Gene Expression: Translation
Reminder

• Genes that encode proteins are transcribed and the transcript is processed to make mRNA.
• Next the base sequence in the mRNA must be translated into amino acid sequences in a polypeptide.
• Once polypeptides are formed, they fold up and combine with other molecules, but this is the realm of biochemistry, not genetics.

Review structure of polypeptides; a protein consists of one or more polypeptides.
The Code Problem

DNA and RNA sequences are written in an alphabet of 4 different bases: G, A, T (U), C.

Protein sequences are written in an alphabet of 20 different amino acids.

Therefore there must be a code that relates a base sequence to an amino acid sequence.
Important features:
• Differ only in side chains
• Each has three-letter and one-letter abbreviations

You should understand that the sequence of amino acids in a polypeptide (protein) can be written two ways:
Arg Thr Ser Ser

Polypeptides have an N- and a C-terminal AA

You don’t have to memorize these but you do have to know there are 20 and recognize amino acid sequences.
The Genetic Code

The code is:

- **Written in RNA** because it is the mRNA sequence that is read.

- **Universal (almost)**  
  Exceptions (small):  
  - Vertebrate mitochondria  
  - Invertebrate mitochondria  
  - Chloroplasts  
  - Ciliate nuclear  
  - *Mycoplasma* nuclear  
  - *Candida* nuclear  
  - etc.

- **Triplet**  
- **Nonoverlapping**  
- **Commaless**  
- **Degenerate**

<table>
<thead>
<tr>
<th>First position (5' end)</th>
<th>Second position</th>
<th>Third position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UUU Phe</td>
<td>UGU Cys</td>
</tr>
<tr>
<td></td>
<td>UUC Phe</td>
<td>UGC Cys</td>
</tr>
<tr>
<td></td>
<td>UUA Leu</td>
<td>UAA Stop</td>
</tr>
<tr>
<td></td>
<td>UUG Leu</td>
<td>UAG Stop</td>
</tr>
<tr>
<td>C</td>
<td>CUC Leu</td>
<td>CGU Arg</td>
</tr>
<tr>
<td></td>
<td>CUC Leu</td>
<td>CCG Arg</td>
</tr>
<tr>
<td></td>
<td>CUA Leu</td>
<td>CAU His</td>
</tr>
<tr>
<td></td>
<td>CUG Leu</td>
<td>CAG Gln</td>
</tr>
<tr>
<td></td>
<td>CUC Leu</td>
<td>CAC His</td>
</tr>
<tr>
<td></td>
<td>CAC Leu</td>
<td>CAA Gln</td>
</tr>
<tr>
<td></td>
<td>CUG Leu</td>
<td>CAG Gln</td>
</tr>
<tr>
<td>A</td>
<td>AUG Met</td>
<td>GGU Trp</td>
</tr>
<tr>
<td></td>
<td>AAU Ile</td>
<td>AAG Lys</td>
</tr>
<tr>
<td></td>
<td>AUC Ile</td>
<td>AAG Lys</td>
</tr>
<tr>
<td></td>
<td>AUA Ile</td>
<td>AAG Lys</td>
</tr>
<tr>
<td>G</td>
<td>GUU Val</td>
<td>GGU Gly</td>
</tr>
<tr>
<td></td>
<td>GUC Val</td>
<td>GGU Gly</td>
</tr>
<tr>
<td></td>
<td>GUA Val</td>
<td>GGU Gly</td>
</tr>
<tr>
<td></td>
<td>GUG Val</td>
<td>GGU Gly</td>
</tr>
</tbody>
</table>

*Note: Each amino acid is given its conventional abbreviation in both the single-letter and the three-letter format. The codon AUG, which codes for methionine (boxed) is generally used for initiation. The codons are conventionally written with the 5' base on the left and the 3' base on the right.*
The Genetic Code

The code is:
• Universal (almost)
• Triplet
• Degenerate
• Nonoverlapping
• Commaless

CACCAUGGUGCACCUGACUCCUGAG...CACUAAGCU

Quadruplet: AUGGUGCACCUGACUC

Overlapping: AUGGUGCACCUGACUC

Comma: AUGCCGUGCCCAACCCUGG
The Genetic Code

UAA, UAG, and UGA are **nonsense** codons; they do not code for any amino acid and hence are **stop** or **termination** codons.

All the rest are **sense** codons.

AUG is the **start** codon and codes for methionine (Met, M).

(only one to memorize)

CACCAUGGUGCACCUGACUC

**Met**  **Val**  **His**  **Leu**  **Thr**  **Pro**  **Glu**  ....  **His**  **Stop**
Open Reading Frames

An open reading frame (ORF) is a string of sense codons starting with the start codon ATG and flanked at the 3’ end by a stop codon.

All genes that code for proteins must have an ORF.

CACCAUGGUGGACCACCUGACUCCUGAG...CACUAAGGCU

Met  Val  His  Leu  Thr  Pro  Glu  ....  His  Stop
Start
Using Open Reading Frames to Find Genes

A computer program can search a sequence of bases for open reading frames. These are candidates for genes encoding proteins.

Problems:
Using Open Reading Frames to Find Genes

A computer program can search a sequence of bases for open reading frames. These are candidates for genes encoding proteins.

Problems:

• A gene can be on either strand, but the sequence is only written for one strand. Solution: search both complementary sequences.

• A random sequence of bases can have an ORF. Partial solutions: look for long ORFs starting with ATG.

• Introns can interrupt ORFs. The introns are spliced out of the mRNA leaving only the exons which form a continuous ORF; but DNA sequences will still have the introns. Partial solution: look for sequences that often flank introns.

Designing computer programs to search complete genome sequences is a major problem in bioinformatics.
The Mechanics of Translation

1. Translation requires:
   - Small ribosomal subunit = SSUrRNA + ribosomal proteins
   - Large ribosomal subunit = LSUrRNA + ribosomal proteins + 5SrRNA (eukaryotes)
   - (Small and large subunits also have S names: 16S, 18S, 23S, etc. S is for Svedberg units describing how fast something moves in a centrifugal field.)
   - Aminoacyl tRNAs = transfer RNAs + amino acids
   - Accessory proteins that promote various steps

2. mRNA is translated 5’ to 3’

3. Polypeptide is made N-terminal to C-terminal
After eukaryotic nuclear genes are transcribed and processed, the mRNA must be moved to the cytoplasm for translation.

Prokaryotic genes, and those in the chloroplasts and mitochondria, are not separated from the sites of protein synthesis. Transcription and translation proceed simultaneously.
Making Aminoacyl tRNAs

Each tRNA has a specific base sequence, including an anticodon that can base pair with a codon.
An aminoacyl tRNA synthase recognizes a tRNA and its corresponding amino acid and joins them.

The anticodon on the aminoacyl tRNA basepairs with its codon on the mRNA.

A new peptide bond is formed to join the amino acids.
Wobble

There are 61 sense codons. However, organisms may not have 61 different tRNAs.

1st (5’) base in anticodon can sometimes pair with 2 or 3 bases:

<table>
<thead>
<tr>
<th>5’ anticodon base</th>
<th>pairs with 3rd (3’) base in codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>E. coli: U</td>
</tr>
<tr>
<td>C</td>
<td>E. coli: G</td>
</tr>
<tr>
<td>U</td>
<td>S. cerevisiae: A or G</td>
</tr>
<tr>
<td>G</td>
<td>S. cerevisiae: C or U</td>
</tr>
<tr>
<td>Inosine</td>
<td>Inosine: A, C, or U</td>
</tr>
</tbody>
</table>

Don’t need to memorize these; just know basic principle.
Translation in More Detail than You Wanted or Need to Know
(see text Figure 10.17 for a better diagram)
Ribosome Binding, Translation Initiation, and Termination Signals on mRNAs

Prokaryotes: a special sequence (the “Shine-Delgarno sequence”) is the ribosome binding site.

Eukaryotes: the 5′ end of the mRNA is modified to form the “5′ cap” that initiates ribosome binding.

Prokaryotes and eukaryotes: the AUG start codon is the signal to initiate translation; the nonsense stop codon binds no tRNA and this stops translation.
Unique Features of Translation in Prokaryotes

One mRNA can encode more than one polypeptide.

With prokaryotic ribosomes, three polypeptides are made because the ribosomes can initiate translation within an mRNA.

With eukaryotic ribosomes, only one polypeptide is made because the ribosomes can initiate translation only at the 5' end.
Unique Features of Translation in Prokaryotes

Translation of an mRNA can begin before transcription is complete, because these processes are not separated by a nuclear membrane.
Where the Machinery Comes From

The machinery for DNA replication and repair, transcription, and translation includes various enzymes and other proteins, as well as rRNAs and tRNAs.

Each of these is encoded by genes, which in turn are transcribed to make RNAs and these are processed in various ways.

The enzymes and proteins are themselves translated from mRNAs using the machinery of which they are components.

Like all genes, they are subject to mutations that can change the rate or manner in which they act.

Like all genes, they are the product of evolution.