BIF401 Mid Term Solve Subjective

Welcome To vuways Study Help

BIF401 Mid Solve

1. **What is the bioinformatics promise to the society? (2 marks)**

Bioinformatics is a multidisciplinary science that solves and analyzes biological problems. With the quantum explosion in biomedical data, the demand of bioinformatics has increased gradually. Present paper provides an overview of various ways through which the biologists or biological researchers in the domain of neurology, structural and functional biology, evolutionary biology, clinical science, etc., use bioinformatics applications for data analysis to summarise their research. A new perspective is used to classify the knowledge available in the field thus will help general audience to understand the application of bioinformatics.

2. **What dose role of 5' cap and 3' role A poly tail? (2)**

The 5′ cap protects the nascent **mRNA** from degradation and assists in ribosome binding during translation. A poly (A) tail is added to the 3′ end of the pre-**mRNA**once elongation is complete.

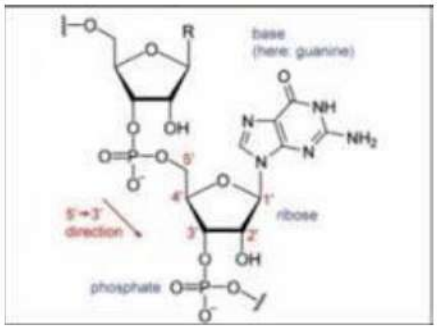
3. **What is dot plots?**

In bioinformatics a dot plot is a graphical method for comparing two biological sequences and identifying regions of close similarity. It is a type of recurrence plot.

4. **What does a Uniprot sequences give any information about protein query?**

5. **Write down step of BLOSUM?**

Step 1: Eliminate sequences that are identical in x% positions Step 2: Compute observed frequency f i,j of aligned pair Ai to Aj . Hence, fi,j becomes the probability of aligning Ai and Aj in the selected blocks. Step 3: Compute fi which is the frequency of observing Ai in the entire block



6. **Hydroxyl group is less stable in RNA as compared to DNA? Explain it with a figure?**

RNA.... Unlike **DNA**, **RNA** in biological cells is predominantly a single-stranded molecule. While**DNA** contains deoxyribose, **RNA** contains ribose, characterised by the presence of the 2'-**hydroxyl group** on the pentose ring (Figure 5). This **hydroxyl group** make **RNA less stable** than **DNA** because it is more susceptible to hydrolysis ...

7. **Diffrentiate between acid and basic amino acids?**

There are three **amino acids** that have **basic** side chains at neutral pH. These are arginine (Arg), lysine (Lys), and histidine (His). Their side chains contain nitrogen and resemble ammonia, which is a**base**. Their pKa's are high enough that they tend to bind protons, gaining a positive charge in the process.

8. How bulge loop formed?

Bulges, are formed when a double-stranded region cannot form base pairs perfectly. • Bulges can be asymmetric with varying number of base pairs on one side of the loop.

9. Differentiate between mass spectrometry and finger printing spectrometer?

Mass spectrometry (MS) is an analytical technique that ionizes chemical species and sorts the ions based on their**mass**-to-charge ratio. In simpler terms, a **mass** spectrum measures the masses within a sample. **Mass spectrometry** is used in many different fields and is applied to pure samples as well as complex mixtures.

Fingerprint refers to the characteristic **spectrum** or image of a test material which can be related to its properties and thus to its authenticity in the same way as a human **fingerprint** is specific of a certain person and unequivocally identify him/her.Oct 7, 2017

10. How tertiary structure of RNA formed?

Nucleic acid tertiary structure is the three-dimensional shape of a **nucleic acid polymer**. RNA and **DNA molecules** are capable of diverse functions ranging from molecular recognition to catalysis. Such functions require a precise three-dimensional tertiary structure.

11. What are pseudoknots?

A **pseudoknot** is a nucleic acid secondary structure containing at least two stem-loop structures in which half of one stem is intercalated between the two halves of another stem.

12. What is the purpose of 3D-1D bowle algorithm? PPT

Algorithm And Examples Fold Recognition/Threading Online Tools for Fold Recognition GOR Algorithm Homology Modelling 3D-1D Bowie Algorithm Machine Learning Approaches to Structure Prediction Neural Networks for Structure Prediction PSIPRED Introduction to Hidden Markov Models Ab initio modeling

13. What is needs of bioinformatics?

The primary goal of **bioinformatics** is to increase the understanding of biological processes .....Common activities in **bioinformatics** include mapping and analyzing DNA and protein sequences, aligning DNA and protein sequences to compare them, and creating and viewing 3-D models of protein structures.

14. Uses of blastx, tblastx and tblastn?

• blastn (nucleotide BLAST) • blastp (protein BLAST) • blastx (translated BLAST) • tblastn (translated BLAST) • tblastx (translated BLAST)

15. Steps involved in FASTA algorithm?

Step 1: Local regions of identity are found. Step 2: Rescore the local regions using PAM or BLOSUM matrix. Step 3: Eliminate short diagonals below a cutoff score. Step 4: Create a gapped alignment in a narrow segment and then perform Smith Watermann alignment

16. Formula of domain and descibeit's method?

In mathematics, and more specifically in naive set theory, **the domain** of **definition** (or simply **the domain**) of a **function** is **the** set of "input" or argument values for ...

17. How to show phylogenetic tree describe the types to visuilaizae it?

A phylogenetic tree or evolutionary tree is a branching diagram or "tree" showing the..... This typeof tree only represents a branching pattern; i.e., its branch spans do not represent time or relative amount of character change.....Exploration This is a programming library to analyze, manipulate andvisualize phylogenetic trees.

18. Difference between .....Peptide and unshot protein

Structurally, **proteins** and **peptides** are very similar, being made up of chains of amino acids that are held together by **peptide** bonds (also called amide bonds). So, what distinguishes a **peptide** from a**protein**? The basic distinguishing factors are size and structure. **Peptides** are smaller than **proteins**.

19. Why we use FASTA algorithem?

FASTA takes a given nucleotide or amino acid sequence and searches a corresponding sequence database by **using** local sequence alignment to find matches of similar database sequences. The **FASTA** program follows a largely heuristic **method** which contributes to the high speed of its execution.

20. What is the acronym meaning/definition of BLAST?

BLAST (basic local alignment search tool) is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA and/or RNA sequences.

21. Threonine is an hydrophobic amino acid?

Polar amino acids include serine, **threonine**, **asparagine**, glutamine, histidine and tyrosine.
The **hydrophobic** amino acids include alanine, valine, leucine, **isoleucine**,**proline**, **phenylalanine**, tryptophane, cysteine and **methionine**.

22. How does protein is sequenced?

Protein sequencing is the practical process of determining the amino acid **sequence** of all or part of a **protein** or peptide..... The two major direct methods of **protein sequencing are** mass spectrometry and Edman degradation using a**protein** sequenator (sequencer).

23. What is the uses and functions of blast?

In bioinformatics, BLAST (basic local alignment search tool) is an **algorithm** for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA and/or RNA sequences.

24. How does protein sequenced is identify?

Predicted **protein sequences are** an important resource for **protein identification**by mass spectrometry..... The **sequence** of the cloned DNA was then determined and used to deduce the full amino-acid **sequence** of the **protein**.

25. What are Substitutions & Indel?

Indel.....Indels can also be contrasted with Tandem Base Mutations (TBM), which may result from fundamentally different mechanisms. A TBM is **defined** as a**substitution** at adjacent nucleotides (primarily **substitutions** at two adjacent nucleotides, but **substitutions** at three adjacent nucleotides have been observed.

26. Why unpaired nucleotides should make the Structure of RNA destabilized?

27. Write formula for finding distance between two Sequences and elaborate it.

28. Top down proteomics?

Measures intact proteins followed by their peptides after fragmentation.

29. 7 Computed each possible alignment?

30. Usage of sequence data?

Sequence data can be used to obtain: • Similarity of sequences • Evolutionary History • Predict the function of molecules

31. Challenges in the field of bioinformatics.

Bioinformatics is full of challenges and opportunities. Amongst other frontiers in bioinformatics, there are protein structures, systems biology and personalized medicine.

32. Give diff b/w Rooted and uprooted also write its uses 5 marks

In a phylogenetic tree, every node represents a species. A **rooted tree** is a **tree in** which one of the nodes is stipulated to be the **root**, and thus the direction of ancestral relationships is determined.
An **unrooted tree**, as could be imagined, has no pre-determined **root** and therefore induces no hierarchy.Jan 1, 2001

33. Difference b/w uniprot & EMBL?

UniProt is the Universal Protein resource, a central repository of protein data created by combining the Swiss-Prot, **TrEMBL** and PIR-PSD databases.....UniProt is a freely accessible database of protein sequence and functional information, many entries being derived from genome sequencing projects.

34. About amino acids and protein which has 3 codons?

The **three** stop **codons** have **names**: UAG is amber, UGA is opal (sometimes also called umber), and UAA is ochre. Stop **codons** are also called "termination" or "nonsense" **codons**.

35. Scoring Matrix & NJ algorithm?

Scoring matrices are used to determine the relative**score** made by matching two characters in a sequence alignment.....There are many flavors of **scoring matrices**for amino acid sequences, nucleotide sequences, and codon sequences, and each is derived from the alignment of "known" homologous sequences.

The neighbor-joining **method** (**NJ**) is a distance based **method** (requires a distance matrix) and uses the star decomposition **method**. **Algorithm**. Neighbor-joining is a recursive **algorithm**.

36. About bioinformatics tools?

Tool	Description
BLAST	It is a search tool, used for DNA or protein sequence search based on identity.
HMMER	Homologous protein sequences may be searched from the respective databases using this tool.
Clustal Omega	Multiple sequence alignments may be performed using this program.
Sequerome	Used for sequence profiling.
ProtParam	Used to predict the physico-chemical properties of proteins.
JIGSAW	To find genes, and to predict the splicing sites in the selected DNA sequences.
novoSNP	Used to find the single nucleotide variation in the DNA sequence.
ORF Finder	The putative genes may be subjected to this tool to find Open Reading Frame (ORF).
PPP	Prokaryotic promoter prediction tool used to predict the promoter sequences present up-stream the gene
Virtual Foorprint	Whole prokaryotic genome (with one regular pattern) may be analysed using this program along with promoter regions with several regulator patterns.
WebGeSTer	This is a database containing sequences of transcription terminator sequences and is used to predict the termination sites of the genes during transcription.
Genscan	Used to predict the exon-intron sites in genomic sequences.

Softberry Tools	Several tools are specialized in annotation of animal, plant, and bacterial genomes along with the structure and function prediction of RNA and proteins.

What is the use of blast & its performance?

It demands high computational resources and this demand keeps increasing.**BLAST uses** a similarity searching heuristic that determines sequences from a database that are most similar to a query sequence. We performed all experiments using a set of queries containing genes or amino acid sequences.

37. How Gibbs energy fold RNA Secondary structure?

38. Difference between global and local alignment?

If you think your sequence is a subsequence of the reference, do a local alignment. But if you think your entire sequence should match your entire reference, you would do a global alignment.

39. Progressive alignment?

Progressive alignment (Feng and Doolittle, 1987) is a heuristic for multiple sequence **alignment** that does not optimize any obvious **alignment** score. The idea is to do a succession of pairwise alignments, starting with the most similar pairs of sequences and proceeding to less similar ones.

40. HOW BULGE ARE FORMED?

Bulges, are formed when a double-stranded region cannot form base pairs perfectly. • Bulges can be asymmetric with varying number of base pairs on one side of the loop.

41. MSA?

A multiple sequence alignment is a sequence alignment of three or more biological sequences, generally protein, DNA, or RNA. In many cases, the input set of query sequences are assumed to have an evolutionary relationship by which they share a linkage and are descended from a common ancestor.

42. PSA?

Prediction of probable secondary structures and fold-class; good for visualizing amphipathic helices, where present.

43. Define system biology and its application?

Systems biology is **the computational** and mathematical modeling of complex **biological systems**. **It** is a **biology**-based interdisciplinary field of study that focuses on complex interactions within**biological systems**, using a holistic approach (holism instead of **the** more traditional reductionism) to **biological** research.

44. Write the methods of clustering?

Cluster analysis or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense) to each other than to those in other groups (**clusters**).....**Clustering** can therefore be formulated as a multi-objective optimization problem.

45. Why RNA has short life span? Illustrateur answer with diagram?

This kind of **RNA** is considered **short** lived. In contrast, **RNA** that is crucial to basic cellular processes like cell structure should stick around in cells for a long time to carry out its function

46. Write the four objectives of the comparing sequencing?

Sequence homology is the biological homology between DNA, RNA, or protein sequences, defined in terms of shared ancestry in the evolutionary history of life. Several conclusions can be drawn on a set of sequences in terms of: • Similarity • Specific Differences • Relationship • Evolutionary Insight

47. Purines and pyrimidins structures?

Purines vs. **Pyrimidines**. **Purines and Pyrimidines** are nitrogenous bases that make up the two different kinds of nucleotide bases in DNA and RNA. The two-carbon nitrogen ring bases (adenine and guanine) are **purines**, while the one-carbon nitrogen ring bases (thymine and cytosine) are**pyrimidines**.

48. Benefits of dot plot?

Data points may be labelled if there are few of them.**Dot plots** are one of the simplest statistical **plots**, and are suitable for small to moderate sized data sets. They are useful for highlighting clusters and gaps, as well as outliers. Their other **advantage** is the conservation of numerical information.

49. EMBL and EBI stands for?

The European Bioinformatics Institute (**EMBL-EBI**) is an IGO which as part of the European Molecular Biology Laboratory (**EMBL**) family focuses on research and services in bioinformatics.

50. Different field of expasy?

Expasy provides access to a variety of online databases and tools. • Depending upon your requirement, you find sequence information from Expasy.

51. Describe the steps of PAM?

Steps to compute PAM matrices

Step 1: Align proteins sequences which are 1-PAM unit diverged

Step 2: Let $A_{i,j}$ be the number of times A_i is substituted by A_j

Step 3: Compute the frequency f_i of amino acid A_i

Then, PAM1 = p_{ij} = $\frac{A_{ij}}{\sum_k A_{ik}}$

PAM'n' = (PAM1) ^{n}

52. Hairpin structure of RNA?

hairpin loop (mRNA) A hairpin loop is an **unpaired**loop of messenger RNA (mRNA) that is created when an mRNA strand folds and forms base pairs with another section of the same strand. The resulting structure looks like a loop or a U-shape. Hairpins are a common type of secondary structure in RNA molecules.

53. Application of bioinformatics?

Genomics • Transcriptomics • Proteomics • Metabolomics • Structural Proteomics • Drug Design • Systems Biology • Personalized Medicine

54. PCR, Electrophoresis and Dynamic programming?

PCR (**polymerase chain reaction**): **PCR** (**polymerase chain reaction**) is a technique in molecular genetics that permits the analysis of any short sequence of DNA (or RNA) even in samples containing only minute quantities of DNA or RNA.**PCR** is used to reproduce (amplify) selected sections of DNA or RNA for analysis.

Gel **electrophoresis** is a technique commonly **used in** laboratories to separate charged molecules like DNA', RNA' and proteins' according to their size.....The movement of charged molecules is called migration. Molecules migrate towards the opposite charge.

Dynamic programming is both a mathematical optimization method and a computer programming method. The method was developed by Richard Bellman in the 1950s and has found applications in numerous fields, from aerospace engineering to economics.

55. DNA, RNA and Protein sequences can be aligned using Needleman Wunsh algorithm?

DNA, RNA and Protein sequences can be aligned using Needleman Wunsh algorithm

56. Local alignment focuses on the sub regions within the sequences. Global alignment can full length?

57. Modification in trace back strategy can help find overlap matches containing leading or trailing sequence?

Yes: Trace back can also be started from high score on the right column and bottom row. • Such trace backs helps find overlap matches containing leading or trailing sequence.

58. CLUSTALW can run Fast and slow modes?

Slow

59. BLAST can search sequence databases and identify unknown sequences by comparing them to the known sequences?

60. Multiple types of FASTA exist which assist in aligning DNA/RNA/Protein sequences (Fasta)

61. Structural information can be get by X-ray Crystallography and NMR?

Structures are obtained from X-Ray Crystallography, Atomic Force Microscopy & Nuclear Magnetic Resonance Spectroscopy. Protein Databank 73.

62. Protein Sequencing can be done b?

Next Generation Sequencers

63. Object based method is?

Least Square Distances

64. Not Clusturing Approch Other than?

UPGMA MPGMA, Nieghbor Joining Single Linkage Complete Linkage

65. Define mutation

Mutations are:

- Replacement of amino acids with certain other amino acids in proteins
- Replacement of nucleotides with certain other nucleotides in DNA/RNA

66. Databases use for proteins?

uniport and swisport

67. Generally used databases and which type of information they store?

Genebank. Pdb, uniport swisport

- • Locus
- Accession number
- • Sequence
- • Molecular mass
- Authors
- Journal etc.

68. Types of RNA and their functions?

RNA can be divided into two categories

- Coding rnas
- Non-Coding rnas
- Coding rnas as is obvious from their name, code for Proteins
- Non-Coding rnas regulate/assist in the process of translation

Types of RNA

1. Messenger RNA (mrna) Messenger RNA (mrna) carries the genetic information copied from DNA in the form of a series of three-**base**code "words," each of which specifies a particular amino acid

2. Transfer RNA (trna) is the key to converting the code words in mrna.

3. Ribosomal RNA (rrna) associates with a set of proteins to form ribosomes. These complex structures, which physically move along an **mrna** molecule, catalyze the assembly of **amino acids** into protein chains. They also bind trnas and various accessory molecules necessary for **protein synthesis**.

4. Micro rnas (mirna)

Functions in RNA silencing and post-transcriptional regulation of gene expression.

5. Small Interfering RNA (sirna)

Small (or short) interfering **RNA** (**sirna**) is the most commonly used **RNA** interference (**rnai**) tool for inducing short-term silencing of protein coding genes.**sirna** is a synthetic **RNA** duplex designed to specifically target a particular mrna for degradation.

70. Uses of clustalw?

CLUSTALW can:
• create multiple alignments,
• optimize existing alignments,
• profile analysis &
• create phylogenetic trees

71. Uniprot swisprot

• • Protein Sequences are reported from sequencing experiments

• This data is stored in protein sequence databases
• The famous ones include uniprot & swissprot

72. Scoringatrix formula?

S(a,b)=1/λlog pab /fafb

73. Pam matric steps5"

Step 1: Align proteins sequences which are 1-PAM unit diverged
Step 2: Let Ai,j be the number of times Ai is suabstituted by Aj
Step 3: Compute the frequency fi of amino acid Ai
Then, PAM1 = pij =
PAM'n' = (PAM1) n

74. Name three hydrophobic amino acid?

glycine (Gly), alanine (Ala), valine (Val), leucine (Leu), isoleucine (Ile), proline (Pro), phenylalanine (Phe), methionine (Met), and tryptophan (Trp).

75. **Dynammic programming to create scoring matrix?R 1 series sequence dia hua tha scoring matrix maloom krny k liy?** Ans: From the large number of possible nucleotide combinations, it is hard to find the optimal one

• Dynamic Programming (DP) helps break the problem into smaller problems
Select a nucleotide from sequence
List all possible complementary position for the selected nucleotide in the complete sequence.
How DP works?
• DP exhaustively makes all possible nucleotide combinations
• DP then recombines such combinations in a process called "Traceback" to ensure that the highest coupled 2' structure is reported

76. Complete the protein sequence .F-x(3)-x-R-F-K-x(4-5)-D-E-R?

FXXXXRFKXXXXDER
FXXXXRFKXXXXXDER

77. Clustering?

Clustering is used to build groups of genes with related expression patterns