04

RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES

B.Sc. Honours in Applied Biology

Third Year Semester II Examination – July 2020

BIO 3204 – BIOINFORMATICS

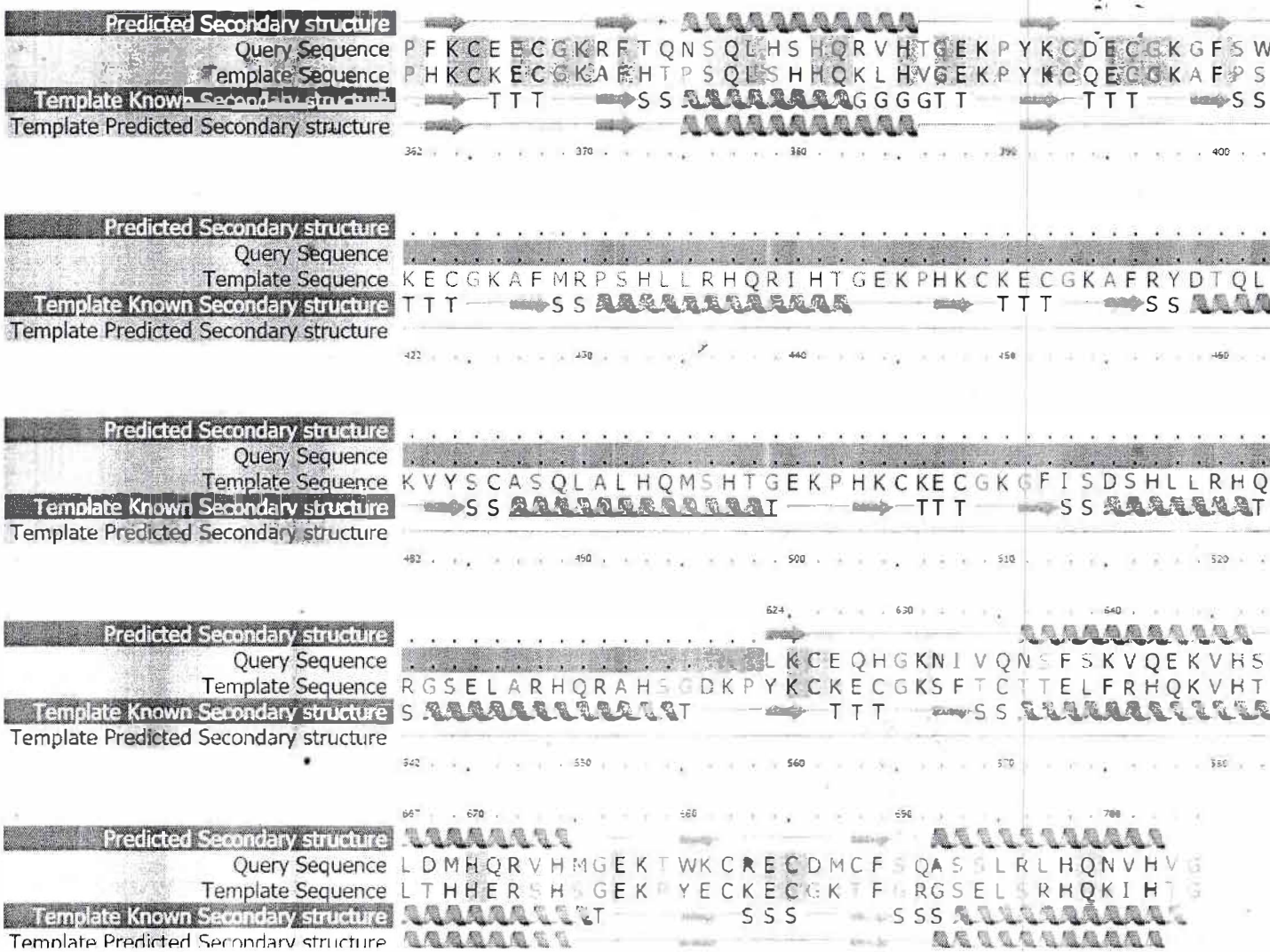
Time: Two (02) hours

Answer **ALL** questions.

1.

a) Describe the two main computational approaches to structure prediction of proteins. Your description should include the strengths and pitfalls of each workflow and example software packages for each. **(60 marks)**

b) The figure below contains the predicted secondary structure of a portion of the human Zinc-finger protein obtained using the protein structure prediction software Phyre2. The RMSD value of the predicted portion is given as 2.06 Å from the template structure. The percentage sequence identity of the output is given as 49%. Citing evidence from the figure, explain whether this model produced by Phyre2 is an accurate representation of the template (native) structure of the human Zinc finger protein. **(40 marks)**


Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure
Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure
Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure
Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure
Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure

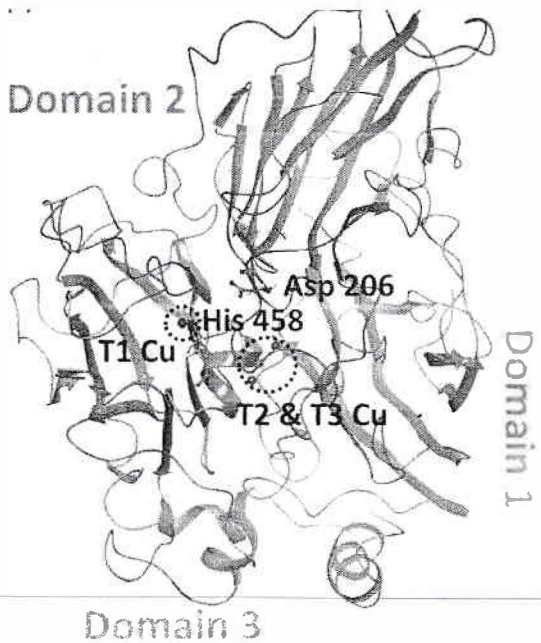
2. a) List the main differences between sequences in the database GenBank and the RefSeq subset of sequences **(20 marks)**

b) Describe how Boolean operators, parentheses, wild-cards and quotation marks are used in narrowing, broadening and refining the search outputs of bioinformatics databases such as GenBank and UniProt. **(50 marks)**

c) List the main factors considered when selecting one isoform of a protein as the "Canonical" sequence in UniProt, when multiple splice-variants (isoforms) of that protein are present.

(30 marks)

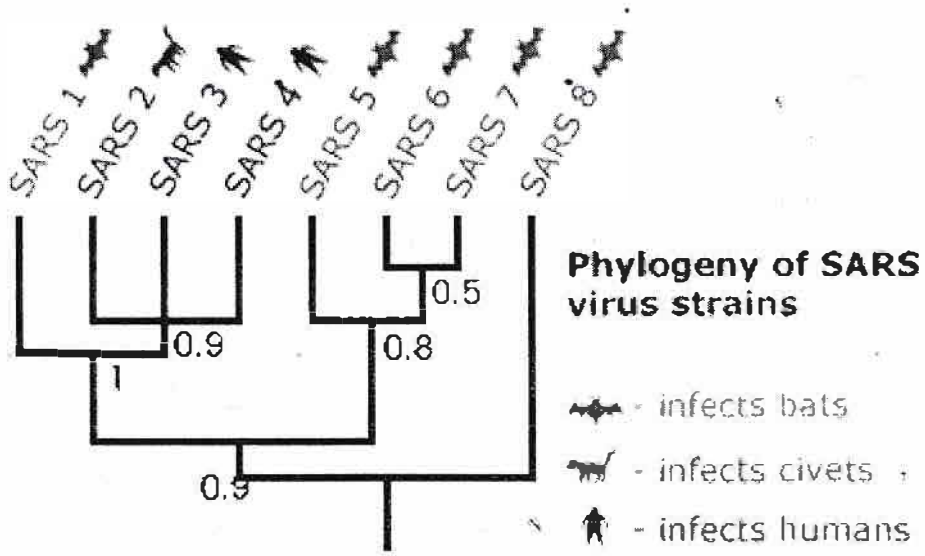
3.



A group of scientists discover that a mutant from of fungal laccase enzyme exhibits higher binding affinity for its' native substrate, lignin. Native form of laccase shown in the figure above is a multi-copper ligand containing enzyme that has a central binding pocket for lignin. A key amino acid residue stabilizing the enzyme-substrate complex is Asp 206. They hypothesize that a single amino acid substitution at position 206 is responsible for the increase in binding affinity. Design an experimental workflow to prove or disprove this hypothesis. If a substitution is found at position 206, describe the likely chemical nature of the substituted amino acid, in the mutant protein.

(100 marks)

4. A zoonotic disease is a type of an infectious disease that can spread to humans from animals. SARS outbreak between 2002 - 2005 is a similar disease that spread from China to several other countries in the world. It is caused by an RNA virus that leads to severe acute respiratory syndrome in humans. To trace the likely origins of the disease, the epidemiologists used a molecular phylogenetic approach using RNA from several different SARS virus strains that were isolated from other animal sources. The resulting consensus tree from the Bayesian analysis is given below.



I. Using your understanding in molecular biology and molecular phylogenetics, describe the possible steps the epidemiologists must have followed to test the origins of SARS 2002-2005 outbreak.

(30 marks)

II. During the time of the outbreak, there were three leading hypotheses regarding the origins of the disease. One suggested that the disease originated from infected bats. Another hypothesized that the disease transmitted to humans from infected civets (a cat-like mammal). The third hypothesis proposed that the virus infected civets from infected bats and evolved within civets and transmitted to humans. Using the above molecular phylogeny, explain which one of these three hypotheses is most likely?

(40 marks)

III. Describe what the numbers at the nodes mean and how they can be used in the interpretation of the tree.

(10 marks)

IV. Explain what is meant by a 'consensus tree' and state how it differs from a 'best scoring tree'.

(20 marks)

-END-