Gene Expression and Protein Synthesis
The Central Dogma

Figure 26-1 The central dogma of molecular biology:

- Information contained in DNA molecules is expressed in the structure of proteins.
- Gene expression is the turning on or activation of a gene.
**Transcription**: The process in which information encoded in a DNA molecule is copied into an mRNA molecule.

- Transcription takes place in the nucleus.
- Transcription starts when the DNA double helix begins to unwind near the gene to be transcribed.
- Only one strand of the DNA is transcribed.
- Ribonucleotides assemble along the unwound DNA strand in a complementary sequence.
- Enzymes called **polymerases (poly)** catalyze transcription: poly I for rRNA formation, poly II for mRNA formation, and poly III for tRNA formation.
Figure 26-2 Transcription of a gene. The information in one DNA strand is transcribed to a strand of RNA. The termination site is the locus of termination of transcription.
In eukaryotes. Three kinds of polymerases catalyze transcription.

- RNA polymerase I (pol I) catalyzes the formation of most of the rRNA.
- Pol II catalyzes mRNA formation
- Pol III catalyzes tRNA formation as well as one ribosomal subunit.
• Figure 26-3  The architecture of yeast RNA polymerase II. Transcription of DNA (helical structure) into RNA (red) is shown. The template strand of DNA (blue) and the coding strand (green) and also shown.
Transcription

- A eukaryotic gene has two parts:
  - A **structural gene** that is transcribed into RNA; the structural gene is made of exons and introns.
  - A **regulatory gene** that controls transcription; the regulatory gene is not transcribed but has control elements, one of which is the **promoter**.

A promoter is unique to each gene.

- There is always a sequence of bases on the DNA strand called an **initiation signal**.
- Promoters also contain **consensus sequences**, such as the **TATA box**, in which the two nucleotides T and A are repeated many times.
A TATA box lies approximately 26 base pairs upstream. All three RNA polymerases interact with their promoter regions via transcription factors that are binding proteins. After initiation, RNA polymerase zips up the complementary bases in a process called elongation. Elongation involves formation of phosphate ester bonds between each ribose and the next phosphate group. Elongation is in the 5’ —> 3’ direction. At the end of each gene is a termination sequence.
Transcription

• The RNA products of transcription are not necessarily functional RNAs.
  • They are made functional by **post-transcription modification**.
  • Transcribed mRNA is capped at both ends.
  • The 5′ end acquires a methylated guanine (7-mG cap).
  • The 3′ end acquires a polyA tail that may contain from 100 to 200 adenine residues.
  • Once the two ends are capped, the introns are spliced out.
  • tRNA is similarly trimmed, capped, and methylated.
  • Functional rRNA also undergoes post-transcription methylation.
Transcription

- Figure 26-4 Organization and transcription of a split eukaryote gene.
Role of RNA in Translation

• mRNA, rRNA, and tRNA all participate in translation.
• Protein synthesis takes place on ribosomes.
• A ribosome dissociates into a larger and a smaller body.
• In higher organisms, including humans, the larger body is called a 60S ribosome; the smaller body is called a 40S ribosome.
• The 5’ end of the mature mRNA is bonded to the 40S ribosome and this unit then joined to the 60S ribosome.
• Together the 40S and 60S ribosomes form a unit on which mRNA is stretched out.
• Triplets of bases on mRNA are called codons.
• The 20 amino acids are then brought to the mRNA-ribosome complex, each amino acid by its own particular tRNA.
• Each tRNA is specific for only one amino acid.
• Each cell carries at least 20 specific enzymes, each specific for one amino acid.
• Each enzyme recognizes only one tRNA.
• The enzyme bonds the activated amino acid to the 3’ terminal -OH group of the appropriate tRNA by an ester bond.
• At the opposite end of the tRNA molecule is a codon recognition site.
• The codon recognition site is a sequence of three bases called an anticodon.
• This triplet of bases aligns itself in a complementary fashion to the codon triplet on mRNA.
tRNA

- Figure 26-5
  The three-dimensional structure of tRNA.

Anticodon loop

CCA terminus

5’

3’
Assignments of triplets is based on several types of experiments.

- One of these used synthetic mRNA.
- If mRNA is polyU, polyPhe is formed; the triplet UUU, therefore, must code for Phe.
- If mRNA is poly ---ACACAC---, poly(Thr-His) is formed; ACA must code for Thr, and CAC for His.
- By 1967, the genetic code was broken
# The Genetic Code

<table>
<thead>
<tr>
<th>First Position (5' end)</th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
<th>Third Position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>Phe</td>
<td>UCU</td>
<td>UAU</td>
<td>UGU</td>
<td>Cys</td>
</tr>
<tr>
<td>UUC</td>
<td>Phe</td>
<td>UCC</td>
<td>UAC</td>
<td>UGC</td>
<td>Cys</td>
</tr>
<tr>
<td>UUA</td>
<td>Leu</td>
<td>UCA</td>
<td><strong>UAA</strong></td>
<td><strong>UGA</strong></td>
<td>Stop</td>
</tr>
<tr>
<td>UUG</td>
<td>Leu</td>
<td>UCG</td>
<td><strong>UAG</strong></td>
<td><strong>UGG</strong></td>
<td>Trp</td>
</tr>
<tr>
<td>CUU</td>
<td>Leu</td>
<td>CCU</td>
<td>CAU</td>
<td>CGU</td>
<td>Arg</td>
</tr>
<tr>
<td>CUC</td>
<td>Leu</td>
<td>CCC</td>
<td>CAC</td>
<td>CGC</td>
<td>Arg</td>
</tr>
<tr>
<td>CUA</td>
<td>Leu</td>
<td>CCA</td>
<td>CAA</td>
<td>CGA</td>
<td>Arg</td>
</tr>
<tr>
<td>CUG</td>
<td>Leu</td>
<td>CCG</td>
<td>CAG</td>
<td>CGG</td>
<td>Arg</td>
</tr>
<tr>
<td>AUU</td>
<td>Ile</td>
<td>ACU</td>
<td>AAU</td>
<td>AGU</td>
<td>Ser</td>
</tr>
<tr>
<td>AUC</td>
<td>Ile</td>
<td>ACC</td>
<td>AAC</td>
<td>AGC</td>
<td>Ser</td>
</tr>
<tr>
<td>AUA</td>
<td>Ile</td>
<td>ACA</td>
<td>AAA</td>
<td>AGA</td>
<td>Arg</td>
</tr>
<tr>
<td><strong>AUG</strong></td>
<td><strong>Met</strong></td>
<td>ACG</td>
<td>AAG</td>
<td>AGG</td>
<td>Arg</td>
</tr>
<tr>
<td>GGU</td>
<td>Val</td>
<td>GCU</td>
<td>GAU</td>
<td>GGU</td>
<td>Gly</td>
</tr>
<tr>
<td>GUC</td>
<td>Val</td>
<td>GCC</td>
<td>GAC</td>
<td>GGC</td>
<td>Gly</td>
</tr>
<tr>
<td>GUA</td>
<td>Val</td>
<td>GCA</td>
<td>GAA</td>
<td>GGA</td>
<td>Gly</td>
</tr>
<tr>
<td>GUG</td>
<td>Val</td>
<td>GCG</td>
<td>GAG</td>
<td>GGG</td>
<td>Gly</td>
</tr>
</tbody>
</table>

*AUG signals translation initiation as well as coding for Met*
Features of the Code

• All 64 codons have been assigned.
• 61 code for amino acids.
• 3 (UAA, UAG, and UGA) serve as termination signals.
• AUG also serves as an initiation signal.
• Only Trp and Met have one codon each.
• More than one triplet can code for the same amino acid; Leu, Ser, and Arg, for example, are each coded for by six triplets.
• The third base is irrelevant for Leu, Val, Ser, Pro, Thr, Ala, Gly, and Arg.
• It is said to be continuous and unpunctuated. There are no overlapping codons and no nucleotides interspersed.
• For the 15 amino acids coded for by 2, 3, or 4 triplets, it is only the third letter of the codon that varies. Gly, for example, is coded for by GGA, GGG, GGC, and GGU.

• The code is almost universal: it the same in viruses, prokaryotes, and eukaryotes; the only exceptions are some codons in mitochondria.
How is Protein Synthesized?

- Activation
- Initiation
- Elongation
- Termination
Protein Synthesis

- **Table 26-2 Molecular Components of Reactions at Four Stages of Protein Synthesis:**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Molecular Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation</td>
<td>Amino acids, ATP, tRNAs, aminoacyl-tRNA synthases</td>
</tr>
<tr>
<td>Initiation</td>
<td>fMet-tRNA$^{f\text{Met}}$, 30S ribosome, initiation factor proteins, mRNA with</td>
</tr>
<tr>
<td>Elongation</td>
<td>Shine-Dalgarno sequence, 50S ribosome, GTP</td>
</tr>
<tr>
<td>Termination</td>
<td>Releasing factors, GTP</td>
</tr>
</tbody>
</table>
End class 10/26/16 W
Amino Acid Activation

- Requires:
  - amino acids
  - tRNAs
  - aminoacyl-tRNA synthetases
  - ATP, Mg$^{2+}$
- Activation of an amino acid (formation of an amino acid–tRNA)

\[
\begin{align*}
\text{Adenosine} & -\text{O} & -\text{P} & -\text{O} & -\text{P} & -\text{O} & -\text{O} & + & \text{O} & -\text{C} & -\text{CH} & -\text{NH}_3^+ & \rightarrow \\
& \text{O} & -\text{O} & -\text{O} & -\text{O} & & & & & & & & \text{R} \\
\text{ATP} & & & & & & & & & & & & \text{An amino acid} \\
\text{Adenosine} & -\text{O} & -\text{P} & -\text{O} & -\text{C} & -\text{CH} & -\text{NH}_3^+ & + & \text{O} & -\text{P} & -\text{O} & -\text{P} & -\text{O} & -\text{O} & -\text{O} & -\text{O} & -\text{O} & & \text{An amino acid–AMP} & \text{Pyrophosphate}
\end{align*}
\]
The activated amino acid is bound to its own particular tRNA by an ester bond between the carboxyl group of the amino acid and the 3’-OH of the tRNA.
Amino Acid Activation

This two-stage reaction allows selectivity at two levels:

- **The amino acid**: The amino acid-AMP remains bound to the enzyme and binding of the correct amino acid is verified by an editing site on the tRNA synthetase.

- **tRNA**: There are specific binding sites on tRNAs that are recognized by aminoacyl-tRNA synthetases.

- This stage is very important and accuracy is vital. Once the amino acid is on its tRNA, there is no other opportunity to check for correct pairing. The anticodon of the tRNA will match up with its correct codon on the mRNA regardless of whether it is carrying the correct amino acid.
Chain Initiation

- Figure 26-6 Formation of the 30s Initiation complex.
- Step 2: The 50S ribosomal subunit is added forming the full complex.
Chain Initiation

- **Figure 26-6 cont’d Formation of an initiation complex.**
Figure 26-7
The steps of chain elongation.
Peptide Bond Formation

- Figure 26-9 Peptide bond formation in protein synthesis.
- Nucleophilic attack of -NH₂ on the peptidyl carbonyl
- Followed by collapse to give the new peptide bond.
Chain Termination

- Chain termination requires:
  - Termination codons (UAA, UAG, or UGA) of mRNA.
  - Releasing factors that cleave the polypeptide chain from the last tRNA and release the tRNA from the ribosome.
Gene Regulation

• **Gene regulation**: The various methods used by organisms to control which genes will be expressed and when.

• As the ribosome moves along the mRNA, it encounters a stop codon.

• Release factors and GTP bond to the A-site.

• The peptide is hydrolyzed from the tRNA.

• Finally, the entire complex dissociates, and the ribosome, mRNA, and other factors are recycled.
  
  • Some regulations operate at the **transcriptional level** (DNA ——> RNA)

  • Others operate at the **translational level** (mRNA ——> protein).
Transcriptional Level

- In eukaryotes, transcription is regulated by three elements: promoters, enhancers, and response elements.

- Promoters:
  - Located adjacent to the transcription site.
  - Are defined by an **initiator** and **conserved sequences** such as TATA or GC boxes.
  - Different **transcription factors** bind to different modules of the promoter.
  - Transcription factors allow the rate of synthesis of mRNA (and from there the target protein) to vary by a factor of up to a million.
Promoters

- Transcription factors find their targeted sites by twisting their protein chains so that a certain amino acid sequence is present at the surface.
- One such conformational twist is provided by **metal-binding fingers** (next screen).
- Two other prominent transcription factor conformations are the **helix-turn-helix** and the **leucine zipper**.
- Transcription factors also possess **repressors**, which reduce the rate of transcription.
Metal-Binding Fingers

- Figure 26-13 Cys$_2$His$_2$ zinc finger motifs. (a) The coordination between zinc and cysteine and histidine residues. (b) The secondary structure.
Figure 26-14 Zinc finger proteins follow the major groove of DNA.
Figure 26-15 Alternate splicing. A gene’s primary transcript can be edited in several different ways where splicing activity is indicated by dashed lines.

(a) Skipped exon

(b) Alternative 5’ splice sites

(c) Alternative 3’ splice sites

Resulting mRNA
Alternate Splicing

- Figure 26-15 cont’d

(d) Retained intron

(e) Mutually exclusive exon retention

Exon always spliced in  Exon alternatively spliced  Introns
Gene Regulation

- Control at the translational level to ensure quality control.
  - 1. The specificity of a tNRA for its unique amino acid.
  - 2. Recognition of the stop codon.
  - 3. Post-translational control.
    - (a) Removal of methionine.
    - (b) Chaperoning
    - (c) Degradation of misfolded proteins.
Mutations and Mutagens

- **Mutation**: An error in the copying of a sequence of bases.
  - It is estimated that, on average, there is one copying error for every $10^{10}$ bases.
  - Mutations can occur during replication.
  - Base errors can also occur during transcription in protein synthesis (a nonheritable error).
  - Consider the mRNA codons for Val, which are CAT, CAC, CAG, and CAA.
  - If the original codon is CAT, it may be transcribed onto mRNA as GUC which codes for Val.
  - Other errors in replication may lead to a change in protein structure and be very harmful.
Mutations and Mutagens

- **Mutagen**: a chemical that causes a base change or mutation in DNA.
- Many changes in base sequence caused by radiation and mutagens do not become mutations because cells have repair mechanisms called nucleotide excision repair (NER).
  - NER can prevent mutations by cutting out damaged areas and resynthesizing the proper sequence.
- Not all mutations are harmful.
  - Certain ones may be beneficial because they enhance the survival rate of the species.
Recombinant DNA

- **Recombinant DNA**: DNA from two sources that have been combined into one molecule.

- One example of the technique begins with plasmids found in the cells of *Escherichia coli*.
  - **Plasmid**: A small, circular, double-stranded DNA molecule of bacterial origin.
  - A class of enzymes called **restriction endonucleases** cleave DNA at specific locations.
  - One, for example, may be specific for cleavage of the bond between A-G in the sequence -CTTAAG-.
Recombinant DNA

• In this example “B” stands for bacterial gene, and “H” for human gene.

\[ \text{B} \quad \text{GAATTTC} \quad \text{B} \quad \text{restriction enzyme} \quad \text{B} \quad \text{G} \quad \text{AATTC} \quad \text{B} \]

\[ \text{B} \quad \text{CTTAAG} \quad \text{B} \quad \text{B} \quad \text{CTTAAC} \quad \text{G} \quad \text{B} \]

• The DNA is now double-stranded with two “sticky ends”, each with free bases that can pair with a complementary section of DNA.

• Next, we cut a human gene (H) with the same restriction endonuclease; for example, the gene for human insulin.

\[ \text{H} \quad \text{GAATTTC} \quad \text{H} \quad \text{restriction enzyme} \quad \text{H} \quad \text{G} \quad \text{AATTC} \quad \text{H} \]

\[ \text{H} \quad \text{CTTAAG} \quad \text{H} \quad \text{H} \quad \text{CTTAAC} \quad \text{G} \quad \text{H} \]
Recombinant DNA

- The human gene is now spliced into the plasmid by the enzyme **DNA ligase**.

\[
\begin{align*}
H\text{---G} & \quad + \quad \text{AATTC---B} \\
H\text{---CTTAA} & \quad \text{DNA ligase} \quad \rightarrow \\
H\text{---GAATTC---B} & \quad H\text{---CTTAAG---B}
\end{align*}
\]

- Splicing takes place at both ends of the human gene and the plasmid is once again circular.
- The modified plasmid is then put back into the bacterial cell where it replicates naturally every time the cell divides.
- These cells now manufacture the human protein, in our example human insulin, by transcription and translation.
Recombinant DNA

- Figure 26-17 The recombinant DNA technique used to turn a bacterium into an insulin “factory”.
Recombinant DNA

- **Figure 26-17 Continued**

DNA with insulin gene inserted

Bacterium with inserted recombinant DNA
Gene Therapy

- **Gene therapy** is a technique whereby a missing gene is replaced by a viral vector.
  - In *ex vivo* gene therapy, cells are removed from a patient, given the missing gene, and then the cells are given back to the patient.
  - In *in vivo* gene therapy, the patient is given the virus directly.
Gene Therapy

- Figure 26-18 Gene therapy via retroviruses. The Maloney murine leukemia virus (MMLV) is used for ex vivo gene therapy.
Gene Therapy

- Figure 16-18
- Cont’d Gene therapy via retroviruses.
End
Chapter 26