Converting a DNA sequence to an Amino Acid Sequence (Human translation):

A *codon* is a series of three nucleotide bases that encode a single amino acid.

1. Three DNA bases specify a single amino acid. These are called a 'codon'. For example, the following codon is translated as follows:
   \[
   RNA \ UGG \rightarrow TGG \rightarrow Trp
   \]

2. The first codon in all genes that encode proteins is *ATG (AUG)* in the (RNA), coding for the amino acid methionine (in eukaryotic cells) and formyl-Met in prokaryotic cells. All internal ATG codons will be translated as Met, e.g. a bacterial mRNA:
   \[
   5'--AUG-----------AUG--------UAA--
   fMet------------Met--
   \]
   (HIV protease does not start with a Met because it is produced from a longer peptide by proteolysis.)

3. Special codons (termination/stop codons) indicate the end of the protein. These are UAA, UAG, UGA. (The HIV protease sequence lacks a stop codon because its carboxy terminus is produced by proteolysis.)

4. The "reading frame" must be defined during the translation of the mRNA to protein. The reading frame is the base that is taken to be the first base of the codon. The rest of the codons are obtained by taking 3 bases at a time. Without knowledge of the reading frame the above sequence could be punctuated in any one of the following three ways, giving three completely different sequences.

   **Frame 1**
   
<table>
<thead>
<tr>
<th>Pro</th>
<th>Gln</th>
<th>Ile</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCT</td>
<td>CAG</td>
<td>ATC</td>
</tr>
</tbody>
</table>

   **Frame 2**
   
<table>
<thead>
<tr>
<th>Leu</th>
<th>Arg</th>
<th>Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTC</td>
<td>AGA</td>
<td>TC</td>
</tr>
</tbody>
</table>

   **Frame 3**
   
   | Ser | Asp-
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>TCA</td>
</tr>
</tbody>
</table>

There is only one correct reading frame. Signals in the mRNA tell the ribosome what reading frame to use. The reading frame from a DNA sequencing experiment can only be established with certainty by comparing the protein sequence predicted from the DNA to the protein sequence determined by chemical methods (e.g. Edman degradation). In the case of HIV protease, the established reading frame is:

**Region of HIV DNA Coding for HIV protease.**

5' -ggagcccctagacagaaaactgtaactcttactactccttcagatcacgtttgcccag-3'

**ProGlnIleThrLeuTrpGl

Example:
1. Given the following sequence data from HIV protease, what is the amino acid sequence?
2. Give the last four (3' end) bases of the primer that was used to obtain these data:

   ![DNA sequence and codon table]
Lec 38: DNA Transcription (mRNA), tRNA & Ribosomes

Gene: A DNA sequence that is transcribed to an RNA sequence. A gene contains a promoter at its 5' end to initiate RNA synthesis followed by codons if a protein is to be made.

Promoter: DNA sequence that RNA polymerase binds.

Transcriptional termination signal: Causes RNA polymerase to stop and leave DNA template so that only the gene of interest is transcribed into mRNA.

RNA polymerase:
- Holoenzyme: $\sigma + \alpha_2 \beta \beta'$
- Core polymerase $= \alpha_2 \beta \beta'$ - sufficient for elongation.
- Binds to promoter (P) sequence in a base specific manner via $\sigma$ subunit.
- Uses DNA as a template, making a copy of the template (lower) strand.
- Generates an RNA copy of the DNA template.
- NTPs are polymerized in the 5'→3' direction. Hydrolysis of P-P', drives reaction ($\Delta G<0$, indirect coupling).
- Does not require a primer (makes its own).
- Processive - continues $5'$→$3'$.
- No error checking. Product is 'disposable'.

Diagram:

1. Reversible transcription to RNA
2. Irreversible transcription to RNA
3. DNA template
4. RNA template
5. mRNA
Steps in the Production of mRNA:
1. **Template binding**: Holoenzyme (R) binds only to promoter sites (P), reversibly.
2. **"Open complex" formation**: An irreversible, committed step, DNA is melted (from -9 to +2).
3. **Chain initiation**: When the RNA chain is about 10 nucleotides long, α-subunit dissociates, leaving core enzyme to elongate the RNA processively (i.e., without dissociating from the DNA template).
4. **Chain elongation**: RNA chain growth is from 5' to 3', and elongation is rapid: about 50 nucleotides/sec.
5. **Chain termination**: Termination occurs at specific DNA sequences, causing release of mRNA.

Inducible Expression of Recombinant Proteins utilizing the Lactose Operon:
- **Lac operator** – DNA sequence that binds the lac repressor protein, utilized as an on/off switch in the control of recombinant proteins.
- **Lac repressor** – protein that binds to the lac operator, released from DNA by lactose or IPTG
- **Lac I gene** – chromosomal, produces the lac repressor protein

1. The constitutive expression of high levels of almost any protein is toxic to the bacteria. The protein itself may be toxic, or the simple competition for cellular resources can lead to poor growth of the bacterial host, and in some cases, cell death. Therefore, it is necessary to regulate the production of high levels of recombinant protein.

2. In the Lactose operon, production of proteins involved in the metabolism of lactose are controlled by the binding of the lac repressor (a protein that is the product of the lacI gene, produced from the bacterial chromosome) to a region of the DNA near the promoter region of the genes that encode the proteins for lactose metabolism. This segment of DNA is called the lac operator. Although, this system usually controls the expression of enzymes required for metabolism of lactose we can use it to control expression of the HIV mRNA (or any other gene) by simply placing the appropriate DNA segments in the correct location in our expression vector.

3. The lac repressor binds to the DNA when lactose is absent and blocks transcription of the DNA. When lactose is present, it binds to the lac repressor, causing an allosteric change that releases it from the DNA. Since lactose would be rapidly degraded by the bacteria, a non-hydrolyzable analog, isopropyl-thio-galactoside (IPTG), is used instead.

4. Once the lac repressor leaves the DNA, RNA polymerase can bind, allowing production of mRNA that can be used by the ribosomes to produce HIV protease.
RNA molecules involved in protein synthesis:
mRNA – messenger RNA is a copy of the DNA that encodes a gene. This was made by RNA polymerase. The mRNA specifies the order of amino acids to be used in making the protein.
tRNA – transfer RNA is the dictionary the converts the codon to a specific amino acid. One part of the tRNA recognizes the codon, the other part contains the amino acid to add.
rRNA – ribosomal RNA is found in the ribosome and is responsible for most of the function in protein synthesis.

tRNA Structure and Function: There are generally 25-45 different tRNAs/organism. The structure of this single chain RNA molecule is stabilized by W-C H-bonds, non-W-C H-bonds, and phosphate-metal (Mg²⁺) interactions.

- **Acceptor stem:** amino acids are attached to the 3' terminus of the tRNA by enzymes called **aminoacyl-tRNA Synthetases (aaRS).** These enzymes attach the correct amino acid to the correct tRNA. There are ~25-30 of these enzymes, essentially one for each amino acid. This process is often referred to as “charging” the tRNA.

- **Anti-codon arm:** contains the anticodon triplet that translates the codon in mRNA to an amino acid. Watson-Crick H-bonds are used here.

"Charging" of tRNA: amino acyl tRNA synthetases (aa tRNA syn) – a two step process.

\[
[\text{AA}_x + \text{ATP} \rightarrow \text{AA}_x\text{-AMP} + 2\text{P}_i] + \text{tRNA}^x \rightarrow \text{tRNA}^x\text{-AA}_x + (\text{AMP} + 2\text{P}_i)
\]

There are separate aminoacyl tRNA synthetases for each amino acid/tRNA combination. They add the correct amino acid to the correct tRNA. The aminoacyl tRNA synthetase shown above will only bind the tRNA for phe, with the anti-codon 3'-AAG-5', and will only add Phe to the bound tRNA. No other tRNA will be bound and no other amino acid will be added to that tRNA.
Step 1. Activation of the amino acid (2 ATP equivalents)

\[
\text{ATP} \rightarrow \text{Adenosine} + \text{pyrophosphate}
\]

Step 2. Transfer of activated AA to tRNA

Ribosome - The ribosomal subunits and their RNA components are named for their sedimentation coefficients, S, which is a measure of how rapidly they move when a centrifugal force is applied. 50S+30S=70S

- 30S – mRNA tRNA interactions.
- 50S – peptide bond formation.
- Exit tunnel – where growing protein chain exits the ribosome, amino term first.
- tRNA binding sites:
  - A site binds the aminoacyl-tRNA (charged tRNA) containing the next amino acid to be added.
  - P site holds the peptidyl-tRNA at the start of each cycle - the growing chain. The peptide shifts to the A site, temporarily, during each addition of an amino acid.
  - E site is the exit site for the uncharged tRNA, after peptide bond formation.

50S Subunit
- 31 Proteins
- 23S and 5S rRNA

30S Subunit
- 21 proteins
- 16S rRNA - binds to Ribosome Binding Site (Shine-Dalgarno) on mRNA.