MUHAMMAD IMRAN	BIF401-Bioinformatics-I
PAST PAPER SOLVED	MID TERM

Rooted and Un-rooted phylogentic tree?

Rooted and Unrooted trees can be used to show phylogenetic relationships between sequences • Several types of algorithms exist which are divided into two classes.

Purpose of bowie algorithm?

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. **(PPT)**

Complete protein sequence? F-x(3)-X-R-F-K-X (4-5) –D-E-R

FXXXXRFKXXXXXDER (not sure)

Enlist salient Feature of block diagram of mass spectrum?

A mass spectrum is a plot of the ion signal as a function of the mass-to-charge ratio. These spectra are used to determine the elemental or isotopic signature of a sample, the masses of particles and of molecules, and to elucidate the chemical structures of molecules and other chemical compounds.

What does uni port sequence give information about protein query?

Types and Function of RNA?

Three major types of RNA are mRNA, or messenger RNA, that serve as temporary copies of the information found in DNA; rRNA, or ribosomal RNA, that serve as structural components of proteinmaking structures known as ribosomes; and finally,tRNA, or transfer RNA, that ferry amino acids to the ribosome to be assembled.

Protein folding?

Protein folding is the physical process by which a protein chain acquires its native 3-dimensional structure, a conformation that is usually biologically functional, in an expeditious and reproducible manner.

Chou Fasman Algorithm?

The Chou–Fasman method is an empirical technique for the prediction of secondary structures in proteins, originally developed in the 1970s by Peter Y. Chou and Gerald D. Fasman.

Protein sequence and database name?

Protein sequencing is the practical process of determining the amino acid sequence of all or part of a protein or peptide. The Protein Data Bank (PDB) is a database for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids. ... The PDB is overseen by an organization called the Worldwide Protein Data Bank, wwPDB.

pH less than pK?

If the pH is less than the pK_a, then the acid form of the compound predominates. If the pH is greater than the pK_a, then the conjugate base predominates. This is shown graphically here: Most ionizable groups fallinto two patterns depending on the charge found on the acidic form.

Difference between RNA and DNA with ref. to nucleotide?

Nucleotide[edit] A nucleotide is composed of a nucleobase (nitrogenous base), a five-carbon sugar (either ribose or 2'-deoxyribose), and one to three phosphate groups. ... In DNA, the purine bases are adenine and guanine, while the pyrimidines are thymine and cytosine. RNA uses uracil in place of thymine.

Redundancy codon, name an amino acids which coded for three different codon?

Phylogenetic tree types?

Scaled Trees: Branch lengths are equal to the magnitude of change in the nodes.
Unscaled Trees: Only representing the relationship between sequences.

Purpose of uni port and their feature?

UniProt is a freely accessible database of protein sequence and functional information, many entries being derived from genome sequencing projects. It contains a large amount of information about the biological function of proteins derived from the research literature.

What do you know about rammananchandrah plot?

Ramachandran Plot is a way to visualize dihedral angles ψ against φ of amino acid residues in protein structure.Ramachandran recognized that many combinations of angles in a polypeptide chain are forbidden because of steric collisions between atoms.

Blastx and tblastx work?

Blastx: Search protein database using translated nucleotide query.
Tblastx: Search translated nucleotide database using translated nucleotide query.

How we can modify dot plot to consider and mis match in alignment?

In the Alignment View locate the first region of mismatch in your alignment. Turn on the ... Youshould realign these sequences with modified settings. Rather than ... In the Dotplot there is also an inverted region (you will need to activate the. Reverse ... sequences and do you think it is a fair representation of their sequence.

Why do we need to use dynamic programming in predicting RNA structure?

RNA sequences comprises of 4 types of nucleotides • G/C, G/U & A/U are complementary & can form H-bonds • RNA sequences may contain hundreds of nucleotides, hence many combinations are possible. • So, given an RNA sequence, there is a large number of possible 2' structures • This presents us with an extremely complex and large problem

Few challenges in field of biotechnology?

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

Step to use FASTA?

Step 1: Specify the tool input (sequence and database).
Step 2: Entering of input sequence.
Step 3: Set up the parameters.
Step 4: Submit the query for processing.

Difference between local and global alignment?

Global alignment - maximizes the number of matches between the query and source sequences along the entire length of both the sequences.
Local alignment - gives the highest scoring local match between both query and sequences

Domain of protein?

Domains are semi-independent functional structures in a protein • Protein may contain multiple domains • Hence, we can try to classify proteins by their domains.
Examples of Protein Domains 1. Alpha Domains 2. Beta Domains 3. Alpha/Beta Domains 4. Alpha + Beta Domains 5. Alpha & Beta Multi-Domains 6. Membrane & cell-surface proteins

Uni port three main functions?

Scoring matrix in detail?

Scoring Matrix. 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.

PAM matrix in detail?

Alignment scoring matrices are very useful in giving suitable scores to matches and mismatches • There are 2 types of scoring matrices i.e. PAM and BLOSUM • PAM means "Point Accepted Mutations" • Point accepted

mutation is a substitution of one amino acid by another such that the protein functions stays conserved. PAM unit is a time in which about 1% of amino acids in a sequence undergo accepted mutations.

Bioinformatics Promise?
???????????

Draw the structure of RNA and DNA?

Newick notation example?

?????????
??
Complete protein sequence formula 3 role of 5'cap and 3'cap?
???????????

Dot plot matrix?
Dot plots employ dot matrix with two sequences plotted on top & left of the matrix • Matches are represented by dots • Dots on diagonals are connected and represent alignments.

Formula of sequence?
?????????

Write field of bioinformatics?

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinaryfield of science, bioinformatics combines computer science, statistics, mathematics, and engineering to analyze and interpret biological data.

How bulge 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.

Why need dynamic programming?

Dynamic programming is both a mathematical optimization method and a computer programmingmethod. ... Likewise, in computer science, if a problem can be solved optimally by breaking it into sub-problems and then recursively finding the optimal solutions to the sub-problems, then it is said to haveoptimal substructure.

Why diagonal form energy matrix of zuker algorithm?
?????????
?
Find out the mis match match and gap final score?
????????
?
Code of protein sequence ATCATCCATAC?
???????????

Basic principle for progressive alignment for MSA?

Progressive alignment methods are efficient enough to implement on a large scale for many (100s to 1000s) sequences. ... They recommend Clustal Omega which performs based on seeded guide trees and HMM profile-profile techniques for protein alignments. They offer different MSA tools forprogressive DNA alignments.

How pseudo knot effect RNA?

The pseudoknot is a potentially important tertiary structural motif of RNA.

How many type of data base in bioinformatics and What type of information they contain?
These formats include text, sequence data, protein structure and links. Each of these can be found from certain sources, for example: Text formats are provided by PubMed and OMIM. Sequence data is provided by GenBank, in terms of DNA, and UniProt, in terms of protein.

Application of bioinformatics?

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

Difference between rooted and un-rooted tree?

Rooted trees shows the most basal ancestor of the tree.Unrooted phylogenetic tree does not show an ancestralroot. ... Unrooted trees represents the branching order but do not indicate the root or location of the last common ancestor. Unrooted trees shows the relatedness oforganisms without indicating ancestry.

Define mutation?

the changing of the structure of a gene, resulting in a variant form which may be transmitted to subsequent generations, caused by the alteration of single base units in DNA, or the deletion, insertion, or rearrangement of larger sections of genes or chromosomes.

Purine and Pyramidne Bases?

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 onecarbon nitrogen ringbases (thymine and cytosine) are pyrimidines.

Diffrence acidic and basic amino acids?

Acidic amino acids have acidic side chains at neutral pH while basic amino acids have basic side chains at neutral pH. carboxylic acid is the side chain foracidic amino acids and basic amino acids contain nitrogen containing groups. ... Lysine, arginine and histidine are basic amino acids.

ORF and FASTA stand for?

ORF (Open Reading Frame)
FASTA (Fast Alignment)

Conserved sequence?

Conserved sequence: A basesequence in a DNA molecule (or an amino acid sequence in a protein) that has remained essentially unchanged, and so has been conserved, throughout evolution.

BLAST function?

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.

BLOSUM steps?

- 1 Collecting sample blocks. ...
- 2 Computing probabilities. ...
- 3 Computing the BLOSUM matrix.

Way of constructing phylogenetic tree?

Several methods exist for constructing phylogenetic trees. Broadly, they belong to objective methods or clustering methods. • We will study UPGMA and Distance Methods.

Genetic mutation?

Genetic mutation is the basis of species diversity among beetles, or any other organism. Mutationsare changes in the genetic sequence, and they are a main cause of diversity among organisms. These changes occur at many different levels, and they can have widely differing consequences.

Dot plot benefits?

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.

Clustring method?

The 5 Clustering Algorithms Data Scientists Need to Know. Clustering is a Machine learning technique that involves the grouping of data points. ... Clustering is a method of unsupervised learning and is a common technique for statistical data analysis used in many fields.

Protein sequence contain?

Protein sequences contain A, R, N, D, C, E, Q, G, H, I, L, K, M, F, P, S, T, W, Y & P

Gibbs Free energy?

A thermodynamic quantity equal to the enthalpy (of a system or process) minus the product of the entropy and the absolute temperature.

Categories of RNA?

- mRNA - Messenger RNA: Encodes amino acid sequence of a polypeptide.
- tRNA - Transfer RNA: Brings amino acids to ribosomes during translation.
- rRNA - Ribosomal RNA: With ribosomal proteins, makes up the ribosomes, the organelles that translate the mRNA.

Diffrence between Blast and FASTA?

BLAST is the most widely used tool for the local alignment of nucleotide and amino acid sequences. FASTA is a fine similarity searching tool which uses sequence patterns or words.

Why RNA is less stable than DNA?

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. This hydroxyl group make RNA less stable than DNA because it is more susceptible to hydrolysis.

Domian of protein?

A protein domain is a conserved part of a given protein sequence and (tertiary) structure that can evolve, function, and exist independently of the rest of the protein chain. Each domain forms a compact threedimensional structure and often can be independently stable and folded.

Step of compute PAM?

Steps to compute PAM matrices

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

Step 2: Let Ai,j be the number of times Ai is substituted by Aj

Step 3: Compute the frequency fi of amino acid Ai Then, PAM1 = pij = PAM'n' = (PAM1) n

Step of FASTA algorithm?

FASTA can search sequence databases and identify unknown sequences by comparing them to the known sequence databases. • This can help obtain information on the parent organism, function and evolutionary history.

Gibbs energy what role play in RNA folding?

Gibbs Free Energy" is the free energy of an RNA molecule available for reaction. The smaller, the better!. RNA structure formation lowers the free energy.

To quantify the similarity achieved by an alignment, scoring matrices are used: they contain a value for each possible substitution, and the alignment score is the sum of the matrix's entries for each aligned amino acid pair.

How biology simulation benefits to society?

- Genomics:
- Evolutionary Studies: Systems Biology:

Conclusion: Bioinformatics not only organizes, stores and analyzes biological data, but can also validate novel hypotheses. Modern day bioinformatics also helps predict disease outcomes as well as drugs to treat them!

Mathematical relationship between phlyogenetic tree?

A phylogenetic tree is a diagram that representsevolutionary relationships among organisms. ... Intrees, two species are more related if they have a more recent common ancestor and less related if they have a less recent common ancestor.Phylogenetic trees can be drawn in various equivalent styles.

How hydroxyl group of RNA make it unstable?

This hydroxyl group make RNA less stable than DNA because it is more susceptible to hydrolysis. RNA contains the unmethylated form of the base thymine called uracil (U), which gives the nucleotide uridine.

Given sequence find scoring matrix?

Wobble hypothesis?

The Wobble Hypothesis explains why multiple codons can code for a single amino acid. One tRNA molecule (with one amino acid attached) can recognise and bind to more than one codon, due to the lessprecise base pairs that can arise between the 3rd base of the codon and the base at the 1st position on the anticodon.

Which tool or databases use for nucleotide sequence?

Sequence Translation is used to translate nucleic acid sequence to corresponding peptide sequences. Backtranslation is used to predict the possiblenucleic acid sequence that a specified peptide sequence has originated from.
NUCLEOTIDE DATABASE – contains sequencedata from GenBank, EMBL, DDBJ as well as from the Genome Sequence Database and the US Patent and Trademark Office. It includes STSs and ESTs.

Sequence alignment by using matrix?

Seconday structure of RNA?

The secondary structures of biological DNA's and RNA's tend to be different: biological DNA mostly exists as fully base paired double helices, while biological RNA is single stranded and often forms complex and intricate base-pairing interactions due to its increased ability to form hydrogen bonds stemming from the ...

Enlist the field of Bioinformatics that can be explored under expasy?

Developed by Bioinformatics Institute (SIB) • Website provides access to databases and tools • Proteomics, Genomics, Phylogeny, Systems biology, Population genetics, transcriptomics etc.
Expasy provides access to a variety of online databases and tools. • Depending upon your requirement, you find sequence information from Expasy.

How pesudoknot alter RNA?

A 1’ RNA structure can fold into 2’ structure • 2’ structures can then form 3’ structures but avoid pseudoknots!

Why uni port query is used?

Batch search with UniProt IDs or convert them to another type of database ID (or vice versa) ... Find sequences that exactly match a query peptide sequence ...

Protien sequence?

Protein sequencing is the practical process of determining the amino acid sequence of all or part of a protein or peptide.

Matrix formula?

A determinant is just a special number that is used to describe matrices and finding solutions to systems of linear equations. The formula for calculating a determinant differs according to the size of thematrix. For example, a 2×2 matrix, the formula is ad-bc.

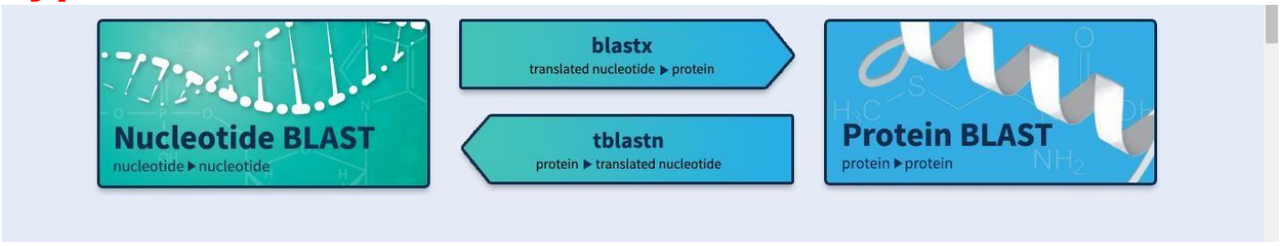
How tertiary structure of protein is less stable than secondary structure?

All proteins have primary, secondary and tertiary structures but quaternarystructures only arise when a protein is made up of two or more polypeptide chains. ... Primary structure is when amino acids are linked together by peptide bonds to form polypeptide chains.

Uses of CLUSTULW?

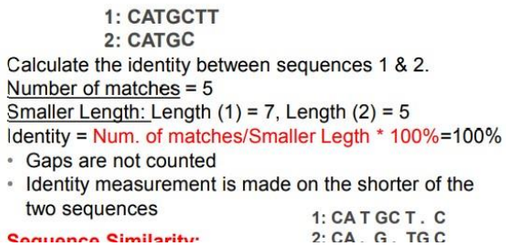
ClustalW is a widely used system for aligning any number of homologous nucleotide or protein sequences. For multi-sequence alignments, ClustalW uses progressive alignment methods.

Types of Blast and Uses?



Identity and Similarty and the formula of identity alignemts?

• Identity is the count of exact matches between two sequences. • Gaps are excluded • Similarity is the comparison between sequences calculated by using alignment approach.



4 Scoring possibility of Nussinov and Jacob algorthem?

4 positions are considered to calculate the NJ scoring matrix
• At every step, we can see score contributions from the 4 possible locations

Scope of bioinofrmatics?

Use of computational algorithms and techniques for: 1. Storage, 2. Organization, 3. Analysis, and 4. Representation of biological information

Why use castalw?

ClustalW is a widely used system for aligning any number of homologous nucleotide or protein sequences. For multi-sequence alignments, ClustalW uses progressive alignment methods. ... These scores are computed using the pairwise alignment parameters for DNA and protein sequences.

BIF401

Solved past and current papers (plz correct if u see any mistake) solved by Zareen Fatima

Define mutation (2) Ans:

Mutations are:

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

Databases use for proteins (2)

Ans: uniport and swisport

Generally used databases and which type of information they store (3) Ans:

Genebank. Pdb, uniport swisport

- Locus
- Accession number

- Sequence
- Molecular mass
- Authors
- Journal etc.

Types of RNA and their functions (5) Ans:

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.

Uses of clustalw 2

Ans: CLUSTALW can:

• create multiple alignments,

• optimize existing alignments,

• profile analysis &

• create phylogenetic trees

Pseudonots 3

Ans: • Tertiary or 3' structures may form pseudoknots

• A simple graph-based approach "Circular Plot" can help us detect the pseudoknots

A **pseudoknot** is a nucleic acid secondary structure containing at least two stemloop structures in which half of one stem is intercalated between the two halves of another stem. Pseudoknots fold into knot-shaped three-dimensional conformations but are not true topological knot

Uniprot swisprot Ans:

- Protein Sequences are reported from sequencing experiments

- This data is stored in protein sequence databases
- The famous ones include UniProt & SwissProt

Scoringatrix formula 5

$$s(a,b) = \frac{1}{\lambda} \log \frac{p_{ab}}{f_a f_b}$$

$s(a,b)=1/\lambda \log p_{ab}/f_a f_b$

Pam matric steps5" Ans:

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 suabstituted by A_j

Step 3: Compute the frequency f_i of amino acid A_i

$$\frac{A_{ij}}{\sum_k A_{ik}}$$

Then, PAM1 = p_{ij} =

PAM'n' = (PAM1) ^{n} **STEP OF**

BLOSUM?

Ans:

Step 1: Eliminate sequences that are identical in x% positions

Step 2: Compute observed frequency $f_{i,j}$ of aligned pair A_i to A_j. Hence, $f_{i,j}$ becomes the probability of aligning A_i and A_j in the selected blocks.

Step 3: Compute f_i which is the frequency of observing A_i in the entire block

$$s_{ij} = \log_2 \frac{f_{ij}}{p_{ij}}, \quad p_{ij} = \begin{cases} f_i f_i & i = j \\ 2 f_i f_j & i \neq j. \end{cases}$$

Name three hydrophobic amino acid?

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

2.rooted and unrooted tree difference?

Ans: Rooted and Unrooted trees can be used to show phylogenetic relationships between sequences

Rooted tree shows the phylogeny of ancestors

Phylogenetic trees take several forms: They can be rooted or **unrooted**, binary or general, and may show, or not show, edge lengths. 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

- Rooted trees reflect the most basal ancestor of the tree in question.
- Unrooted trees do not imply a known ancestral root.

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

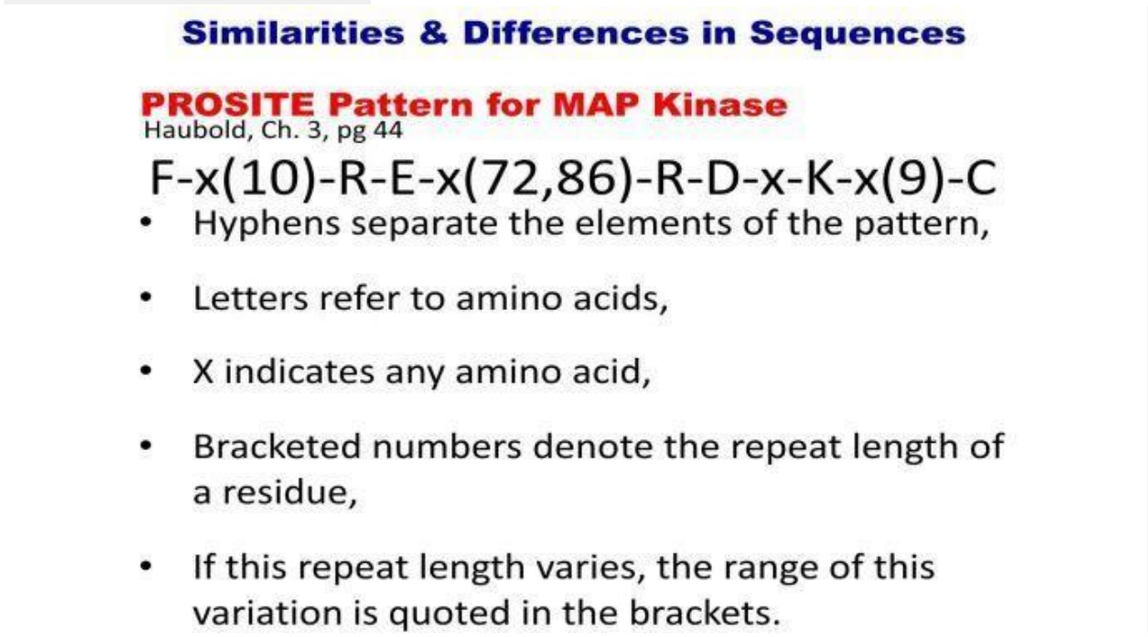
List all possible complementary position for the selected nucleotide in the complete sequence.

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

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

FXXXXRFKXXXXDER
FXXXXRFKXXXXXDER



Clustering? clustering is used to build groups of genes with related expression patterns **Define system biology and its application?**

. Systems Biology

Systems biology is the computational and mathematical modeling of complex biological systems

Application

- Model protein and gene interactions
- Dynamical analysis of such models

- Understand system properties • Predict system level behaviors **Need for bioinformatics ?**

1 An interdisciplinary field
2. New but rapidly developing field

3. Low requirement on infrastructure and research equipment

4. Vast opportunities for scientific discovery

Life, evolution and disease can be better understood
• Drugs can be better developed after understanding molecular basis of life

Possible contributions
• Solve important problems related to life

• Handle mammoth quantities of experimental data

• Develop efficient solutions towards these goals **Role of 5' cap and poly a tail?**

1. The 5' cap helps identify mRNAs at the Ribosomes

2. The 5' cap also acts as a shield against 5' exonuclease, thus leading to an increase in mRNA stability

The 3'-end of mRNA has a polyA tail (around 30-200 adenylate residues) which help shield against 3' exonucleases

Mid BIF401 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.

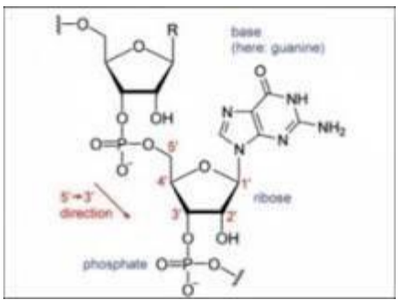
3. **What is dot plots?**

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 singlestranded 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. 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

9. 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.

10. 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 upstream 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.

37. 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.

38. How Gibbs energy fold RNA Secondary structure?

39. 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.

40. 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.

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.

43. PSA?

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

44. 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.

45. 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.

46. 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

47. 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

48. 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**.

49. 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.

50. 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.

51. 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.

52. 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}

53. 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.

54. Application of bioinformatics?

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

55. 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.

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

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

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

58. 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.

59. CLUSTALW can run Fast and slow modes?

Slow

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

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

62. 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.

63. Protein Sequencing can be done b?

Next Generation Sequencers

64. Object based method is?

Least Square Distances

65. Not Clusturing Approch Other than?

UPGMA MPGMA, Nieghbor Joining Single Linkage Complete Linkage

66. 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

67. Databases use for proteins?

uniport and swisport

68. Generally used databases and which type of information they store? Genebank. Pdb, uniport swisport

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

69. 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?

$S(a,b)=1/\lambda \log pab /fafb$

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