B10 **OSNOVE BIOINFORMATIKE**
KOLOKVIJ, 1. rok, 28.1.2015

S pomočjo tehnik, ki ste jih spoznali na praktičnih vajah odgovorite na spodaj zastavljena vprašanja. Točke za posamezno vprašanje so podane v oklepajih pri vsakem vprašanju. Odgovarjajte v slovenščini!

Datoteko shranjujte na namizje.
Ko boste nalogo rešili, jo oddajte preko sistema Moodle in nato datoteko izbrišite iz diska.

Vpisna številka

Številka računalnika zapisana na listu

1.) V podatkovni zbirki NCBI Gene pridobite zapis za človeški gen za protein Cathepsin Z.

Na katerem kromosomu se nahaja (2)

20

Dolžina gena z introni in eksoni (2)

12068bp

Število eksonov tega gena (2)

6

Od katerega do katerega nukleotida v genu za Cathepsin Z se nahaja drugi ekson? (2)

Od 1453 do 1615 nukleotida

1.1) Prilepite nukleotidno zaporedje gena za protein Cathepsin Z z introni in eksoni v obliki zapisa FASTA (2):

```
>gi|568815578:c59007254-58995187 Homo sapiens chromosome 20, GRCh38 Primary Assembly
AAAGTGCGGGGTCGGCCGGGTGCTGGGCCGAGGCCGAGGCCGGGGCGGGATCCAGAGCGGGAGCCGGCGC
GGGATCTGGGACTCGGAGCGGGATCCGGAGCGGGACCCAGGAGCCGGCGCGGGGCCATGGCGAGGCGCGG
GCCAGGGTGGCGGCCGCTTCTGCTGCTCGTGCTGCTGGCGGGCGCGGCGCAGGGCGGCCTCTACTTCCGC
CGGGGACAGACCTGCTACCGGCCTCTGCGGGGGGACGGGCTGGCTCCGCTGGGGCGCAGGTGGGCACCGG
CGGGGAGGCAGCCCTGGGGGGACGGGAGGCCCATCGCCCGGGCCAGGCGCGCGCTCTGCCTCCTGGGGGC
CCTGCGTCGCCAACAGGCCTTGGAAGGTCAGATCCAGCCGCTTCGTTTTGGGGGTCGCTCCTTCTCTTCC
ACATCCAGGGTGAGCCCTGATGTTTGTGGGGTTGGACAAGCCCCACCTGGCCAGAAAGCCAGGCCAGCCC
CACCTCCTCGGCCTCTGGGGGTGGCCCTTACGGCGCGGCATGGCTGGGGGAGCAACCCGGTTCCCTCCTT
TCGGGGGTGCCCTGCCTCCTTCACATTCTGAGAGCGCTGGAGCCTCCTAGTGTGAGGTGGGCCCACACCT
GGCTCACCTCTGGTGATTGTGGCCCCACTCCCATGCCCTCAGGTGTTGGGAAAGGCGGGAGGGCCCTACC
GAGGGCTTCCTGGGCCCACGGCCCCCGGGAGCCCCAGCACCGATCTGGGTGGGGATGATCTCTTCTCAGC
ACATACCCCCGGCCTCATGAGTACCTGTCCCCAGCGGATCTGCCCAAGAGCTGGGACTGGCGCAATGTGG
ATGGTGTCAACTATGCCAGCATCACCCGGAACCAGCACATCCCCCAATACTGCGGCTCCTGCTGGGCCCA
CGCCAGCACCAGCGCTATGGCGGGTGAGTGGCCGCCCTTTGACTGTGGGCCAGGAAAGCCCATCCCGGCT
GCCTGTTTATAGGCCCAGGCAGCTCAGCCCGCGGTTCCAGGGGCCTGTGATTCAAGGCCAGGTCAGCCTG
GGCTGGGGTGCTAAGGCCGAGGAGCGCGGGCCTTTGCCCTCTGCTGAGGTGACCCAAGCTGGGTCATTCT
GGGGCTTTTGGGTGACAGGCAGGACTAGGACTGAATCCCAGTGACAAGTGGCTGCCAGGTGCCCTCTCTC
CCCCAAAGCTTGTGACCTCAGCCATTGGGATAGAAACAGAAATGCAAAATCACTGAGTCCCCAGGCTGTA
TTTGAGGATACAGAAGTAAACTGCTTCTCAGCACAGATGAAAGCCTGTCCTGGGCTTGGAATATTGGGTG
AGGCTAGTTGGCTACAGCCCATGGCCTAACCTGGTCACTACCTCCAGCCTGGTGGGTTTCTATCTTAGGC
TTTTGGCAGGAGTAGGGTAGGGAGGCTCTTAGCAAAGAGGAAGGGGAAACTAAGGCCCTCCCTATCAACG
AGGCTGCTGCACCCTGCTTTGGGAGAGAAGGCGGGAGAAGGGACAGCTGCAAACCAGCCCCAGGCCCACC
CCACAGCATTTGGAAGGGACTCTGCTATCCCCTCTCCAGTCACAGCACAGGGTCCAGGAGCCTGGATTTC
```

CCCTGGAGATGGGACTTGCGTGGCCTGGGAGTTTGGCCCATGAGCACTGTGGTGCTGTTGCCTCTTCCCT
GGCCTTGGTCCTCATGGGTCAGGCCTGCTGTACACCCATGTGAAAAGGATAGACTACAGTGTGGCTGCCG
CTGTGGGCCAAGGCTGGGGGACATGCGTTTGCGTGTGCTCCATCTCACTGACGCCTCCTCAATGCCATGA
TTGCTTCCAGTTTGCAGATGAGGAAGCAGGCTTTTTGCAGAGAGATGGGGAGGCTTGAGCATAGTGTGAG
TGACAGCTGAGATTTGAGCCCGGAAGCCTGGGCTTTGTGCCAGGATGTCTGATGGGCTCAGTCTTGTGAA
TGAGGGGTTCTTGCTCTGGGGGTAGGGTCTGGACCCCAGAGCCAGCCAGCTGAAGCTGTGGCATCCGTGG
CGTCTCCCCCGGGGGCTGGAACACACTTCTGATCGTTCAAAGGTCCCTTTGATGTGGATGGAGCTGGAAG
GTCTGGGAGACCTTGGGCGGGCTGCAGGGTGTGGGTGGGGCTAGAGGCAGAGGACTGGGCAGAGTCTAGG
GGAAGTGTCCAGGCCTGGCACCCGCTGAGGAAGTGAAAGAGCCTTTGAGGGTGGGGTGAGGGTGAGGGGA
GAATAGAAAGGTCAACAGGTGCCTGTGGGAAGCCTTGAACTCTTTGAGAGGGAGTCAGGTGACAAGATGT
GCGTCTGTGAAGGACTGCGGGTAGAAAGGAGATGGCTTTCAGATGGATGTGGGACCCTGGCCAGGGGTCA
GCCCTGCCGCATTTCCCCAGGGGACTCTCTCTGGGCGGATGACTGCTTCACCCCACTTTCTCCTGAAACC
TCCTTCTCAGCGCACCCAGGCCCAGGCACTCGCTTCCTCTTTCACTGAGGAATAGAAGCCACCAGAAGGG
GACTTCCACGGGGCCCTGGCCATTTGCCAGCCAAGGCAGCAGGTCCTCCATTTGTCTTCCCTCCTCTCCC
CCAGGATCAACTGTCTTGTCCCCTCACTGAGCCCGTCCCTTCCCACTCAGGGAAGCTCTCCCACCTGGAA
AAAAGGAACACAAAGCCCCCGGCCCCACCCCGCCCCAATTCTGGGCTCCCTAGGGAAGTATCTGGAGTCC
CTCCACTTCCTCTTCTCTACCCTTGAACCCACTCCAGCCTCAGTCCCACCTCTGCCACTCCACCGGCACT
GTTCTTTCCACTGCCACCTGGGGCCCACCGTGAAGCATTCTCCTGATCACGTCTCCCCAGGATGCTTCCT
TCTCCTGCACCATACTGCCCGCCTCCACCTCCCTGGCCACTCCTTCCCCTCTCCCCAGCTGTCTGCTCCT
CAGAGTTCTGAGTGTCAGCTTCCCCAGGACCAACCCTTCACCTGAACTCAGCTTGACATCCACCTGCCCA
CTCGCCCGCTGTCCCGGTCACCTGACACCTCCAGCCTCAGACCTTCAGGTCTGGATGGCTCCTTCCACTA
CTGTCACCTGACAGCCCATTTCTTCCCCAGTCCTCCCTGGCTGTCCTTCCGGCTATGCAGGCAAAAGCCT
GGACTGGTACCCGACTCCTCTGTTTTCTGTCCCTCATATAATCCACCCACAAGTCTTGTCGGTTCTACCT
TCGAAGCAGATGCAGAAATCACGCACCGCCCACCGCCTCCTCCAGGTCTAAGCCACACCGCCCCACTGCC
GACTGCCCTCACTGGTCTCCTGGCCTCTGCCCCTGAGGCCACAGTGAGCCTGTGACAACCCATCAGAACA
TGTTCTTTGGCTTGAAACCCTCTCAGGATTCCACTCAGATAAAAGACAGAACCCTCTGGCCTTCAAGAGC
CTTCACCATCGGGCTCCCTCCCAGTCCGCCCTTGCTCCCCTGCTCCTGCGGAGAGTGGAGCAGGGAGGTA
GCTTCTCCCACCAGGCACCTCCTCTCTCCCTGAGGTCCCATGGCTCCCTCCCTCCCTCCTATGGATGTCA
CCTCAGCGGGGTCTTCCTCAGCCACCCTATGCATATGACACCCCACTTCCTGAATTATCTCCCGGATTTA
TTTTTCTTCATGGTGCTGTAGACTTTGCACAGTTACAGATGCTCCTTGACTTAACTGTGGGGTCATGTCC
CGACAACCCCCTCGTAAGTTGAAAACACCATGAGTGGAAAATGTGTTTGACACACCCAGCCTCCTGCACA
CCCTGGCCTGGCTTTGTCTGCATTAAACGTATTCAGGACACTTGTGTTAGCCCACAGTGTGAGTAATAAG
CCCGTTCCCAGATGGGGCTGGTGGGATGAAGGATTGAGTTTGGCCGTGGGAGAAGCTTGTGCATACCTCA
GTGAGCTGGTTAGGCCGGGTCCAGATGCTTCAGGCAGGAGTAGACGGAGAGGCCAGAACCGAAACTGCTC
ATCCATCCGAGCTCTGCCTCTGGGCCAGGACTGCTGTCGCCGGGTTGGGTTGGGCATAGAAGGATAGGGC
TCGGGGGCAGGCCATGGGACGCCCTGCACTGCCAAGTGGAACTCCAGAAGCCCAGGCAGGAGGTGCTTCC
AGAAGGATGGAGCGACCGCTGTGCTGACTGCTGCCGGAGAGTGGTCAGCTCGGATGACAGTGGAGAAGTG
ACCTTTAGCAAGAGAGAAAGGGCTGAAGAGGAGCGGGGCAGTAGGGGTCGGGGCACTGTCTCAAACCATT
GAGTGAAGCAGAAAACCGTGATAGTGAGTGGGAGGGGCCGGGGGCTTCCTCTGTTGTTGTGAGAAAGGCA
AGTTTGGGGTGTGCAGATCTGCTTAGAGGGGGCATAGCTGACAAGGGAGCGGGGAGAAGGAAAGCTGGAG
GGAGCCAAGTCCTTGGGAGGAAACAGGATGGAACCTAGGGTGGGTCGGTCAGGTGGGAGGGGACAAACCA
GCCACCGGCTGCTCCAGGAAGAGGGTTTGGGGTGCAGAGCATGGGCAGGCTGGGTGGCGGCAGGATGGAG
GTCTGCTTTGGGACTTTTCAGTGATGCGTGAAGTCACCAGAGCAGGGACAGTTCCAGAAGGCCCCTGGCT
GGGCGACCGGGAGGAGGCCAGGTGGAGGAGCGGACTGTGGGGCAGTGTGGGAGAGTCTGGAAGGTGACAA
GAGATGGAGGGGATGTGCAGGTCCAGAGGGGGCTGCCAATGGGCAGAGGGACTCCCTGCTCATTGTGCCC
TGGGGGAGGGACAGGTGAGGCTGCAGGAAGAGTTCCGACTTGAACGTGGGGTCAGGAGATGAGGCCTGAG
GCAGGTGCAGGGGTGTGTGGCTGTGAAATCATTGTCCTGGGAGGGCAGAGTGAGATGACCCTGAACATGG
GCCCGGATTTCCTTCTAGGATCTGTTTTATTGCATTTGTGCATGGGAAGGGAAGCAGGTCCCTCGGTGAC
CACCCTGAGCTGTCCCTGGCTGCACCTTGGGATTTAGAGCCCAGCCACCTGGGCAGTGAGGGCATGGTCA
GCCAGCCGCTGGGAGCCCTCTGCATCCCCCAGTGCCTTCGGGGCTGGGTCCCTGCTGGCCTTCCCCAGCC
CTCAGGACCGCATCTGGCCCCTCACTGGGGCTCATGCGGCTCCCTGGGCTATACTCGTTCACAGAGCTGA
CACCTTGGCCCCCTGTCCACGAGGGCAGGGACGGTCTCTCCTGGCTGGCTTTGGGGTGTGGCCCACTTGC
CATGGGTGGCTGAGTCTCTGTCCATTCAGCCCTTCATTCAGCAGGCACTTCTCAGGTGTCTGCAGGGCCA
GATGCTGGTGACCGCCTGGCCCTGGCCCTGCTCTCATGGCATTTGTAGCGGGAGGTGTGGCAGACAGGAC
AGTGCCCAAAGAAGGCCAGGAGAGGCCTCTCTGCAGAGCGACGTCTGAGCGGAGACCTCCATGAAGAGGT
CAGCAGGGAGAAGGAGACACAGGCAGCTGGGCAGAGCTGAAGGCAGGCCAGGCTGGGCAGCGCCTGCATC
CCGCCTGAGGCGTCCAGGTCCGTTCGGGGAAGTGGGTGGAGAGTGGGTGAGTGCTGACTGGCAGGTGCTG
ACTGTTGCAGATCGGATCAACATCAAGAGGAAGGGAGCGTGGCCCTCCACCCTCCTGTCCGTGCAGAACG
TCATCGACTGCGGTAACGCTGGCTCCTGTGAAGGGGGTAATGACCTGTCCGTGTGGGACTACGCCCACCA
GCACGGCATCCCTGACGAGACCTGCAACAACTACCAGGCCAAGGACCAGGGTAGGCTGCTGCCCGTGCCC

CCACTCCACCCTCCTGCCCTCCACTCCCTCTTCCACAGGCCTTCGTGGCTGAGGACCCAGCCTGACCCCG
TGCTCCTCACATTGGCTGGTGGCTGGGGAGCTGCGGGAGGGCAGAGGTTGCAGCCCCCAGGGGCGCTGGT
GCTCTTGACCATGAGGGGTAGTGTTCCCAGTGCCAGCTCCAGCCTCTCGCTGCAGAAGCCCCCTCTGCTA
TCTTGCACCCGCTTGGGCTGGGGATCAGAAAACTGGACCCCGTAGTCCCTGTTGAGAAGCCCAACGAAAA
AGCCTGCTGCAGGAGCGGTCTGAAGAGCATCCTTGGCCATGCGCGCGGTGGCTCATGCCTGTAATCTCAG
CACTTTGGGAGGCCGAGACAGGCAGATCACTGGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACACGGTG
AAATCCCGTCCCTACCAAAAGCACAAAAATTAGCTGGGTGTGGTGGTGCACACCTGTAATCCCAGGTACT
CGGGCAGCTGAGGCAGGAGCATTGCTTGACCCGGGAGGCAGAGGTTGCAGTGAGCCAAGATCGCGCCACT
GCACTCTAGCCTGGGTGACAGAGCAAGACTACGGCTCAAAAAAAAAAAAAAAAAAAATACCGGCCCTGTCC
CCATGTCTCCAGCCTCGTGGAAGAGTGTCTTGCTCCTGTGCCCCTTTGGCTGGGCCTGAATGTGGTGTTT
GAGTTCTGACATGCTGGCATTTTGGTGGCTTTTTGGGCCTTTGTAGGGGAGAGGTGTAGGGGGCCTTGTG
GAGGGAGAGTTCATTTTCCAGGAGCTTTGGTGGAGAGCAGGTGGGGAGGTGCTGCCCCAAACTGGAAGGA
GCCATTGGCTTGGTTCCTGTTCTGGTGACCCAGGGCCCAGGTCCCCCACATTCAGGTGCGGCTACGGCCA
AGGGGGTGCTGTGTGGATGAGCAGACCCAAGCTAGCCTGGAATCCCTAGAGCTGGCACATTCCTTTCAGG
CTGCTCCCATTAGAAACAGAAGGAAAGAAAAGTCCAGACCCTAATCCCATAATTGGTAGGTTTTGTTTGT
TTGTTTGTTTGTTTTTGAGAGGGAGTTTTGCTCTTGTTGCCCAGGCTGGAGTGCAATGGCACGATCTCTG
CTCACTGCAACCTCCGCCTCCCGGGTTTGAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTAC
AGGCATGCGCCACCACGCCCAGCTACTTTTGTATTTTTTAGTAGAGACGGGGTTTCTCCATGTTGTTCAG
GCTGGTCTTAAACTCCCAACCTCAAGTGATCCACCTGCCTTGGCCTCCCAAAATGGTGGGATTACAGGCA
TGAGCTGCTGCGCCCGGCCATAACTGGTAGTTTGTTTTTTTTTTTTGAGACGGAGTCTCGCTCTTTCGCC
CAGGCAGGACTGCAGTGGCGCGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCACTCTCCT
GCCTCAGCCTCCGGAGTAGCTGGGACTACAGGCGCCCGCCACCGCGCCCGGCTAATTTTTTGTATTTTTA
GTAGAGGCGGGATTTCACTGTGTTAACCAGGATGGTCTTGATCTCCTGACCTCGTGATATGCCTGCCTCG
GCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGTGCCCGGCCATAACTGGTAGTTTTAAGCCACC
TGACTCAGATCTCCCTCACAGCCAGAGCAGAGAGCTGGGAAAGCCCAGCAGCTGCTTGCTTTGCTGCCTC
CACCCCTGCCCCAGGCTATATAGGCTGGCCCCAGCGCCCTGTGCACTCTGAAGTGGCAGAAGGTCCTCTG
GGCACCACCCTCTGCCAACGGGGCTCCCTTGCCTTGCCTCTCAGGCCTTATTGTTCAGCCAGAAATTCTC
CTGGACTGGCTGAGTGAGTCCTTCGAGAAGGGGGACAGATGGAAGGTGCAGAGGGAAGGGCTATCAGATG
GCTCTGGAAGCAGGAAGCAGGTATCAGGTGGTCTCCAGGGCCACGCCCACTCCACGCCAAAGTGCTCTCG
CCCAGCTGAGGCCAGGCCCCTCCAACCCACCCAGCCTGACAGGAGCAGATGGAGCTCATCAGAAACCTGA
TTGTGTGACTGGTGACCGCGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCGGAT
CACTTGAGACCAGGAGTTCGAGACCAGACTGGCCAACATGGTAAAACCCTGTCCCTACTAAAAATACAAA
AAATAGCCGAGCGCGGTGGCATGCGCCTGTAATCCCAGCTACTCAGGAGGCTAAGGCAGGAGAATCACTT
GAACCCAGGAGGCAGAAGTTGCAGTGAGCCAGGATCATACCACTGAACTCCAGCCTGGGTGACAGAGTGA
GACTCTGTCTCCAAAAAAAAAAGAAAAGAAAAGAATAAAAAAACCTGATTGCGATAAATAGTCATCTTTG
TAGTGTTAGCAAAAGTGTGGCCACACATCCACCATTTTCCCGTAGTACTGGGCACTGGGCACTTGGTGTG
CAAGGAACAAATTGAACCAGTGTCTGAGCCTCAAGGTGCTCATGTTCCTGGGGGAGGAATCCCGACAGAT
GTGTAACTATGACAGACAGGTCGCGCCAAAGCAGCGAATGGCAGGCCGAGGGCCGCCGTGTGGATGGAGT
GAAAAGTGGGAACACAGTGACTCTGAGCCTTGCAAAACCGAGAGCTGTACTTTTTCTTTTTTTTAAATTA
ATTTATAAAATTTTTTTTCTGTAGAGATGAGGTCTCGCTCTGTTGTCCAGGCTGGTCTCAGACTCCTGGT
ATCAAGCCATCTTCCTGCCTCGGCCTCCCGAAGTGCTGGGATTACAGGTGTGAGCCACCACGCCTGTCCT
TTCTTTCTTTCTTTCTTTCTTTTTTTTTGAGACGGAGTCTCGCTCTTGTCGCTCAGGCTGGAGTGCAATG
GCGCGATCCTGGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCACTGA
GTAGCTGGGATTACAGGCATCCACCACCACACCCAGCTAATTTTTGTACTTTATTAGTAGAGACAGGGGT
TTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAG
AGCTGGGATTACAGGCGCCTGGCCCTCTTTATTTTATTTTTTAAGCCTAATTCTTGATAACCTCTTAAGT
CGAAATAACCAGAATTTTTTTTTTTGGTCCCCTCAGCGTGGGCACAGAAGAAAATGCTTTTTTGTGGTTG
TTGTTACCCAAAGTATTGGCCTATTTATTCACCAGCCAGCAATAGAGACATTTTATTAAGTGCTAATCAT
TTGATTATGCTCCATTTTCAGCTAGAAAGTATTGTGGAAAGGTCCTTCTCCAGCCATTAGAGGAAAAAGG
CGAACTCCTGGGTGCCTACTATGTGCTAATTAATGTCTGTCTCTTTTCTTCTCTGCCAAGGCCTGGGGAA
GGGCTTTTCATAACTCCCTGGCAGCTGAAAGAAGAGCGTCATTAAGCAGCACCTAATAGCAAGATGATTT
CCGTTCTTAGCTGCTGAGTGGTAGGCTGCATGATGAGAGTGGAGAGTGGGCTTGGCGGGCTTCTTTGTCA
GTGGGTTGATGAGGAGACGGCCTCATGCTGACTTTTCTTGCTCTCCCCCAGAGTGTGACAAGTTTAACCA
ATGTGGGACATGCAATGAATTCAAAGAGTGCCACGCCATCCGGAACTACACCCTCTGGAGGGTGGGAGAC
TACGGCTCCCTCTCTGGGAGGGAGAAGATGATGGCAGAAATCTATGCAAATGGTCCCATCAGGTGAGGAG
AGGTCCCGCGGGCAAAGAGAGGTTTGCGGGTGAGGAAAGGTCCCGCGGGTGAGGAAAGGTCCCGCGGGTG
AGGAGAGATCCGCGGTGACAGCGCCAGTGTGGGCTCATGTCGGTGCTCATGTAGGCATCTCCGTCCTTCG
CCACTTTCCTTCTGGCCGGTTGTACTCAGTAGGCTCGAGACAGCCACATCATTTCCCCTCTAAGAACTTC
AGTTTTCTCAGAGTTAACACTCAGCTGATAGCAAGTCTGCCCCTATGGATGGTTGTGCGGAATGCGGATG
AGACGCTCCAAGCTGTACCGAGGGGGCAGTTTCAGCAGGGAAGACACTATGTGAATAACAACAGAAGTTT

ATCCAGAACCGTCAGATATACACTACCTGAGTCGGGTTTACACTTTGTTGTCATACCAGATTTTTTTTTT
TTTTTTTTTTTTTGAAACACAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTATGATCTTCGCTCA
CTGCAGCCTTGACCGCCCGGGCTCAAGCTATCCTCCCGCCTCAGCCCCCTCCACCGAGTAGCTGGGACTA
CTGGCGCACACCACCACACCCAGCTAATTTTTGTATTTTCTGTAGAGGTGGGGTTTCGCCATCTTGCCCA
GGCTGGTCTTGAATTCCTGGACTCAAGCGATCCACCTGCCTCGGCCTCCCAAAGTGCTGGAATTACAGGC
GTGAGCCACGGCGCCCAGCCATATCAGATTATTTCTAATGACGGTAAATTCTCTCAATTTAAAAGTGGTA
ATTCTTTAACTGCTTGCTCTCAGCTGTGGAATAATGGCAACAGAAAGACTGGCTAACTACACCGGAGGCA
TCTATGCCGAATACCAGGACACCACATATATAAACCATGTCGTTTCTGTGGCTGGGTGGGGCATCAGTGA
TGGGACTGAGTACTGGATTGTCCGGAATTCATGGGGTGAACCATGGGTAAGATGTTTTCATTTTCTTAAG
GTCATTCCTAGAAAAAAACGTCCTGGTACTCCTCTCGCTTGAATGGTTCTGTTTACGTTTAGCTAAAAAC
TGGGGCCTTGACAGAGGGTTCCTTGCTCCAGCTGTCTCAGCCACCGCGCGAGTGGCTTTTGGGGCCAGAC
CCTTCTCTGACATGCGCGGCTGCCGCCATCCTGTGCACTTTGGGTGTTCAGCAGCATCCCTGGCCCCTCC
ACCCACTAGGTGACTCCCCGCAATCCCAGTGTGACCAAAACCGTCTCCAGATGGTGCCTGGTGCCCTGGG
GAAGTCCCTGTGAGCACTGCTGGAGAGGGGCTCTGCCACCCTCTTCCTTTACAACCAGCTCATTCTTTTT
CCCATCATGGTTTTTGGGCTCCCCTAAAATACTGTGTCGGTTTTTCAAAAATACTCGAGTTGGGCCAAAA
ATCCAGGAGAGAGAAATTAACTGCCAGACCTAGTCTTAGCCATTAATAATTTAAGACACAGAAATTCATG
ATTAAGAAATGGGGGGAAAGCGGTTCATGTGCAGGTGATGAAATGTCGTCATTTTTTACTAGGTGACTAA
TGGGGGTAGTAGTAGCTGGAGGAGGCACCAGTCTAGAGTCAACATGTTTTGTGAGATGAATTCACTTCTG
GCATAGAGGGCCCTCATGTTAAACATTAACCAAGTGAAATGTTTCACAAAAGGTATTTCAGAACAGGGGT
CATGGGAAAGCTGAGGGGCAGCTGAGTGGGCTGTCTGGATCCCCCAAGGCTGGCCTGAGAGATGGAGGCA
GAAAGCAGGGGGCAGGGGGCTGACGGCAGCAAGGGGAGCGGGAGACGGCTGCAGATGGGCTGACGCGCGA
TCCCACTTCTCACAGGGCGAGAGAGGCTGGCTGAGGATCGTGACCAGCACCTATAAGGATGGGAAGGGCG
CCAGATACAACCTTGCCATCGAGGAGCACTGTACATTTGGGGACCCCATCGTTTAAGGCCATGTCACTAG
AAGCGCAGTTTAAGAAAAGGCATGGTGACCCATGACCAGAGGGGATCCTATGGTTATGTGTGCCAGGCTG
GCTGGCAGGAACTGGGGTGGCTATCAATATTGGATGGCGAGGACAGCGTGGCACTGGCTGCGAGTGTTCC
TGAGAGTTGAAAGTGGGATGACTTATGACACTTGCACAGCATGGCTCTGCCTCACAATGATGCAGTCAGC
CACCTGGTGAAGAAGTGACCTGCGACACAGGAAACGATGGGACCTCAGTCTTCTTCAGCAGAGGACTTGA
TATTTTGTATTTGGCAACTGTGGGCAATAATATGGCATTTAAGAGGTGAAAGAGTTCAGACTTATCACCA
TTCTTATGTCACTTTAGAATCAAGGGTGGGGGAGGGAGGGAGGGAGTTGGCAGTTTCAAATCGCCCAAGT
GATGAATAAAGTATCTGGCTCTGCACGA

2.) Zrela oblika mRNA za Cathepsin Z ima dostopno kodo AF032906:

V kolikšnem odstotku se elementi gena z introni in eksoni (Točka 1.1) ujemajo z elementi mRNA zapisa gena (AF032906) (2)

Zaporedji se ujemata v 92% (query cover).

3.) Ali glede na nukleotidno zaporedje AF032906 obstajajo nukleotidna zaporedja homologov proteina Cathepsin Z v organizmih Pan troglodytes (taxid:9598) ali Oryctolagus cuniculus (taxid:9986)? Utemeljite vaš odgovor! (6)
Da, homologa obstajata. Pri obeh organizmih je E vrednost 0, stopnja identičnosti pa nad 70%, kar je meja za homologijo.

4.) V enem/dveh stavkih navedite vlogo proteina Cathepsin Z pri človeku in dostopno kodo aminokislinskega zaporedja. Ali je zaporedje dobro anotirano, pojasnite zakaj tako menite! (4)
Ima karboksi-monopeptidazno in tudi karboksi-dipeptidazno aktivnost. Ima katalitično aktivnost, šibko endopeptidazno aktivnost. Zaporedje je dobro anotirano, saj je pregledano s strani strokovnjakov in vnešeno v zbirko swissprot.

5.) Ali obstaja eksperimentalno določena 3D struktura mišjega (Mus musculus) Cathepsin Z? (6)
3D struktura ne obstaja, obstajajo samo strukture mišjega proteina Cathepsin S.

V kolikor obstaja navedite dostopno kodo in s katero metodo je bila struktura izdelana:

V kolikor struktura ni izdelana navedite ali bi lahko na podlagi homolognega modeliranja izdelali model 3D strukture proteina Cathepsin Z iz miši? Utemeljite vaš odgovor!

Da, model bi lahko izdelali in je tudi že narejen in vnešen v zbirko swissmodel depository. Lahko bi ga izdelali tudi s pomočjo homologije s človeškim proteinom (E=1e-164, I=85%), katerega struktura je že eksprimentalno določena.

6.) Aminokislinsko zaporedje Cathepsin Z človeka v obliki FASTA priložite k spodaj navedenim zaporedjem in izvedite poravnavo več zaporedij. (2)

```
>Protein1
MARRGPGWRPLLLLVLLAGAAQGGLYFRRGQTCYRPLRGDAPLGRSTYPRPHEYLSPADLPKSWDWRNVDGVNYASITRNQHIPQYCGS
CWAHASTSAMADRINIKRKGAWPSTLLSVQNVIDCGNAGSCEGGNDLSVWDYAHQHGIPDETCNNYQAKDQECDKFNQCGTCNEFKECH
AIRNYTLWRVGDYGSLSGREKMMAEIYANGPISCGIMATERLANYTGGIYAEYQDTTYINHVVSVAGWGISDGTEYWIVRNSWGEPWGE
RGWLRIVTSTYKDGKGARYNLAIEEHCTFGD
>Protein2
MARRGPGWRPLLLLVLLAGAAQGGLYFRRGQTCYRPLRGDAPLGRSTYPRPHEYLSPADLPKSWDWRNVDGVNYASITRNQHIPQYCGS
CWAHASTSAMADRINIKRKGAWPSTLLSVQNVIDCGNAGSCEGGNDLSVWDYAHQHGIPDETCNNYQAKDQECDKFNQCGTCNEFKECH
AIRNYTLWRVGDYGSLSGREKMMAEIYANGPISCGIMATERLANYTGGIYAEYQDTTYINHVVSVAGWGISDGTEYWIVRNSWGEPWGE
RGWLRIVTSTYKDGKGARYNLAIEEHCTFGD
>Protein3
MARRGPGWRPLLLLVLLAGAAQGGLYFRRGQTCYRPLRGDAPLGRSTYPRPHEYLSPADLPKSWDWRNVDGVNY
ASITRNQHIPQYCGSCWAHASTSAMADRINIKRKGAWPSTLLSVQNVIDCGNAGSCEGGNDLSVWDYAHQHGIP
DETCNNYQAKDQECDKFNQCGTCNEFKECHAIRNYTLWRVGDYGSLSGREKMMAEIYANGPISCGIMATERLAN
YTGGIYAEYQDTTYINHVVSVAGWGISDGTEYWIVRNSWGEPWGERGWLRIVTSTYKDGKGARYNLAIEEHCTF
GD
>CathpesinZ
MARRGPGWRPLLLLVLLAGAAQGGLYFRRGQTCYRPLRGDGLAPLGRTTYPRPHEYLSPA
DLPKSWDWRNVDGVNYASITRNQHIPQYCGSCWAHASTSAMADRINIKRKGAWPSTLLSV
QNVIDCGNAGSCEGGNDLSVWDYAHQHGIPDETCNNYQAKDQECDKFNQCGTCNEFKECH
AIRNYTLWRVGDYGSLSGREKMMAEIYANGPISCGIMATERLANYTGGIYAEYQDTTYIN
HVVSVAGWGISDGTEYWIVRNSWGEPWGERGWLRIVTSTYKDGKGARYNLAIEEHCTFGD
PIV
```

Poravnava:

```
CLUSTAL 2.1 multiple sequence alignment


Protein1        MARRGPGWRPLLLLVLLAGAAQGGLYFRRGQTCYRPLRGD--APLGRSTYPRPHEYLSPA 58
Protein2        MARRGPGWRPLLLLVLLAGAAQGGLYFRRGQTCYRPLRGD--APLGRSTYPRPHEYLSPA 58
Protein3        MARRGPGWRPLLLLVLLAGAAQGGLYFRRGQTCYRPLRGD--APLGRSTYPRPHEYLSPA 58
CathpesinZ      MARRGPGWRPLLLLVLLAGAAQGGLYFRRGQTCYRPLRGDGLAPLGRTTYPRPHEYLSPA 60
                ************************************  *****:************

Protein1        DLPKSWDWRNVDGVNYASITRNQHIPQYCGSCWAHASTSAMADRINIKRKGAWPSTLLSV 118
Protein2        DLPKSWDWRNVDGVNYASITRNQHIPQYCGSCWAHASTSAMADRINIKRKGAWPSTLLSV 118
Protein3        DLPKSWDWRNVDGVNYASITRNQHIPQYCGSCWAHASTSAMADRINIKRKGAWPSTLLSV 118
CathpesinZ      DLPKSWDWRNVDGVNYASITRNQHIPQYCGSCWAHASTSAMADRINIKRKGAWPSTLLSV 120
                ************************************************************
```

```
Protein1        QNVIDCGNAGSCEGGNDLSVWDYAHQHGIPDETCNNYQAKDQECDKFNQCGTCNEFKECH 178
Protein2        QNVIDCGNAGSCEGGNDLSVWDYAHQHGIPDETCNNYQAKDQECDKFNQCGTCNEFKECH 178
Protein3        QNVIDCGNAGSCEGGNDLSVWDYAHQHGIPDETCNNYQAKDQECDKFNQCGTCNEFKECH 178
CathpesinZ      QNVIDCGNAGSCEGGNDLSVWDYAHQHGIPDETCNNYQAKDQECDKFNQCGTCNEFKECH 180
                ************************************************************

Protein1        AIRNYTLWRVGDYGSLSGREKMMAEIYANGPISCGIMATERLANYTGGIYAEYQDTTYIN 238
Protein2        AIRNYTLWRVGDYGSLSGREKMMAEIYANGPISCGIMATERLANYTGGIYAEYQDTTYIN 238
Protein3        AIRNYTLWRVGDYGSLSGREKMMAEIYANGPISCGIMATERLANYTGGIYAEYQDTTYIN 238
CathpesinZ      AIRNYTLWRVGDYGSLSGREKMMAEIYANGPISCGIMATERLANYTGGIYAEYQDTTYIN 240
                ************************************************************

Protein1        HVVSVAGWGISDGTEYWIVRNSWGEPWGERGWLRIVTSTYKDGKGARYNLAIEEHCTFGD 298
Protein2        HVVSVAGWGISDGTEYWIVRNSWGEPWGERGWLRIVTSTYKDGKGARYNLAIEEHCTFGD 298
Protein3        HVVSVAGWGISDGTEYWIVRNSWGEPWGERGWLRIVTSTYKDGKGARYNLAIEEHCTFGD 298
CathpesinZ      HVVSVAGWGISDGTEYWIVRNSWGEPWGERGWLRIVTSTYKDGKGARYNLAIEEHCTFGD 300
                ************************************************************

Protein1        ---
Protein2        ---
Protein3        ---
CathpesinZ      PIV 303
```

Preglejte poravnavo zaporedij Ephrin-A1 človeka in proteinov 1, 2 in 3 ter za vzstavitev vrzeli upoštevali težo -10, za podaljševanje vrzeli pa -1. Koliko je vrednost za težo vrzeli, če primerjate poravnavo Cathepsin Z s kateremkoli od drugih treh aminokislinskem zaporedju proteinov? (4)

Zaradi vrzeli in njihovega podaljševanja je vrednost pri primerjavi -23

7.) Zakaj uporabljamo vrednost rezultata poravnave (S) dveh zaporedij? Kako je vrednost rezultata poravnave povezana z vrednostjo e? (4)

Vrednost rezultata poravnave nam pove kako kakovostna je poravnava, je seštevek vseh vrednosti, ki jih pridobimo s pomočjo izbrane matrike in kazni za vstavitev in podaljševanje vrzeli. E vrednost pa nam pove kako relevanten je naš podatek, pove nam koliko je verjetnost, da nastane zaporedje z enako ali večjo S vrednostjo po naključju. Nižja kot je E vrednost in višji kot je S, bolj kakovostna je poravnava.

8.) Zakaj bi uporabili Blastx? (4)

Blastx uporabljamo ko imamo znano nukleotidno zaporedje in ga želimo primerjati z zbirko aminokislinskih zaporedji, torej ko nas zanima, kakšen protein bi lahko zapisoval naš neznan gen. Program nukleotidno zaporedje prevede v aminokislinsko in išče po bazi aminokislinskih zaporedij.

9.) V kolikor z nukleotidnim zaporedjem ProteinaX preiščete podatkovno zbirko nukleotidnih zaporedij ugotovite, da ne dobite nobenega zadetka. Zakaj menite, da je temu tako? (4)

Zaradi napak pri določanju zaporedja, lahko da protein še ni vnešen v zbirko.

10.) Navedite katero podatkovno zbirko bi uporabili, da bi pridobili najnovejše članke o strokovni tematiki iz področja biologije, ki vas zanima. Navedite katere podatke hrani podatkovna zbirka, ki bi jo uporabili. (4)

Za pridobitev podatkov o npr. določenem genu, ki vpliva na razvoj neke bolezni bi uporavila zbirko PubMed ali OMIM. Zbirka PubMed vsebuje podatke naravoslovne in medicinske povzetke in citate. OMIM pa je zbirka človeških genov in fenotipov v katerih se izražajo.

11.) ZaporedjeY je bilo pridobljeno v laboratoriju, s sklopitvijo dveh genov različnih organizmov. Prvih nekaj več kot 600 nukleotidov je sklopljenih preko 24 nukelotidov z genom drugega organizma, dolgim nekaj več kot 800 nukleotidov.

.

>zaporedjeY

ATGGAGTTCCTCTGGGCCCCTCTCTTGGGTCTGTGCTGCAGTCTGGCCGCTGCTGATCGCCACACCGTCTTCTGGA
ACAGTTCAAATCCCAAGTTCCGGAATGAGGACTACACCATACATGTGCAGCTGAATGACTACGTGGACATCATCTGT
CCGCACTATGAAGATCACTCTGTGGCAGACGCTGCCATGGAGCAGTACATACTGTACCTGGTGGAGCATGAGGAGT
ACCAGCTGTGCCAGCCCCAGTCCAAGGACCAAGTCCGCTGGCAGTGCAACCGGCCCAGTGCCAAGCATGGCCCGG
AGAAGCTGTCTGAGAAGTTCCAGCGCTTCACACCTTTCACCCTGGGCAAGGAGTTCAAAGAAGGACACAGCTACTA
CTACATCTCCAAACCCATCCACCAGCATGAAGACCGCTGCTTGAGGTTGAAGGTGACTGTCAGTGGCAAAATCACTC
ACAGTCCTCAGGCCCATGACAATCCACAGGAGAAGAGACTTGCAGCAGATGACCCAGAGGTGCGGGTTCTACATA
GCATCGGTCACAGTGCTGCCCCACGCCTCTTCCCACTTGCCTGGACTGTGCTGCTCCTTCCACTTCTGCTGCTGCAA
ACCCCGTGAACCGAGGCTACCGGAGGCTCGGGAATGGTCATCCAGAAAGAGAAGAAGAGCTGCGGGCAGGTGGT
TGAGGAGTGGAAGGAGTTCGTGTGGAACCCGAGGACGCACCAGTTCATGGGCCGCACAGGGACCAGCTGGGCCT
TTATCCTCCTCTTCTACCTCGTCTTCTATGGCTTCCTCACCGCCATGTTCACCCTCACCATGTGGGTGATGCTGCAGA
CAGTCTCCGACCACACCCCCAAGTACCAGGACCGATTGGCTACACCAGGCTTGATGATTCGCCCTAAGACTGAGAA
CCTCGATGTCATTGTCAATGTCAGTGACACTGAAAGCTGGGACCAGCATGTTCAGAAACTCAATAAGTTCTTGGAGC
CTTACAATGACTCCATCCAAGCCCAGAAGAATGATGTCTGCCGCCCTGGTCGCTATTACGAACAACCAGATAACGGA
GTTCTAAACTACCCAAAACGTGCTTGCCAGTTCAACCGGACCCAGCTGGGTGACTGCTCTGGCATTGGGGACCCCA
CCCACTATGGTTACAGCACTGGACAGCCCTGTGTCTTCATCAAGATGAACCGGGTCATCAGCTTCTATGCAGGAGC
AAACCAGAGCATGAATGTTACCTGCGTGGGGAAGCGAGATGAAGATGCCGAGAATCTCGGCAACTTCGTTATGTTC
CCTGCAAACGGCAATATCGACCTCATATACTTCCCCTACTATGGCAAGAAGTTTCACGTGAACTACACACAGCCCCT
GGTGGCCGTCAAGTTCCTGAATGTGACCCCCAACGTGGAGGTGAACGTGGAGTGCCGCATCAATGCTGCCAACATT
GCCACTGACGATGAGCGAGACAAGTTCGCCCCCCGTGTGGCCTTCAAACTCCGCATCAACAAAACCTGA

11.1) Ugotovite za katera proteina in iz katerih organizmov nosi zapis nukleotidno zaporedjeY? (4)

Protein 1/organizem	Protein 2/organizem
Ephirin-A1/Homo sapiens	sodium/potassium-transporting ATPase subunit beta-2 [Bos taurus]