Biology 212 General Genetics **Spring 2007**
Lecture 18: Gene Expression I

Reading: Chap. 8 pp. 282-291; pp. 302-305

Lecture Outline:
1. **Principles of gene expression**
2. **Transcription**
3. **Translation: elucidating the genetic code**

1. Gene Expression Principles

- Gene expression involves processes of transcription and translation, which result in the production of proteins
- The primary structure of proteins is a linear sequence of amino acids held together by peptide bonds
    - Peptide bonds link the carboxyl group of one amino acid to the amino group of the next amino acid
    - There are twenty naturally occurring amino acids, the building blocks of proteins
    - each amino acid has a unique side chain = R group
    - The linear sequence of amino acids in proteins is specified by the coding information in specific genes
        - Colinearity: the linear order of amino acids is encoded in a DNA base sequence
        - The base sequence in DNA specifies the base sequence in mRNA→ decoded in blocks of 3 nt → amino acid sequence

2. Transcription

- Transcription = production of messenger RNA (mRNA) complementary to the base sequence of specific genes
- mRNA differs from DNA in that it is single stranded, contains ribose sugar instead of deoxyribose and has the pyrimidine uracil in place of thymine

A. RNA Synthesis

- The nucleotide sequence in the transcribed mRNA is complementary to the base sequence in DNA
- RNA is copied from the template strand which is 3'-to-5' in the 5'-to-3' direction = antiparallel
- RNA synthesis does not require a primer and proceeds by the sequential addition of nucleotides to form an mRNA chain

B. Process of Transcription (Prokaryotes)
Three main phases:
Initiation
Elongation
Termination

Initiation:

- Promoter = nucleotide sequence 5' to the transcription start site
  - binding site of RNA polymerase initiation factor (sigma subunit, σ)
  - Promoter recognition by RNA polymerase is a prerequisite for transcription initiation
  - Many promoters contain a similar DNA sequence = TATAAT = "TATA" box, at -10
  - Another consensus promoter sequence is at -35 = TTGACA

Elongation:

- Growth of RNA chain 5' → 3' direction by sequential addition of nucleotides catalyzed by RNA polymerase

Termination:

- Transcription termination sites are often inverted repeat sequences which can form hairpin loops in RNA

C. Eukaryotic Gene Structure

- In many eukaryotic genes, the coding regions are interrupted by noncoding segments = "split genes"
- Coding regions = exons
- Noncoding regions = introns
- Primary transcript contains exons and introns; introns are subsequently removed = "splicing"

D. Eukaryotic Transcription

- Eukaryotic transcription involves the synthesis of RNA specified by DNA template strand to form a primary transcript
- **Protein-coding genes** are transcribed by **RNA polymerase II**
- separate RNA polymerases transcribe rRNA genes (RNA polymerase I) and tRNA genes (RNA polymerase III)
- Primary transcript is processed to form mRNA which is transported to the cytoplasm
  - The first processing step adds 7- methylguanosine to 5' end = "cap"
  - Splicing removes introns and links exons.
  - Additional processing involves the addition of a series of 150-200 adenines at the 3' end of the transcript = "poly A tail"

The processed transcript contains a 5' cap (7-methylguanosine), adjacent exons, and a poly A tail

3. Translation: Genetic evidence for the triplet code

A. Experiments of Francis Crick and Sidney Brenner 1961

T4: virus of E. coli
r: rapid lysis mutant; forms a different appearing plaque from wild type
rII: mutations in the rapid lysis gene II locus

i) induce mutations with proflavin

rII+ → rII-

- "intercalating agent"= sits between stacked base pairs on DNA
- when DNA is replicated in the presence of proflavin, tend to get insertions or deletions of one to a few nucleotides
- insertions or deletions of one or two nucleotides typically cause "frameshift" mutations that shift the reading frame, often a stop codon is encountered in the new reading frame

Analysis of frameshift mutations: Fig 8.25

+/- 1 nucleotide: get frameshifts
+/- 2 nucleotides: get frameshifts
+/- 3 nucleotides: often not harmful; reading frame restored

ii) back mutation (reversion)

rII- → rII+

theseoften result from a mutation that restores the reading frame

B. Biochemical evidence for the code: Work of Niremberg and Matthaei 1961 followed by Khorana

Use synthetic mRNAs (homopolymers and heteropolymers) in an in vitro protein synthesis system.

Niremberg and Matthaei: Poly U directs the synthesis of polypeptide containing only phe subunits; CCC encodes proline, AAA encodes lysine, GGG encodes glycine

Khorana: poly UC directs the synthesis of NH2…ser-leu-ser-leu…COOH