

Virtual University of Pakistan

(Bif401)

(Bioinformatics)

Midterms Past Papers solved

Email address

bif401@vu.edu.pk

Course Instructor

(Dr. Safee Ullah Choudry)

Created By Zareen Fatima

Group of VU Biologists

www.facebook.com/vubiologists

www.facebook.com/groups/vubiologists

VU Biologists

www.facebook.com/groups/vubiologists

BIF401

Solved 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: uniprot and swisport

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

Ans:

Genbank. Pdb, uniprot 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 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

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 matrix 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 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} =

$\text{PAM}'n' = (\text{PAM1})^n$

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

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

How are Bulges Formed?

Ans:

- 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

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?

FXXXXRFXKXXXXDER

FXXXXRFXKXXXXDER

Similarities & Differences in Sequences

PROSITE Pattern for MAP Kinase

Haubold, Ch. 3, pg 44

F-x(10)-R-E-x(72,86)-R-D-x-K-x(9)-C

- Hyphens separate the elements of the pattern,
- Letters refer to amino acids,
- X indicates any amino acid,
- Bracketed numbers denote the repeat length of a residue,
- If this repeat length varies, the range of this variation is quoted in the brackets.

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

Please write down the complete names of organisms renowned as "EMBL"

and "EBJ"

Ans: European Molecular Biology Laboratory

Describe the types of phylogenetic tree

Name any 3 hydrophobic amino acids.

Elaborate the 4 steps to use Fasta algorithm to your nucleotide amino acid

Formula for scoring matrix.

Ans: topic 67

domain of protein ?

Ans: Examples of Protein Domains Background • Alpha Domains • Beta Domains • Alpha/Beta Domains • Alpha + Beta Domains • Alpha & Beta Multi-Domains

• Membrane & cell-surface proteins

what is positive & negative release of energy in molecule?

Ans:

Application of bioinformatics

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

• 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

- Dynamical analysis of such models
- Understand system properties
- Predict system level behaviors

q1: indel and substitution

Ans: Insertions or deletions ("indels") result in gaps in alignments

- Substitutions result in mismatches

q2: EXPASY STAND

Ans: export protein analysis system

Q3: NEWICK ZRUTH

q4: need of bioinformatic

Ans: Motivation

1. An interdisciplinary field
2. New but rapidly developing field
3. Low requirement on infrastructure and research equipment
4. Vast opportunities for scientific discovery

Conclusion

- Since genome and proteome information is publically available, you can process this information after downloading it
- If you carefully search this online data, with a little bit of biology background, it is a gold mine!

Quote Biology easily has 500 years of exciting problems to work on.

Donald Knuth, Professor,
Stanford University

What can bioinformatics deliver? • 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

Conclusion

- From gene sequences to protein sequences

Bioinformatics is the way forward

- Protein structure, protein-protein interactions and systems biology are other research areas in bioinformatics

q5: field of bioinformatic

q6: similarity and identity and formula

Ans: topic 28

Enumerate the clustering and objective based methods for constructing phlogenetic tree.

Ans:

Clustering Approach	Objective based Methods
UPGMA	Least Square Distances
WPGMA	Maximum Likelihood
Neighbor Joining	Maximum Parsimony
Single Linkage	
Complete Linkage	