

CSC / BIO 310

Bioinformatics

Instructor: Dr. Laurie J. Heyer

Project #2

In Class Presentations: Tuesday, April 29

Begin with a copy of the current gene splitter web page and perl program. You will build an alternative version of this page that will let the user upload the DNA sequence of a gene, and (optionally), a PDB structure file for the gene. Note that functionality of the PDB aspect is required for your projects.

You will need to either write the code to translate the DNA sequence to protein, or learn how to extract the primary amino acid sequence from the PDB file. Less desirable is another input on your web page for uploading the protein sequence.

The user should also be able to enter a DNA sequence to be inserted in the gene in the “least conspicuous” place. The heart of your project is deciding how to make the insertion have the least impact on the functionality of the gene. You might minimize number of DNA bases inserted, but you also want to minimize changes in codon usage. This will require some choices on your part that you should be prepared to defend in your presentation. The user should also be able to limit the part of the sequence in which they want to insert (start and stop base numbers).

Choose a selectable marker gene with known structure in the Protein Data Bank to test your page with. The DNA and PDB inputs for your chosen gene should be the default on your page, but ultimately you want the user to be able to upload different files.

Return a web page with the “top 10” matches as a list of DNA base numbers (and corresponding amino acid numbers) where the desired DNA sequence can begin. These should be live links to html pages showing the PDB file of your chosen gene, rendered in color to show the insertion site.