Chapter 17: From Gene to Protein

1. Overview of Gene Expression
2. Transcription
3. The Genetic Code
4. Translation
5. Mutations
1. Overview of Gene Expression

Chapter Reading – pp. 334-337
How are Genes related to DNA?

Genes are segments of DNA that code for a particular protein (or RNA molecule)

• the human genome contains ~3 billion base pairs (bps) and ~25,000 genes

• most genes **encode** proteins
  • when we talk about “genes” we will focus on those that express proteins

• the gene products of some genes are RNA molecules that play a variety of roles in cells
### EXPERIMENT

Growth: Wild-type cells growing and dividing

No growth: Mutant cells cannot grow and divide

### RESULTS

#### Classes of *Neurospora crassa*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Wild type</th>
<th>Class I mutants</th>
<th>Class II mutants</th>
<th>Class III mutants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal medium (MM) (control)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM + ornithine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM + citrulline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM + arginine (control)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Summary of results**
- Can grow with or without any supplements
- Can grow on ornithine, citrulline, or arginine
- Can grow only on citrulline or arginine
- Require arginine to grow

### CONCLUSION

- **these classic experiments led to the “one gene-one enzyme” hypothesis**
- **we now know that genes code for more than just enzymes**

**Beadle & Tatum, 1941**

**Srb & Horowitz, 1944**
Gene Expression

The expression of most genes involves two distinct processes:

1) **Transcription** of a gene into RNA
   - RNA is a nucleic acid very similar to DNA (RNA uses “U” instead of “T”)
   - this is essentially creating a “photocopy” of the gene
   - occurs in the nucleus

2) **Translation** of the RNA transcript into protein
   - accomplished by **ribosomes**, in the cytoplasm
Prokaryotes vs Eukaryotes

(a) Bacterial cell  
(b) Eukaryotic cell
• gene expression ends at transcription for genes encoding RNA gene products
Comparison of DNA & RNA

RNA
- sugar = Ribose
- single-stranded
- A, C, G & U (uracil)

DNA
- sugar = Deoxyribose
- double-stranded
- A, C, G & T (thymine)
2. Transcription

Chapter Reading – pp. 340-345
Stages of Transcription

INITIATION
RNA polymerase binds to the promoter of a gene, separates DNA

ELONGATION
RNA polymerase makes RNA complementary to the template strand

TERMINATION
RNA polymerase stops transcribing when end of gene is reached.
Initiation of Eukaryotic Transcription

Eukaryotic promoters contain a short sequence called a **TATA box**.

With the help of an array of transcription factors, RNA polymerase binds to the promoter and starts transcription.
In self-termination, the transcription of DNA terminator sequences cause the RNA to fold, loosening the grip of RNA polymerase on the DNA.

In enzyme-dependent termination, a termination enzyme pushes between RNA polymerase and the DNA, releasing the polymerase.

- triggered by stem/loop structure in RNA or termination factors such as the Rho protein (prokaryotes)
Primary RNA Transcripts

Newly made RNA molecules in eukaryotes are called primary transcripts because they need to be processed before they can carry out their function:

mRNA 1\textsuperscript{o} transcript

- Protein-coding segment
- Polyadenylation signal

Messenger (mRNA) primary transcripts require the following modifications:

- a 5’ cap
- a poly-A tail
- splicing out of introns
RNA Splicing

The coding region of the primary RNA transcript contains intervening sequences called **introns** that need to be removed or “spliced out”.

The regions that are retained are called **exons** which after splicing form a continuous coding region.
How is Splicing Carried Out?

Spliceosomes are structures made of protein and snRNA

snRNA in spliceosome base pairs with ends of intron sequences resulting in their being “spliced out”
Exons tend to code for distinct substructures called **domains** in proteins.

**Exon shuffling** is thought to be an evolutionary process by which new genes are made.
Various Roles of RNA Transcripts

1) messenger RNA (mRNA)
   - RNA copy of a gene that encodes a polypeptide

2) ribosomal RNA (rRNA)
   - RNA that is a structural component of ribosomes

3) transfer RNA (tRNA)
   - delivery of “correct” amino acids to ribosomes during translation

For some genes, the end-product is the RNA itself (rRNA, tRNA)
3. The Genetic Code

Chapter Reading – pp. 337-340
How do Genes code for Proteins?

Recall that proteins are linear polymers made of 20 different amino acids.

Genes need simply to encode the identity of each amino acid in a given protein!

- i.e., genes must be capable of encoding 20 different amino acids and their order in a protein

- although DNA contains only 4 different nucleotides, this is more than sufficient to specify 20 different amino acids...
Basis of the Genetic Code

Each amino acid in a protein is specified by 3 nucleotide sequences called **codons**

- each of the 20 amino acids is coded for by a unique set of codons:
  - e.g. \( \text{ATG} = \text{methionine (start codon)} \)
  - \( \text{GGN} = \text{glycine} \)
  - \( \text{CAA or CAG} = \text{glutamine} \)

- there are 64 possible “codon” triplets (\(4 \times 4 \times 4\))

- more than enough to encode 20 amino acids and the signal to “stop” or end the protein (TGA, TAA or TAG)
The Genetic Code

If the DNA sequence is:
5’-CATGCCTGGGCAATAG-3’
3’-GTACGGGACCCGTTATC-5’

(transcription)

The mRNA transcript is:
5’-CAUGCCUGG-GCAAUAG-3’

(translation)

The polypeptide is:
*Met-Pro-Gly-Gln (stop)

all proteins begin w/Met

<table>
<thead>
<tr>
<th>First mRNA base (5’ end of codon)</th>
<th>Second mRNA base</th>
<th>Third mRNA base (3’ end of codon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU UUC UUA UUG</td>
<td>Phe</td>
<td>UCU UCC UCA UCG</td>
</tr>
<tr>
<td>UUU UUC UUA UUG</td>
<td>Leu</td>
<td>CCU CCC CCA CCG</td>
</tr>
<tr>
<td>AUU AUC AUA AUG</td>
<td>Ile</td>
<td>ACU ACC ACA ACG</td>
</tr>
<tr>
<td>GUU GUC GUA GUG</td>
<td>Val</td>
<td>GCU GCC GCA GCG</td>
</tr>
<tr>
<td>UAU UAC UAA AAG</td>
<td>Tyr</td>
<td>CAU CAC CAA CAG</td>
</tr>
<tr>
<td>UAU UAC UAA AAG</td>
<td>UAA Stop</td>
<td>CGU CGC CGA CGG</td>
</tr>
<tr>
<td>UUG CUG</td>
<td>UAG Stop</td>
<td>UGG Trp</td>
</tr>
<tr>
<td>CUA CUG</td>
<td>Met or start</td>
<td></td>
</tr>
<tr>
<td>GUU GUC GUA GUG</td>
<td>Ala</td>
<td>GAU GAC GAA GAG</td>
</tr>
<tr>
<td>GUU GUC GUA GUG</td>
<td>Asp</td>
<td>GGU GGC GGA GGG</td>
</tr>
<tr>
<td>UGU GUC</td>
<td>Cys</td>
<td></td>
</tr>
<tr>
<td>UGA</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>UAA Stop</td>
<td>His</td>
<td></td>
</tr>
<tr>
<td>UGA</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td>UAA Stop</td>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td>UAG Stop</td>
<td>Met or start</td>
<td></td>
</tr>
</tbody>
</table>

© 2011 Pearson Education, Inc.
The Genetic Code is Universal

Genes from one species can be expressed in another!

(a) Tobacco plant expressing a firefly gene

(b) Pig expressing a jellyfish gene
4. Translation

Chapter Reading – pp. 83, 345-354
Overview of Translation

Ribosomes facilitate the production of polypeptides by:

1) matching codons in mRNA with complementary anticodons in tRNA

2) catalyzing peptide bonds between amino acids carried by tRNAs
Ribosome Structure

(a) Computer model of functioning ribosome

(b) Schematic model showing binding sites

(c) Schematic model with mRNA and tRNA
Transfer RNA (tRNA)

- Amino acid attachment site
- tRNA anticodon will base pair with a complementary codon in mRNA
- Two-dimensional structure
- Three-dimensional structure
- Symbol used in this book

Anticodon

- Hydrogen bonds
- Amino acid attachment site

(a) Two-dimensional structure
(b) Three-dimensional structure
(c) Symbol used in this book
Aminoacyl-tRNA Synthetases

Each tRNA is loaded with the correct amino acid due to the action of these enzymes.
Initiation of Translation

- small ribosomal subunit aligns with start codon of mRNA
- initiator tRNA\textsubscript{met} and large subunit then join the complex
Amino end of polypeptide

Ribosome ready for next aminoacyl tRNA

Elongation Cycle of Translation

E

mRNA

5’

3’

P site

A site

GTP

GDP + P_i

GTP

GDP + P_i
When a stop codon is reached there is no tRNA with a complementary anticodon.

Instead a **release factor** fits in the A site and catalyzes the dissociation of all components.
Multiple Ribosomes translate the same mRNA

Referred to as polyribosomes.
Chaperonins are large protein structures that assist newly made polypeptides to ensure they fold properly.

- capture improperly folded proteins inside its cavity, forcing them to unfold and hopefully refold correctly
Proteins destined for the “secretory pathway” have a **signal peptide** at the N-terminus that targets the protein to the ER lumen.
Gene Expression in Prokaryotes

Transcription and translation are not segregated in prokaryotes.

Since prokaryotic RNA transcripts don’t need to be processed, translation can begin before transcription is finished!
Summary of Gene Expression
5. **Mutation**

Chapter Reading – pp. 355-357
Mutations

A mutation is *any* change in DNA sequence:

- change of one nucleotide to another
- insertion or deletion of nucleotides or DNA fragments
- inversion or recombination of DNA fragments

What causes mutations?

- errors in DNA replication or DNA repair
- chemical mutagenesis
- high energy electromagnetic radiation
  - UV light, X-rays, gamma rays
Sickle-cell Anemia

Wild-type hemoglobin DNA

3' C T T T
5' G A A A

Mutant hemoglobin DNA

3' C A T T
5' G T A A

mRNA

5' G A A A
3'

5' G U A A
3'

Normal hemoglobin

Glu

Sickle-cell hemoglobin

Val

- a single nucleotide change causes a single amino acid change resulting in a malformed protein
Types of Mutations

**Silent mutations:**
- have no effect on amino acid specification

**Missense mutations:**
- result in the change of a single amino acid

**Nonsense mutations:**
- convert a codon specifying an amino acid to a stop codon
  - results in premature *truncation* of a protein

**Insertion/deletion mutations:**
- cause a shift in the reading frame of the gene
  - all codons downstream of insertion/deletion will be incorrect
Silent Mutations

Wild type

DNA template strand

3’ T A C T T C A A A A C C G A T T T
5’ A T G A A G T T T T G G C T A A

mRNA5’

A U G A A G U U U U G G C U A A

Protein

Met – Lys – Phe – Gly

Amino end

Carboxyl end

Stop

(a) Nucleotide-pair substitution: silent

3’ T A C T T C A A A A C C A A T T
5’ A T G A A G T T T T G G T T A A

A instead of G

5’ A U G A A G U U U U G G U U A A

U instead of C

Met – Lys – Phe – Gly

Stop

© 2011 Pearson Education, Inc.
Missense Mutations

Wild type

DNA template strand
3' T A C T T C A A A A C C G A T T T
5' A T G A A G T T T T G G C T A A

mRNA
5' A U G A A G U U U U G G C U A A A
3' Met Lys Phe Gly

Protein
Amino end
Stop Carboxyl end

(a) Nucleotide-pair substitution: missense

T instead of C
3' T A C T T T C A A A A T C G A T T T
5' A T G A A G T T T T A G C T A A A

A instead of G
5' A U G A A G U U U U A G C U A A A
3' Met Lys Phe Ser Stop

© 2011 Pearson Education, Inc.
Nonsense Mutations

Wild type

DNA template strand
3' T A C T T C A A A A C C G A T T T
5' A T G A A G T T T T G G C T A A

mRNA
5' A U G A A G U U U U G G C U A A A
3'

Protein
Met Lys Phe Gly

Amino end
Stop Carboxyl end

(a) Nucleotide-pair substitution: nonsense

A instead of T
3' T A C A T C A A A A C C G A T T T
5' A T G A G T T T T G G C T A A

T instead of C
5' A U G A A G U U U U G G C U A A

U instead of A
5' A U G U A G U U U U G G C U A A

Stop
Frameshift – Insertion/Deletion

Wild type

DNA template strand 3’ T A C T T C A A A A C C G A T T T 5’
5’ A T G A A G T T T T G G C T A A A 3’

mRNA 5’ A U G A A G U U U U G G G G C U A A 3’

Protein

Amino end

Met  Lys  Phe  Gly

Stop  Carboxyl end

(b) Nucleotide-pair insertion or deletion: frameshift causing immediate nonsense

Extra A

3’ T A C A T T C A A A A C G G A T T T 5’
5’ A T G T A A A G T T T G G C T A A A 3’

Extra U

5’ A U G U A A A G U U U U G G G C U A A 3’

Met  Stop

1 nucleotide-pair insertion

Can cause a premature stop codon...
Frameshift – Insertion/Deletion

Wild type

DNA template strand
3’ T A C T T C A A A C C G A T T T
5’ A T G A A G T T T G G C T A A

mRNA
5’ A U G A A G U U U G G G C U A A
3’

Protein
Met Lys Phe Gly

Amino end
Stop Carboxyl end

(b) Nucleotide-pair insertion or deletion: frameshift causing extensive missense

…or extended shift in reading frame…

1 nucleotide-pair deletion
Frameshift – Insertion/Deletion

Wild type

DNA template strand

3’ T A C T T C A A A A C C G A T T 5’

5’ A T G A A G T T T G G C T A A 3’

mRNA

5’ A U G A A G U U U G G C U A A 3’

Protein

Met — Lys — Phe — Gly

Amino end — Stop — Carboxyl end

(b) Nucleotide-pair insertion or deletion: no frameshift, but one amino acid missing

…or the addition/deletion of a single amino acid.

3 nucleotide-pair deletion

Must be a multiple of 3!
Key Terms for Chapter 17

• primary transcript, mRNA, tRNA, rRNA, snRNA

• transcription, promoter, TATA box, RNA polymerase

• 5’ cap, poly-A tail, intron, exon, splicing, spliceosome

• rho protein, stem-loop, codon, anti-codon, translation

• aminoacyl tRNA synthetase, polyribosome, signal peptide, release factor, chaperonin

• mutation: substitution, deletion, insertion

• silent, missense, nonsense mutations

• reading frame, frameshift

Relevant Chapter Questions 1-9, 11