## Student ID <br> Name

1) From the sequences (a)-(e) shown below, choose all sequences which agree with the following sequence motif.
[ ] is disjunctive OR. Any one element in [ ] can be matched.
$\mathrm{x}(a, b)$ is a series of any spacing characters, at least $a$ and up to $b$ characters.
$[\mathrm{C} \mathrm{G}]-\mathrm{A}-[\mathrm{C} \mathrm{A}]-\mathrm{x}(1,3)-\mathrm{A}-\mathrm{G}-[\mathrm{A} \mathrm{T}]$

Sequences:
(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 |
| A | 6 | -3 | -3 | 0 | -3 |
| C | -9 | 0 | -5 | -3 | 6 |
| G | -3 | 7 | -4 | -7 | 0 |
| T | 2 | -4 | 0 | 0 | -4 |

highest score sequence :
highest score value :
lowest score sequence :
lowest score value :
Ans.
3) Calculate relative entropy score for the four sequences shown below.

Sequence motif length $L$ is 8 .

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

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

Sequences
$\begin{array}{llllllll}1 & 2 & 3 & 4 & 5 & 6 & 7 & 8\end{array}$


Score: Score Sum :

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

Ans.

