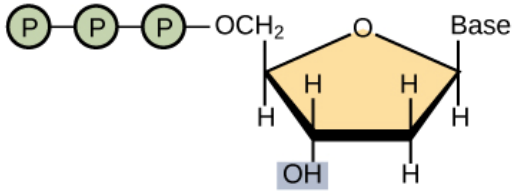
DNA Sequencing - Determining the Order of Bases Added by DNA Polymerase

2. Use a mixture of normal bases (dNTPs) and dideoxy bases (ddNTP) for polymerization. Ratio of dNTP to ddNTP is (100:1), most of the time elongation occurs.

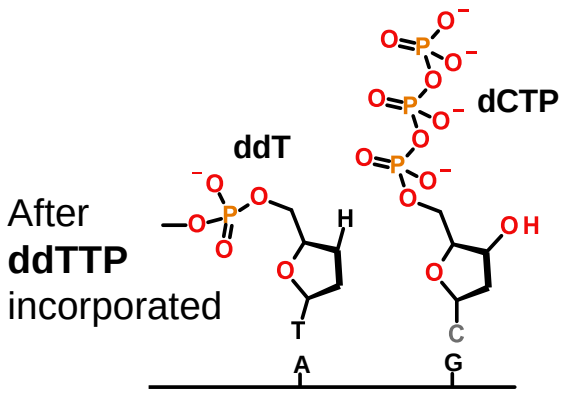
- ddNTPs can be added to the DNA since they have a 5'-triphosphate but ***terminate*** the chain due to the lack of a 3'-OH. ~ 1 in 100 chains terminate at each base addition

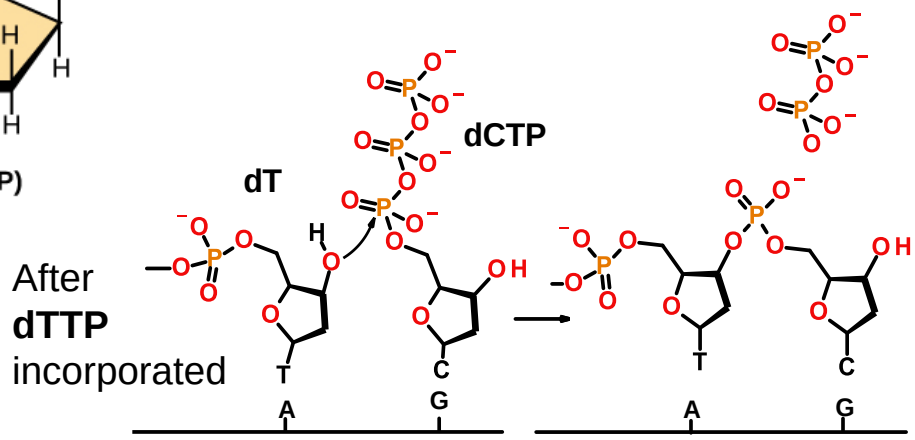


Dideoxynucleotide (ddNTP)



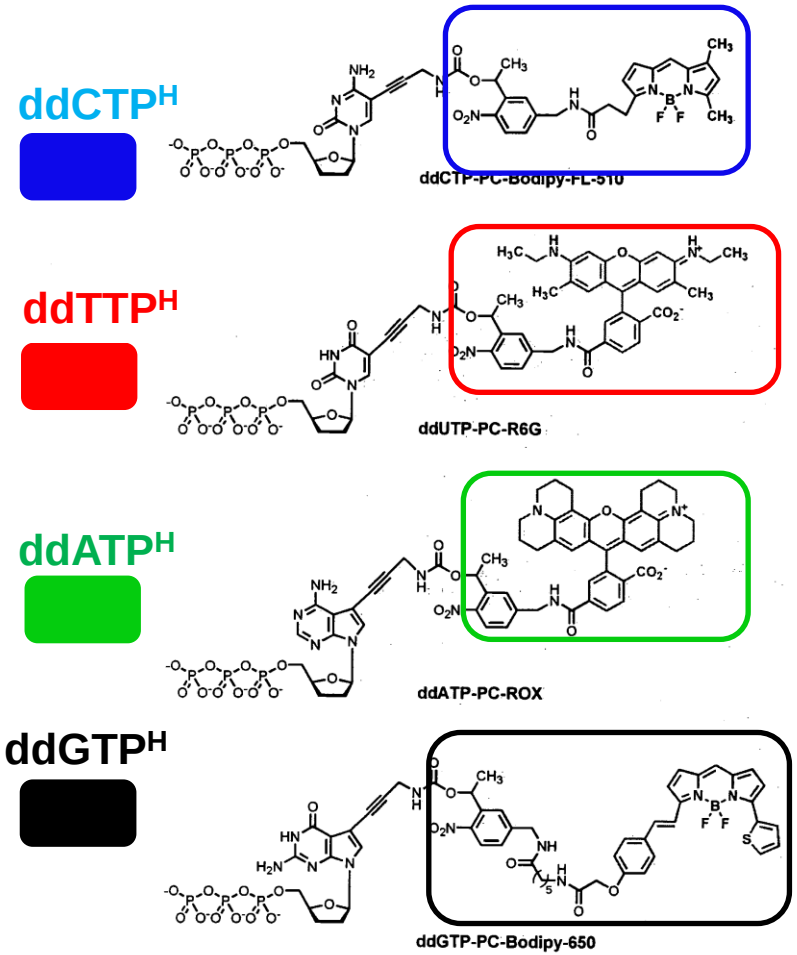
Deoxynucleotide (dNTP)



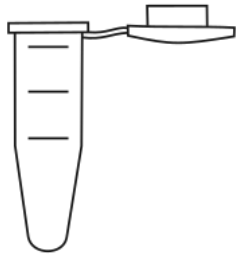


3. The ddNTPs are color coded by different fluorescent emission wavelengths.

The ddNTP that terminated the chain is known from its fluorescent color.



DNA Sequencing – Generation of Fragments



Template
Primer
DNA Pol
dTNP, ddNTP

5’-C-T-T-A
3’-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C
(1000 molecules)

Length=5, Black color

(10) 5’-C-T-T-A-**G**^{H}
3’-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C

(990) 5’-C-T-T-A-G^{OH}
3’-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C

5’-C-T-T-A-G^{OH}
3’-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C
(990 molecules)

Length=6, Red color

(10) 5’-C-T-T-A-G-**T**^{H}
3’-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C

(980) 5’-C-T-T-A-G-T^{OH}
3’-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C

5’-C-T-T-A-G-T^{OH}
3’-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C
(980 molecules)

Length=7, Green color

(10) 5’-C-T-T-A-G-T-**A**^{H}
3’-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C

(970) 5’-C-T-T-A-G-T-A^{OH}
3’-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C

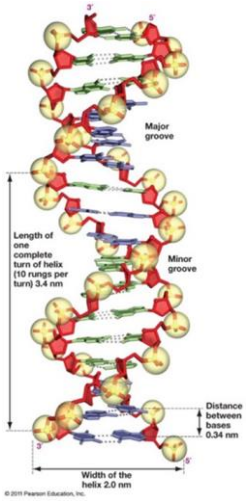
***All* Possible Fragments are Made:**

1. Each begins with the primer
2. Each ends with a *known* ddNTP, based on the color of the fluorescence.
3. Each is one longer than the previous.

C-T-T-A-**G**
C-T-T-A-G-**T**
C-T-T-A-G-T-**A**
C-T-T-A-G-T-A-**A**
C-T-T-A-G-T-A-A-**T**
C-T-T-A-G-T-A-A-T-**C**
C-T-T-A-G-T-A-A-T-C-**C**
C-T-T-A-G-T-A-A-T-C-C-**G**

Primer Added by Pol.

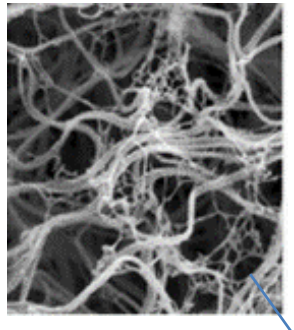
Size Determination of Fragments from DNA Sequencing
Capillary Electrophoresis

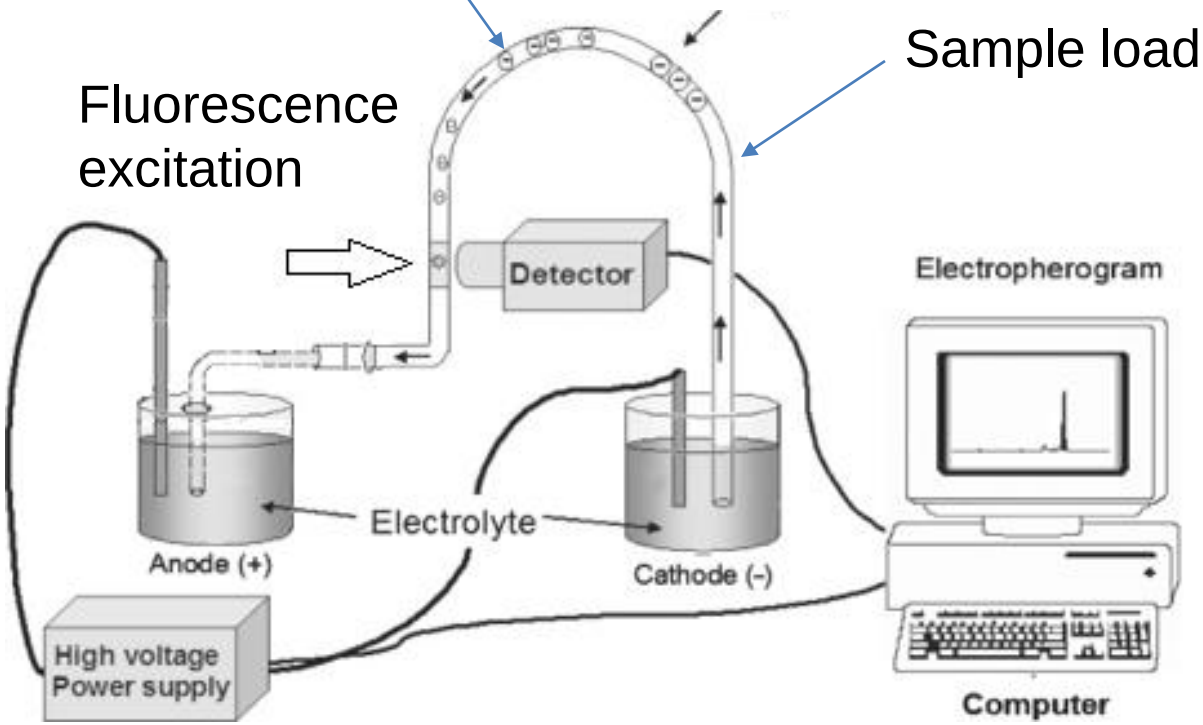


DNA has a negative charge.
It will migrate towards the anode.

Capillary is filled with a gel that causes separation by size.

DNA molecules that are smaller migrate _____.

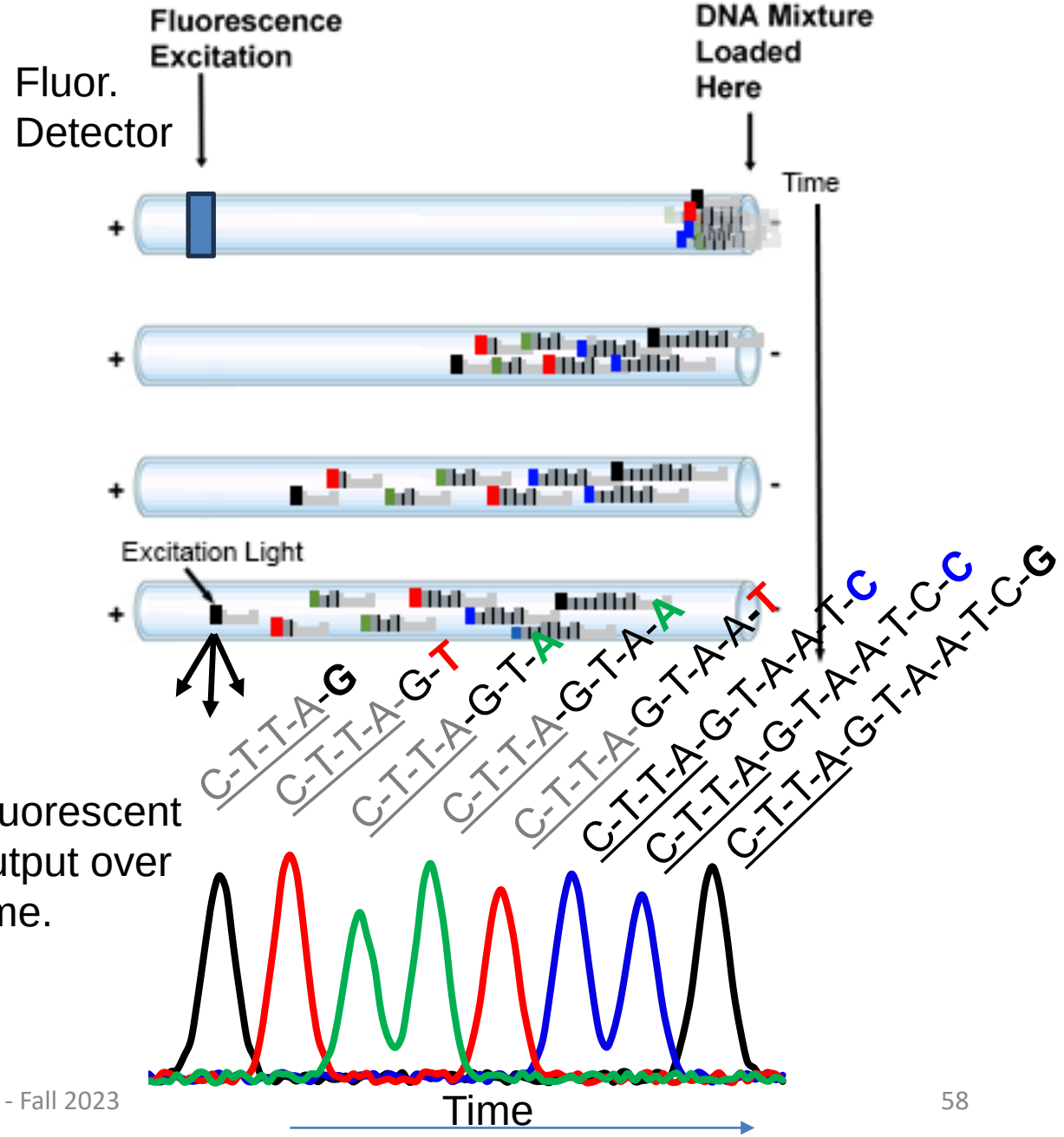




DNA Sequencing – Analysis of Fragments to Determine Order of Addition

4. Capillary Gel Electrophoresis

- Migration due to the voltage because of the neg. charge on DNA phosphates
- Separation of DNA molecules by size, smaller travel through gel faster.
- Fragments reach the detector in the order of their length: primer+1 first, primer+2 second, etc.
- At the detector, a laser excites the fluorescence.
- Only fluorescent DNA molecules (terminated with ddNTP) give a signal.
- The color of the emitted fluorescence gives the dideoxy base at the 3' end of the DNA fragment.
- The order of peaks gives the sequence that is complementary to the template (= strand with primer).

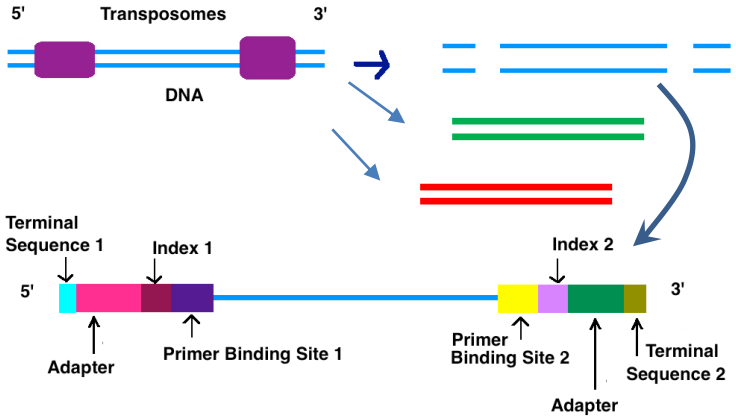


5'C-T-T-A G-T-A-A-T-C-C-G

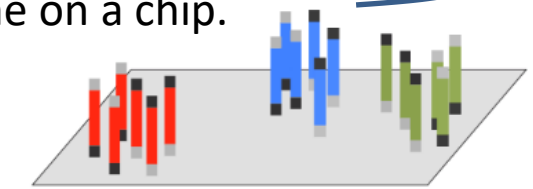
3'-A-G-T-C-G-A-A-T-C-A-T-T-A-G-G-C---

Newer Sequencing Methods-Illumina Dye Sequencing – Next Generation High Throughput

A. Obtaining the DNA

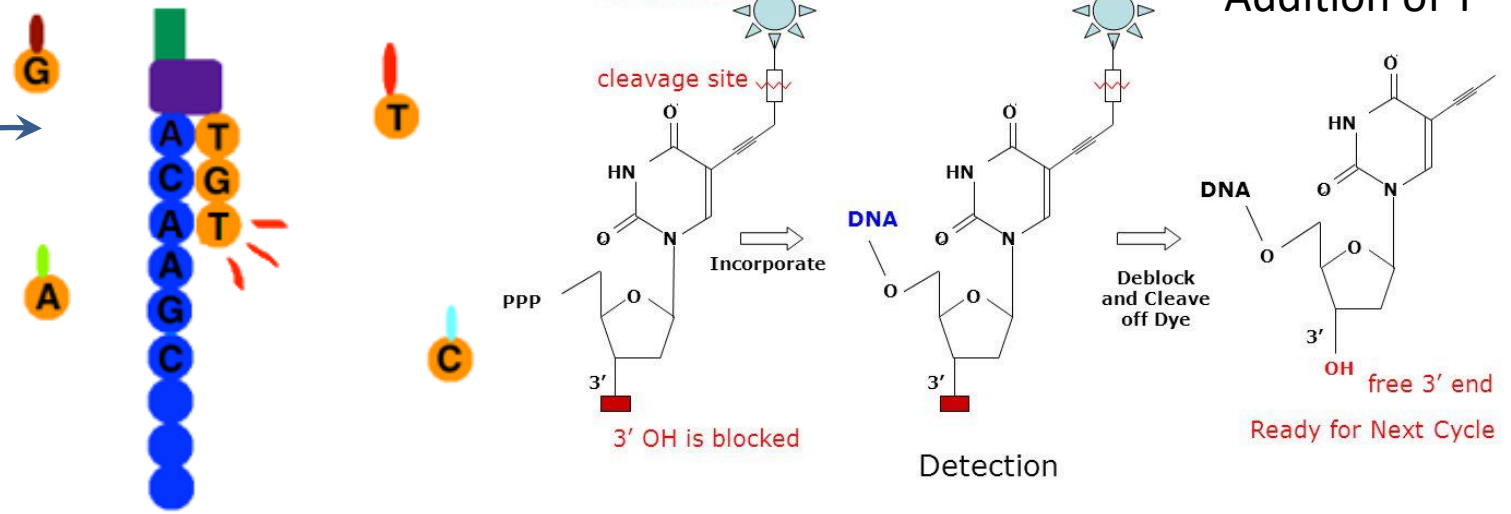


- The entire genome can be sequenced.
- The DNA is fragmented into small 100 base pieces.
- Synthetic DNA is added to the ends (sites for primers for sequencing)
- Different fragments are bound to different location on a chip.
- All fragments are sequenced at the same time on a chip.



Cluster formation

B. Sequencing by synthesis – Fluorescent labeling & *reversible* 3'-OH blocking



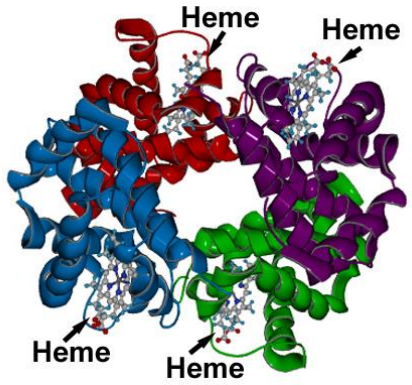
1. Only one base is added at a time (3'-OH is blocked)
2. The base that is added is determined by the color of the fluorescent base.
3. 3'-OH blocking group and the fluorescent group are removed prior to the next addition. ~100 cycles can be performed.

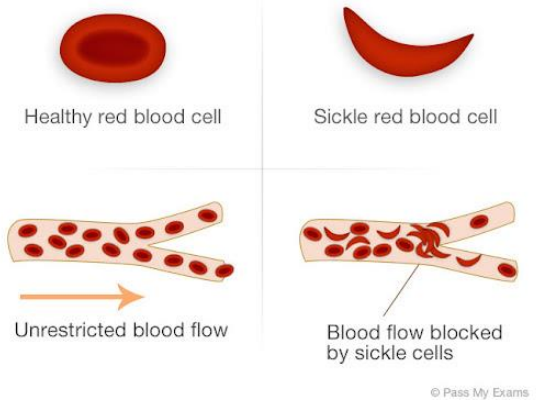
By DMLapato - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=43777596

Method	Read Length	Samples Processed
Sanger	~1000	1
Illumina	~100	~1000s

Genotyping at the Molecular Level with DNA Sequencing.

- Sickle cell anemia is caused by a single mutation in the beta chain of hemoglobin
- This mutation causes the hemoglobin to form long polymers that distort the shape of the red blood cell.
- Determining whether someone has the mutation can be useful for treatment.





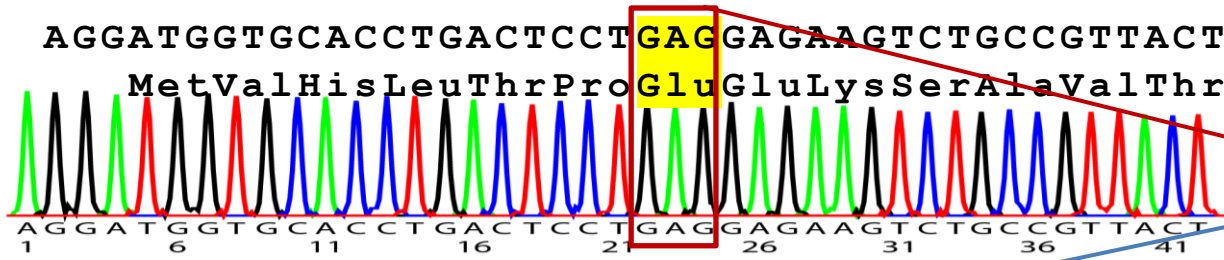
The 5' end of the Hb gene is shown on the right (ATG=start). Using GGTGCCAG as a sequencing primer gives the following sequences for the normal and mutant genes:

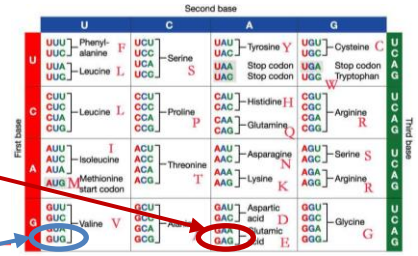
First dd-base added by polymerase

```
GGTGCCAGAGGATGGTGCACCTGACTCCTGAGGAGAAGTC . .
CCACGGTCTCCTACCACGTGGACTGAGGACTCCTCTTCAG . .
```

The sequencing data for the normal beta chain is:

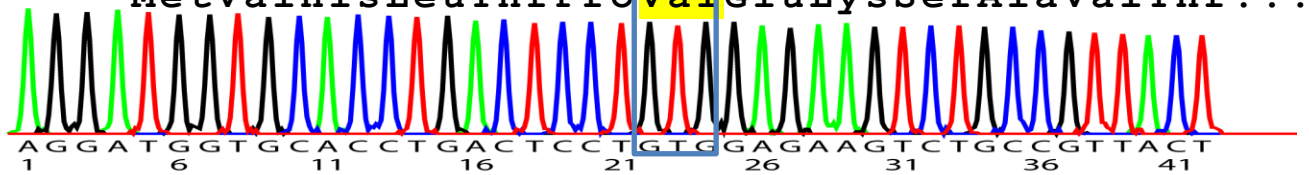
```
AGGATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACT . .
  MetValHisLeuThrProGluGluLysSerAlaValThr . .
```





The sequencing data for the mutation with sickle cell is:

```
AGGATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTTACT . . .
  MetValHisLeuThrProValGluLysSerAlaValThr . . .
```



False color code:

A=Green

G=Black

T=Red

C=Blue

Sequencing Summary & Expectations

Sanger Sequencing:

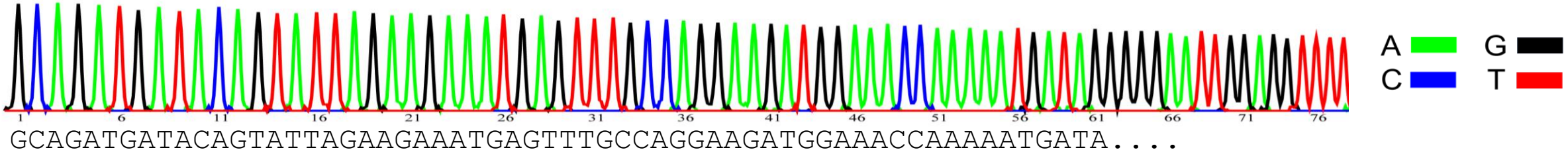
- Gives the sequence that is complementary to the template strand = “top” strand, same strand at the primer.
- The start of the sequencing information is defined by a primer that anneals to the template (therefore some of the sequence has to be known, how this is done will be described later)
- Dideoxy sequencing is carried out by adding both dideoxynucleotide triphosphates (ddNTPs) and deoxyribonucleotide triphosphates (dNTPs) to the synthesis reactions, at a ratio of 1:100. Most growing chains do not terminate.
- ddNTPs are identical to dNTPs except that they lack the 3' hydroxyl group. Because of the missing 3’-OH, DNA polymerization stops once one ddNTP is added to a growing strand.
- The type of the added base is determined by “color coding” each base.
- The location of added bases is determined by measuring the size of the DNA fragment that was terminated by the ddNTP.
- It is possible to sequence approximately 1000 bases by this method.

Next Gen-Sequencing:

- Simultaneous sequencing of a large number of fragments
- Shorter “reads” 100 versus 1000 bases/template

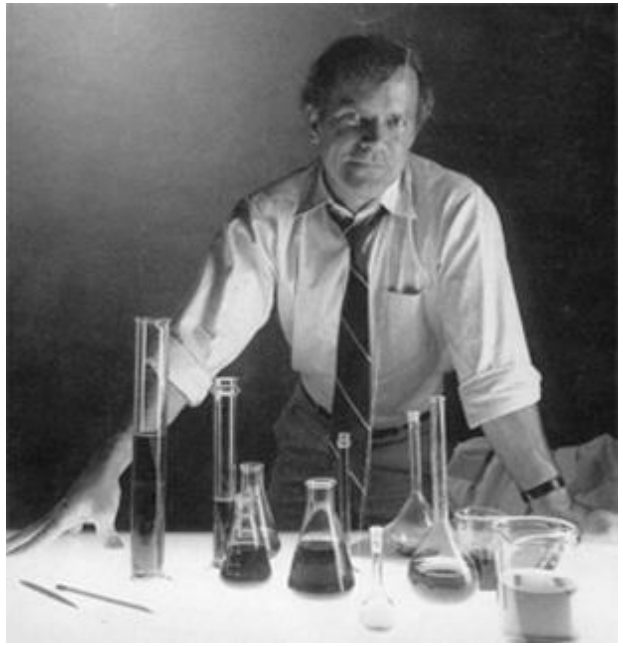
Expectations:

- Can you explain how the colored peaks are generated by random termination using dideoxybases in Sanger Sequencing?
- Can you compare and contrast major features of Sanger and Next-Gen Sequencing?



Polymerase Chain Reaction -PCR

- In 1983, Kary Mullis developed the molecular biology technique that has since revolutionized genetic research, earning him the Nobel Prize in 1993.
- PCR had an impact on four main areas of biotechnology: gene mapping, cloning, DNA sequencing, and gene detection (e.g. coronavirus).
- PCR is now used as a medical diagnostic tool to detect specific mutations that may cause genetic disease, in criminal investigations and courts of law to identify suspects on a molecular level, and in the sequencing of the human genome.

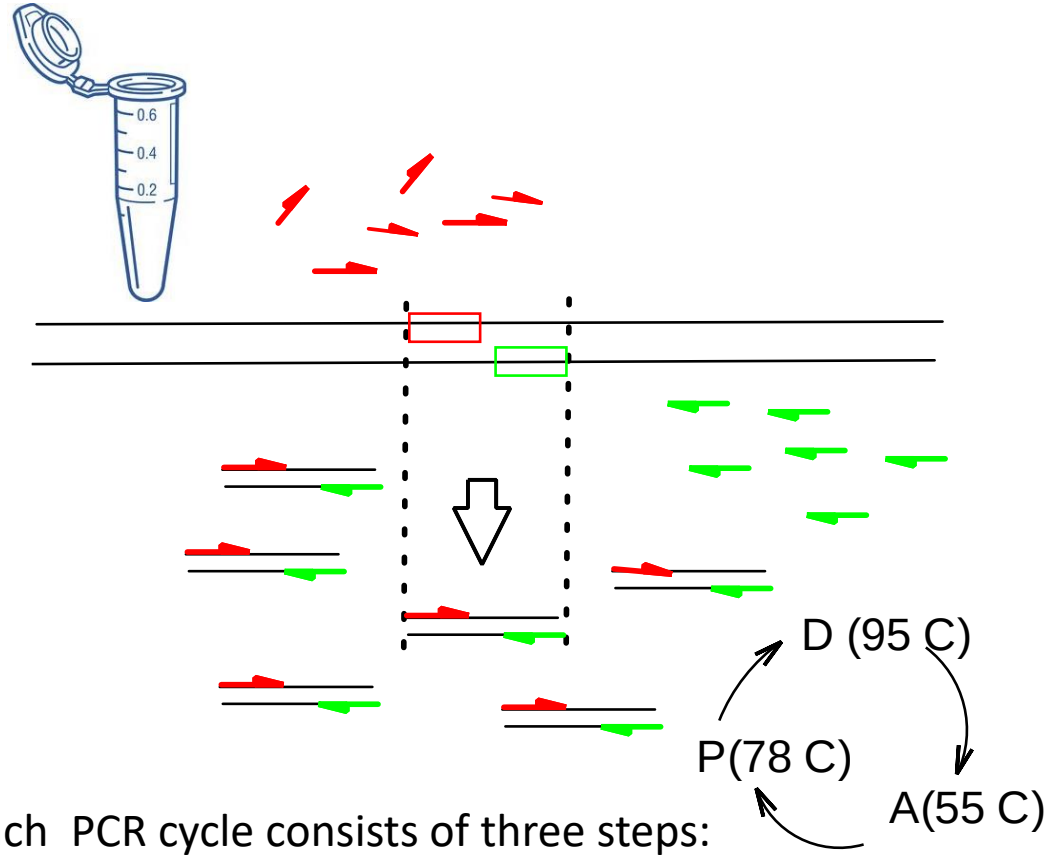


Expectations:

1. Be able to explain how PCR works to amplify a segment of DNA.
2. Be able to give the left and right primers.
3. Apply PCR approaches to determine genotype and detection of viruses.

Polymerase Chain Reaction

- PCR is an *in vitro* DNA synthesis reaction in which a specific section of DNA is replicated over and over generating exponentially large amounts of a specific piece of DNA from trace amounts of starting material (template).
- Template can be trace amounts of DNA from a drop of blood, a single hair follicle, or a cheek cell.
- The region of DNA that is copied is specified by the sequence of two primers, which are short ssDNA that initiate polymerase activity. The primers are in vast excess over the DNA.
- The location of the amplified segment is *defined* by two primers (left = upstream, right = downstream):
  - they anneal to their templates according to Watson-Crick pairing rules (A-T, G-C),
  - initiate polymerization from those sites,
  - they are incorporated into the final PCR product.
  - Left primer = sequence of top strand at left boundary
  - Right primer = sequence of bottom strand at right boundary
- The primers are DNA and are synthesized chemically, they can be any desired sequence.
- If there is no homology between the primers and the input DNA, then no PCR product will be formed.



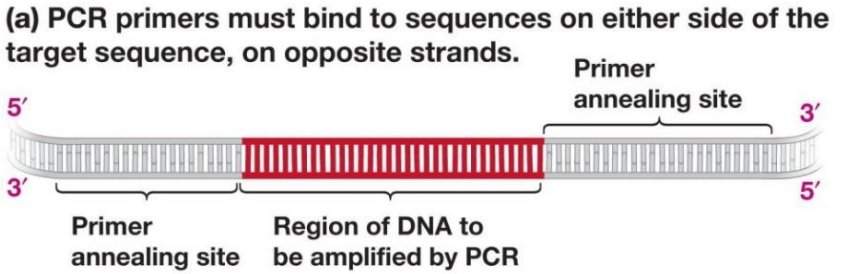
Each PCR cycle consists of three steps:

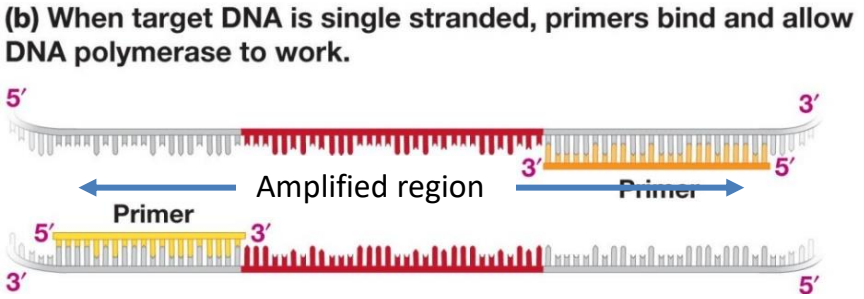
1. Denaturation of the DNA to make it single stranded (2 min at 98 C)
2. Lowering of temperature to let the primers form double-stranded DNA (1 min at 55 C)
3. Elongation by DNA polymerase (1 min/kb at 78 C)

PCR – Primer Design

- Before a region of DNA can be amplified, one must identify and determine the sequence of a piece of DNA upstream and downstream of the region of interest.
- These areas are then used to determine the sequence of oligonucleotide primers that will be synthesized and used as starting points for DNA replication.
- Primers are complimentary to the up- and down-stream regions of the sequence to be amplified, so they stick, or anneal, to those regions.
  - Left primer = sequence of top strand on the left. This primer will anneal to the bottom strand.
  - Right primer = sequence of bottom strand on the right. This primer will anneal to the top strand.
- Primers are in large excess over the template DNA, they are never used up and they are incorporated into the final PCR product.

Know these rules!





Amplified region

5′--AAC**CTGAC**TAGTCGATGCGAATGTGCGGTGC—3′
3′--TTCGACTGATCAGCTACGCTT**ACACG**CCACG—5′

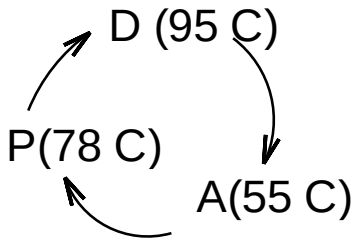
5′CTGAC

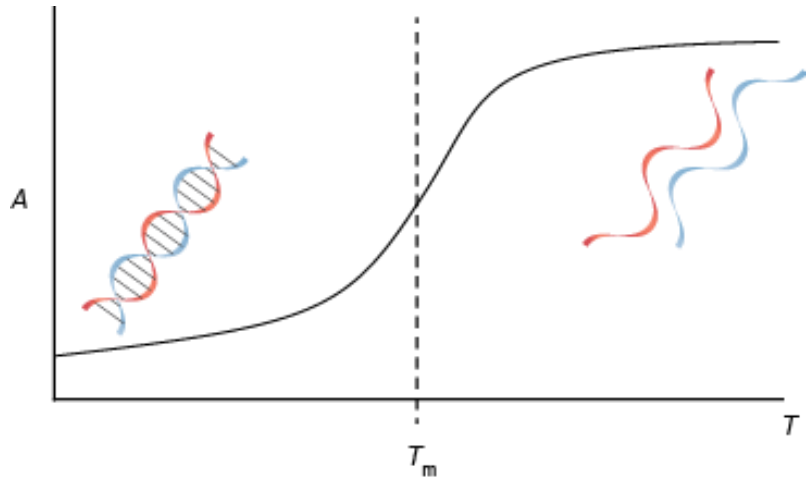
ACACG5′ = 5′GCACA

CTGACTAGTCGATGCGAATGTGC
GACTGATCAGCTACGCTT**ACACG**

Note: Actual primer lengths are 20-30 bases, in the illustrations here and on problem sets, much shorter primers are used.

PCR Step 1 - Thermal Stability of Double Stranded DNA (dsDNA)





5′ A-T-C-A
3′A-T-G-C-C-G-T-A-G-T-C-G-T-A-C-A-G-T-A-C-G-T-G-C-A

5′ A-T-C-A
3′A-T-G-C-C-G-T-A-G-T-C-G-T-A-C-A-G-T-A-C-G-T-G-C-A

1. *How does the DNA sequence affect T_m? Explain the difference in Tm for the following two DNAs:*

Duplex A: Tm=16 C

AAAATTTT
TTTTAAAA

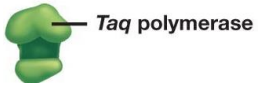
Duplex B: Tm= 32 C

GCGCGCGC
CGCGCGCG

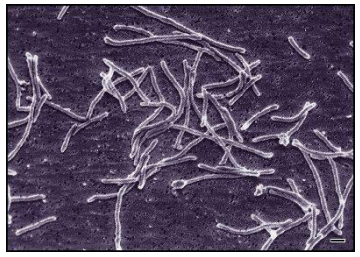
PCR Step 1 - Thermostable Polymerases

- If we heat up the DNA to temperatures high enough that it denatures into single stranded form, (temperatures of between 60°C and 94°C) what will happen to our DNA polymerases?
- Most DNA polymerases are destroyed at this high temperature.
- How can we synthesize DNA is all of our DNA polymerases are destroyed?

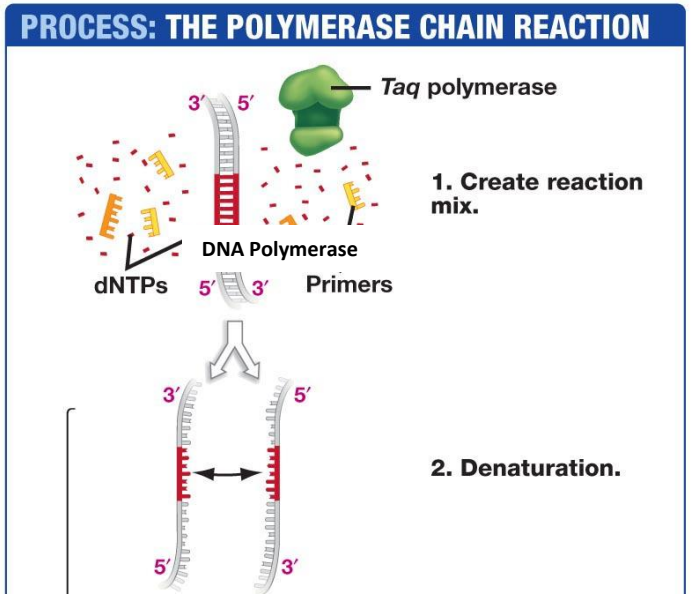
- Utilize a thermostable polymerase



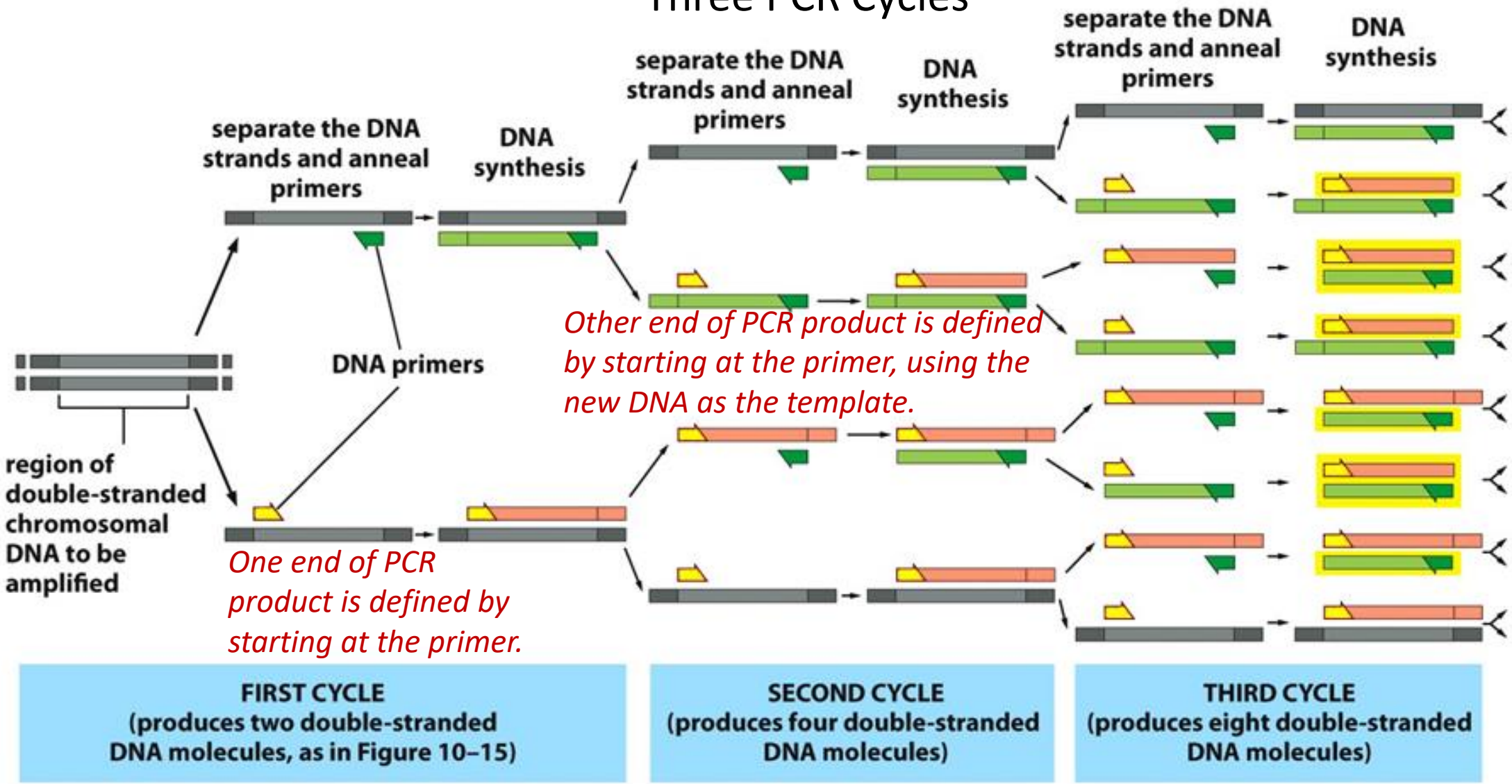
Thermus Aquaticus



http://www.mun.ca/biology/scarr/Thermus_aquaticus.html



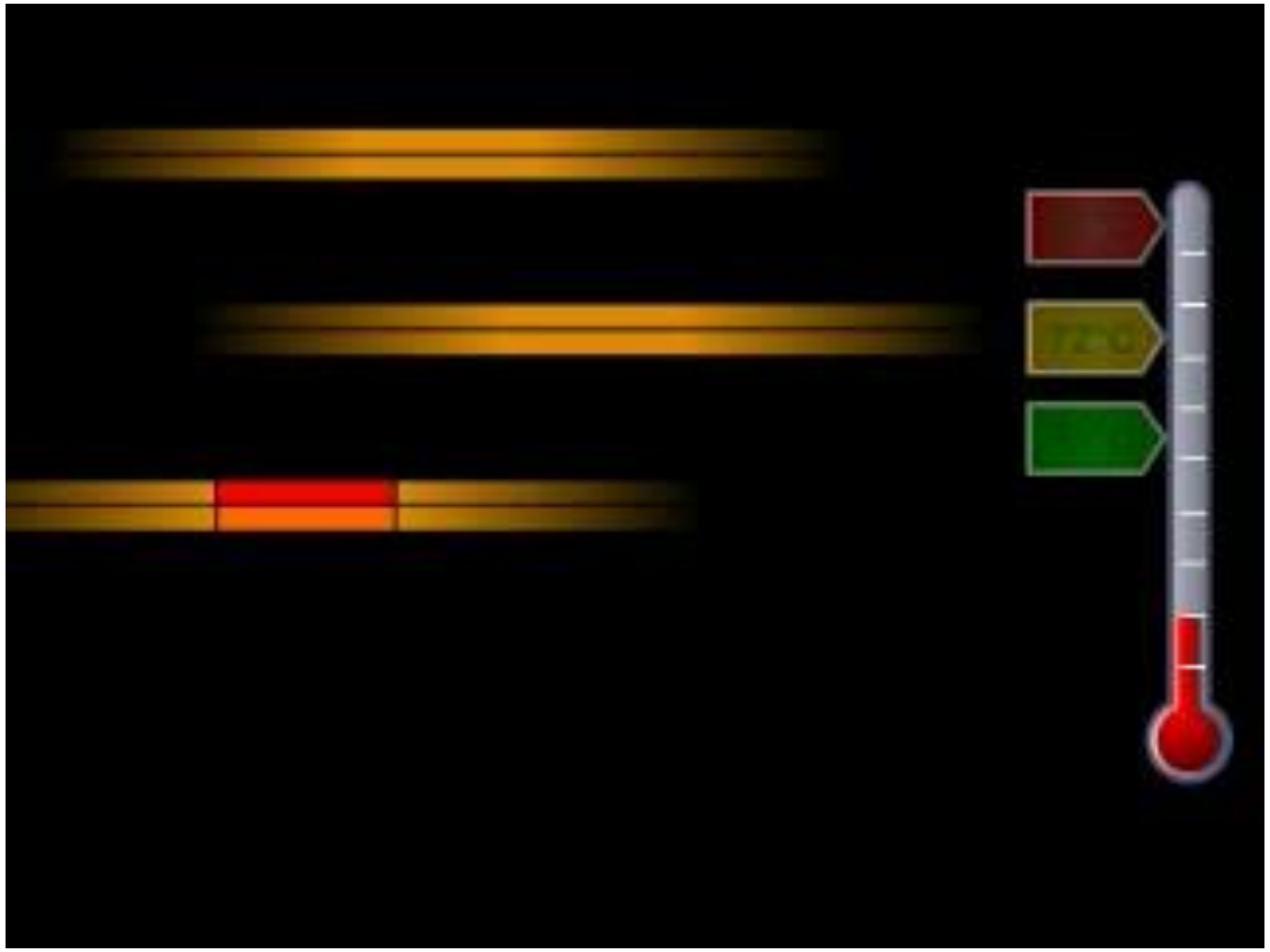
Three PCR Cycles



After 30 cycles there will be 2^{30}, or over 1 billion times more copies than at the beginning!!!

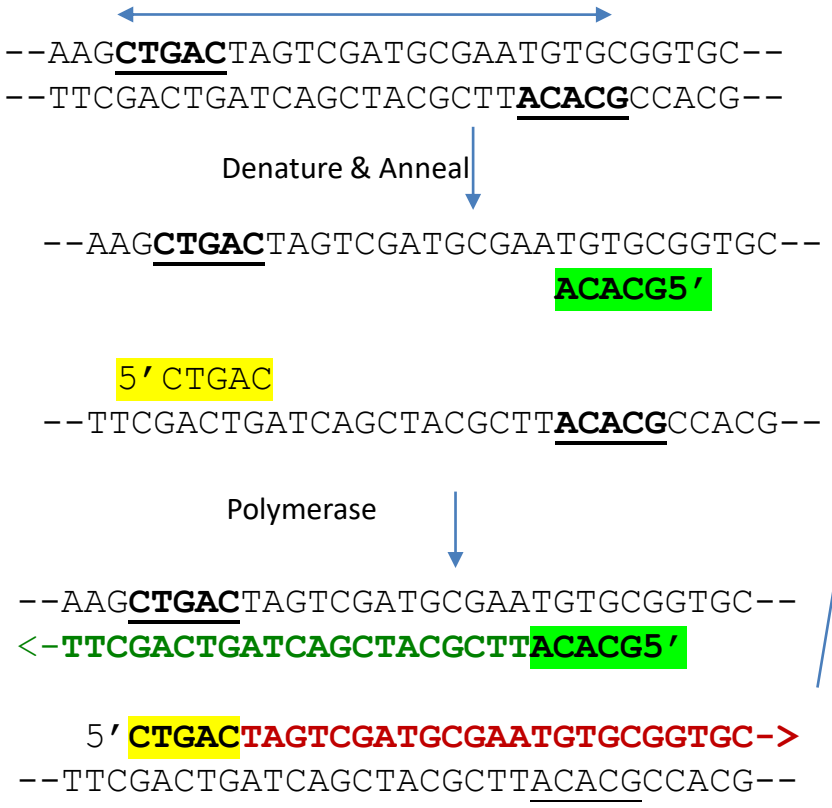
PCR Animation

Watch Me!



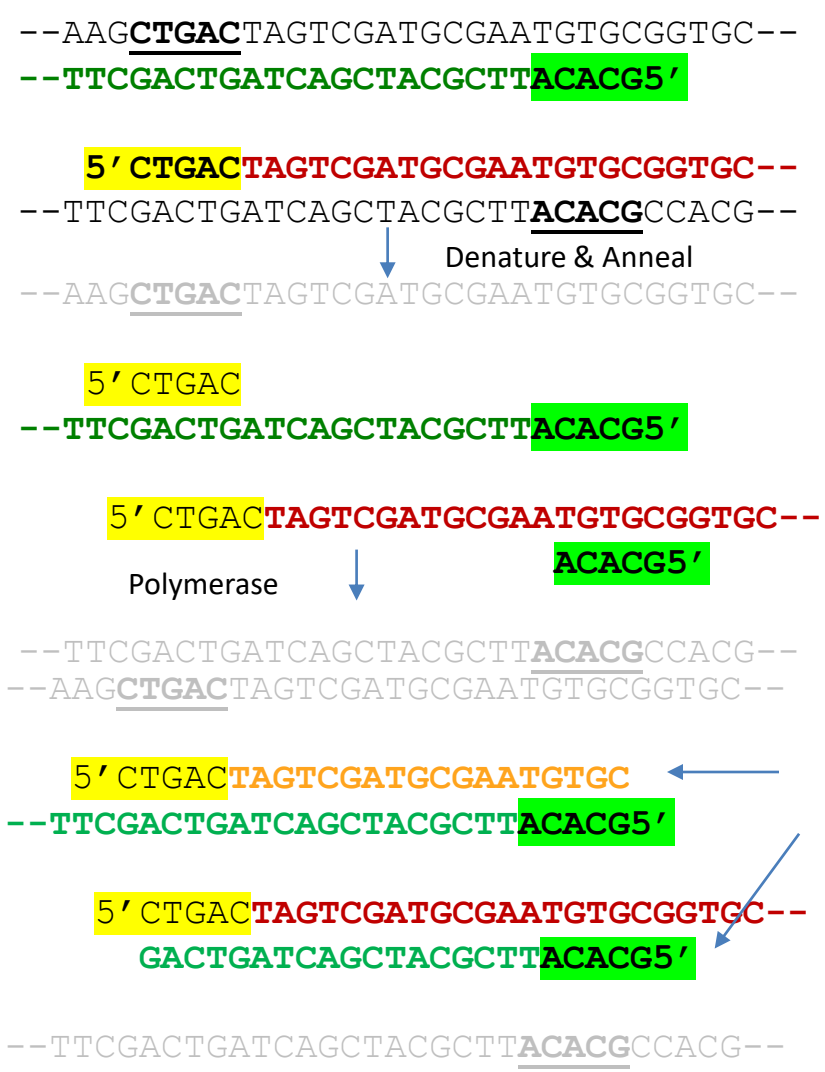
Detailed Events during first Three PCR Cycles

Cycle I



So far - defined one end of the product

Cycle II



Now have one strand of the product

Final Product

CTGACTAGTCGATGCGAATGTGC
GACTGATCAGCTACGCTTACACG

Detailed Events during first Three PCR Cycles

Example – follow the PCR cycles for the following template with primers 5’ AATT (left) and 5’ GGCC (right)

```
----AATT------------------------GGCC----
----TTAA------------------------CCGG----
```

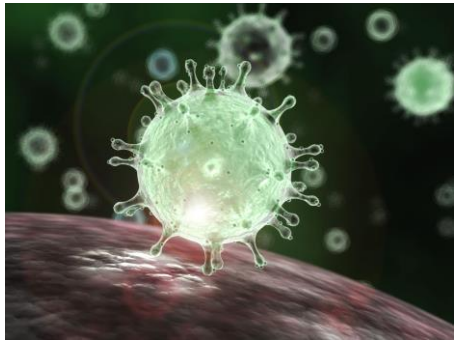
Cycle 3

```
    5’CTGACTAGTCGATGCGAATGTGC
--TTCGACTGATCAGCTACGCTTACACG5’

    5’CTGACAGTCGATGCGAATGTGCGGTGC--
      GACTGTCAGCTACGCTTACACG5’
Denature & Anneal
    5’CTGACTAGTCGATGCGAATGTGC
                    ACACG5’

--TTCGACTGATCAGCTACGCTTACACG5’

    5’CTGACAGTCGATGCGAATGTGCGGTGC--

    5’CTGAC
      GACTGTCAGCTACGCTTACACG5’
Polymerase
  5’CTGACTAGTCGATGCGAATGTGC
    GACTGATCAGCTACGCTTACACG5’

--TTCGACTGATCAGCTACGCTTACACG5’

    5’CTGACAGTCGATGCGAATGTGCGGTGC--

    5’CTGACAGTCGATGCGAATGTGC
      GACTGTCAGCTACGCTTACACG5’
```

Now have complete PCR product. This is doubled in each of the following cycles.
Note that the primers are the first bases at the beginning of each strand.

PCR & Detection of Viruses

Coronavirus



Sequence of Covid-19 (top strand only)

```
1     attaaaggtt tataccttcc caggtaacaa accaaccaac tttcgatctc ttgtagatct
61    gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcatgc ttagtgcact
121   cacgcagtat aattaataac taattactgt cgttgacagg acacgagtaa ctcgtctatc
181   ttctgcaggc tgcttacggt ttcgtccgtg ttgcagccga tcatcagcac atctaggttt

28261 cgaacaaact aaaatgtctg ataatggacc ccaaaatcag cgaaatgcac cccgcattac
28321 gtttggtgga ccctcagatt caactggcag taaccagaat ggagaacgca gtggggcgcg
28381 atcaaaacaa cgtcggcccc aaggtttacc caataatact gcgtcttggt tcaccgctct
28441 cactcaacat ggcaaggaag accttaaatt ccctcgagga caaggcgttc caattaacac

29701 gggaggactt gaaagagcca ccacattttc accgaggcca cgcggagtac gatcgagtgt
29761 acagtgaaca atgctaggga gagctgccta tatggaagag ccctaatgtg taaaattaat
29821 tttagtagtg ctatccccat gtgattttaa tagcttctta ggagaatgac aaaaaaaaaa
29881 aaaaaaaaaa aaaaaaaaaa aaa
```

CDC Recommended PCR Primers

2019-Novel Coronavirus (2019-nCoV) Real-time rRT-PCR Panel Primers and Probes

Name	Description	Oligonucleotide Sequence (5'>3')	Label[1]	Working Conc.
2019-nCoV_N1-F	2019-nCoV_N1 Forward Primer	5'-GAC CCC AAA ATC AGC GAA AT-3'	None	20 µM
2019-nCoV_N1-R	2019-nCoV_N1 Reverse Primer	5'-TCT GGT TAC TGC CAG TTG AAT CTG-3'	None	20 µM

dsSeq of above bold & circled region

```
28271 aaaatgtctgataatgGACCCCAAAATCAGCGAAATgcaccccgcattacgtttggtggaccctcagattcaactggcagtaaccagaatggagaacgca
      ttttacagactattacctggggttttagtcgctttacgtggggcgtaatgcaaaccacctgggaGTCTAAGTTGACCGTCATTGGTCTtacctcttgcgt
```

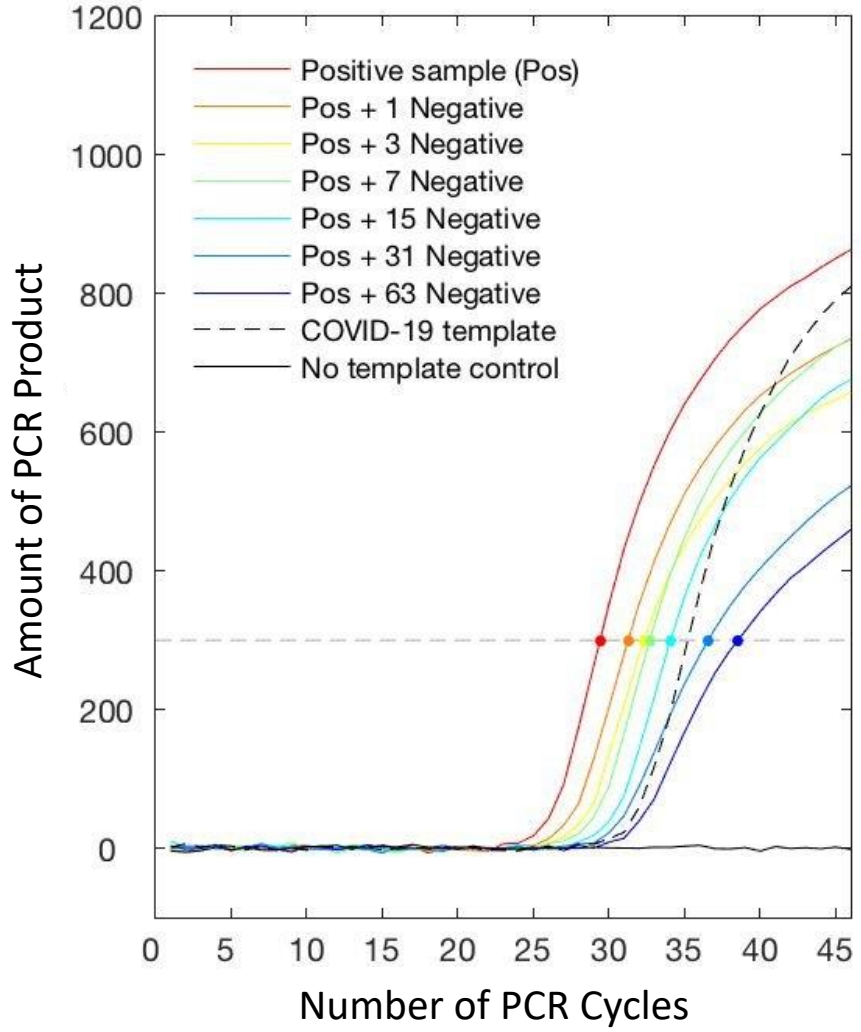
PCR Product

```
GACCCCAAAATCAGCGAAATGCACCCCGCATTACGTTTGGTGGACCCTCAGATTCAACTGGCAGTAACCAGA
CTGGGGTTTTAGTCGCTTTACGTGGGGCGTAATGCAAACCACCTGGGAGTCTAAGTTGACCGTCATTGGTCT
```

Will PCR generate products if the viral DNA is not present?

Covid 19 PCR Test: Detection of the PCR Product.

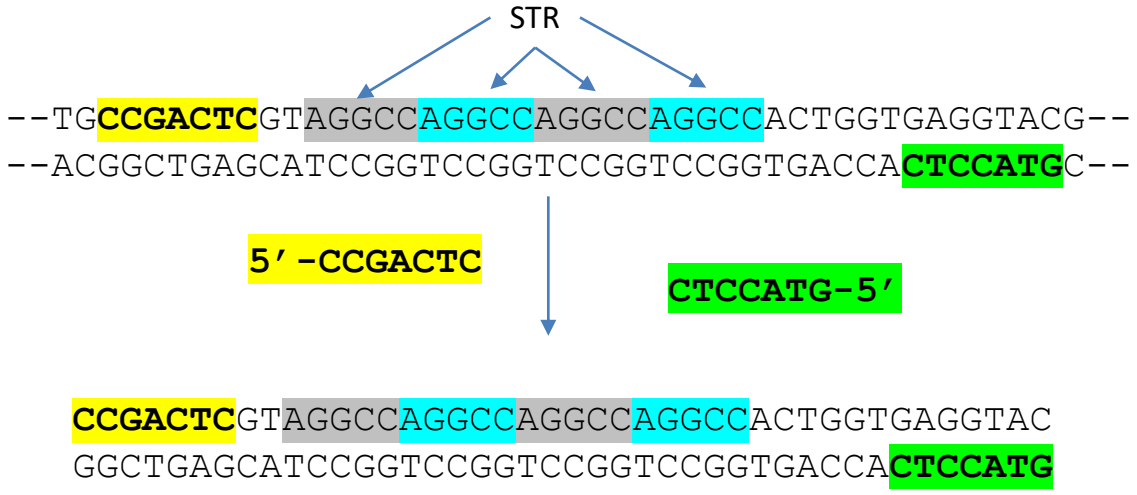
https://www.medrxiv.org/content/10.1101/2020.03.26.20039438v1



- Production of PCR products (double stranded DNA) causes an increase in signal (fluorescence)
- Dots represent the cross point of the fluorescence threshold (threshold = 300, gray dashed line).
- Red curve (Positive sample) shows a threshold level of PCR product after 27 cycles.
- Next 6 samples are the positive sample mixed with up to 63 negative samples, showing that it is possible to test pooled samples.
- - - - is a ***positive control*** amount of Covid template. It shows that you can detect a PCR product if the covid genome is present.
- Solid black line is a ***negative control***, no Covid DNA. It shows that addition of covid template will lead to a signal.

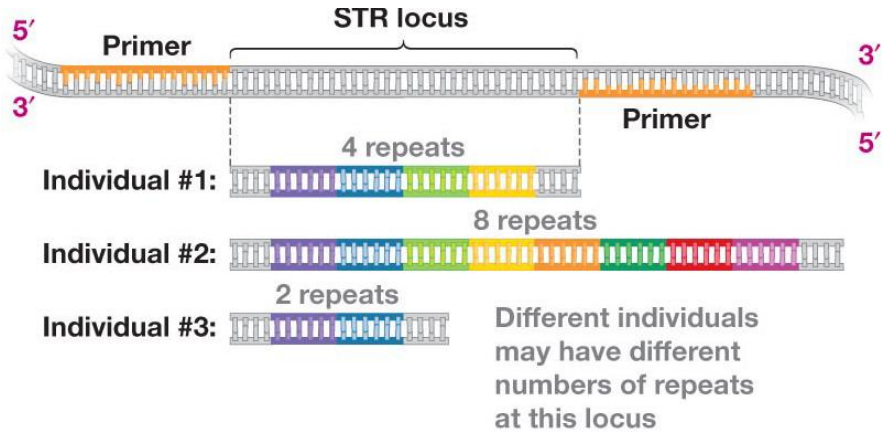
Application of PCR – Identification of Individuals

- Regions of DNA have variable numbers of repeated DNA sequences (Short tandem repeats, STR) that differ from one person to the next.
  - Individuals will inherit one copy of the repeat from each parent. The length of the inherited DNA can be the same or different.
  - PCR Primers are designed to be outside the repeated region, so that they will anneal to a single location on the chromosome and then amplify the region containing the STR
  - PCR Product length = primer lengths + number of tandem repeats (+ any DNA between the primers and the repeats). *Individuals can be differentiated by the length of the PCR product if they have different numbers of STR*



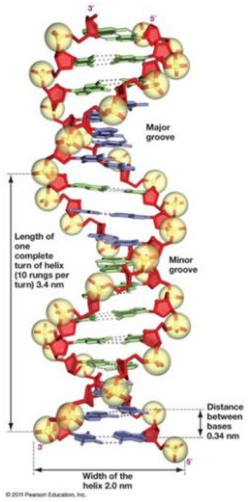
Which individual has the shortest PCR product?

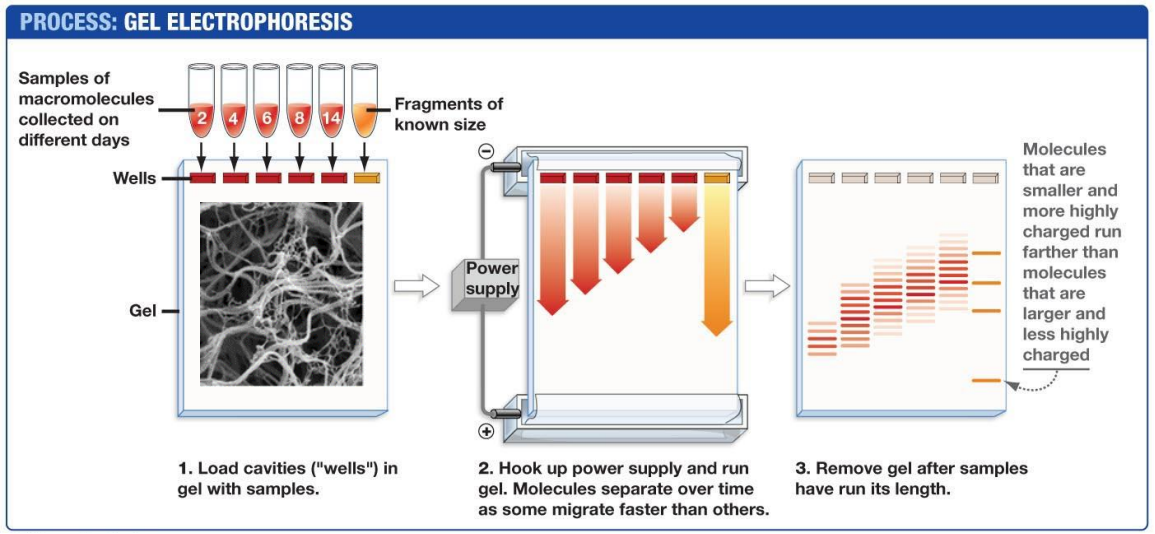
Which has the longest?

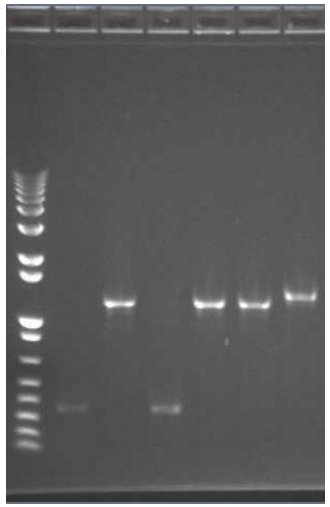


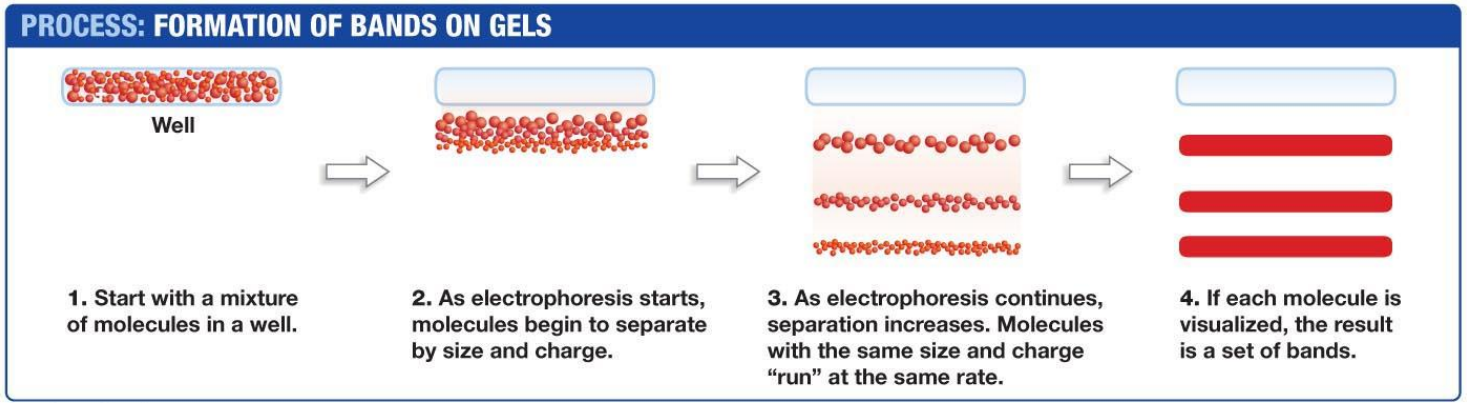
Size Determination of PCR products - Agarose Gel Electrophoresis.

https://dnalc.cshl.edu/resources/animations/gelelectrophoresis.html



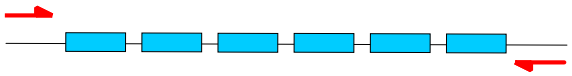






Which are the smallest PCR fragments?

Short Tandem Repeats to Test Paternity

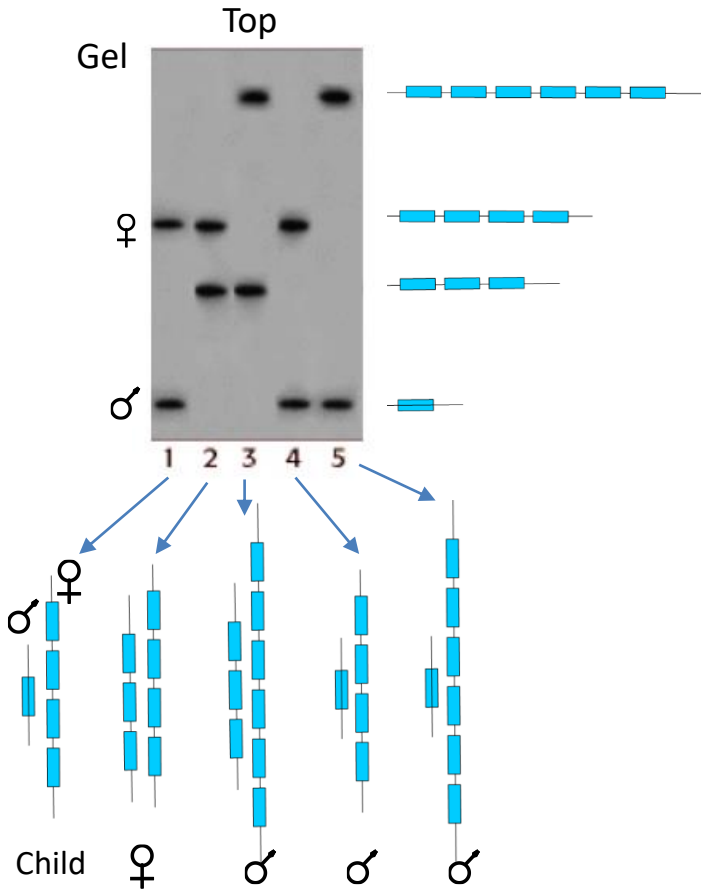


PCR primers:

Repeat:

Lane 1: Child
Lane 2: Mother

Lanes 3-5: Possible Fathers



1. Why are there two band in each lane?

*2. Who is **not** the father?*

*3. Who **may** be the father?*

Repeat Expansions Related to Diseases

Handbook of Clinical Neurology
Volume 147, 2018, Pages 105-123

Chapter 9 - Repeat expansion diseases

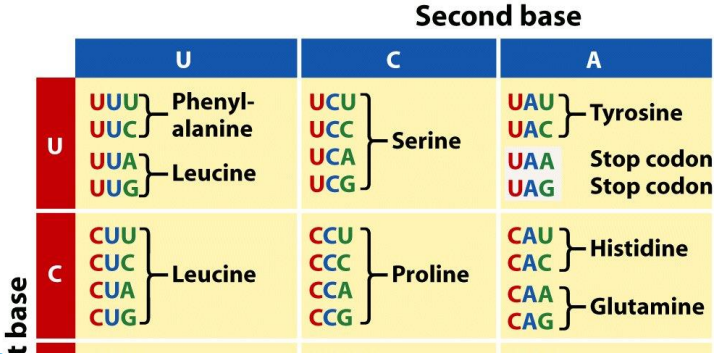
Henry Paulson

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/B978-0-444-63233-3.00009-9

Get rights and conten



- CAG – at least 10 diseases (Huntington disease, spinal and bulbar muscular atrophy, dentatorubral-pallidoluysian atrophy and seven SCAs)
- CGG – fragile X, fragile X tremor ataxia syndrome, other fragile sites (GCC, CCG)
- CTG – myotonic dystrophy type 1, Huntington disease-like 2, spinocerebellar ataxia type 8, Fuchs corneal dystrophy
- GAA – Friedreich ataxia
- GCC – FRAXE mental retardation
- GCG – oculopharyngeal muscular dystrophy
- CCTG – myotonic dystrophy type 1
- ATTCT – spinocerebellar ataxia type 10
- TGGAA – spinocerebellar ataxia type 31
- GGCCTG – spinocerebellar ataxia type 36
- GGGGCC – C9ORF72 frontotemporal dementia/amyotrophic lateral sclerosis
- CCCCGCCCCGCG – EPM1 (myoclonic epilepsy)

- These repeats can grow during replication in the cell.
- Longer repeats can be inherited from the parent
- Longer repeats can also occur with cells of the individual
- The number of repeats can be detected by:
  - DNA sequencing
  - PCR