

2015

BIOCHEMISTRY**Paper – BCT-103****(Bioinformatics)****Full Marks – 25***The figures in the margin indicate full marks**Candidates are required to give their answers in their own words as far as practicable*Answer **any twelve** questions

1. Which MATLAB statement is correct for obtaining a vector where the elements are formed using

$$y = \cos(x) + \frac{\sin(x)}{4 + \cos(x)}$$

where x has been defined as $x = (0 : 0.01 : 6)$?

- (a) $y = x * \cos(x) . + \sin(x) / (4 + \cos(x))$
 (b) $y = x .* \cos(x) + \sin(x) ./ (4 + \cos(x))$
 (c) $y = x .* \cos(x) + \sin(x) / (4 + \cos(x))$
 (d) $y = x * \cos(x.) . + \sin(x.) / (4 + \cos(x))$

2. What is the value of **d** after the following MATLAB commands are executed:

```
a = [1 2; 3 4];
b = [0 1; 2 1];
c = a.*b;
d = b*c;
```

- (a) $d = [0 \ 1; \ 12 \ 4]$
 (b) $d = [6 \ 4; \ 6 \ 8]$
 (c) $d = [0 \ 3; \ 16 \ 7]$
 (d) $d = [4 \ 2; \ 8 \ 10]$.

3. Consider the following MATLAB function: function output = midterm (z)
 $x = z(1)$; $y = z(2)$; $w1 = x*x + y*y$; $w2 = x$; output = [w1 ; w2]. When
 $z = (2 \ 1)$, what is the value of output?

- (a) [5 ; 2]
 (b) [5 ; -1]

[Turn Over]

(c) [5 2]

(d) [5 -1]

4. If ACC is the accession number of a given protein and we write $S = \text{getgenpept}(S)$, the one line command to find out the percentage x of lysine in the amino acid will be :

2

(a) $x = 100 * S.\text{Accession.K}$

(b) $x = 100 * \text{aaccount}(S).\text{K}$

(c) $g = \text{aaccount}(S)$; $L = \text{length}(S.\text{Sequence})$; $x = g.K/L$

(d) $x = 100 * (\text{length}(\text{find}(S.\text{Sequence} == 'k')))/\text{length}(S.\text{Sequence})$

5. Assume that there are two vectors $a = [0 \ 0 \ 1]$; $b = [1 \ 0 \ 0]$; The cross product of the two vectors will be :

2

(a) [0 1 0]

(b) 0

(c) [1 0 1]

(d) [0 0 0]

6. If '1HHO' is PDB ID of a protein, the expressions $z = \text{getpdb}('1HHO')$; $s1 = z.\text{Sequence}(1).\text{Sequence}$; $s2 = z.\text{Sequence}(2).\text{Sequence}$, null value of $\text{strcmp}(z.\text{Sequence}(2).\text{Sequence}, z.\text{Sequence}(1).\text{Sequence})$ implies :

2

(a) Sequences are non-identical

(b) There is only one chain in the protein

(c) The protein chains have no homology

(d) None of these.

7. Ubiquitin is a small protein of approximately 76 amino acids, found in all eukaryotic cells and very well conserved among species. We can call the structure of the protein using $\text{ubi} = \text{getpdb}('lubi')$; The protein can be seen in molviewer using the command $\text{h} = \text{molviewer}(\text{ubi})$; Now consider the following script : $\text{evalrasmolscrip}(\text{hl}, ['\text{measure}' \text{ num2str}(\text{initHelix}) \text{ '}' \text{ num2str}(\text{endHelix}) \text{ '}' \text{ '}]$; Indicate whether it measures :

2

(a) length of all possible helices

(b) end points of the first helix

(c) length of the first helix in angstrom unit

(d) length of all helices in angstrom unit.

8. Consider a curve fitting problem. Assume $\text{coef} = \text{polyfit}(x, y, 4)$; $\text{coef}(2)$ is the coefficient of which power of x ?

2

(a) 2

(b) 3

(c) 1

(d) 0

9. Fplot command in matlab plots a function . The x - $\sin(x)$ plot in the limit $x = -\pi$ to $+\pi$ will be `fplot(@(x) [-pi + pi])`. Fplot can be used to plot coulomb and van der waals potentials using the distance range ($0 < r, 0.001$) and taking all constants as unity :

- `fplot(@(x)1/(x^2), [0, 0.001]); fplot(@(x)(1/x^12 - 1/x^6), [0, 0.001])`
- `fplot(@(x)1/(x), [0, 0.001]); fplot(@(x)(1/x^12 - 1/x^6), [0, 0.001])`
- `fplot(@(x)1/(x), [0, 0.001]); fplot(@(x)(1/(x^6 - 1/x^12)), [0, 0.001])`
- None of these

10. Write a program to test Chargaff's rule for a randomly generated DNA and an actual DNA (say the mitochondrial). The rule states that :

- $\frac{A+T}{G+C} = 1$
- $\frac{A}{T} = \frac{G}{C} = 1$
- $\frac{A}{G} = \frac{C}{T} = 1$
- $\frac{A+T}{A+T+G+C} = \frac{G+C}{A+T+G+C} = 1$

11. If z is a matrix expressing an image the command to generate a matrix inverse to z will be :

- `w = inv(double(z))`
- `w = inv(uint8(z))`
- `w = inv(z)`
- None of these

12. The command to generate the histogram of the red plane of a colored image z is given by :

- `imhist(red(z))`
- `imhist(z(:,:,1))`
- `imhist(z(:,1))`
- `hist(z(:,:,1))`

13. Which of the following cases are commonly used ?

- gap opening penalty = -2 , gap extension penalty = -0.5
- gap opening penalty = -0.5 , gap extension penalty = -2.0
- gap opening penalty = -100 , gap extension penalty = 0
- gap opening penalty = -100 , gap extension penalty = -100 .

[Turn Over]

14. Which is the default scoring matrix used in BLAST ? 2
(a) PAM62 (b) BLOSUM 62
(c) BLOSUM 60 (d) BLOSUM 80
15. PAM matrices are derived by noting evolutionary changes in protein sequences that are more than : 2
(a) 80 (b) 60
(c) 40 (d) 25
16. Which alignment is used to predict whether two sequences are homologous or not ? 2
(a) Local (b) Global
(c) Pair-wise (d) Multiple
17. In Needleman Wunsch algorithm of pairwise alignment of sequences with lengths n and m , the computational time is proportional to : 2
(a) $n \times m$ (b) $(n + 1) \times (m + 1)$
(c) $n + m$ (d) $n \times (m + 1)$
18. What is PROSITE ? 2
(a) a database of protein structures
(b) a database of interacting proteins
(c) a database of protein motifs
(d) a search tool
19. If you want literature information, which is the best website to visit? 2
(a) OMIM (b) Entrez
(c) PubMed (d) PROSITE
20. You have two distantly related proteins. Which of the following sets is the best for comparing them ? 2
(a) BLOSUM45 or PAM250
(b) BLOSUM45 or PAM1
(c) BLOSUM80 or PAM250
(d) BLOSUM80 or PAM1
21. The Smith-Waterman algorithm was developed for 2
(a) Local pairwise sequence alignment
(b) Global pairwise sequence alignment
(c) Multiple sequence alignment
(d) Structural alignment.

[Neatness – 1]
