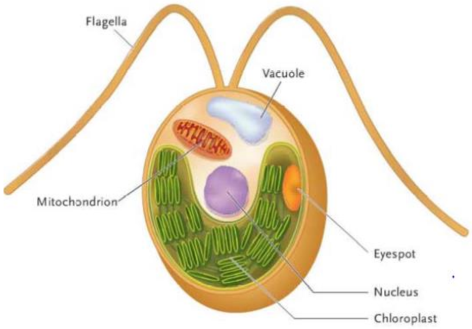
1.1 – The Physical Nature of Light (p. 2-4)

- 2 functions of light: source of energy that directly or indirectly sustains all organisms, AND it provides organisms with info about the physical world that surrounds them
- The green alga Chlamydomonas is an organism that uses light for both energy and information
    - Single-celled photosynthetic eurkaryote
    - Found in ponds and lakes
    - Each cell contains a single large chloroplast that harvests light energy and uses it to make nutrient-rich molecules through the photosynthesis
    - Each cell contains a light sensor called an "eyespot" that allows individual cells to gather information about the location and intensity of a light source

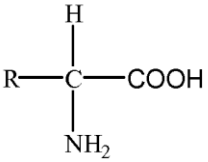


- Regardless of whether the light is used a source of energy or as a source of information, both uses require light energy to be captured by the organism
- <u>What is light?</u>
    - By converting hydrogen into helium at an immensely high rate, the sun converts over 4 million tonnes of matter into energy every second
        - This energy is given off as electromagnetic radiation, which travels in the form of a wave at the speed of light
        - Different types of ER are distinguished by their wavelength
    - Light is defined as the portion of the electromagnetic spectrum that humans can detect with their eyes (visible light → 400 to 700 nm)
    - Light both behaves as a wave that travels through space, AND a stream of energy particles called photons (particle-wave duality)
    - The longer the wavelength, the lower the energy of the photons it contains
- <u>Light interacts with matter</u>
    - Even though light has no mass, it can still interact with matter and cause change
        - This change is what allows the energy of light to be used by living things
    - When a photon of light hits an object, the photon has 3 possible fates:
        - It can be reflected off the object
        - It can be transmitted through the object
        - It can be absorbed by the object
    - Absorption must take place for the light to be used as a source of energy or information
    - Absorption of light occurs when the energy of the photon is transferred to an electron within a molecule
        - This excites the electron, moving it from its ground state to a higher-energy level that is referred to as an excited state
    - A photon can be absorbed by an electron of a molecule ONLY if the photon's energy equals the energy difference between the electron's ground state and an excited state
        - If the energies don't match, then the photon is transmitted through the molecule or reflected
        - It is the excited-state electron that represents the source of energy required for processes such as photosynthesis and vision
    - Pigments are a class of molecules that are very efficient at absorbing photons

- E.g. chlorophyll a, retinal, indigo
- Pigments are all very structurally different from each other, yet they all share a common feature critical to light absorption: a region where carbon atoms are covalently bonded to each other with alternating single and double bonds
  - This bonding arrangement is called a **conjucated system**, and it results in the delocalization of electrons
  - None of these electrons are closely associated with a particular atom or involved in bonding and are thus able to interact with a photon of light
- Pigments absorb light and distinctly different wavelengths
  - This is because they differ in the number of excited states available to the excitable electrons
  - While some pigments can absorb only blue photons because they have only one high-energy state, others can absorb two or more different wavelengths because they have two or more excited states
- Pigment absorption is intimately related to the concept of colour
  - A pigment’s colour is the result of photons of light that it does NOT absorb
  - Instead of being absorbed, these photons are reflected off the pigment or transmitted through the pigment to reach your eyes

Proteins (p. F-29-35)

- Proteins are polymers of amino acids and most diverse group of biological macromolecules
- Generalized structure of an amino acid has a central atom attached to an amino group, a carboxyl group and a hydrogen atom



$R-CH(NH_2)-COOH$

- The remaining bond of the central atom is to 1 of 20 different side groups (R group), which ranges from a single hydrogen atom to complex carbon chains and rings
  - Differences in side groups give amino acids their individual properties
  - 20 amino acids are categorized into different groups:
    - Non-polar
    - Uncharged polar
    - Negatively charged (acidic) polar
    - Positively charged (basic) polar
- Covalent bonds link amino acids into linear chains called **polypeptides**
  - The link between each pair of amino acids is called a peptide bond and is formed by a dehydration synthesis reaction between the $-NH_2$ group of one amino acid and the –COOH group of a second
  - In cells, amino acids are added only to the –COOH end of the growing peptide strand
  - A protein is a polypeptide that has folded into the specific 3-D shape that is required for most proteins to be functional
  - One end of the chain has an N-terminus, and the other end has a C-terminus

- <u>4 Levels of Protein Structure:</u>
    - **Primary structure** – particular & unique sequence of amino acids forming a polypeptide
        - Determined by the nucleotide sequence of the coding region of the protein's corresponding gene
    - **Secondary structure** – twists and turns of the amino acid chain
        - Based on hydrogen bonds between atoms of the backbone
            - Form between the hydrogen atom attached to the nitrogen of the backbone and the oxygen attached to one of the carbon atoms of the backbone
        - Alpha helix and beta sheet are two highly regular structures
            - A third, less regular arrangement is a random coil or loops
        - Alpha helix → coil shape formed when hydrogen bonds form between every N-H group of the backbone and the C=O group of the amino acid
        - Beta sheet → side-by-side alignment of beta strands, formed by hydrogen bonds between atoms of each strand
    - **Tertiary structure** – folding of the amino acid chain, with its secondary structures, into the overall 3-D shape of a protein
        - 4 major interactions between R groups that occur are: ionic bonds, hydrogen bonds, hydrophobic interactions and disulfide bridges
        - Protein undergoes conformational changes
    - **Quaternary structure** – optional structure that refers to the arrangement of polypeptide chains in a protein that is formed from more than one chain
- ***Cofactors/Prosthetic groups***
    - Non-protein chemical compounds that are bound to a protein and required for the protein to function
    - Can be either organic or inorganic molecules
    - Vitamins are essential to life because they act as cofactors
    - E.g. heme in hemoglobin is a cofactor as it is responsible for binding molecules of oxygen
- In many proteins, folding of the polypeptide chains produces distinct, large structural subdivisions called domains
    - One domain of a protein is often connected to another by a segment of random coil
    - The random coil is flexible and allows domains to move with respect to one another
    - Different domains of a protein are both structurally and functionally distinct

<u>Proteins (p. F-29-35)</u>

- We can detect energy only through its ability to do *work* – move objects against opposing forces, such as friction, gravity, or pressure, or to push chemical reactions toward completion
- **Energy** – the capacity to do work
- <u>*Energy exists in different forms and states*</u>
    - Different forms of energy includes heat, chemical, electrical and mechanical
    - These forms can be converted or transformed readily from one form to another
    - All forms are categorized into one of two different types:
        - KINETIC energy → energy possessed by an object because it is in motion

        - POTENTIAL energy → stored energy; the energy an object has because of its location or chemical structure
- *The laws of thermodynamics describe energy and its transformation*
    - **Thermodynamics** – the study of energy and its transformations
    - The object being studied is the *system* and everything outside the system is called the *surroundings* (*Universe* = total of the system and surroundings)
    - **Isolated system** – does not exchange matter or energy with its surroundings (e.g. thermos bottle)
    - **Closed system** – exchanges energy, but not matter, with its surroundings (e.g. greenhouse)
    - **Open system** – both energy and matter can move freely between the system and the surroundings (e.g. ocean)
- *The two laws of thermodynamics*
    - **First law** → energy can be transformed or transferred, but not created or destroyed
        - Also known as the principle of the conservation of energy
        - E.g. Niagara Falls
    - **Second law** → the total disorder of a system and its surroundings always increases
        - Each time energy is transformed from one form into another, some of the energy is lost and unavailable to do work
        - Therefore, nothing is 100% efficient
            - E.g. cells are only able to convert 40% of the potential energy in glucose into a form usable for metabolism
            - Heat cannot usually be harnessed to do work; instead, it is simply lost to the environment
        - The unusable energy that is produced during energy transformations results in an increase in the disorder of randomness of the universe
            - This randomness or disorder is a quantity called **entropy**
            - Systems will move spontaneously toward arrangements with greater disorder – greater entropy
            - It takes energy to maintain low entropy! (maintain order)
- *Life obeys the second law of thermodynamics*
    - Life displays order; the molecules and structures that define life are very precisely arranged in a non-random manner
    - Living cells have the ability to create ordered structures out of less ordered starting materials (e.g. a molecule of DNA, protein, ribosome are all ordered structures)
        - This suggests that life goes against the 2nd law; things become more ordered in a cell rather than random
        - How is this possible??? → it takes energy to maintain low entropy
    - Living cells are open systems, not isolated
        - They bring in energy and matter and use them to generate order out of disorder
    - According to the 2nd law, things are constantly breaking down
        - Cellular components become damaged and therefore need to be constantly repaired or replaced
        - New cells need to be made and old ones maintained by the synthesis of proteins, carbs and lipids
        - Therefore, we EAT to maintain low entropy (order within our cells)

        - Much of the food energy we ingest is used to maintain our cells in their highly ordered state
    - According to the 2nd law, isn't the entropy supposed to increase???
        - In the course of the many chemical reactions that take place to generate order, living things give off heat and byproducts of metabolism such as CO_2 that are much less ordered and increase the entropy of the surroundings
        - The entropy of a system is allowed to decrease as long as the entropy of the universe as a whole increases
            - This is why living organisms are thought of as islands of low entropy in a sea that is constantly becoming more disordered

5.1 – An Overview of the Structure of Membranes (p. 93-96)

- **Plasma membrane** – as a selectively permeable barrier, it allows for the uptake of key nutrients and elimination of waste products while maintaining a protected environment for cellular processes to occur
    - Subsequent development of internal membranes allowed for compartmentalization of processes and increased complexity
- *A membrane consists of proteins in a fluid of lipid molecules*
    - **Fluid mosaic model** – membranes are not rigid with molecules locked into place; instead, they consist of proteins within a mixture of lipid molecules the consistency of olive oil
    - Lipid molecules exist in a bilayer
    - Lipid molecules of bilayer vibrate, flex back and forth, spin around, move sideways, exchange places, etc. within the *same* bilayer half
        - Exchanging places occurs very fast, making the lipid molecules very dynamic
    - Membranes contain an assortment of different types of proteins
        - Includes proteins involved in transport and attachment, signal transduction, and processes such as electron transport
    - Proteins move more slowly since they are larger than lipid molecules
    - A small number of proteins anchor cytoskeleton filaments to the membrane and do not move
    - A number of lipid and protein components of some membranes have carbohydrate groups linked to them, forming glycolipids and glycoproteins
    - The proportions of lipid and protein within a membrane vary considerably depending on the type of membrane
        - Membranes that contain protein complexes involved in ET (e.g. inner mitochondrial membrane) contain large amounts of protein (76% protein, 24% lipid)
        - Plasma membrane contains nearly equal amounts of protein and lipid
    - Proteins and other components of one half of the lipid bilayer are different from those that make up the other half
        - This is known as **membrane asymmetry**, and it represents differences in the functions performed by each side of the membrane
        - E.g. glycolipids and carbohydrates are attached to proteins on external side of plasma membrane; cytoskeleton components bind to inner side

- *Experimental evidence in support of the fluid mosaic model*
    - **Membranes are fluid** – membrane bilayer appears to be about as fluid as olive oil
    - **Membrane symmetry** – particles on either side of the membrane differ in size, number and shape, providing evidence that the two sides are distinctly different
- *Phospholipids are the dominant lipids in membranes*
    - Lipids are a diverse group of water-insoluble molecules that include fats, phospholipids (which are the dominant lipids in membranes), and steroids
    - The lipid bilayer is formed of phospholipids
    - Each phospholipid consists of a head group attached to 2 long hydrocarbon chains called a fatty acid
        - Head group consists of glycerol linked to an alcohol or amino acid by a phosphate group
    - All phospholipids are amphipathic → contains both hydrophobic and hydrophilic regions
    - Fatty chains are non-polar, and the phosphate-containing head group is polar
    - Phospholipids can differ in the degree of unsaturation of their fatty acids
        - Presence of a C=C bond imparts a kink or bend to fatty acid tail
    - When added to an aqueous solution, phospholipids self-assemble into one of three structures:
        - A micelle, a liposome, or a bilayer
        - Which structure forms depends on the phospholipid concentration
        - Phospholipids spontaneously form these structures in an aqueous environment because of the **hydrophobic effect** → the tendency of polar molecules to exclude hydrophobic molecules
            - This results in the aggregation of lipid molecules in structures where fatty acid tails interact with each other and the polar head groups associate with water
            - These arrangements are favoured because they represent the lowest energy state and are more likely to occur over any other arrangement

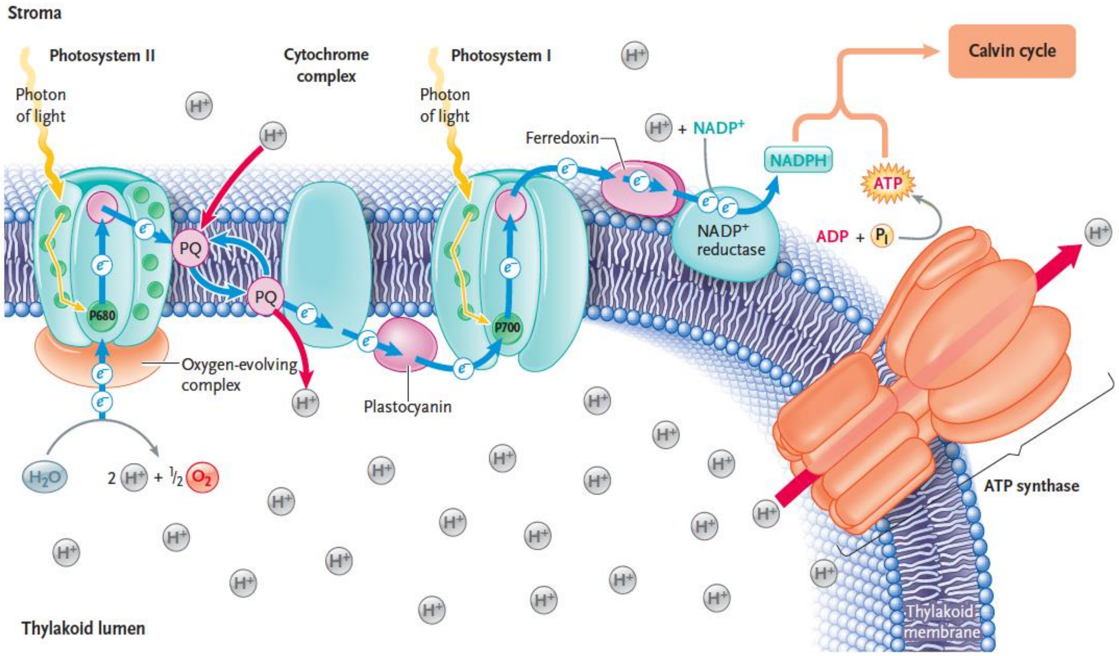
Lipids (F40-41)

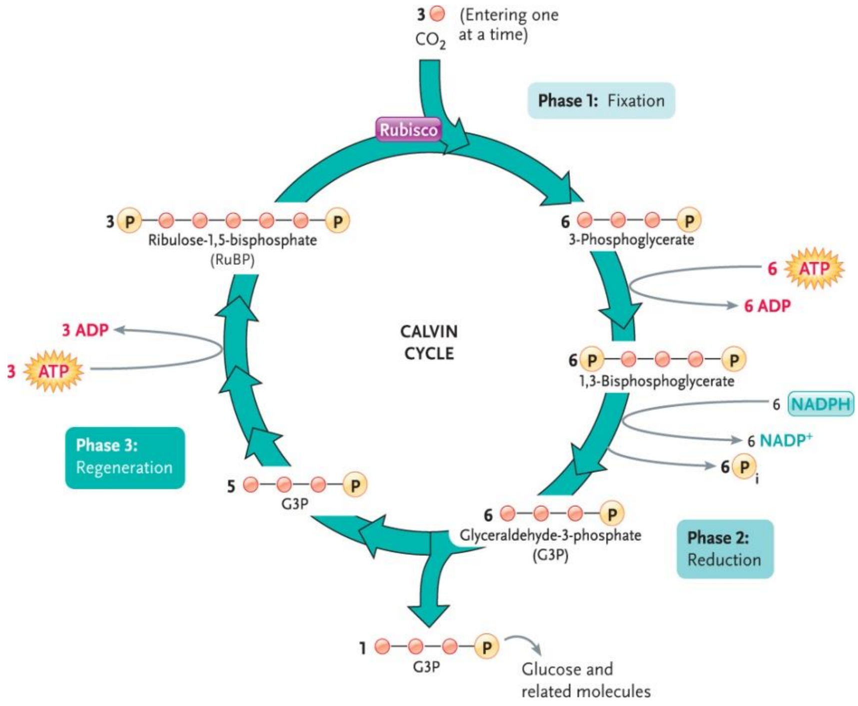
- Lipids are composed mostly of C and H
- Not large enough to be considered true macromolecules and unlike nucleic acids and proteins, are not considered polymers of defined monomeric subunits
- As a result of their non-polar character, lipids typically dissolve much more readily in non-polar solvents, than in water
    - Their insolubility in water underlies their ability to form cell membranes
- Some lipids are stored and used in cells as an energy source; other lipids serve as hormones that regulate cellular activities
- Lipids are grouped into: fats, phospholipids and steroids
- *Isoprenes and Fatty Acids*
    - The structural backbone of all lipids is derived from one of two hydrocarbon molecules: isoprene and fatty acids
    - Isoprenes → 5 C molecules that when linked together can form long hydrocarbon chains
        - The structural unit in steroids and a number of phospholipids
    - Fatty acid → single hydrocarbon chain with a carboxyl group (-COOH) linked at one end

        - Carboxyl group gives it its acidic properties
        - As chain length increases, fatty acids become progressively less water soluble and more solid
        - Saturated → maximum number of possible hydrogen atoms (single bonds)
        - Unsaturated → one or more double bonds
        - Monounsaturated → one double bond; Polyunsaturated → 2+ double bonds
- *Phospholipids*
    - Primary lipids of cell membranes
    - Glycerol forms the backbone for the molecule as in triglycerides , but only two of its binding sites are linked to fatty acids
        - The 3rd is linked to a polar phosphate group, which also binds to another polar unit
    - Thus, a phospholipid contains 2 hydrophobic fatty acids at one end attached to a hydrophilic polar group (head group)

Photosynthesis (Figures 7.11 & 7.14)

- **The Light Reactions:**
    1. Absorption of light energy by PSII results in oxidation of P680
        - The free electron is used to reduce the primary acceptor
        - $P680^+$ is rapidly reduced back to P680 by an electron from H_2O transferred from the oxygen evolving complex
    2. From the primary acceptor, the electron is passed to the mobile carrier molecule PQ
        - As it accepts an electron from PSII, it picks up a proton from the stroma
        - PQ diffuses through the membrane before binding to the cytochrome complex, at which point it donates an electron and releases a proton into the thylakoid lumen
        - From the cytochrome complex, the electron is donated to plastocyanin
    3. Absorption of light energy from PSI results in the oxidation of P700
        - The free electron is used to reduce the primary acceptor before being passed to ferredoxin
        - This single electron is then held by the $NADP^+$ reductase complex
        - $P700^+$ is reduced back to P700 by the electron that is coming from plastocyanin
        - Once a second electron travels along the chain and reaches $NADP^+$ reductase complex, $NADP^+$ is reduced to NADPH
    4. Proton pumping by PQ creates a concentration gradient of H^+ across the thylakoid membrane
        - The gradient is dissipated as H^+ diffuses back into the stroma through the ATPase complex, which drives the synthesis of ATP from ADP and Pi
- **The Calvin Cycle:**
    - 3 phases – fixation, reduction, regeneration
    - For every 3 molecules of CO_2 that are fixed, one molecule of a 3-C sugar G3P is synthesized





4.4C – Energy Coupling (p. 83-84)

- Enzymes combine with reactants and are released unchanged after reaction is complete
- The reactant that an enzyme acts on is called the enzyme's **substrate**
    - Enzyme is much larger than its substrate

- Each type of enzyme catalyzes the reaction of only a single type of molecule or a group of closely related molecules
- This enzyme specificity explains why the metabolism of a typical cell is represented by 4000 different enzymes
- Substrate interacts with only a small region of the enzyme → the active site (specific site on an enzyme where catalysis takes place)
- OLD → *Lock & key hypothesis* – explains specificity of the substrate-enzyme interaction
    - Somewhat similar substrates (keys) were able to bind to the same enzyme (lock) to cause catalysis (unlocking of door)
- RECENT → *Induced-fit hypothesis* – enzymes are not rigid objects (like locks); instead, they are flexible
    - Just before substrate binding, the enzyme changes its conformation so that the active site becomes even more precise in its ability to bind substrate
- Enzyme binds to the substrate, forming an enzyme-substrate complex
    - Catalysis occurs when the two are joined, with the action of the enzyme converting the substrate(s) into one or more products
- Since enzymes are released unchanged after reaction, enzyme molecules can rapidly bind to other substrate molecules, catalyzing the same reaction again, repeating the enzyme cycle
- The rates at which enzymes catalyze reactions varies widely depending on the specifics of the enzyme and substrates involved
- Many enzymes require a **co-factor** – a non-protein group that binds very precisely to the enzyme
    - These are often metals (e.g. iron, copper, zinc) that are required in very small amounts for the catalytic activity of the enzyme
    - Organic cofactors called **co-enzymes** play similar roles and are often derived from vitamins

6.6D – Cellular Respiration is Controlled by Supply & Demand (p. 131-2)

- Respiration includes many enzymes and transport systems, and in eukaryotic cells, many cellular compartments
- The overall state of respiration, which is often measured as the rate of oxygen consumption, is tightly controlled so that ATP synthesis matches the requirements of the cell for chemical energy
- Supply & demand → the cell does not waste valuable resources making more of a substance than it needs
    - Most metabolic pathways are regulated by supply and demand through the process of **feedback inhibition** – the end products of the pathway inhibit an enzyme early in the pathway
- Rate of glucose oxidation by glycolysis is closely regulated by mechanisms to match cellular demands for ATP
    - Early in glycolysis, phosphofructokinase is a key enzyme that is a major site of regulation
        - Catalyses the conversion of fructose 6-phosphate to fructose 1,6-biphosphate
        - Allosteric enzyme, and thus activity can be adjusted by the binding of certain metabolic activators and inhibitors
- Two of the key regulators of phosphofructokinase are ATP and AMP

- ATP, ADP and AMP are interconvertible
- Therefore, when ATP levels in the cell are low, ADP and AMP levels are higher and vice versa
- ATP is an allosteric inhibitor of phosphofructokinase, so if excess ATP is present in the cytosol, it binds to phosphofructokinase, inhibiting its action
    - The resulting decrease in the concentration of fructose 1,6-biphosphate slows or stops the subsequent reactions of glycolysis and, as a consequence, the remainder of respiration as well
    - The enzyme becomes again active again when metabolic demands consume the excess ATP and the inhibition of phosphofructokinase is released
    - The increase in phosphofructokinase activity is not due solely to the release of ATP inhibition
    - AMP, which accumulates when ATP is being consumed for metabolism, is an allosteric activator of the enzyme
    - Besides ATP and AMP, phosphofructokinase activity is also sensitive to the levels of citrate (first product of citric acid cycle)
        - If the products of the citric acid cycle are in high demand, then citrate shouldn't accumulate in the cell
        - Increased citrate concentrations suggest that the demand for ATP is low, which may occur under conditions of limited oxygen when rate of oxidative phosphorylation is restricted
            - OR it may indicate that citrate is not required as a carbon skeleton for anabolic reactions
    - Overall, through various metabolic activators and inhibitors altering phosphofructokinase activity, the functional state of glycolysis and the citric acid cycle can be kept balanced

15.1 – DNA Cloning (p. 333-341)

- **Metagenomics** – DNA from an entire community of organisms in a particular niche is harvested collectively, sequenced and analyzed using DNA technologies
    - Before, we were limited to the very small proportion of species that could be cultivated in the laboratory
    - With metagenomics, we gain access to the genomes of a whole new world of previously inaccessible organisms
    - It revealed the vast majority of microbial biodiversity that had been missed by cultivation-based methods
- **DNA technologies** – techniques used to isolate, purify, analyze, and manipulate DNA sequences
    - Used for both basic and applied research
- **Genetic engineering** – use of DNA technologies to alter genes for practical purposes
    - This is the latest addition to **biotechnology** – any technique applied to biological systems or living organisms to make or modify products or processes for a specific purpose
    - Biology can include manipulations that don't involve DNA technologies
- **Clone** – line of genetically identical cells or individuals derived from a single ancestor
- **Cloning** – method of producing many copies of a piece of DNA
    - This piece of DNA is referred to as a *gene of interest*, which is a gene that is to be studied or manipulated
    - A gene is difficult to study in its natural state, which is why it is usually cloned, so the researcher can produce a sample large enough for scientific experimentation
    - Cloned genes are used in basic research to find out about their biological functions
        - E.g. determine the DNA sequence of a cloned gene; manipulate the gene and induce mutations in it to show information about how its expression is regulated
    - Proteins encoded by cloned genes can be produced in quantity and purified
- One method of cloning → uses bacteria and plasmids (small circular DNA that replicates separately from bacterial chromosome)
    - DNA is extracted from cells containing gene of interest, then cut into fragments
    - Each fragment is inserted into a plasmid, thus producing a *recombinant DNA molecule*
    - Recombinant plasmids are then introduced into bacteria; each bacterium receives a different plasmid
    - Bacterium continues to grow and divide, and as it does, the recombinant plasmid DNA is also replicated
    - Final step is to identify which bacterium contains the plasmid carrying the gene of interest and isolate it for further study
- ***Bacterial enzymes called restriction endonucleases form the basis of DNA cloning***
    - The key to DNA cloning is the specific joining of two DNA molecules from different sources, which is made possible by bacterial enzymes called **restriction endonucleases**
    - These enzymes recognize sort, specific DNA sequences called *restriction sites* (4-8 bp), and cut the DNA at specific locations within those sequences
    - The DNA fragments produced by cutting a long DNA molecule with a restriction enzyme are known as **restriction fragments**
    - The enzymes defend against viral attack by breaking down (restricting) the DNA molecules of infecting viruses

        - Why don’t they break down the cell’s own DNA?
        - The bacterium “hides” the restriction sites in its own DNA by methylating bases in those sites, which blocks the action of its restriction enzyme
    - Each restriction enzyme cuts DNA at a specific restriction site
    - Most restriction sites are symmetrical in that the sequence of nucleotides read in 3’ → 5’ direction on one strand is the same as the sequence read in the 3’ → 5’ direction on the complementary strand
    - A given enzyme always recognizes the same DNA sequence as its cut site and always cuts in the same place within the sequence
    - Restriction enzymes most used in cloning cleave the sugar-phosphate backbones of DNA to produce DNA fragments with single-stranded ends
        - The ends are called **sticky ends** because the short-single stranded regions can form hydrogen bonds with complementary sticky ends on any other DNA molecules cut with the same enzyme
    - The pairing leave nicks in the sugar-phosphate backbones of DNA strands which are sealed by *DNA ligase*
- ***Bacterial plasmids illustrate the use of restriction enzymes in cloning***
    - Bacterial plasmids are examples of *cloning vectors* – DNA molecules into which a DNA fragment can be inserted to form a recombinant DNA molecule for cloning
    - Plasmid cloning vectors are engineered to contain 2 genes that are useful in the final steps of a cloning experiment for distinguishing bacteria that have recombinant plasmids from those that do not
    - Amp^R gene encodes an enzyme that breaks down the antibiotic ampicillin
        - When the plasmid is introduced into Ecoli and the ampR gene is expressed, the bacteria become resistant to ampicillin
    - $LacZ^+$ gene encodes beta-galactosidase, which hydrolyzes the sugar lactose
        - Restriction sites are located within the lacZ+ genes, but do not alter the gene’s function
    - Cloning a gene of interest:
        - Genomic DNA isolated from the organism in which the gene is found is cut with a restriction enzyme, and then, using the same enzyme, a plasmid cloning vector is cut within the lacZ+ gene
        - Mixing the DNA fragments and cut plasmid together with DNA ligase produces various joined molecules as the sticky ends pair and the enzyme seals them together
            - Some are recombinant plasmids → consist of a DNA fragment inserted into plasmid cloning vector
            - Others are non-recombinant plasmids resulting from the cut plasmid being resealed (by DNA ligase) into a circle without an inserted fragment
            - Only recombinant plasmids are important in cloning
        - Next, the DNA molecules are introduced into ampicillin-sensitive $lacZ^-$ E coli (which cannot make B-galactosidase), and these transformed bacteria are spread on a plate of agar growth medium containig ampicillin and C-gaactosidase substrate
        - Only bacteria with a plasmid can grow and form colonies because expression of the plasmid’s amp^R gene is needed to make the bacteria resistant to ampicillin

        - Within each cell of a colony, the plasmids replicate until there are 100 of them
        - X-gal in the medium distinguishes between bacteria that have been transformed with recombinant plasmids vs. Non-recombinant plasmids by *blue-white screening*
            - If a colony produces B-galactosidase, it converts X-gal to a blue product and the colony turns blue
            - If a colony doesn't produce the enzyme, X-gal is unchanged and the colony remains white
        - Colonies containing non-recombinant plasmids have an intact lacZ+ gene, produce the enzyme and turn blue
        - Colonies containing recombinant plasmids are white because they contain a DNA fragment inserted into the lacZ+ gene, so they do NOT produce a functional enzyme
        - White colonies are then further examined to find the one containing a recombinant plasmid with the gene of interest
    - <u>Identifying the clone containing the gene of interest:</u>
        - **DNA hybridization** – gene of interest is identified in the set of clones when it base-pairs with a short, single-stranded complementary DNA or RNA molecule called *nucleic acid probe*
        - Probe is labelled with a radioactive or non-radioactive tag so investigators can detect it
        - Once a colony containing plasmids with the gene of interest has been identified, that colony can be used to produce large quantities of the cloned gene
- ***<u>DNA libraries contain collections of cloned DNA fragments</u>***
    - **Genomic library** – collection of clones that contains a copy of every DNA sequence in a genome
        - Can be made using any kind of cloning vector
    - Number of clones in a library increases with the size of the genome
    - cDNA library:
        - Convert single-stranded mRNA to double-stranded DNA for cloning by using *reverse transcriptase* (made by retroviruses) to make a single-stranded DNA that is complementary to the mRNA
        - Then, the mRNA strand is degraded with an enzyme and DNA polymerase is used to make a second DNA strand that is complementary to the first
            - Results in complementary DNA (cDNA)
        - After adding restriction enzymes to each end, the cDNA is inserted into a cloning vector for the cDNA library
    - Not all genes are active in every cell
    - cDNA library is limited in that it includes copies of only the genes that were active in the cells used as the starting point for the creation of the library
        - This limitation is an advantage in identifying genes active in one cell type and not another
    - cDNA libraries are useful for providing clues to the changes in gene activity that are responsible for cell differentiation and specialization
        - Advantage for those who want to insert eukaryotic genes into bacteria, particularly when bacteria are to be used as "factories" for making the protein encoded in the gene

        - Genes in eukaryotic nuclear DNA contain many introns
        - Because bacterial DNA doesn't contain introns, bacteria aren't equipped to process eukaryotic genes correctly
        - However, the cDNA copy of a eukaryotic mRNA already has the introns removed, so bacteria can transcribe and translate it accurately to make eukaryotic proteins
- ***<u>The polymerase chain reaction amplifies DNA in vitro</u>***
    - **PCR** – produces a large number of copies of specific DNA sequence from a DNA mixture without having to clone the sequence in a host organism
    - *Amplication* process because it increases the amount of DNA to the point where it can be analyzed or manipulated easily
    - DNA polymerase replicates just a portion of a DNA molecule rather than whole
    - Enzymes add nucleotides only the 3' end of an existing chain (primer)
    - For replication to begin, a primer must be available, base-paired to a template chain
    - After 20-30 cycles of priming and replication, PCR amplifies the target sequence, producing millions of copies
    - Since the primers used in PCR are designed to bracket only the sequence of interest, the cycles replicate only this sequence from a mixture of essentially any DNA molecules
    - Notice that:
        - Primers are made of DNA, not RNA
        - Left primer binds to one strand while the right primer binds to the opposite strand of the original DNA
        - Of all the DNA sequences put into the PCR reaction tube, only the target sequence (the sequence between the primers) is amplified
        - DNA polymerase reads the template in a 3' → 5' direction
    - PCR allows extremely small samples to be amplified to concentrations high enough for analysis (e.g. analysis of hair, blood, etc.)
    - A successful outcome of PCR is shown by analyzing a sample of the amplified DNA using agarose gel electrophoresisto see if the copies are the same length as the target
    - **Gel electrophoresis** – technique by which DNA, RNA or protein molecules are separated in a gel subjected to a electric field
        - The gel functions as a molecular sieve to separate the macromolecules based on size, electrical charge, etc.
        - Size of amplified DNA is determined by comparing the position of the DNA band with the positions of DNA fragments of known size separated on the gel at the same time
            - If that size matches the predicted size for the target DNA, PCR was successful
    - Limitation of PCR is the primers
        - To design a primer for PCR, you must have sequence information about the target DNA
        - By contrast, cloning can be used to amplify DNA of UNKNOWN sequence

<u>13.2 – Transcription: DNA-Directed RNA Synthesis (289-90)</u>

- **Transcription** – process by which information coded in sequential DNA bases is transferred to a complementary RNA strand
    - For a given gene, only one of the two DNA nucleotide strands acts as a template for the synthesis of a complementary copy, instead of both, as in replication
    - Only a relatively small part of a DNA molecule – the sequence encoding a single gene – serves as a template, rather than all of both strands, as in replication
    - RNA polymerases catalyze the assembly of nucleotides into an RNA strand, rather than the DNA polymerases that catalyze replication
    - The RNA molecules resulting from transcription are single polynucleotide chains, not double ones as in replication
    - Uracil, instead of thymine is matched with adenine in the chain
- ***<u>Transcription proceeds in 3 steps</u>***
    - The gene consists of 2 parts: a **promoter** (control sequence for transcription), and a **transcription unit** (section of the gene that is copied into an RNA molecule)
    - INITIATION → molecular machinery that carries out transcription assembles at the promoter and begins synthesizing an RNA copy of the gene
    - ELONGATION → RNA polymerase moves along the gene extending the RNA chain
    - TERMINATION → transcription ends and the RNA molecule and RNA polymerase are released from the DNA template
- ***<u>Similarities and differences in transcription of eukaryotic and bacteria PCGs</u>***
    - Gene organization is the same, although the specific sequences in the promoter where the transcription apparatus assembles differ
    - Eukaryotes → RNA polymerase II, the enzyme that trascribes PCGs, cannot bind directly to DNA; it is recruited to the promoter once proteins called **transcription factors** have bound
    - Bacteria → RNA polymerase binds directly to DNA; it is directed to the promoter by a protein factor that is then released once transcription stops
    - Elongation is identical in both organisms
    - Prokaryotic cells → 2 types of specific DNA sequences called **terminators** that trigger the end of transcription of the gene and release of the RNA and RNA polymerase from the template
        - Both types of sequences act *after they are transcribed*
        - 1st case: terminator sequence on mRNA uses complementary base pairing with itself to form a "hairpin"
        - 2nd case: protein binds to a particular terminator sequence on the mRNA
    - Eukaryotes → have no equivalent transcription terminator sequences
        - Instead, the 3' end of mRNA is specified by a different process
    - Once an RNA polymerase has started transcription, another molecule of RNA polymerase may start transcribing as soon as there is room at the promoter
        - This process continues until there are many RNA polymerase molecules spaced closely along a gene, each making an RNA transcript
- ***<u>Transcription of non-protein-coding genes occur in a similar way</u>***
    - Non- PCGs include those for tRNAs and rRNAs
    - In eukaryotes, RNA polymerase II transcribes PCGs, RNA polymerase III transcribes tRNA genes and the gene for one of the 4 rRNAs, and RNA polymerase I transcribes the genes for the 3 other rRNAs

    - The promoters for these non- PCGs are different from those of PCGs, being specialized for the assembly of the transcription machinery that involves the correct RNA polymerase type
  - In bacteria, a single type of RNA polymerase transcribes all types of genes
    - Promoters for bacteria non- PCGs are the same as those of PCGs

13.4 – Translation: mRNA-Directed Polypeptide Synthesis (294-305)

- **Translation** – Assembly of amino acids into polypeptides on ribosomes
- Prokaryotes → translation takes place throughout the cell
  - mRNA after translation is available immediately for translation since it is not confined in a nucleus
- Eukaryotes → translation occurs in the cytoplasm (& few genes in the chloroplast/mitochondria)
  - mRNA produced by splicing of the pre-mRNA first exits the nucleus and is then translated in the cytoplasm
    - Either free in cytoplasm OR attached to membranes of ER
- tRNA brings amino acids to the complex to the ribosome complex to be joined, one by one, into the polypeptide chain
  - Sequence of amino acids in polypeptide chain is determined by the sequence of codons in mRNA
  - mRNA is read from 5' to 3' end (ribosome moves codon by codon in 5' → 3')
  - Polypeptide is assembled from the N- to C-terminal end
    - First amino acid in chain is Met
- Overview of translation:
  - A tRNA molecule with an amino acid bound to it enters ribosome on the right
  - Anticodon on tRNA pairs with codon in mRNA
  - Its amino acid is then added to the growing polypeptide currently attached to the tRNA in the middle of the ribosome
  - As it assembles a polypeptide chain, the ribosome moves from one codon to the next along the mRNA in 5' → 3' direction
- ***<u>tRNAs are small RNAs of a highly distinctive structure that bring amino acids to the ribosome</u>***
  - Unlike mRNAs which are 100s of nucleotides long, tRNAs are about 75-90
  - They have a highly distinct structure
  - Can base-pair with themselves to wind into 4 double-helical segments
    - At one of these segments is the **anti-codon**, the 3-nucleotide segment that base-pairs with a codon in mRNAs
    - Opposite the anti-codon end is a free 3' end of the molecule that links to the amino acid corressponding to the anti-codon
  - Anti-codon and codon pair in an anti-parallel manner
    - Anti-codons are 3' to 5' and codons are 5' to 3'
  - 61 out of the 64 codons of the genetic code specify an amino acid
    - However, there aren't 61 different tRNAs
    - **Wobble hypothesis** – pairing of the anti-codon with the first 2 nucleotides is always precise, but the anti-codon has more flexibility in pairing with the 3rd nucleotide of the codon
  - The correct amino acid must be present on the tRNA for translation to be correct

  - **Aminoacylation** (charging) is the addition of an amino acid to tRNA
    - The finished product is an **aminoacyl-tRNA**
  - **Aminoacyl-tRNA synthases** are the enzymes that catalyze aminoacylation
  - Energy in aminoacyl-tRNA drives the formation of the peptide bond linking amino acids during translation
- ***<u>Ribosomes are rRNA-protein complexes that work as automated protein assembly machines</u>***
  - Made up of *large* and *small ribosomal subunits*
  - Each subunit is made up of a combination of rRNA and ribosomal proteins
  - Mitochondria and chloroplasts still code for their own prokaryotic ribosomes that are distinct from those in the cytoplasm
  - Prokaryotic and eukaryotic ribosomes are similar in structure and function
  - **A Site** → where the incoming aminoacyl-tRNA binds to mRNA
  - **P Site** → where the tRNA carrying the growing polypeptide chain is bound
  - **E Site** → where an exiting tRNA binds as it leaves the ribosome
- ***<u>Translational initiation brings the ribosomal subunits, an mRNA, and the first aminoacyl-tRNA together</u>***
  - The small ribosomal subunit, the initiator Met-tRNA and GTP bind directly to the region of the mRNA with the AUG start codon
  - This initiation complex is then guided by the **ribosome binding site** – a short specific RNA sequence – just upstream of the start codon on the mRNA that base-pairs with a complementary sequence of rRNA in the small ribosomal subunit
  - Large ribosomal subunit binds to the small ribosomal subunit; GTP hydrolysis then begins translation
  - After the initiator tRNA pairs with the AUG start codon, the subsequent stages of translation simply read the codons one at a time on the mRNA
    - This establishes the correct **reading frame** – series of codons for the polypeptide encoded by the mRNA
- ***<u>Polypeptide chains grow during the elongation stage of translation</u>***
  - Amino acids are added one at a time to a growing polypeptide chain
  - The P site can bind only to a **peptidyl-tRNA** – a tRNA linked to a growing polypeptide chain containing two or more amino acids
    - Exception is the initiator tRNA
  - The A site can bind only to an aminoacyl-tRNA
    - Binding is facilitated by a protein **elongation factor (EF)** that is bound to the aminoacyl-tRNA and that is released once the tRNA binds to the codon
    - Another ER is used when the ribosome translocates along the mRNA to the next codon
    - Each EF is released after its job is completed
    - GTP hydrolysis powers the ribosome along the mRNA
  - A peptide bond is formed between the C-terminal end of the growing polypeptide on the P site tRNA and the amino acid on the A site tRNA
    - Peptidyl transferase catalyzes this reaction
  - The tRNA that was previously in the P site is shifted to the E site and then leaves the ribosome
  - Elongation is faster in prokaryotes than eukaryotes
- ***<u>Termination releases a completed polypeptide from the ribosome</u>***
  - Similar in prokaryotic and eukaryotic cells

    - Takes place when one of the stop codons (UAG, UAA, or UGA) on the mRNA arrives at the A site of a ribosome
    - No tRNA can pair with the stop codon
    - A protein **release factor (RF)** binds in the A site and causes the ribosome to disassemble into its subunits, and the polypeptide to be released
        - RF has a similar shape as tRNA
- ***Multiple ribosomes simultaneously translate a single mRNA***
    - Once the first ribosome begins translation, another one can assemble with an initiator tRNA as soon as there is room at the 5'UTR of the mRNA
    - **Polysome** – multiple ribosomes attached to one mRNA
        - This increases the overall rate of polypeptide synthesis from a single mRNA
    - In prokaryotic cells, transcription and translation are coupled
        - As soon as the 5' end of a new mRNA emerges from the RNA polymerase, ribosomal subunits attach and initate translation
        - By the time the mRNA is fully transcribed, it is covered with ribosomes from end to end, each undergoing translation
        - Meanwhile, other RNA polymerases have begun transcribing the same gene
- ***Newly synthesized polypeptides are processed and folded into finished form***
    - Processing reactions convert the new proteins released from ribosomes into their finished form
    - Amino acids are removed from the ends or interior of the polypeptide chain
    - Larger organic groups (carbs, lipids) are added
    - Proteins fold into 3-D shape
    - Chaperones may assist in folding process
    - Alternative processing → increases the # of proteins encoded by a single gene
    - Other proteins are processed into an initial, inactive form that is later activated ata particular time or location by removal of a covering segment of the polypeptide chain
- ***Finished proteins are sorted to the cellular locations where they function***
    - Without a sorting and delivery system, all proteins would be flowing freely in the cytoplasm... we don't want this
    - 3 types of final destination compartments where the final proteins are needed:
        - The cytosol
        - The endomembrane system (ER, golgi, lysosomes, secretory vesicles, nuclear envelope, plasma membrane)
        - Other membrane-bound organelles distinct from the endomembrane system (nucleus, mitochondria, chloroplasts, microbodies)
    - Protein sorting is similar in both eukaryotic and prokaryotic cells
    - **Protein sorting to the cytoplasm:**
        - Simply released from ribosomes once translation is completed
    - **Protein sorting to the endomembrane system:**
        - Begin synthesis on free ribosomes in cytosol and produce a short segment of amino acids called a signal sequence near their N-terminal ends
        - Signal sequence is recognized by a signal recognition particle that initiates a series of steps that ultimately result in the polypeptide entering the ER interior
            - This is called cotranslational import because import of the polypeptide into the ER occurs simultaneously with translation of the mRNA encoding the polypeptide

        - Once inside the ER, proteins fold into their final form
        - They obtain a 'tag' that targets each protein for sorting into its final destination
        - Some proteins are transported to the Golgi, where they are modified further and packaged into vesicles, which may deliver them to lysosomes, or secrete them from the cell, or deposit them in plasma membrane
    - **Protein sorting to other membrane-bound organelles:**
        - These proteins are imported after they have been made on free ribosomes in the cytosol
        - Mechanism for sorting is called post-translational import
        - These proteins have short amino acid sequences called transit sequences at their N-terminal ends that take them to the appropriate organelle
        - Transit peptidase enzyme within the organelle then removes the transit sequence
        - Proteins sorted to the nucleus have short amino acid sequences called nuclear localization signals
            - This signal is never removed because they need to reenter the nucleus each time the nuclear envelope breaks down and reforms during cell cycle
- ***<u>Mutations can affect protein structure and function</u>***
    - Mutations are changes in the sequence of bases in the genetic material
    - **Base-pair substitution mutations** – the change of one particular base to another
        - Causes a change in a base in a codon in mRNA
        - 4 different types
    - <u>Missense mutation</u> → alters the codon to specify a different amino acid, resulting in the protein to have a different amino acid sequence
        - Can be functional or non-functional, depending on which amino acid is changed
    - <u>Non-sense mutation</u> → changes a sense codon to a nonsense (termination) codon in the mRNA, which results in a premature 'stop" and a shorter-than-normal polypeptide
        - Partially functional
    - <u>Silent mutation</u> → does not alter the amino acid specified by the gene because the changed codon specifies the same amino acid as in the normal polypeptide
        - Still functional
    - <u>Frameshift mutation</u> → when the insertion or deletion of a single base pair alters the reading frame of the resulting mRNA, which produces a complely different amino acid sequence in the polypeptide from then on
        - Usually non-functional

<u>14.1 – Regulation of Gene Expression in Prokaryotic Cells (309-313)</u>

- Rather than the complex patterns of long-term cell differentiation and development typical of multicellular eukaryotes, prokaryotic cells typically undergo rapid and reversible alterations in biochemical pathways that allow them to adapt quickly to changes in their environment
- ***<u>The operon is a unit of transcription</u>***
    - For a typical metabolic process, several genes are involved, and they must be regulated in a coordinated fashion

    - **Operon** – cluster of prokaryotic genes and the DNA sequences involved in their regulation
        - Promoter → region where RNA polymerase begins transcription
        - Operator → a short segment that is a binding sequence for a regulatory protein
    - A gene that is separate from the operon encodes the regulatory protein
    - **Repressor** – a regulatory protein that when bound to the DNA, it reduces the likelihood that genes will be transcribed
    - **Activator** – a regulatory protein that when bound to the DNA, increases the likelihood that genes will be transcribed
    - Many operons are controlled by more than one regulatory mechanism, and a number of the activators/repressors may control more than one operon
    - Each operon, which can contain several genes, is transcribed as a unit from the promoter into a single mRNA, and as a result, the mRNA contains codes for several proteins
    - The cluster of genes transcribed into a single mRNA is called a **transcription unit**
    - A ribosome translates the entire mRNA from one end to the other, sequentially making each protein encoded in the mRNA
    - Typically, the proteins encoded by genes in the same operon catalyze steps in the same process, such as enzymes acting in sequence in a biochemical pathway
- ***The lac operon for lactose metabolism is transcribed when an inducer inactivates a repressor***
    - 3 genes encode proteins for the metabolism of lactose by E coli
        - *lacZ, lacY,* and *lacA*
        - Absence of lactose: 3 genes are transcribed very little
        - Presence of lactose: 3 genes are transcribed actively
    - The genes are transcribed as a unit into a single mRNA starting with the *lacZ* gene; the promoter for the transcription unit is upstream of *lacZ*
    - *lacZ* – encodes the enzyme B-galactosidase, which catalyzes the conversion of the disaccharide sugar, lactose, into the monosaccharide sugars, glucose & galactose
    - *lacY* – encodes a permease enzyme that transports lactose actively into the cell
    - *lacA* – encodes a transacetylase enzyme, which metabolizes compounds other than lactose
    - For the lac operon, the operator is between the promoter and *lacZ* gene
    - Lac operson was controlled by the regulatory protein, *lac repressor*, which is encoded by the regulatory gene, *lacI*
    - When lactose is ABSENT → lac repressor binds to the operator, blocking RNA polymerase from binding to the promoter
        - Sometimes the repressor occasionally comes off, allowing polymerase to successfully transcribe
        - As a result, there is always a low concentration of lac operon gene products in the cell
    - When lactose is PRESENT → lac operon is turned on and all 3 enzymes are synthesized rapidly
        - As lactose enters, B-galactosidase converts some of it to allolactose, an isomer of lactose which is an **inducer** for the lac operson
        - The inducer binds to the lac repressor, altering its shape so it can no longer bind to operator DNA
        - With the repressor out of the way, RNA polymerase can do its job

  - Since an inducer molecule increases its expression, the lac operson is called an **inducible operon**
  - As the lactose is used up, the lac operson is switched off
    - Absence of lactose means there are no inducer molecules to inactivate the repressor, so it binds to the operator, reducing transcription
- ***Transcription of lac operon is also controlled by a positive regulatory system***
  - Makes the expression of the lac operon responsive to the availability of glucose
  - Lac operon is sensitive to the availability of glucose through the binding of an activator protein called CAP
  - CAP binding site is on the DNA, just upstream of the lac promoter
    - When bound at this site, CAP bends the DNA in ways that makes the promoter more accessible to RNA polymerase and transcription increases
  - NOTE:
    - (I) CAP is synthesized in an inactive form that can only bind to DNA after it is activated by binding with cAMP
    - (II) cAMP levels are inversely related to the uptake of glucose from the growth medium
      - When glucose is abundant, cAMP levels tend to be low (meaning CAP is mostly inactive)
      - When glucose is absent from the environment, cAMP concentrations are high inside the cell, leading to an increased level of activated CAP
  - Together, the negative control by the lac repressor and positive control by CAP/cAMP ensure that cells express the lac operon MOST strongly when lactose is present and glucose is not
  - What if cells are growing on glucose only? → Very little cAMP is available to bind to CAP, so CAP/cAMP binding will be rare and there will be very little stimulation of expression
  - Absence of lactose → lac repressor will be bound to operator most of the time, and very little synthesis of lac genes will occur (lac operon expression will be at LOWEST level)
  - When both glucose and lactose are present, level of transcription is higher than when lactose is absent, but far lower than when lactose is present and glucose is absent
  - lac repressor and CAP are DNA-binding proteins
  - *Whether gene expression is under negative or positive control depends on the impact of the respective DNA-binding proteins, not on the impact of the available substrates such as glucose or lactose!!!*
    - If the binding of a protein to DNA results in decreased gene expression → **negative control**
    - If the binding of a protein to DNA results in increased gene expression → **positive control**

13.1c – The Genetic Code (p. 286-288)

- The genetic code is written in 3 letter words using a 4-letter alphabet
- DNA alphabet consists of 4 letters A, T, G, C, representing 4 nucleotide bases
- RNA alphabet consists of 4 letters A, U, G, C
- Both DNA and RNA share 3 of the 4 bases but differ in the other one: T is equivalent to U in RNA

- There are 20 amino acids... so how is nucleotide information in an mRNA translated into the amino acid sequence of a polypeptide?
- *Breaking the Genetic Code:*
    - **Genetic code** – nucleotide information that specifies the amino acid sequence of a polypeptide
    - 4 bases in mRNA have to be used in combinations of at least 3 to provide the capacity for 20 amino acids
    - If only one-letter words were used in the code, then only 4 different amino acids could be specified (4^1) and if two-letter words were used, only 16
    - But for 3 letter words → 64 different amino acids
    - Each three letter word in the genetic code is called a **codon**
    - Genetic info in DNA is first transcribed into complementary 3-letter RNA codons
        - RNA polymerase reads 3'→5' DNA template strand to make complementary RNA molecule (5'→3' of mRNA matches 5'→3' of DNA non-template strand)
        - In translation, ribosomes read mRNA 5'→3', and each codon designates an amino acid sequence in resulting polypeptide
        - **REMEMBER:** template strand can either be the top OR bottom strand; doesn't matter! It just depends on what direction RNA polymerase is going
    - tRNAs are the molecules that bring amino acids to the ribosome for assembly into the polypeptide chain
        - They complementary base pair with the mRNA
- *Features of the Genetic Code:*
    - Codons are written in the 5'→3' direction as they appear in mRNAs
    - Of the 64 codons, 61 specify amino acids
    - 1 of these, **AUG**, specifies the amino acid methionine which is always the first codon translated in any mRNA in both prokaryotes and eukaryotes
        - Called the start codon
    - 3 codons that don't specify amino acids – UAA, UAG and UGA are stop codons
        - They indicate the end of a polypeptide-encoding sequence
        - When a ribosome reaches a stop codon, polypeptide synthesis stops and the new polypeptide chain is released from the ribosome
    - Most amino acids are represented by at least two or three codons
        - Therefore, there are many synonyms in the nucleic acid code, a feature known as **degeneracy**
    - The genetic code is also **commaless** – that is, the words of the nucleic acid code are sequential (no commas, spaces, etc. to make the end of one codon and the beginning of the next)
        - Therefore, the code can be read correctly only by starting at the right place – at the first base of the first 3-letter codon at the beginning of a coded message (in other words, the start codon)
        - Thus, there is only one correct **reading frame** for each mRNA
    - The code is also **universal** – the same codons specify the same amino acids in all living organisms, and even viruses
        - This suggests that the genetic code was established in its present form very early in the evolution of life and had remained virtually unchanged overtime
        - Consistency in the genetic code is what makes genetic engineering possible

14.2 – Regulation of Transcription in Eukaryotes (p. 315-322)

- In eukaryotes, there is no need to organize genes under the control of a single promoter in an operon
- 2 general categories of eukaryotic regulation:
    - Short-term regulation → regulatory events in which gene sets are quickly turned on or off in response to changes in environmental or physiological conditions in the cell's organism's environment (similar to prokaryotic gene regulation)
    - Long-term regulation → regulatory events required for an organism to develop and differentiate
- ***In eukaryotes, regulation of gene expression occurs at several levels***
    - More complicated gene expression than prokaryotes since eukaryotic cells are more complex, the nuclear DNA is organized with histones into chromatin, and since eukaryotes produce large numbers and different types of cells
        - Also, the nuclear envelope separates transcription and translation
    - Chromatin → DNA → pre-mRNA → mature mRNA → leaves nucleus → initiation of protein synthesis → new polypeptide chains → finished proteins → protein breakdown
    - Transcriptional regulation
        - Regulation of transcription initiation
        - Chromatin remodelling to make genes accessible for transcription
    - Posttranscriptional regulation
        - Variations in pre-mRNA processing
        - Removal of masking proteins
        - Variations in rate of mRNA breakdown
        - RNA interference
    - Translational regulation
        - Variations in rate of initiation of protein synthesis
    - Posttransational regulation
        - Variations in rate of protein processing
        - Removal of masking segments
        - Variations in rate of protein breakdown
- ***Regulation of transcription initiation involves the effects of proteins binding to a gene's promoter and regulatory sites***
    - **Organization of a eukaryotic protein-coding gene:**
        - Upstream of the transcription unit is the promoter
        - Promoter contains a TATA box (sequence of 25 bp upstream of the start point of transcription)
            - Plays an important role in transcription initiation
        - RNA polymerase II itself can't recognize the promoter sequence; instead, proteins called **transcription factors** recognize and bind to the TATA box and then recruit the polymerase
        - Once the RNA polymerase II-transcription factor complex forms, the polymerase unwinds the DNA and transcription begins
        - Adjacent to the promoter, farther upstream, is the **promoter proximal region** – contains regulatory sequences called promoter proximal elements

            - Regulatory proteins that bind to the promoter proximal elements may stimulate or inhibit the rate of transcription initiation
        - More distant from the beginning of the gene is the **enhancer**
            - Regulatory proteins binding to regulatory sequences within an enhancer also stimulate or inhibit the rate of transcription initiation
    - **Activation of Transcription:**
        - General transcription factors (proteins) bind to promoter in area of TATA box
        - These factors recruit RNA polymerase II, which alone can't bind to the promoter, and orient the enzyme to start transcription at the correct place
        - Combination of the factors with RNA polymerase is the **transcription initiation complex**
        - **Activators** – regulatory proteins that play a role in a positive regulatory system that controls the expression of one or more genes
            - Activators that bind to the promoter proximal elements interact directly with the transcription factors at the promoter to stimulate transcription initiation
        - House-keeping genes have promoter proximal elements that are recognized by activators present in all cell types
            - In constrast, genes expressed only in particular cell types or at particular times have promoter proximal elements that are recognized by activators found only in those cell types, or at those times when transcription of these genes needs to be activated
        - The DNA-binding and activation functions of activators are properties of two distinct domains in the proteins
        - The 3-D arrangement of amino acid chains within and between domains also produces highly specialized regions called **motifs**
            - Several types of motifs, each with a specialized function, are found in proteins, including motifs that insert into the DNA double helix
            - Motifs found on the DNA-binding domains of regulatory proteins, such as activators, include the helix-turn helix, zinc finger, and leucine zipper
            - **Helix-turn-helix motif** – part of a protein bound to DNA
                - One of the alpha helices binds to base pairs in the major groove of the DNA
                - A looped region of the protein connects to a second alpha helix that helps hold the first helix in place
            - **Zinc finger motif** - part of a protein in the presence of a bound zinc atom
                - Bind to specific base pairs in the grooves of DNA
            - **Leucine zipper proteins** – dimers, with each monomer consisting of alpha-helical segments
                - Hydrophobic interactions between leucine residues within the leucine zipper motif hold two monomers together
                - Other alpha helices bind to DNA base pairs in the major groove
        - Activators binding at the enhancer greatly increase transcription rates
            - Enhancers of different genes have different sets of regulatory sequences, which bind particular activators

        - **Coactivator** – larger multiprotein complex, forms a bridge between the activators at the enhancer and the proteins at the promoter and promoter proximal region, causing the DNA to form a loop
        - The interactions between the activators at the enhancer, the coactivator, the proteins at the promoter, and the RNA polymerase greatly stimulate transcription upto its maximal rate
- **Repression of Transcription:**
    - Repressor genes block or reduce the rate of transcription
    - Some may bind to the same regulatory sequences which activators bind (often in enhancer), thereby preventing activators from binding to that site
    - Others bind to their own specific site in the DNA near where the activator binds and interact with the activator so that it can't interact with the coactivator
    - Others bind to specific sites in the DNA and recruit **corepressors**, multiprotein complexes analogous to coactivators except that they are negative regulators, inhibiting transcription initiation
- **Combinatorial Gene Regulation –** a small # of regulatory proteins control transcription of all protein-coding genes. Different combinations of activators bind to enhancer regulatory sequences to control rate of transcription of each gene
    - General transcription factors bind to certain promoter sequences such as the TATA box and recruit RNA polymerase II, resulting in a basal level of transcription
    - Specific activators bind to promoter proximal elements and stimulate rate of transcription
    - Activators also bind to the enhancer to stimulate transcription of gene
    - Any given gene has a specific # and types of promoter proximal elements
        - The # and types of regulatory sequences in the enhancer is also specific for each gene
    - Both promoter proximal regions and enhancers are important in regulating transcription of gene
        - Each regulatory sequence in those two regions binds a specific regulatory protein
        - Since some regulatory proteins are activators and others are repressors, the overall effect of regulatory sequences depends on the particular proteins that bind to them
        - If activators bind both to the regulatory sequences in the proximal region and to the enhancer, transcription is activated maximally
        - But if a repressor binds to the enhancer and an activator binds to the promoter proximal element, the amount of gene expression depends upon the relative effects of those two regulatory proteins
            - E.g. if repressor is strong, gene expression will be reduced
    - If each gene were regulated by a single, distinct protein, the # of genes encoding regulatory proteins would have to equal the # of genes to be regulated
        - Regulating the regulators would require another set of genes equal in number, and so on until the coding capacity of any chromosome set would be exhausted

        - But since different genes require different combinations of regulatory proteins, the # of genes encoding regulatory proteins can be much lower than the # of genes the regulatory proteins control
  - **Coordinated Regulation of Transcription of Genes with Related Functions:**
    - All genes that are coordinately regulated have the same regulatory sequences associated with them
      - Therefore, with one signal, the transcription of all the genes can be controlled simultaneously
    - E.g. control of gene expression by steroid hormones in mammals
      - Steroid hormones act on specific target tissues within the body because only cells in those tissues have steroid hormone receptors in their cytoplasm
      - Hormone moves through plasma membrane into cytoplasm and the receptor binds onto it
      - Hormone-receptor complex enters nucleus and binds to specific regulatory sequences that are adjacent to the genes whose expression is controlled by the hormone
      - This binding activates transcription of those genes, and proteins encoded by the genes are synthesized rapidly
      - A single steroid hormone can regulate many different genes because all of the genes have an identical DNA sequence – a **steroid hormone response element** – to which the hormone-receptor complex binds
- ***Methylation of DNA can control gene transcription***
  - DNA methylation is when enzymes add a methyl group to cytosine bases in DNA
  - Methylated cytosines in promoter regions can regulate transcription through **silencing**, in which transcription of genes controlled by those promoters is greatly reduced
  - This is an example of **epigenetics** – a change in gene expression is achieved without a change in the DNA sequence of the gene or genome
  - Common among vertebrates
  - Can silence large blocks of genes, or even chromosomes
  - **Genomic imprinting** – the expression of an allele is determined by the parent that contributed it
    - Methylation permanently silences transcription of either the inherited maternal or the inherited paternal allele of a particular gene
- ***Chromatin structure plays an important role in whether a gene is active or inactive***
  - Eukaryotic DNA is organized into chromatin by combination with histone proteins
  - DNA is wrapped around histones, forming a nucleosome
  - Negative DNA and positive histones attract each other, making it stable
  - Genes in regions of the DNA that are tightly wound around histones in chromatin are less active, because their promoters are less accessible to the proteins that initiate transcription
  - For the gene to be activated, the chromatin structure must be altered in the vicinity of the promoter to provide access to the general transcription factors for transcription initiation
  - Process of changing chromatin structure = **chromatin remodelling**
    - 1st type → An activator may bind to regulatory sequence upstream of the gene's promoter and recruit a nucleosome remodelling complex, which uses energy

        from ATP hydrolysis to slide the nucleosome along DNA to expose the promoter, OR to restructure the nucleosome without moving it to allow transcription factors to bind
    - 2nd type → The activator can bind to a regulatory sequence upstream of the gene's promoter and recruit an enzyme that acetylates lysine amino acids in the tails of histones in the nucleosome, where the promoter is located. Acetylation makes histone less attractive to DNA, so promoters become accessible
  - Once mRNAs are transcribed from active genes, further regulation occurs at each of the major steps in the pathway from genes to proteins (next section)

14.3 – Posttranscriptional, Translational and Posttranslational Regulation (p. 322-226)

- ***Post-transcriptional Regulation Controls mRNA Availability to Ribosomes***
  - **Variations in Pre-mRNA Processing:**
    - These variations can regulate which proteins are made in cells
    - Pre-mRNAs can be processed by alternative splicing – produces different mRNAs from the same pre-mRNA by removing different combinations of exons along with the introns
    - The resulting mRNAs are translated to produce a family of related proteins with various combinations of amino acid sequences derived from the exons
    - Alternative splicing itself is under regulatory control
      - Regulatory proteins specific to the type of cell control which exons are removed from pre-mRNA molecules by binding to regulatory sequences within those molecules
    - Result is that appropriate proteins within a family are synthesized in cell types or tissues in which they function optimally
  - **Masking Proteins:**
    - Proteins that bind to mRNAs are made unavailable for protein synthesis
    - These controls are important in animal eggs (e.g. keeping mRNAs in inactive form until the egg has been fertilized and embryonic development is underway)
    - When an mRNA is to become active, other proteins remove the masking proteins and allow mRNA to enter protein synthesis
  - **Variations in the Rate of mRNA breakdown:**
    - Involves a regulatory molecule, such as a steroid hormone, directly or indirectly affecting the mRNA breakdown steps, either slowing or increasing the rate of those steps
  - **Regulation of Gene Expression by Small RNAs:**
    - RNA interference (RNAi) – non-coding single-stranded RNAs can bind to mRNAs and affect their translation
    - 2 groups of small regulatory RNAs are involved → miRNAs and siRNAs
    - **miRNA** – produced from RNA that is encoded in the cell's genome
      - RNA polymerase transcribes miRNA gene, producing a hairpin loop called pre-miRNA
      - Pre-miRNA is imported into cytoplasm
      - Dicer enzyme removes loop from pre-miRNA hairpin to leave a double-stranded RNA that base paired with itself

            - Protein complex then binds to double-stranded RNA
            - An enzyme in the protein complex degrades one of the RNA strands, leaving the miRNA
            - Together, the miRNA and the complex are called the **miRNA-induced silencing complex (miRISC)**
            - miRNA in the miRISC binds to sequences in the 3' UTRs of target mRNAs
            - If the miRNA and mRNA pair imperfectly, the double-stranded segment formed between the miRNA and mRNA blocks ribosomes from translating the mRNA
                - In this case, mRNA's expression is silenced (but the mRNA is not destroyed)
            - If they pair perfectly, an enzymes cleaves the target mRNA, destroying it and silencing its expression
            - Imperfect pairing more common in animals and perfect pairing more common in plants
        - **siRNA** – produced from double-stranded RNA that is not encoded by nuclear genes
        - Cells attacked by viruses with RNA genomes of a double-stranded RNA stage can defend themselves using siRNA that they produce from the virus's own RNA
            - Double-stranded RNA is cut by Dicer into short double-stranded RNA molecules, and then a protein complex binds to the molecules and degrades one of the RNA strands to produce single-stranded siRNA
            - The protein complex (siRNA-induced silencing complex) is similar to one that acts on double-stranded RNA precursors of miRNAs
            - Target RNA is cleaved and the pieces are then degraded
        - The expression of any gene can be knocked down to low levels or even completely when involving RNAi with siRNA
            - Equivalent to creating a mutated version of that gene, but without changing the gene's DNA sequence
            - To silence a gene, a double-stranded DNA is introduced into the cell that can be processed by Dicer and the protein complex into an siRNA complementary to the mRNA transcribed from that gene
            - Used to identify the functions of genes whose presence has been detected by sequencing complete genomes, but whose function is completely unknown
            - After an siRNA specific to gene of interest is introduced into cell, researchers look for a change in phenotype
- ***<u>Translational Regulation Controls the Rate of Protein Synthesis</u>***
    - Adjusting the length of the poly(A) tail of mRNA:
        - This is a string of adenine-containing nucleotides that is added to the 3' end of the pre-mRNA and is retained on the mRNA produced from the pre-mRNA after introns are removed
    - Enzymes can change the length of the poly(A) tail on an mRNA in the cytoplasm in either direction – by shortening it or lengthening it
    - Increases in length → results in increased translation
    - Decreases in length → results in decreased translation
- ***<u>Post-translational Regulation Controls the Availability of Functional Proteins</u>***

- **Chemical modification** – addition or removal of chemical groups, which reversibly alters the activity of the protein
- **Processing** – proteins are synthesized as inactive precursors, which are converted to an active form under regulatory control
- **Degradation** – also under regulatory control
    - Some proteins last for the lifetime of the individual, whereas others persist for only minutes
    - Proteins with short cellular lives include many of the proteins regulating transcription (these are usually marked for breakdown by enzymes by a ubiquitin tag)
        - This tag labels the proteins so that they are recognized and attacked by a proteasome, which unfolds the protein and protein-digesting enzymes then break it down into peptides
        - Peptides are released from proteasome

13.3 – Processing of mRNAs in Eukaryotes (p.291-294)

- mRNAs contain non-coding regions that play key roles in the process of protein synthesis, even though they don't specify any amino acids
- In prokaryotic mRNAs, the coding region is flanked by untranslated ends, the 5' UTR & 3' UTR
    - These same elements are found in eukaryotic mRNAs, along with additional types of non-coding elements
- ***Eukaryotic protein-coding genes are transcribed into a precursor mRNA that is modified in the nucleus***
    - **Modification of pre-mRNA and mRNA ends:**
        - At the 5' end of the pre-mRNA is the 5' guanine cap, consisting of a guanine-containing nucleotide that is reversed so that its 3'-OH group faces the beginning rather than the end of the molecule
            - A capping enzyme adds this 5' cap to the pre-mRNA (without the need for complementary base pairing) soon after RNA polymerase II begins transcription
            - The cap, which is connected to the rest of the chain by 3 phosphate groups, remains when pre-mRNA is processed to mRNA
            - The cap protects mRNA from degradation and is the site where ribosomes attach at the start of translation
        - Transcription of a eukaryotic protein-coding gene is terminated differently from that of a prokaryotic gene
            - Eukaryotic gene has no terminator sequence in the DNA that, after transcription into RNA, signals RNA polymerase to stop transcribing
            - Instead, near the 3' end of the gene is a DNA sequence that is transcribed into the pre-mRNA
                - Proteins bind to this polyadenylation signal in the RNA and cleave it just downstream
                - This signals the RNA polymerase to stop transcription
            - Then, poly(A) polymerase adds a chain of 50 to 250 adenine nucleotides, one at a time, to the newly created 3' end of the pre-mRNA

            - No complementary base-pairing with a template is needed for this type of RNA synthesis (no poly(T) sequence in the DNA corresponding to poly(A) in pre-mRNA)
         - The string of adenine nucleotides (called **poly A tail**) enables the mRNA produced from the pre-mRNA to be translated efficiently and protects it from attack by RNA-digesting enzymes in the cytoplasm
   - **Sequences interrupting the protein-coding sequence:**
      - The transcription unit of a protein coding gene – the RNA-coding sequence – also contains non-protein-coding sequences called **introns** that interrupt the protein-coding sequence
      - Introns are transcribed into pre-mRNAs but are removed from pre-mRNAs during processing in the nucleus
      - The amino acid-coding sequences that are retained in finished mRNAs are called **exons**
- ***<u>Introns are removed during pre-mRNA processing to produce translatable mRNA</u>***
   - **mRNA splicing** occurs in the nucleus
      - Removes introns from pre-mRNAs and joins exons together
      - Takes place in a **spliceosome** – a complex formed between the pre-mRNA and a handful of snRNPs
         - An snRNP particle is a complex of snRNA and proteins
         - snRNPs involved in mRNA splicing are located in the nucleus
         - They bind in a particular order to an intron in the pre-mRNA and form the active spliceosome
         - Spliceosome cleaves the pre-mRNA to release the intron, and joins the flanking exons
      - Complementary base-pairing between regions of snRNA and mRNA ensures that the cutting and splicing are so exact that not a single base of an intron is retained in the finished mRNA, nor is a single base removed from the exons
         - Without this precision, removing introns would change the reading frame of the coding portion of the mRNA!
- ***<u>Introns contribute to protein variability</u>***
   - Introns seem wasteful.. why are they present??
      - Provide a selective advantage to organisms by increasing the coding capacity of existing genes through a process called alternative splicing and in a process that generates new proteins called exon shuffling
   - **Alternative splicing**:
      - The removal of introns from a given pre-mRNA is not absolute
      - In particular tissues or sexes, or under certain environmental conditions, different regions of a given pre-mRNA may be identified as introns and removed in different combinations to produce mature mRNAs
      - *Regions that are exon in one situation may well be removed as intron in another situation*
         - E.g. alternative splicing of the same pre-mRNA for both smooth muscle and skeletal muscle is distinct in terms of its final mRNA
            - The same introns are removed, but different exons
            - These two forms of the mRNA

    - This mechanism greatly increases the number of variety of proteins encoded in the cell nucleus without increasing the size of the genome
    - The different mRNAs produced from the parent pre-mRNA are translated to produce a family of related proteins with various combinations of amino acid sequences derived from the exons
      - Each protein in the family, then, varies in its function
    - This helps explain why humans with only 20 000 genes can make over 100 000 different proteins
      - Ultimately it is the diversity of the proteins available, not the amount or diversity of DNA sequence, that determines the relative complexity of an organism's functions
    - Due to alternative splicing, we know that for some genes, one gene may specify a number of polypeptides, each of which has a related function
- **Exon Shuffling:**
  - Intron-exon junctions often fall at points dividing major functional regions in encoded proteins
  - The functional divisions may have allowed new proteins to evolve by exon shuffling, a process by which existing protein regions or domains, already selected for due to their useful functions, are mixed into novel combinations to create new proteins
  - Evolution by this mechanism would produce new proteins with novel functions much more quickly than by changes in individual nucleotides at random points

13.1 – The Connection between DNA, RNA and Protein (p. 284-286)

- Proteins are encoded by genes made of DNA
- *Genes Specify Either Protein or RNA Products*
  - Initially the One-gene-one enzyme hypothesis → shows direct relationship between enzymes and genes
  - Many proteins consist of more than one subunit; each of these subunits is a separate molecule, called a polypeptide, that is coded by a separate gene
  - Polypeptides can assemble to create a functional cluster of molecules called a protein
  - Later turned into One-gene-One Polypeptide hypothesis
  - **Protein** – the functional collection of polypeptides
  - **Polypeptide** – the molecule encoded by a gene
- *The Pathway from Gene to Polypeptide Involves Transcription and Translation*
  - **Transcription** – mechanism by which the info encoded in DNA is made into complementary RNA copy
    - **RNA polymerase** – creates an RNA sequence that is complementary to the DNA sequence of a given gene
    - The process follows the same basic rules of complementary base-pairing and nucleic acid chemistry from DNA replication
    - For each of the several 1000 genes that will be appropriate to express in a given cell, one DNA strand or the other is the **template strand** and is read by the RNA polymerase
    - The RNA transcribed from a gene encoding a polypeptide is called **mRNA**

    - **Translation** – the use of the info encoded in the RNA to assemble amino acids into a polypeptide
        - An mRNA associates with a **ribosome** – a particle on which amino acids are linked into polypeptide chains
        - As the ribosome moves along mRNA, amino acids specified by mRNA are joined one by one to form the polypeptide encoded by the gene
    - **Central Dogma** – flow of information from DNA to RNA to protein
    - Transcription and translation are similar in prokaryotes and eukaryotes
        - Key difference → prokaryotic cells can transcribe and translate a given gene simultaneously, whereas eukaryotic cells transcribe and process mRNA in the nucleus before exporting it to the cytoplasm for translation on ribosomes

11.5b – Genomic Imprinting (p. 253-254)

- **Genomic imprinting** – the expression of an allele of a gene is determined by the parent that contributed it
    - In some cases, the paternally derived allele is expressed; others, the maternal
- The allele inherited from one of the parents is expressed whereas the other allele is silent
- The silent allele (one that isn't expressed) is called the **imprinted allele**
    - The imprinted allele is not inactivated by mutation
    - Rather, it is silenced by chemical modification (**methylation**) of certain bases in its sequence
- Example of imprinting → Angelman syndrome (AS) and Prader-Willi Syndrome (PWS) are caused by imprinting of a particular gene on a chromosome inherited from one parent, coincident with deletion of the same gene on the homologous chromosome inherited from the other parent
    - The syndromes differ with respect to the gene imprinted
    - PWS is caused when someone has a normal maternally derived chromosome 15 and a paternally derived chromosome 15 with a deletion of a region that includes the PWS gene
        - The PWS gene is imprinted, and therefore silenced, on maternally derived chromosomes
        - So, when there is no PWS gene on the paternally derived chromosome, there is no PWS gene activity and PWS results
    - AS is caused when someone has a normal paternally derived chromosome 15 and a maternally derived chromosome 15 with a deletion of the same region that includes the AS gene
        - Imprinting silences the AS gene on the paternally derived chromosome
        - Because there is no AS gene on the maternally derived chromosome, there is no AS gene activity and AS syndrome develops
- Although imprinted traits can show a parent of origin effect, it is not the same as sex-linkage
    - Imprinted alleles aren't necessarily carried on sex chromosomes and any given sex-linked allele can be inherited from either a mother or father
- Mechanism of imprinting involves the modification of the DNA in the region that controls the expression of a gene by the addition of methyl (-CH_3) groups to cytosine nucleotides
    - The methylation of the control region of a gene prevents it from being expressed

- Imprinting occurs in gametes where the allele is destined to be inactive in the new embryo after fertilization is methylated
    - The methylated (silenced) state of the gene is passed on as the cells grow and divide to produce the somatic (body) cells of the organism
- Cancers are associated with the failure to imprint genes
    - **Loss of imprinting** – imprinting mechanism for a gene doesn't work, resulting in both alleles of the gene to be active
    - Disrupts cell division cycle, increasing the risk of uncontrolled growth and cancer

11.2e – Inactivation of One X Chromosome Evens out Gene Effects (p. 244-245)

- Females have twice as many copies of genes carried on the X chromosome as males, but they don't even require all of them
- Products from genes on the X chromosome could be equalized in males and females if one X chromosome were turned off in females (genes inactivated)
- As a result of this, the activity of most genes carried on the X chromosome is essentially the same in males and females
- Inactivation occurs by a condensation process that folds and packs the chromatin of one of the two X chromosomes into a tightly coiled state
    - The inactive, condensed X chromosome can be seen within the nucleus in cells of females as a dense mass of chromatin called the **barr body**
- Inactivation occurs during embryonic development
- Which of the 2 X's becomes inactive in a particular embryonic cell line is a random event
    - But once one of the X chromosomes is inactivated, the same X is inactivated in all descendants of the cell
- If the 2 X chromosomes carry different alleles of a gene, one allele will be active in cell lines in which one X chromosome is active, and the other allele will be active in cell lines in which the other X chromosome is active
- For many sex-linked alleles, random inactivation of either X chromosome has little overall wholebody effect in heterozygous females because the dominant allele is active in enough of the critical cells to produce a normal phenotype
    - However, for some genes, the inactivation of either X chromosome in heterozygotes produces recognizably different effects in distinct regions of the body
    - E.g. orange and black patches of fur in cats results from inactivation of one of the two X chromosomes in regions of the skin of heterozygous females (males have either orange or black fur since they only get one of the two alleles)

12.5a – Histones Pack Eukaryotic DNA (p. 278-279)

- **Histones** – small positively charged proteins that are complexed with DNA in the chromosomes of eukaryotes (links to negative phosphate groups in DNA)
    - They pack DNA molecules into the narrow confines of the cell nucleus
    - Combination with histones compacts the huge length of DNA so much that it fits into nuclei
    - 5 types of histones
- Histones pack DNA at several levels of chromatin structure:

  - **Nucleosome** – 2 molecules each of 4 histones combine to form a bead-like, 8-protein nucleosome core particle around which DNA winds for almost two turns
  - **The linker** is a short segment of DNA that extends between one nucleosome and the next
  - Each nucleosome and linker includes about 200 bp of DNA
  - Nucleosomes compact DNA by a factor of 7
- The 5th type of histone (H1) binds both to the nucleosomes and to the linker DNA
  - This binding causes nucleosomes to be packaged into a coiled structure (**solenoid**)
- The arrangement of DNA in nucleosomes and solenoids compacts DNA and protects it from chemical and mechanical damage
  - Makes it more resistant to degradation by deoxyribonuclease
- In interphase nuclei, chromatin fibres are loosely packed in some regions and densely packed in others
  - Loosely packed regions – **euchromatin**
  - Densely packed regions – **heterochromatin**
    - Heterochromatin represents large blocks of genes that have been turned off and placed in a compact form
    - E.g. as one of the 2 X chromosomes becomes inactive, it packs down into a block of heterochromatin called the Barr body

<u>15.3 – Genome Analysis (p. 353-358)</u>

- <u>*DNA Sequencing techniques are based on DNA replication*</u>
  - Used to determine the sequence of individual genes that have been cloned or amplified by PCR
  - **Sanger method** – based on dideoxy sequencing
    - Dideoxyribonucleotides have a single –H bond to the 3’ C of the deoxyribose sugar, instead of the –OH normally at this position in deoxyribonucleotides
    - DNA polymerases recognize the dideoxyribonucleotides and place them in the DNA, just like they would normally
    - However, since a dideoxyribonucleotide has no 3’-OH group available for addition of the next base, replication of a nucleotide chain stops when one of these nucleotides is added to a growing nucleotide chain
    - In a dideoxy sequencing reaction, a mixture of dideoxyribonucleotides and normal nucleotides is used so that chain termination will occur randomly at each position where a particular nucleotide appears in the population of DNA molecules being replicated
    - Each chain-termination event generates a newly synthesized DNA strand that ends with the dideoxyribonucleotide
    - Hence, for this particular strand, the base at the 3’end is known, and thus, the base on the template strand being sequenced can be deduced
    - Once they know the base at the end of each terminated DNA strand, researchers can figure out the complete sequence of the template DNA strand
    - This method can be used with any pure piece of DNA
  - Genome analysis consists of 3 main areas:

        - **Genome sequence determination and annotation** – obtaining the sequences of complete genomes and analyzing them to locate putative genes and other functionally important sequences within the genome
        - **Functional genomics** – the study of the functions of genes and other parts of the genome, including regulation of gene expression, role of proteins, etc.
        - **Comparative genomics** – the comparison of entire genomes to understand evolutionary relationships and the basic biological similarities and differences among species
- *Genome sequence determination and annotation involves obtaining and analyzing the sequences of complete genomes*
    - **Whole-genome shutgun method** – entire genome is broken into thousands of random, overlapping fragments
        - Each fragment is cloned and sequenced
        - The genome sequence is then assembled by computer on the basis of the sequence overlaps between fragments
        - Once the complete sequence of a genome has been determined, the next step is annotation → the identification of genes and other sequences of importance
    - Protein-coding genes are of particular interest in genome analysis
        - Using computer analysis, researchers identify possible protein-coding genes by searching for **open reading frames (ORFs)**
            - This is a start codon (ATG) separated by a multiple of 3 nucleotides from one of the stop codons (TAG, TAA, TGA)
        - This process is easy for prokaryotic genomes since the genes have no introns
    - With many genomes sequenced, they can be compared to learn about genome sizes and the number of PCGs
        - In bacteria and archaea, genes are densely packed in the genomes
            - Larger genomes of organisms in these domains tend to reflect increased gene number
        - Members of eukarya vary in form an complexity, and their genomes also show great differences in size
            - No rules for relating organism complexity with size
        - In general, genes in eukaryotes are packed less densely than prokaryotes
        - Eukaryotes have many non-coding sequences
    - Protein-coding genes occupy less than 2% of human genome, yet they code for over 100,000 proteins (due to alternative splicing, etc.)
        - Introns occupy 24% of the genome (non-coding)
        - The rest of the DNA is basically the spaces between genes
            - Some contains promoters/enhancers; others are just repeated sequences with no function
- *Functional Genomics focuses on the functions of genes and other parts of the genome*
    - The complete genome sequence for an organism is a long string of A, T, G, and C, but this means little without further analysis
    - **Assigning gene function by sequence similarity:**
        - Computer analysis of a genome sequence will reveal its putative genes
        - For a newly sequenced genome, researchers use databases to assign functions to the putative genes identified in the initial computer analysis

        - They look for sequence matches in sequence similarity searches, in which an input sequence is compared with all sequences in a database
        - Sequence similarity searching can assign probable functions to genes in a newly obtained genome sequence because homology – descent from a common ancestor – reflects evolutionary relationships
            - So, two genes from different organisms will be similar if they are homologous genes with a common ancestor
            - Differences between the genes will have resulted from mutations overtime
        - However, there are still many putative genes with unknown functions
    - **Assigning gene function experimentally:**
        - You can knock out or knock down the function of a gene and determine which phenotypic process changes
        - The altered phenotype informs the researcher about the function of the normal gene
        - 2 main methods used → gene knockouts and RNA interference (RNAi)
        - *Gene knockout* – uses molecular techniques to disrupt the gene in the chromosome
        - *RNA interference* – knocks down the expression of a gene at translation level
            - A small regulatory RNA is transcribed from an expression plasmid introduced into cell
            - The sequence of that regulatory RNA can form complementary base pairs with the mRNA of the gene of interest
            - Base-pairing triggers the RNA interference molecular mechanisms, which knock down the expression of the gene by causing degradation of that gene's mRNA by blocking its translation

8.5 – Cell Cycle Regulation (p. 174-178)

- As part of the internal controls, the cell cycle has built-in **checkpoints** to prevent critical phases from beginning until the previous phases are completed
- Hormones, growth factors and other external controls coordinate the cell cycle with the needs of an organism by stimulating or inhibiting division
- *Cyclins and Cyclin-dependent kinases are the internal controls that directly regulate cell division*
    - **CDKs** are protein kinases that add phosphate groups to target proteins
    - CDKs are called cyclin-dependent because they are switched on only when combined with another protein called **cyclin**
    - Since the concentration of the cyclins rises and falls during the cell cycle, so does the enzyme activity of CDKs (even though its concentration remains constant)
    - Several different cyclin/CDK combinations regulate cell cycle transitions at different checkpoints
    - As cyclin levels rise, CDK becomes active
    - Cyclin is degraded in the end
- *Internal checkpoints stop the cell cycle if stages are complete*
    - While cyclin/CDK directly controls the cell cycle, other factors within the act act as indirect controls by altering the activity of the cyclin/CDK complexes

    - At each checkpoint, regulatory events block the cyclin/CDK complex from triggering the associated cell cycle transition until the actions of a previous phase are successfully completed
- *External controls coordinate the mitotic cell cycle of individual cells with the overall activities of the organism*
    - The internal controls that regulate the cycle are modified by signal molecules that originate from outside the dividing cells
    - These signal molecules include peptide hormones and growth or death factors
    - These external factors bind to receptors at the cell surface, which respond by triggering reactions inside the cell
        - Reactions include steps that add phosphate groups to the cyclin/CDK complexes, thus affecting their function
    - Overall effect is to speed, slow or stop the progress of cell division
    - Cell-surface receptors in animal cells also recognize contact with other cells or with molecules of the extracellular matrix
        - The contact triggers internal reaction pathways that inhibit division by arresting the cell cycle, usually in G1 phase
        - The response, called **contact inhibition**, stabilizes cell growth in fully developed organs and issues
        - As long as the cells of most tissues are in contact with one another or with the extracellular matrix, they are shunted into G0 phase and prevented from dividing
- *Most cells in a multicellular body cannot divide indefinitely*
    - **Cellular senescence** – the loss of proliferative ability over time
        - Caused by the Hayflick factors: DNA damage and telomere shortening
    - The progressive accumulation of random damage to a cell's DNA sequence, or its chromosome structure, or even the genes coding for the enezyme machinery needed to repair such damage
        - "Older" cells would have diminished function if they have suffered mutations in genes controlling critical activities
    - Telomeres are repetitive DNA sequences added to the ends of chromosomes by telomerase
        - Since DNA replication is unable to replicate entire ends of linear chromosomes, telomere sequences are lost at each round of replication
        - Once telomeres diminish to a certain minimum length, cells stop dividing and may die
    - Senescence is an important anti-tumour mechanism
        - Faster senescence = resistance to cancer
- *Cell cycle controls are lost in cancer*
    - Cancer occurs when cells lose the normal controls that determine when and how often they will divide
    - Cancer cells divide continuously and uncontrollably, producing a rapidly growing mass called a tumour
    - Cancer cells also lose their adhesions to other cells and become actively mobile
        - Metastasis occurs, in which they tend to break loose from an original tumour, spread throughout the body, and grow into new tumours in other body regions
    - Growing tumours damage surrounding normal tissues by compressing them and interfering with blood supply and nerve function

  - Cancer cells have accumulated mutations in a variety of different genes that promote uncontrolled cell division or metastasis
    - Before they undergo mutation, many of these genes codes for components of cyclin/CDK system that regulates cell division; others encode proteins that regulate gene expression
    - When mutated, the genes, called **oncogenes**, encode altered versions of these products
- *Some cells are programmed to die*
  - **Apoptosis** – programmed-cell death
  - Initiation of cell death can result from either internal or external signals
  - "Executioner" enzyme is one of a family of normally inactive proteases, called **caspases** and is coded by CED-3
  - If a cell is destined to die by apoptosis, the cascade begins when internal developmental cues stimulate expression of a gene called EGL01
  - EGL-1 protein binds to CED-9, resulting in the release of bound CED-4 protein
    - CED-3 caspase is thus activated and cell death ensures
  - The causes of death are nuclear DNA degradation and disrupted mitochondrial function
  - Removing cells that are surplus for development is one function of apoptosis
  - Other cells are programmed to die because it would be beneficial for an organism to provoke apoptosis in cells suffering severe DNA damage, viral infection or mutations leadings to uncontrolled division

34.7 – Genetic and Molecular Control of Development (p. 845-851)

- *Genetic control of development*
  - Gene expression regulates changes that occur through determination and differentiation
  - Molecular mechanisms involved in determination and differentiation usually depend on regulatory genes that encode regulatory proteins that control the expression of other genes
  - Regulatory genes act as master regulators and, in most cases, the expression of the regulatory genes is controlled by induction
- *Gene control of pattern formation: developing a body*
  - As part of the signals guiding differentiation, cells receive positional information that tells them where they are in the embryo
  - This info is vital to **pattern formation** – the arrangement of organs and body structures in their proper 3-D relationships
  - Positional info is laid down primarily as concentration gradients of regulatory molecules produced by genetic control
  - In most cases, gradients of several different regulatory molecules interact to tell a cell, or cell nucleus, where it is in the embryo
- *Embryogenesis in Drosphila: Fruit Fly Model*
  - Following fertilization, division of the nucleus begins by mitosis
  - This produces a multinucleate blastoderm because the cytoplasm does not divide in the early embryo
  - The nuclei then migrate to the periphery of the embryo, where they are organized into separate cells

  - At this stage, the embryo is a **cellular blastoderm**, which then develops into a segmented embryo
  - Egg eventually hatches into a larva that will undergo 3 moults before becoming a pupa
  - The adult fly then emerges after metamorphosis
  - When homozygous, recessive embryonic lethal mutations result in embryo death during development
- *Maternal-effect genes and segmentation genes: segmenting the body*
  - Both of these types of genes work sequentially to regulate the expression of other genes and control the establishment of the embryo's body plan
  - **Maternal-effect genes** – expressed by the mother during oogenesis
    - These control egg polarity and thus embryo polarity
    - Some control formation of embryonic anterior structures, whereas others control formation of posterior structures
    - BICOID GENE – responsible for the development of the head and thorax
      - Transcribed in the mother during oogenesis, and the resulting mRNAs are deposited in the egg
      - After fertilization, translation of mRNAs produces BICOID protein, which diffuses through the zygote to form a gradient with concentration highest at the anterior end
      - BICOID is a transcription factor that activates some genes and represses others along the anterior-posterior axis of the embryo
      - Embryos with mutations in the BICOID gene lack thoracic structures but have posterior structures at each end
      - In normal embryos, the BICOID gene is a master regulator gene controlling the expression of genes for the development of anterior structures (head and thorax)
    - NANOS GENE – responsible for the development of posterior structures
      - When this gene is mutated, embryos lack abdominal segments
  - Once the axis of the embryo is set, expression of **segmentation genes** progressively subdivides the embryo into regions, determining the segments of the embryo and the adult
    - Gradients of BICOID and other proteins encoded by maternal-effect genes regulate expression of the embryo's segmentation genes differentially
    - So each segmentation gene is expressed at a particular time and location during embryogenesis
    - 3 sets of segmentation genes are regulated in a cascade of gene activations
      - **Gap genes** are the first to be expressed – they divide the embryo along the anterior-posterior axis into broad regions
        - Activated based on their positions in maternally-directed anterior-posterior axis of the zygote by reacting to the concentrations of BICOID and other proteins
        - Mutations in gap genes result in the loss of one or more body segments in the embryo
      - Products of gap genes are transcription factors that activate **pair-rule genes** – divide the embryo into units of two segments each
        - Mutations in these genes lead to the deletion of every other segment of the embryo

            - Pair-rule genes regulate expression of **segment polarity genes** – set boundaries and the anterior-posterior axis of each segment in the embryo
                - Mutations in the segment polarity genes produce segments in which one part is missing and the other part is duplicated as a mirror image
                - The products of segment polarity genes are transcription factors and other molecules that regulate other genes involved in laying down the pattern of the embryo
- *Homeotic Genes: structure and determining outcomes*
    - Once the segmentation pattern has been set, **homeotic genes** of the embryo specify what each segment will become after metamorphosis
    - **Homeotic genes –** master regulatory genes controlling development of structures such as eyes, antennae, legs, and wings on particular segments
    - These genes regulate development by encoding transcription factors that regulate expression of genes responsible for the development of adult structures
    - Each homeotic gene has a common region called a **homeobox** that corresponds to an amino acid section of the encoded transcription factor called the homeodomain
        - The homeodomain of each protein binds to a region in the promoters of the genes whose transcription it regulates
    - 8 homeobox (hox) genes in drosophila are organized along a chromosome in the same order as they are expressed along the anterior-posterior body axis
    - Hox genes are present in all major animal phyla, where they control the development of segments or regions of the body and are arranged in order in the genome
    - Homeobox sequences in hox genes are highly conserved, indicating common function in the wide range of animals in which they occur
    - Homeobox sequences of mammals are the same as or similar to those of the fruit fly
        - These genes are also found in plants
- *Cell-Death Genes: Apoptosis*
    - Apoptosis results from gene activation in response to molecular signals from receptors on the surfaces of marked cells
        - The signals are death notices, delivered at specific times during embryonic development
        - A death signal molecule that binds to a receptor in the plasma membrane of the target cell results in apoptosis
        - When the receptor is activated, it lead to inactivation of CED-9
        - In the absence of CED-9, CED-4 is activated, which in turn activates CED-3, triggering a cascade of reactions that kills the cell
    - The killing proteins remain inactive in the absence of the death signal
        - This allows CED-9, a protein associated with the outer mitochondrial membrane, to inhibit CED-4 and CED-3
    - Cells with CED-9 active normally survive in the adult stage
- *Epigenetics*
    - If each cell in a multicellular organism has identical DNA, how does the animal end up with so many different cell phenotypes?
    - **Epigenetics** – non-genetic cellular memory that reflects developmental and environmental conditions

    - Epigenetic modifications alter the expression of DNA but not its sequence! → they are NOT mutations
        - They basically just cause genes to be switched on and off
        - May involve modifications to histones
            - Methylation tightens packing of genes around histones, affecting down-regulating and inhibiting transcription
            - Acetylation loosens packing and encourages transcription and translation
        - Without these modifications, all genes would be functional in every cell, severely inhibiting or preventing cell differentiation
    - In mammals, epigenetic processes occur at 2 stages during development:
        - Primordial germ cells AND zygote after fertilization (but before implantation)
        - These changes can involve erasure of DNA methylation throughout the genome in the zygote and in primordial germ cells
    - Epigenetic changes account for different patterns of activation of the X chromosome in mammals
        - Each female receives two X's (one maternal, one paternal), but often only one is activated
        - Details of activation varies among mammals, and also among cell lines within a developing embryo
        - Epigenetic effects can account for striking differences among cells an in whole organism phenotypes

<u>**Experimental Evolution – Article**</u>

- An experiment studying bacterial populations over thousands of generations shows that a novel trait can evolve through rearrangement and amplification of a few pre-existing genes
- Lenski took a clone of fast-growing bacterium E coli and founded 12 independent populations, each resident in a flask containing a simple medium, with glucose as the sole source of carbon
- Everyday for almost 25 years, a sample of each population has been transferred to a fresh flask – which means that these populations have gone through 55 000 generations
- In addition to daily transfer, the researchers have been periodically storing samples at -80 degrees C, thereby ensuring access to a detailed history of ancestral states
    - Unlike fossils, these frozen ancestors are readily brought back to life
- After about 31 000 generations, one of the populations (A-3) evolved the capacity to use citrate as a nutrient
    - Citrate is a chelating agent that facilitates iron uptake; it is a component of the growth medium
    - Although E coli can digest citrate when oxygen is absent, it typically cannot do so under the aerobic conditions of this experiment
    - The bacterium has the building blocks necessary for the evolution of aerobic citrate utilization, but realization of this capacity requires a rewiring and refinement of regulatory connectivities
- To understand the nature of the genetic changes conferring the ability to use citrate in the presence of oxygen (a trait dubbed the Cit+ phenotype), they resuscitated ancestral Cit- populations from earlier time points in the A-3 lineage

- When the researchers tried to replay the evolution of the trait, they found that the Cit+ phenotype appeared again only in populations derived from the most recent Cit- lineages
  - This result suggested that genetic changes had occurred in these lineages that made it possible for the Cit+ phenotype to evolve
  - Although the precise nature of such ‘potentiating’ mutations is unclear, the important fact is that they took place
- The step that made the cells weakly Cit+ was the **actualization** step
  - A genomic region that carries the gene CitT was examined, which encodes a protein (citrate transporter) needed for citrate to enter the cell
  - In the original bacterial strain, citT is located downstream of citG (another gene required for citrate utilization) and RNK (a functionally unrelated gene that participates in energy metabolism)
  - In each of the derived Cit+ cells, they found a distinctive genomic rearrangement that fused RNK to citG
    - As a result, expression of citG and citT came under the control of the promoter of RNK, which allowed their expression in the presence of oxygen
  - They also found that a single copy of the genetic rearrangement was insufficient to generate the cit+ phenotype, and that there was a tandem array of between two and nine copies in Cit+ cells
    - Therefore, gene duplications took a very low level of citrate transport and propelled it to life-sustaining levels
    - All cells that inherited the genetic fusion (the actualizing mutation), followed by amplification (a refining mutation) were cit+