

TASK 1

Write all questions and answers in a PDF file

1. For the human SIRT7 gene:

a. Download the sequence of the human SIRT7 gene. Paste the first 20 nucleotides of the gene (complementary strand). Paste the same 20 nucleotides in the main strand. Both sequences should be pasted from 5' to 3'

Main strand: CTGCCGTGTGAGGCGGAAGC

Complementary strand: GCTTCCGCCTCACACGGCAG

b. Paste 100 nucleotides upstream of the gene (in the promoter region)

```
gtggttttggtgaaaggaaccaacacattaacgatttttccccagaagcc
actgaataattcttttttggtgtatttttgccttctctgttggtgtctggc
cttgagcaggtcgaggagtagagtacataatcagtgccacacagaggc
agggcgtgtctgagagactaatgcctgtcctgcctctggcctgtgtcctg
ggtgtgtaccataacctgctggtggccagtcacacacatttccacctgcccc
tgagcagcagcctttgataccatgggtgtcccttgagctgtgagctagagt
gtgtccccgataaaactgtgagacctcctgttcttggtgaccccagttgg
gctttggccctccagcaaccttgtgtcccagccctgccttcttcccac
cccctcaatctgctagtcctgaagcctttaacaaacgggagtgggcac
agaaagccttctcctggcaacaggacaagggctgcccgcgtgtccagccc
tgttgctctcctggcgctgagaggtgggtccaagcagagttgatcagtc
ctgcctgcccctgcctaggtctagccagggaggttgtgtgtgtgggagaa
gaccccgggatctgctgggtgggatggaggtatttaaagatctaagtga
ggaggggctgggttttggtgctgggggtgagggccgagcgctcacacttcg
tgtagggcgggcaggggctgacctcttgagggcagtgcccttggtgcc
ctacccctgcccctgcgcagtatattattgtctaaattattgtccaggaggg
cagcactgggctggccccccgggtatttattgtgtacatagtgatgt
ttgtgatatataagggttttctttattttgtatatgatcaataaaccttt
aaagagactggcaggtgcttcttggggcgacgcggcgggcgagcgga
aggggtgggtggggcggggcccaccttgcgcagtgccctgagcgcggc
CTGCCGTGTGAGGCGGAAGCGGAAGAGCAGGTCTCCAGGGGAGCGATGGC
AGCCGGGGGTCTGAGCCGCTCCGAGCGCAAAGCGGCGGAGCGGGTCCGGA
GGTTGCGGGAGGAGCAGCAGAGGGAGCGCCTCCGCCAGgtacgcgcgcgc
cgctccccggccccggccatgccccggccccgcgcgcgcgtcacgcgtccgc
tgccccagGTGTGCGGCATCCTGAGGAAGCGGCGGCGGAGCGCAGCGC
CGAGGAGGGCCGGCTGCTGGCCGAGAGCGCGGACCTGGTAACGGAGCTGC
AGGGCCGGAGCCGGCGGCGGAGGGCCTGAAGCGGCGGAGGAGGAGgcg
agttccgcgtgcggcgcgcggcgcccccggtttcgggagcagctggggc
gacgggcggtccccgggtggggcgggccccggggcggtgaccacctggcgtc
ttggcagGTGTGCGACGACCCGGAGGAGCTGCGGGGGAAGTCCGGGAGC
TGGCCAGCGCCGTCCGGAACGCCAAATACTTGGTCGTCTACACAGGCGCG
GGAATCAGCACGgtagggaaggagggcgagggcgtaacccaggacggagta
tgagctccagtaatcgcgaaaaactcgccctttaaagcagctctaaggttt
tttctcttaaagaacgaaatgacaaaaacttacctaaggtaaagcgttt
ttaaagcgttggcctctgtgttacagccagttaaaaaaaacaaggagtag
agatacgaatggggtgtagtagccgactgctcgcaggcacccccagggtta
tgtggacagagctaagcccaaagttgtgattttccactctgttctgtcca
tgtcgagggaagataagtagaaagtgcacagtaagagccagaataacacc
aggtgaaggagagaattgcattgtgttttgagaagtttactgacaagtt
atcctgggctgtgggacatcactagctttgaaagtgtagctggcacctcg
tccatctaatttgatgggtgtgtgtgggggtgttgggcacgcgtcggccta
gcagatctgaaccaggtgatttctgttctcaggaagcttttaggtgaca
aggatcaggcatgtgaacaaataaccatactgtaaagctggctgtgtctgg
```

Bet Bardají
Biomedical engineering

gtcgctagagcaaaactcagacacacactctgcgctcttgaggttgggaaa
cccacctgcggttggtttttgtgggaggtggccttgattgggccttgaag
gatgggtgagatttacagaaggttggaatgagggcactccaagcaaagaac
agcagaggtcagcagcaagaatgcaaaaaggaggttccactactgactca
aatacccgagccctgggttttagtctcctccctccacaagtccatgta
aaagtccagtcagtgctgggcgctgggtcatgcctataatcctagcact
ttgggagggcaggtgggcggatcatgaggtcagtagttcaagaacaacc
tggccaacacggtgaaaccctgtctctactaaaaatacaaaaattagtcg
ggcatggtggcgggcacctgcaatcccagctactcgagaggtgaggcag
gagaatcacttgaaactggaaggcggaggttgacgtgagccgagatcatg
ccactgcattccagcctgggcgaaagagctaaactccgtctctcaaaaaa
aaaaaaaaaaaaaaaaagtcagtcagtaattatgtaacagtcacgtgac
ctgttatggaacttccaatggcaactaaaagcacatgcagctagtggatt
tcacgcgagtggttgaggttcccgctcttgaaatgtgactgtcggaactact
gtccgggggggtggtgcatttttctgagtttaagcaggagtcgggagtc
ccaaaagggaacacaagacaccttgatcctggcatatcttggtgcccctc
tgtgggcctcagtttggttcacaggattgaaaacctgggaagttagatgc
tcactcactctgaagaagttgttctgccttggtaagggtggagcgggaat
agtcagcactgggacatgagaatggacagtcgcctggaccacactaggga
ttcaccatttgctaaatgtgtgagctgtgggctcgccctgggggcactt
taggaacatgactagtcttcccctgcagtggtggaggacacatgtgccaca
gagcccagctttgtgctcggtgccagagaggttccggaggcaggcaggg
ctgctgtgacgcctggaaggatgagccaggccaggcgggaacggaagtcc
aggtagaaggaggagccgaattggggtacactccatatgggtcaggga
ggtcagcctgtggaatgaatagaggccaacatgcaggccagcccgaatg
cggcaggagtgacagtggtttccggtttctgggaattctgccagtaccta
cagtggtgcttttgacttggttacctttttctcgacatgcagGCAGC
GTCTATCCCAGACTACCGGGGCCCTAATGGAGTGTGGACACTGCTTCAGA
AAGGGAGAAGCGTTAGGtaagcggggccaggcatggcctcccataggt
gggcagcggcagcagggcctgagctccagctctcctcaccttgccttcc
tttctgcctggcagTGCTGCCGACCTGAGCGAGGCGGAGCCAACTCAC
CCACATGAGCATCACCCGTCTGCATGAGCAGAAGCTGgtaagagccctgg
gtggctggtacacttgccagggaccaggcagagcaccttggtgccagtg
ggcaactaactgcacccgcctctgtctgccagttgactcccatgatgag
caccaccaagcgggttaggcgcgggtttgatcctcctgtgcttgactc
tccaggacagaaggggagctccccctctgaaccatcccatccgcagccag
cctcagcctcagagctgctggtggcctttcccttttgatgattcccttgca
tttttctgtggacctcagaagccatcctagtccaggggaggcttatagg
acatctctggggaccttctgctgaacacctcaaccaatggggtgtagtg
gttgatctgcctaggtccccagggacctgaaatgtcatgaccaagagag
catggatctggggcagagtggccttgccaaaccccgagccacttcccaa
ccttgccgggacggtgcaacctcttgccctcccagccactccagggtgat
cagggttgagggaaggacagccctccccaccacaggccctcttgactcc
tggtggttgagacctgttggtgtgttttactttctaaaggctctctggggga
cggagctgcccctggttttgggagccatcggcggggctcagaacagcctc
gctgtggccgggtactgacctccccaccacctggcagccatccctccc
cttgccaccggtctgtgccttccagGTGCAGCATGTGGTGTCTCAGAACTG
TGACGGGCTCCACCTGAGGAGTGGGCTGCCGCGCACGGCCATCTCCGAGC
TCCACGGGAACATGTACATTGAAgtgagcagtcctgcagggaacccagggt
ctccatgggcagggcgggtcccaactcactgtgcctcttgccctctagGTCT
GTACCTCTGCGTTCCCAACAGGGAGTACGTGCGGGTGTTCGATGTGACG
GAGCGCACTGCCCTCCACAGACACCAGACAGGCCGACCTGCCACAAGTG
TGGGACCCAGCTGCGGGACACCATTGTGCACTTTGGGGAGAGGGGGACGT
TGGGGCAGCCTTTGAACTGGGAAGCGGCGACCGAGGCTGCCAGCAGAGCA
GACACCATCCTGTGTCTAGGGTCCAGCCTGAAGgtacgtgccgatgacac
aatgagtgaaccgagcccctgcccggccgagggtgtccagctctgcgcc
cagcactgtacagacttgcccttggtgtgtgtgaggtgtctgtctgtctg
cttccacagGTTCTAAAGAAGTACCCACGCCTCTGGTGCATGACCAAGCC
CCCTAGCCGGCGGCCGAAGCTTTACATCGTGAACCTGCAGgtaactcggg
tgctgagagccacgtccttagatctgggtcttagaacgcacagccagaga
caccacacacccatgcaccaggggcgtctgatagggcccccgtgggtgct
caggggagcaccgactgagcccgtagggggccaaggctgacaggccaccggg
aagggttgggtgctgttactctcactcggctttccctgtcctcagTGA
CCCCGAAGGATGACTGGGCTGCCCTGAAGCTACATGGGAAGTGTGATGAC
GTCATGCGGCTCCTCATGGCCGAGCTGGGCTTGAGATCCCCGCCTATAG
CAGgtgagtgagccgctgcagcagcctgctccccgcacctctgtgtgct
gggccttgctgtcttctctcgtgagctgagtggtggaggaagctctgagg
tggttgagtggtgcctgagggcatgactgaagcgtggtggtctccagagg

```
gcctgacctcggtggttgccggagaccctgcgtgtgccactcctgccctg
gctgatgtggcacacacaatccccgcggggagagggattctgcccgcgtg
ctcctgtctccaggcctccccgtggagctctccgagatgctggtgggaag
catctggaggggacgagcactcggcagctctggtcagacagaatctgtgt
gcttggttttgggagttggcgtactttgggaaagcttaaaactgtgc
cttaatacagaatttgtgataatttagacttggtgtatgtattgagtaaa
aagtttacactctctttctctgtgaatttccagggctttataggggaaat
caataacttcttttaataaagggttcaagaaattaaggatcccttcacc
ttctgggacctggcacttctgtatgttatgtgtgtggtgttctgtgatgt
gggctatcgtgtactgtattttttttacattaacttagctcattttcc
ttatcagtgcgatctgtatcttaagttatgatctgtggttctgcattctc
cgtcagacacatgctttcttcacggggtcgtctgtaggccacgcctccct
agtcagctgggaagggggagaggggtctggtccacctgccccagcggtaca
agtggaaaggtggggcccagagttgctagtactcatccctggagacggag
gcagccctggggccactgctgccccaccctgtgtgtgcacgccgctcagt
ggtggacaaggacacggagttttgaggagaccgagctagtgtgggtgccg
acctttgagtcaccacctaagaggtgacctctccacatccgttctgcag
cttggttaacaatgaagctgccgccaaccagagccccgccgcagttgacac
gggaggggaaggggatgggaaggcagggaccgcagacagctttcccgagct
ggggcaggtgtgactgcgagagggtcccaggcccgctgatgccgctttc
cctttttggcagGTGGCAGGATCCCATTTTCTCACTGGCGACTCCCCTGC
GTGCTGGTGAAGAAGGCAGCCACAGTCGGAAGTCGCTGTGCAGAAGCAGA
GAGGAGGCCCCGCCTGGGGACCGGGGTGCACCGCTTAGCTCGGCCCCCAT
CCTAGGGGGCTGGTTTGGCAGGGGCTGCACAAAACGCACAAAAGGAAGA
AAGTGACGTAATCACGTGCTCGATGAAGAACAGTTGGCACTTTGCAGATG
GCCAGTGTACGGTGAAGGCTGGGTTGCCCCACGGGTCTAGGGAGAACG
AACTCTTTGGGGATGACATTTTCACCGTGACATTTTAGCCATTTGTCCT
TGAGGAAGCCCCCTTGCACTGCTGCGGTTGTACCCTGATACGGCCTGGCCA
TCGAGGACACCTGCCCATCCGGCCTCTGTGTCAAGAGGTGGCAGCCGCAC
CTTTCTGTGAGAACGGAACCTCGGGTTATTTAGCCCCGGCCTGCAGAGTG
GAAGCGCCAGCGGCCTTTCCTCGCTCACCAGGCCAGTCTCAGGGCCTCA
CCGTATTCTACTACTACTTAATGAAAAAGTGTGAACCTTATAGAATCCT
CTCTGTACTGGATGTGCGGCAGAGGGGTGGCTCCGAGCCTCGGCTCTATG
CAGACCTTTTATTCTATTAAACGTTTCTGCACTGGC
```

c. Write the positions of the SIRT7 gene in the genomic sequence (the positions should be referenced to the NC_ sequence)

NC_000017.11 (81911939..81918176, complement)

d. Write the positions of the exons in the human SIRT7 gene

Exon count: 10

1. Exon 1: 1-138
2. Exon 2: 139-276
3. Exon 3: 277-381
4. Exon 4: 382-452
5. Exon 5: 453-525
6. Exon 6: 526-624
7. Exon 7: 625-861
8. Exon 8: 862-942
9. Exon 9: 943-1049
10. Exon10: 1050-1725

- e. Use the SIRT7 gene sequence and the positions of the exons to make the corresponding SIRT7 transcript (mRNA). Paste the transcript sequence. Indicate the code used to make the transcript (any programming language can be used)

```
GCUGCCGUGUGAGGCGGAAGCGGAAGAGCAGGUCUCCAGGGGAGCGAUGGCAGCCGGGGGUCUG
AGCCGCUCCGAGCGCAAAGCGGCGGAGCGGGUCCGGAGGUUGCGGGAGGAGCAGCAGAGGGAGC
GCCUCCGCCAGGUACGCCCGCCGCCUCCCCGGCCCGGCCAUGCCCGGCCCGCCGCCGCUCA
CCGUCCGCCUGCCCGCAGGUGUCGCGCAUCCUGAGGAAGGCGGCGGCGGAGCGCAGCGCCGAGG
AGGGCCGGCUGCUGGCCGAGAGCGCGGACCUGGUAACGGAGCUGCAGGGCCGGAGCCGGCGGGCG
CGAGGGCCUGAAGCGGCGGCAGGAGGAGGCGAGUUCGCGUGCGGCGCGCGGGCGCCCCCGGUU
UCGGGAGCAGCUGGGGCGACGGGCGGUCCCGGUGGGGCGGCCCGGGGCGGUGACCACCCUGGC
GUCUUGGCAGGUGUGCGACGACCCGGAGGAGCUGCGGGGGAAGGUCCGGGAGCUGGCCAGCGCC
GUCCGGAACGCCAAAUACUUGGUCGUCUACACAGGCGCGGGAUACAGCACGGUAGGGAGGGAGG
CGGAGGCGUACCCCAGGACGGAGUAUGAGCUCCAGUAAUCGCGAAAAACUCGCCUUUAAAGCAG
CUCUAAGGUUUUUUCUCUUAAGAAACGAAUAGCCAAAACUUAACUAAGGUAACGCUUUUUA
AACGCUUGGCCUCUGUGUACAGCCAGUUAACAAAAACAAGGAGUAGAGUACGAAUGGGGUGU
AGUAGCCGACUGCUCGCGAGGCACCCCGAGUUUAUGUGGACAGAGCUAAGCCCAAAGUUGUGAU
UUCCACUCUGUUCUGUCCAUUGUCGAGGGAAGUAAGUAGAAAGUGACACAGUAAGAGCCAGAAU
ACACCAGGUGAAGGAGAGAAUUGCAUUGUGUUUUGAGAAGUUUCACUGACAAGUUUUAUCCUGGGC
UGUGGGACAUCACUAGCUUUGAAAUGUAGCUGGCACCUCGUCCAUCUAAUUGAUGGGUGUGU
GUGGGGUGUUGGGCACGCGUCGGCCUAGCAGAUUCGAACCCAGGUGAUUUCUGUUCUACAGGAAG
CUUUUAGGUGACAAGGAUCAGGCAUGUGAACAAAUAACCAUACUGUAAAGCUGGCUGUGCUGGG
UCGCUAGAGCAAACUCAGACACACACUCUGCGCUCUUGGAGUUGGGAAACCCACCUGCGUUGGC
UUUUUUGGGGAGGUGGCCUUGAUUGGGCCUUGAAGGAUGGGUGAGAUUUACAGAAGGUUGGAU
GAGGCACUCCAAGCAAAGAACAGCAGAGGCUCAGCAGCAAGAAUGCAAAGGGAGUUCACUAC
UGACUCAAAUACCCGGAGCCUUGGGGUUAGUCUCCUCCUCCACAAGUCACAUGUAAAAGUCC
AGUCAUGCUGGGCGCGGUGGCUCAUGCCUAUAAUCCUAGCACUUUGGGAGGCCGAGGUGGGCGG
AUCAUGAGGUCAGUAGUUAAGAACAACCUGGCCAACACGGUGAAACCCUGUCUCUACUAAAAA
UACAAAAAUUAGUCGGGCAUGGUGGCGGGCACCUGCAAUCCAGCUACUCGAGAGGCUGAGGCA
GGAGAAUACAUUGAAAACUGGAAGGCGGAGGUUGCAGUGAGCCGAGAUCAUGCCACUGCAUCCA
GCCUGGGCGAAAAGAGCUAAAACUCCGUCUCUAAAAAAAAAAAAAAAAAAGUCCAGUCAUG
UAAUUAUGUAAACAGUCACGUGACCUGUUAUGGAACUCCAAUGGCAACUAAAAGCACAUGCAGC
UAGUGGAUUUCAUCGGAGUGUUUGAGGUUCCCGUCUUGAAUGUGACUGUCGGAACUACUGUCCG
GGGGGUGGUGCAUUUUUCUGAGUUUAAGCAGGAGUCGGGAGUCCCAAAGGGAACACAAGAC
ACCUUGAUCCUGGCAUAUCUUGUGUGCCUCUGUGGGCCUCAGUUUGUUUACAGGAUUGAAAA
CCUGGGAAGUUAGAUGCUCUACUACUGAAGAAGUUGUUCUGCCUUUGUUUAGGUGGAGCGGG
AAUAGUCAGCACUGGGACAUGAGAAUGGACAGUCGCCUGGACCCACCUAGGGAUUCACCAUUUG
CUAAAUGUGUGAGCUGUGGGCUCGGCCUGGGGGCACUUUAGGAACAUAGCUAGUCUCCCCUG
CAGUGUGGAGGACACAUGUGCCACAGAGCCCAGCUUUGUGCUCGGUGCCAGAGAGGCUUCCGGA
GGCAGGCAGGGCUGCGUGCAGCCUGGAAGGAUGAGCCAGGCCAGGCGGGAACGGAAGUCCAGG
UAGAAGGGAGGAGCCGAAUUGGGGUACACUCCAUAUGGGCUCAGGCAGGUCAGCCUGUGGAUG
AAUAGAGGCCAACAUAGCAGGCCAGCCGGAUUGCGGCAGGAGUGACAGUGGCUUCCGUUUCUG
GGAAUUCUGCCAGUACCUACAGUGGUGCCUUUUGACUUGGCUUACCUUUUUUCUGGACAUGCAG
GCAGCGUCUAUCCAGACUACCGGGGCCUAAUGGAGUGUGGACACUGCUUCAGAAAGGGAGAA
GCGUUAGGUAAAGCGGGCCAGGCAUGGCCUCCCAUAGGCUGGGCAGCGGCAGCACGGGCCUGA
GCUCCAGCUCUCCUACCUUGCCUUCUUCUGCCUGGCAGUGCUGCCGACCUGAGCGAGGCCG
AGCCAAACCCUACCCACAUGAGCAUACCCGUCUGCAUGAGCAGAAGCUGGUAAGAGCCUGGG
UGGCUGGUACAUUGCCAGGGACCAGGCAGAGCACCUGGUGCCCAGUGGGCAACUACUGCAC
CCGCCCUCUGUCUGCCAGUUGACUCCCAUGAUGAGCACCCACCAAGCGGGUUAAGCCGCGGGU
UGAUCCUCCUGUGCUUGACUCUCCAGGACAGAAGGGGAGCUCCCCUCUGAACCAUCCCAUCCG
CAGCCAGCCUCAGCCUCAGAGCUGCUGGUGGCCUUUCCCCUUUGUAUCCCUUGCAUUUUUCU
GUGGACCUCAGAAGCCAUCCUAGUCACAGGGGAGGCUUAUAGGACAUCUCUGGGGACCUUUCUG
UGAACACCCUCAACCAAUGGGGUGUAGUGGUUGAUUCUGCCUAGGUCCCCAGGGACCUGAAUUGU
CAUGACCCAAGAGAGCAUGGAUCUGGGGCGAGAGUGGCCCUUGCCAAACCCCGAGCCACUCCCCA
ACCUUGCCGGGACGGUGCAACCUUCUUGCCUCCAGCCACUCCAGGUGCAUCAGGGCUGGAGGA
AGGACAGCCCCUCCCCACCACAGGCCUCUUGACUCCUGGUGGUUGGACCUGUUGUGUUUUA
CUUUCUAAGGCUUCUGGGGGGACGGAGCUGCCCCUGGUUUUGGGAGCCAUCGGCGGGGCUAG
AACAGCCUCGCUUGGCGGGGUACUGACCUCCCCACCACCACUGGCAGCCAUCCCUCCCCUUGC
CACCCGUUCUGCCUUCAGGUGCAGCAUGUGGUGUCUCAGAACUGUGACGGGCUCCACCUGAGG
```

AGUGGGCUGCCGCGCACGGCCAUCUCCGAGCUCCACGGGAACAUGUACAUUGAAGUGAGCAGUC
CUGCAGGGACCCAGGGUCCUCCUAGGGCAGGCGGGUCCACUCACUGUGCCCUCUUGCCUCUAGG
UCUGUACCUCUGCGUUCCCCAACAGGGAGUACGUGCGGGUGUUCGAUGUGACGGAGCGCACUGC
CCUCCACAGACACCAGACAGGCCGGACCUGCCACAAGUGUGGGACCCAGCUGCGGGACACCAU
GUGCACUUUGGGGAGAGGGGGACGUUGGGGCAGCCUUUGAACUGGGAAGCGGCACCGAGGCUG
CCAGCAGAGCAGACACCAUCCUGUGUCUAGGGUCCAGCCUGAAGGUACGUGCCGAUGACACAAU
GAGUGAACCAGAGCCCCUGCCCGCCGAGGGUGUCCAGCUCUGCGGCCAGCACUGUACAGACUU
GUCCCUUGUGUGUGUGCGGUGUCUGUCUGUCUGCUUCCACAGGUUCUAAAGAAGUACCCACGCC
UCUGGUGCAUGACCAAGCCCCUAGCCGGCGGCCGAAGCUUUACAUCGUGAACCUGCAGGUAA
UCGGGUGCUGAGAGCCACGUCCUAGAUUCUGGGUCUUAGAACGCACAGCCAGAGACACCCACAC
CCCAUGCACACAGGGCGGUCUGAUAGGGCCCCCGUGGGUGCUCAGGGAGCACCGACUGAGCCCGU
AGGGGCCAAGGCUGACAGGCCACCGGGAAGGGUUGGGCUGCUGUUACUCUCACUGGCCUUUCCC
UGUCCUCAGUGGACCCCGAAGGAUGACUGGGCUGCCCUGAAGCUACAUGGGAAGUGUGAUGACG
UCAUGCGGCUCUCUAGGCCGAGCUGGGCUUGGAGAUCCCCGCCUAUAGCAGGUGAGUGAGCCG
CUGCAGCAGCCUGCUUCCCCGCACCUCUGUGUGCUGGGCCUUGUCUGUCUUCUCUGUGAGCUG
AGUGUGGAGGAAGCUCUGAGGUGUUUGCAGUGGUGCCUGAGGCAUGACUGAAGCGUGGUGGUCU
CCAGAGGGCCUGACCUCGGUGGUUGGCGGAGACCUCGUGUGGCCACUCCUGCCCUGGCUGAUG
UGGCACACACAAUCCCCGCGGGGAGAGGGAUUCUGCCCGCUGCUCUCCAGGCCUCCCCG
UGGAGCUCUCCGAGAUGCCUGGUGGGAAGCAUCUGGAGGGGACGAGCACUCGGCAGCUCUGGUC
AGACAGAAUCUGUGUGCUUGGUUUUGGGAGUUGGCGUACUUUGGGAAGCUUAAACAAACUGUG
CCUAAAUACAGAAUUGUGAUAAUUAAGACUUGGUGUAUGUAUUGAGUAAAAAGUUUACACUCU
CUUUCUCUGUGAAUUUUCAGGGUCUUAUAGGGGAAAUCAAUAACUUCUUUAAUCAAAGGGUUC
AAGAAAUAAAGGAUCCCUUACCCUUCUGGGCCUGGCACUUCUUGUAUGUUUUGUGUGUGGUGU
CUGUGAUGUGGGCUAUCGUGUACUGUAUUUUUUUUUACAUAACUAGCUCAUUUUCCUUAUC
AGUGCGUAUCUGUAUCUUAAGUUAUGAUCUGUGGUUCUGCAUCUCCGUCAGACACAUGCUUUCU
UCACGGGGUCGUCUGUAGGCCACGCCUCCCUAGUCAGCUGGGAAGGGGGAGAGGGUCUGGUCCA
CCUGCCCCACCGGUACAAGUGGAAGGUGGGGCCAGAGUUGCUAGUGACUCAUCCUGGAGACG
GAGGCAGCCCCUGGGGCCACUGCUGCCCCACCCUGUGUGGCACGCCGCUAGUGGUGGACAAGG
ACACGGAGUUUUGAGGAGACCGAGCUAGUGUGGGUGCCGACC UUUGAGUCACCACCUAAGAGGU
GACCUCUCCCAUAUCCGUUCUGCAGCUUGGUAACAAUGAAGCUGCCGCCAACCAGAGCCCCGCC
GCAGUUGACACGGGAGGGAAGGGGAUGGGAAGGCAGGGACCGCAGACAGCUUCCCCGAGCUGGG
GCAGGUGUGACUGCGAGAGGCUCCCAGGCCCGCCUGAUGCCGCUUCCCCUUUUGGCAGGUGGC
AGGAUCCCAUUUUCUACUGGCGACUCCCCUGCGUGCUGGUGAAGAAGGCAGCCACAGUCGGAA
GUCGCUUGUCAGAAGCAGAGAGGAGGCCCGCCUGGGGACCGGGGUGCACCGCUUAGCUCGGCC
CCCAUCCUAGGGGGCUGGUUUUGGCAGGGGCUGCACAAAACGCACAAAAGGAAGAAAGUGACGU
AAUCACGUGCUCGAUGAAGAACAGUUGGCACUUUGCAGAUUGGCCAGUGUCACGGUGAAGGCUGG
GUUGCCCCCACGGGUCUAGGGGAGAACGAACUCUUUGGGGAUUGACAUUUUCACCGUGACAUUUU
UAGCCAUUUGUCCUUGAGGAAGCCCCUUGCACUGCUGCGGUUGUACCCUGAUACGGCCUGGCCA
UCGAGGACACCUGCCCAUCCGGCCUCUGUGUCAAGAGGUGGCAGCCGCACCUUUCUGUGAGAAC
GGAACUCGGGUUAUUUCAGCCCCGGCCUGCAGAGUGGAAGCGCCCAGCGGCCUUUCCUCGCUCA
CCAGGCCAGUCUCAGGGCCUACCGUAUUUCUACUACUACUUAUGAAAAAGUGUGAACUUUAU
AGAAUCCUCUCUGUACUGGAUGUGCGGCAGAGGGGUGGCUCGAGCCUCGGCUCUAUGCAGACC
UUUUUAUUUCUAUUAACGUUCUGCACUGGC

AA
AA
AAA

Code:

You have the dna code with T instead of U:

with open ("p:/arm.txt", "r") as myfile:

```
arm=myfile.readlines()
```

```
def transcribe_dna_to_rna(arm):
```

```
    return arm.upper().replace("T", "U")
```

```
def mrna(exons, seq):  
    mrna1 = make_prev_mrna(seq)  
    mrna = ""  
    for x in exons:  
        mrna += mrna1[x[0]:x[1]+1]  
    new_mrna = 'G' + mrna + 100*'A'  
    return new_mrna  
  
exons= [(1, 138), (139, 276), (277, 381), (382, 452), (453, 525),  
(526, 624), (625, 861), (862, 942), (943, 1049), (1050, 1725)]  
  
final_transcript = mrna(exons, arnm)
```

f. Write the positions of the 5'UTR, 3'UTR, and CDS of the transcript

5'UTR: 1-45
3'UTR: 1249-1725
CDS: 46-1248

g. Extract the CDS of the transcript sequence and paste it. Indicate the code used.

CDS : AUGGCAGCCGGGGGUCUGAGCCGCUCCGAGCGCAAAGCGGCGGAGCGGGU
CCGGAGGUUGCGGGAGGAGCAGCAGAGGGAGCGCCUCCGCCAGGUGUCGC
GCAUCCUGAGGAAGGCGGGCGGCGGAGCGCAGCGCCGAGGAGGGCCGGCUG
CUGGCCGAGAGCGCGGACCUGGUAAACGGAGCUGCAGGGCCGGAGCCGGCG
GCGCGAGGGCCUGAAGCGGCGGCAGGAGGAGGUGUGCGACGACCCGGAGG
AGCUGCAGGGGGAAGGUCCGGGAGCUGGCCAGCGCCGUCCGGAACGCCAAA
UACUUGGUCGUCUACACAGGCGCGGGAAUCAGCACGGCAGCGUCUAUCCC
AGACUACCGGGGCCCCUAAUGGAGUGUGGACACUGCUUCAGAAAGGGAGAA
GCGUUAUGUCUGCCGACCUGAGCGAGGCCGAGCCAACCCUCACCCACAUG
AGCAUACCCCGUCUGCAUGAGCAGAAGCUGGUGCAGCAUGUGGUGUCUCA
GAACUGUGACGGGCUCCACCUGAGGAGUGGGCUGCCGCGCACGGCCAUCU
CCGAGCUCCACGGGAACAUGUACAUAUGAAGUCUGUACCUCUGCGUUCCC
AACAGGGGAGUACGUGCGGGUGUUCGAUGUGACGGAGCGCACUGCCCUCCA
CAGACACCAGACAGGCCGGACCUGCCACAAGUGUGGGACCCAGCUGCGGG
ACACCAUUGUGCACUUUGGGGAGAGGGGGACGUUGGGGCAGCCUUUGAAC
UGGGAAGCGGCGACCGAGGCUGCCAGCAGAGCAGACACCAUCCUGUGUCU
AGGGUCCAGCCUGAAGGUUCUAAAGAAGUACCCACGCCUCUGGUGCAUGA
CCAAGCCCCCUAGCCGCGGCGCGAAGCUUUACAUCGUGAACCGCAGUGG
ACCCCGAAGGAUGACUGGGCUGCCCUGAAGCUACAUGGGAAGUGUGAUGA
CGUAUGCGGCUCCUCAUGGCCGAGCUGGGCUUGGAGAUCCCCGCCUAUA
GCAGGUGGCAGGAUCCCAUUUUCUACUGGCGACUCCCCUGCGUGCUGGU
GAAGAAGGCAGCCACAGUCGGAAGUCGUGUGCAGAAGCAGAGAGGAGGC
CCCGCCUGGGGACCGGGGUGCACCGCUUAGCUCGGCCCCCAUCCUAGGGG
GCUGGUUUGGCAGGGGCGUCACAAAACGCACAAAAAGGAAGAAAGUGACG
UAA

Code:

```
print ('Do you have an RNA string?')
```

```
RNAS= 'RNA'
Resp=input().upper()
if "RNA" in Resp:
    print ('Input your string')
    RNAs=input().upper()
```

coding region (CDS)

```
def codr (seq, n):
    for i in range(0, len(seq), n):
        yield seq[i:i+n]

def method(seq, start=['AUG'], stop=['UAA','UAG','UGA']):
    response = ''
    started = False
    for e in codr(seq, 3):
        if e in start:
            started = True
            response += ''
        if e in stop:
            started=False
        if started:
            response += e
```

Response

```
for result in method(RNAs):
    l=result
print (' ')
print ('The coding part of the RNA is:')
print(result)
print (' ')
print ('the nucleotids of the coding part of the RNA is:')
f=result.split()
coding_regionslist=[e for e in f if len(e)>=int(30)]
coding_regions=(separator.join(coding_regionslist))
print(coding_regions)
```

**h. Find the corresponding amino acid sequence from the CDS.
Indicate the code used.**

AA sequence ->Protein (400aa):

MAAGGLSRSEKAAERVRLREEQQRERLRQVSRILRKAAAERSAEEGRLLAESADLVTELQGRSRRREG
LKRRQEEAASIPDYRGPNGVWTL LQGRSVSAADLSEAEP TLHMSITRLHEQKL VQHVV SQNCDGLHLR
SGLPRTAISELHGNMYIEVCTSCVPNREYVRVFDV TERTALHRHQ TGR TCHKCGTQLRDTIVHFGERTL
GQPLNWEAATEAASRADTILCLGSSLKVLKKYPRLWCMTKPPSRRPKLYIVNLQWTPKDDWAALKLHGKC
DDVMRL LMAELGLEIPAYSRWQDPIFSLATPLRAGEEGSHSRKSLCRSREEAPPGDRGAPLSSAPILGGW
FGRGCTKRTRKRVKT

Code:

```
def translate(seq):
    dict = {
        'AUA':'I', 'AUC':'I', 'AUU':'I', 'AUG':'M',
        'ACA':'T', 'ACC':'T', 'ACG':'T', 'ACU':'T',
        'AAC':'N', 'AAU':'N', 'AAA':'K', 'AAG':'K',
        'AGC':'S', 'AGU':'S', 'AGA':'R', 'AGG':'R',
```

```
'CUA': 'L', 'CUC': 'L', 'CUG': 'L', 'CUU': 'L',  
'CCA': 'P', 'CCC': 'P', 'CCG': 'P', 'CCU': 'P',  
'CAC': 'H', 'CAU': 'H', 'CAA': 'Q', 'CAG': 'Q',  
'CGA': 'R', 'CGC': 'R', 'CGG': 'R', 'CGU': 'R',  
'GUA': 'V', 'GUC': 'V', 'GUG': 'V', 'GUU': 'V',  
'GCA': 'A', 'GCC': 'A', 'GCG': 'A', 'GCU': 'A',  
'GAC': 'D', 'GAU': 'D', 'GAA': 'E', 'GAG': 'E',  
'GGA': 'G', 'GGC': 'G', 'GGG': 'G', 'GGU': 'G',  
'UCA': 'S', 'UCC': 'S', 'UCG': 'S', 'UCU': 'S',  
'UUC': 'F', 'UUU': 'F', 'UUA': 'L', 'UUG': 'L',  
'UAC': 'Y', 'UAU': 'Y', 'UGG': 'W', 'UGU': 'C',  
'UGC': 'C', '': '' }  
protein = ""  
if len(l)%3 == 0:  
  
    for i in range(0, len(seq), 3):  
        codon = seq[i:i + 3]  
        protein+= dict[codon]  
    return protein  
p = translate(l)  
  
p = translate(coding_regions)  
print ('')  
print ('The aa sequence is:')  
print (p)
```