1. Does your protein have an HMM available in the PFAM database?

Our protein 6XG6 (Full-length human mitochondrial Hsp90 (TRAP1) with ADP-BeF3) has a HMM available in the PFAM database.
This are the steps we followed to find out:

1. We got into **Uniprot** and downloaded the fasta (canonical) file for our protein 6XG6

```
>6XG6_1|Chains A, B|Heat shock protein 75 kDa, mitochondrial, Fibronectin
binding protein fusion|Homo sapiens (9606)
GIDPFTSTQTAEDKEEPLHSIISSTESVQGSTSKHEFQAETKKLLDIVARSLYSEKEVFIRELISNASDALEKLRHKLVS
DGQALPEMEIHLQTNAEKGTITIQDTGIGMTQEELVSNLGTIARSGSKAFLDALQNQAEASSKIIGQFGVGFYSAFMVAD
RVEVYSRSAAPGSLGYQWLSDGSGVFEIAEASGVRTGTKIIIHLKSDCKEFSSEARVRDVVTKYSNFVSFPLYLNGRRMN
TLQAIWMMDPKDVGEWQHEEFYRYVAQAHDKPRYTLHYKTDAPLNIRSIFYVPDMKPSMFDVSRELGSSVALYSRKVLIQ
TKATDILPKWLRFIRGVVDSEDIPLNLSRELLQESALIRKLRDVLQQRLIKFFIDQSKKDAEKYAKFFEDYGLFMREGIV
TATEQEVKEDIAKLLRYESSALPSGQLTSLSEYASRMRAGTRNIYYLCAPNRHLAEHSPYYEAMKKKDTEVLFCFEQFDE
LTLLHLREFDKKKLISVETDIVVDHYKEEKFEDRSPAAECLSEKETEELMAWMRNVLGSRVTNVKVTLRLDTHPAMVTVL
EMGAARHFLRMQQLAKTQEERAQLLQPTLEINPRHALIKKLNQLRASEPGLAQLLVDQIYENAMIAAGLVDDPRAMVGRL
NELLVKALERHGGSGSGSSAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWIS
DGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHI
```

2. We used **hmmscan** to find what the HMMs that would match our sequence, hence finding a model.

```
hmmscan ~/Documents/databases/Pfam-A.hmm 6XG6.fasta > 6XG6_hmm.pfam.out
```

```
Query:       6XG6_1|Chains  [L=774]
Description: A, B|Heat shock protein 75 kDa, mitochondrial, Fibronectin binding protein fusion|Homo sapiens (9606)
Scores for complete sequence (score includes all domains):
   --- full sequence ---    --- best 1 domain ---     -#dom-
    E-value  score  bias     E-value  score  bias     exp  N  Model       Description
    -------  ------ -----     -------  ------ -----    ----  --  --------    -----------
      3e-80  270.0   2.7     2.1e-77  260.6   1.3     2.1  2  HSP90        Hsp90 protein
    6.1e-13   48.4   0.1     1.9e-12   46.9   0.0     1.9  2  HATPase_c_3  Histidine kinase-, DNA gyrase B-, and HSP90-like
    1.2e-12   47.2   0.6     2.7e-12   46.1   0.6     1.6  1  Cna_B        Cna protein B-type domain
    6.9e-09   35.2   0.1     2.1e-08   33.7   0.1     1.9  1  HATPase_c    Histidine kinase-, DNA gyrase B-, and HSP90-like
      4e-07   29.3   1.4     8.1e-07   28.3   1.4     1.5  1  Fn_bind      Fibronectin binding repeat
  ------ inclusion threshold ------
       0.02   14.4   0.2       0.23   11.0   0.0     2.4  2  GYD          GYD domain
```

In PFAM

ACCESSION	NAME	SOURCE DATABASE	MATCHES
PF00183	Hsp90 protein	Pfam	200 400 600
PF02986	Fibronectin binding repeat	Pfam	200 400 600
PF13085	2Fe-2S iron-sulfur cluster binding domain	Pfam	200 400 600
PF13589	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase	Pfam	200 400 600
PF17802	Prealbumin-like fold domain	Pfam	200 400 600

3. Using **hmmfetch** to extract a profile from the PFAM that corresponds to our model.

```
hmmfetch ~/Documents/databases/Pfam-A.hmm HSP90 > 6XG6.hmm
```

Now we have a hidden markov model from the PFAM database that is the one used for the domain which is most similar to our target.
We have chosen to study the domain of **HSP90** since it has the best e-value, either on the full sequence and on the full sequence, so it shall be the most representative.

2. Choose a set of 6 to 8 amino acid sequences that belong to the protein family you are studying. These sequences should represent the evolutionary history of your protein family, so you want them to have some diversity between them and avoid redundant or highly similar pairs of sequences. You will use these sequences to build a multiple sequence alignment. From what database should you retrieve these sequences? Why?

We are choosing to retrieve these sequences from the Uniprot database, since it is not redundant and not biased and it has a lot of proteins since we have the ones with available sequences. On the other hand PDB does contain the proteins with available structure therefore there are less proteins, it is also biased and redundant.

In order to avoid redundancy and highly similar sequences que decided to obtain the sequences from Uniprot using psiblast, since our protein has a large sequence. Usually HMM is better when it comes to small sequences and to look into small amounts of domains.

To get a better understanding of the evolutionary history of our protein family, we have decided to split it in two multiple alignments; one for paralogous and another for orthologous.
These are the steps we used to do so:

1. Using **psiblast** to search sequences for similar protein families than ours.

```
psiblast -query rcsb_pdb_6XG6.fasta -num_iterations 5 -out_pssm 6xg6_sprot5.pssm
-out 6xg6_sprot_5.out -db ~/Documents/databases/uniprot_sprot.fasta
```

2. Choose **candidates**, those sequences of amino acids that are going to be relevant for the evolutionary history of our protein.

Paralogs

In humans, the Hsp90 family includes several paralogs like Hsp90AA1 and Hsp90AB1 in the cytosol, Grp94 in the endoplasmic reticulum, and Trap1 in the mitochondria. Each paralog serves distinct functions, adapting to the unique needs of their respective cellular compartments, from folding and stabilizing proteins to maintaining cellular integrity and quality control. This diversification underscores the specialized roles of Hsp90 chaperones in human biology.

sp|Q12931|TRAP1_HUMAN Heat shock protein 75 kDa, mitochondrial ...	**1328**	**0.0**
sp|P14625|ENPL_HUMAN Endoplasmin OS=Homo sapiens GN=HSP90B1 PE=...	330	6e-100
sp|Q58FF7|H90B3_HUMAN Putative heat shock protein HSP 90-beta-3...	183	3e-48
sp|Q14568|HS902_HUMAN Putative heat shock protein HSP 90-alpha ...	176	7e-48
sp|Q58FF3|ENPLL_HUMAN Putative endoplasmin-like protein OS=Homo...	104	9e-23

Orthologs

Heat-shock protein 90 (Hsp90) molecular chaperones are conserved across different species, yet exhibit organism-specific dynamic behaviors. This is exemplified by the diversity among Hsp90 orthologs, such as bacterial Hsp90 (HtpG) and eukaryotic cytosolic Hsp90 (Hsp82).

sp|Q12931|TRAP1_HUMAN Heat shock protein 75 kDa, mitochondrial ...	**1328**	**0.0**
sp|Q24VT7|HTPG_DESHY Chaperone protein htpG OS=Desulfitobacteri...	511	6e-172
sp|A4SLY0|HTPG_AERS4 Chaperone protein htpG OS=Aeromonas salmon...	408	8e-132
sp|P0A6Z3|HTPG_ECOLI Chaperone protein htpG OS=Escherichia coli...	385	6e-123
sp|Q69QQ6|HSP82_ORYSJ Heat shock protein 81-2 OS=Oryza sativa s...	345	6e-107
sp|O02705|HS90A_PIG Heat shock protein HSP 90-alpha OS=Sus scro...	189	1e-49
sp|P82995|HS90A_RAT Heat shock protein HSP 90-alpha OS=Rattus n...	189	3e-49
sp|P11501|HS90A_CHICK Heat shock protein HSP 90-alpha OS=Gallus...	187	8e-49
sp|Q76LV1|HS90B_BOVIN Heat shock protein HSP 90-beta OS=Bos tau...	183	2e-47

The sequence for our target protein is the one highlighted in bold.

3. Make a sequence alignment with the sequences you just obtained in the previous step. To create this alignment, use the HMM you found in PFAM and the programs from the HMMer package.

These are the steps we followed to compute the alignments both for paralogs and orthologs of the HSP90.

1. Retrieve all the sequences, in fasta canonical format, from **Uniprot**. Put all the sequences that are going to be used in the same alignment into one single file.

```
perl ~/Documents/perl_scripts/FetchFasta.pl -i human.list -d
~/Documents/databases/uniprot_sprot.fasta -o human.fasta

perl ~/Documents/perl_scripts/FetchFasta.pl -i other.list -d
~/Documents/databases/uniprot_sprot.fasta -o other.fasta
```

2. Put all the sequences found before and our target sequence in the same file.

```
cat 6XG6.fasta > human_pssm.fasta
cat human.fasta >> human_pssm.fasta

cat 6XG6.fasta > other_pssm.fasta
cat other.fasta >> other_pssm.fasta
```

3. Use **hmmalign** to perform the alignment using the HMM we obtained before and the fasta file with all the sequences we selected.

Paralogs

```
hmmalign 6XG6.hmm human_pssm.fasta > human.sto
```

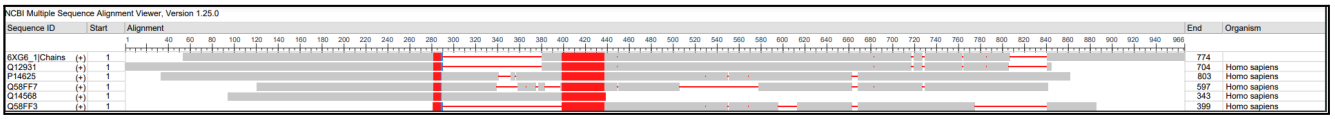
Orthologs

```
hmmalign 6XG6.hmm other_pssm.fasta > other.sto
```

It makes sense that the alignment for the orthologs is better than the paralogs one, since the paralogs are proteins which have evolved to adapt to the environment in which they are found, for example different cells, this means that they would have slightly different functions to which we expect they would have different structures.
The orthologs selected have some variability on the alignment, which makes sense since the location of the HSP90 in the individual is also a telling of the functionality and so, there are some regions of variability in the alignment. Yet we can see that the alignment between orthologs are more similar than the alignment between paralogs.

4. Search for conserved regions in your alignment. Do these regions correspond with the essential regions you described in the previous assignment (question 6)? Why do you think this is happening? Provide images of your alignment to support your explanation. In this images, the alignments should be in clustalw format, use the perl script we learnt in practice 2 to change the format of the alignments produced by hmmer programs.

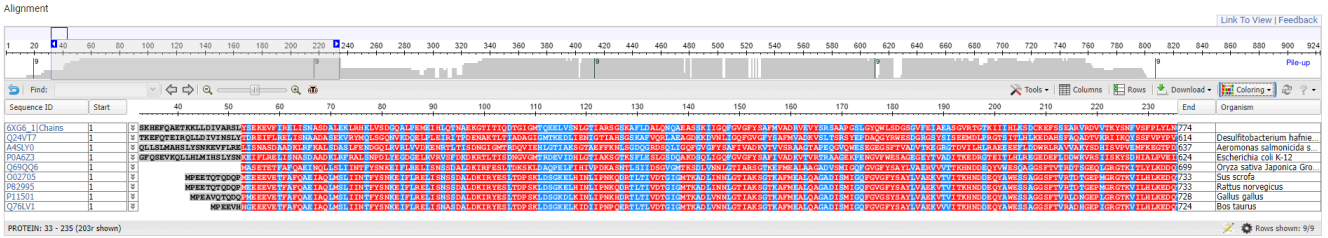
The highly conserved regions are colored in red (if there's no gap in the alignment), those that are colored in blue indicate lower conservation, so conserved regions which have some slight variability.
In paralogs, less conserved regions:

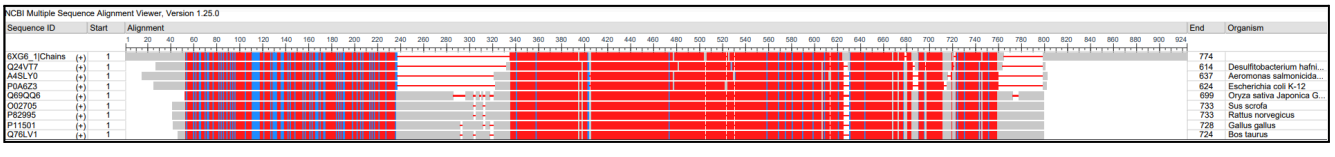


CORRECTIONS:
https://www.ncbi.nlm.nih.gov/projects/msaviewer/?key=N4ihWVeKgKuspLash7Wlotvf2d7R3P3W9c775G7pyKsSnXuOM3kCVD4YMWFkfT1IL014TDtSK0MxWSFGF2suaQ5ZKA,9klgmJZLQWptZXdtRnRJYxoeGB8QHTwXNA86Ja8oCWrTXMoyWTFoztWuRtcTy0rTWPsP-kzkXPVG71bwYN1Z33nvXw

took out some of the sequences (so that the gaps match the Chains sequence, this is from the previous alignment, so I'd have to remake the alignment):

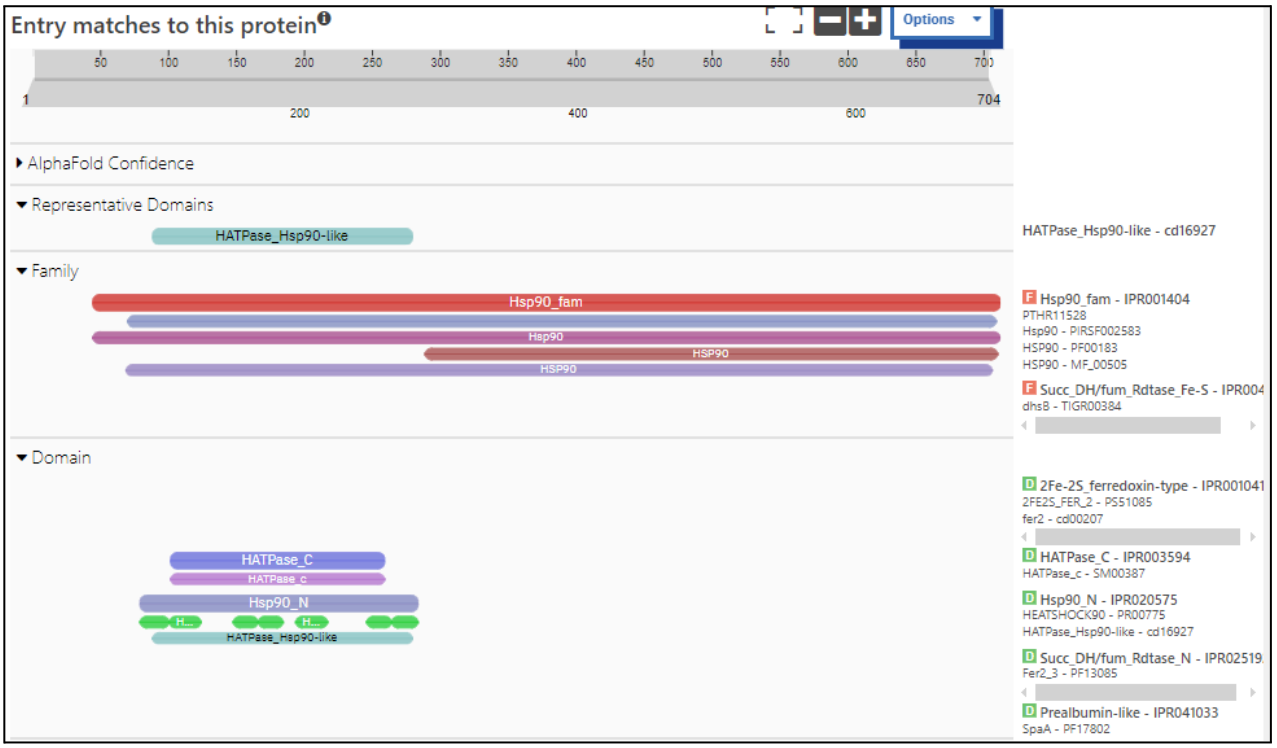
In orthologs, more conserved regions:
https://www.ncbi.nlm.nih.gov/projects/msaviewer/?key=hzgR6ec6MBscFAYcNwU4EmtvaW5hbE1mRX5LVN5ZeByXKmJDHFltdBmNFPRB6BjwCthd2R7HDtYUzATTMv4L_CvMDQ,FKuCenSpo4iPh5WPpJargfj8-v3y_9711u3Yx03K648EuU3F-A_J5-Z8dgUjGXoBaCk_KHw2bCd2PWYiUA9pDUk9bw
Segun lo que hemos escrito abajo esto seria
N-terminal: 33-232





These regions do correspond with some of the essential regions we depicted on the previous assessment. In the paralogs there are few regions which are conserved which would make sense since the function of these proteins have adapted to the environment they occupy, they may have evolved to perform other functions.

Here we can see some of the essential regions highlighted in InterPRO. We have used this to compare and contrast if the conserved regions are essential or not.



To change the format of the alignments, we use this command:

```
perl ~/Documents/perl_scripts/aconvertMod2.pl -in h -out c <human.sto> human.aln
perl ~/Documents/perl_scripts/aconvertMod2.pl -in h -out c <other.sto> other.aln
```

Paralogs

```
6XG6_1|Chains          ------------------------------------------------------gidpfts
sp|Q12931|TRAP1_HUMAN  marelralllwgrrlrpllrapalaavpggkpilcprrttaqlgprrnpawslqagrlfs
sp|P14625|ENPL_HUMAN   ---------------------------------mralwvlglccvlltfgsvraddevdv
sp|Q58FF7|H90B3_HUMAN  ------------------------------------------------------------
sp|Q58FF3|ENPLL_HUMAN  ------------------------------------------------------------

6XG6_1|Chains          tqtaedkeeplhsiisstesvqgstskhefqaetkklldivarslysekevfirelisna
sp|Q12931|TRAP1_HUMAN  tqtaedkeeplhsiisstesvqgstskhefqaetkklldivarslysekevfirelisna
sp|P14625|ENPL_HUMAN   dgtveedlgksregsrtddevvqreeeaiqldglnasqirelreksekfafqaevnrmmk
sp|Q58FF7|H90B3_HUMAN  ------------------------------------------------------------
sp|Q58FF3|ENPLL_HUMAN  ------------------------------------------------------------

6XG6_1|Chains          sdaleklrhklvsdgqalpemeihlqtnaekgtitiqdtgigmtqeelvsnlgtiarsgs
sp|Q12931|TRAP1_HUMAN  sdaleklrhklvsdgqalpemeihlqtnaekgtitiqdtgigmtqeelvsnlgtiarsgs
sp|P14625|ENPL_HUMAN   liinslyknkeiflrelisnasdaldkirlisltdenalsgneeltvkikcdkeknllhv
sp|Q58FF7|H90B3_HUMAN  mpeevhhgeeevetfafqaeiaqlisliintfysneeiflqelisnasdaldkiryeslt
sp|Q58FF3|ENPLL_HUMAN  ------------------------------------------------------------

6XG6_1|Chains          kafldalqnqaeasskiigqfgvgfysafmvadrvevysrsaapgslgyqwlsdgsgvfe
sp|Q12931|TRAP1_HUMAN  kafldalqnqaeasskiigqfgvgfysafmvadrvevysrsaapgslgyqwlsdgsgvfe
sp|P14625|ENPL_HUMAN   tdtgvgmtreelvknlgtiaksgtseflnkmteaqedgqstseligqfgvgfysaflvad
sp|Q58FF7|H90B3_HUMAN  dpskldsgkelkidiipnpqertlalvdtgigmtkadlinnlrtiaksgtkacmealqae
sp|Q58FF3|ENPLL_HUMAN  ------------------------------------------------------------

6XG6_1|Chains          iaeasgvrtgtkiiihlksdckefssearvrdvvtkysnfvsfplylng-----------
sp|Q12931|TRAP1_HUMAN  iaeasgvrtgtkiiihlksdckefssearvrdvvtkysnfvsfplylng-----------
sp|P14625|ENPL_HUMAN   kvivtskhnndtqhiwesdsnefsviadprgntlgrgttitlvlkeeasDYLELDTIKNL
sp|Q58FF7|H90B3_HUMAN  klvvitkhnddeqyawessaggsftvhadhgepigrgtkvilhlkedqtEYLEERRVKEV
sp|Q58FF3|ENPLL_HUMAN  ----------------------------------------maetiqev-----------

6XG6_1|Chains          ------------------------------------------------------------
sp|Q12931|TRAP1_HUMAN  ------------------------------------------------------------
sp|P14625|ENPL_HUMAN   VKKYSQFINFPIYVWSSKTETVEEPMEEEEAAKEEKEESD---------DEAAVEEEEEE
sp|Q58FF7|H90B3_HUMAN  VKKHSQFIGYPITLYLEKEQDKEISDDEAEEEKGEKEE---------------EDKDDEE
sp|Q58FF3|ENPLL_HUMAN  ------------------------------------------------------------
```

```
6XG6_1|Chains          ---------------RRMNTLQAIWMMDPKDVGEWQHEEFYRYVAQAHDKPRYTLHYKTD
sp|Q12931|TRAP1_HUMAN  ---------------RRMNTLQAIWMMDPKDVREWQHEEFYRYVAQAHDKPRYTLHYKTD
sp|P14625|ENPL_HUMAN   KKPKTKKVEKTVWDWELMNDIKPIWQRPSKEVEEDEYKAFYKSFSKESDDPMAYIHFTAE
sp|Q58FF7|H90B3_HUMAN  -KPKIKDVG---SDEED---------------DSKEYGEFYKSLTSDWEDHLAVKHFSVE
sp|Q58FF3|ENPLL_HUMAN  --------------------------------EDEYKAFCKSFSKESDDPVACIHFTAE

6XG6_1|Chains          APLNIRSIFYVPDMKPS-MFDVS-rELGSSVALYSRKVLIQTKATDILPKWLRFIRGVVD
sp|Q12931|TRAP1_HUMAN  APLNIRSIFYVPDMKPS-MFDVS-rELGSSVALYSRKVLIQTKATDILPKWLRFIRGVVD
sp|P14625|ENPL_HUMAN   GEVTFKSILFVPTSAPRGLFDEYgsKKSDYIKLYVRRVFITDDFHDMMPKYLNFVKGVVD
sp|Q58FF7|H90B3_HUMAN  GQLEFRALLFSPRRAPFDLFENK--KKKNNIKLYVRRVFIMDSCDELIPEYLNFIHGVVD
sp|Q58FF3|ENPLL_HUMAN  GEVTFKSILFVPTFVPRGLFDEYgsKKSDYIKLYVRCVFITDDFRDTMPKNLNFVKGVVD

6XG6_1|Chains          SEDIPLNLSRELLQESALIRKLRDVLQQRLIKFFIDQSkKDAEKYAKFFEDYGLFMREGI
sp|Q12931|TRAP1_HUMAN  SEDIPLNLSRELLQESALIRKLRDVLQQRLIKFFIDQSkKDAEKYAKFFEDYGLFMREGI
sp|P14625|ENPL_HUMAN   SDDLPLNVSRETLQQHKLLKVIRKKLVRKTLDMIKKIA-DDKYN-DTFWKEFGTNIKLGV
sp|Q58FF7|H90B3_HUMAN  SEDLPLNISREMLQQSKILK-----------------------------------------
sp|Q58FF3|ENPLL_HUMAN  SGGLSLNVSCETLQQHKLLKVIRKKLVHKTLDMIKKIA-DEKYN-DTFWKEFGTNIKLGV
```

```
6XG6_1|Chains          VTATeqEVKEDIAKLLRYESSALpSGQLTSLSEYASRMRAGTRNIYYLCAPNRHLAEHSP
sp|Q12931|TRAP1_HUMAN  VTATeqEVKEDIAKLLRYESSALpSGQLTSLSEYASRMRAGTRNIYYLCAPNRHLAEHSP
sp|P14625|ENPL_HUMAN   IEDH--SNRTRLAKLLRFQSSHH-PTDITSLDQYVERMKEKQDKIYFMAGSSRKEAESSP
sp|Q58FF7|H90B3_HUMAN  --------------------------------YVSHMKETQKSTYYITGESKEQVANSA
sp|Q58FF3|ENPLL_HUMAN  IEDH--SNRTCLAKLLRFQSSHH-PADITSLHQDVERMKEKQDKICLMAG----------

6XG6_1|Chains          YYEAMKKKDTEVLFCFEQFDELTLLHLREFDKKKLISVETD-IVVDHYKEEKFEDRSpaa
sp|Q12931|TRAP1_HUMAN  YYEAMKKKDTEVLFCFEQFDELTLLHLREFDKKKLISVETD-IVVDHYKEEKFEDRSpaa
sp|P14625|ENPL_HUMAN   FVERLLKKGYEVIYLTEPVDEYCIQALPEFDGKRFQNVAKEGVKFDESEKTKESREA---
sp|Q58FF7|H90B3_HUMAN  FVERVRKQGFEVVYMTEPIDEYCVQQLKEFDGKSLVSVTKEGLELPEDEEEKKKMEE---
sp|Q58FF3|ENPLL_HUMAN  --------GYEVIYLTEPVVEYCIQALPEFDGKRFQNVAKEGVKFDDSEKTKESHEA---

6XG6_1|Chains          eclSEKETEELMAWMRN-VLGSRVTNVKVTLRLDTHPAMVTVLEMGAARHF---LRMQQL
sp|Q12931|TRAP1_HUMAN  eclSEKETEELMAWMRN-VLGSRVTNVKVTLRLDTHPAMVTVLEMGAARHF---LRMQQL
sp|P14625|ENPL_HUMAN   ---VEKEFEPLLNWMKDkALKDKIEKAVVSQRLTESPCALVASQYGWSGNMERIMKAQAY
sp|Q58FF7|H90B3_HUMAN  ---SKEKFENLCKLMKE-ILDKKVEKVTISNRLVSSPCCIVTSTYGWTANMEQIMKAQAL
sp|Q58FF3|ENPLL_HUMAN  ---VEKEFEPLPNWVKDkAIKDKIEKAMVSQCLTESLCALVASQYGWSGNMERIMKAQAY
```

5666666

Orthologs

```
6XG6_1|Chains          ------------------------------------------------------------
sp|Q24VT7|HTPG_DESHY   ------------------------------------------------------------
sp|A4SLY0|HTPG_AERS4   ------------------------------------------------------------
sp|P0A6Z3|HTPG_ECOLI   ------------------------------------------------------------
sp|Q69QQ6|HSP82_ORYSJ  ERRLKDLVKKHSEFISYPISLWTEKTTEKEISDDEDEEEKKDAEE----------GKVE
sp|O02705|HS90A_PIG    ERRIKEIVKKHSQFIGYPITLFVEKERDKEVSDDEAEEKEDKEEEKEKEEKeseDKPEIE
sp|P82995|HS90A_RAT    ERRIKEIVKKHSQFIGYPITLFVEKERDKEVSDDEAEEKEEKEEEKEKEEKesdDKPEIE
sp|P11501|HS90A_CHICK  ERRIKEIVKKHSQFIGYPIRLFVEKERDKEVSDDEAEEKEEEKEEKEEKTE---DKPEIE
sp|Q76LV1|HS90B_BOVIN  ERRVKEVVKKHSQFIGYPITLYLEKEREKEISDDEAEEEKGEKEEEDKDDE---EKPKIE

6XG6_1|Chains          ----------------------------------RRMNTLQAIWMMDPKDVGEWQHEEF
sp|Q24VT7|HTPG_DESHY   --------------------------------VGEEKINTVQALWTKNKNEISEEEYKEF
sp|A4SLY0|HTPG_AERS4   --------------------EEDGETVVGTPGEWEQVNRATALWTRNPKEIKDEEYQEF
sp|P0A6Z3|HTPG_ECOLI   --------------------KREEKDGETVISWEKINKAQALWTRNKSEITDEEYKEF
sp|Q69QQ6|HSP82_ORYSJ  DV---DEE--K---EEK----EKKKKKIKEVSHEWNVMNKQKPIWLRKPEEITKEEYAAF
sp|O02705|HS90A_PIG    DV---GSDEEE---EEKkdgdKKKKKKIKEKYIDQEELNKTKPIWTRNPDDITNEEYGEF
sp|P82995|HS90A_RAT    DV---GSDEEE---EEKkdgdKKKKKKIKEKYIDQEELNKTKPIWTRNPDDITNEEYGEF
sp|P11501|HS90A_CHICK  DVgsdEEEEKK---DGD----KKKKKKIKEKYIDEEELNKTKPIWTRNPDDITNEEYGEF
sp|Q76LV1|HS90B_BOVIN  DV---GSDEEDdsgKDK----KKKTKKIKEKYIDQEELNKTKPIWTRNPDDITQEEYGEF
```

```
6XG6_1|Chains          ISVETD-IVVDH--YKEEKFEDRSpaaeclSEKETEELMAWMRNVLGSRVTNVKVTLRLD
sp|Q24VT7|HTPG_DESHY   LSADHGDLNLAD--EDGASAEEKL-----lSEEELKEFNDWLKEVLGEKVTEVRESKRLV
sp|A4SLY0|HTPG_AERS4   VSVTRGELDLGDleDEASKQAQEE------AEKANAGLVERVKTSLGEAVKEVRVTHRLT
sp|P0A6Z3|HTPG_ECOLI   QSVSKVDESLEK-lADEVDESAKE------AEKALTPFIDRVKALLGERVKDVRLTHRLT
sp|Q69QQ6|HSP82_ORYSJ  VSATKEGLKLDE--SEDEKKRQEE------LKEKFEGLCKVIKEVLGDKVEKVVVSDRVV
sp|O02705|HS90A_PIG    VSVTKEGLELPE--DEEEKKKQEE------KKTKFENLCKIMKDILEKKVEKVVVSNRLV
sp|P82995|HS90A_RAT    VSVTKEGLELPE--DEEEKKKQEE------KKTKFENLCKIMKDILEKKVEKVVVSNRLV
sp|P11501|HS90A_CHICK  VSVTKEGLELPE--DEEEKKKQEE------KKAKFENLCKIMKDILEKKVEKVVVSNRLV
sp|Q76LV1|HS90B_BOVIN  VSVTKEGLELPE--DEEEKKKMEE------SKAKFENLCKLMKEILDKKVEKVTISNRLV

6XG6_1|Chains          THPAMVTVLEMGAARHF---LRMQQLAKTQEERAQLLQPTLEINPRHALIKKLNQLRASE
sp|Q24VT7|HTPG_DESHY   DSPAIIL-SHYG-THSMQRMMQLM---NRDLQDVP--AGILEINPKHVLIQRLNDLRKQE
sp|A4SLY0|HTPG_AERS4   DSPSCIVTDNHGMSTQMIKLMRAA------GQPVPEQKYILELNPDHALVKKLDT----I
sp|P0A6Z3|HTPG_ECOLI   DTPAIVSTDADEMSTQMAKLFAAA------GQKVPEVKYIFELNPDHVLVK----RAADT
sp|Q69QQ6|HSP82_ORYSJ  DSPCCLVTGEYGWTANMERIMKAQALRDSSMAGYMSSKKTMEINPENSIMDELRKRADAD
sp|O02705|HS90A_PIG    TSPCCIVTSTYGWTANMERIMKAQALRDNSTMGYMAAKKHLEINPDHSIIETLRQKAEAD
sp|P82995|HS90A_RAT    TSPCCIVTSTYGWTANMERIMKAQALRDNSTMGYMAAKKHLEINPDHSIIETLRQKAEAD
sp|P11501|HS90A_CHICK  TSPCCIVTSTYGWTANMERIMKAQALRDNSTMGYMAAKKHLEINPDHSIIETLRQKAEAD
sp|Q76LV1|HS90B_BOVIN  SSPCCIVTSTYGWTANMERIMKAQALRDNSTMGYMMAKKHLEINPDHPIVETLRQKAEAD

6XG6_1|Chains          P--GLAQLLVDQIYENAMIAAGL-VDDPRAMVGRLNELLVKALE----------------
sp|Q24VT7|HTPG_DESHY   --DSFAPLAAEQLFANAQIAAGIIV-DPRSMVSRLNEILEKAL-----------------
sp|A4SLY0|HTPG_AERS4   EDEALFGEWVTLLHEQAQLAEQGGLNDPASFVSRINRLL---------------------
sp|P0A6Z3|HTPG_ECOLI   EDEAKFSEWVELLLDQALLAERGTLEDPNLFIRRMNQLL---------------------
sp|Q69QQ6|HSP82_ORYSJ  KNDKSVKDLVMLLFETALLTSGFSLEDPNTFGTRIHRMLKLGLSIDEDESAE------AD
sp|O02705|HS90A_PIG    KNDKSVKDLVILLYETALLSSGFSLEDPQTHANRIYRMIKLGLGIDEDDPTADDSSAAVT
sp|P82995|HS90A_RAT    KNDKSVKDLVILLYETALLSSGFSLEDPQTHANRIYRMIKLGLGIDEDDPTVDDTSAAVT
sp|P11501|HS90A_CHICK  KNDKSVKDLVILLYETALLSSGFSLEDPQTHANRIYRMIKLGLGIDEDDTAAEEASPAVT
sp|Q76LV1|HS90B_BOVIN  KNDKAVKDLVVLLFETALLSSGFSLEDPQTHSNRIYRMIKLGLGIDEDEVTAEEPSAAVP
```

```
6XG6_1|Chains          --------------------rhggsgsgssamvdtlsglsseqgqsgdmtieedsathikf
sp|Q24VT7|HTPG_DESHY   --------------------r-------------------------------------------
sp|A4SLY0|HTPG_AERS4   --------------------lqa-----------------------------------------
sp|P0A6Z3|HTPG_ECOLI   --------------------vs------------------------------------------
sp|Q69QQ6|HSP82_ORYSJ  ADMPPLEDDAGESKMEEVD-----------------------------------------------
sp|O02705|HS90A_PIG    EEMPPLEGDDDTSRMEEVD-----------------------------------------------
sp|P82995|HS90A_RAT    EEMPPLEGDDDTSRMEEVD-----------------------------------------------
sp|P11501|HS90A_CHICK  EEMPPLEGDDDTSRMEEVD-----------------------------------------------
sp|Q76LV1|HS90B_BOVIN  DEIPPLEGDEDASRMEEVD-----------------------------------------------
```

N-Terminal:
The N-Terminal is well conserved in general
There is an active sita (atp-binding site) which is also conserved (hydrophobic)
Hydrophobic residues involved also well conserved

Middle domain:
Also a well conserved domain
An arginine well conserved

C-Terminal:
….

5. Work with the mutation you choose in the previous assignment (assignment 1, question 7). Find where this mutation would happen in the alignment you created in question 3. Compare the mutated amino acid with the amino acids that you find at that position in your alignment,

do they share similar properties or not? Make a hypothesis of how this mutation is affecting the function of the protein. Provide images of your alignment to support your explanation.

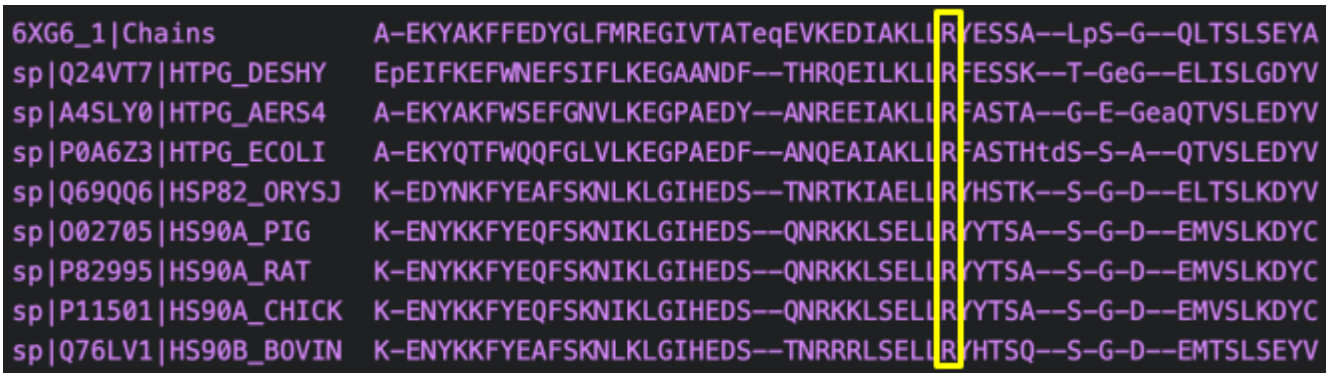
Our mutation is in position **469** :

COSM215648 RCV000676236 RCV001849425 rs144787542	469	R>H	central_nervous_system (Cosmic) Imported Congenital anomaly of kidney and urinary tract (ClinVar) Imported	Pathogenic (Ensembl, ClinVar)	cosmic curated ClinVar 1000Genomes ESP ExAC TOPMed dbSNP gnomAD

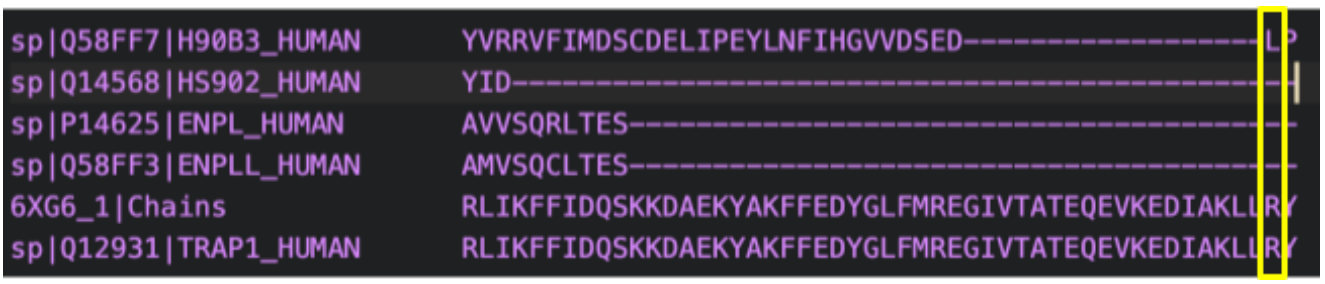
PDBe-KB: UniProt Coverage View

Range of the domains:

- **N-terminal:** SER33 - LEU232
- **Middle domain (Ribosomal-like):** LEU242 - THR499
- **C-terminal:** ASP513 - HIS651



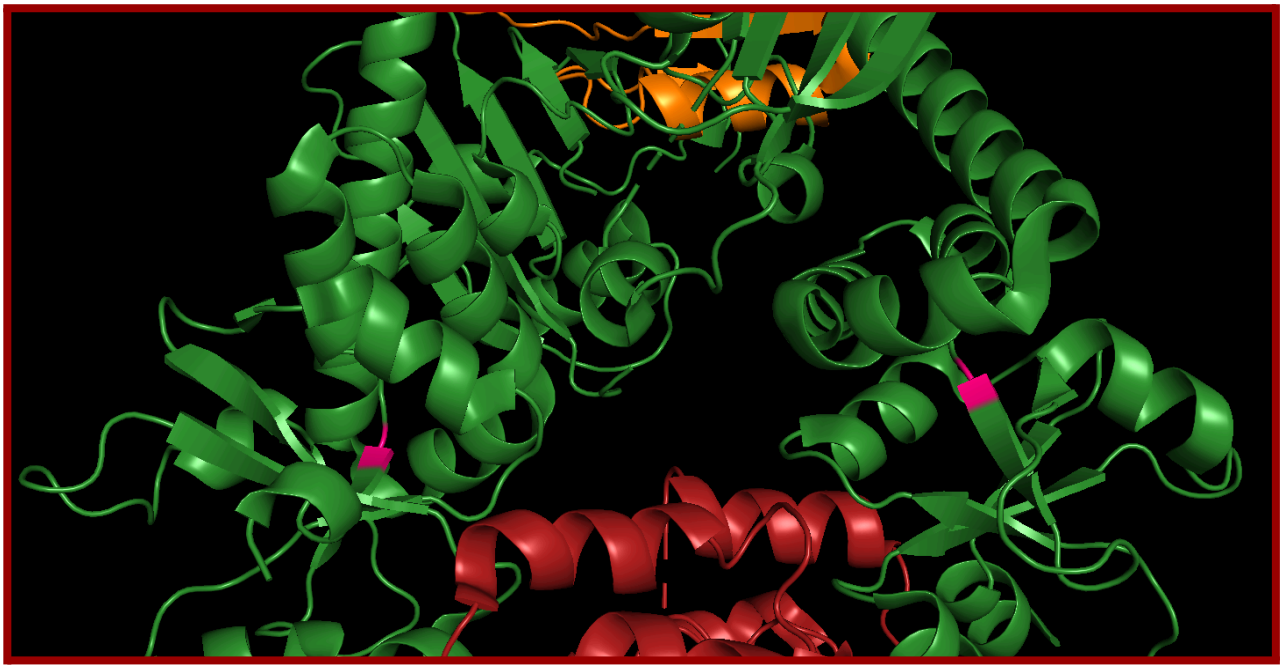
We can see here that it's in a conserved region among all species



But between paralogs is just in the mitochondrial protein

PyMOL:

We can see that the **mutation** (in pink) is happening in the middle domain:



N-terminal
Middle Domain
C-Terminal