Green Fluorescent Protein (GFP) Exercise
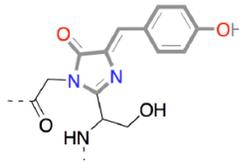
Learning Objectives
In this exercise, you will use the StarBiochem and StarORF software tools to explore:
- the sequence of the GFP gene and the structure of its protein
- the fluorophore that accounts for GFP's fluorescence

Background
GFP is a fluorescent protein isolated from several organisms including the pacific jellyfish *Aequoria Victoria*. GFP converts two specific wavelengths of blue light into green fluorescent light by energy transfer.

Proteins that fluoresce contain components called fluorophores. In GFP the fluorophore originates from an internal tripeptide sequence (Ser-Tyr-Gly) which is post-translationally modified to the structure shown below, a 4-(p-hydroxybenzylidene)-imidazolidin-5-one.
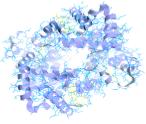

It is worth noting that the formation of GFP's fluorophore is an autocatalytic process that requires no cofactor or enzymatic reaction. This fluorophore is highly stable even at high temperatures and various pH conditions.

The gene for GFP has been isolated and has become a useful tool for making fusion proteins in which GFP is linked to other proteins and functions as a fluorescent protein tag. GFP tagging can be used in a wide range of applications: as a tracer for cell lineage, as a reporter of gene expression, or as a reporter of protein-protein interactions.

Getting started with StarORF
StarORF is a six-frame translation software tool for exploring gene sequences.
- To get to StarORF, please navigate to **http://web.mit.edu/star/ORF**.
- Click on the **Start** button to launch the application.
- Click **Trust** when a prompt appears asking you if you trust the certificate.
- In the **Input** box, paste the sequence provided below.

The **cDNA sequence** of the GFP gene (5'→ 3' direction) is shown below.

tacacacgaataaaagataacaaagatgagtaaaggagaagaactttcactggagttgtcccaattcttgttgaattagatggcgatgttaatgggc
aaaaattctctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactgggaagctacctgttccatggc
caacacttgtcactactttctcttatggtgttcaatgctttttcaagatacccagatcatatgaaacagcatgacttttcaagagtgccatgcccgaaggt
tatgtacaggaaagaactatattttacaaagatgacgggaactacaagacacgtgctgaagtcaagtttgaaggtgatacccttgttaatagaatcga
gttaaaaggtattgattttaaagaagatggaaacattcttggacacaaaatggaatacaactataactcacataatgtatacatcatggcagacaaac
caaagaatggaatcaaagttaacttcaaaattagacacaacattaaagatggaagcgttcaattagcagaccattatcaacaaaatactccaattgg
cgatggccctgtcctttaccagacaaccattacctgtccacacaatctgccctttccaaagatcccaacgaaaagagagatcacatgatccttcttgag
tttgtaacagctgctgggattacacatggcatggatgaactatacaaataaatgtccagacttccaattgacactaaagtgtccgaacaattactaaat
tctcagggttcctggttaaattcaggctgagactttatttatatatttatagattcattaaaattttatgaataatttattgatgttattaatagggggctattt
tcttattaaataggctactggagtgtat

Using the StarORF software tool answer the following questions.

1 *What is the length of the GFP cDNA?*

Answer

2 *What is the percentage of each nucleotide base in the GFP cDNA?*
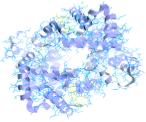
Answer

3 *What is the sequence of the first ten bases of GFP's noncoding DNA strand (5'→ 3' direction).*

Answer

4 *What is the sequence of the first 15 bases of GFP's mRNA (5'→ 3' direction)?*

Answer

5 Using the cDNA sequence provided in this exercise, you estimate GFP's mRNA length. In your laboratory, you then isolate total GFP RNA from jellyfish and resolve it on a gel based on the RNA size difference. You find two different GFP RNAs: one of the RNAs is bigger than your estimate and the other RNA is the same size as your estimate. ***How do you explain this result?***

Answer

6 *The cDNA sequence provided in this exercise has only one open reading frame (ORF).* Explain why
- ORFs are visually represented by green lines within the **Six-frame translation** box.

Answer

7 *What are the first 10 amino acids of the GFP protein?*
- Under the **Six frame translation** box, click on the **green line**, which indicates a potential ORF within the sequence provided.
- The full translated amino acid sequence, represented by the **green line**, is indicated within the **Putative ORF protein sequence** window.
- The total number of amino acids of the translated sequence is indicated below the **Putative ORF protein sequence** window.
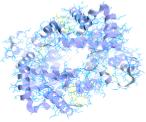- The amino acid sequence can be represented within StarORF in the **1 letter** or **3-letter** amino acid code.

Answer

We will now compare the GFP protein sequence deduced from the cDNA sequence provided in this exercise with the protein sequence you will obtain from GFP's protein structure.

To explore the structure of GFP and its primary sequence, we will use StarBiochem, a protein 3D-viewer.
- To get to StarBiochem, please navigate to **http://web.mit.edu/star/biochem**.
- Click on the **Start** button to launch the application.
- Click **Trust** when a prompt appears asking you if you trust the certificate.
- Under **File**, click on **Open/Import** and select "1EMA".
- Click **Open**.

You are now viewing the structure of GFP (structure ID: "1EMA") with each bond in the protein drawn as a line ("bonds only" view).

Practice changing the viewpoint of this protein in the view window.

	Mac	**PC**
TO ROTATE	click and drag the mouse	left-click and drag the mouse
TO MOVE UP/DOWN RIGHT/LEFT	apple-click and drag the mouse	right-click and drag the mouse
TO ZOOM	option-click and drag the mouse	Alt-left-click and drag the mouse

Take a moment to look at the structure of GFP (1EMA) from various angles in this "bonds only" view.

8 *What is the total length of the GFP protein sequence that you obtained using StarORF and StarBiochem software tools? Is it the same or different?* Explain your answer and reasoning.
   • In StarBiochem, under **Structure**, click on **Primary** which shows the protein's amino acid sequence.

Answer

9 The fluorophore in GFP originates from the Ser65 -Tyr66 -Gly67 tripeptide. For this tripeptide provide the following:

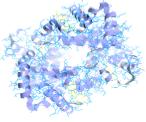**a)** *A possible DNA sequence with labeled 5' and 3' ends.*

Answer

b) *A possible mRNA sequence (5'→ 3' direction).*
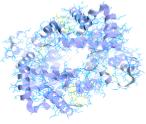
Answer

c) *The corresponding anti-codon sequence (5'→ 3' direction).*

Answer

10 You come across <u>four</u> GFP cDNAs. Each of these has one of the following <u>point</u> mutations. ***Complete the following table.***

Mutation (mRNA sequence)	What is the corresponding mutation in the cDNA?	Does the fluorophore form? Explain briefly.	Is the resulting GFP protein functional (fluoresces)? Explain briefly.
$G^{\#27}$ to $U^{\#27}$			
Extra "A" added after the 896th nucleotide, prior to "G"			
$A^{\#740}$ to $G^{\#740}$			
$U^{\#220}$ to $G^{\#220}$			
$U^{\#222}$ to $C^{\#222}$			

11 We will now take a closer look at the structure of GFP (1EMA).

a) *Draw the <u>chemical</u> structure of the fluorophore in GFP (1EMA).* Indicate the <u>atoms</u> within the structure and the parts that are contributed by Ser65, Tyr66, and Gly67. Label the alpha carbon contributed by each of these three amino acids.

• In StarBiochem, under **PDB tree**, click on the **1EMA** folder and then click on "Cro_66".
• In the **View Controls** panel, set the **Unselected** transparency slider to "**0**".
• Click on **Draw** within the **Atoms** box to see what atoms are present. Each atom is color-coded: Carbon is grey, Nitrogen is blue, and Oxygen is red.

Answer

b) *State the most likely interaction between Ser65, Tyr66 and Gly67 that contributes to the formation of the fluorophore.* Your choices are **'**hydrogen bond', 'ionic bond', 'hydrophobic interaction', 'covalent bond', 'peptide bonds', or 'van der Waals forces'.

Answer

c) *Which secondary structure surrounds the fluorophore: helices, sheets or coils?*

• Within the main menu go to **View**.
• Click on **Reset Molecule**.
• Under **Structure**, click on **Secondary**.
• Explore the different secondary structures by individually clicking on **Helices**, **Sheets** or **Coils** within the **Show Ribbons** box. Alternatively, you can click on **All Ribbons** within the **Show Ribbons** box to view all structures at the same time.

Answer

d) *How may this secondary structure be important for fluorophore function*?

Answer