BSCI 348S – Comparative Bioinformatics

Midterm I, Fall 2001

Answer each question fully and to the best of your ability. Be sure to read the question carefully, and respond to all components of the question without unnecessary elaboration. Excessively lengthy or disorganized responses will be penalized.

1. We have examined three pairwise sequence alignment methods in detail. For each of these, discuss briefly (in the space below) the design of the algorithm, assumptions made by each method, and circumstances under which this software tool would be useful:

a. Needleman-Wunsch

i. Algorithm design (5)

ii. Assumptions (5)

iii. Uses (5)

b. Smith-Waterman

i. Algorithm design (5)

ii. Assumptions (5)

iii. Uses (5)

c. BLAST
i. Algorithm design (5)

ii. Assumptions (5)

iii. Uses (5)

2. Consider the figure below. It is intended to illustrate the advantage of using a "two-hit heuristic" in a recent improvement of BLAST. Explain why the two-hit heuristic is preferable to the one-hit approach (note: don't get hung up on the term "heuristic", which means "an educational procedure," and is used to describe algorithms that are helpful without being guaranteed to find the best possible solution). (10)

[Figure 2 from Nucleic Acids Research 25:3391]

3. What is the difference between *similarity* and *homology*? (10)

4. Consider BLOSUM matrices.

a. What are the two key characteristics of amino acids represented in BLOSUM matrices? (5)

b. Under what circumstances would it be preferable to use a BLOSUM 30 matrix rather than BLOSUM 62 matrix? (5)

5. Consider the *Haemophilus influenzae* genome.

a. Outline the "random" strategy used to sequence this genome.(15)

*b. What are two important advantages of this approach over the map-based approach?
   (5)*

c. What are two key disadvantages of the random approach?(5)

d. Is H. influenzae *free-living? (5)*