

**Problem: Starting from a given DNA sequence, find the protein sequence it may code for, as well as its predicted secondary structure content.**

**Input strand S:**

**5'-TAATGGCCTTAGAAGAGGGTCTCGCGAAACACTAAGG-3'**

**Solution:**

1) The DNA strand given as input may not be the coding strand. We first generate the sequence of the complementary strand, from 5' to 3':

**Strand cS:**

**5'-CCTTAGTGTTTCGCGAGACCCTCTTCTAAGGCCATTA-3'**

2) Once we have the DNA sequences of the DNA strand and its complementary, we transcribe them into RNA sequences:

**RNA sequence R corresponding to the input strand S:**

**5'-UAAUGGCCUAGAAGAGGGUCUCGCGAAACACUAAGG-3'**

**RNA sequence cR corresponding to the complementary strand cS:**

**5'-CCUAGUGUUUCGCGAGACCCUCUUCUAAGGCCAUUA-3'**

3) Identify start codons (in red) and then stop codons (in blue) in the two sequences:

**RNA sequence R:**

**5'-UA **AUG** GCC UUA GAA GAG GGU CUC GCG AAA CAC **UAA** GG-3'**

Note that there is a UAA (stop codon) at the 5' of the sequence R, which we ignore.

No start codon in the sequence cR, so this sequence does not contain a coding region.

4) Derive the corresponding ORF (Open Reading Frame), if the start and stop codon are in phase:

**ORF:**

****AUG** GCC UUA GAA GAG GGU CUC GCG AAA CAC **UAA****

5) Translate ORF to protein sequence:

Protein sequence:

**Met Ala Leu Glu Glu Gly Leu Ala Lys His**

Or, in one-letter code:

**M A L E E G L A K H**

6) Secondary structure prediction:

*Start with prediction of helix:*

<b>M</b>	<b>A</b>	<b>L</b>	<b>E</b>	<b>E</b>	<b>G</b>	<b>L</b>	<b>A</b>	<b>K</b>	<b>H</b>
<b>1.47</b>	<b>1.29</b>	<b>1.30</b>	<b>1.44</b>	<b>1.44</b>	<b>0.56</b>	<b>1.30</b>	<b>1.29</b>	<b>1.23</b>	<b>1.22</b>

**Nucleation site:** MALEEG

**Elongation:** we can add L ( $1.44+1.44+0.56+1.3>4$ ), A ( $1.44+0.56+1.30+1.29>4$ ), K ( $0.56+1.30+1.29+1.3>4$ ), and H ( $1.3+1.29+1.23+1.22>4$ ).

**Final check:**  $(1.47+1.29+1.30+1.44+1.44+0.56+1.30+1.29+1.23+1.22)/10 = 1.254 > 1$   
The whole peptide can be helical.

*Repeat with prediction of strands:*

<b>M</b>	<b>A</b>	<b>L</b>	<b>E</b>	<b>E</b>	<b>G</b>	<b>L</b>	<b>A</b>	<b>K</b>	<b>H</b>
<b>0.97</b>	<b>0.9</b>	<b>1.02</b>	<b>0.75</b>	<b>0.75</b>	<b>0.92</b>	<b>1.02</b>	<b>0.90</b>	<b>0.77</b>	<b>1.08</b>

**Nucleation site:** None

**The peptide is therefore predicted to be helical.**