| Bioinformatics | (Akiyama) | Exercise #07 |
|----------------|-----------|--------------|
|                |           |              |

1) From the sequences (a)-(e) shown below, choose <u>all</u> sequences which agree with the following sequence motif.

Name

[ ] is disjunctive OR. Any one element in [ ] can be matched.

x(a,b) is a series of any spacing characters, at least *a* and up to *b* characters.

[C G] - A - [C A] - x(1,3) - A - G - [A T]

Sequences :

Student ID

(a) CACTAAGT(b) GAACTAAGA(c) CACACAGA(d) CACCTAGAGA(e) GAACACAGT

Ans.

2) The following PSSM (Position Specific Score Matrix) represents a DNA sequence motif. Show a DNA sequence which gets the highest score with this PSSM, and its highest score. Additionally, show a DNA sequence for lowest score, and the lowest score.

|   | position |    |    |    |    |  |
|---|----------|----|----|----|----|--|
|   | 1        | 2  | 3  | 4  | 5  |  |
| А | 6        | -3 | -3 | 0  | -3 |  |
| С | -9       | 0  | -5 | -3 | 6  |  |
| G | -3       | 7  | -4 | -7 | 0  |  |
| Т | 2        | -4 | 0  | 0  | -4 |  |

highest score sequence : highest score value :

lowest score sequence : lowest score value :

Ans.

Calculate relative entropy score for the four sequences shown below.
Sequence motif length L is 8.

Relative Entropy Score = 
$$\sum_{j=1}^{L} \sum_{a \in \Sigma} f_j(a) \cdot \log_2 \frac{f_j(a)}{p(a)}$$

Here, background probabilities are defined as p(A)=p(G)=p(T)=p(C) = 0.25. Note that  $0 \log 0 = 0$ .

Sequences

|                 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----------------|---|---|---|---|---|---|---|---|
| <b>Seq. 1</b> : | G | Α | т | Α | т | G | Α | Α |
| Seq. 2 :        | Т | Α | Т | Т | т | G | т | т |
| Seq. 3 :        | Α | Α | Α | G | т | С | С | С |
| Seq. 4 :        | С | Α | Α | С | т | С | G | С |

Score :

Score Sum :

Show partial score for each column. And then calculate the score sum.

Ans.