Chapter 9
Gene Expression:
From Genes to Proteins
DNA Codes for Protein

The information required to produce proteins is encoded in the nucleotide sequence of DNA.
Relationship Between Genes, Proteins, and Phenotype

• Archibald Garrod (early 1900s)
  – Inborn errors of metabolism
• Alkaptonuria altered metabolism of alkapton (homogentisic acid)
• Urine turns black
• Using pedigree analysis he determined it is an autosomal recessive trait
Metabolic Pathways

• Chains of chemical reactions

• In a pathway, an enzyme is required to convert one compound into another

• Each reaction is controlled by a different enzyme
One Gene, One Enzyme

• Beadle and Tatum (1930s)

• Experiments on *Neurospora*, a common bread mold

Fig. 9.2
DNA Stores Genetic Information

- Phenotype is the result of protein function
- When the function is altered or absent, the result is a mutant phenotype
- Proteins are products of genes
- Genes are made of DNA
- Changes in DNA may change protein functions
Genetic Code

- Information is stored in the sequence of nucleotides in the DNA
### The Genetic Code

![Fig. 9.3](image)

<table>
<thead>
<tr>
<th>First letter</th>
<th>Second letter</th>
<th>Third letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>U</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>U</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>U</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>UU</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>UU</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>UU</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>UU</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>UUC</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>UUC</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>UUC</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>UUG</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>UUG</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>UUG</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>CU</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>CU</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CU</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>CU</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>CUC</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CUC</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>CUC</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>CUDA</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CUDA</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>CUDA</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>CUG</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CUG</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>CUG</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>AU</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>AU</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>AU</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>AU</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>AUC</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>AUC</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>AUC</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>AUA</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>AUA</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>AUA</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>AUG</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>AUG</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>AUG</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>AG</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>AG</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>AG</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>AG</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>AGG</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>AGG</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>AGG</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>GU</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>GU</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>GU</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>GU</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>GUC</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>GUC</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>GUC</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>GUA</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>GUA</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>GUA</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>GUG</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>GUG</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>GUG</td>
<td>G</td>
<td>G</td>
</tr>
</tbody>
</table>

- **UUU, UUC, UUA, UUG**: Phe (F)
- **UCU, UCC, UCA, UCG**: Ser (S)
- **UAU, UAC**: Tyr (Y)
- **UGU, UGC, UGA, UGG**: Cys (C), Stop
- **AUU, AUC, AUA**: Ile (I)
- **ACG**:thr (T)
- **GCU, GCC, GCA, GCG**: Ala (A)
- **GAU, GAC**: Asp (D)
- **GGU, GGC, GGA, GGG**: Gly (G)

**Legend:**
- **Pink** = Chain termination codon (stop)
- **Yellow** = Initiation codon

---

Chapter 9 *Human Heredity* by Michael Cummings ©2006 Brooks/Cole-Thomson Learning
About the genetic code:

1. Triplet code: 3 bases encode an amino acid.

2. Continuous code.

3. Non-overlapping open reading frames

AAGCTTCAATTC  AAGCTTCAATTC
Lys-Leu-Gln-Phe  Ser-Phe-Asn

AAGCTTCAATTC
Ala-Ser-Ile
About the genetic code cont’d.

4. Code is universal.

5. Code is degenerate.

Exceptions: AUG - Met
UGG - Trp

6. Wobble hypothesis.

7. Nonsense or STOP codons

61 sense codons
3 nonsense codons
<table>
<thead>
<tr>
<th><strong>Skin cell</strong></th>
<th><strong>Liver cell</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene 1 $\rightarrow$</td>
<td>Protein 1</td>
</tr>
<tr>
<td>Gene 2 $\rightarrow$</td>
<td></td>
</tr>
<tr>
<td>Gene 3 $\rightarrow$</td>
<td>Protein 3</td>
</tr>
<tr>
<td>Gene 4 $\rightarrow$</td>
<td>Protein 4</td>
</tr>
<tr>
<td>Gene 5 $\rightarrow$</td>
<td></td>
</tr>
<tr>
<td>Gene 1 $\rightarrow$</td>
<td>Protein 2</td>
</tr>
<tr>
<td>Gene 2 $\rightarrow$</td>
<td></td>
</tr>
<tr>
<td>Gene 3 $\rightarrow$</td>
<td>Protein 4</td>
</tr>
<tr>
<td>Gene 4 $\rightarrow$</td>
<td></td>
</tr>
<tr>
<td>Gene 5 $\rightarrow$</td>
<td>Protein 5</td>
</tr>
</tbody>
</table>
RNA

- **Messenger (mRNA)** – single-stranded, complementary copy of DNA

- **Transfer RNA (tRNA)** contains a binding site for mRNA codon and a binding site for a specific amino acid

- **Ribosomal RNA (rRNA)** – component of the ribosome
Protein Synthesis

• Linear sequence of nucleotides of DNA is transferred to a linear sequence of amino acids of the protein

Two steps

• Transcription
  – DNA to mRNA

• Translation
  – mRNA to protein
The Flow of Genetic Information

Central dogma

Chapter 9  *Human Heredity* by Michael Cummings ©2006 Brooks/Cole-Thomson Learning
Transcription

- Occurs in the nucleus
- One strand is used as a template to produce mRNA
- Three stages
  - Initiation
  - Elongation
  - Termination
Transcription

(a) RNA polymerase initiates transcription at a promoter region in the DNA. It will recognize the base sequence located downstream from that site as a template for linking together the nucleotides adenine, cytosine, guanine, and uracil into a strand of RNA.

(b) All through transcription, the DNA double helix becomes unwound in front of the RNA polymerase. Short lengths of the newly forming RNA strand briefly wind up with its DNA template strand. New stretches of RNA unwind from the template (and the two DNA strands wind up again).

(c) What happened at the assembly site? The RNA polymerase catalyzed the base-pairing of RNA nucleotides, one after another, with exposed bases on the DNA template strand.

(d) At the end of the gene region, the last stretch of the new mRNA transcript is unwound and released from the DNA.
Transcription: Initiation

- RNA polymerase binds to the promoter region
- DNA unwinds
Transcription: Elongation

- 30–50 nucleotides/second

Fig. 9.5c
Transcription: Termination

- RNA polymerase reaches the **terminator** and mRNA is released

Fig. 9.5d
Product of Transcription: Pre-mRNA

- Large pre-mRNA molecules are composed of exons and introns
- **Exons** are nucleotide sequences that are transcribed and translated
- **Introns** are nucleotide sequences that are transcribed but not translated
  - 0–75 per gene
  - Size ranges from 100–100,000 nucleotides
Organization of an Eukaryotic Gene

Fig. 9.6
Processing Pre-mRNA

- Introns are removed
- Exons are spliced together
- Cap is added to the 5’ end
  - The cap helps attach mRNA to the ribosome
- Poly-A tail is added to the 3’ end
- The mRNA is transported out of the nucleus to the ribosome for translation
Processing Pre-mRNA

Unit of transcription in DNA strand

3’ Exon Intron Exon Intron Exon 5’

Transcription into pre-mRNA

Cap

5’

Snipped out Snipped out

Poly-A tail

3’

Mature mRNA transcript

Fig. 9.7
Nobel Prize in Medicine, 1993

Introns

Phil Sharp

Rich Roberts
Proteins

- Proteins are composed of amino acids joined together by peptide bonds
- 20 different amino acids can be used to produce proteins
- Amino acids have a Carboxyl (acid) group, an amino group, and an unique R group

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>ala</td>
</tr>
<tr>
<td>Arginine</td>
<td>arg</td>
</tr>
<tr>
<td>Asparagine</td>
<td>asn</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>asp</td>
</tr>
<tr>
<td>Cysteine</td>
<td>cys</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>glu</td>
</tr>
<tr>
<td>Glutamine</td>
<td>gln</td>
</tr>
<tr>
<td>Glycine</td>
<td>gly</td>
</tr>
<tr>
<td>Histidine</td>
<td>his</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>ile</td>
</tr>
<tr>
<td>Leucine</td>
<td>leu</td>
</tr>
<tr>
<td>Lysine</td>
<td>lys</td>
</tr>
<tr>
<td>Methionine</td>
<td>met</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>phe</td>
</tr>
<tr>
<td>Proline</td>
<td>pro</td>
</tr>
<tr>
<td>Serine</td>
<td>ser</td>
</tr>
<tr>
<td>Threonine</td>
<td>thr</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>trp</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>tyr</td>
</tr>
<tr>
<td>Valine</td>
<td>val</td>
</tr>
</tbody>
</table>
Amino Acids Are the Subunits of Proteins

\[ \alpha\text{-carbon atom} \quad R \text{ group (differs in each amino acid)} \]

\[ \begin{array}{c}
\text{Amino group} \\
\text{Carboxyl group}
\end{array} \]

Structures common to all amino acids

Peter J. Russell, iGenetics: Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

Amino acids to polypeptides

Polypeptide formation

AA1 + AA2 + AA3

H₂O → Peptide bonds

N-terminus

C-terminus
Primary sequence - sequence of amino acids in a protein.

Secondary structure - depends on primary sequence. Beta sheets and alpha helices. Can be usually predicted from sequence.

Tertiary structure - 3-D structure of how the protein folds. Cannot be always predicted. Domains - can bring residues far apart in primary sequence in contact with each other.

Quaternary structure - multimers. Many subunits each of which is a polypeptide.
Tertiary structure

Quaternary structure

Hemoglobin
Diseases arising from defects in protein folding

• May cause changes in protein folding
• Examples
  – A form of Alzheimer disease
  – MPS VI
  – Cystic fibrosis
  – Prion diseases
Prions

Stanley Pruisiner won the Nobel Prize in 1997 for his work with prions and their role in disease
Ribosomes

- The site of protein synthesis
- May be free in the cytoplasm or attached to endoplasmic reticulum
- Contains two binding sites
  - $A$ (Amino) and $P$ (peptide) site
- Composed of
  - Two subunits
  - rRNA and protein

Fig. 9.9
Transfer RNA (tRNA)

- Adapters that bond to an amino acid and recognize specific mRNA codons
- Small, single-stranded RNA that folds back, forming a cloverleaf shape
HIV infection

Retrovirus

RNA genome

Cytoplasm

Host cell

RNA

Reverse transcription

Viral DNA

Nucleus

Chromosome
Nobel Prize in Medicine, 1975

Retroviral mechanisms

Renato Dulbecco    David Baltimore    Howard Temin
Nucleotide analogs used as anti-AIDS drugs

3’-Azido-2’,3’-dideoxythymidine (AZT)  2’,3’-Dideoxyinosine (DDI)
Translation

- Converts mRNA sequence to amino acid sequence
- Occurs within the ribosomes
- Requires various factors: enzymes, amino acids, energy, mRNA, rRNA and tRNA
- Three steps
  - Initiation
  - Elongation
  - Termination
Steps in Translation: Initiation

- Small ribosomal subunit binds to mRNA
- tRNA carrying methionine binds to the start codon (AUG) with in P site
- Forms initiation complex
- Large ribosomal subunit binds to small subunit
Translation, part I

Initiation

Eukaryotic

5'-untranslated leader
Initiating codon
5'-untranslated leader
Initiating codon

5' Methylated mRNA cap
Small ribosomal subunit (40s)

Large ribosomal subunit (60S)
P site
A site

A mRNA

5' → 3' → 5'

Scanning
Met
Translation: part 2

Elongation

Peptidyl transferase

Direction of ribosome movement
Translation : part 2

Elongation

Peptidyl transferase

Growing polypeptide

Direction of ribosome movement
Steps in Translation: Elongation

- tRNA for the second amino acid binds to the mRNA within the second ribosomal binding site (A site)
- Peptide bonds forms between methionine and second amino acid
- Ribosome moves down mRNA
- tRNA brings in the third amino acid into the A site
- Peptide bond forms
Termination
Steps in Translation: Termination

- Elongation continues until the ribosomes reaches a stop codon
- No tRNA binds
- Synthesis is completed and mRNA, tRNA, are released from the ribosome
- Polypeptide is folded into 3-dimensional shape
- Many antibiotics interfere with steps in protein synthesis
Mechanism of antibiotic action

**Specifically target protein translation in bacteria.**

Since prokaryotes and eukaryotes use different proteins for translation, antibiotics can specifically affect bacterial, but not cellular translation.
Puromycin prevents elongation of polypeptide. Causes premature termination.
Streptomycin

- Binds to 30s subunit and prevents initiation of protein synthesis

Tetracycline

- Prevents elongation by preventing charged tRNA from binding to A site.

Chloramphenicol

- Binds 50s subunit and blocks peptidyl transferase reaction.

Erythromycin

- Prevents translocation
Proteome

• Set of proteins present in a cell under specific environmental conditions
• It is estimated that humans can make over 100,000 proteins
• This is greater than the number of genes in the human genome

• How is this possible?