BIF401_Past Mid term Subjective

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.

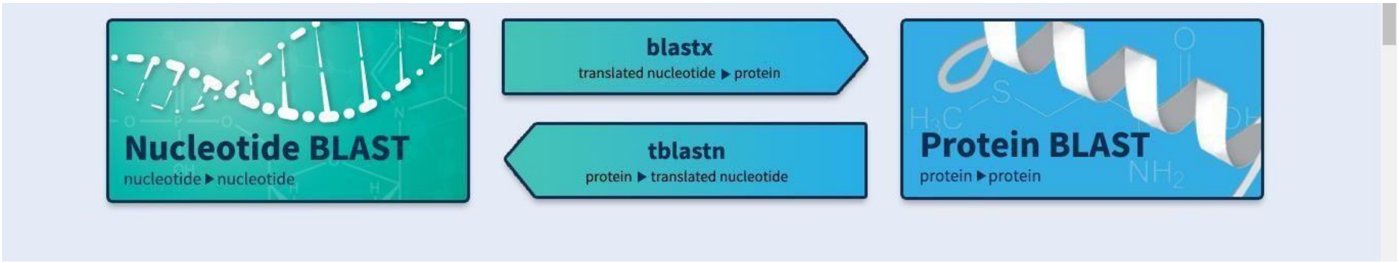
How tertiary structure of protein is less stable than secondary structure?

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?

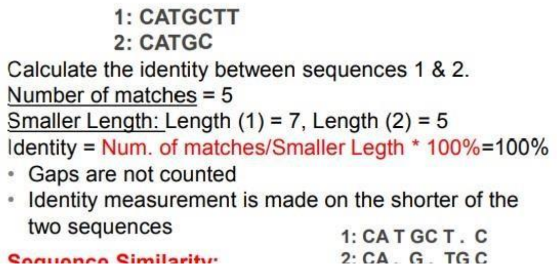
ClustalW is a widely used system for aligning any number of homologous nucleotide or protein sequences. For multi-sequence alignments, ClustalW uses progressive alignment methods.

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.



1: CATGCTT
2: CATGC

Calculate the identity between sequences 1 & 2.

Number of matches = 5

Smaller Length: Length (1) = 7, Length (2) = 5

Identity = Num. of matches/Smaller Legth * 100%=100%

- Gaps are not counted
- Identity measurement is made on the shorter of the two sequences

1: CA T GC T . C
2: CA . G . TG C

Sequence Similarity:

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.