Bioinformatics Assignment Worksheet

Exercise 2

|  |
| --- |
| **Query Sequence** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** |
| S | T | W | G | E | R | G | L | **M** | **P** | **Y** | R | G | L | A | C | E | G | H | I |

|  |
| --- |
| **Protein Sequence 1** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** | **30** | **31** | **32** | **33** | **34** | **35** | **36** | **37** | **38** | **39** | **40** | **Total score** |
| M | V | G | V | M | N | S | A | Y | L | N | N | R | L | M | P | Y | G | G | S | G | M | P | E | C | D | Y | D | C | C | W | C | I | M | T | M | N | H | Y | C |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |
| --- |
| **Protein Sequence 2** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** | **30** | **31** | **32** | **33** | **34** | **35** | **36** | **37** | **38** | **39** | **40** | **Total score** |
| R | D | R | N | L | S | S | I | K | L | G | P | Q | V | Q | T | N | Q | M | P | C | G | E | R | D | A | T | N | T | R | C | I | I | N | Y | T | Y | I | S | R |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |
| --- |
| **Protein Sequence 3** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** | **30** | **31** | **32** | **33** | **34** | **35** | **36** | **37** | **38** | **39** | **40** | **Total score** |
| Y | A | W | F | Q | F | R | Q | D | S | F | S | R | E | T | Q | V | T | D | M | P | Y | T | V | W | K | L | I | R |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |
| --- |
| **Protein Sequence 4** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** | **30** | **31** | **32** | **33** | **34** | **35** | **36** | **37** | **38** | **39** | **40** | **Total score** |
| M | A | Q | Y | L | A | F | M | P | Y | N | N | S | V | L | H | Y | K | R | A | N | S | I | K | E | Q | H | P | F | R | M | P | Y | E | A | I | A | L | P | Y |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |
| --- |
| **BLOSUM-62 Substitution Matrix** |
|  | C | S | T | P | A | G | N | D | E | Q | H | R | K | M | I | L | V | F | Y | W |
| C | **9** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| S | -1 | **4** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| T | -1 | 1 | **5** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| P | -3 | -1 | -1 | **7** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A | 0 | 1 | 0 | -1 | **4** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| G | -3 | 0 | -2 | -2 | 0 | **6** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| N | -3 | 1 | 0 | -2 | -2 | 0 | **6** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| D | -3 | 0 | -1 | -1 | -2 | -1 | 1 | **6** |  |  |  |  |  |  |  |  |  |  |  |  |
| E | -4 | 0 | -1 | -1 | -1 | -2 | 0 | 2 | **5** |  |  |  |  |  |  |  |  |  |  |  |
| Q | -3 | 0 | -1 | -1 | -1 | -2 | 0 | 0 | 2 | **5** |  |  |  |  |  |  |  |  |  |  |
| H | -3 | -1 | -2 | -2 | -2 | -2 | 1 | -1 | 0 | 0 | **8** |  |  |  |  |  |  |  |  |  |
| R | -3 | -1 | -1 | -2 | -1 | -2 | 0 | -2 | 0 | 1 | 0 | **5** |  |  |  |  |  |  |  |  |
| K | -3 | 0 | -1 | -1 | -1 | -2 | 0 | -1 | 1 | 1 | -1 | 2 | **5** |  |  |  |  |  |  |  |
| M | -1 | -1 | -1 | -2 | -1 | -3 | -2 | -3 | -2 | 0 | -2 | -1 | -1 | **5** |  |  |  |  |  |  |
| I | -1 | -2 | -1 | -3 | -1 | -4 | -3 | -3 | -3 | -3 | -3 | -3 | -3 | 1 | **4** |  |  |  |  |  |
| L | -1 | -2 | -1 | -3 | -1 | -4 | -3 | -4 | -3 | -2 | -3 | -2 | -2 | 2 | 2 | **4** |  |  |  |  |
| V | -1 | -2 | 0 | -2 | 0 | -3 | -3 | -3 | -2 | -2 | -3 | -3 | -2 | 1 | 3 | 1 | **4** |  |  |  |
| F | -2 | -2 | -2 | -4 | -2 | -3 | -3 | -3 | -3 | -3 | -1 | -3 | -3 | 0 | 0 | 0 | -1 | **6** |  |  |
| Y | -2 | -2 | -2 | -3 | -2 | -3 | -2 | -3 | -2 | -1 | 2 | -2 | -2 | -1 | -1 | -1 | -1 | 3 | **7** |  |
| W | -2 | -3 | -2 | -4 | -3 | -2 | -4 | -4 | -3 | -2 | -2 | -3 | -3 | -1 | -3 | -2 | -3 | 1 | 2 | **11** |

**Question 1**

Which of the proteins from the database is most similar to our query? Which is the least similar?

**Question 2**

What problems did you encounter that may have affected your calculated scores?

**Question 3**

How similar do you expect their H4 proteins to be? Briefly explain the reason for your answer.

**Question 4**

If we were to compare the *nucleotide* sequences for the gene encoding this protein between humans and chimps, do you think they would be identical? Explain.

**Nucleotide Alignment screenshots**

**Question 5**

How many differences did you find between the query and subject? What percent identity is there between these sequences? Is this at all surprising to you? Explain.

**Question 6**

Imagine that a colleague asks you to align a conserved metabolic protein coding gene sequence from a dog to its human homolog. Which sequence type, DNA or protein, would you expect to exhibit the highest percentage of identities? Why?