October 16th, 2023   Name: _____

Question 1. (1 pt.) True or False. Suppose S_1 and S_2 are protein sequences *of the same length*. Then Needleman-Wunch and Smith-Waterman give you the same alignment.

Question 2. (1 pt.) Which of the following substitution matrices exists? (circle one)
   (A) BLOSUM60
   (B) BLOSUM120
   (C) BLOSUM240
   (D) BLOSUM480

Question 3. (3 pts.) Consider the following block. When constructing a BLOSUM50 substitution matrix, various rows of the block must be clustered. Identify the clusters. In other words, how many clusters are there, and specify each cluster by stating which row numbers belong to it?

$$
\begin{array}{c|ccccc}
\text{row } 1 & G & R & S & R & T \\
\text{row } 2 & G & R & S & V & T \\
\text{row } 3 & W & Q & S & V & T \\
\text{row } 4 & W & Q & S & L & S \\
\text{row } 5 & G & T & L & T & L \\
\end{array}
$$

Question 4. (2 pts.) The Markov Chain described by the following transition matrix can produce an infinitely long sequence from the alphabet of three letters $\{A, B, C\}$. True or False, this sequence can have infinitely many C's.

$$
\begin{array}{c c c c}
 & A & B & C \\
A & \left[\begin{array}{ccc} 0.3 & 0.5 & 0 \\ \end{array}\right. \\
B & \left.\begin{array}{ccc} 0.1 & 0.6 & 0 \\ \end{array}\right. \\
C & \left.\begin{array}{ccc} 0.2 & 0.7 & 0.1 \\ \end{array}\right] \\
\end{array}
$$

Question 5. (3 pts.) Suppose the following block of data has been clustered into four clusters, so that the first four sequences comprise one cluster and the bottom three are each in their own cluster. So one cluster of size 4 and three clusters of size 1.

Count the number of times an A aligns with a D for the goal of constructing a BLOSUM matrix. So you have to count properly taking into account the clusters, as is done in the BLOSUM method.

$$
\begin{array}{cccc}
A & B & A & A \\
A & B & A & A \\
A & B & A & A \\
A & D & A & A \\
A & A & B & D \\
A & C & B & A \\
D & A & B & A
\end{array}
$$

Answer: _____