Biology Crash Course

7.546/15.480/20.586 The Science and Business of Biotechnology

Catherine (Kate) Koch

Biology Crash Course Outline

Part 1: Biology Overview

- Three Important Biological Entities
  - Eukaryotes, Prokaryotes, Viruses
- Central dogma of biology
  - DNA → RNA → Protein
- “Building Blocks” of life
  - Nucleic acids + nucleotides
  - Proteins + amino acids
- Making proteins
  - Overview: DNA → RNA → Protein
  - What is a gene?
  - The genetic code
  - Transcription (DNA→RNA)
    - RNA Polymerase + transcription factors
    - Splicing + Alternative splicing
  - Translation (RNA → Protein)
    - Ribosomes + Transfer RNA (tRNA)
  - Post-translational modifications
- Gene expression

Part 2: Useful topics

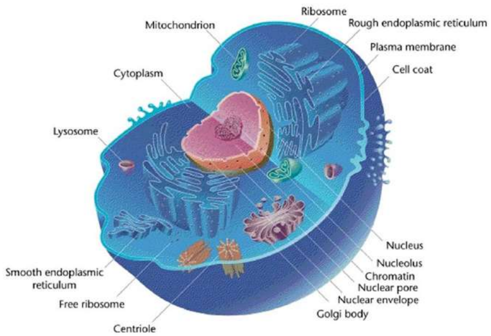
- Inheritance: dominant, recessive, X-linked
- What is disease?
- Questions and future topic requests
- Lecture 1/Recitation 1 important topics:
  - Recombinant proteins
  - Biosimilars vs. generics

Appendix

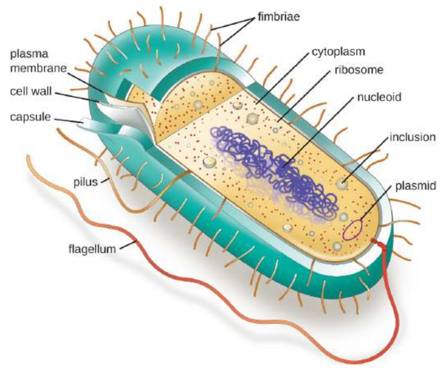
- Questions asked during/after the session

Part 1: Biology Overview

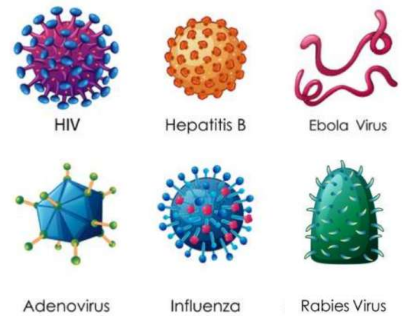
Three important Biological entities



Eukaryotic cells
(Ex: human cells)



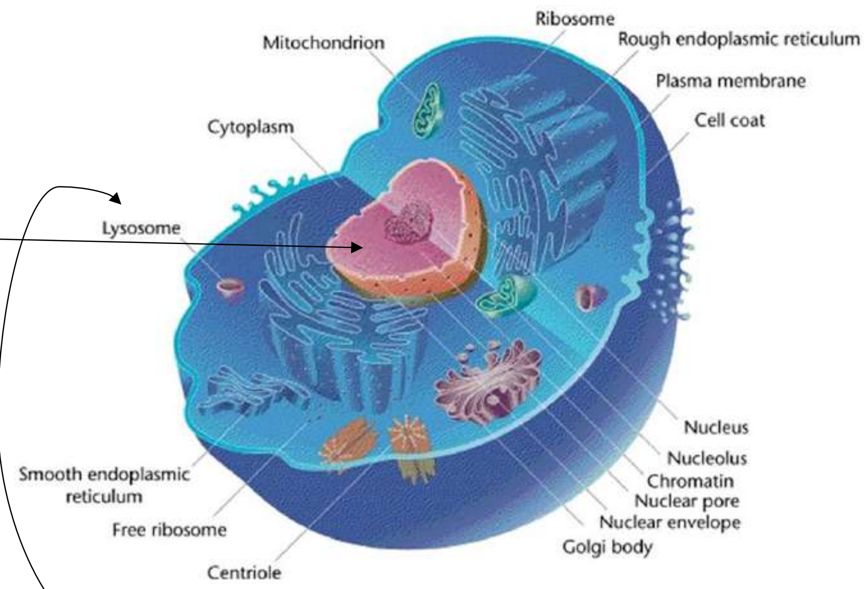
Prokaryotic cells
(Ex: bacteria)



Viruses
(Ex: Herpes virus)

Eukaryotic cells

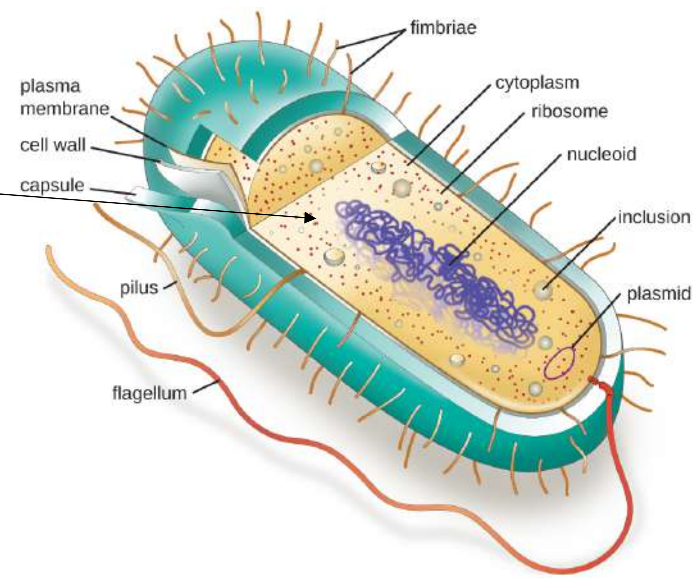
- Tend to be more "complex" and larger than prokaryotic cells (like bacteria)
- **DNA:**
  - Linear and organized into chromosomes
  - Located in nucleus
- **Internal membrane-bound compartments**
  - Membrane bound compartments exist inside cell (called "organelles" ex: lysosome)
- **Contain mitochondria:**
  - Mitochondria generate energy
- Can exist as **single celled organisms** (like yeast) **or** join together to form **multi-cellular organisms** (like humans)
- **Examples:**
  - Mammalian cells: human cell, mouse cell, Chinese ovarian hamster (CHO) cell
  - Yeast (ex. Candida albicans)
  - Parasites: plasmodium falciparum (causes malaria)



Organelle: compartment inside cell surrounded by its own membrane that generally serves some specific function (ex: lysosomes perform digestion of cellular waste material)

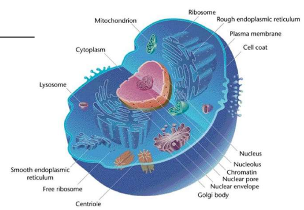
Prokaryotic cells

- Tend to be less "complex" and smaller than eukaryotic cells (like human cells)
- **DNA:**
  - Circular (have one or more massive loops of DNA)
  - DNA floats around cytoplasm – no nucleus
- **Membrane bound compartments**
  - NO membrane bound compartments inside cell (no organelles)
- **Do not contain mitochondria:**
  - Mitochondria do not serve as energy source
- Exist as **single celled organisms** like bacteria
- **Examples:**
  - E. coli (bacteria that can cause food poisoning)
  - Streptococcus pyogenes (bacteria that causes Strep throat)
  - Bacillus anthracis (bacteria that causes Anthrax)
  - Salmonella Typhi (bacteria that causes Typhoid Fever)

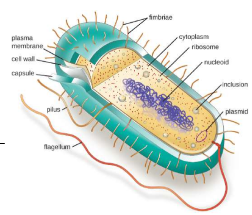


If you take anything away from this…

Mammalian cell (Eukaryotic)



Both are **cells,** both have **DNA,** and both can **make proteins**



Bacterial cell (Prokaryote)

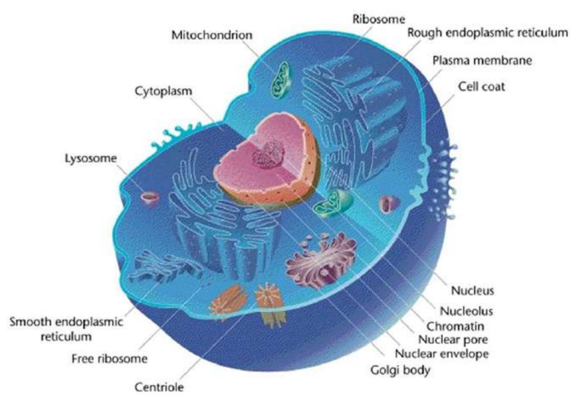
In the same way that these are not exactly the same, mammalian and bacterial cells are not exactly the same.

Van Gogh

Both are **artists**, both have **paint**, and both can **make art**

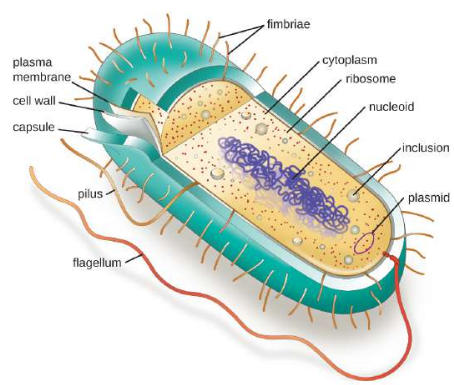
Child

If you take away anything from this…



Mammalian cell
(Eukaryotic)

≠



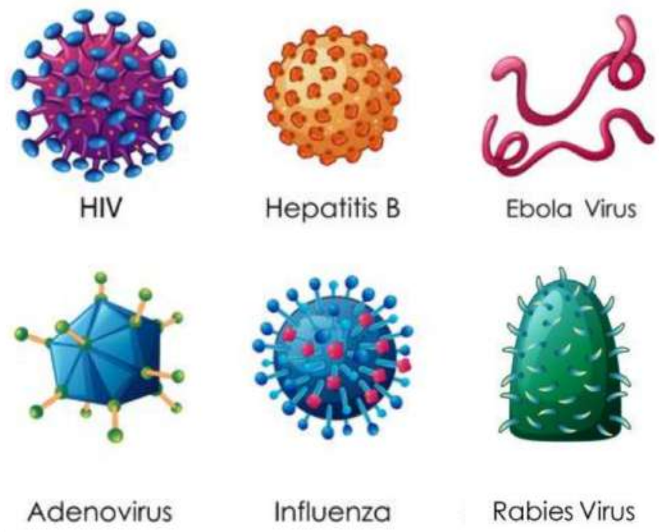
Bacterial cell
(Prokaryote)

Summary

Eukaryotic cells	Prokaryotic cells
• More complex and larger • **DNA:** linear (chromosomes) • **DNA location:** nucleus • **Organelles:** has defined membrane bound compartments • **Mitochondria:** has mitochondria • Exist as single cell organisms **or** join together to from multi-cellular organisms • **Examples:** mammalian cells (human cell, mouse cell), yeast cell, etc.	• Less complex and smaller • **DNA:** circular (large loop or loops) • **DNA location:** cytoplasm (no nucleus) • **Organelles:** does NOT have defined membrane bound compartments • **Mitochondria:** has no mitochondria • Exist as single celled organisms • **Examples:** bacterial cell like E. coli, Bacillus anthracis, Salmonella typhi, etc.

**Not an exhaustive list but sufficient for the moment

Viruses


HIV
Hepatitis B
Ebola Virus
Adenovirus
Influenza
Rabies Virus

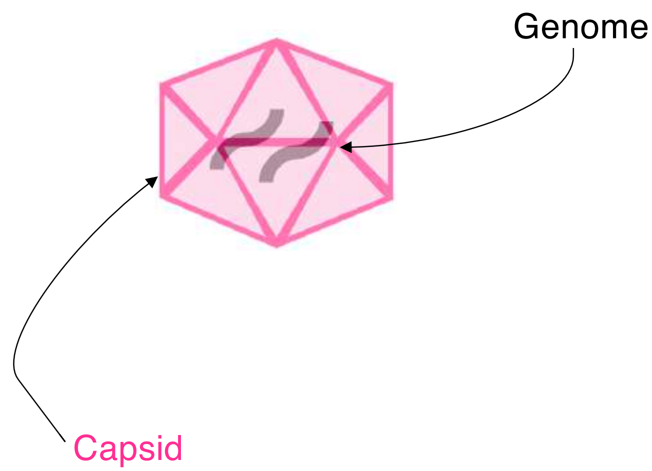
Viruses: Structure

- **Virus:** a small, obligate intracellular parasite
  - Viral particle (virion) = nucleic acid (RNA *or* DNA) enclosed in a protein shell (capsid)

- **Genome:** can be made of single stranded or double stranded DNA or RNA
  - Simple viruses: can contain only enough RNA or DNA to encode 4 proteins
  - Complex viruses: can contain enough RNA or DNA to encode ~200 proteins

- **Capsid:** surrounds viral genome and is composed of multiple copies of virus-encoded protein(s)

Virus particle (virion)



Viruses: Structure

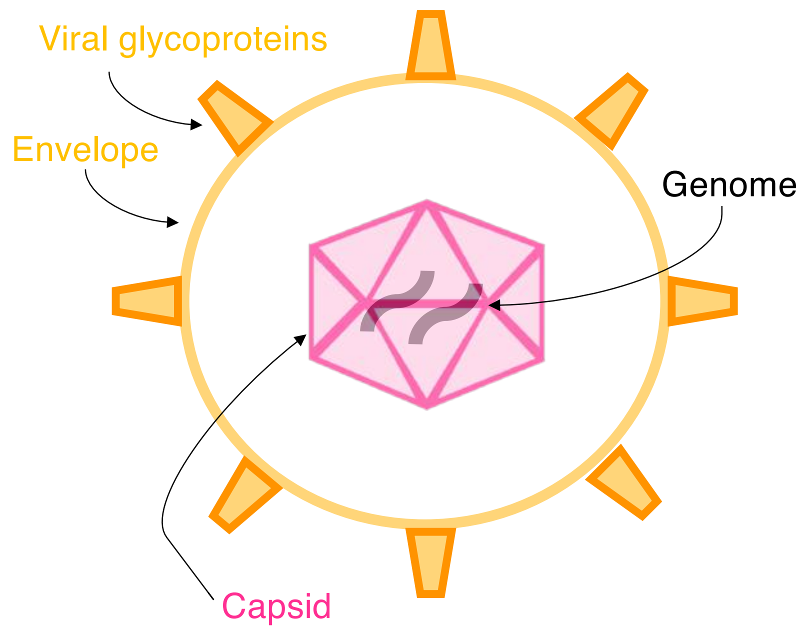
- **Virus:** a small, obligate intracellular parasite
  - Viral particle (virion) = nucleic acid (RNA *or* DNA) enclosed in a protein shell (capsid)

- **Genome:** can be made of single stranded or double stranded DNA or RNA
  - Simple viruses: can contain only enough RNA or DNA to encode 4 proteins
  - Complex viruses: can contain enough RNA or DNA to encode ~200 proteins

- **Capsid:** surrounds viral genome and is composed of multiple copies of virus-encoded protein(s)

- **Envelope:** outer envelope that surrounds the capsid containing the viral genome
  - Viral genome does not provide instructions for envelop – it's "borrowed" from host cell membrane
  - Envelopes contain viral glycoproteins (virus provides instructions for synthesis of these)

Virus particle (virion)



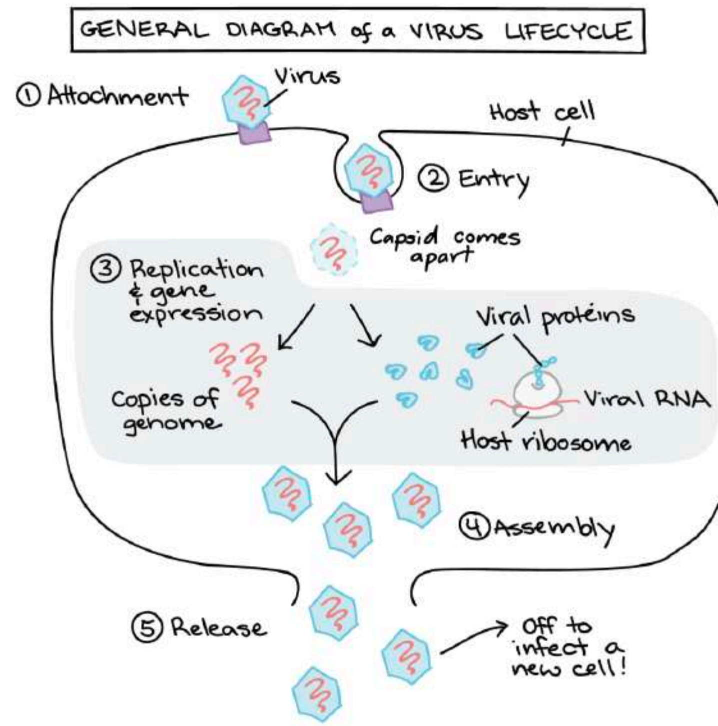
**Not all viruses have envelope

Viruses: general outline of replication

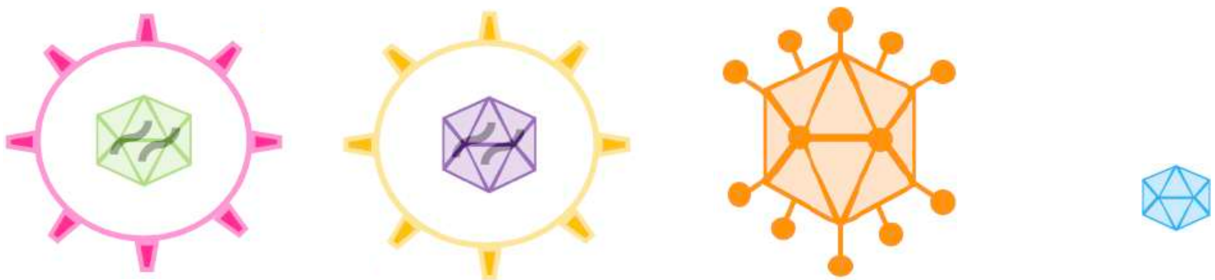
Viruses can't reproduce by themselves:
- Need to take over a host cell's machinery to make viral proteins and replicate their genome
- Viruses are not actually considered to be "living"

General steps:
1. Attachment to cell surface
2. Entry into cell
3. Replication: genome replication and viral protein production
4. Assembly of viral particles
5. Release of more viral particles

**This is a generalized outline of viral replication – specific viruses deviate from this process in their own special ways



Viruses: four groups important for gene therapy



Retrovirus | **Lentiviruses** | **Adenovirus** | **Adeno-associated viruses (AAV)**

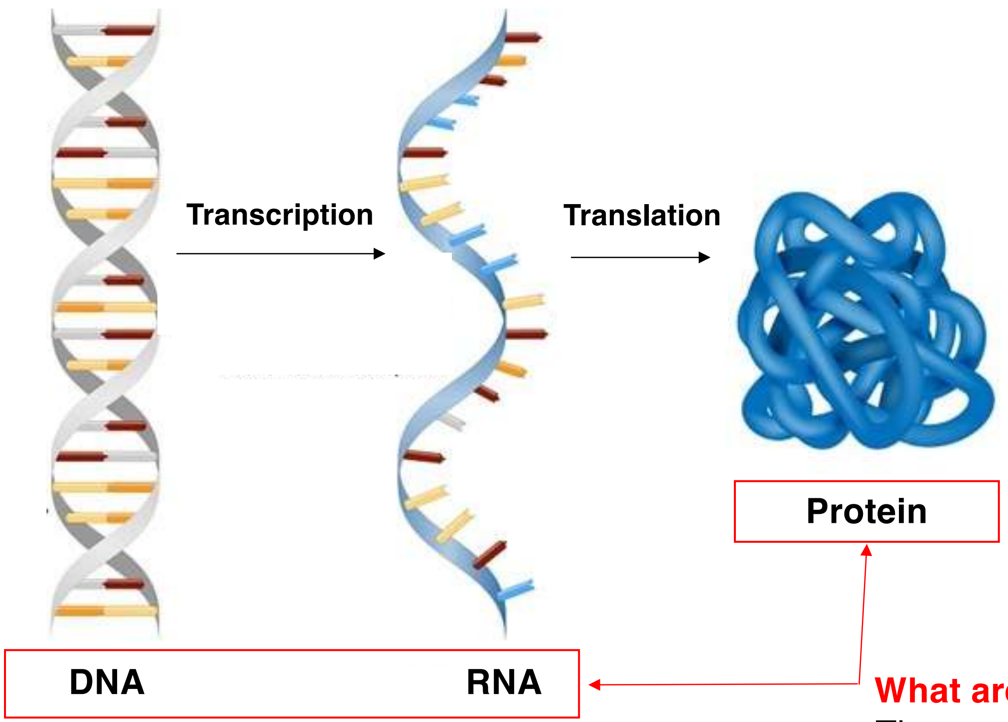
**Next 2 recitations will cover in detail

Summary

Retroviruses	Lentiviruses	Adenoviruses	Adeno-associated viruses (AAV)
• Can only infect dividing cells • Viral DNA integrates into host genome • Can package a gene that is up to: 8kb DNA	• Can infect dividing and non-dividing cells • Viral DNA integrates into host genome • Can package a gene that is up to: 8kb DNA	• Can infect dividing and non dividing cells • Generally does not integrate into host genome • Can package a gene that is up to: 37kb DNA	• Can infect dividing and non dividing cells • Generally does not integrate into host genome • Can package a gene that is up to: 4.5kb DNA

**Next 2 recitations will cover in detail

How do we go from DNA to protein?

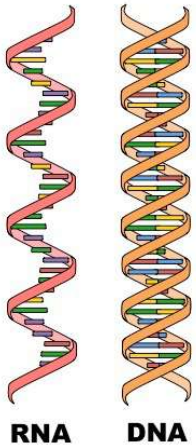


Central Dogma:

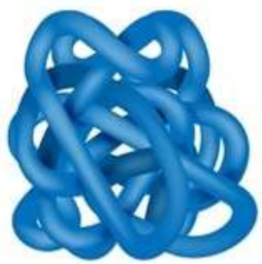
- Framework for how genetic information flows from DNA to a protein
- **Transcription:** DNA is transcribed into RNA
- **Translation:** RNA is translated into protein

What are these and what are they made of?
These entities are the basis of life and biology.

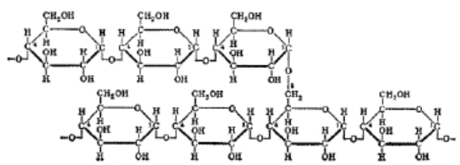
Macromolecules important for life



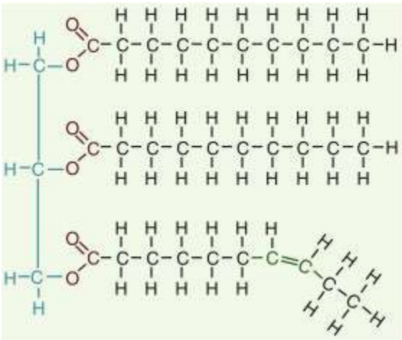
Nucleic acids



Proteins

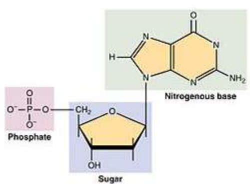


Carbohydrates

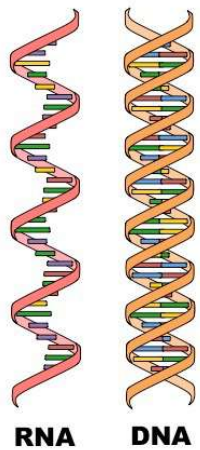


Lipids

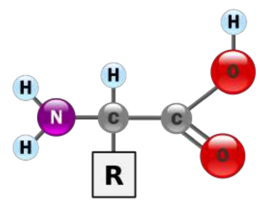
“Building Blocks”



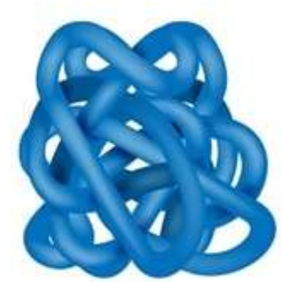
Nucleotides



Nucleic acids



Amino Acid

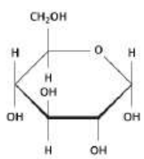


Protein

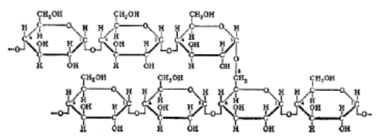
Proteins are “built” of amino acids

Nucleic acids are “built” of nucleotides

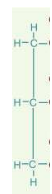
“Building Blocks”



Monosaccharide

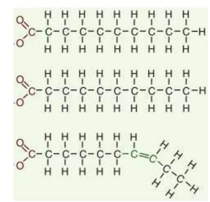


Carbohydrate
(Polysaccharide)

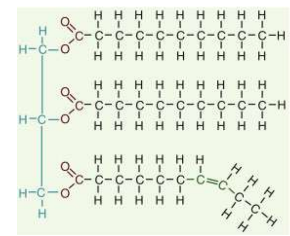


Glycerol

+



Fatty acid
chains

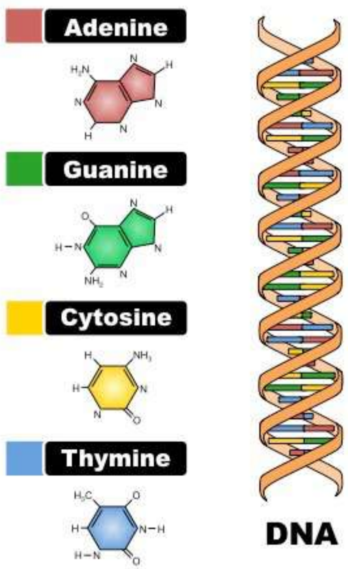


Triglyceride

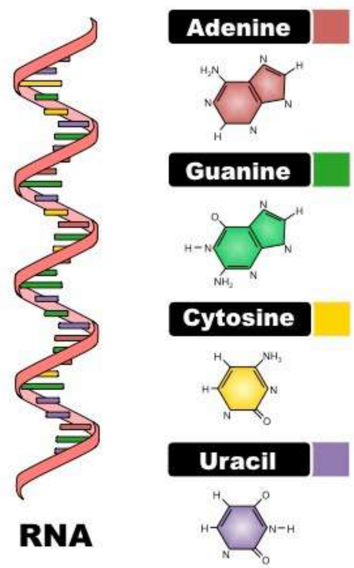
Carbohydrate are “built” of monosaccharides

Lipids are “built” of fatty acids + other components (eg glycerol)

Nucleic Acids



DNA
(Deoxyribonucleic Acid)



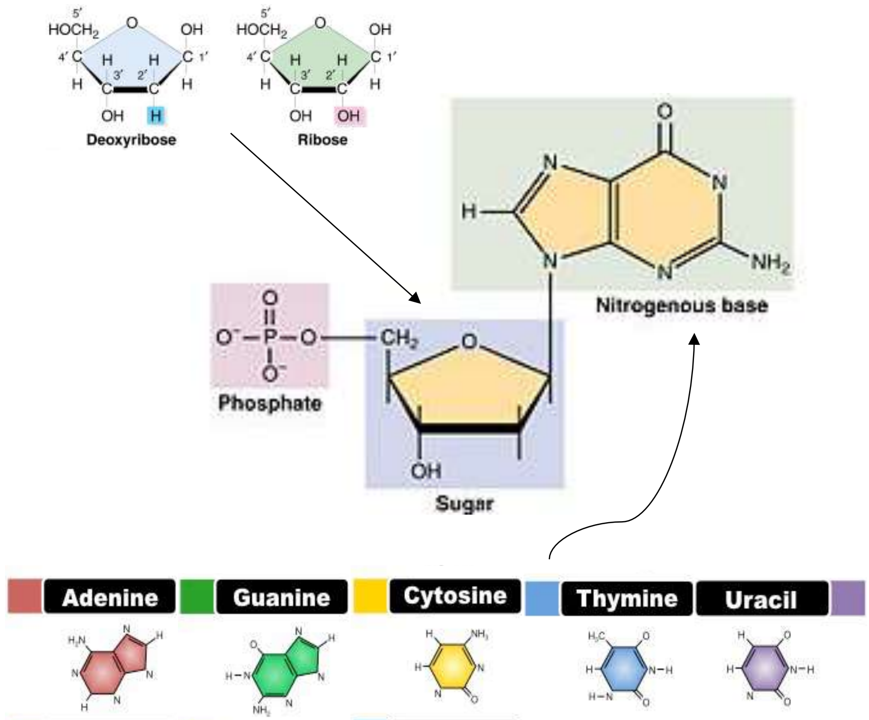
RNA
(Ribonucleic Acid)

Nucleic Acids

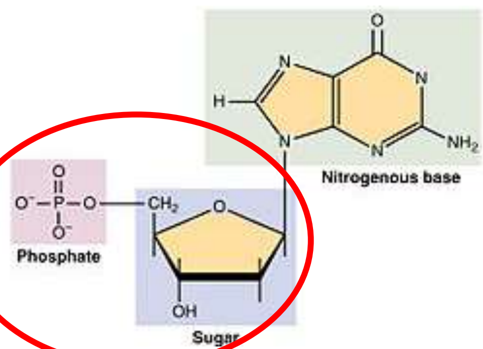
- **Nucleic acids (DNA and RNA) are composed of "units" called nucleotides**
  - Nucleotides are the "building blocks" of nucleic acids (DNA and RNA)
  - 5 different nucleotides (differentiate them based on their **nitrogenous base**: A, G, C, T, U)

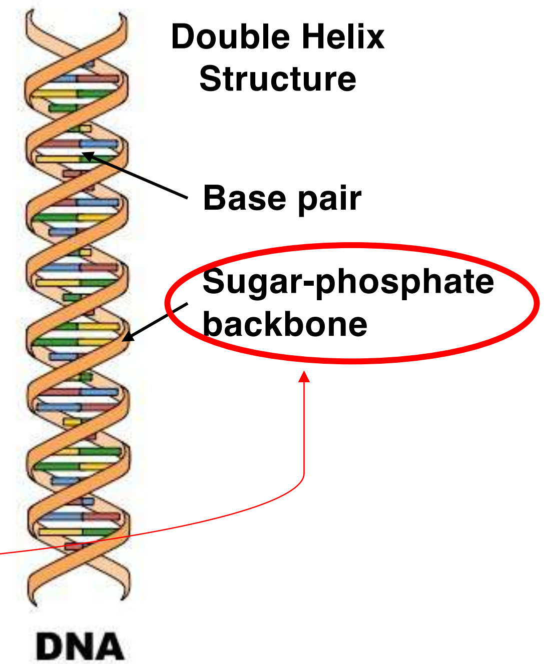
- **Nucleotide "building blocks" have 3 parts:**
  - **Phosphate**
  - **Sugar:** ribose in RNA, deoxyribose in DNA
  - **Base:** Adenine (A), Guanine (G), Cytosine (C), and Thymidine (T) (DNA) or Uracil (U) (RNA)
  - Different nucleotides are defined by their base (A, G, C, T, U)

- **Function: Nucleic acids (DNA and RNA) carry information about making proteins**
  - Information is coded by the nucleotide bases A, C, G, and T (DNA) or U (RNA)
  - The order or sequence of the nucleotides provides information



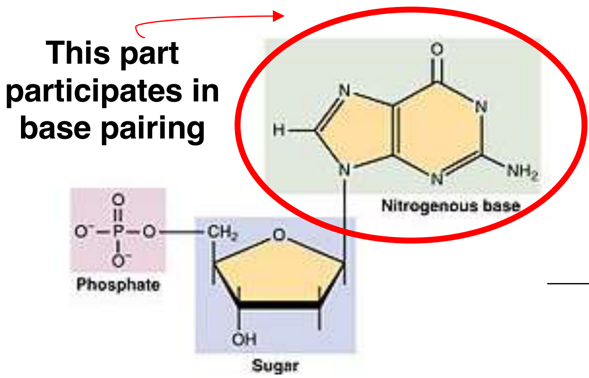
DNA: breaking down the structure

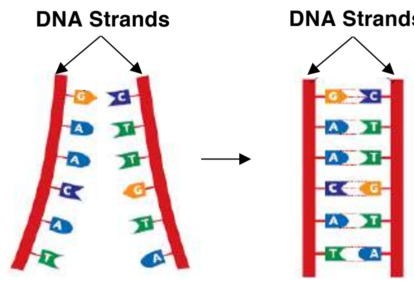




- Nucleotides with different bases (A, G, T, C) link together via sugar/phosphate regions to form a single DNA strand
- **These links form the sugar-phosphate backbone in the DNA molecule**

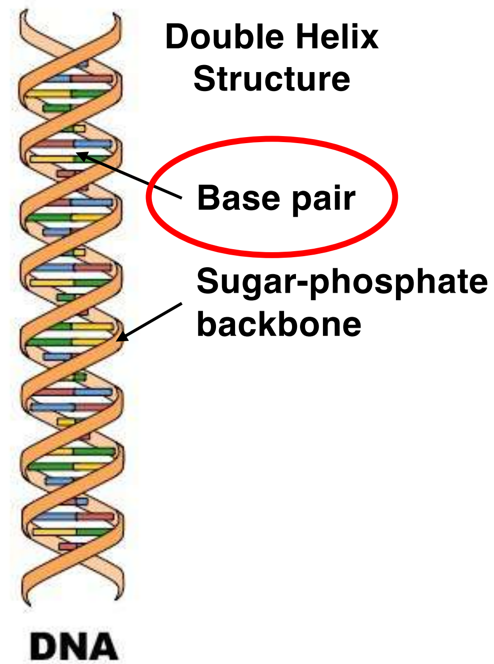
DNA: breaking down the structure





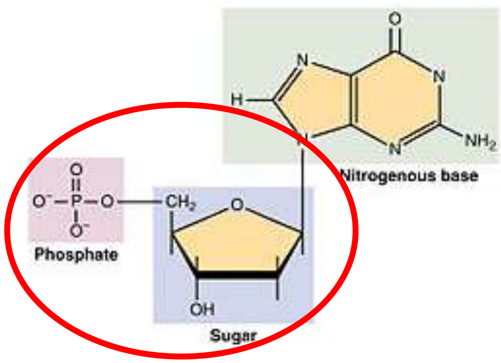
Base pairing

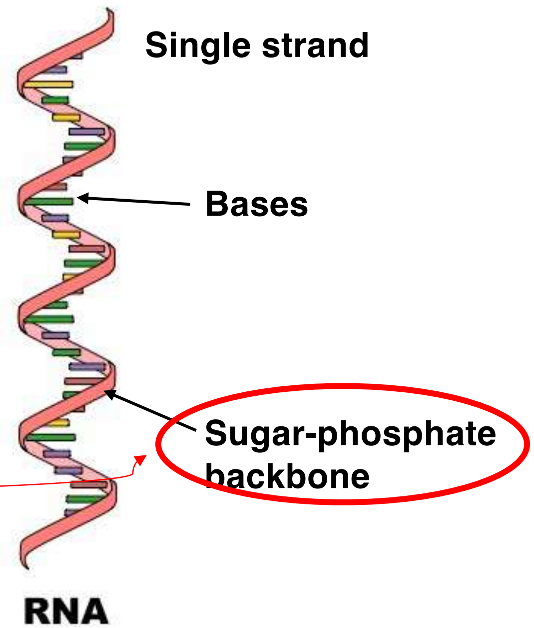
(the interactions that happen between the bases on each strand joins the 2 strands of linked nucleotides together)



- Two DNA strands join together via base pairing (A with T, C with G) and twist to form a double helix
- **Interactions between base parts of nucleotides on each strand allow the 2 strands to join together and form double helix**

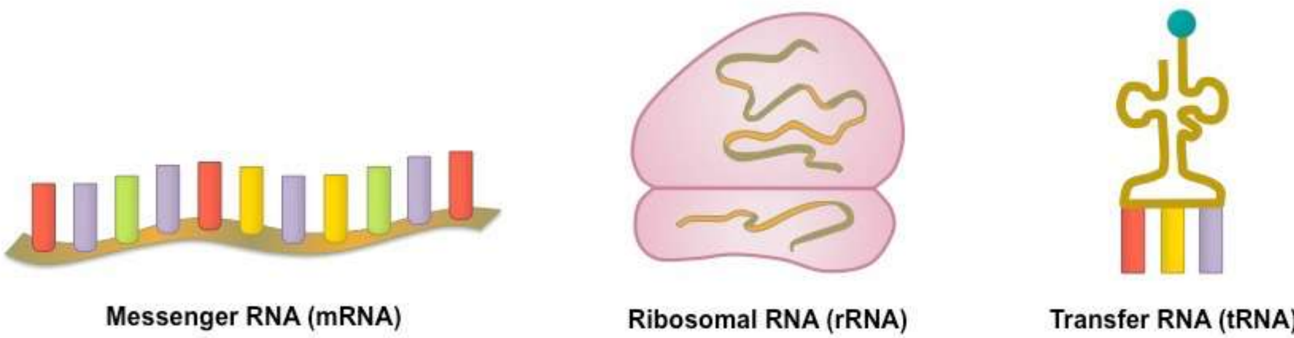
RNA: breaking down the structure





- Nucleotides with different bases (A, G, C, U) link together via sugar/phosphate regions to form a RNA strand

RNA: three main types



mRNA, rRNA, and tRNA all play important roles in the process of protein production

*We're going to loop back to this when these things come up in protein production

Summary

<table>
<tr><th>DNA</th><th>RNA</th></tr>
<tr><td>

Structure: double-stranded, double-helix
Nucleotide sugar: Deoxyribose
Nucleotide Bases:
Adenine (A)
Guanine (G)
Cytosine (C)
Thymine (T)
Location: nucleus (usually)
Function: DNA replicates when cells divide and stores a cell’s genetic information. It acts as a blueprint for all genetic information contained within an organism.

</td><td>

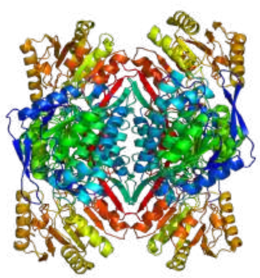
Structure: single stranded
Nucleotide Sugar: Ribose
Nucleotide Bases:
Adenine (A)
Guanine (G)
Cytosine (C)
Uracil (U)
Location: cytoplasm (usually)
Function: Three main types (mRNA, rRNA, and tRNA) which are all involved in steps of protein synthesis. Importantly, RNA helps convert genetic information contained in DNA into a format that can be used to make proteins.

</td></tr>
</table>

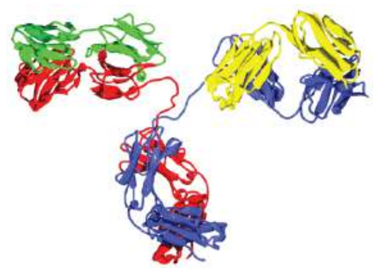
Proteins

What are proteins?

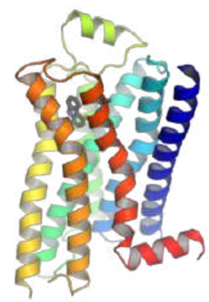
They are essential components of all living organisms and are made up of one or more chains of amino acids.



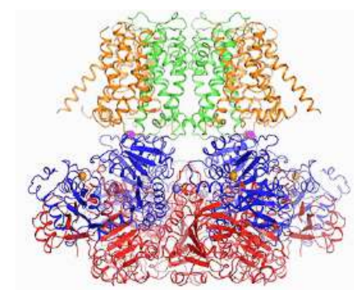
Enzyme



Antibody

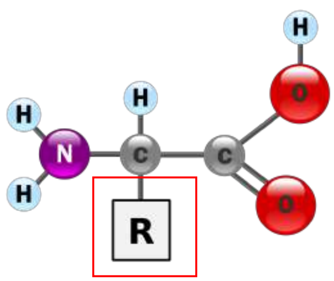


Surface Receptor

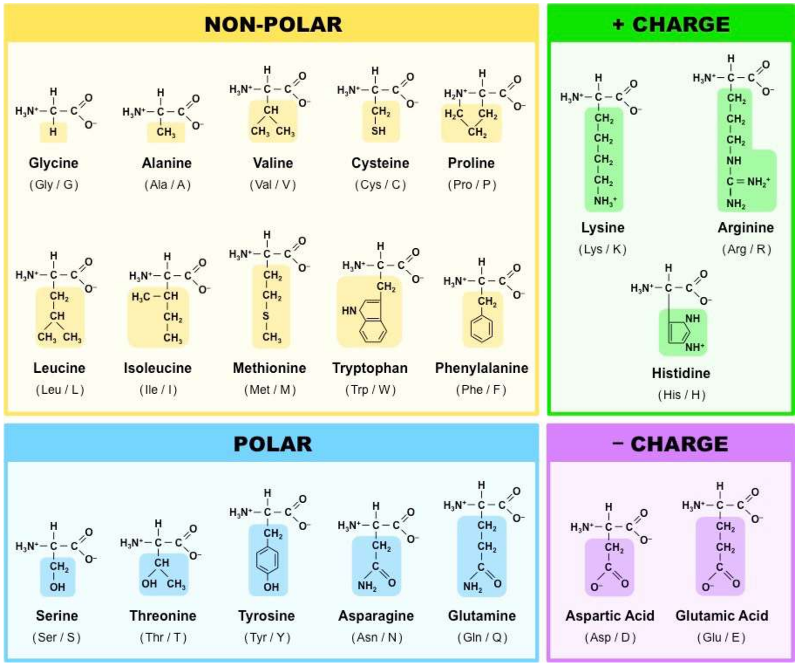


Ion channel

Proteins: amino acids



Generic structure of amino acid



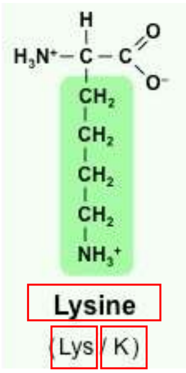
- Proteins are comprised of chains of amino acids linked together
- 20 different amino acids
- Each amino acid has the same basic structure with a different “R” group

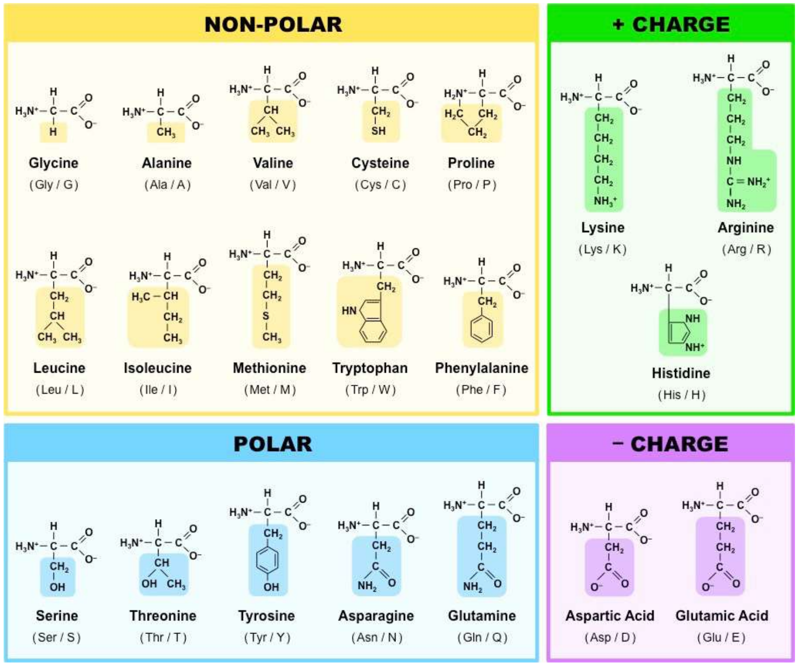
Proteins: amino acids

Naming Amino Acids:

- Full name for each amino acid
- Three letter abbreviation
- Single letter abbreviations

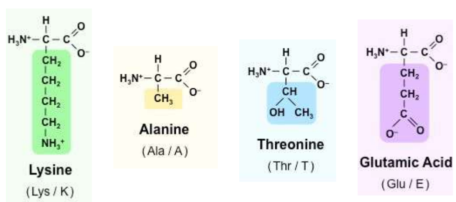
Example:





Proteins: formation

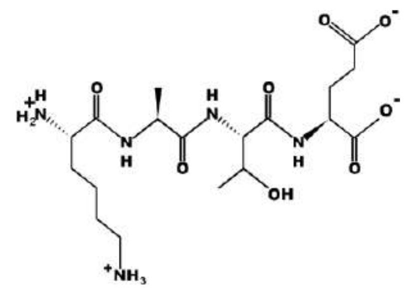
Have 4 amino acids: Lys (K), Ala (A), Thr (T), Glu (E)



In the same way that DNA and RNA are comprised of nucleotides linked together, proteins are made of amino acids linked together.

Join amino acids together
(via peptide bond)

K-A-T-E Peptide



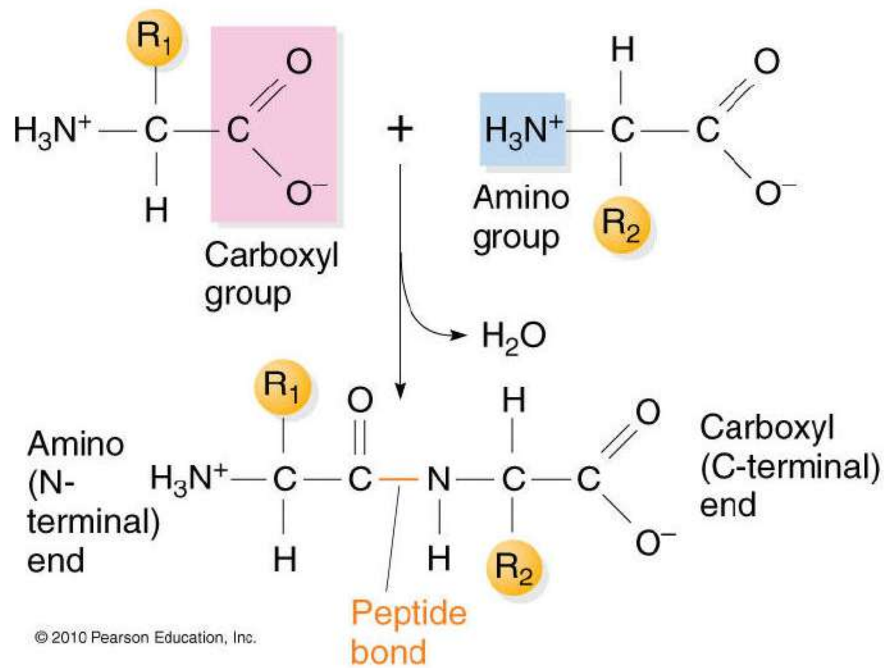
What is a peptide vs. a protein?

They're both made of amino acids, peptides are just smaller. Proteins are MANY amino acids linked together while peptides are a few amino acids linked together. (It is basically just a size thing.)

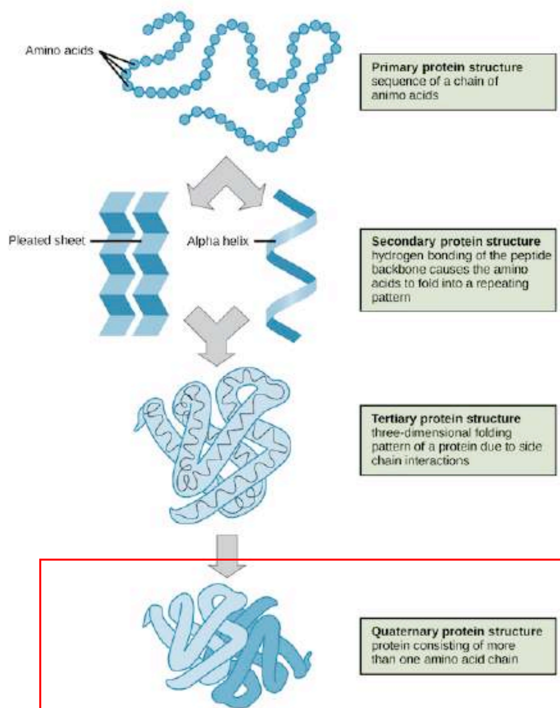
Proteins: peptide bond formation

If you are really curious…

- Peptide bonds are what link amino acids together
- Chains of amino acids linked together form peptides/proteins



Proteins: levels of structure

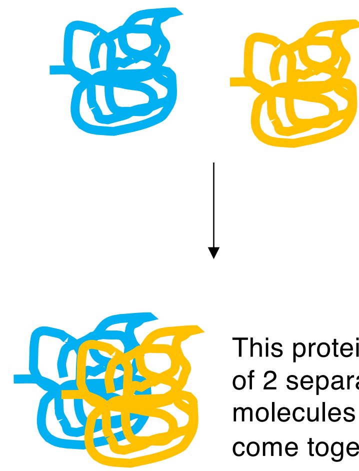


Amino acids linked together

Linked together amino acids interact with each other
+
Causes protein to fold into various shapes

Different linked chains of amino acids can interact

Separate protein molecules



This protein consists of 2 separate protein molecules that have come together

Multi-subunit protein

Example: multi-subunit protein

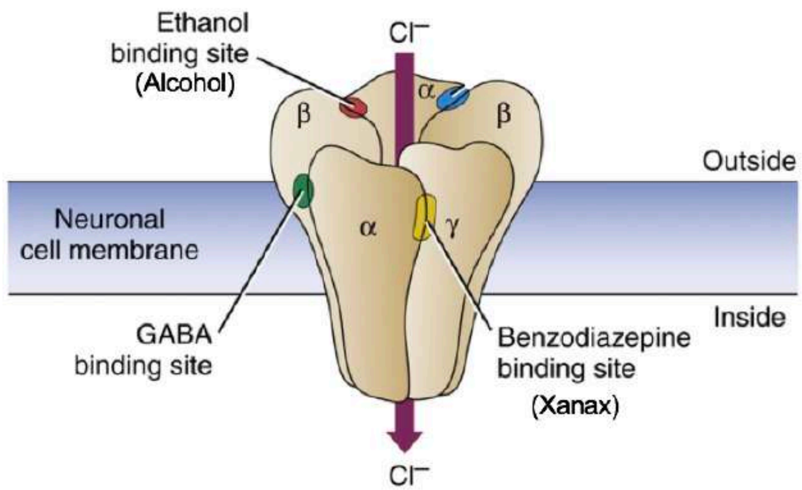
GABA$_A$ Receptor

- Ion channel and receptor that sits in neuronal cell membrane (ligand gated ion channel)

- Comprised of multiple protein subunits that come together (α, β, $\boldsymbol{\gamma}$)
    - Each subunit is a separate protein molecule comprised of linked amino acids

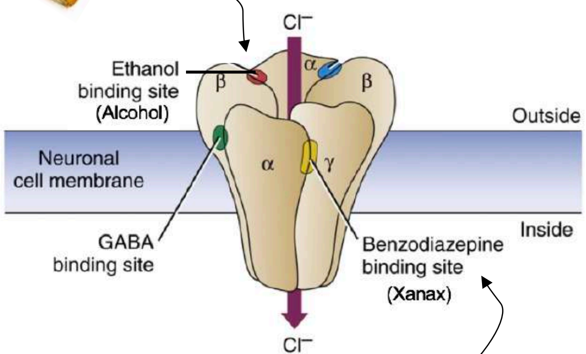
- GABA, the major inhibitory neurotransmitter in the brain, binds, channel opens, and Cl- flows through
    - Inhibitory neurotransmitter: makes neurons less like to ”fire” (i.e. signal to each other)
    - Decreases neuron activity essentially

- **GABA$_A$ receptor also has binding sites for ethanol and benzodiazepines**



Example: multi-subunit protein

Alcohol (Ethanol): Binding increases frequency and duration of channel opening.



Xanax (benzodiazepine): Binding increases frequency of channel opening.

(Moral of the story: don't drink on benzos.)

Central Nervous System (CNS) Depression

(Can have respiratory depression, unconsciousness, etc.)

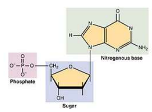
Potential death.

Don't mix "downers"

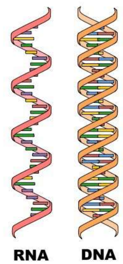
Because you might literally die.

"Downers": benzos, alcohol, opioids, etc.

Summary: “Building Blocks” and Macromolecules

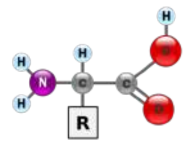


Nucleotides

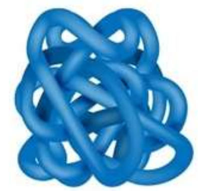


Nucleic acids

Nucleic acids are “built” of nucleotides

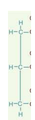


Amino Acid



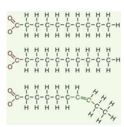
Protein

Proteins are “built” of amino acids

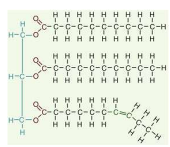


Glycerol

+

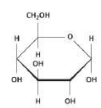


Fatty acid chains

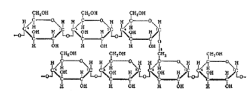


Triglyceride

Lipids are “built” of fatty acids + other components (eg glycerol)



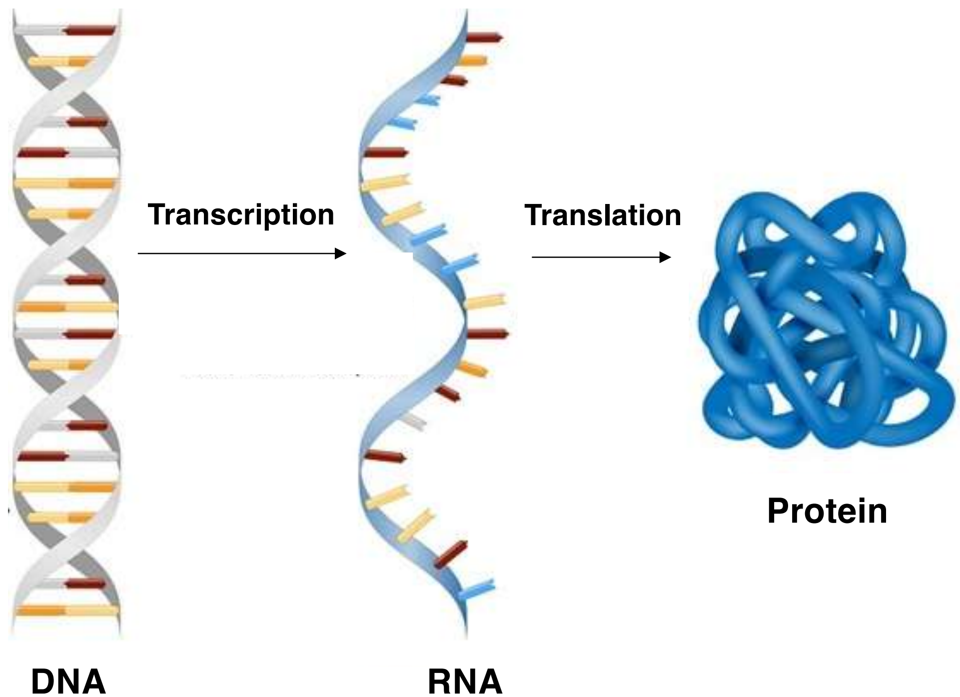
Monosaccharide



Carbohydrate (Polysaccharide)

Carbohydrate are “built” of monosaccharides

How do we go from DNA to protein?



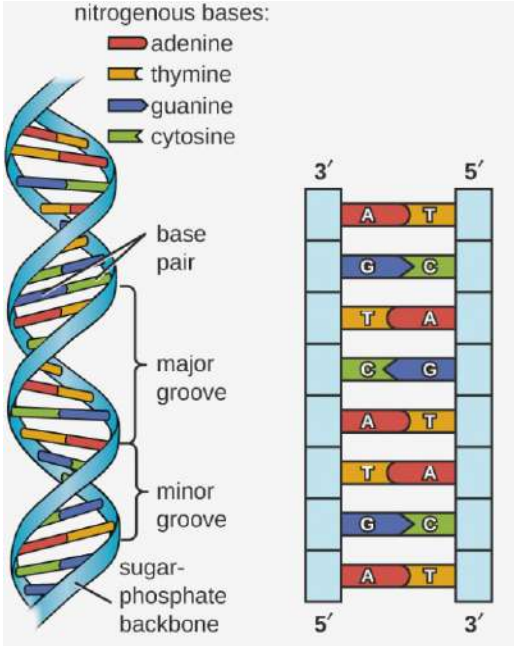
Central Dogma:

- Framework for how genetic information flows from DNA to a protein
- **Transcription:** DNA is transcribed into RNA
- **Translation:** RNA is translated into protein

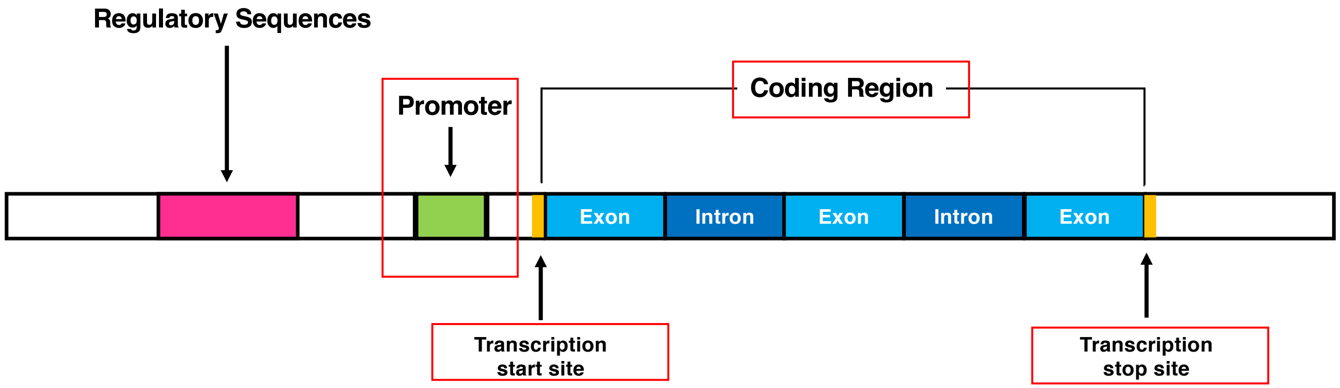
What is a Gene?

- **DNA carries information about how to make proteins:**
  - The order or sequence of bases (A, C, T, G) functions as the “code” for a protein
  - DNA essentially contains the ”instructions” to make various proteins

- **Genes are made up of DNA**
  - Genes are just **stretches of DNA** that **contain a code for a given protein**
  - They function as the basic unit of heredity - they can be passed down to future generations (how you inherit traits)



Three important parts of a gene



Regulatory Sequences: Don't worry about this (for now)

(1) Promoter: Can affect rate and frequency at which protein is made and can also turn gene "on" and "off"

(2) Coding region: Contains actual information (sequence) to make the protein and is made up of exons and introns (introns are removed from RNA via "splicing" but are present in the DNA)

Exon: contain information coding for a protein
Intron: does not contain protein information

(3) Transcription start/stop sites: Where transcription starts and ends respectively

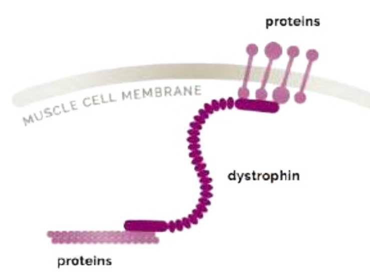
How big are genes?

Gene size really varies - they can be really big or really tiny! Here are some actual stats though…

Transfer RNA (tRNA)

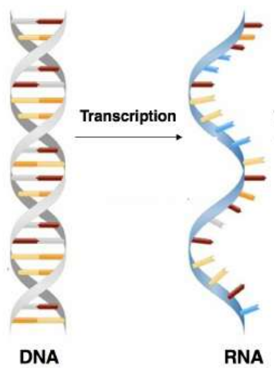
Smallest genes: the transfer RNAs (tRNAs) mentioned briefly in earlier slides have genes that are just **76 base pairs long.**

Largest gene: the dystrophin protein (important in Duchenne's muscular dystrophy) is **2.4 million base pairs long.**



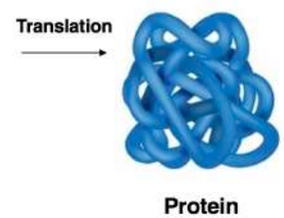
What is the “genetic code”?

I keep saying that DNA carries information about how to make proteins and that it is coded in the base sequence – how does this actually work?



Nucleotide bases in a specific sequence: A, G, C, T (DNA) or U (RNA)

How do we turn nucleotide base sequence into amino acid sequence?

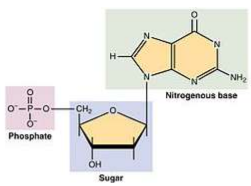


Amino acids linked together

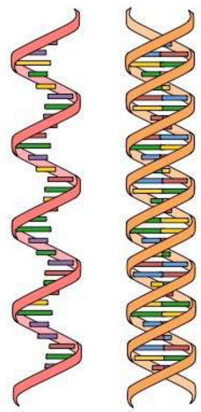
What is the “genetic code”?

(Reminder)

A, C, G, T (DNA)
or U (RNA)



Nucleotides

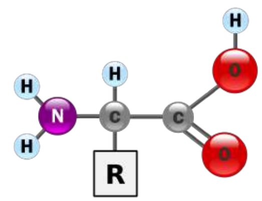


RNA **DNA**

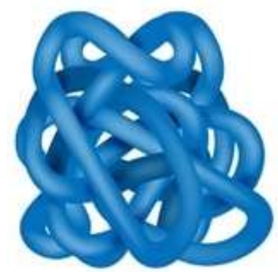
Nucleic acids

Nucleic acids are “built” of nucleotides

20 different
amino acids



Amino Acid

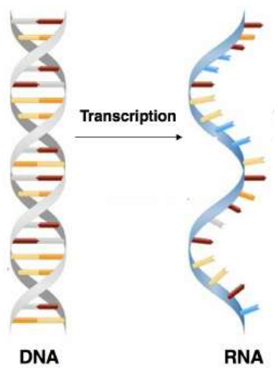


Protein

Proteins are “built” of amino acids

What is the “genetic code”?

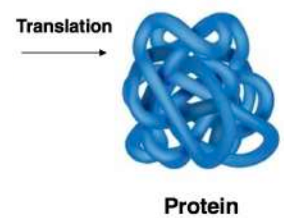
I keep saying that DNA carries information about how to make proteins and that it is coded in the base sequence – how does this actually work?



Nucleotide bases in a specific sequence: A, G, C, T (DNA) or U (RNA)

	U	C	A	G
U	UUU = phe UUC = phe UUA = leu UUG = leu	UCU = ser UCC = ser UCA = ser UCG = ser	UAU = tyr UAC = tyr UAA = stop UAG = stop	UGU = cys UGC = cys UGA = stop UGG = trp
C	CUU = leu CUC = leu CUA = leu CUG = leu	CCU = pro CCC = pro CCA = pro CCG = pro	CAU = his CAC = his CAA = gln CAG = gln	CGU = arg CGC = arg CGA = arg CGG = arg
A	AUU = ile AUC = ile AUA = ile AUG = met	ACU = thr ACC = thr ACA = thr ACG = thr	AAU = asn AAC = asn AAA = lys AAG = lys	AGU = ser AGC = ser AGA = arg AGG = arg
G	GUU = val GUC = val GUA = val GUG = val	GCU = ala GCC = ala GCA = ala GCG = ala	GAU = asp GAC = asp GAA = glu GAG = glu	GGU = gly GGC = gly GGA = gly GGG = gly

There’s a code.



Amino acids linked together

What is the “genetic code”?

Example: how to ”read” the code

--AACGTCGACTGG--
--TTGCAGCTGACC--

DNA sequence

AACGUCGACUGG

mRNA sequence

mRNA is copied from the DNA
*Only difference is that T’s are now U’s

What is the “genetic code”?

Example: how to ”read” the code

--AACGTCGACTGG--
--TTGCAGCTGACC--

DNA sequence

→

AAC GUC GAC UGG

mRNA sequence

Let’s organize the mRNA sequence into groups of 3

These groups of 3 are called “codons”

+

Each codon corresponds to an amino acid

What is the “genetic code”?

Example: how to ”read” the code

--AACGTCGACTGG--
--TTGCAGCTGACC--

DNA sequence

AAC GUC GAC UGG

mRNA sequence

Each codon corresponds to a certain amino acid

	U	C	A	G
U	UUU = phe UUC = phe UUA = leu UUG = leu	UCU = ser UCC = ser UCA = ser UCG = ser	UAU = tyr UAC = tyr UAA = stop UAG = stop	UGU = cys UGC = cys UGA = stop UGG = trp
C	CUU = leu CUC = leu CUA = leu CUG = leu	CCU = pro CCC = pro CCA = pro CCG = pro	CAU = his CAC = his CAA = gln CAG = gln	CGU = arg CGC = arg CGA = arg CGG = arg
A	AUU = ile AUC = ile AUA = ile AUG = met	ACU = thr ACC = thr ACA = thr ACG = thr	AAU = asn AAC = asn AAA = lys AAG = lys	AGU = ser AGC = ser AGA = arg AGG = arg
G	GUU = val GUC = val GUA = val GUG = val	GCU = ala GCC = ala GCA = ala GCG = ala	GAU = asp GAC = asp GAA = glu GAG = glu	GGU = gly GGC = gly GGA = gly GGG = gly

***There is redundancy in the system – single amino acid may be coded for by more than 1 codon**

Ex: UUA, UUG, CUU, CUC, CUA, CUG all code for leucine (leu)

What is the “genetic code”?

Example: how to ”read” the code

--AACGTCGACTGG--
--TTGCAGCTGACC--

DNA sequence

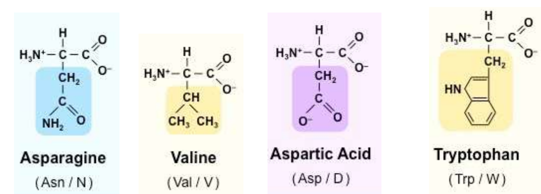
→ AAC GUC GAC UGG

mRNA sequence

→ **Each codon corresponds to a certain amino acid**

	U	C	A	G
U	UUU = phe UUC = phe UUA = leu UUG = leu	UCU = ser UCC = ser UCA = ser UCG = ser	UAU = tyr UAC = tyr UAA = stop UAG = stop	UGU = cys UGC = cys UGA = stop UGG = trp
C	CUU = leu CUC = leu CUA = leu CUG = leu	CCU = pro CCC = pro CCA = pro CCG = pro	CAU = his CAC = his CAA = gln CAG = gln	CGU = arg CGC = arg CGA = arg CGG = arg
A	AUU = ile AUC = ile AUA = ile AUG = met	ACU = thr ACC = thr ACA = thr ACG = thr	AAU = asn AAC = asn AAA = lys AAG = lys	AGU = ser AGC = ser AGA = arg AGG = arg
G	GUU = val GUC = val GUA = val GUG = val	GCU = ala GCC = ala GCA = ala GCG = ala	GAU = asp GAC = asp GAA = glu GAG = glu	GGU = gly GGC = gly GGA = gly GGG = gly

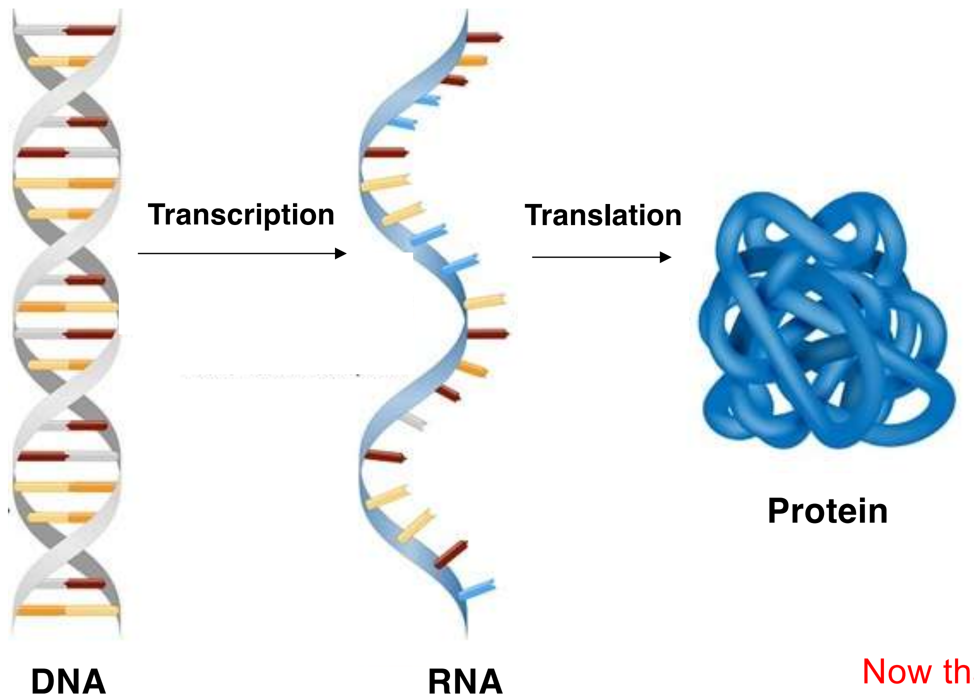
→



Asparagine – Valine –Aspartic Acid –Tryptophan

Protein sequence

How do we go from DNA to protein?



Central Dogma:

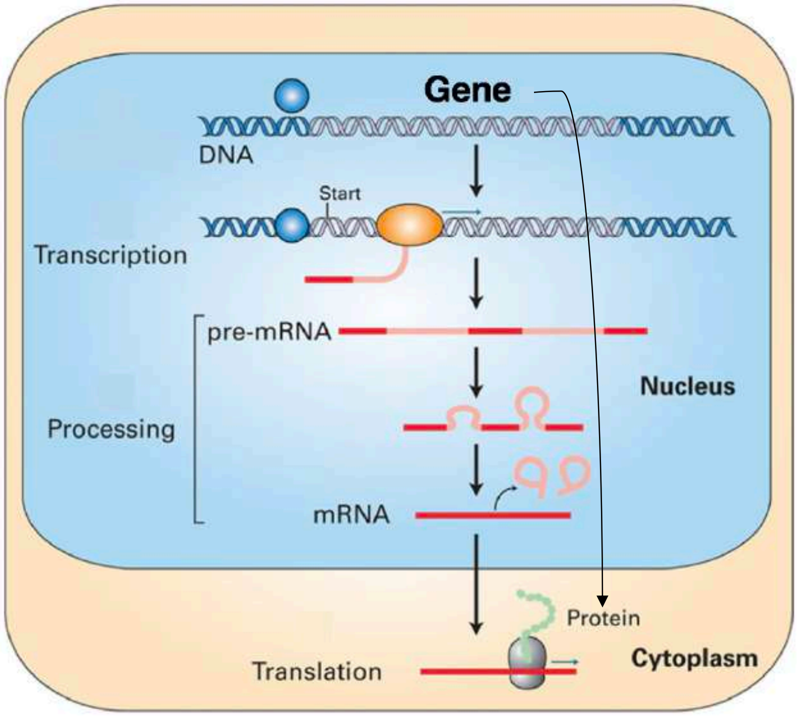
- Framework for how genetic information flows from DNA to a protein
- **Transcription:** DNA is transcribed into RNA
- **Translation:** RNA is translated into protein

Now that we know how the “code” works let’s go through this whole process.

Protein production: DNA to protein

Goal:

Get from coded information in DNA (in the genes) to fully formed functional protein

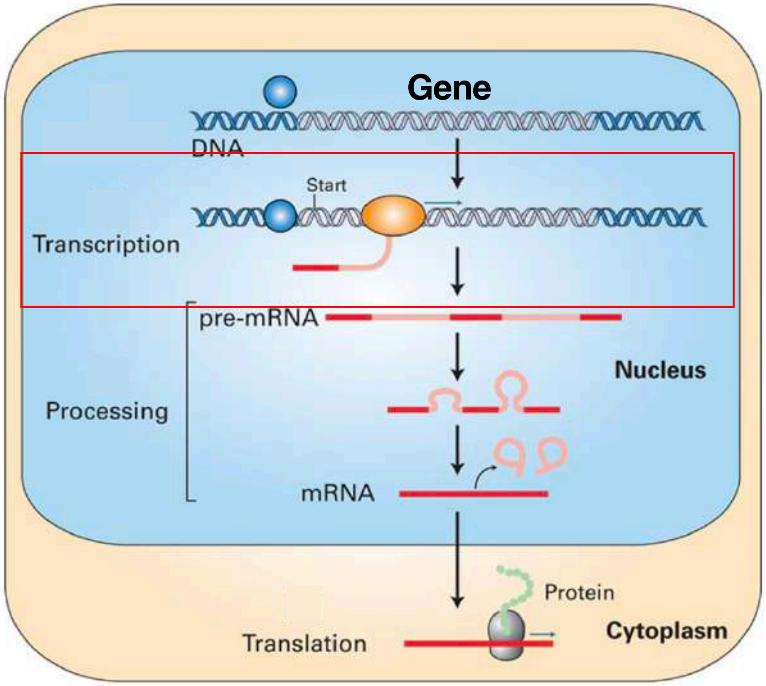


Protein production step 1: transcription (DNA to RNA)

Transcription goal: convert genetic information in DNA to RNA (mRNA form)

- **DNA in gene coding for protein is "transcribed"**
  - Transcription: protein coding region of a gene is copied into mRNA
  - mRNA sequence = same as one of DNA strands

- DNA contains introns and exons thus, **pre-mRNA** also contains introns and exons



Transcription: a closer look

RNA Polymerase: enzyme that synthesizes mRNA molecule using the DNA strand as a template

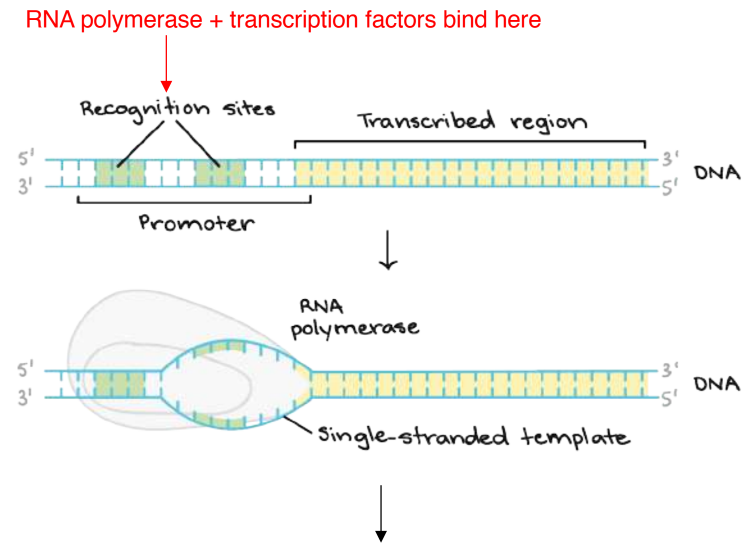
Transcription factor: protein that binds to specific DNA sequence and controls rate of transcription – in general they function to regulate genes (turn them on/off)

Three big steps:

(1) Initiation:
- RNA polymerase + 1 or more transcription factors bind to promoter region
- RNA polymerase separates 2 strands of the DNA helix ("transcription bubble")

(2) Elongation
- RNA polymerase begins transcription at start site
- RNA polymerase adds RNA nucleotides to mRNA strand (copied off of DNA strand)

(3) Termination
- RNA polymerase releases from template DNA
- Complete pre-mRNA molecule dissociates

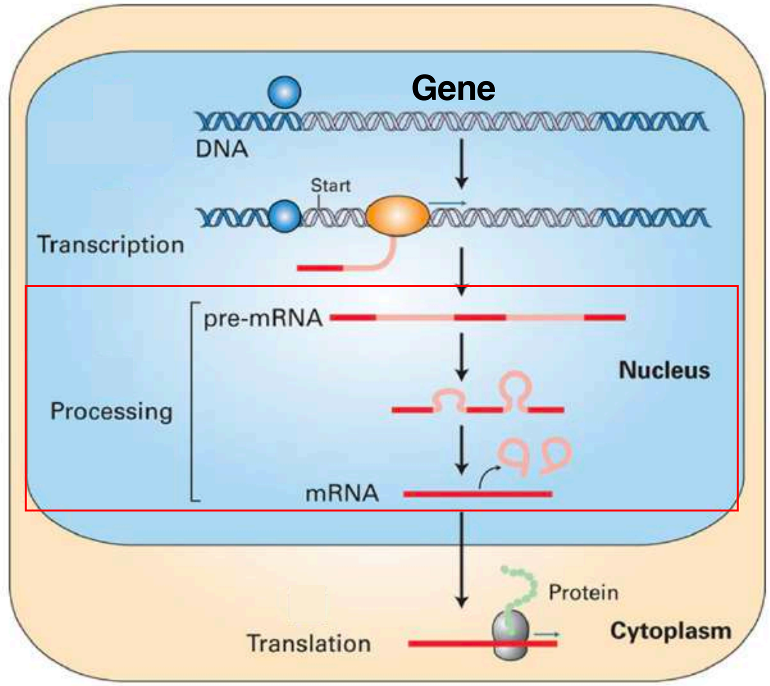


Pre-mRNA molecule

Protein production step 1: transcription (DNA to RNA)

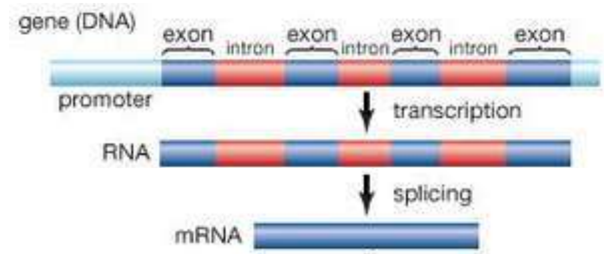
Transcription goal: convert genetic information in DNA to RNA (mRNA form)

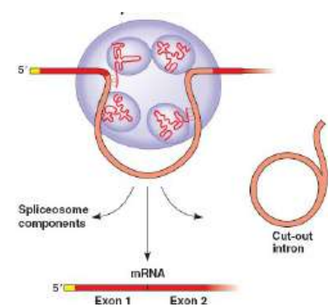
- **DNA in gene coding for protein is "transcribed"**
  - Transcription: protein coding region of a gene is copied into mRNA
  - mRNA sequence = same as one of DNA strands
- DNA contains introns and exons thus, **pre-mRNA** also contains introns and exons
- **Pre-mRNA processing**
  - Introns get cut out and exons are joined together
  - Term for this is **"splicing"**
  - Other modifications also made



Splicing: removing introns from pre-mRNA

- **Splicing:** removing introns from a pre-mRNA transcript to form a mature mRNA transcript
- Exons contain the actual protein coding information
- Exons are made up of codons which can then be read by the ribosome
- Introns do not contain coding information
- **Spliceosome:** the machinery that actually removes the introns (can recognize characteristic “splice sites” between introns and exons)





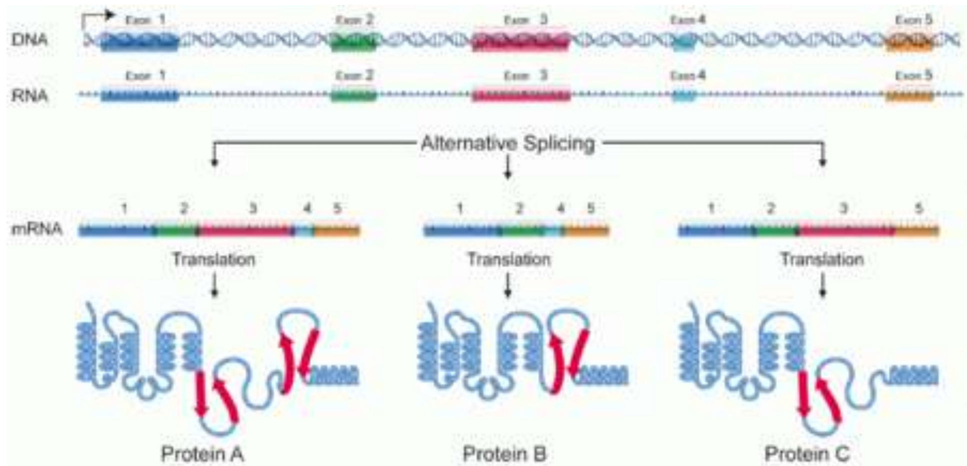
Spliceosome: large molecular “machine” comprised of proteins and small nuclear RNAs

Alternative splicing

Why even have introns if they just get cut out?

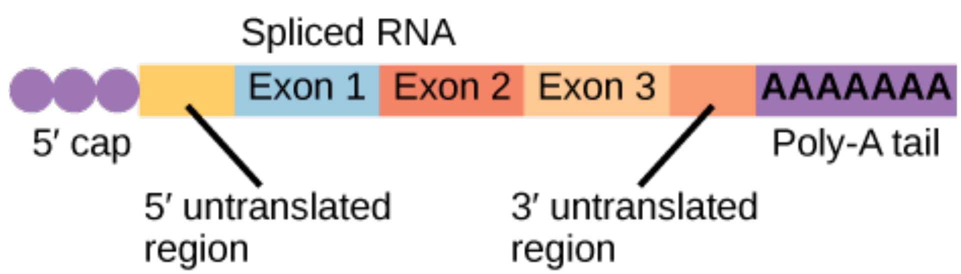
Having introns though allows us to “splice” things a lot of different ways in order to combine different exons. Combining different exons results in different proteins being created from the **same** gene.

- Create different proteins by combining different exons
- Without introns, exons could not be differentially combined
- This process is called “alternative splicing”



Other Modifications

7-methyl Guanosine cap



Poly-A tail

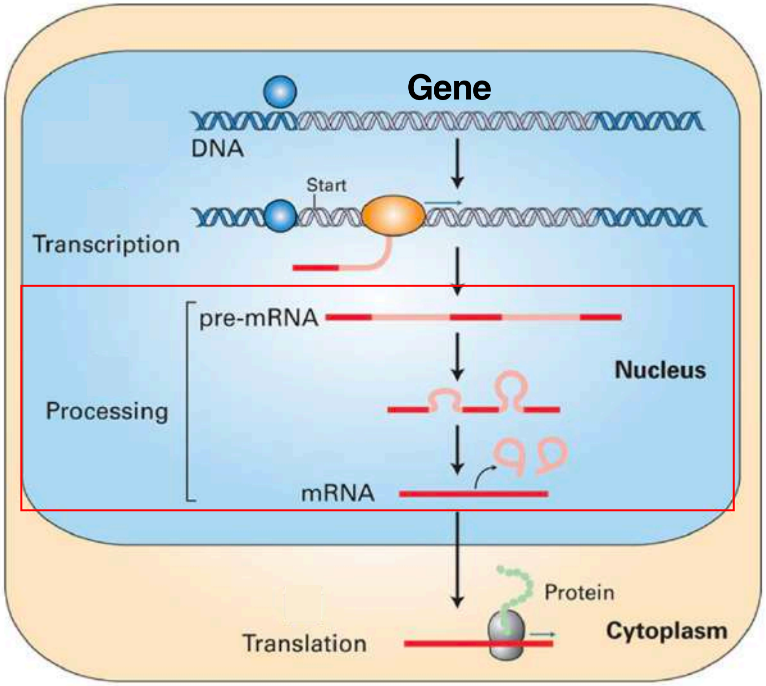
These end modifications:

- Increase mRNA stability (prevent degradation by enzymes)
- Assist with mRNA transport out of nucleus
- Promote translation (mRNA → ribosome)

Protein production step 1: transcription (DNA to RNA)

Transcription goal: convert genetic information in DNA to RNA (mRNA form)

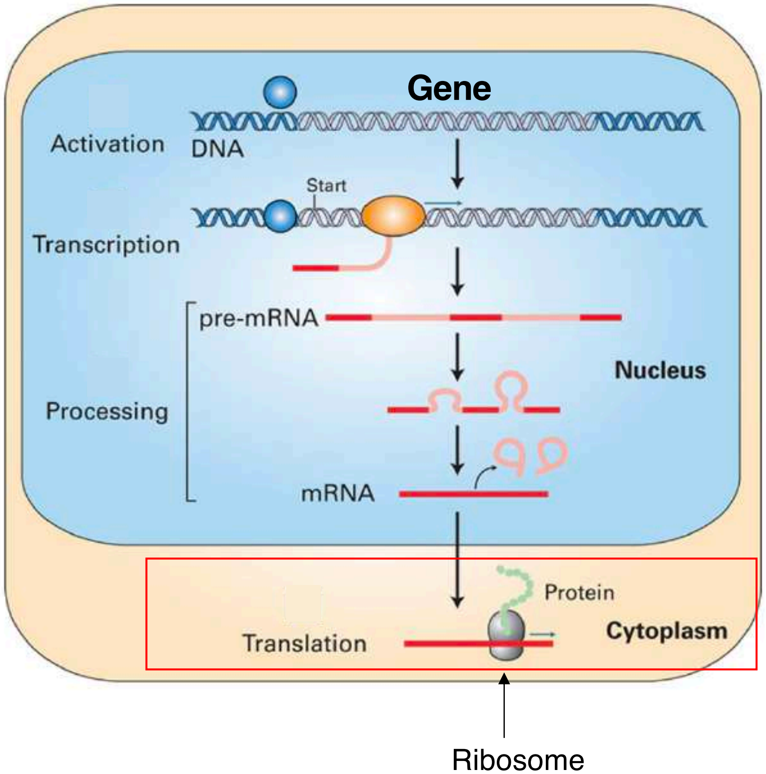
- **DNA in gene coding for protein is "transcribed"**
  - Transcription: protein coding region of a gene is copied into mRNA
  - mRNA sequence = same as one of DNA strands
- DNA contains introns and exons thus, **pre-mRNA** also contains introns and exons
- **Pre-mRNA processing**
  - Introns get cut out and exons are joined together
  - Term for this is **"splicing"**
  - Other modifications also made
- **Mature mRNA is ready for the next step:**
  - Translation (RNA→ protein)



Protein production: translation (RNA to protein)

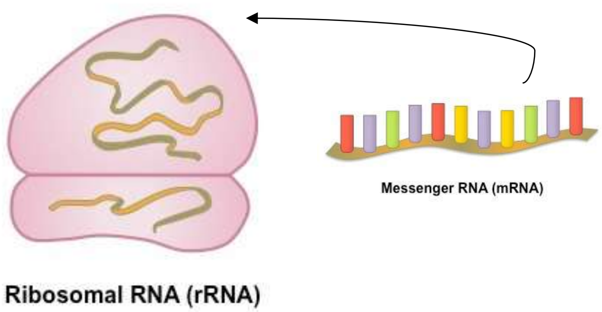
Translation goal: use information in the RNA (mRNA) molecule to make a protein

- Processed mRNA leaves nucleus and goes to the cytoplasm
- Ribosome and mRNA bind to each other
- Ribosome "reads" mRNA code and assembles a protein by linking amino acids together in a chain
- Ribosome reads until a stop signal in mRNA

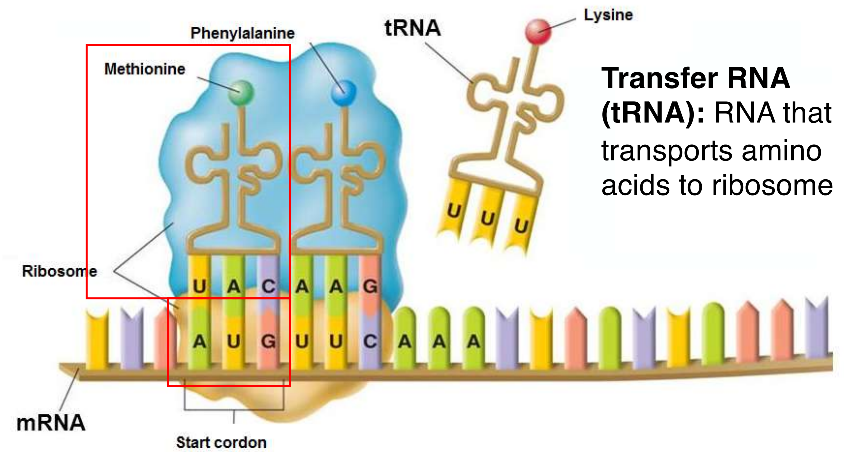


Translation: a closer look

Ribosome: rRNA + protein complex that binds mRNA and synthesizes protein



Ribosome and mRNA bind



Transfer RNA (tRNA): RNA that transports amino acids to ribosome

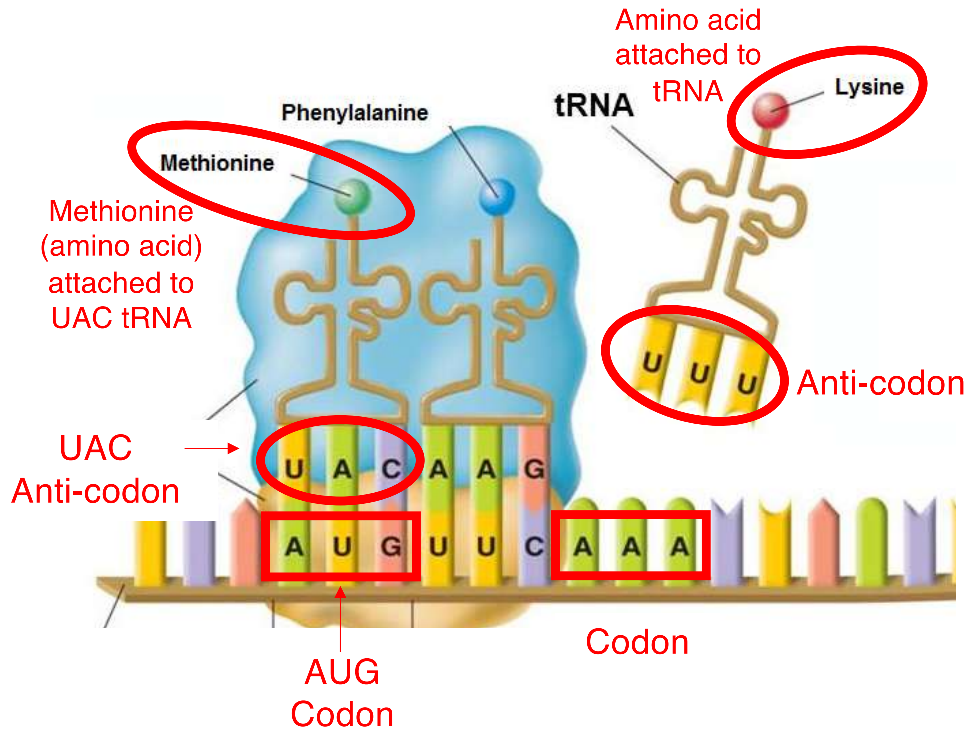
(1) Ribosome beings "reading" mRNA at start codon
Eg: A U G → Methionine

(2) tRNA brings amino acid corresponding to that codon to the ribosome
Eg: Methionine brought by tRNA

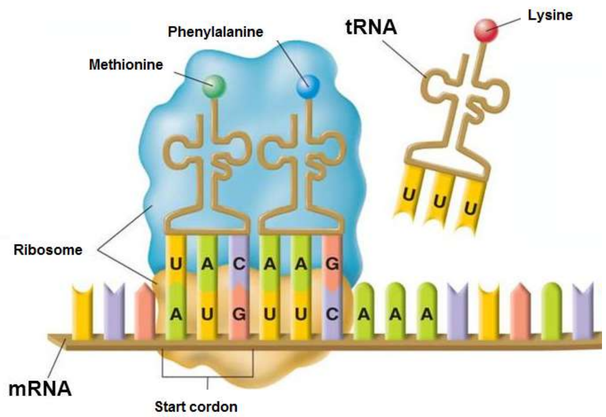
Protein production: translation (RNA to protein)

Transfer RNAs (tRNAs)

- Remember: codons (groups of 3 nucleotides) code for amino acids
- tRNAs on one end contain an "anti-codon" which can base pair with a given mRNA codon
- On the other end, tRNAs carry the amino acid that their anti-codon base pairs with
- **For example:**
  - **Codon:** A-U-G (codes for methionine)
  - **tRNA anti-codon:** U-A-C
  - tRNA with U-A– anticodon will be carrying methionine amino acid



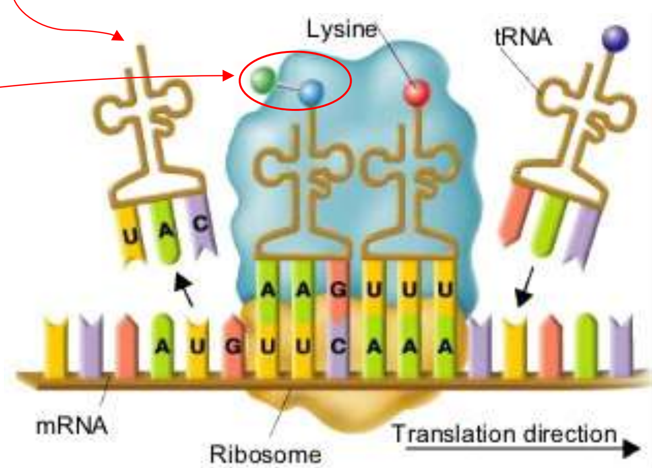
Translation: a closer look



Ribosome continues to read codons and tRNAs bring amino acids

Ribosome joins amino acids together and breaks bond between tRNA + amino acid

Once "unloaded" tRNAs dissociate leaving room for the next one to bind

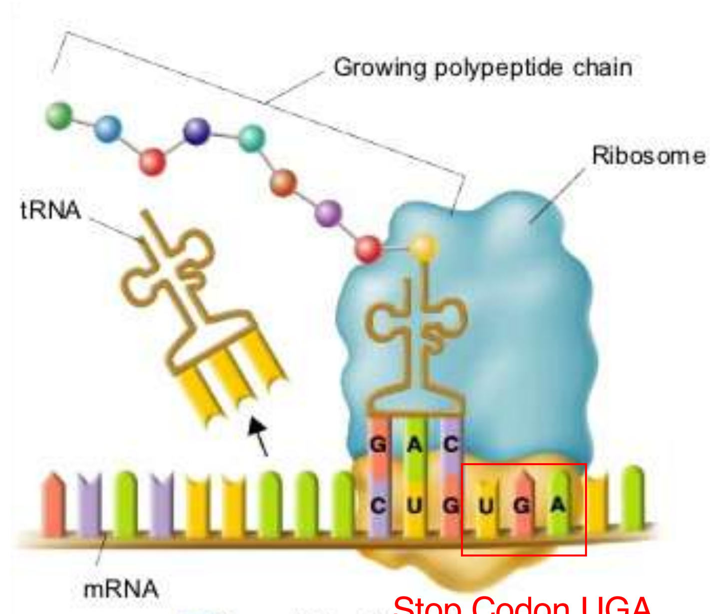


This process is repeated
(Read codon, tRNA brings in corresponding amino, ribosome links amino acid to growing peptide chain, tRNA dissociates)

Translation: a closer look

- Ribosome will keep "reading" codons in mRNA transcript until it reaches a "stop codon" – signals that translation is over
- **Three stop codons:** UAG, UGA, UAA

	U	C	A	G
U	UUU = phe UUC = phe UUA = leu UUG = leu	UCU = ser UCC = ser UCA = ser UCG = ser	UAU = tyr UAC = tyr UAA = stop UAG = stop	UGU = cys UGC = cys UGA = stop UGG = trp
C	CUU = leu CUC = leu CUA = leu CUG = leu	CCU = pro CCC = pro CCA = pro CCG = pro	CAU = his CAC = his CAA = gln CAG = gln	CGU = arg CGC = arg CGA = arg CGG = arg
A	AUU = ile AUC = ile AUA = ile AUG = met	ACU = thr ACC = thr ACA = thr ACG = thr	AAU = asn AAC = asn AAA = lys AAG = lys	AGU = ser AGC = ser AGA = arg AGG = arg
G	GUU = val GUC = val GUA = val GUG = val	GCU = ala GCC = ala GCA = ala GCG = ala	GAU = asp GAC = asp GAA = glu GAG = glu	GGU = gly GGC = gly GGA = gly GGG = gly

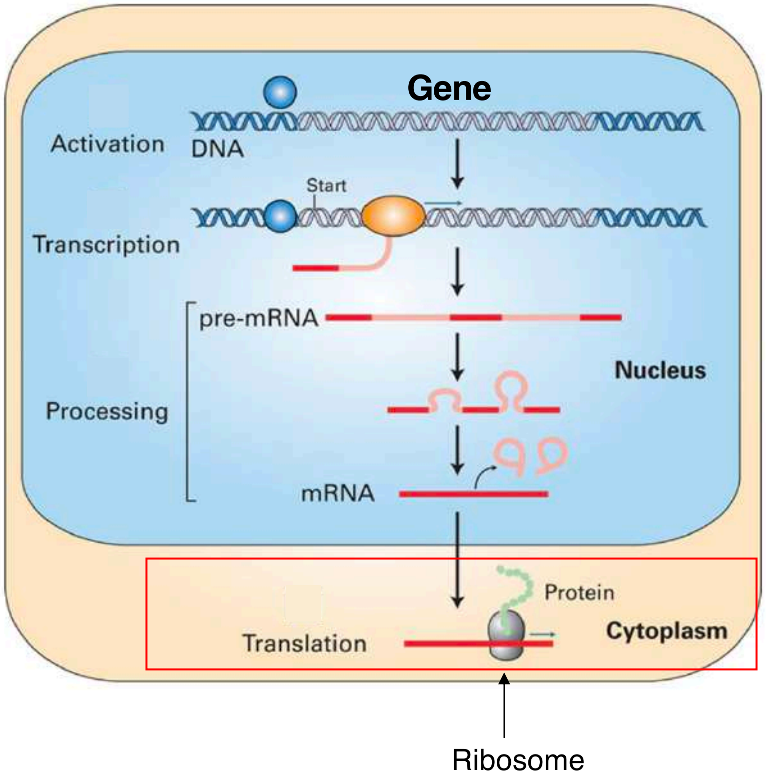


Stop Codon UGA

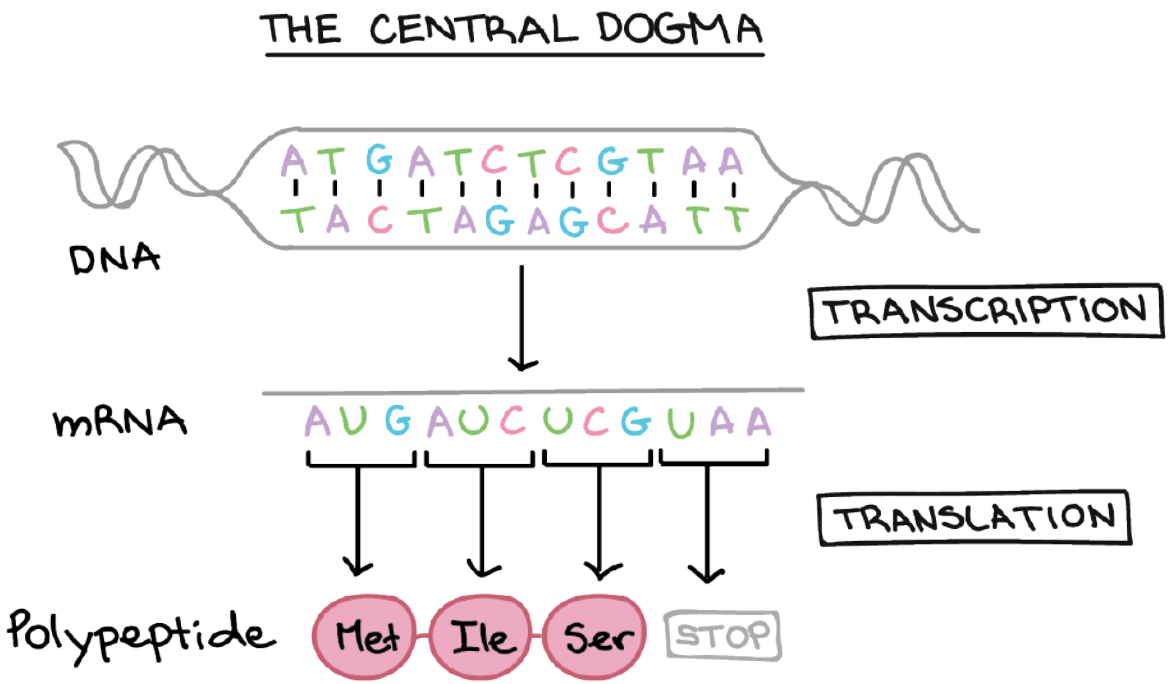
Protein production: translation (RNA to protein)

Translation goal: use information in the RNA (mRNA) molecule to make a protein

- Processed mRNA leaves nucleus and goes to the cytoplasm
- Ribosome and mRNA bind to each other
- Ribosome “reads” mRNA code and assembles a protein by linking amino acids together in a chain
- Ribosome reads until a stop signal in mRNA
- **Protein is produced!**
- Other modifications can be made to the protein i.e. post-translational modifications (PTMs) (discussed on future slide)



Summary


THE CENTRAL DOGMA
A T G A T C T C G T A A
T A C T A G A G C A T T
DNA
TRANSCRIPTION
mRNA
A U G A U C U C G U A A
TRANSLATION
Polypeptide
Met
Ile
Ser
STOP

Post-translational modifications (PTMs)

What are post translational modifications (PTMs)?

PTMs are basically things (modifications or processing) that happen to proteins after they are made. It's called "post translational' because it happens after translation.

Important Fact about PTMS:

- Some proteins are modified after they are made (cut, have things like carbohydrates added to them, etc.)

- **Different examples of PTMs:**
  - Formation of disulfide bonds
  - Specific cleavages
  - Addition of carbohydrates (glycosylation):

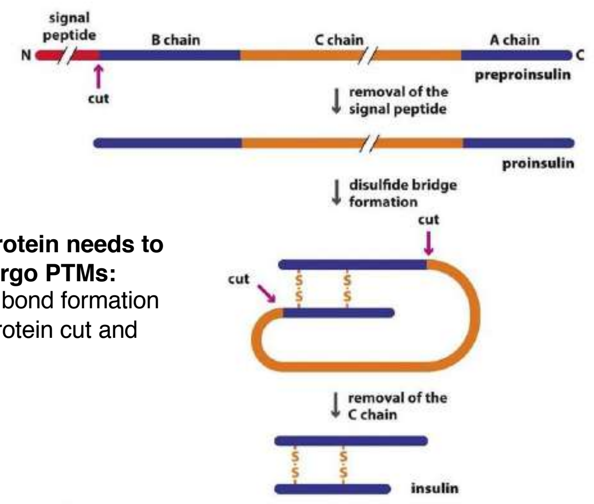
    Short O- linked chains

    Long N- linked chains

Example: Insulin protein

Insulin protein needs to undergo PTMs:

- Disulfide bond formation
- Part of protein cut and removed



Post-translational modifications (PTMS)

All kinds of cells do not carry out same PTMs – why is this important?

- Modifications serve a functional purpose and there are consequences when they are not correct

- If you make a protein in bacteria or yeast to treat a disease in humans it may not have the right PTMs.
    - This can affect efficacy/functionality (protein might not work in human)
    - Could also result in rapid clearance (body will recognize it as non-human or potentially pathogenic)

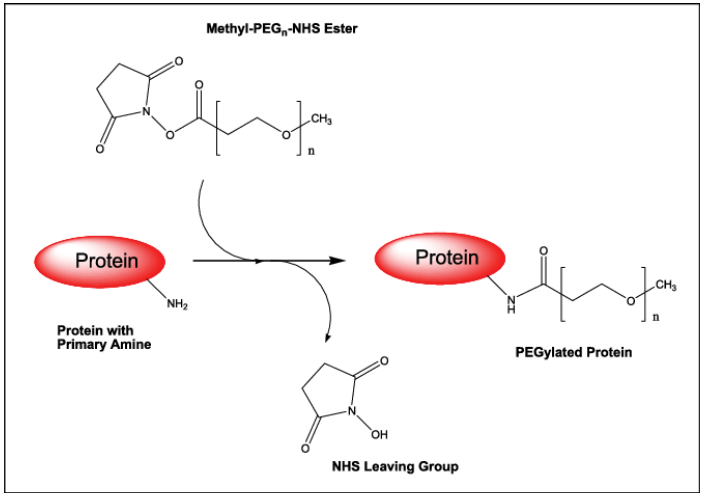
Why not make all recombinant therapeutic proteins in bacterial cells instead of mammalian cells?

- While it would be cheaper/cost effective, some recombinant protein therapeutics cannot be made in bacteria because bacteria do not have the PTM capabilities to make them correctly.
- PTMs can be just as important to a protein functioning as the actual original string of amino acids.

In short: some recombinant therapeutics need to undergo certain PTMs that bacterial cells can't do. (**The bacteria basically can't even make the correct protein which is why we don't use them all the time**.)

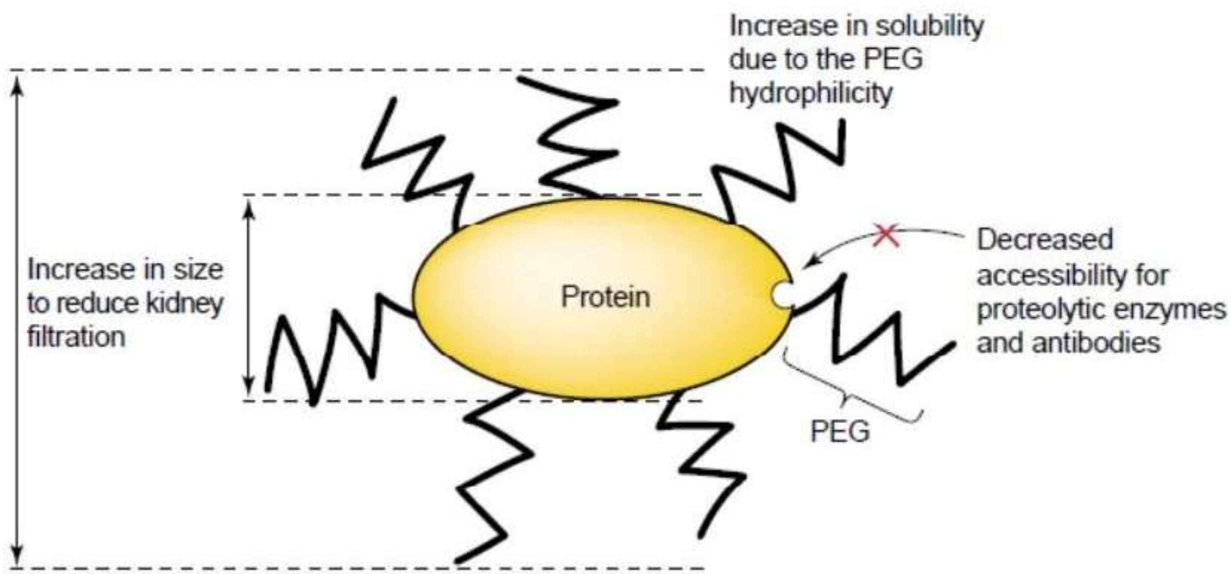
Artificial Post-translational modification (PTM)

- PTMs happen naturally to proteins being made in a cell
- PTMs can also be added when engineering a protein therapeutic
- **Engineered PTMs can be:**
  - ”Natural” i.e. they occur in nature (ex: glycosylation)
  - Can be artificial polyethylene glycol (PEG) - cells cannot naturally add this to proteins.



Artificial Post-translational modification (PTM)

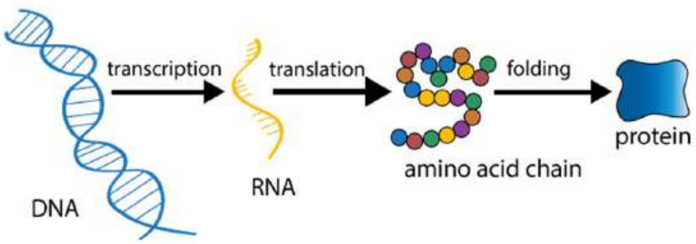
Why would we want to add PTMs to protein therapeutics?



Gene Expression

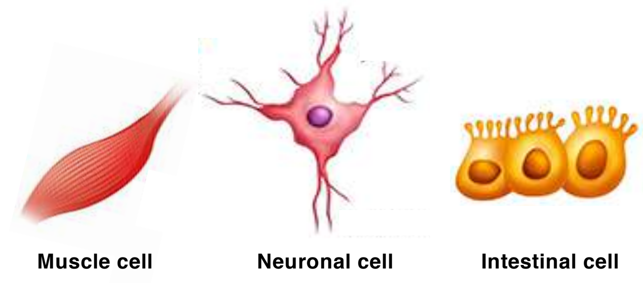
What does it mean to "express" a gene?

- Expression just means that whatever that gene codes for is being made (the gene product is being produced)
    - DNA sequence → mRNA → protein



Why is gene expression important?

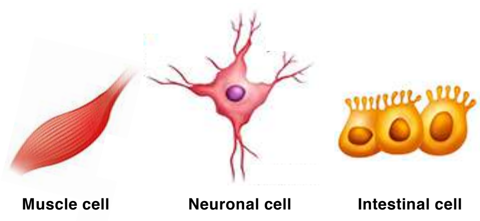
- The genes that a cell expresses dictate what proteins that cell will make
- The proteins a given cell makes determine the cell's identity and functionality
    - (i.e. whether it's a muscle cell, a neuron, a liver cell, an intestinal cell, etc.)



Gene Expression

Do all of our cells have the same DNA?

We have many different cells types in our body and each cell (**for the most part) contains the *same* genetic information – i.e. the same set of genes and the same DNA.



A muscle cell, a neuron, and an intestinal cell all make different proteins and look and do very different things but, if you looked inside each one, you would find the same set of genes (the same DNA).

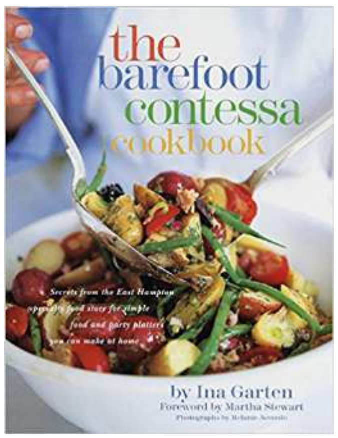
How do we get different cell types then?

Short answer: Differential gene expression. All cells have the same set of genes/genetic information but different cells express different genes (turn different genes "on") which in turn produce different proteins and result in cells that look different and do different things.

Very stupid analogy on the next slide….

Gene Expression Analogy

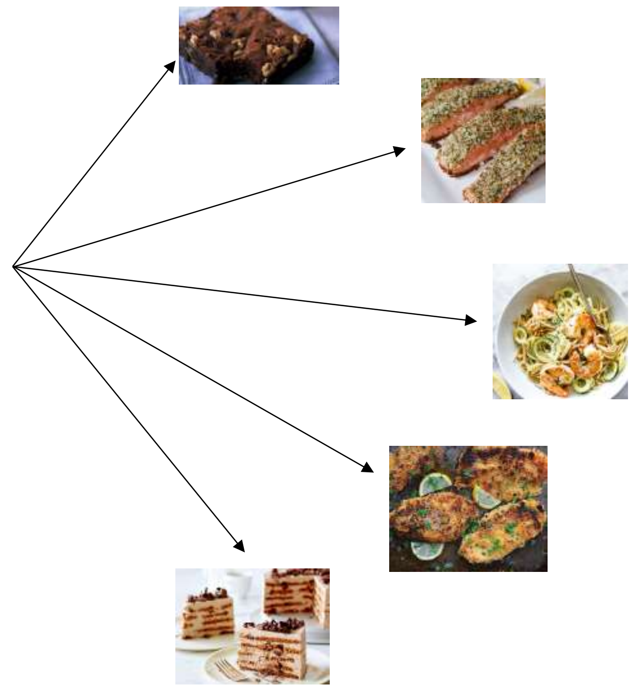
Gene expression/different cell types analogy…



All cells** have the same "cookbook" (genome) with the same set of "recipes" (genes)

Different cells choose to "look at" and "read" different "recipes" (genes)

Because of this, they make different "foods" (proteins) and are different from each other

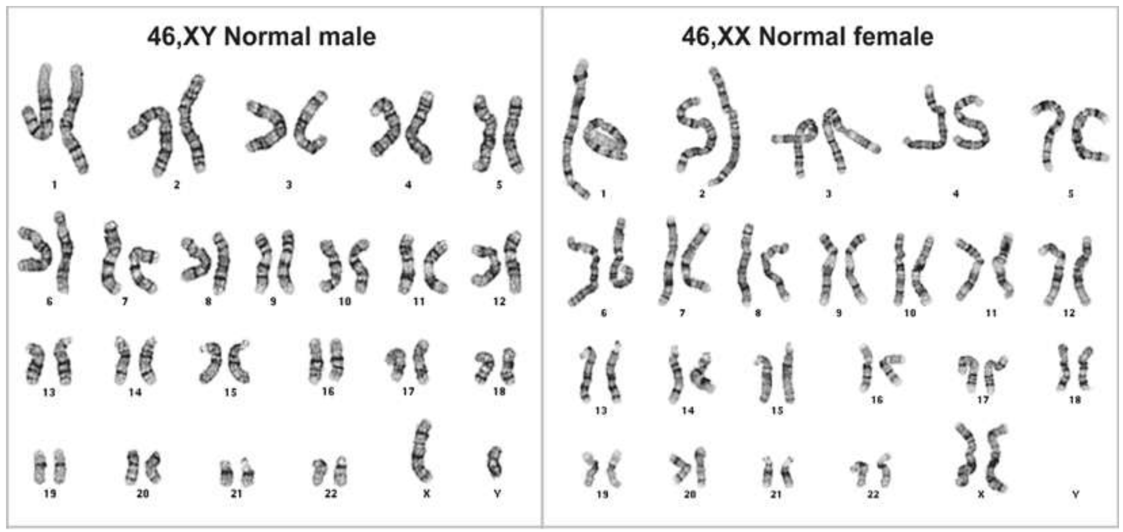


**Almost all cells with a few exceptions for example Red blood cells contain no DNA

Part 2: Useful Topics

Inheritance

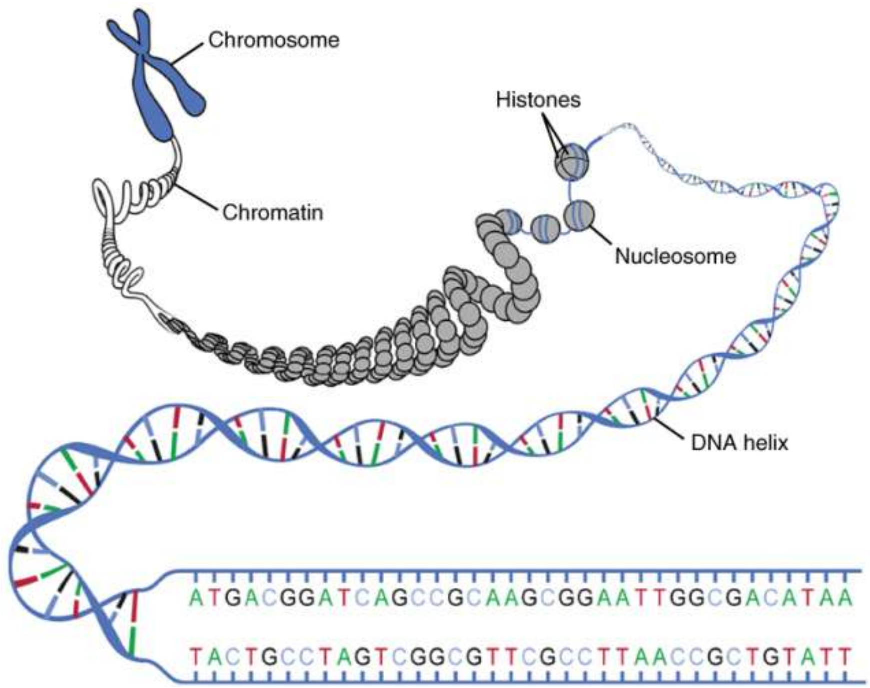
- Each cell in the body** contains 23 pairs of chromosomes (46 individual ones)
- One chromosome from each pair is inherited from your mother and one is inherited from your father.
- Chromosomes contain the genes you inherit from your parents.
- Male: XY sex chromosomes
- Female: XX sex chromosomes
- The term “Autosomal” refers to chromosomes 1- 22
- The term “sex-linked” or “X-linked” refers to the X chromosome(s)



**Sperm and eggs contain half the number of chromosomes (they have 23 single chromosomes vs. 46 total)

Inheritance: what is a chromosome?

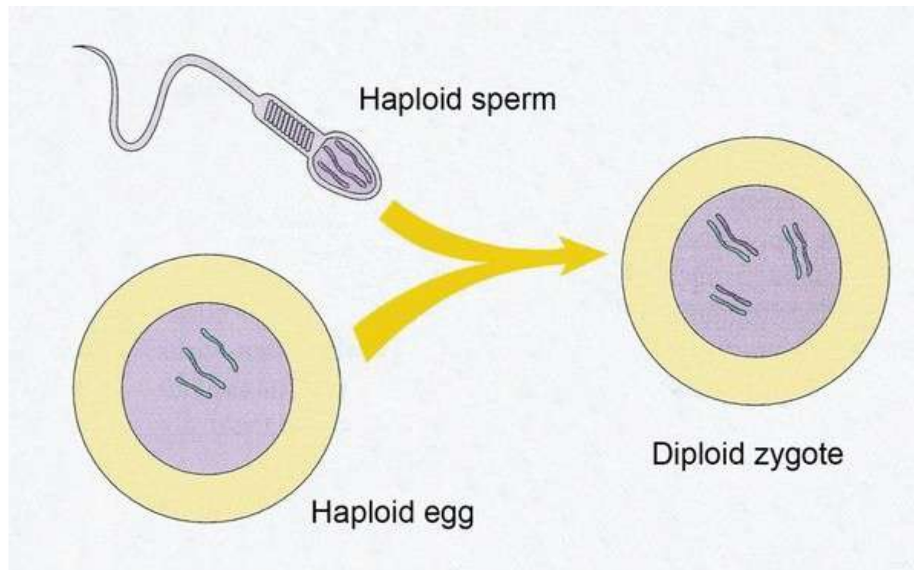
- Each of us has enough DNA to reach from here to the sun and back, more than 300 times.
- Needs to fit all the DNA inside the nucleus (and inside the cell)
- DNA is wound around groups of proteins called histones – these then complex to form nucleosomes
- Nucleosomes fold up – this then the structure condenses even more
- Final condensed form of DNA is a chromosome (in the way we typically draw it and think of it)



Chromosome = DNA + various packaging proteins that help condense it

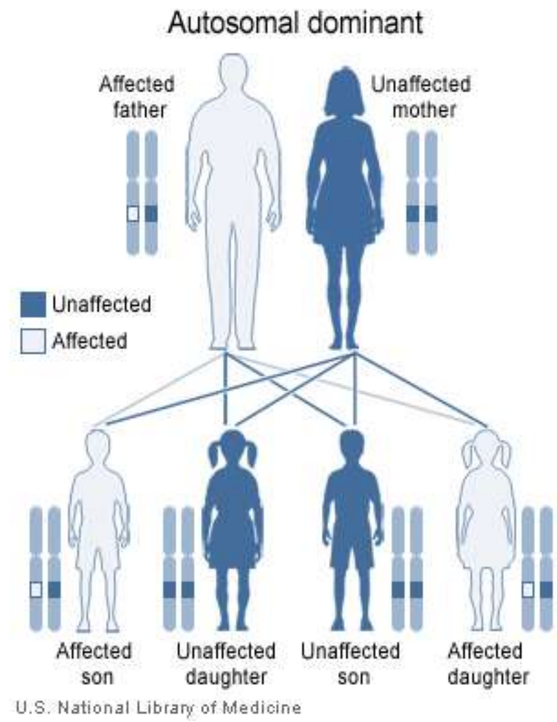
Inheritance: fertilization

- Sperm with 23 chromosomes joins together with an egg with 23 chromosomes
- Forms something called a "zygote" (this will eventually develop into an embryo)
- The zygote contains 46 chromosomes
- Chromosomes carry traits that we inherit from our parents (eg. Eye color)
- Chromosomes can also carry disease traits (genes)



Inheritance: (Autosomal) Dominant

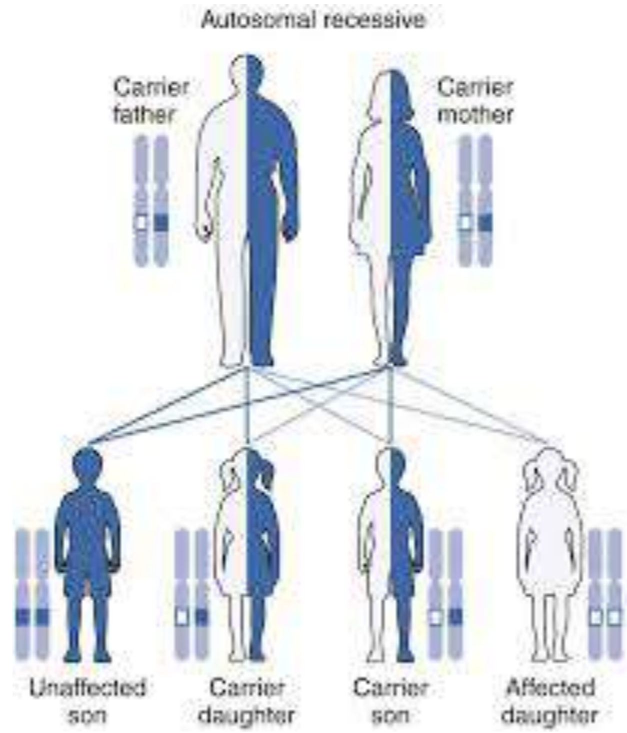
Autosomal Dominant:
- If you inherit the disease gene you will have the disease
- You only need one “bad” gene to have the disease
- **Example:** Huntington’s disease



Inheritance: (Autosomal) Recessive

Autosomal Recessive:

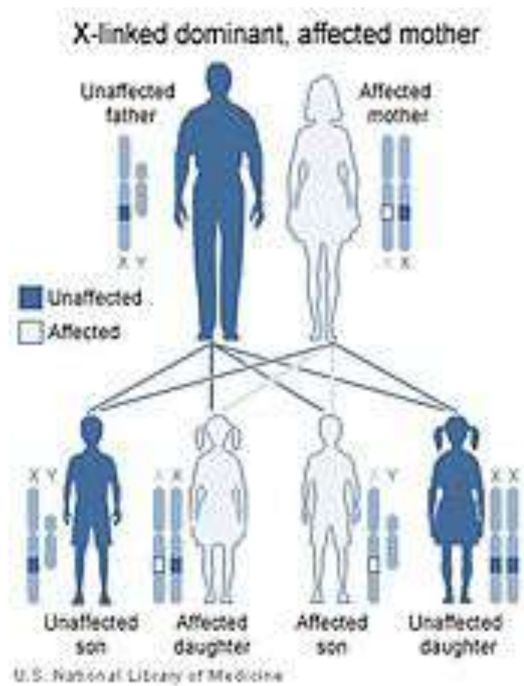
- If you inherit one disease gene + one normal gene you will not be affected
- Need 2 “bad” genes to get the disease
- Carrier’s are individuals who carry a copy of the disease gene but also have a copy of normal gene and are not affected but can pass it down
- Example: Gaucher’s disease



Inheritance: X-linked Dominant

X-linked Dominant inheritance

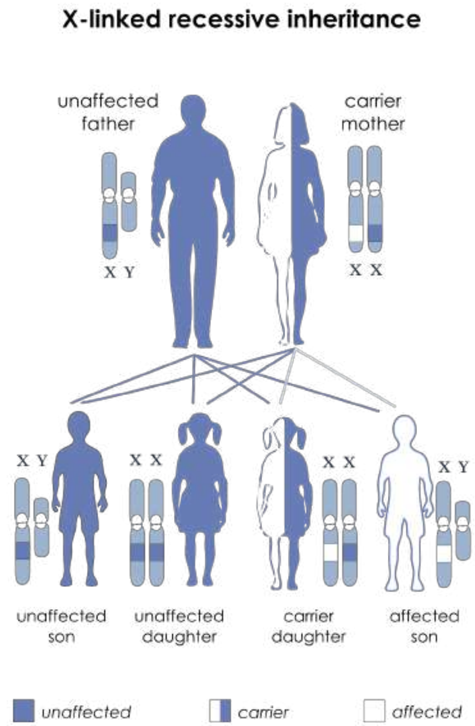
- X-linked just means talking about genes on the X chromosome
- If you inherit an X-chromosome with the disease gene you will have the disease
- Only need 1 “bad” X chromosome to get the disease
- Will affect boys and girls
- Example: fragile X syndrome



Inheritance: X-linked Recessive

X-linked Recessive inheritance

- X-linked just means talking about genes on the X chromosome
- If you are a man and you receive an X chromosome with a recessive X disease gene you will have the disease
- If you are female and you get one disease copy X and one normal X you will not have the disease
- Example: Hemophilia



What is human disease?

How and why do diseases happen?

Nature

Something is wrong at a genetic or DNA level

**Could be an issue in the DNA sequence or could be an issue with the DNA structural organization (epigenetic issue)

Ex. Hemophilia or SMA

We will mostly be dealing with these diseases in this class.

Nature and Nurture

+

Ex. Multiple Sclerosis (studies suggest there are both genetic and environmental components)

Nurture

Includes: diet, exercise, environment, drug use, Lifestyle, geography, etc.

Ex. Type 2 Diabetes or Coronary artery disease

Appendix

Does bigger gene always mean bigger protein?

No, not necessarily! Here is an example below.

Dystrophin

- **Gene size:** ~2.4 million base pairs long
- Contains 79 exons

- **Size of dystrophin protein:**
  - 3,684 amino acids long
  - Molecular weight ~427 kDa

Vs.

Titin

- **Gene size:** 80,781 base pairs long
- Contains 312 exons

- **Size of titin protein:**
  - 34,350 amino acids long
  - Molecular weight ~3816 kDa

Titin protein is ~9.3 times as big as dystrophin BUT dystrophin gene is ~7.9 times as big as the titin gene!

*kilodaltons (kDa) – unit of mass used to used to describe the molecular weight of proteins

Ok, bigger gene doesn't always mean bigger protein but why, how?

It has to do with the number of exons. Remember, exons actually contain the instructions for a protein to get made while the introns do not. Introns are cut out and only exist in the DNA version of the gene!

Dystrophin

- **Gene size:** ~2.4 million base pairs long
- **Contains 79 exons**
- **Size of dystrophin protein:**
  - 3,684 amino acids long
  - Molecular weight ~427 kDa

Vs.

Titin

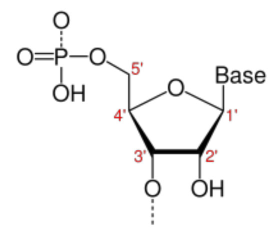
- **Gene size:** 80,781 base pairs long
- **Contains 312 exons**
- **Size of titin protein:**
  - 34,350 amino acids long
  - Molecular weight ~3816 kDa

Titin protein is ~9.3 times as big as dystrophin BUT dystrophin gene is ~7.9 times as big as the titin gene!

What do 5’ and 3’ mean?

In a nucleotide (unit of DNA/RNA)…

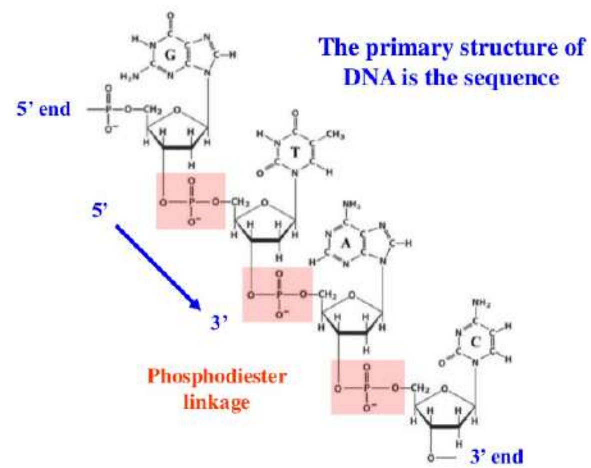
- The carbons in the sugar portion are numbered 1’ to 5’
- **5' carbon** has a phosphate group attached
- **3' carbon** a hydroxyl (-OH) group



Nucleotide with carbons of sugar component (ribose in RNA, deoxyribose in DNA) numbered

When nucleotides are linked together to form a DNA/RNA strand it goes in a “5’ to 3’” direction meaning:

- The 3’ group of the first nucleotide is linked to the 5’ group of the second nucleotide
- 3’ group of the second nucleotide is linked to the 5’ group of the third nucleotide (etc.)



When we label RNA and DNA strands we can label the 5’ and 3’ ends

When DNA strands base pair together they run “opposite” directions:

5’-ATGCTGCG-3’
3’-TACGACGC-5’