

1. [/ 5] Use the DNA sequence below for the following question. Genomic DNA is double-stranded, but by convention only one of the strands is shown below. The presence of a complementary strand is assumed.

```
ATGGGCAGAGACAAAAAATACAGCGCTTCTAGATATTGCAAGAGATATAGG
AGGAGATGAAGCTGTAGAAGTTGTAAAAGCCTTAGAAAAGAAAGGAGAAGCAA
CAGATGAGGAATTGGCAGAATTAAGTGGAGTAAGAGTTAATACGGTGAGAAAA
ATCTTATACGCCCTGTACGATGCTAAGCTTGCAACCTTTAGAAGAGTTAGAGA
TGACGAGACTGGTTGGTATTATTACTGGCGCATTGATACTAAAAGATTAC
CGGAGGTTATTAGAACAAGAAAGTTGCAAGAGCTTGAAAAGTTAAAGCAGATG
CTTCAGGAAGAAACCAGCGAGACCTATTATCACTGTGGAAGTCCAGGGCATCC
AAAGCTAACATTTGACGAGGCCTTTGAGTACGGGTTCCAATGTCCAATATGTG
GAGAGATACTTTACGAGTATGACAACTCAAAAATAATTGAAGAACTCAAAAAG
CGAATTGAAGAATTAGAGATTGAAGTCCGACTGAGAAGTCCACCAAAGAAGA
AAAACCAAAAAAAGCAACAAGAAGAAAAAAGTCAAGATCAGGGAAAAAGAAGA
AATAA
```

Your lab has obtained and sequenced genomic DNA from a hyperthermophilic microbe. You want to amplify the full sequence above from the genomic DNA, using a PCR protocol.

(a) [/ 2] Use the Primer3Plus software (or other method of your choice) to identify two oligonucleotide primers that will amplify the entire sequence (hint: be sure to use the "Cloning" option in the upper left task selector, and be sure to select the entire sequence). Indicate the actual primer sequences, where they would anneal to the above strand (or its complement), and the 5' and 3' ends of each primer. Indicate the size of the DNA product expected. Are the primers 'acceptable' according to the software? If not, what are the potential problems with the primers?

(b) [/ 3] You perform the PCR and run an agarose gel to view the results. You expected to see a single band for your PCR product, but you see multiple bands of different sizes, including the size that you expected. You do see bands for a marker DNA size standard, so you know that the gel and staining worked properly. What might have gone wrong with the PCR? Indicate changes you could make to your PCR protocol that should help to solve the problem.

2. [/ 5] Go to the [UCSC genome browser](#), select the Dec. 2013 (GRCh38/hg38) assembly, and navigate to the human CFTR gene.

(a) [/ 1] Which chromosome is the gene on, and which arm of the chromosome?

(b) [/ 2] Use the Gencode V36 'full' track to answer the following (note: this track is found in the "Genes and Gene Predictions" category and can be activated/alterd if it is not initially visible). Use the gene version that is highlighted (blue background) to answer the following:

- i. How many exons make up the CFTR gene?
- ii. How much space does the entire gene occupy, including both exons and introns?

(c) **[2 / 2]** Activate the OMIM genes track to identify the OMIM entry for CFTR (note: this track is found in the "Phenotype and Literature" category and can be activated/alterd if it is not initially visible). Visit the OMIM gene page to determine the function of CFTR, and whether it is disease associated.

- i. What is the normal biological function of this gene?
- ii. Which, if any, diseases is it associated with?

3. **[5 / 5]** Prediction of protein function and conservation using BLAST.

Example polypeptide sequence:

```
MGRDKKNTALLDIARDIGGDEAVEVVKALEKKGEATDEELAELTGVRVNTVRKILY
ALYDAKLATFRRVRDDETGWYYYYWRIDTKRLPEVIRTRKLQELEKLKQMLQEET
SETYYHCGTPGHPKLTDFEAFEYGFQCPICGEILYEYDNSKIIEELKKRIEELEIELGL
RSPPKKEEPPKATRRKKSRSRGKKKK
```

(a) **[2 / 2]** Use <https://blast.ncbi.nlm.nih.gov/Blast.cgi> to determine the probable function of the protein encoded by the example polypeptide sequence above. Choose the Basic BLAST program "protein BLAST". Paste the query sequence, and search the non-redundant protein sequences within "archaea" (by typing this into the "Organism" search set).

- i.
- ii.
- iii.

(b) **[1 / 1]** Go back to the <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, and use standard protein BLAST determine if there are proteins with biologically relevant similarity in Homo sapiens (note: it may take a few minutes for results to arrive).

Indicate the most similar protein, and identify the organism this gene is from. To assess the significance of the alignment, give the E-value as well as the identity and positive percentages.

(c) **[2 / 2]** Next, perform a BLAST search again for similarity in Homo sapiens, but this time using the DeltaBLAST option (note: Delta stands for 'Domain Enhanced Lookup Time Accelerated').

- i. Does this modify your previous answers in part b), and if so, how? How likely is it that the function of the top hit is the same as the function of the query protein sequence? Which information in the BLAST results allowed you to make this judgement?

