

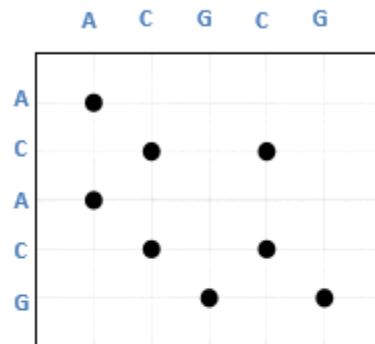
## BIF401 Current midterm paper

### QNO#1 How DP creates function to deal with matches mismatches and gapes? (5)

**ANSWER:**

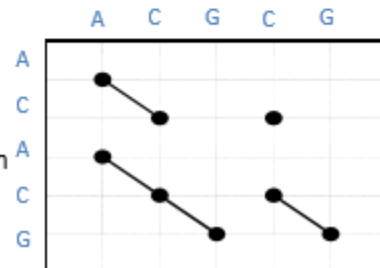
#### Representing Matches:

- For each nucleotide/amino acid which matches any counterpart nucleotide/amino acid, a dot is placed in the grid position across each row, one at a time!



#### Representing Matches:

- Similar regions are represented by a diagonal row of dots
- Non-diagonal broken dots are random matches



**Gaps:**-Dot plots cannot capture insertions, deletions and gap indications. How can we deal with them? • Modify the dot plot

### QNO#2 What are the four scoring possibilities for computing a position of matrix?

**ANSWER:**

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

## Scoring Matrices

Alignment is used to compare biological sequences

- Amino acids / nucleotides are more easily substituted based on their chemical similarity
- Since amino acids are substituted with different probabilities, we need a flexible match and mismatch scoring strategy “Scoring Matrices” contain such flexible scores for use during alignment How to build scoring matrices?
- We analyze the observed frequency with which each amino acid / nucleotide is substituted by another one in similar proteins/genes

## QNO#3 How protein sequence can be identified?

### ANSWER:

Protein sequence can be identified by uniprot and Swissport

- Protein Sequences are reported from sequencing experiments
- This data is stored in protein sequence databases
- The famous ones include uniprot & Swissport

## QNO#4 Difference local and global alignment?

### ANSWER:

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

### Types of pairwise alignments

Global	Local
<ul style="list-style-type: none"><li>• Maximizing sequence matches over the entire length of two sequences, by introducing gaps.</li><li>• Used to determine overall similarity, conservation.</li></ul>	<ul style="list-style-type: none"><li>• Find regions between two sequences that have the strongest similarity, excluding less similar regions.</li><li>• Finding similar domains, motifs, detecting distant homology</li></ul>

## QNO#5 Challenges of bioinformatics?

### ANSWER

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

### **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.

### **QNO#6 Gene mutation?**

#### **ANSWER**

A gene mutation is a permanent alteration in the DNA sequence that makes up a gene, such that the sequence differs from what is found in most people. Mutations may also result from insertion or deletion of segments of DNA due to mobile genetic elements. Mutations may or may not produce discernible changes in the observable characteristics of an organism.

### **QNO#7 Newick notation? Give example?**

#### **ANSWER**

In mathematics, **Newick tree format** (or **Newick notation** or **New Hampshire tree format**) is a way of representing graph-theoretical trees with edge lengths using parentheses and commas. It was adopted by James Archie, William H. E. Day, Joseph Felsenstein, Wayne Maddison, Christopher Meacham, F. James Rohlf, and David Swofford, at two meetings in 1986, the second of which was at Newick's restaurant in Dover, New Hampshire, US. The adopted format is a generalization of the format developed by Meacham in 1984 for the first tree-drawing programs in Felsenstein's PHYLIP package.

### **QNO#8 Why tertiary structure of RNA is less stable than secondary structure? (2)**

#### **ANSWER**

Because tertiary structure has more complexity than secondary structure in tertiary structure one strand of RNA not only collides over itself but also with other RNA molecules. Removing structures like bulges, loops they produce junctions which less expose the open strand of template RNA.

### **QNO#9 Difference b/w purine and pyrimidine bases (2)**

#### **ANSWER**

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

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

### QNO#10 GenBank define and uses(3)

#### ANSWER

**gene bank** a collection of clones containing all the **genes** of a particular organism, such as E. coli. The **bank** can be prepared by isolating the DNA from an organism, digesting it with a RESTRICTION ENZYME and cloning the restriction fragments. It can be maintained for many years, to provide a repository of cloned **genetic** material.

1. Gene bank. In an effort to **conserve agricultural biodiversity**, gene banks are used to store and conserve the plant genetic resources of major crop plants and their crop wild relatives. There are many gene banks all over the world, with the Svalbard Global Seed Vault being probably the most famous one.

### QNO#11 Redundancy codons and one example (5)

#### ANSWER

**Redundancy in the** genetic code means that most amino acids are specified by more than one mRNA **codon**. ... Tryptophan is unique because **it is the** only amino acid specified by **a single codon**. **The** remaining 19 amino acids are specified by between two and six **codons** each

### QNO#12 d/w rooted and unrooted trees and their uses (5)

#### ANSWER

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

## What is the difference between Rooted and Unrooted Phylogenetic Tree?

Rooted vs Unrooted Phylogenetic Tree	
A rooted phylogenetic tree is a diagram which shows the last common ancestor of the groups.	An unrooted phylogenetic tree shows the relationships between organisms without showing the common ancestor.
Node	
It has a node (root).	It does not have a node.
Direction of Evolution	
It has a direction to indicate the evolutionary time.	It does not specify an evolutionary relationship.
Attitude towards Others	
The tree allows to determining the ancestor – descendant relationship between groups.	The tree does not allow to talk about ancestor – descendant relationship.

### Summary – Rooted vs Unrooted Phylogenetic Tree

A phylogenetic tree represents the evolutionary pathways and connections between organisms using a branched tree-like diagrams. Phylogenetic trees can be rooted or unrooted. A rooted tree has a node at the base, representing the common ancestor which connects all interest groups. An unrooted tree shows the relationships between organisms. However, it does not depict the common ancestor that all the groups share. This is the difference between rooted and unrooted phylogenetic tree.

### USES:

Since the **rooted tree** depicts the direction of evolutionary time, it is easy to find the older or newer groups it has. A rooted tree can be used to study the entire groups of organisms. Accurate rooting of a phylogenetic tree is an important and crucial factor since inaccurate rooting can result in wrong interpretations of genetic changes between organisms and their directionality of the evolution.

#### **unrooted tree:**

This type of a tree does not indicate the origin of evolution of the groups of interest. It depicts only the relationship between organisms irrespective of the direction of the evolutionary time line. Therefore, it is difficult to study the evolutionary relationships of the groups with respect to time using an unrooted tree.

### QNO#13 Define gene

#### ANSWER

A gene is a locus (or region) of DNA that encodes a functional RNA or protein product, and is the molecular unit of heredity. The transmission of genes to an organism's offspring is the basis of the inheritance of phenotypic traits. Most

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

biological traits are under the influence of polygenes (many different genes) as well as the gene–environment interactions

### QNO#14 Natural plasmid

#### ANSWER

#### 1. Natural Plasmid

A plasmid is a small DNA molecule within a cell that is physically separated from a chromosomal DNA and can replicate independently. They are most commonly found in bacteria as small, circular, double-stranded DNA molecules; however, plasmids are sometimes present in archaea and eukaryotic organisms.

. Plasmids found in nature often give their hosts beneficial traits that allow them to survive in competitive environments. Plasmids derived directly from the environment are sometimes called ‘**natural**’ plasmids, to distinguish them from the modified versions we usually work with in the lab.

### QNO#15 Electroporation

#### ANSWER

1. the action or process of introducing DNA or chromosomes into bacteria or other cells using a pulse of electricity to briefly open the pores in the cell membranes.

**Electroporation**, or electro permeabilization, is a microbiology technique in which an electrical field is applied to cells in order to increase the permeability of the cell membrane, allowing chemicals, drugs, or DNA to be introduced into the cell (also called electro transfer).

### QNO#16 Secondary structure of protein

#### ANSWER

**Secondary Structure** refers to the coiling or folding of a polypeptide chain that gives the protein its 3-D shape. There are two types of secondary structures observed in proteins. One type is the **alpha ( $\alpha$ ) helix** structure. This structure resembles a coiled spring and is secured by hydrogen bonding in the polypeptide chain. The second type of secondary structure in proteins is the **beta ( $\beta$ ) pleated sheet**. This structure appears to be folded or pleated and is held together by hydrogen bonding between polypeptide units of the folded chain that lie adjacent to one another.

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

## QNO#17 Griffith experiment

### ANSWER

Griffith's experiment was an experiment done in 1928 by Frederick Griffith. It was one of the first experiments showing that **bacteria can get DNA through a process called transformation**. Griffith used two strains of Streptococcus pneumoniae. These bacteria infect mice.

**bacteria are capable of transferring genetic information through a process known as transformation**

## QNO#18 Why we need dynamic programming? 2 marks

### ANSWER

Sequence comparison is a costly process • Dynamic programming helps reduce this cost

**Dynamic programming** is a very powerful algorithmic **paradigm** in which a problem is solved by identifying a collection of subproblems and tackling them one by one, smallest first, using the answers to small problems to help figure out larger ones, until the whole lot of them is solved. In **dynamic programming** we are not given a dag; the dag is implicit. Its nodes are

## QNO#19 Tools use in RNA structure prediction? 2 marks

### ANSWER

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.

## QNO#20 Different fields in expasy? 5 marks

### ANSWER

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

Expasy provides access to a variety of online databases and tools.

• Depending upon your requirement, you find sequence information from Expasy.

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

## QNO#21 steps in Pam matrix? 5 marks

### ANSWER

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  $A_i$

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

Then, PAM1 =  $p_{ij}$  =

PAM'n' = ( PAM1 )<sup>n</sup>

## QNO#22 Gibbs energy? and how it is help in folding of Protein?

### ANSWER

“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

“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 of RNA structures RNA structures have the lowest (or close) quantity of free energy. In cases where RNA can take two structural forms, one can select the one with lower energy state.

- 5 nucleotides formed H-Bonds • This bond formation released energy (-12.0 kcal/mol) • RNA molecule took up a 2' structure • Hence became more stable

## QNO#23 conserved sequences?

### ANSWER

In evolutionary biology, **conserved sequences** are identical or similar **sequences** in nucleic acids (DNA and RNA) or proteins across species (orthologous **sequences**), or within a genome (paralogous **sequences**), or between donor and receptor taxa (xenologes **sequences**).

## QNO#24 What is the purpose of 3D-1D bowel algorithm? PPT?

### ANSWER

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

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	



Structure Prediction PSIPRED Introduction to Hidden Markov Models Ab initio modeling

**QNO#25 Orf and fasta stand for 3 marks**

**ANSWER**

Fast Alignment Algorithm,  
Open reading frame

**QNO#26 tblastn and tblastx 3 marks Three Purpose of tblastn and tblastx 2marks**

**ANSWER**

Blastx: Compares a nucleotide query sequence against a protein sequence database.

Helps find potential translation products of unknown nucleotide sequences

- tblastn: Compares a protein query sequence against a nucleotide sequence database

Nucleotide sequence dynamically translated into all reading frames

- tblastx: Compares the six-frame translated proteins of a nucleotide query sequence against the six frame translated proteins of a nucleotide sequence database.

**Amino acids codon 5 marks**

**ANSWER**

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.

**applications of bioinformatics?**

**ANSWER**

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

Ans: Introduction • Where can bioinformatics be applied specifically? • What are the avenues in biology that can benefit from this science? • What benefits can it deliver to the society?

1. Genomics

- DNA Sequencing • Gene Finding • Genome Assembly • Variation in Genomes • Transcription Data • Databases

2. Evolutionary Studies

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

- Evolutionary relationships • Evolutionary distances • Phylogenetics • Tree of life
3. Proteomics
- Protein Sequencing • Protein Structures • Post-translational Modifications • Protein-Protein Interactions • Database Development
4. Systems Biology
- Model protein and gene interactions
- VU Biologists [www.facebook.com/groups/vubiologists](http://www.facebook.com/groups/vubiologists)
- Dynamical analysis of such models • Understand system properties • Predict system level behaviors

## RNA And DNA Difference with respect to Nucleotide?

### ANSWER

Structurally, **DNA** and **RNA** are nearly identical. As mentioned earlier, however, there are three fundamental **differences** that account for the very **different** functions of the two molecules. **RNA** has a ribose sugar instead of a deoxyribose sugar like **DNA**. **RNA nucleotides** have a uracil base instead of thymine.

## What is pseudoknots?

### ANSWER

- 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 stem loop 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

## 3)Difference between Purines and pyrimidine?

### ANSWER

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.

## 4)uniport database

### ANSWER

## Difference acidic and basic amino acids?

### ANSWER

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

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

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.

## Basic amino acid

Acidic amino acids have acidic side chains at neutral pH while basic amino acids have basic side chains at neutral pH.

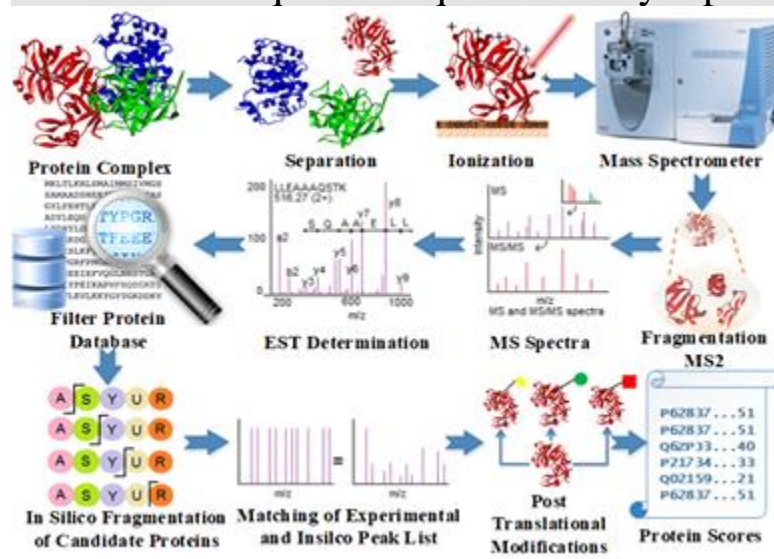
carboxylic acid is the side chain for acidic amino acids and basic amino acids contain nitrogen containing groups.

**Aspartic acid and glutamic acid are acidic amino acids. Lysine, arginine and histidine are basic amino acids.**

### Protein sequence identification?

#### ANSWER

Protein Sequence Identification • Mass spectrometers are used to measure the molecular weight of proteins and peptides • Given that several proteins may have the same mass, how can we identify the protein sequence Conclusions: • The flowcharts discussed above can help us arrive at the sequence of the protein in question • Scoring schemes are required to quantitatively represent the quality of results



### Hairpin structure of RNA?

#### ANSWER

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

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.

## Difference b/w protein mass fingerprinting & shotgun proteomics

### ANSWER

#### protein mass fingerprinting

Peptide mass **fingerprinting** (PMF) (also known as **protein fingerprinting**) is an analytical technique for **protein** identification in which the unknown **protein** of interest is first cleaved into smaller peptides, whose absolute masses can be accurately measured with a mass spectrometer such as MALDI-TOF or ESI-TOF.

#### Shotgun proteomics

refers to the use of bottom-up **proteomics** techniques in identifying **proteins** in complex mixtures using a combination of high-performance liquid chromatography combined **with mass spectrometry**

### Why there is need of bioinformatics

### ANSWER

Ans: The need for bioinformatics is on a rapid rise as biological data is rapidly increasing and becoming available online, free of any cost. The number of online tools for processing genomics and proteomics information are rapidly increasing. This is just a reflection of the need for bioinformatics in modern day biology.

Why we use CLUSTALW? (2) Enlist few points comes under the scope of bioinformatics (3) Why we use **multiple ORF while coding protein (3)**

### ANSWER

Codons of 3 nucleotides code for each Amino Acid • There are 1 start and 3 stop codons • How to search and select the valid start and stops?

Important: • Six ORFs exist for each DNA sequence • The valid ORF will have the longest DNA read • The first stop codon will mark the end of the protein sequence.

Given the DNA/RNA sequence, ORFs can be extracted and protein sequence can be determined • Codons of 3 nucleotides code for each Amino Acid • There are 1 start and 3 stop codons • How to search and select the valid start and stops? Conclusion: • Both forward and reverse RNA sequences are to be considered. • Multiple ORFs may code for proteins. • Select the ORF which produces the longest protein SEQUENCE

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	

## Mathematical relationship for joining two clusters in phylogenetic tree (2)

### ANSWER

Combining Clusters: Cluster X + Cluster Y = Cluster Z Calculate the distance of each cluster (e.g. W) to the new cluster Z Building trees using UPGMA

$$d_{ZW} = \frac{N_X d_{XW} + N_Y d_{YW}}{N_X + N_Y}$$

## 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).

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

Ans:

Gene Bank. Pdb, uniprot Swissport

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

ye ppr aya aj bioinformatics ka ....  
gibbs free energy  
generic mutations  
RNA folding  
protein sequence  
acidic and basic amino acids  
9:44 AM

Forwarded

[6/15, 9:20 AM] +92 [REDACTED]  
BIF 401 subjective „ How DP  
create function to deal with  
matches.mismatches and gapes(5)  
[6/15, 9:20 AM] +92 [REDACTED]:  
What are the four scoring  
possibilities for computing a position  
of matrix  
9:44 AM

VU APEX CAMPUS	<a href="http://vuapex.com.pk">vuapex.com.pk</a>	<a href="http://vuapex.pk">vuapex.pk</a>
Contact Us:	0322-8877744	