NAME: SUNDAY THANKGOD IDORENYIN

MATRIC NO: 18/MHS01/343

DEPT: ANATOMY

COURSE: CSC261

ASSIGNMENT:

(a)  List and discuss any three (3) criterion of database searching

1(b) Explain briefly Basic Local Alignment Search Tool (BLAST) as used in database similarity searching

2(a) Briefly discuss the differences between Dayhoff PAM Matrices and BLOSUM Matrices

2(b) Explain briefly Heuristic Database Searching

3(a) Define the following (i) Sequence Homology(ii) Sequence Similarity (iii) Sequence Identity

3(b) (i)         Give any three (3) methods of Alignment Algorithm (ii) Discuss briefly Pairwise Sequence Alignment

4(a) Differentiate between Global Alignment and Local Alignment

4b        Distinguish between the following

1. Sequence Homology and Sequence Similarity
2. Sequence Similarity and Sequence Identity

ANSWERS:

**1a**

1. Sensitivity
2. Selectivity
3. Speed
4. Sensitivity: sensitivity, which refers to the ability to find as many correct hits as possible. It is measured by the extent of inclusion of correctly identified sequence members of the same family. These correct hits are considered “true positives” in the database searching exercise.
5. Selectivity: selectivity, also called specificity, which refers to the ability to exclude incorrect hits. These incorrect hits are unrelated sequences mistakenly identified in database searching and are considered “false positives.”
6. Speed: which is the time it takes to get results from database searches. • Depending on the size of the database, speed sometimes can be a primary concern. Ideally, one wants to have the greatest sensitivity, selectivity, and speed in database searches.

**1b**. Basic local alignment search tool as used in database similarity searching: It is a sequence similarity search program that can be used to quickly search a sequence database for matches to a query sequence. BLAST provides an exact value, statistical information about the significance of each alignment.

Its main function is to compare a sequence of interest, the query sequence, to sequences in a large database, BLAST then reports the best matches, or “hits" found in the database.

**2a** Dayhoff PAM (Point Accepted Mutations) was created by Margaret Dayhoff. PAM matrices are regularly used as substitution matrices to score sequence alignments between closely related protein sequences. They are based on global alignments, mutations in global alignments are very significant. Their alignments have high similarity than BLOSUM alignments. Different PAM matrices correspond to different lengths of time in the evolution of protein sequence. Higher numbers in the PAM matrix naming denotes greater evolutionary distance. Example: PAM 250 is used for more distant sequences than PAM 120.

**WHILE**

BLOSUM (Blocks Substitution Matrix) which was created by Steven and Henikoff is a substitution matrix used to score alignments between evolutionarily divergent protein alignments. They are based on local alignments. Their alignments have low similarity than PAM alignments. Based on highly conserved stretches of alignments. All BLOSUM matrices are based on observed alignments, they are not extrapolated from comparisons of closely related proteins like the PAM matrices. Higher numbers in the BLOSUM matrix naming denotes higher sequence similarity and smaller evolutionary distance.

Example: BLOSUM 80 is used for closely related sequences than BLOSUM 62.

**2b Heuristic database** searching refers to a search strategy that attempts to optimize a problem by iteratively improving the solution based on a given heuristic function or a cost measure. Heuristic algorithms have been designed to reduce the time required to build an alignment that has a reasonable chance to be the best one. Such algorithms have been implemented as fast and efficient programs (Blast, FastA) available in different types to address different kinds of problems. The heuristic algorithms perform faster searches because they examine only a fraction of the possible alignments

examined in regular dynamic programming.

**3a. i. SEQUENCE HOMOLOGY:** It is an inference or a conclusion about a common ancestral relationship. When two sequences are descended from a common evolutionary origin, they are said to have a Homologous relationship or share Homology. Homology sequences usually have the same, or very similar functions so new sequences can be relatively assigned functions if homologous sequences with known functions can be identified.

1. **SEQUENCE SIMILARITY:** Sequence similarity is a measure of an empirical relationship between sequences. Its common objective is establishing the likelihood for sequence homology. It is the likeness (resemblance) between two sequences in comparison.
2. **SEQUENCE IDENTITY:** Sequence identity is the amount of characters which match exactly between two different sequences. The gaps are not counted and the measurement is relational to the shorter of the two sequences. It is the extent to which two (nucleotide or amino acid) sequences have the same residues at the same positions in an alignment, often expressed as a percentage.

**3B. (i)Give any three (3) methods of Alignment Algorithm (ii)  Discuss briefly Pairwise Sequence Alignment**

1. Dot matrix method

Dynamic Programming method

Word Method.

1. **Pairwise Sequence Alignment:** Pairwise sequence alignment methods are used to find the best-matching piecewise (local or global) alignments of two query sequences. Pairwise alignments can only be used between two sequences at a time, but they are efficient to calculate and are often used for methods that do not require extreme precision (such as searching a database for sequences. extremely useful in structural, functional, and evolutionary analyses of sequences. Pairwise sequence alignment provides inference for the relatedness of two sequences.

**4a. Differences between global alignment and local alignment;**

1. In Global alignment, two sequences to be aligned are assumed to be generally similar over their entire length WHILE In local alignment, does not assume that the two sequences in question have similarity over their entire length.
2. In Global alignment, entirety of both sequences are taken into consideration when finding alignment WHILE in local alignment small portions are taken into account.
3. In global alignment, alignment is carried out from beginning to the end of both sequences to find the best possible alignment across the entire length WHILE in Local alignment it only finds local regions with the highest level of similarity between the two sequences and align these regions without regard for the alignment of the rest of the sequence regions.
4. In global alignment algorithm starts at the beginning of the two sequences and add gaps to each until the end of one is reached WHILE In Local alignment algorithm finds the region or regions of the highest similarity between the two sequences and build the alignment outward from there.

**4b i. Differences between Sequence homology and Sequence similarity**;

1. In Sequence homology when two sequences are descended from a common evolutionary origin, they are said to have a homologous relationship or share homology WHILE In sequence similarity it is a related but different term is sequence similarity, which is the percentage of aligned residues that are similar in physiochemical properties such as size, charge, and hydrophobicity.
2. Sequence homology is an inference or a conclusion about a common ancestral relationship drawn from sequence similarity comparison when the two sequences share a high enough degree of similarity. On the other hand, Sequence similarity is a direct result of observation from the sequence alignment.
3. Sequence homology is a qualitative statement; it is either homologous or non-homologous WHILE Sequence similarity can be quantified using percentages.

**4b ii. Differences between sequence similarity and Sequence Identity;**

1. In a protein sequence alignment, sequence identity refers to the percentage of matches of the same amino acid residues between two aligned sequences. WHILE Sequence Similarity refers to the percentage of aligned residues that have similar physicochemical characteristics and can be more readily substituted for each other.
2. Sequence similarity is a measure of an empirical relationship between sequences. Its common objective is establishing the likelihood for sequence homology i.e chance that sequences has evolved from a common ancestor WHILE Sequence identity is the amount of characters which match exactly between two different sequences. A similarity score is therefore aimed to approximate the evolutionary distance between a pair of nucleotide or protein sequences.