**Bioinformatics Worksheet #2: BLAST Practice - Assignment**

**Group Number:\_\_\_\_\_2\_\_\_\_\_\_\_\_**

**Names of individuals in group:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Mystery bacterial operon #: Operon 2**

**Table 1: Table of results for 8 largest proteins in operon**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Protein #** | **Length (base pairs)** | Length (amino acids) | E-value | Query Cover | % Identity | Protein name – brief description |
| 1 | 3075 | 1024 | 0.0 | 100 | 100 | [beta-D-galactosidase [Escherichia coli]](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi#alnHdr_446100051) - enzyme |
| 2 | 1254 | 417 | 0.0 | 100 | 100 | [MULTISPECIES: galactoside permease [Enterobacteriaceae]](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi#alnHdr_446213694) |
| 3 | 960 | 319 | 0.0 | 100 | 100 | [lac repressor [Escherichia coli str. K-12 substr. MG1655]](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi#alnHdr_1657541) |
| 4 | 606 | 201 | 2e-145 | 82 | 100 | [galactoside O-acetyltransferase [Escherichia coli]](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi#alnHdr_485785328) |
| 5 | 438 | 145 | 8e-95 | 100 | 100 | [cyanate transporter [Escherichia coli O91:NM str. 2009C-3745]](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi#alnHdr_607751876) |
| 6 | 318 | 105 | 5.8 | 52 | 36 | No protein found - <100% identity |
| 7 | 315 | 104 | 2e-40 | 65 | 99 | No protein found - <100% identity |
| 8 | 312 | 103 | 4e-64 | 100 | 100 | [hypothetical protein OG1X\_1862 [Enterococcus faecalis OG1X]](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi#alnHdr_429512671) |

1. What organism did this operon come from? What evidence did you use to come to this conclusion?

**Escherichia coli – (http://www.ncbi.nlm.nih.gov/nuccore/J01636.1)**

**“E.coli lactose operon with lacI, lacZ, lacY and lacA genes” (Gilbert & Maxam 1973).**

**Evidence from the fact that all of the best matching proteins were from this organism.**

1. Are the functions of the genes found in this operon related? What cellular system do you think this operon and the genes contained are involved in?

**By visiting the actual site describing the operon itself (http://www.ncbi.nlm.nih.gov/nuccore/J01636.1) and scrolling through the genes in the operon, they are labeled as “E.coli lactose operon with lacI, lacZ, lacY and lacA genes” and “beta-D-galactosidase” “galactoside permease” and “lac repressor”, all indicating that this operon is required for the full digestion of the sugar lactose. Involved in metabolism, to break sugars into smaller parts to use as building blocks or use to generate ATP.**

1. Do some research using NCBI (<http://www.ncbi.nlm.nih.gov>) and other internet resources to learn about the system. What role does this system have in this organism? Create a hypothesis about what might happen to the organism if these genes were deactivated.

**From Wikipedia, “The lactose permease, which sits in the cytoplasmic membrane, transports lactose into the cell. β-galactosidase, a cytoplasmic enzyme, subsequently cleaves lactose into** [**glucose**](http://en.wikipedia.org/wiki/Glucose) **and** [**galactose**](http://en.wikipedia.org/wiki/Galactose)**. However, it would be wasteful to produce the enzymes when there is no lactose available or if there is a more preferable energy source available, such as glucose. Gene regulation of the *lac* operon was the first genetic regulatory mechanism to be understood clearly, so it has become a foremost example of**[**prokaryotic**](http://en.wikipedia.org/wiki/Prokaryote)[**gene regulation**](http://en.wikipedia.org/wiki/Gene_regulation)**. It is often discussed in introductory molecular and cellular biology classes at universities for this reason.” (Wikipedia, Date accessed, June 10 2014).**

**Hypothesis/Prediction: If this protein was mutated or broken, then the bacterium would not be able to use lactose as an energy source.**

1. Share what you learned with the other groups and write down what you learned about the other systems.

**Various answers**

**Bonus**: Go to RCSB Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>) and search for one of the proteins you found by name. You may not be able to find it for your organism, or at all, but here you can see the structure of the protein determined by protein crystallography. First, view the **Summary** tab. Count the number of beta pleatings and alpha helixes (can also be found under the **Sequence tab**). Are there any ligands associated with this protein? Are there any other units for this protein? If so, name them. Be sure to look at the protein in the **3-D view**.

**Name of protein you investigated**:\_\_ beta-D-galactosidase \_\_\_\_\_\_\_\_\_\_\_\_\_

**Number of Beta-pleated sheets**:\_\_\_\_ 40% beta sheet (78 strands; 410 residues) \_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Number of alpha helix sections**:\_\_\_ 13% helical (22 helices; 138 residues) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Associated ligands?**\_\_Sodium, Magnesium, Dimethyl sulfoxide, D-galactonolactone \_\_\_\_\_

**Additional subunits for protein?\_\_\_**none\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Provide a rough 3-D sketch of your protein: