Computer Science 4762 (Winter 2002):
Assignment #3
Due: April 9, 2002

1. [Question #2, Assignment 2] (70 marks) Implement (in C, C++, or Java) the dynamic programming algorithm for computing the optimal Sum-of-Pairs (SP) alignment under edit distance, i.e., $smatch = 0$ and $smis = sspace = 1$, of three sequences. This algorithm is described in Section 14.6.1 of Gusfield (1997). You may assume that all input strings will be over the DNA alphabet, e.g., $\{A,C,G,T\}$. Write a main program that takes strings $s_1$, $s_2$, and $s_3$ as input and produces an optimal alignment as output. Run this main program against the following test cases:

1. $s_1 = AACGGA$, $s_2 = TACGGGA$, $s_3 = ACGTAA$
2. $s_1 = GGGGGG$, $s_2 = GAAGCCGAG$, $s_3 = AAACGGA$
3. $s_1 = GGGGGGGG$, $s_2 = ATGGAAGC$, $s_3 = GTGGLCCA$

Hand in your code and copies of your output on these test cases.

2. (30 marks) Trace the execution of the independent base-pairs RNA structure prediction algorithm as described in class, i.e., no pairing of adjacent bases, on the sequence GAGCUCUA relative to the following free-energy functions:

1. Canonical base-pairs, i.e., G-C and A-U, have score -2 and all other base-pairs have score 0.
2. G-C base-pairs have score -2, A-U base-pairs have score -1, and all other base-pairs have score 0.
3. G-C and G-U base-pairs have score -2, A-U base-pairs have score -1, and all other base-pairs have score 0.

Fill in the provided dynamic-programming tables (including cell backpointers) and show any one of the paths of backpointers that indicates an optimal base-pairing as well as the optimal base-pairing corresponding to that path.
Figure 1: Answer to Question 2(a).
Figure 2: Answer to Question 2(b).
Figure 3: Answer to Question 2(c).