Genetics Exam 4: Transcription, Translation, and Gene Regulation

The central dogma: DNA Replication → Transcription → Translation

- DNA Replication: makes DNA copies that are transmitted from cell to cell and from parent to offspring
  - Chromosomal DNA: stores information in units called genes
  - Start with 2 antiparallel strands of DNA – only one of which is used for transcription
- Transcription: produces an RNA copy of a gene
  - Messenger RNA: a temporary copy of a gene that contains information to make a polypeptide
- Translation: produces a polypeptide using the information in mRNA
  - Polypeptide: becomes part of a functional protein that contributes to an organism's phenotype

Translation:

- *Gene Expression:* is the overall process by which the information within a gene is used to produce a functional product which can, in concert with environmental factors determine a trait
  - Gene expression is converting information contained within a genome into other forms of information
  - Functional product will always be RNA but this may go on to make other forms of RNA or other types of products
  - For gene expression/ transcription to occur proteins must be able to recognize and act on DNA
  - DNA base sequences define the beginning and end of a gene and regulate the level of RNA synthesis
- Organization of a bacterial gene:
  - DNA
    - *Regulatory Sequences*: located upstream from the promotor; site for regulatory proteins; the role of regulatory proteins is to influence the rate of transcription; regulatory sequences can be found in a variety of locations
    - *Promotor*: Site for RNA polymerase binding; signals the beginning of transcription; sequence of DNA (retained across organisms) that defines start boundaries for genes; located between the regulatory sequence and the body of the gene
      - This sequence is not transcribed; the +1 transcription start site is the first base added and the RNA is transcribed in the 5' – 3' direction
    - *Terminator*: signals the end of transcription
  - mRNA
    - *Ribosome-Binding Site:* site for ribosome binding; translation begins near this site in the mRNA; in eukaryotes the ribosome scans the mRNA for a start codon
    - *Start Codon:* specifies the first amino acid in a polypeptide sequence, usually formyl-methionine (in bacteria) or a methionine (in eukaryotes)
    - *Codons:* 3 nucleotide sequences within the mRNA that specify particular amino acids; the sequence of codons within mRNA determines the sequence of amino acids within a polypeptide
    - *Stop Codon:* specifies the end of polypeptide synthesis
    - Bacterial mRNA may be polycistronic, which means it encodes two or more polypeptides
- Gene expression requires base sequences:
  - The DNA strand that is used as a template is termed the *template strand* or the *anti-sense strand*
    - The RNA transcript is complementary to the template strand
    - RNA transcript is what "makes sense" because it is what is used to make the proteins
  - The opposite strand is called the *coding strand* or the *sense strand* as well as the *non-template strand*
    - This strand is not used to build the RNA
    - The base sequence is identical to the RNA transcript (has the same base sequence as the RNA)
      - except for the substitution of uracil in the RNA for thymine in the DNA
  - *Transcription Factors* recognize the promotor and regulatory sequences to control transcription
    - Proteins, not the RNA polymerase, aid the polymerase in localizing specific sequences in the transcription start site (recognize the promotor regions, etc.) – polymerase otherwise does not recognize sequences and just synthesizes RNA

- When you know that you have a triplet code, you can quickly count the number of possible triplets and compare this to the 20 amino acids to show that we have a degenerate code

- The genetic code:
  - The start codon in all organisms is AUG which codes for methionine which is the start of every protein
  - When there is a stop codon, translation stops
  - Molecular genetics is coded in sequences (at the DNA level and RNA level)
  - Stop codons do not receive a tRNA leading to termination
  - Codes only work when they are in frame – based on the start codon at the translation +1 start site
  - *Degenerate:* multiple codons may encode the same amino acid
    - in most cases the third base is the variable base
  - Triplet codons correspond to a specific amino acid. These are *sense codons*
  - Proteins are translated from 5' to 3' on the mRNA strand and are synthesized from N terminal to C terminal (analogous to 3' to 5')
  - Anticodons are complementary to codons
  - tRNA has the same sequence as the template strand
  - mRNA has the same sequence as the coding strand
  - in the mRNA there is a 5' and 3' untranslated regions (starts at the first AUG and with the stop codon that is in frame)
  - Remember that genes can run both ways in DNA
- Polypeptides:
  - Polypeptide synthesis has directionality that parallels the 5'-3' orientation of mRNA
  - During each elongation cycle, a peptide bond is formed between the carboxyl group of the last amino acid in the polypeptide chain and the amino group in the amino acid being added
  - Synthesis occurs in the N to C direction
  - The first amino acid has an exposed amino – this is the *N terminal* or *amino terminal*
  - The last amino acid has an exposed carboxyl terminal – this is the *C terminal* or the *carboxyl terminal*
  - Polypeptides involve *dehydration synthesis* in which water is lost when the Oxygen from the carboxyl terminal of the existing polypeptide attacks the 2 H's on the amino terminal of the incoming amino acid
  - Amino acids with OH groups can be phosphorylated by kinases – ex. TFIIH acts as a kinase by phosphorylating the CTD of RNA polymerase II
  - *Nonpolar Amino Acids* are hydrophobic and are often buried within the interior of a folded protein
    - They cluster in the globular portion of the protein
  - *Polar* and *charge amino acids* are hydrophilic and they are more likely to be on the surface of the protein
  - *Nonstandard amino acids* are found at a low frequency
  - Levels of structure:
    - *Primary structure* is a proteins amino acid sequence – while being translated, further folding will occur that is dictated by the amino acid sequence and this folding might be aided by chaperones
    - *Secondary structure* is the regular repeating shapes – there are two secondary structures alpha helices and beta sheets
      - Certain amino acids are good candidates for each structure
      - Secondary structures are stabilized by hydrogen bonds in the polypeptide backbone
      - Sigma factor is an example of secondary structure because it uses its alpha helix in the major groove of DNA to identify the promotor region in bacteria – the side chains poke out from the helix to contact the bases
    - *Tertiary structure* is the short regions of secondary structure in a protein folding into a 3D structure
      - Structure is determined by hydrophobic and ionic interactions as well as hydrogen bonds and van der waals interactions
    - *Quaternary structure* is proteins that are made up of two or more polypeptides
      - This is formed when the various polypeptides associated with one another to make a functional protein

        - The repressor protein inhibits the binding of TFIID to the core promotor or inhibits its function silencing transcription
    - Mediator
        - Transcriptional activator stimulates the function of mediator enabling RNA pol to form a pre-initiation complex – promotes phosphorylation of RNA pol II
        - Transcription repressor inhibits the mediator suppressing transcription
        - Can change the ability of the CTD to be phosphorylated by interacting with TFIIH
        - Some transcription factors stabilize/destabilize mediator to enhance/suppress phosphorylation
    - Recruiting proteins that affect nucleosome composition
- Modulation: activators and repressors may be modulated by the binding of
    - small effector molecules
    - protein-protein interactions
        - example: Glucocorticoids: uses steroid hormone receptors on cell
            - influence nutrient metabolism by promoting glucose utilization, fat mobilization, and protein break down
            - *Glucocorticoid Response Elements:* these function as enhancers; located near dozens of different genes so the hormone can activate many genes
            - Steps:
                1. Glucocorticoids enter the cell and bind to the glucocorticoid receptor displacing heat shock proteins
                2. This unblocks the dimer binding interface
                3. The homodimer can now enter the nucleus
                4. And target transcription in a particular gene
    - covalent modification
        - an extracellular molecule causes conformational changes in proteins.
        - Example: the cAMP response element binding (CREB) protein – cAMP is activated which binds to PKA which phosphorylates KREB protein dimer which then can bind to regulator sequences to stimulate transcription (phosphorylate CTD in RNA Pol II)
            - CREB protein becomes activated in response to cell signaling molecules that cause an increase in the cytoplasmic concentration of cAMP
            - The CREB protein recognizes a consensus sequence
- Regulatory proteins might alter nucleosomes near the promotor
    - Chromatin is a dynamic structure that can alternate between heterochromatin and euchromatin
        - Closed conformation (*Heterochromatin*): chromatin is very tightly packed and transcription might be difficult or impossible
        - Open conformation (*euchromatin)*: chromatin is accessible to transcription factors so transcription can take place
        - Histone modification is catalyzed by histone acyltransferase and histone deacetylase – they change the histone/DNA interaction
            - Lysine residues become acetylated reducing protein positivity and thus affinity to nucleosomes is reduced allowing accessibility to transcription proteins – these changes are reversible
        - ATP dependent chromatin refers to the dynamic changes in chromatin structure
            - The energy of ATP is used to alter the structure of nucleosomes and thus make the DNA more accessible
            - Chromatin Remodeling Complexes (*chromatin remodelers)* control accessibility to promotors and cis-acting regulatory elements
            - General mechanisms of remodeling:
                - Changes in nucleosome position: sliding nucleosomes around
                - Histone/Nucleosome eviction: eviction of nucleosome(s) to allow free DNA