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 begin using StarORF, please navigate to [http://mit.edu/star/ORF/](http://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.

```
tacacagctaaagcttgcttggtgcctcatcttttactctcatgggctgggcaacatggctggctgctgctgtttcttctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc
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 ten 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 RNA transcripts: one transcript is bigger than your estimated GFP mRNA and the other transcript is of the same size. How do you explain this result?

Answer:

6 Explain why the cDNA sequence provided in this exercise has only one open reading frame (ORF).
   • 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 begin using StarBiochem, please navigate to [http://mit.edu/star/biochem](http://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.
- In the top menu, click on **Samples** → **Select from Samples**. Within the **Amino Acid/Proteins** → **Protein** tab, select “Green Fluorescent Protein – A. victoria (1EMA)”. “1EMA” is the four character unique ID for this structure.

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

Take a moment to look at the structure from various angles by rotating and zooming on the structure.

- Instructions for changing the view of structure can be found in the top menu, under **Help** → [Structure viewing instruction](http://mit.edu/star/biochem).

The current way you are viewing the structure is by seeing each atom and bond in the protein drawn as a ball and a line, respectively. This way of representing a structure is called the **ball-and-stick** model and is the default model in StarBiochem. The **ball-and-stick** model allows you to see how atoms in the structure bond together. However, the space each atom occupies IS NOT accurately represented. To see a more realistic representation of the atoms in the structure you can use the **space-filled** model, where each atom is drawn as a **sphere**, whose size represents the physical space an atom occupies.

You can switch from the **ball-and-stick** model to the **space-filled** model in StarBiochem by increasing the size of the atoms in the structure:

- Notice that different atoms are slightly different in size.
- Gray = Carbon, Blue = Nitrogen, Red = Oxygen, Yellow = Sulfur, Orange = Iron.
- Click on the **Primary** tab. The default atom size is 20% (**ball-and-stick** model).
- Move the **Atoms Size** slider to 100% (**space-filled** model).

The last page, Reference page, contains a series of terms and useful information that you will refer to during this exercise.

**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.

- In the default view (**Structure** → **Protein** → **Primary**) carefully scroll down the [**Amino acid Position:Chain**](http://mit.edu/star/biochem) in the **Sequence** window to see the number of amino acids in each polypeptide chain/monomer.

**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 four GFP mRNAs. Each of these has one of the following point mutations. Complete the following table.

<table>
<thead>
<tr>
<th>Mutations in mRNA sequence</th>
<th>Is GFP produced (Yes/No)?</th>
<th>If GFP is produced, is it functional (fluoresces)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>G#27 to U#27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra “A” added after the 896th nucleotide, prior to “G”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A#740 to G#740</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U#220 to G#220</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U#222 to C#222</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Please take a look at the codon chart on the following page to complete the table.
We will now take a closer look at the structure of GFP (1EMA).

a) Draw the chemical structure of the fluorophore in GFP (1EMA). Indicate the atoms 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.
   - Reset the structure by clicking on Reset → Reset structure in the top menu.
   - From the top menu select View → View Specific Regions / Set Center of Rotation. This will open a smaller window, which enables you to set specific region(s) of the structure visible and centered in the viewer.
   - In the Non-Peptide tab of the View Specific Regions / Set Center of Rotation window, select the fluorophore molecule, [CRO]66:A, in the Sequence Window.
   - Move the VDW Radius slider all the way to the left (1 Van der Waals radii). Close the View Specific Regions / Set Center of Rotation window. Zoom in as needed.

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:

---

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:

---

Ver. 7 – D. Sinha and L. Alemán
• Reset the structure by clicking on **Reset → Reset structure** in the top menu.
• In the **Protein → Non-Peptide** tab, select the fluorophore, [CRO]66:A, and move the **Atoms Size** slider completely to the right (100%) to visualize the fluorophore.
• In the **Protein → Primary** tab, move the **Atoms Size** slider completely to the left and the **Bonds Translucency** slider to the right (95%) to minimize the appearance of all the amino acids in the hOGG1 protein.
• In the **Protein → Secondary** tab, explore the different secondary structures either one at a time or all together by checking the box beside the desired structure (ex: helices) and moving the **Structures Size** slider to the right to increase the size of the secondary structure(s). View additional secondary structures that may be present by checking the boxes next to each structure.

**Answer:**

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

**Answer:**
Reference

CHEMICAL STRUCTURES OF THE AMINO ACIDS
The 20 amino acids share a common backbone and are distinguished by different side chains, also called ‘R’ groups, highlighted by the various colors below.

PROTEIN STRUCTURE BASICS
All proteins have the following three levels of protein structure:

Primary structure
Describes the order of the amino acids in the protein chain but does not describe its shape.

Secondary structure
Describes shapes that form from local folding of regions within the amino acid chain. These smaller structures can be divided into two main types: helices and sheets. Coils are made of amino acids that do not form regular secondary structures (helices and sheets) but play important roles in protein folding.

Tertiary structure
Describes the entire folded shape of a protein chain.

In addition, some proteins interact with themselves or with other proteins to form larger protein structures. These proteins have an additional level of protein structure:

Quaternary structure
Describes how multiple protein chains interact and fold to form a larger protein complex.