Genetics, L2 Biology Module provided by: Medjani S,

TD 6 (protein synthesis)

Exercise 1:

1. A portion of the coding strand (sense strand) for a given gene has the sequence

5′ ATGAGCGACTTTGCGGGATTA 3

a) Deduce the non-coding strand (antisense strand), which strand is the transcribed strand?

5′ ATGAGCGACTTTGCGGGATTA3' sense strand (coding strand)

3'TACTCGCTGAAACGCCCTAAT 5'antisense strand (non-coding strand)

The transcribed strand is the antisense strand (non-coding strand)

(b) What are the mRNA and protein sequences that would be produced during transcription and translation of this DNA segment?

mRNA 5′ A U GAGCGAC UUU GCGGGA UU A3'

Protein **N-ter** Met Ser Asp Phe Ala Gly Leu **C-ter**

Exercise 2:

1) What does this diagram represent?

tRNA

2) Determine the ends of this molecule?

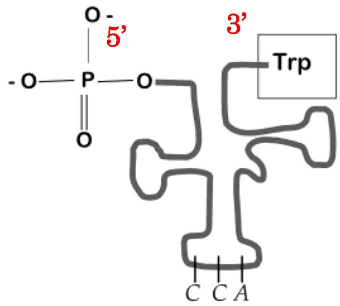
See figure (put the 3'front end tryptophan)

3) What are the three nucleotides found at the end of the 3' end?

The three nucleotides are CCA 3' (note: this is not the anti-codon, it's just a coincidence)

4) Give the nucleotides of the anticodon?

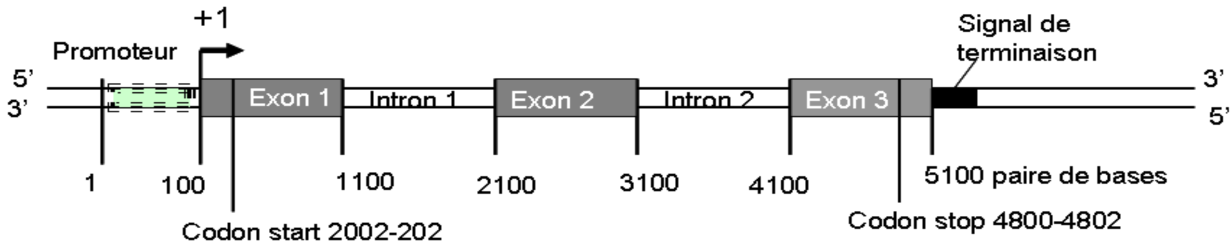
The Trp anti-coder is 5'CCA 3'(**codon 5'UGG3' on mRNA**)



Exercise 3:

(WARNING: The answers to this exercise will need to be explained to students who sh take notes. It is not necessary to dictate the answers to them as they are)

Below is a schematic representation of the Y gene, which encodes protein X. Transcription begins immediately after the promoter.



1) The primary transcript of this gene produced by RNA polymerase will be approximately how many nucleotides long?

The primary transcript (immature mRNA) produced by this gene would be about 5000 nucleotides (From base 100 to base 5100 **THE PROMOTER IS NOT EXPOSED**). Transcription begins at nucleotide +1 which is **the first nucleotide of the transcript** .

2) Two different transcripts are produced from this gene, one is about 2000 nucleotides long, the other is about 3000 nucleotides long. Explain how two different transcripts can be produced from this gene?

This is possible due to alternative splicing (post- transcriptional modifications). Alternative splicing could produce two different transcripts. One transcript may contain exons 1, 2 and 3 (3000 nucleotides), the other may contain only exons 1 and 3 (2000 nucleotides).

3) Suppose each transcript produces a protein. Given the diagram above, what is the approximate size of the protein produced by the main transcript of this gene?

The main transcript is the transcript that contains all the exons, so here exons 1, 2 and 3. Its size is 3000 nucleotides (after post- transcriptional modifications (elimination of introns 1 and 2)

Caution : The region from nucleotide 100 to 200 is an untranslated region 5'because it is

upstream of the **start codon** (the 5'UTR region (5' untranslated region)) and the region from nucleotide 4802 to 5100 is an untranslated region 3'because it is downstream of the **stop codon** (the 3'UTR region (3' untranslated region)
So the mature mRNA of this transcript contains all three exons that encode a protein starting at the ATG at position 200 and ending at the stop codon. This is about 2600 nucleotides (3000-100 (5'UTR) -300 (3'UTR) =2600), which is about 866 codons. So the expected protein would be about 866 amino acids

4) Suppose that gene Y was mutated such that the G/C base pair found at position 2200 awas replaced by an A/T. Would transcription and translation of the mutated gene Y still occur?
This mutation changes a single base pair well after the promoter, so it should not affect transcription and translation. The primary transcript should be the same length.

5) Would the protein produced be the same length, shorter or longer than the protein produced by the wild-type Y gene? Explain your answer.

This mutation changes a single base pair of a single codon. There are two possibilities. If the new codon still codes for an amino acid, then the resulting protein will be the same length. If the new codon creates a stop codon, the protein will be the shorter one.

Exercise 4:
ATTENTION: The sense strand is the strand that has the same direction of the transcribed RNA AND ALSO the same nucleotide composition except that instead of T there is a U.
Here is a DNA sequence consisting of 210 consecutive base pairs that corresponds to the beginning of the X gene sequence. Transcription begins at the T/A base pair at position 60 (In bold)
1) What are the first 6 nucleotides of the mRNA of gene X?
The first transcribed base is at position 60. The first 6 nucleotides of the mRNA are 5'UAG G CU3'

2) What are the first 4 amino acids coded for by gene X? (A genetic code table can be found on the last page)
The start codon (the first ATG codon (AUG on RNA) is at position 133. The first 4 amino acids encoded by the X N-ter met arg arg leu C-ter

```
    1        10        20        30        40        50        60        70
    I--------I---------I---------I---------I---------I---------I---------I
5'  ATCGGTCTCGGCTACTACATAAACGCGCGCATATATCGATATCTAGCTAGCTATCGGTCTAGGCTACTAC
3'  TAGCCAGAGCCGATGATGTATTTGCGCGCGTATATAGCTATAGATCGATCGATAGCCAGATCCGATGATG
```

Pomoteur

```
             80        90       100       110       120       130       140
    ---------I---------I---------I---------I---------I---------I---------I
5'  CAGGTATCGGTCTGATCTAGCTAGCTTCTCTTCTCTCTCCCCCGCGGGGCTGTACTATCATGCGTCG
3'  GTCCATAGCCAGACTAGATCGATCGAAGAGAAGAGAGAGGGGGCGCCCCGACATGATAGTACGCAGC
```

```
            150       160       170       180       190       200       210
    ---------I---------I---------I---------I---------I---------I---------I
5'  TCTCGGCTACTACGTAAACGCGCGCATATATCGATATCTAGCTAGCTATCGGTCTCGGCTACTACGTAAA
3'  AGAGCCGATGATGCATTTGCGCGCGTATATAGCTATAGATCGATCGATAGCCAGAGCCGATGATGCATTT
```

You found two different mutations of gene X, mutation 1 and mutation 2.
3) For mutation 1, there is an insertion of the three base pairs immediately after the C/G base pair at position 100 (in bold). 5'TGT3'
3'ACA5'

a) Would the mRNA expressed from this version of gene X be longer, shorter, or identical to the mRNA produced by the normal gene X? Explain, and if longer or shorter, indicate the number of bases.

The insertion is after the promoter and after the first transcribed nucleotide. So the mRNA would be longer by 3 nucleotides.

b) If the mRNA is translated, would the protein produced be longer, shorter, or the same size as that produced by the normal X gene? Will the protein produced have the same function as the normal X protein ? Explain your thinking.
The protein would be identical to the normal protein because the insertion is before the start codon .
The protein produced is identical to the normal protein, so it will have the same structure and function.
4) For mutation 2, there is a four base pair insertion immediately after the A/T base pair at position 130 (in bold). 5'ATGT3'
3'TACA5'
In this case:
c) What are the first four amino acids produced?
after insertion there is creation of a new ATG start codon just before the ATG codon of the wild gene and a reorganization of the codons. The first four amino acids N-ter met phe met arg C-ter

d) The protein produced would be longer, shorter, or the same size as that produced by the normal X gene? The protein produced will have the same function as the normal X protein ? Explain your thinking?
The protein encoded by mutant 2 will be 2 amino acids longer.
This mutant protein is almost identical to the normal protein. The only difference is that it has two extra amino acids at the N-terminus, but these two terminal amino acids are unlikely to change the overall tertiary structure of the protein, so it is unlikely that this insertion will change the function of the protein.

le code génétique

Première lettre (côté 5')	Deuxième lettre U		Deuxième lettre C		Deuxième lettre A		Deuxième lettre G		Troisième lettre (côté 3')
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
U	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
U	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A
U	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
C	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
C	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
C	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
A	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
A	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
A	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
G	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
G	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
G	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

codon d'initiation codon de terminaison