Computer Science 480 Project 2
Due December 9, 2008 at 10AM

This is a group project. You should assign responsibilities to each member of the group and turn in that list to me by October 30th.

The Background

Dr. Kerry in the Biology department at SMCM studies *Caenorhabditis elegans*, a worm which can be used as a model organism for understanding the affects of various proteins in cells. The genome for *C. elegans* has been sequenced and is available from wormbase.org. As part of her research, Dr. Kerry has obtained several DNA sequences. She is studying the affect of a particular protein on the worm as a whole. She has a specific transcription factor binding site that she's looking for.

The Question

How many times does the specific transcription factor binding site appear in the promoter sequence upstream of a gene?

Answers to More Questions

There won't be any reversals of the transcription factor binding site. We also know exactly what the start and end of the promoter sequence of the gene looks like. If this protein doesn't affect the gene we're searching, then the transcription factor binding site won't appear. A gene which is very affected by the protein will have the transcription factor binding site appear approximately 5 times. Finding the transcription factor binding site more than 10 times is highly unlikely. You should ignore any base pairs beyond 3000 (ie if you go past 3000 base pairs, you've gone too far).

Your Purpose

Develop a program which will answer Dr. Kerry's question. This needs to be a robust program, so that if she starts looking for a different transcription factor binding site on different genes, she can still use your program.

A rubric will be distributed after Thanksgiving Break.