With each question, I want you to submit a small, typed, nicely-formatted report stating what you accomplished, what problems you faced, and how you can make your results better if you had all the time in the world.

1. You will implement algorithms for aligning two sequences at a time. The sequences can be amino acid sequences or nucleotide sequences.

   (a) We want to align two sequences *globally*. Implement the Needleman-Wunsch algorithm. Allow it to be accessible over the Internet.

   (b) Implement the Smith-Waterman algorithm for *local alignment*. Allow access over the Web.

   (c) Compare the results you obtain with the results you get on alignment sites you find on the Web. Are the results the same or different? Why?

   (d) You should allow for searches for both nucleotide sequences and amino acid sequences. Perform checks for correctness of input.

   (e) Allow for choices in the scoring matrix to be used. Find some commonly used matrices on the Web and use them.

   (f) Provide for simple help to users of your site.

   (g) Improve your implementation any way you see fit. Extra credit will be given for substantial extra work.

2. Once again consider the human chromosome 22 that you analyzed for homework assignment 1. This assignment asks you to build a simple database for this chromosome. I know it’s not a database class, but databases are extremely important for bioinformatics and learning how to do simple things with databases.

   **Requirements:** We want the users of our database to be able to search for “genes” in the chromosome. We will allow for searches

   - **By the “name” of a gene:** Make up any names you want for the “genes” you discovered in the chromosome.

   - **By the location of a gene in the chromosome:** Provide location in terms of starting base pair number and end base pair number.
You need to also provide for a simple graphical display for the chromosome and the genes.

To do:

(a) Build a graphical user interface accessible on the Web that allows for the searches discussed above.
(b) Research into how genes are named. Write a paragraph or two on it. Research into how locations are specified in genome databases. Write a paragraph or two on it.
(c) Create a simple database schema for the requirements given above. Create the tables using PostgreSQL on bioinfo.uccs.edu. Use any language of your choice to program. Allow for Web access to your database.
(d) Provide for simple visualization of the database and the results of your searches.
(e) Add anything else want. Extra credit will be given for substantial extra work you do beyond what’s asked.