**Lecture 38: DNA Sequence Interpretation, DNA Transcription**

**Codon Table:**

|  |  |  |
| --- | --- | --- |
| **5' Base** | **Middle Base** | **3'** |
|  | **T(U)** | **C** | **A** | **G** |  |
| **T(U)** | Phe | Ser | Tyr | Cys | **T(U)** |
|  | Phe | Ser | Tyr | Cys | **C** |
|  | Leu | Ser | **Term** | **Term** | **A** |
|  | Leu | Ser | **Term** | Trp | **G** |
| **C** | Leu | Pro | His | Arg | **T(U)** |
|  | Leu | Pro | His | Arg | **C** |
|  | Leu | Pro | Gln | Arg | **A** |
|  | Leu | Pro | Gln | Arg | **G** |
| **A** | Ile | Thr | Asn | Ser | **T(U)** |
|  | Ile | Thr | Asn | Ser | **C** |
|  | Ile | Thr | Lys | Arg | **A** |
|  | **Met** | Thr | Lys | Arg | **G** |
| **G** | Val | Ala | Asp | Gly | **T(U)** |
|  | Val | Ala | Asp | Gly | **C** |
|  | Val | Ala | Glu | Gly | **A** |
|  | Val | Ala | Glu | Gly | **G** |

*A codon* is a series of three nucleotide bases that encode a single amino acid. DNA = RNA with U (RNA) replacing T.

1. Three DNA bases specify a single amino acid. These are called a 'codon'. For example, the following codon is translated as follows:

5’TGG3’ =

1. The first codon in all genes that encode proteins start with ATG (AUG in the RNA), or the amino acid methionine.

(HIV protease does not start with a Met because it is produced from a longer peptide by a cleavage reaction.)

1. Special codons (termination/stop) indicate the end of the protein. These are TAA /UAA, TAG/UAG, TGA/UGA.
2. Many amino acids have multiple codons, i.e. Phe. These codons usually differ in the third base, e.g. TTT/TTC
3. 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. **Only 1 of the 3 reading frames will generate the correct amino acid sequence of the protein.**

*Frame 1(correct) Frame 2 Frame 3*

**C**CT CAG ATC ***or*** C **C**TC AGA TC ***or*** CC **T**CA GAT C

Pro-Gln-Ile Leu-Arg-Ser -Ser-Asp-

The reading frame from a DNA sequencing experiment is established by comparing the predicted protein sequence to the protein sequence determined by chemical methods.

**Worked Example - Finding Mutations**

1. Read sequence in region of gel that shows changes.

2. Find the location of the sequence in the HIV protease gene, whose sequence, *and reading frame*, has already been determined and the

3. Identify 1st base of a codon, based on HIV protein sequence (*determining reading frame*)

4. Translate each codon.

|  |  |
| --- | --- |
| WT: TA TTA GAA GAA AMut:TA CTC GAG GAA A | **Region of HIV DNA Coding for HIV protease.**5'-ggagccgatagacaaggaactgtatcctttaacttccctcagatcactctttggcaa57 **ProGlnIleThrLeuTrpGln7**cgacccctcgtcacaataaAgataggggggcaactaaaggaagctctatta**gat**acagga117**ArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGly27**gcagatgatacag**tattagaagaaat**gaGtttgccaggaaGatggaaaccaaaaatgata177**AlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIle47**gggggaattggaggttttatcaaagtaagacagtaTgatcagatacTCAtagaaatctgt237**GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCys67**ggacataaagctataggtacagtattagtaggacctacacct**gtc**aacataattggaaga297**GlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg87**aatctgttgactcagattggttgCactttaaatttTcccattagccctattgagact354-3'**AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe** |

**DNA Transcription (DNA→mRNA→Protein) :**

**Key Terms:**

**Promoter:** DNA sequence that RNA polymerase binds.

**Lac operator**: Region of DNA that the lac repressor binds – inducible on/off switch for mRNA production.

**Optimal E. coli promoter**

-35 region -10 region Start of mRNA

TTGACA -------17 bp--------TATAAT---6 bp--- A(or G)—mRNA---------

AACTGT-----------------------ATATTA-------------T(or C)--template stand

**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: σ + α2ββ’ Core: α2ββ’

* Binds to promoter (P) sequence in a base specific manner via σ subunit.
* Uses DNA as a template
* Does not require a primer (makes its own).
* Generates an RNA copy of the DNA template.
* NTPs are polymerized in the 5’→3’ direction
* No error checking.

**1. Template binding**: Holoenzyme (R) binds only to promoter sites (P), reversibly.

**2. "Open complex" formation**: A irreversible, committed step, DNA is melted (from -9 to +2).

**3. Chain initiation:** When the RNA chain is about 10 nucleotides long, σ-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:

 HIV Protease->

 mRNA------------------------------------------->

**TTGACA**TTTATGCTTCCGGCTCG**TATAAT**GTGTG**TGAGCGGATAACAATTTCACACA**GGAAACAGCT**ATG**..

 -35 -10 <--lac operator---------> Met…

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.

1. 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.
2. 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.

1. Once the lac repressor leaves the DNA, RNA polymerase can bind, allowing production of mRNA that can be used by the ribosome to produce HIV protease.

