ANSWER SHEET

Exercise 2

Name and surname: Michal Piecuch

Student id: s32484

[Task 1]

The following Python script takes a user-supplied DNA sequence (case-insensitive) and computes the complementary sequence (handling standard bases A, C, G, T, U and ambiguous base N):

```
def get_complement(dna_seq: str) -> str:
    """Returns the complementary DNA strand (A<->T, C<->G)."""
    trans_table = str.maketrans('ACGTUacgtuNn', 'TGCAAaccaaNn')
    return dna_seq.translate(trans_table)

# Read DNA from user
user_sequence = input("Enter DNA sequence: ").strip()
complementary_dna = get_complement(user_sequence)

print("Complementary strand (3'->5'): ", complementary_dna)
print("Reverse complement   (5'->3'): ", complementary_dna[::-1])
```

[Task 2]

Protein sequence: Generated protein sequence (translated from Reading Frame 1 of the viral nucleic acid, 75 amino acids in length):

```
MYSFVSEEIGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV
```

Translation Frame: Translated from Reading Frame +1 of the 227 bp viral nucleic acid, yielding a full 75-amino-acid peptide with zero premature stop codons.

[Task 3]
Full Scientific Name: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

Database Match: The 75-aa sequence matches the Envelope small membrane protein (E protein) of Severe acute respiratory syndrome coronavirus 2 (NCBI Taxonomy ID: 2697049, Betacoronavirus) across the non-redundant protein database (nr).

[Task 4]
Alignment of Virus_1, Virus_2, and Virus_3: Pairwise sequence alignment generated via NCBI BLAST (blastp, bl2seq) with 'Flat query-anchored with dots for identities' and line length 120:

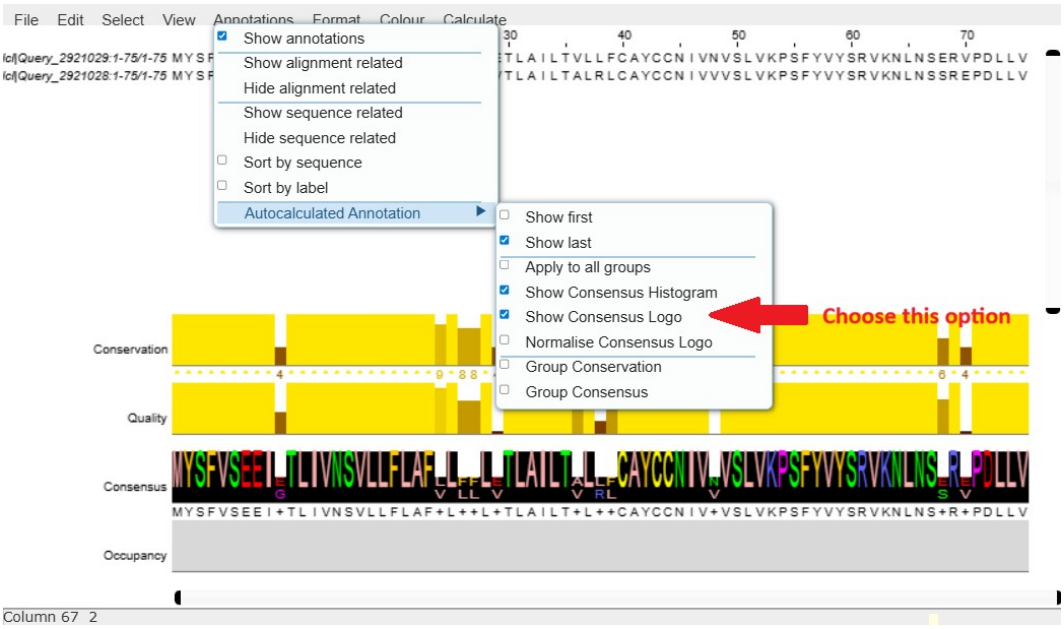
```
Query range 1: 1 to 75
Query     1  MYSFVSEEIGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV  75
Virus_2   1  .....................LL.....................V.........................  75
Virus_3   1  ......................LLF.A.....V.LF.................................  75
```

Interpretation: 1. Virus_1 vs Virus_2: 96.00% sequence identity (72 of 75 matches) with 3 substitutions: V24L, V25L, and N48V.
2. Virus_1 vs Virus_3: 90.67% sequence identity (68 of 75 matches) with 7 substitutions: V25L, F26L, L27F, V29A, A36V, R38L, and L39F.
3. Biological interpretation: Both terminal ends (residues 1 to 23 and 49 to 75) are 100% conserved across all three viruses. All sequence variations are located in the central hydrophobic transmembrane pore region (residues 24 to 40), where conservative hydrophobic exchanges preserve the ion channel pore.
4. Evolutionary relationship: Virus_1 is much closer to Virus_2 (96.0% identity) than to Virus_3 (90.7% identity).

[Task 5]
Similarity Matrix (BLOSUM62): Identified from the 'params' section of the single-file JSON report, with default gap penalties of 11 for existence and 1 for extension.

Similarity Matrix (BLOSUM62): Identified from the 'params' section of the single-file JSON report, with default gap penalties of 11 for existence and 1 for extension.

[Task 6]
Alignment Conservation: The JalviewJS visualization shows 100% sequence conservation in yellow across the N-terminal and C-terminal ends, with visible drops in conservation and multi-letter consensus representation across the transmembrane region (columns 24 to 29 and 36 to 39) and at residue 48.



Alignment Conservation: The JalviewJS visualization shows 100% sequence conservation in yellow across the N-terminal and C-terminal ends, with visible drops in conservation and multi-letter consensus representation across the transmembrane region (columns 24 to 29 and 36 to 39) and at residue 48.

[Task 7]

RCSB PDB Molecule of the Month: Molecule of the Month (May 2024): CFTR (Cystic Fibrosis Transmembrane Conductance Regulator).
Associated Disease: Cystic Fibrosis (mukowiscydoza).

RCSB PDB Molecule of the Month: Molecule of the Month (May 2024): CFTR (Cystic Fibrosis Transmembrane Conductance Regulator).
Associated Disease: Cystic Fibrosis (mukowiscydoza).

[Task 8]

Disease-Associated Amino Acid Deletion (PHE508): Amino acid name: Phenylalanine | Three-letter abbreviation: PHE | One-letter abbreviation: F | Position in chain: 508 (standard notation: PHE508, commonly written as delta F508 or p.Phe508del).
The deletion of Phenylalanine at position 508 destabilizes the CFTR protein. Endoplasmic reticulum quality-control systems in lung cells recognize the misfolded channel and destroy it before it reaches the cell surface.

Disease-Associated Amino Acid Deletion (PHE508): Amino acid name: Phenylalanine | Three-letter abbreviation: PHE | One-letter abbreviation: F | Position in chain: 508 (standard notation: PHE508, commonly written as delta F508 or p.Phe508del).

The deletion of Phenylalanine at position 508 destabilizes the CFTR protein. Endoplasmic reticulum quality-control systems in lung cells recognize the misfolded channel and destroy it before it reaches the cell surface.

[Task 9]
Type of Inheritance (OMIM #219700): In the OMIM database for CFTR (Entry #219700), the Phenotype-Gene Relationships table documents the mode of inheritance as Autosomal recessive. Symptoms develop only when both copies of the gene carry pathogenic mutations.

Type of Inheritance (OMIM #219700): In the OMIM database for CFTR (Entry #219700), the Phenotype-Gene Relationships table documents the mode of inheritance as Autosomal recessive. Symptoms develop only when both copies of the gene carry pathogenic mutations.

[Task 10]
Selected Diagram: Selected Diagram: a.
Justification: Diagram 'a' illustrates autosomal recessive inheritance. Two healthy carrier parents (genotype Aa) have a 25% chance of an affected child (aa), a 50% chance of carrier children (Aa), and a 25% chance of an unaffected child (AA). Diagram 'b' depicts autosomal dominant inheritance, where an affected parent passes the condition to 50% of the offspring.

Selected Diagram: Selected Diagram: a.
Justification: Diagram 'a' illustrates autosomal recessive inheritance. Two healthy carrier parents (genotype Aa) have a 25% chance of an affected child (aa), a 50% chance of carrier children (Aa), and a 25% chance of an unaffected child (AA). Diagram 'b' depicts autosomal dominant inheritance, where an affected parent passes the condition to 50% of the offspring.

[Task 11]
ClinVar Genomic Variant: Correct Answer: 2. c.1521_1523del.
In ClinVar (Variation ID: 7105), the classic cystic fibrosis mutation is an in-frame deletion of three nucleotides (CTT), designated c.1521_1523del, which removes codon 508 (Phenylalanine) without disrupting the remaining reading frame.

ClinVar Genomic Variant: Correct Answer: 2. c.1521_1523del.
In ClinVar (Variation ID: 7105), the classic cystic fibrosis mutation is an in-frame deletion of

three nucleotides (CTT), designated c.1521_1523del, which removes codon 508 (Phenylalanine) without disrupting the remaining reading frame.

[Task 12]
dbSNP Identifier (rsID): The reference SNP cluster identifier linked directly from ClinVar variation 7105 is rs113993960, located on chromosome 7 (GRCh38 coordinates 117559591-117559594).

dbSNP Identifier (rsID): The reference SNP cluster identifier linked directly from ClinVar variation 7105 is rs113993960, located on chromosome 7 (GRCh38 coordinates 117559591-117559594).

[Task 13]
Landmark Science 1989 Publication: PMID: 2570460 | Title: Identification of the cystic fibrosis gene: genetic analysis | Authors: Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC | Journal: Science (September 8, 1989, volume 245, issue 4922, pages 1073-1080) | Percentage of disease cases: Approximately 70% (~70% of CF mutant chromosomes worldwide).

[Task 14]
Function & Heme Group: Function of Neuroglobin (UniProt ID: Q9NPG2, NGB_HUMAN): 1. Protects brain and retinal neurons against ischemic stroke, hypoxia, and reperfusion damage. 2. Promotes oxygen delivery to mitochondria in active neurons under low oxygen. 3. Scavenges reactive oxygen and nitrogen species, including nitric oxide. 4. Acts as a nitrite reductase in its deoxy form to generate NO, slowing cell respiration and preventing apoptosis.
Role of the Heme Group: Contains an iron-protoporphyrin IX (protoheme) group coordinated by distal His64 and proximal His96. The central iron atom reversibly binds diatomic gases (O2, CO, NO) and undergoes reversible redox reactions (Fe2+ and Fe3+) needed for oxygen storage and antioxidant activity.

Role of the Heme Group: Neuroglobin contains an iron-protoporphyrin IX (protoheme) prosthetic group in a hexacoordinated bis-histidyl coordination state (coordinated by distal His64 and proximal His96). The heme iron atom is the active functional site that reversibly binds diatomic ligands (O2, CO, NO) and undergoes reversible redox cycles (Fe2+ <-> Fe3+) required for ligand binding, oxygen homeostasis, and enzymatic antioxidant defense.

[Task 15]

3D Structure of Human Neuroglobin (PDB ID: 1OJ6): Crystal structure of human brain neuroglobin (RCSB PDB ID: 1OJ6, solved at 1.95 Å resolution by Vallone et al.), showing the rainbow cartoon ribbon representation and the central protoheme prosthetic group:



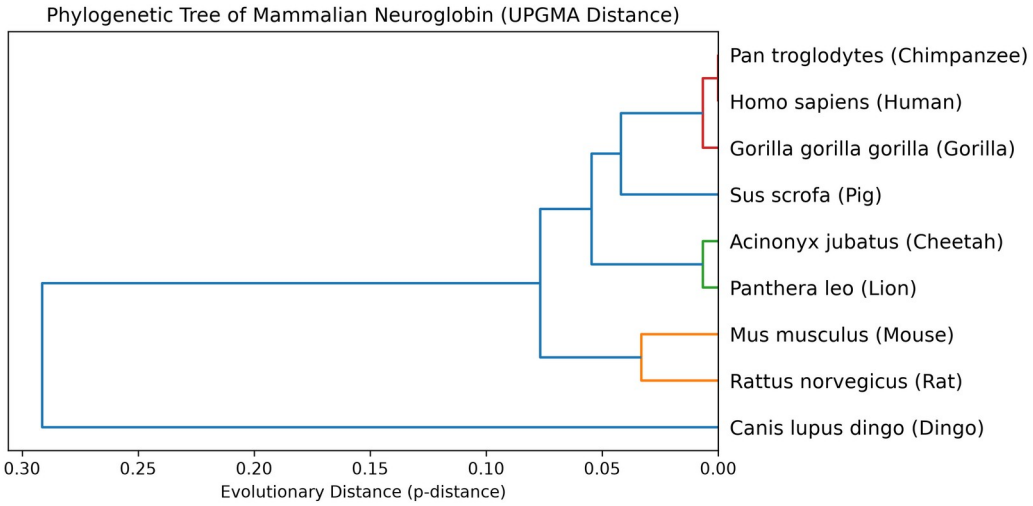
Secondary Structure Analysis (PDB: 1OJ6): Crystal structure of human neuroglobin (PDB ID: 1OJ6, 1.95 Å resolution by Vallone et al.).
Identified Secondary Structures: 1. Alpha-helices: the main structural element, forming a classic all-alpha globin fold of 8 alpha-helices (A through H) that make up 70% to 75% of the protein. 2. Turns and loops: flexible segments that connect the alpha-helices around the central heme pocket. 3. Beta-sheets: completely absent (0%). Neuroglobin contains zero beta-strands.

[Task 16]

Closest Organism & Sequence Identity: Organism with Neuroglobin Most Similar to Human: Pan troglodytes (Chimpanzee) with 100.00% identity (151 of 151 identical residues, 0 substitutions).
Comparative Sequence Identity vs Human (151 aa): Pan troglodytes (100.00%, 0 substitutions), Gorilla gorilla (99.34%, 1 substitution), Sus scrofa (96.03%, 6 substitutions), Panthera leo (95.36%, 7 substitutions), Acinonyx jubatus (94.70%, 8 substitutions), Rattus norvegicus (94.70%, 8 substitutions), Mus musculus (94.04%, 9 substitutions), Canis lupus dingo (<92% identity).

Closest Organism & Sequence Identity: Organism with Neuroglobin Most Similar to Human: Pan troglodytes (Chimpanzee) with 100.00% identity (151 of 151 identical residues,

0 substitutions).
Comparative Sequence Identity vs Human (151 aa): Pan troglodytes (100.00%, 0 substitutions), Gorilla gorilla (99.34%, 1 substitution), Sus scrofa (96.03%, 6 substitutions), Panthera leo (95.36%, 7 substitutions), Acinonyx jubatus (94.70%, 8 substitutions), Rattus norvegicus (94.70%, 8 substitutions), Mus musculus (94.04%, 9 substitutions), Canis lupus dingo (<92% identity).



Closest Organism & Sequence Identity: Organism with Neuroglobin Most Similar to Human: Pan troglodytes (Chimpanzee) with 100.00% identity (151 of 151 identical residues, 0 substitutions).
Comparative Sequence Identity vs Human (151 aa): Pan troglodytes (100.00%, 0 substitutions), Gorilla gorilla (99.34%, 1 substitution), Sus scrofa (96.03%, 6 substitutions), Panthera leo (95.36%, 7 substitutions), Acinonyx jubatus (94.70%, 8 substitutions), Rattus norvegicus (94.70%, 8 substitutions), Mus musculus (94.04%, 9 substitutions), Canis lupus dingo (<92% identity).