

ARHAM (+923351328979) | Dir: SAFI KHAN (+923339666686) | Co. Dir: KAINAT CH.

OBJECTIVES

1. Which statement is not true about optimal energy? **Native Structure Is Always at The Global Minimum**
2. Which of the following main component of mass spectroscopy component according to their mass to change ratio? **Detector System**
3. _____ is exhaustive search approach. **SEQUEST Algorithm**
4. Bioinformatics requires _____. **Smart Mind & Internet Connected Computer**
5. Which of the following does not affect the stability of alpha helix? **Bulkiness**
6. A sequence of genome is meaningless without___. **Assembly**
7. Brute Force is _____Pattern finding Algorithms. **Approximate**
8. LCS scoring mechanism is for _____ **Sequence Similarity Analysis**
9. Physiological base approach _____ **Disease Centric**
10. An association rule has two parts _____ **If and Then**
11. Transfer of gene among organism is called _____ **Horizontal Gene Transfer**
12. Eco R1 breakdown DNA at _____ **GAATTC**
13. Programming language used for modeler _____ **Python**
14. It is to compute the prefix function _____ **O(M)**
15. Prob used by sequencing in hybridization have size _____ **8-32**
16. Which is correct_____? **Homology Vs Paralogy + Orthology**
17. _____ helps in finding global alignment. **Needleman Wunsch**
18. Brute force algorithm uses every possible _____ to find the solution. **L MER**
19. LCS scoring award 1 for _____ **Match**
20. __programming approach does not require prior sequence alignment**.** **Dynamic Programming**
21. Brute force is a _____ **Motif Finding Problem**
22. DNA array is also known as _____ **DNA Chip**
23. GCG stand for _____ **Genetics Computer Group**
24. Gene mapping is determining the_____ on chromosome. **Location**
25. MS2 genome has _____ amino acid.**129**
26. Method use fr lesd validation_____ **Virtual Screening**
27. Use for embryonic development _____ **Homeobox Genes**
28. Concerned with the effect of genetic factors on reactions to drugs ___ **Pharmacogenetics**
29. Sequence in EMBL format start with _____ **SQ**
30. An ordered sequence of symbols ___ **Pattern**



ARHAM (+923351328979) | Dir: SAFI KHAN (+923339666686) | Co. Dir: KAINAT CH.

31. Deletion in alignment in Edit graph shows _____ **East (Horizontal Arrow)**

32. How many methods for lead validation? **5**

33. Chou FASMAN algorithm helps to find _____ **Secondary Structure of Protein**

34. Plant mitochondria contain genes _____ **150 – 200kb**

35. Brute force is an _____ Pattern finding algorithm. **Approximation Algorithm**

36. Which identify the structure with minimum free energy _____ **Ab Initio**

37. Length of each line for FASTA algorithm should be _____ **Shorter Than 80 Or (60)**

38. Genome rearrangement results in a change of _____ **Gene Ordering**

39. Human genome consists of repeats _____ **20 %**

40. Responsible for protein function _____ **Protein Structure**

41. _____ pattern matches with k mismatches or differences. **Approximate**

42. We look for molecule which need to be selected as target and those molecules that cause disease _____ **Target Discovery**

43. GEO for _____. **Gene Expression Omnibus**

44. Intron start from5 end is____ ans 3′ end is_____. **GT, AG**

45. SINE is _____ bp long. **80 – 300**

46. Needle man publish in _____ **1970**

47. Chou FASMAN algorithm was 1st used in _____ **1974 – 1978**

48. The notion of spectral similarity with shared peaks count _____ **P1, P2-S1 and S2**

49. Human genome project _____ **1988**

50. TCGGGATTTCC is in drosophila for activate _____ **Immunity**

51. BLOSUM 6 is good for _____ **Un Gaped Alignment**

52. Disease that are hereditary _____ **3000 – 4000**

53. Computational technique uses for protein not present in database _____. **DENOVO**

54. Motif are found in _____. **Promoter**

55. A simple way (but not the best) to score an alignment is to count 1 for each match and 0 for each mismatch. **Simple Alignment Scores**

56. Substitution matrices are the log-odds matrices used for scoring____ substitutions in pairwise alignments. **Amino Acid**

57. When amino Acids are dissimilar, mutations from one into the other occur less during evolution it will be__? **Negative Score**

58. Substitution scores can be derived from**, (Match Model, Probabilistic Model)**

59. The probability of two sequences is the product of probability of each amino acid, in the case Of. **(Unrelated Model, Random Model)**



ARHAM (+923351328979) | Dir: SAFI KHAN (+923339666686) | Co. Dir: KAINAT CH.

60. Day Hoff, Schwartz and Orcutt presented their famous PAM matrix in? **1978**

61. _____do not capture true difference between short time substitutions and long-term ones. **Day Hoff Matrices**

62. Which one can perform well only in case of closely related proteins. **PAM Matrices**

63. BLOSOM or BLOCKS database was derived in ____by Henikoff & Henikoff. **1992**

64. BLOSUM62 is good for ___Alignments. **Ungapped**

65. PAM and BLOSUM are famous ___ matrices. **Substitution**

66. Smith-Waterman algorithm finds the best ____ alignment between two sequences. **Local**

67. Dynamic programming is used to find an _alignment of two sequences and its scores. **Optimal**

68. Needleman Wunch algorithm was developed by Saul B. Needleman and Christian D. Wunsch and published in. **1970**

69. In needle man wunch algorithm while doing matrix filling is____ the sum of the cell to the left and the gap score. **Horizontal Gap Score**

70. ___was awarded by Noble prize after sequencing a virus having size 5386 nucleotides. **Fred Sanger**

71. Human genome project was started in? **1988**

72. After completion of genome project approximately ___ nucleotides were sequenced. **3 Billion**

73. How many nucleotides could be reads in starting sequencing methods? **500 - 700**

74. After completion of H.G.P, Shotgun sequencing was coming into being after? **20 Years**

75. Shot gun sequencing is used when there will be_____ Amount of DNA sequence. **Huge**

76. In shot gun we use ____ for the breakdown of DNA into small pieces. **Sonication**

77. For obtaining shortest super string we use a technique. **Shortest Super String Problem**

78. Poly nomial algorithm is ____ for shortest super string problem. **Unlikely**

79. In sequencing hybridization to identify a fragment there was a small prob of ___ nucleotides size. **8 – 32**

80. Which of the following is time consuming technique? **Sequencing Hybridization, Human Genome Project**

81. Light directed polymer synthesis was develop by. **Steve Fodor**

82. Steve gave his sequencing technique in? **1991**

83. 64-kb DNA array was built in 1994 by a company of California named as? **Affymetrix**

84. In an array of DNA chip each prob locate at a distance with? **Known Location**

85. What we use to determine the prob which is hybridized to DNA. **Spectroscopic Detector**

86. Fragment assembly is actually assemblage of fragments into? **Entire Genome**

87. How many %age of human genome consists of repeat sequence? **20%**



ARHAM (+923351328979) | Dir: SAFI KHAN (+923339666686) | Co. Dir: KAINAT CH.

88. How many times the 300bp ALU is present in human genome? **1 Million**
89. In human T cells Locus, there are 5 different closely related fragment of ____gene. **Tryptosine**
90. LINE and SINE present more than ____ times in human genome. **2 Laks**
91. _____ is a special design vector in which large size of fragment can be ligated. **BAC**
92. The size of fragments that we can use in BAC is almost? **150,000bp**
93. We often use 30,000 bp BAC to sequence the? **Human Genome**
94. Whole genome sequencing using BAC to BAC strategy was advocated by**?** **James Weber, Gene Myers**
95. Clear biotechnology company announced 1st human draft in___. **2001**
96. The problem of repeated units found in a sequence was resolved by ____strategy. **Mate**
97. In Mate strategy we actually _____the size of read sequence. **Increase**
98. Who sequence the insulin 1st time in 1940? **Sanger**
99. We can cut the terminal amino acids for sequencing by using. **Edman Degradation**
100. Sanger use _____ enzyme for the breakdown of insulin. **Protease**
101. Which of following is sulfur containing amino acid? **Cystene**
102. _____purified spliceosome complex? **Mattias Man**
103. To identify apoptosis in higher organisms a commonly used technique is__. **Mass Spectrometer, Mass Spectrophotometer**
104. A programmed cell death also known as___. **Apoptosis**
105. In the Tandem mass spectrometry most frequent C terminal ion of peptide is ____ ion and N terminal ion is ____ ion. **B – Y**
106. If we have access to data base for all protein from a genome then**. No Need Of MS/MS Spectrum**
107. To search sequence of protein from already present data base there is an algorithm called? **SEQUEST**
108. SEQUEST develop by___? **John Yates**
109. Over a million protein sequences we know only about _____protein structures. **100,000**
110. Mostly proteins cannot be___? **Crystallized**
111. _____is used to determine the structure of protein. **NMR**
112. In nature proteins either fold spontaneously or by___? **Enzymes, Chaperons**
113. Which can determine the protein structure? **Amino Acid**
114. We can obtain a large protein sequence dataset from___? **Uniprot**
115. To predict the protein structure 1st step is to predict ____**Structure, Secondary**
116. If we have several _____ in the sequence, then we can anticipate that a helix



ARHAM (+923351328979) | Dir: SAFI KHAN (+923339666686) | Co. Dir: KAINAT CH.

may be formed by them. **Alanines**

117. To predict secondary structure of amino acids 1st algorithm that was use is___. **Chou-Fasman Algorithm**

118. SEQUEST algorithm is a _____approach. **Exhaustive**

119. Product of the propensity values is computed for __For each amino acid. **Overall Propensity**

120. For Alpha Helices ____contiguous amino acids are required. **4**

121. Alpha-Helix propensity should be more than ___ Once this propensity falls, Alpha-Helix stops.**1.0**

122. _____amino acids are needed to start Alpha Helices. **4**

123. Chou FASMAN algorithm help us to predict**. (Alpha Helices, Beta Sheets, Turns)**

124. In Chou FASMAN algorithm Alpha Helices can be finalized if there is ____ the propensity from Beta Sheets. **Higher**

125. Beta sheets can be predicted from ____ amino acid sequences. **Primary**

126. How many amino acids are needed to start a Beta Sheets? **5**

127. Chou FASMAN Algorithm is based on statistical occurrence of __ in known structure. **Amino Acids**

128. Protein sequence determined the structure, and structure determines the___. **Function**

129. Protein Secondary structure form due to__. **Hydrogen Bonds**

130. Alpha helix form by H+ bonds b/w ____ terminal. **C – N**

131. For the prediction of protein 2ndry structure we have to see the propensity of ___ amino Acids. **Neighboring**

132. Chou FASMAN algorithm introduced by Chou Et Al in___. **1974**

133. We use propensity table to see___. **Which Amino Acid Found in Which 2´Structure, Amount of Amino Acid Found In 2´Structure**

134. Glutamine and Alanine are more likely found in. **Alpha Helix**

135. An Alpha helix will be start in which of the following condition. **P(A) > 1.0**

136. _____are very flexible in nature. **Turns**

137. If we have 4 amino acids in a sequence having specific propensity then we can construct. **Tetra Peptide**

138. F (total) means product of? **Propensity of Amino Acids**

139. Which condition is not including for forming a turn**?** **F(Total) < 0.000075**

140. Chou FASMAN algorithm is based on? **Statistical Occurrence of Amino Acids**

141. For the formation of B – sheets, 1st we need to scan the. **Entire Sequence**

142. For scanning a sequence, we have to start it from. **Left Side**

143. If 4 out of 6 amino acids have propensity > 0.1 we call it as. **Alpha – Helix**



ARHAM (+923351328979) | Dir: SAFI KHAN (+923339666686) | Co. Dir: KAINAT CH.

144. If 4 contiguous amino acids have p < 0.1 and >also > p(a) it must be a ___ **B – Sheets**

145. If 4 out of 6 amino acids have propensity for a – helix then. **(We Have No Need to Check It for B – Sheets, We Should Simply Extend It till The End)**

146. According to Chou FASMAN, if a sequence does not have propensity for a – helix or B – sheets then it will be computed for. **B – Turn**

147. For the improvement of Chou FASMAN algorithm, we must be improving its statistical numbers that we obtain from. **PDB**

148. For which we cannot rely on 4 amino acids bcs it has variable sizes. **Loops**

149. For improvement of Chou Fasman algorithm according to latest protein data set we must be considered. **Nucleation Regions, Membrane Proteins, Hydrophobic Domains**

150. Functional evaluation of proteins can be performed by____. **Structure Visualization, Classification, Prediction**

151. X-Ray Crystallography and NMR Spectroscopy are used to find the __ of protein. **Structure**

152. It is then possible to identify unknown protein structures by just examining the____ protein sequences. **Homologous**

153. In homology modelling, proteins with similar_____ are considered. **1' Sequences**

154. Homology modelling is used to predict structures of proteins having high sequence similarity with other proteins with____. **Known Structures, Similar Sequence**

155. What is true about target protein from following? **Protein with Known Sequence but Unknown Structure**

156. By visualization of a structure of protein we actually see. **Biological Role of Protein**

157. Which is not true about experimental determination of overall protein structure, that it is? **Cheep**

158. Which is used to determine the amino acids sequence in protein? **Mass Spectrometry, Admen Degradation**

159. For structural prediction of protein, we can use. **NMR, X – Crystallography**

160. Which of the following statement is not true about Twilight Zone? **It Require High Sequence Alignment**

161. _____is 1st requirement to visualization and classification of protein according to its function. **Structural Prediction**

162. Protein classification can be done by different hierarchies like. **SCOP, CATH**

163. _____is the function unit of protein. **Domain**

164. _____is a software for homology modelling. **Modeler**

165. There are _____ stable folds of protein in nature. **5,000**

166. Several tools are available to perform homology modelling in a programmatic

or automated way such as___. **Swiss Model Roberta 3d Jigsaw**

167. Fold recognition is also called_____. **Threading**

168. Homology modelling fails to predict quality structures _____ is a technique for predicting these structures. **Fold Recognition &Threading**

169. Threading Scoring typically involves using a _____ function based on energy of molecule. **Z-Score**

170. _____cannot be predicted using threading. **Novel Proteins**

171. _____Is online tool for threading. **iTASSER**

172. In a structure less energy present in amino acids make it more___ **Stable**

173. What is the basic principle of fold recognition? **Align and Compare Sequence on The Base of Folding**

174. A common fold is a_____ **Bundle and Trim Barrel.4 Helix**

SUBJECTIVES

From Mid Portion

1. XML Stand for What?

XML stands for Extensible Markup Language.

2. Which Algorithm Avoids the Back Tracking?

KMP Algorithm:

A linear time algorithm for string matching which does not involve backtracking of strings Is called KMP algorithm.

3. What Is the Run Time of Brute Force Algorithm?

Brute force algorithm to solve the Motif Finding problem running time of O(l · nt) . Cannot runs it on biological samples. It is a Faster greedy technique not correct but have good performance. It is an Approximation algorithm, running time of this algorithm is O (ln2 +lnt), which is vastly better than the O(lnt) of IMPLEMOTIFSEARCH or even the O(4lnt) of BRUTEFORCEMEDIANSTRING

4. KMP Matcher

KMP stands for Knuth Morris Pratt. ... KMP algorithm was the first linear time complexity algorithm for string matching. KMP algorithm is one of the string-matching algorithms used to find a Pattern in a Text. KMP algorithm is used to find a "Pattern" in a "Text".

5. Define Suffix Trees?

It is a compressed tree containing all the suffixes and allows many problems on strings to be solved quickly.

6. Explain Endosymbiont Hypothesis?

There is a hypothesis that mitochondria actually evolved from bacteria and is known as endosymbiont hypothesis).

According to this hypothesis, these organelles originated as separate prokaryotic organisms that were taken inside a primordial eukaryotic cell. Such symbiotic relationships in which two species are dependent upon one another to varying extents served as crucial elements of the evolutionary progression of eukaryotic cells.

7. GENEQUIZ-

Focuses on deriving a predicted protein function based upon the available evidence; including evaluation of similarity to the closest homologue in the database (i.e. it is good tool for functional annotation).

8. Swiss PROT

SWISS-PROT is a curated protein sequence database which strives to provide a high level of annotations (such as the description of the function of a protein, structure of its domains, posttranslational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases.

9. What Is the Difference Between Protein Sequence and DNA Sequence?

- A DNA sequence has a chain of nucleotides while A protein sequence has chain of amino acids
- The nucleotides are A, C, G and T for the DNA and A, C, G and U for RNA while There are around 20 amino acids for protein
- Nucleotides are joined together through phosphodiester linkage while Amino acids are join together by peptide bonds

10.What Is the Run Time of Brute Force Algorithm?

Brute force algorithm to solve the Motif Finding problem running time of O(l · nt).

Cannot run it on biological samples. It is a Faster greedy technique- not correct but have good performance. It is an Approximation algorithm, Running time of this algorithm is O(ln2 +lnt), which is vastly better than the O(lnt) of IMPLEMOTIFSEARCH or even the O(4lnt) of BRUTEFORCEMEDIANSTRING

11.What is denovo protein sequencing?

It is used for the protein that are not present in any data base, So its sequence is not known. Or if it is known then it may be modified by some mutation or by some chemical modification.

12.Primary Biological Database

- Examples of Primary database- Nucleic Acid Databases are GenBank and DDBJ
- Protein Databases are PDB, Swiss Prot, PIR, TrEMBL, Metacyc, etc.

13.Hamming Distance and Its Calculation

The number of positions that two codewords of the same length differ is the Hamming distance. In case of DNA sequences, we define this distance as the number of bases by which they differ.

String 2: ATCGCTCAT

String 2: ACCGTTGAT

These two strings have 3 mismatches, so the hamming distance

14.Satellite DNA & Mini-satellite

Satellite DNA has following features:

⇒ It may be one to several thousand bp long and it can also be present as Tandem; array of 100 million bases long.
⇒ They are present near centromere and telomere and
⇒ They can be classified as Mini-satellite and Micro-satellite.

Mini-Satellite:

Mini-satellite features are as follows:



⇒ They are 15 bases long in array of several hundred to thousands kb.
⇒ They are typically present in euchromatin region
⇒ Example is VNTR and is used to identify human individuals in forensics.

15.Write Approaches for Prediction of Exon.

Exons can be predicted within ORF from the information gathered about

- Splice Junctions
- Approach:

For each segment [acceptor, donor], we get three scores (coding potential, donor score, acceptor score). Various possibilities

- All Three Scores Are High – Probably True Exon.
- All Three Scores Are Low – Probably Not A Real Exon.
- All in The Middle?
- Some scores are high and some are low

16.Gene Expression Omnibus (GEO):

Genes are expressed into mRNA, and whenever we talk about gene expression, we generally mean the mRNA sequences so we can normally get those mRNA from techniques like microarray and another famous technique nowadays which is being established is known as RNA seq. And microarray data and RNA seq can be classified into Gene Expression Data which is stored in Gene Expression Databases.

17.Edit distance

Edit distance between two strings as the minimum number of editing operations needed to transform one string into another, where the edit operations are insertion, deletion, and substitution of one symbol for another

18.Write the Five tasks of genome analysis.

Genome Analysis:

In Genome Analysis we mainly perform the following tasks;

Sequencing

- Assembly (since the sequencing is done in a way that whole genome is broken down into short fragments and once those fragments are sequenced, we need to put them together, this step-in genome analysis is known as Assembly).
- Repeat identification and masking out (once we assemble that genome, we try to find out the regions in which we have large number of repeats because assemblies jumble up where we have those repeats so we need to find those regions and it is one of the important tasks to go and look into those assemblies while keeping in mind those regions in which we have those repeats).
- Gene prediction (after we have assembled a finished genome, now we can go for the prediction of the genes where we can find the genes by using different patterns or features of those genes).
- Looking for EST (Expressed Sequenced Tags) and cDNA (complementary DNA) sequences.
- (EST and cDNA are basically originated from the DNA where the genes that are expressed are transcribed into mRNA which is then reversed transcribed back into
- cDNA. So, by the help of cDNA, we can look into where those expressing regions are present in the genes that will give us the idea of the gene expression or the regions from where the mRNAs are made).
- Genome annotation (in which we can find out similar functions performed by different genes)
- Expression analysis (once we have the idea about the regions of the gene in which we can have the gene expression then we can explore the quantification i.e., how much those genes are being expressed).

19.What is the In Farm hexamer count?

While finding exons in the transcription start sites the frame is shifted from left to right by leaving one nucleotide every time and adding next nucleotide in the frame. Then all the frequencies of codons are counted which are likely to code an amino acid more often. This frequency count is called as in frame hexamer count.

20. What Are Cluster Linkage And Subgroups?

Clustering by Linkage:

There is another technique, where the clustering is done by linkage (almost similar to the previous ones except some changes). The method of doing it is as follows:

- Each sequence is a vertex.

- Significant alignment score is an edge.
- Trimming by removing weak edges (High P/E).
- Or remove > e-6 (we remove the ones which are not linked).
- Remaining sub-graph should share 2/3rd of edges.

21. Longest Common Subsequences

The simplest form of a sequence similarity analysis is the Longest Common Subsequence (LCS) problem, where we eliminate the operation of substitution and allow only insertions and deletions.

22. What Is the Function of DELETE (Y, L)?

In Brute Force Algorithm we make a List of pairwise distances, L, and uses the function DELETE (y, L). It means the value of "y" will be deleted from the subset "L".

23. Annotations Steps

Annotations are divided into two types and are as follows:

- Structural annotation
- Functional annotation

Structural Annotation:

Structural Annotation is where we try to identify certain gene features like;

- ⇒ Promoters
- ⇒ Terminators
- ⇒ Shine-Dalgarno sites; the ribosomal binding sites during the protein synthesis)
- ⇒ DNA motifs (patterns of nucleotides within the genes)
- ⇒ Co-transcription units
- ⇒ Operons in microbes (in micro-organisms, lots of genes are transcribed together known as operons)

Functional Annotation

The attributing biological information to the genes is called as functional Annotations and we can have it via;

- Biological function

- Biochemical function
- Gene expression (transcription of the gene is considered as gene expression)
- Regulation and interactions among different genes

24. List Four Databases?

- Swiss-Prot and PIR for protein sequences
- GenBank and DDBJ for genome sequences
- Protein Databank for protein structures

Lectures 95 to 108

25. l=3, TATGGTGC find the spectrum?

If l = 3 and s = TATGGTGC, then its spectrum will be as follow,
Spectrum (s, l) = {TAT, ATG, TGG, GGT, GTG, TGC} and its Lexicographic order will be,
Lexicographic order, like ATG, GGT, GTG, TAT, TGC, TGG

26. What would be the lexicographic order of, (TAT, ATG, TGG, GGT, GTG, TGC)

Lexicographic order of that spectrum is, ATG,GGT,GTG, TAT, TGC, TGG

27. What Are Substitution Matrices? Also Mention Its Types.

Scoring or Substitution Matrices:

It is use for scoring amino acid substitutions in pairwise alignments. They reflect substitution rates that are originated by evolutionary events. For this we align the sequence and count the mutation at each position.

Types:

Some of the substitution matrices to compute sequence alignments are:

- **PAM**: Point Accepted mutations
- **BLOSUM**: BLOCK Substitution Matrix

28. What is MATE and MATE strategy?

During fragment assembly, the reads of Sequences whose suffix and prefix are overlapping and they are likely to be consecutives are called as mate fragments or simply mates.

Mates' strategy:

Overlap: Finding potentially overlapping reads

Layout: Finding the order of reads along DNA

Consensus: Deriving the DNA sequence from layout Best match between the suffix of one read and the prefix of another. By using this strategy, we can sequence the whole fragment into single genome. Large number of reads to ensure that experimental errors are reduced to minor noise.

29. Mate Pair Sequencing Strategy Sequencing?

Mate-pair sequencing can also provide a powerful alternative for whole genome sequencing as the longer DNA fragments can be used to bridge together DNA fragments even in the presence of highly repetitive DNA sequences which are rife in human and many other complex genomes.

30. How Sanger solved Edman Degradation Problem of Protein Identification

Edman solved this problem by using an enzyme protease to breakdown the proteins insulin into the peptides and then he identified the peptides, and same method was used to digest the overlapping peptides.

31. Smith-Waterman Algorithm

- It is use to finds the best local alignment between two subsequences.
- Traceback starts with maximum value in the matrix and then go backwards,

32. Three steps of Needleman Wunch algorithm?

Three steps

- Initialization
- Matrix filling
- Traceback.

Initialization

- The cell of first row and first column of the matrix is initially filled with zero
- Add gap penalty for each shift to the right.

Matrix Fill

- Move through the cells row by row, calculating the score for each cell
- Compute three scores: – A match score – Vertical gap score – Horizontal gap score.

Traceback

- The final step in the algorithm is the trace back for the best alignment
- Start at the bottom-right corner
- Follow where maximum value comes from

33. What method used for sequence?

DNA sequencing methods Fred Sanger and Walter Gilbert Cells make copies of DNA. DNA fragments of different lengths- if one base is missing.

34. Explain fragment assembly algorithms

Strategy for Sequencing

Clone into BAC, sequenced each one-minigenome, Simplifies the computational assembly problem, Human Genome project, Whole-genome assembly paradigm (James Weber and Gene Myers) Celera

Genomics Mate-pair reads sequencing technique.

Inserts of length L- both ends are sequenced, Mates at distance L- length is larger than most repeats.

Overlap: Finding potentially overlapping reads

Layout: Finding the order of reads along DNA

Consensus: Deriving the DNA sequence from layout Best match between the suffix of one read and the prefix of another. By using this strategy, we can sequence the whole fragment into single genome. Large number of reads to ensure that experimental errors are reduced to minor noise

35. Explain spliceosome?

- Spliceosome is a large and complex molecular machine that is basically used in the splicing process.
- It is approximately containing more than 100 proteins.
- The spliceosome removes introns from a transcribe pre-mRNA which is a type of primary

transcript.

Lectures 109 to 132

36. What is denovo protein sequencing?

It is used for the proteins that are not present in any data base, So its sequence is not known. Or if it is known then it may be modified by some mutation or by some chemical modification.

37. Explain two computational problems which are used to identify proteins

- Protein Identification Problem:
- Modified Protein Identification Problem:

Protein Identification Problem:

Find a protein from a database that best matches the experimental spectrum.

Input: A database of proteins, an experimental spectrum S, a set of ion types and a parent mass m.

Output: A protein of mass m from the database with the best match to spectrum S

Modified Protein Identification Problem:

Find a peptide from the database that best matches the experimental spectrum with up to k modifications.

Input: A database of proteins, an experimental spectrum S, a set of ion types Δ, a parent mass m, and a parameter k capping the number of modifications

Output: A protein of mass m with the best match to spectrum S that is at most k modifications away from an entry in the database. Modified Protein Identification problem P1 and P2- S1 and S2, Notion of spectral similarity, Shared peaks count, Limitations in detecting similarities by database search.

38. Secondary Structure prediction process?

Secondary structure prediction is a set of techniques in bioinformatics that aim to predict the local secondary structures of proteins based only on knowledge of their amino acid sequence.

Lectures 133 to 144

39. Tools for Homology Modelling

MODELLER (Webb and Sali, 2014) is one of the widely used computational tools for predicting protein 3D structures using homology modeling. MaxMod, PyMod, and PRIMO are other recent methods/servers for homology modeling

Lectures 145 to 155

40. Cystic Fibrosis Gene

Cystic fibrosis is caused by mutations in the gene that produces the cystic fibrosis transmembrane conductance regulator (CFTR) protein. In people with CF, mutations in the CFTR gene can disrupt the normal production or functioning of the CFTR protein found in the cells of the lungs and other parts of the body

41. Homology, Paralogy and Orthalogy? Homology Modeling?

Homology modeling, also known as comparative modeling of protein, refers to constructing an atomic resolution. Model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template").

Orthology and paralogy relationships between two sequences are derived from such trees based on which evolutionary event (speciation or duplication, respectively) is assigned for their last common ancestor.

Lectures 156 to 166

42. What are personalized medicines?

The drugs which are made specifically for a patient with respect to its genome, physiology and system biology are called personalized medicine. It's a hot cake today to grow in this field being a bio-informatician.

43. Explain Primary Tertiary and Quaternary Structure of Protein?

Primary Structure:

The primary structure of a protein is the linear sequence of amino acids.

Together, this linear sequence is referred to as a polypeptide chain. The amino acids in the primary structure are held together by covalent bonds, which are made during the process of protein synthesis (translation).

Tertiary Structure:

The tertiary structure will have a single polypeptide chain "backbone" with one or more protein. secondary structures, the protein domains. Amino acid side chains may interact and bond in a number of ways. The interactions and bonds of side chains within a particular protein determine its tertiary structure.

Quaternary Structure:

Protein quaternary structure is the number and arrangement of multiple folded protein subunits in a multi-subunit complex. It includes organizations from simple dimers to large homo oligomers and complexes with defined or variable numbers of subunits.

44. RNA Structure and types and Secondary Structure and Process?

RNA, abbreviation of ribonucleic acid, complex compound of high molecular weight that functions in cellular protein synthesis and replaces DNA (deoxyribonucleic acid) as a carrier of genetic codes in some viruses. RNA consists of ribose nucleotides (nitrogenous bases appended to a ribose sugar) attached by phosphodiester bonds, forming strands of varying lengths. The nitrogenous bases in RNA are adenine, guanine, cytosine, and uracil, which replace thymine in DNA RNA Structure: RNA is typically single stranded and is made of rib nucleotides that are linked by phosphodiester bonds. A ribonucleotide in the RNA chain contains ribose (the pentose sugar), one of the four nitrogenous bases (A, U, G, and C), and a phosphate group.

45. RNA Types and Function:

⇒ 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.
⇒ snRNA - Small nuclear RNA: With proteins, forms complexes that are used in RNA processing in eukaryotes. (Not found in prokaryotes.)

46. Structure of Protein?

Protein structure is the three-dimensional arrangement of atoms in an amino acid-chain molecule. Proteins are polymers – specifically polypeptides – formed from sequences of amino acids, the monomers of the polymer. Very large aggregates can be formed from protein subunits.

Lectures 166 to 177

47. Supervised and Unsupervised Machine Learning?

Supervised Learning

⇒ Prediction of future cases
⇒ Knowledge extraction
⇒ Compression
⇒ Outlier detection

Un-supervised Learning

⇒ Learning "what normally happens"
⇒ No output
⇒ Clustering: Grouping similar instances

Lectures 178 to 190

48. Three Applications of Pharmacogenomics

The list below provides a few more commonly known applications of pharmacogenomics:

Improve drug safety, and reduce ADRs; Tailor treatments to meet patients' unique genetic pre-disposition, identifying optimal dosing; improve drug discovery targeted to human disease

49. Two target discovery approaches?

Target discovery to clinical application saga

⇒ Physiology-based approach
⇒ Target-based approach.

Physiology-Based Approach: Is a disease-centric approach in which target is not identified, multiple targets are involved. In vivo screening is done by using drugs, siRNA or antisense oligonucleotides.

Target-Centric Approach: Target based discovery starts with the identification of genes and their protein products. Aim to develop drugs affecting one gene or a molecular mechanism.

50. Lead Optimization – Technologies

♥ Medicinal chemists conduct extensive SARs to improve potency and selectivity.
♥ Improve physicochemical and drug-like properties

- Best molecules are advanced to animal Models & preliminary toxicology

51. Target Validation Tools

Helps in assessing their genetic association with disease. The following tools are:

- Antisense agents
- Ribozymes

52. Pharmacogenetics applications

- Detection of genetic variability of drug effects on the genome level
- Agent selection
- Analysis of drug
- Reactions and drug toxicity on gene expression
- Development of new indications for already approved drugs
- Discovery of new drug targets
- Identification of (non) responders in clinical trials of phase I-IV
- Identification of genotype dependent adverse drug reactions
- Identification of individuals at risk for severe adverse drug effects

53. DNA profiling processes

⇒ RFLP analysis.
⇒ PCR analysis.
⇒ STR analysis.
⇒ Amp FLP.
⇒ Y-chromosome analysis.

54. DNA Finger Printing / DNA Profiling & its Advantages

It is a form of forensic identification that is used primarily to identify people.

Advantages

Even though all humans share 99.9% of their DNA sequences, the remaining 0.01% of sequences is unique enough to differentiate people. Because DNA is in every cell of a person's body is present and same, any part of a human's body including dead skin skills, hair, saliva, and more contain DNA sequences.



"ک ت ے بھونکتے ہیں اپنے ہونے کا
احساس دلانے کے لئے ہی۔۔۔۔
جنگل کا سناٹا شیر کی موجودگی بیان کرتا ہے۔"

~ SAFI ~