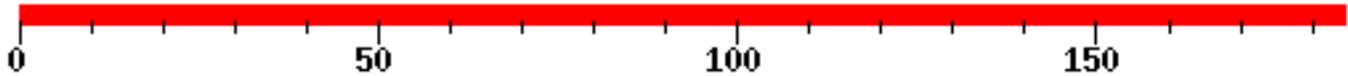# 7.02 BLAST TUTORIAL WORKSHEET

****

1. What is the sequence of Segment 4?

2. Indicate on the ruler bar the location of the *ara C* sequence, *lacZ* sequence of mini Tn *10,* and mini Tn *10* inverted repeat sequence.

3. What is the range of this "subject "sequence? Why does the numbering of the "subject" sequence start high and then go lower?

4. What is the accession number of this sequence?

5. How many base pairs is this sequence? Does the size of the nucleotide sequence correspond to the number of base pairs noted under 'FEATURES?"

6. The *ara C* gene is within this sequence. What is the length of the *ara C* gene? [Look beside "CDS" under "FEATURES."] What is the size of the arabinose C gene (as described in your RDM Lab manual)? Does the length of the *ara C* gene correspond to this size?

7. How many base pairs is the coding region of the arabinose C gene? How many amino acids are in the Ara C protein?

8. Knowing that the coding sequence starts at nucleotide 270 of this sequence, and that your query sequence matches up to 342 of this sequence, how many ara C <u>coding</u> nucleotides are in the translational fusion? How many ara C <u>amino acids</u> are in the fusion?

9 What are the first four sequences that produce significant alignments? What are their scores? Give an example of non-arabinose protein that shows significant homology to the AraC protein that has a score of 50 or greater. From these related proteins it is often possible to deduce the function and mechanism of action for your query sequence. Why might you suspect that AraC carries out its regulatory function by binding DNA?