COMS4761: ASSIGNMENT 2.2: ezLife MSA

[Questions marked with (*) are more challenging]

Problem 1: Computing MSA

Consider the ezRNA sequences {X1, X2, X3, X4, X5}.

```
X1=00101
X2=01010
X3=01100
X4=10110
X5=10101
```

A.   Compute MSA of these sequences using progressive alignment as follows:
   (i)   Compute pairwise alignments using the scoring matrix of problem 1.
   (ii) Use the Hamming distance $H(X,Y)^1$ to compute a UPGMA Guide Tree.
   (iii) Use the Guide Tree to merge the pairwise alignments into an MSA.
   (iv) *Does this progressive alignment always lead to a unique MSA?
        If so, prove it and if not show an example where multiple MSAs are generated.

B. (15 points) Use the Barton-Sternberg iterative alignment algorithm to improve the MSA of A
   (i)   Go through one round of iterations, sampling once each of the sequences
        {X1,X2,X3,X4,X5} to compute an improved MSA; explain your results.
   (ii) Can the order at which you sample the sequences influence the resulting?
   (iii) *Can you find examples where the iterative technique oscillates between different MSAs
        rather than converge?

---

[1] Normally, the distance metric is computed from the scoring matrix. The Hamming distance is
used here to merely simplify the computations of distances. H(X,Y)=# of symbols (including
indels) which are different in X from Y, e.g., H(0_1010, 00101_)=2.