KENYATTA UNIVERSITY

UNIVERSITY EXAMINATIONS 2011/2012

SECOND SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE IN MOLECULAR AND CELL BIOLOGY

SBC427 / SBC251: FUNDAMENTALS OF BIOINFORMATICS

DATE: TUESDAY, 27TH MARCH 2012 **TIME: 8.00 A.M. – 10.00 A.M.**

SECTION A: Answer all questions.

1. Which of the following is NOT a symbol that you would find in a standard nucleotide sequence?

(A) Y

(B) S

(C) O

(D) X

2. Which of the following symbols represents Glutamate in a polypeptide sequence?

(A) B

(B) G

(C) D

(D) E

3. In a PDB database, three of the following information can be found in the PDB sequence file header. Which one is unlikely to be included in the file format?

(A) Disulphide bonds present in the protein

(B) Solvation properties of the protein (i.e. how the protein interacts with water molecules)

(C) The hydrogen atoms in the protein

(D) The name of the people who resolved the protein structure by x-ray crystallography

4. The process of receiving, reviewing, accepting and appropriately storing incoming sequences by a dedicated member of staff is called

(A) Curation
(B) Submission
(C) Querying
(D) Alignment

5. How do we refer to the process of adding information tags to a sequence in a databank, for example highlighting the coding sequence or indicating where the promoter is located within a submitted sequence?

(A) Summation
(B) Allocation
(C) Annotation
(D) Distinction

6. Which of the following databases is NOT curated?

(A) TrEMBL
(B) SWISS PROT
(C) DDBJ
(D) PIR

7. Which of the following is not a type of sequence alignment?
(A) Focal alignment
(B) Global alignment
(C) Multiple alignment
(D) Pair-wise alignment

8. Which of following is not a motif that can be used by a gene finding software to predict a gene?

(A) The -35 sequence of a promoter
(B) A splicing consensus sequence

(C) The nucleotides coding for the first 20 amino acids in the polypeptide chain
(D) The ATG codon

9. A substitution mutation in the nucleotide sequence that does not lead to a change in amino acid sequence or structure of the encoded polypeptide is said to be

(A) Reversible
(B) Homologous
(C) Synonymous
(D) Anonymous

10. Which algorithm creates a global alignment between two nucleotide sequences?

(A) Markov Chain Monte Carlo
(B) Smith-Waterman's
(C) Blossum
(D) Needleman-Wunsch

11. The different conformations of the side chain of an amino acid in a polypeptide during computer protein design are called
(A) Scaffold
(B) Rotamers
(C) Ligand
(D) Docking
12. What is an intractable algorithmic computer simulation?
(A) One that requires very high levels of processor speeds to complete in an reasonable time
(B) One that requires infinitesimal number of calculations that cannot be achieved in rational amount of time using the existing computer technologies
(C) One that cannot be performed by a stand alone desktop PC
(D) One that causes the computer to crash

13. Colors can be used in the representation of a folded protein to show

(A) The active sites
(B) The hydrophobic sites

(C) Hydrophilic sites

(D) All of the above

14. In molecular computer graphics, cylinders normally represent

(A) The polypeptide bonds of the protein

(B) The alpha helix of protein structures.

(C) The N terminal of the proteins

(D) The beta sheets of protein structure

15. In drug design by computer simulation, which of the following is not true?

(A) One can search a chemical library for possible new hit compounds

(B) One can simulate a docking interaction to show how ligands bind

(C) One can determine most of the potential side effects of the new drug

(D) One can identify closely related drug compounds

16. The Kimura parameter 2 correction for genetic distance takes into account

(A) Different rates of evolutionary accumulation of mutations

(B) Differences in rates of transversion mutations

(C) Different rates of duplications

(D) Differences in rates of reversion mutations

17. In nature, we rarely find very large proteins made of a single polypeptide. Instead, we encounter _____ structures made of several sub units bound together by non-covalent interactions.

18. Nexus is a human readable file format. True or false? _____

19. Some protein databases basically contain amino acid translations from primary nucleotide databases. True or false? _____

20. Gaps in sequences being aligned normally represent areas of insertions and/or _____ kinds of mutations.

SECTION B: All questions in this section are compulsory and each question totals to 5 marks.

Question 1

(a). List three forms of representation of biomolecules used in molecular graphics software (3marks)

(b). Give the colors that are frequently used to represent any 2 specific atoms in molecular graphics (2 marks)

Question 2

Explain assumptions of flexibility or rigidity made during computer docking simulations between two macromolecules and the importance of such assumptions (5 marks).

Question 3

(a) What are the main advances in computer hardware components that have enabled major advances in computer applications in biomedical sciences? (3 marks)

(b) Give any two examples of applications where computers are indispensable in the management of huge volumes of biological data (2 marks)

Question 4

(a). Explain the difference between an accession number and an identifier in a sequence file from a sequence database? (3 marks)

(b). Using an imagined sequence of about 20 bp, describe the popular FASTA (Pearson) format for sequence files. (2 marks).

Question 5

(a). Discuss 3 factors that are important to consider during the design of primers (3 marks).

(b). Give two examples of primer design software (2marks).

Question 6.

(a) Name the three main databases for DNA sequence repository (1.5 marks).

(b) List three kinds of web-based resources that we can retrieve using the Entrez system of the NCBI. (1.5 marks).

(c) Use two points to explain SWISS-PROT (2 marks).

SECTION C: Select and answer any two questions only in this section. A single question may be divided into parts numbered a, b, c, d.

Question 1.

(a). Discuss the non-covalent bonds that we would expect in a protein quaternary structure and that facilitate docking (8 marks).

(b). Give two examples of natural interactions that can be computer-simulated as docking interactions (2 marks).

Question 2.

(a). Outline the traditional "pipeline" for drug discovery and development (4 marks).

(b). Then explain how computer simulations can facilitate drug discovery and development (4 marks).

(c). Give two weakness of the Computer aided drug design and development (2 marks).

Question 3.

With examples, discuss the role of motifs in pattern searches and gene predictions in molecular

****