BD PROJECT

Data :

UniProt : P54315

PfamID : PF00151

Domain Position : 18-353

Organism : Homo sapiens (Human)

Pfam Name : Lipase/vitellogenin

Domain Sequence : KEVCYEDLGCFSDTEPWGGTAIRPLKILPWSPEKIGTRFLLYTNENPNNFQILLLSDPSTIEASNFQMDRKTRFIIHGFIDKGDESWVTDMCKKLFEVEEVNCICVDWKKGSQATYTQAANNVRVVGAQVAQMLDILLTEYSYPPSKVHLIGHSLGAHVAGEAGSKTPGLSRITGLDPVEASFESTPEEVRLDPSDADFVDVIHTDAAPLIPFLGFGTNQQMGHLDFFPNGGESMPGCKKNALSQIVDLDGIWAGTRDFVACNHLRSYKYYLESILNPDGFAAYPCTSYKSFESDKCFPCPDQGCPQMGHYADKFAGRTSEEQQKFFLNTGEASNF

Task 1

Use UniRef90 :

1. Domain Characteristics:

* PF00151 is a well-characterized domain family
* Lipases are found across many species
* Your sequence is from human (P54315), but you want to capture diversity

1. Why UniRef90 is particularly good here:

* It clusters sequences at 90% identity, which:
  + Removes redundant sequences (like multiple human variants)
  + Keeps enough diversity to capture functional variations in lipases
  + Will include homologs from different species
  + Reduces bias from overrepresented species

Blast Service :

<https://www.uniprot.org/blast>

Using these Parameters : (E-Thresh 0.001 to be more stringent and get high quality homologous , increased Hits to 500 (half of maximal possible, we want high quality homologous, but not too much which may result in not actual homologous and mess up our model)

