Birzeit University

Biology and Biochemistry Department

MOLECULAR BIOLOGY LAB

BIOL313

Instructor: Dr. Mahmmoud A. Srour

Aseel Abudayah : Teaching Assistance

BIOINFORMATICS sheet

By Roa Abudayah 1192923

The sequences:

DNA sequence/ Query DNA:

ACTTTCACAATCTGCTAGCAAAGGTTATGCAGCGCGTGAACATGATCATGGCAGAATCACCAGGCCTCAT

CACCATCTGCCTTTTAGGATATCTACTCAGTGCTGAATGTACAGTTTTTCTTGATCATGAAAACGCCAAC

AAAATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAAGGGAACCTTGAGA

GAGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACAAC

TGAATTTTGGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGTTGC

AAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATTAGATG

TAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAAGGTGGTTTG

CTCCTGTACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACCAGCAGTGCCATTTCCATGT

GGAAGAGTTTCTGTTTCACAAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTTCCTGATGTGGACTATG

TAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCAC

Amino acid sequence/ Query Amino acid sequence:

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRP

KRYNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNPC

LNGGSCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEG

YRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQST

QSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKIT

Part 1: blast search query DNA and protein sequence

nucleotide blast – blastn >>>

* Gene : F9
* Protein : Homo sapiens coagulation factor IX (F9)
* 100% similar to Query DNA .
* Source: Homo sapiens (human)
* Tissue : could be Blood

|  |  |  |  |
| --- | --- | --- | --- |
| Most similar sequences | Name of gene | SOURSE | % identity |
| Second | F9 | Homo sapiens | 100.00% |
| Third | F9 | Pan paniscus | 99.71% |
| fourth | factor IX | Pan troglodytes | 99.57% |

protein blast – blastp >>>

because it didn’t have exons or introns while the nucleotides query sequences have

|  |  |  |  |
| --- | --- | --- | --- |
| Most similar sequences | Name of gene | SOURSE | % identity |
| First | coagulation factor IX | Homo sapiens | 100.00% |
| Second | coagulation factor IX | Gorilla gorilla gorilla | 100.00% |
| Third | coagulation factor IX isoform 1 preproprotein | Homo sapiens | 100.00% |
| fourth | factor IX | Homo sapiens | 100.00% |

What animals, besides the ones listed previously , have this protein and have a high score?

* Gorilla gorilla gorilla
* Pan troglodytes
* Hylobates moloch

Blatx >>>

No it shows different answers

Part II: sequence analysis:

* How many nucleotides are present in the sequence with accession number X03236?

1948 bp

- Look at "Features" section of the file. The region of the sequence that encodes the a-amylase precursor is indicated by "CDS". What are the numbers of the starting and ending nucleotides for the sequence that encodes the α -amylase precursor?

181>>>1719

* Part of the precursor protein is a signal peptide, which is cleaved as the protein is secreted from the cell. Following removal of signal peptide, the protein is referred to as the mature protein

sig\_peptide 181..267 \\\ mat\_peptide 268..1716

* Using the "Features" section of the file, identify the part of the gene that encodes the signal peptide (identified as "sig peptide")

atgaaacaac aaaaacggct ttacgcccga ttgctgacgc tgttatttgc gctcatcttc

ttgctgcctc attctgcagc agcggcg

* Using the "Features" section of the file, identify the part of the gene that encodes the mature peptide (identified as "mat peptide")

gca aatcttaatg ggacgctgat gcagtatttt

301 gaatggtaca tgcccaatga cggccaacat tggaagcgct tgcaaaacga ctcggcatat

361 ttggctgaac acggtattac tgccgtctgg attcccccgg catataaggg aacgagccaa

421 gcggatgtgg gctacggtgc ttacgacctt tatgatttag gggagtttca tcaaaaaggg

481 acggttcgga caaagtacgg cacaaaagga gagctgcaat ctgcgatcaa aagtcttcat

541 tcccgcgaca ttaacgttta cggggatgtg gtcatcaacc acaaaggcgg cgctgatgcg

601 accgaagatg taaccgcggt tgaagtcgat cccgctgacc gcaaccgcgt aatttcagga

661 gaacaccgaa ttaaagcctg gacacatttt cattttccgg ggcgcggcag cacatacagc

721 gattttaaat ggcattggta ccattttgac ggaaccgatt gggacgagtc ccgaaagctg

781 aaccgcatct ataagtttca aggaaaggct tgggattggg aagtttccaa tgaaaacggc

841 aactatgatt atttgatgta tgccgacatc gattatgacc atcctgatgt cgcagcagaa

901 attaagagat ggggcacttg gtatgccaat gaactgcaat tggacggttt ccgtcttgat

961 gctgtcaaac acattaaatt ttcttttttg cgggattggg ttaatcatgt cagggaaaaa

1021 acggggaagg aaatgtttac ggtagctgaa tattggcaga atgacttggg cgcgctggaa

1081 aactatttga acaaaacaaa ttttaatcat tcagtgtttg acgtgccgct tcattatcag

1141 ttccatgctg catcgacaca gggaggcggc tatgatatga ggaaattgct gaacagtacg

1201 gtcgtttcca agcatccgtt gaaagcggtt acatttgtcg ataaccatga tacacagccg

1261 gggcaatcgc ttgagtcgac tgtccaaaca tggtttaagc cgcttgctta cgcttttatt

1321 ctcacaaggg aatctggata ccctcaggtt ttctacgggg atatgtacgg gacgaaagga

1381 gactcccagc gcgaaattcc tgccttgaaa cacaaaattg aaccgatctt aaaagcgaga

1441 aaacagtatg cgtacggagc acagcatgat tatttcgacc accatgacat tgtcggctgg

1501 acaagggaag gcgacagctc ggttgcaaat tcaggtttgg cggcattaat aacagacgga

1561 cccggtgggg caaagcgaat gtatgtcggc cggcaaaacg ccggtgagac atggcatgac

1621 attaccggaa accgttcgga gccggttgtc atcaattcgg aaggctgggg agagtttcac

1681 gtaaacggcg ggtcggtttc aatttatgtt caaaga

Analyze the sequence and identify the ORF >>>

* Do all three forward reading frames of the DNA sequence contain ORFs?

No

* List all the ORFs linger than 50 amino acids:

**5'3' Frame 1**:

MKQQKRLYARLLTLLFALIFLLPHSAAAAANLNGTLMQYFEWYMPNDGQHWKRLQNDSAYLAEHGITAVWIPPAYKGTSQADVGYGAYDLYDLGEFHQKGTVRTKYGTKGELQSAIKSLHSRDINVYGDVVINHKGGADATEDVTAVEVDPADRNRVISGEHRIKAWTHFHFPGRGSTYSDFKWHWYHFDGTDWDESRKLNRIYKFQGKAWDWEVSNENGNYDYLMYADIDYDHPDVAAEIKRWGTWYANELQLDGFRLDAVKHIKFSFLRDWVNHVREKTGKEMFTVAEYWQNDLGALENYLNKTNFNHSVFDVPLHYQFHAASTQGGGYDMRKLLNSTVVSKHPLKAVTFVDNHDTQPGQSLESTVQTWFKPLAYAFILTRESGYPQVFYGDMYGTKGDSQREIPALKHKIEPILKARKQYAYGAQHDYFDHHDIVGWTREGDSSVANSGLAALITDGPGGAKRMYVGRQNAGETWHDITGNRSEPVVINSEGWGEFHVNGGSVSIYVQR

**5'3' Frame 2:**

MIHSRGNRLSRLSKHGLSRLLTLLFSQGNLDTLRFSTGICTGRKETPSAKFLP

**5'3' Frame 3**:

MSRKRHIGAFLLEENIGKMVFVKNSEYLYNIICFTLKGEENHETTKTALRPIADAVICAHLLAASFCSSGGKS

**3'5' Frame 1**:

MSHRRFAGRHTFALPHRVRLLLMPPNLNLQPSCRLPLSSRQCHGGRNNHAVLRTHTVFSLLRSVQFCVSRQEFRAGSLLSSRTYPRRKPEGIQIPL

**3'5' Frame 2**:

MIKICFVQIVFQRAQVILPIFSYRKHFLPRFFPDMINPIPQKRKFNVFDSIKTETVQLQFIGIPSAPSLNFCCDIRMVIIDVGIHQIIIVAVFIGNFPIPSLSLKLIDAVQLSGLVPIGSVKMVPMPFKIAVCAAAPRKMKMCPGFNSVFS

**3'5' Frame 3**:

MTTGSERFPVMSCHVSPAFCRPTYIRFAPPGPSVINAAKPEFATELSPSLVQPTMSWWSK

The longest ORF corresponds to the coding region of the a-amylase gene. How long is this ORF?

It is 5'3' Frame 1

- Does the BLAST search indicate that this ORF corresponds to the gene used for this analysis?

Yes for , Bacillus licheniformis 584 alpha-amylase gene.

- Do any other genes also correspond to this sequence?

No

* How similar are these other genes?

510/510(100%) , the score = 1016bits

Part III: Primer-BLAST

* How many pairs of primers were suggested by the program?

10 pairs of primers

* what is the most appropriate pair of primers for this ORF? Based on the criteria of digestion Primers set 3 :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Sequence | Length | Tm | GC% | Self complementarity | Self 3' complementarity |
| Forward primer | TTCACATTGAAAGGGGAGGAGAA | 23 | 59.60 | 43.48 | 6.00 | 0.00 |
| Reverse primer | CCTTCAGGAAATCCGTCCTCT | 21 | 59.17 | 52.38 | 4.00 | 0.00 |

Because : short primers (ideal 18-28) the annealing temperature For the forward and reversed primer is close and have a high GC composition in addition to the self complementary and the 3’ complementary that = 0

* List the next most appropriate two pairs of primers along with their parameters?

Primer set 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sequence (5'->3') | sequence | Tm | GC% | Self complementarity | Self 3' complementarity |
| Forward primer | TTTCACATTGAAAGGGGAGGAGA | 177 | 59.60 | 43.48 | 8.00 |
| Reverse primer | TCCTTCAGGAAATCCGTCCTC | 1729 | 59.17 | 52.38 | 6.00 |

Primer set 8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sequence (5'->3') | sequence | Tm | GC% | Self complementarity | Self 3' complementarity |
| Forward primer | TTTCACATTGAAAGGGGAGGAGAA | 60.14 | 41.67 | 8.00 | 0 |
| Reverse primer | TCCTTCAGGAAATCCGTCCTCT | 60.56 | 50.00 | 6.00 | 0 |

Part IV: Locate and map restriction enzyme cleavage sites

Select two enzymes that can be used for cloning of the ORF for a-amylase gene?

HaeIII , EaeI

* For the following enzymes, identity the number of cuts and nucleotide positions for each cut in this query DNA sequence as well as the restriction site?

|  |  |  |  |
| --- | --- | --- | --- |
| Restriction enzyme | Restriction site  (specificity) | Number of cuts | Cut positions |
| EcoRI | Ghttp://nc2.neb.com/NEBcutter2/cut5sm.gifAATThttp://nc2.neb.com/NEBcutter2/cut3sm.gifC | 0 | -------------- |
| Clal | AThttp://nc2.neb.com/NEBcutter2/cut5sm.gifCGhttp://nc2.neb.com/NEBcutter2/cut3sm.gifAT | 1 | 869/871 |
| Kpnl | Ghttp://nc2.neb.com/NEBcutter2/cut3sm.gifGTAChttp://nc2.neb.com/NEBcutter2/cut5sm.gifChttp://nc2.neb.com/NEBcutter2/pix.gif | 1 | 741/737 |
| Haelll | GGhttp://nc2.neb.com/NEBcutter2/cut53sm.gifCChttp://nc2.neb.com/NEBcutter2/pix.gif | 2 | 323, \*1589 |
| Ddel | Chttp://nc2.neb.com/NEBcutter2/cut5sm.gifTNAhttp://nc2.neb.com/NEBcutter2/cut3sm.gifGhttp://nc2.neb.com/NEBcutter2/pix.gif | 3 | 15/18, 1343/1346, 1830/1833 |
| PstI | Chttp://nc2.neb.com/NEBcutter2/cut3sm.gifTGCAhttp://nc2.neb.com/NEBcutter2/cut5sm.gifG | 1 | 258/254 |
| Sa1I | Ghttp://nc2.neb.com/NEBcutter2/cut5sm.gifTCGAhttp://nc2.neb.com/NEBcutter2/cut3sm.gifC | 1 | 1275/1279 |