RNA and Protein Synthesis

DNA → RNA → Protein

Ch. 13.1
What do genes code for?

- **How does DNA code for cells & bodies?**
  - how are cells and bodies made from the instructions in DNA
The “Central Dogma”

- Flow of genetic information in a cell
  - How do we move information from DNA to proteins?

DNA gets all the glory, but proteins do all the work!
Beadle & Tatum

1941 | 1958

one gene : one enzyme hypothesis

George Beadle
Edward Tatum

"for their discovery that genes act by regulating definite chemical events"
**Beadle & Tatum**

Wild-type *Neurospora*

Minimal medium

Asexual spores

Select one of the spores

Test on **minimal medium** to confirm mutation

**Negative control**

Grow on **complete medium**

**Positive control**

Minimal media supplemented only with:
- Arginine
- Riboflavin
- Thiamine
- Niacin
- Folic acid
- Pantothenic acid
- Inositol
- Phosphoric acid
- Choline
- Biotin
- Pteroyl
- Pyridoxine
- Nucleic acid
- Pyridoxine
- Phosphoric acid
- Choline
- Niacin
- Folic acid
- Riboflavin
- Thiamine

**Mutations** identified with:
- Pyridoxine
- Choline
- Nucleic acid
- Folic acid
- Arginine
- Riboflavin
- Thiamine
- Niacin

X rays or ultraviolet light create mutations

Growth on complete medium negative control experimentals amino acid supplements
From gene to protein

DNA → mRNA → protein

transcription

nucleus → cytoplasm

trait
Transcription

From DNA nucleic acid language to RNA nucleic acid language
RNA

• ribose sugar
• N-bases
  – uracil instead of thymine
  – U : A
  – C : G
• single stranded
• lots of RNAs
  – mRNA, tRNA, rRNA, siRNA…

DNA \rightarrow \text{transcription} \rightarrow RNA
Transcription

• Making mRNA
  – transcribed DNA strand = template strand
  – untranscribed DNA strand = coding strand
• same sequence as RNA
  – synthesis of complementary RNA strand
• enzyme
• RNA polymerase

![Diagram of transcription process]

DNA → \text{coding strand} → \text{template strand} → mRNA

build RNA 5′→3′
RNA polymerases

- 3 RNA polymerase enzymes
  - RNA polymerase 1
    - only transcribes rRNA genes
    - makes ribosomes
  - RNA polymerase 2
    - transcribes genes into mRNA
  - RNA polymerase 3
    - only transcribes tRNA genes
  - each has a specific promoter sequence it recognizes
Which gene is read?

- **Promoter region**
  - binding site before beginning of gene
  - **TATA box binding site**
  - binding site for RNA polymerase & transcription factors

- **Enhancer region**
  - binding site far upstream of gene
  - turns transcription on HIGH
Matching bases of DNA & RNA

- Match RNA bases to DNA bases on one of the DNA strands
Eukaryotic genes have junk!

- Eukaryotic genes are not continuous
  - **exons** = the real gene
    - expressed / coding DNA
  - **introns** = the junk
    - In between sequence

**intron** = noncoding (inbetween) sequence

**exon** = coding (expressed) sequence
mRNA splicing

- Post-transcriptional processing
  - eukaryotic mRNA needs work after transcription
  - **primary transcript** = pre-mRNA
  - mRNA splicing
    - edit out introns
  - **make mature mRNA transcript**

- intron = noncoding (inbetween) sequence
  - eukaryotic DNA
  - primary mRNA transcript
  - mature mRNA transcript
  - spliced mRNA

- ~10,000 bases
- ~1,000 bases
Splicing must be accurate

- No room for mistakes!
  - a single base added or lost throws off the **reading frame**

```
AUGC|GGCT|ATGG|GUCC|GAUA|AAGGGG|CCAU
AUGC|GGGU|CCGAUA|AAGGGG|CCAU
AUG|CGG|UCC|GAU|AAG|GGC|CAU
Met|Arg|Ser|Asp|Lys|Gly|His
```

```
AUGC|GGCT|ATGG|GUCC|GAUA|AAGGGG|CCAU
AUGC|GGGU|CCGAUA|AAGGGG|CCAU
AUG|CGG|GUC|CGA|UAA|GGG|CCA|U
Met|Arg|Val|Arg|STOP|
```
RNA splicing enzymes

- **snRNPs**
  - small nuclear RNA
  - proteins

- **Spliceosome**
  - several snRNPs
  - recognize splice site sequence
    - cut & paste gene

[Diagram of RNA splicing process]

- snRNPs
- snRNA
- snRNPs
- exon
- intron
- lariat
- exon
- mature mRNA
- spliceosome
Alternative splicing

• Alternative mRNAs produced from same gene
  – when is an intron not an intron...
  – different segments treated as exons

Starting to get hard to define a gene!
More post-transcriptional processing

- Need to protect mRNA on its trip from nucleus to cytoplasm
  - enzymes in cytoplasm attack mRNA
    - protect the ends of the molecule
    - add 5’ GTP cap
    - add poly-A tail
      - longer tail, mRNA lasts longer: produces more protein
Translation 13.2

From **nucleic acid** language to **amino acid** language
How does mRNA code for proteins?

DNA: TACGCACACATTACGTACGCGG

mRNA: AUGCGUGUAAAUGCAUGCGG

How can you code for 20 amino acids with only 4 nucleotide bases (A, U, G, C)?

protein: Met Arg Val Asn Ala Cys Ala

How can you code for 20 amino acids with only 4 nucleotide bases (A, U, G, C)?
mRNA codes for proteins in triplets

DNA molecule

Gene 1

Gene 2

Gene 3

mRNA codes for proteins in triplets

TACGCACATTTTACGTACGCGG

DNA

AUG

CGU

GUA

AAU

GCA

UGC

GCC

mRNA

Met

Arg

Val

Asn

Ala

Cys

Ala
The code

- Code for **ALL** life!
  - strongest support for a common origin for all life

- Code is redundant
  - several codons for each amino acid
  - 3rd base “wobble”

### Why is the wobble good?

- **Start codon**
  - AUG
  - methionine

- **Stop codons**
  - UGA, UAA, UAG

### Codon Table

<table>
<thead>
<tr>
<th>First base</th>
<th>Second base</th>
<th>Third base</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>Phe</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>UUC</td>
<td>UCC</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>UUA</td>
<td>UCA</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>UUG</td>
<td>UCG</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>CUU</td>
<td>CUU</td>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>CUC</td>
<td>CCC</td>
<td>His</td>
<td></td>
</tr>
<tr>
<td>CUA</td>
<td>CCA</td>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td>CUG</td>
<td>CCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUU</td>
<td>ACU</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>ACC</td>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>AUA</td>
<td>ACA</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>AUG</td>
<td>ACG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUU</td>
<td>GCU</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>GUC</td>
<td>GCC</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>GUA</td>
<td>GCA</td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td>GUG</td>
<td>GCG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Start codon**: AUG

**Stop codons**: UGA, UAA, UAG
How are the codons matched to amino acids?

DNA: TACGCACATTTTACGTACGCGG

mRNA: AUGCGUGUAUAAUGGCAUGCC

Amino acid: Met, Arg, Val

tRNA: UAC, GCA, CAU

Anti-codon: anti-codon
Transfer RNA structure

- “Clover leaf” structure
  - anticodon on “clover leaf” end
  - amino acid attached on 3’ end
Loading tRNA

- **Aminoacyl tRNA synthetase**
  - enzyme which bonds amino acid to tRNA
  - bond requires energy
    - ATP → AMP
  - bond is unstable
  - so it can release amino acid at ribosome easily

![Diagram of tRNA loading process]

- tRNA<sup>Trp</sup>
- anticodon
- activating enzyme
- tryptophan attached to tRNA<sup>Trp</sup>
- mRNA
- tRNA<sup>Trp</sup> binds to UGG condon of mRNA
Ribosomes

• Facilitate coupling of tRNA anticodon to mRNA codon
  – organelle or enzyme?

• Structure
  – ribosomal RNA (rRNA) & proteins
  – 2 subunits
    • large
    • small
Ribosomes

• **A site** (aminoacyl-tRNA site)
  – holds tRNA carrying next *amino acid* to be added to chain

• **P site** (peptidyl-tRNA site)
  – holds tRNA carrying growing *polypeptide* chain

• **E site** (exit site)
  – *empty* tRNA
    leaves ribosome from *exit* site
Building a polypeptide

- **Initiation**
  - mRNA, ribosome subunits, initiator tRNA come together

- **Elongation**
  - adding amino acids based on codons

- **Termination**
  - STOP codon = Release factor
Can you tell the story?

DNA

pre-mRNA

exon

intron

5' GTP cap

poly-A tail

mature mRNA

large ribosomal subunit

small ribosomal subunit

RNA polymerase

amino acids

tRNA

aminoacyl tRNA synthetase

polypeptide

tRNA

ribosome
The Transcriptional unit (gene?)

- Enhancer (1000 b)
- Promoter (20-30 b)
- Transcription start
- Exons
- Introns
- Transcription stop
- TATA box
- 5' RNA polymerase
- DNA
- UTR
- Transcription start
- Translation start
- Translation stop
- Mature mRNA
- Pre-mRNA
- GTP
- Mature mRNA
- AAAAAAAAAA
Protein Synthesis in Prokaryotes

Psssst... **no** nucleus!
Prokaryote vs. Eukaryote genes

- Prokaryotes
  - DNA in cytoplasm
  - circular chromosome
  - naked DNA
  - no introns

- Eukaryotes
  - DNA in nucleus
  - linear chromosomes
  - DNA wound on histone proteins
  - introns vs. exons

*exon = coding (expressed) sequence
*intron = noncoding (inbetween) sequence

introns come out!
Translation in Prokaryotes

- Transcription & translation are simultaneous in bacteria
  - DNA is in cytoplasm
  - no mRNA editing
  - **ribosomes** read mRNA as it is being transcribed
Translation: prokaryotes vs. eukaryotes

• Differences between prokaryotes & eukaryotes
  – time & physical separation between processes
    • takes eukaryote ~1 hour from DNA to protein
  – no RNA processing
Mutations
(Ch. 13.3)

(b) Sickled red blood cells and the primary structure of sickle-cell hemoglobin

DNA

Transcription

RNA

mRNA

Translation

Protein

Val His Leu Thr Pro Val Glu ···

1 2 3 4 5 6 7

10 μm
Mutations

• **Point mutations**
  – single base change
  • **silent mutation**
    – no amino acid change
    – redundancy in code
  • **missense**
    – change amino acid
  • **nonsense**
    – change to stop codon

When do mutations affect the next generation?
Point mutation lead to Sickle cell anemia

What kind of mutation?

Wild-type hemoglobin DNA

3’ C T T 5’

Mutant hemoglobin DNA

3’ C A T 5’

mRNA

5’ G A A 3’

mRNA

5’ G U A 3’

Normal hemoglobin

Glu

Sickle-cell hemoglobin

Val

Missense!
Mutations

- **Frameshift**
  - shift in the **reading frame**
    - changes everything “downstream”
  - insertions
    - adding base(s)
  - deletions
    - losing base(s)

Where would this mutation cause the most change: beginning or end of gene?
Mucus blocks air sacs (alveoli) in the lungs. Mucus blocks pancreatic ducts.

Chromosome 7

Sequence of nucleotides in CFTR gene

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Amino Acid sequence of CFTR protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Isoleucine 506</td>
</tr>
<tr>
<td>AT</td>
<td>Isoleucine 507</td>
</tr>
<tr>
<td>CT</td>
<td>Phenylalanine 508</td>
</tr>
<tr>
<td>GT</td>
<td>Glycine 509</td>
</tr>
<tr>
<td>GT</td>
<td>Valine 510</td>
</tr>
</tbody>
</table>

DELETED IN MANY PATIENTS WITH CYSTIC FIBROSIS
What's the value of mutations?
Review Questions
1. If proteins were composed of only 12 different kinds of amino acids, what would be the smallest possible codon size in a genetic system with four different nucleotides?

A. 1
B. 2
C. 3
D. 4
E. 12
2. A portion of the genetic code is UUU = phenylalanine, GCC = alanine, AAA = lysine, and CCC = proline. Assume the correct code places the amino acids phenylalanine, alanine, and lysine in a protein (in that order). Which of the following DNA sequences would substitute proline for alanine?

A. AAA-CGG-TTA
B. AAT-CGG-TTT
C. AAA-CCG-TTT
D. AAA-GGG-TTT
E. AAA-CCC-TTT
3. What is the relationship among DNA, a gene, and a chromosome?

A. A chromosome contains hundreds of genes, which are composed of protein.
B. A chromosome contains hundreds of genes, which are composed of DNA.
C. A gene contains hundreds of chromosomes, which are composed of protein.
D. A gene is composed of DNA, but there is no relationship to a chromosome.
E. A gene contains hundreds of chromosomes, which are composed of DNA.
4. A particular triplet of bases in the coding sequence of DNA is AAA. The anticodon on the tRNA that binds the mRNA codon is

A. TTT.
B. UUA.
C. UUU.
D. AAA.
E. either UAA or TAA, depending on wobble in the first base.
5. A part of an mRNA molecule with the following sequence is being read by a ribosome: 5' CCG-ACG 3' (mRNA). The following activated transfer RNA molecules are available. Two of them can correctly match the mRNA so that a dipeptide can form.

<table>
<thead>
<tr>
<th>tRNA Anticodon</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGC</td>
<td>Proline</td>
</tr>
<tr>
<td>CGU</td>
<td>Alanine</td>
</tr>
<tr>
<td>UGC</td>
<td>Threonine</td>
</tr>
<tr>
<td>CCG</td>
<td>Glycine</td>
</tr>
<tr>
<td>ACG</td>
<td>Cysteine</td>
</tr>
<tr>
<td>CGG</td>
<td>Alanine</td>
</tr>
</tbody>
</table>

The dipeptide that will form will be:
A. cysteine-alanine.
B. proline-threonine.
C. glycine-cysteine.
D. alanine-alanine.
E. threonine-glycine.
6. This figure represents tRNA that recognizes and binds a particular amino acid (in this instance, phenylalanine). Which of the following triplets of bases on the mRNA strand codes for this amino acid?

A. UGG
B. GUG
C. GUA
D. UUC
E. CAU
7. How is the template strand for a particular gene determined?

A. It is the DNA strand that runs from the 5' → 3' direction.
B. It is the DNA strand that runs from the 3' → 5' direction.
C. It depends on the orientation of RNA polymerase, whose position is determined by particular sequences of nucleotides within the promoter.
D. It doesn’t matter which strand is the template because they are complementary and will produce the same mRNA.
E. The template strand always contains the TATA box.
8. A biologist inserts a gene from a human liver cell into the chromosome of a bacterium. The bacterium then transcribes this gene into mRNA and translates the mRNA into protein. The protein produced is useless. The biologist extracts the protein and mature mRNA that codes for it. When analyzed you would expect which of the following results? *

A. the protein and the mature mRNA are longer than in human cells
B. the protein and mature mRNA are shorter than expected
C. the protein is longer and the mRNA is shorter than expected
D. the protein is shorter and the mRNA is longer than expected
9. If the structure of a TV show is analogous to the structure of a gene, then the introns of a gene would be analogous to
A. the opening theme music.
B. the segments of the show.
C. the commercials between segments of the show.
D. the commercials between shows.
E. the closing credits.
10. Each of the following is a modification of the sentence THECATATETHERAT.

A. THERATATETHECAT
B. THETACATETHERAT
C. THECATARETHERAT
D. THECATATTHERAT
E. CATATETHERAT

Which of the above is analogous to a frameshift mutation?

– A
– B
– C
– D
– E
11. Each of the following is a modification of the sentence THECATATETHERAT.

A. THERATATETHECAT
B. THETACATETHERAT
C. THECATARETHERAT
D. THECATATTHERAT
E. CATATETHERAT

Which of the above is analogous to a single substitution mutation?

— A
— B
— C
— D
— E