

Bioinformatics model final exam with solutions

The solutions are typeset in red.

1, The DNA and RNA code the genetic information with the sequence of 4 bases. What are these bases in DNA?

Adenine, Guanine, Thymine, Cytosine.

2, Which bond is stronger: the covalent or the hydrogen bond?

The covalent bond.

3, DNA replication occurs in protein synthesis. Is this true?

No. DNA replication copies the double-stranded DNA when the cell is dividing.

4, What is the difference in messenger RNA (mRNA) and transfer RNA (tRNA)?

mRNA encodes a whole protein of hundreds of amino acids, it is much longer than the tRNA that simply transports amino acids one-by-one.

5, How many nucleotides encode an amino acid? Three.

6, What is the final product of (i) DNA replication? a copied DNA; and (ii) transcription with translation? A protein.

7, What is the complexity of simple (naive) sequence search algorithm, when we want to find a length-m word in a length-n text? $O(mn)$.

8, How can I identify the longest common subsequence in linear time with the help of suffix trees? What type of suffix trees need to be applied? Generalized suffix tree needs to be applied; the deepest node need to be found (from the root, the distance should be measured as follows: a length of an edge is the number of characters written onto the edge; the length of the path is the sum of the lengths of the edges) that has two nodes (leaves) under it from different sequences.

9, Why does the Hamming-distance satisfy the triangle inequality? The H-distance of a and b is the number of coordinates that need to be exchanged in a in order to make it equal to b. We need to prove that $H(a,b)+H(b,c)$ is larger or equal to $H(a,c)$. Clearly, if we make changes on a to get it equal to b, then on b to get it equal to c then we need to change at least as many coordinates as we changed when the coordinates of a are changed to be equal to c: in both cases a becomes equal to c, but in the first case it need to go through on b.

10, In the Needleman-Wunsch algorithm we build a matrix in the course of the algorithm. How can we interpret the number in row i and column j? The edit- or Lewenshtein distances of the ith and jth prefixes of the two sequences.

11, Is the BLAST a heuristic or an exact algorithm? Is it faster or slower than Needleman-Wunsch? BLAST is heuristic; much faster than N-W, approximately linear time algorithm.

12, What are the two most important secondary structures of proteins? **Alpha-helix and beta-sheet.**

13, xxx-bonds connect amino acid to amino acid in proteins. What word needs to be written instead of xxx? **Peptide**