# GEN242 Homework 1
Thanin Chantarachot 861129943

Final query syntax from Details page:

P450[All Fields] AND hydroxylase[All Fields] AND "Homo sapiens"[Organism] AND swissprot[filter]

The number of retrieved GIs:

3182968
215274267
117275
117283
6686268
117225
116241312
6166044
20137526
544084
56757430
399288
218511987
84028191
19862747296439388
12585217
3041666
296434466
10719963

Multiple alignment and tree of the sequences using [MultAalin](#):
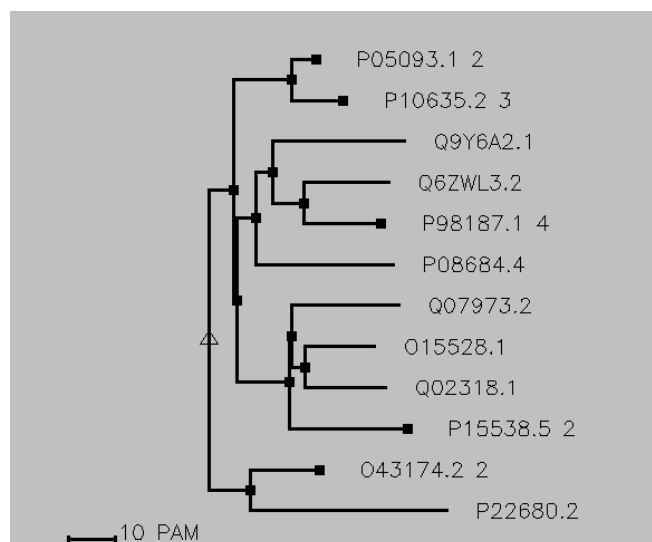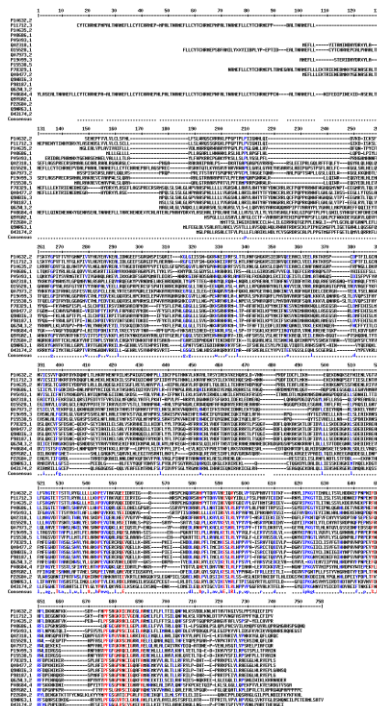
Putative heme binding cysteine:

```
                651    660       670       680       690       700       710       720       730       740       750    758
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+-------|
P10632.2   HFLDKNGNFKK-------SDY--FMPFSAGKRICAGEGLARMELFLFLTTILQNFNLKSVDDLKNLNTTAVTKGIVSLPPSYQICFIPV
P11712.3   HFLDEGGNFKK-------SKY--FMPFSAGKRICVGEALAGMELFLFLTSILQNFNLKSLVDPKNLDTTPVVNGFASVPPFYQLCFIPV
P10635.2   HFLDAQGHFVK-------PEA--FLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVSPSP-YELCAVPR
P08686.1   RFLEPGKNSRA--------------LAFGCGARVCLGEPLARLELFVVLTRLLQAFTL-LPSGDALPSLQPLPHCSVILKMQPFQVRLQPRGMGAHSPGQNQ
P05093.1   RFLNPAGTQLI-------SPSVSYLPFGAGPRSCIGEILARQELFLIMAWLLQRFDLEVPDDGQLPSLEGIPKVVFLIDSFKVKIKVRQAWREAQAEGST
Q02318.1   RWLRNSQPATPR-----IQHPFGSVPFGYGVRACLGRRIAELEMQLLLARLIQKYKYVLAPETG-E-LKSVARIVLVPNKKVGLQFLQRQC
O15528.1   RWL--GEGPTP---------HPFASLPFGFGKRSCMGRRLAELELQMALAQILTHFEVQPEPGAA-P-VRPKTRTVLVPERSINLQFLDR
Q07973.2   RWLQEKEKI----------NPFAHLPFGVGKRMCIGRRLAELQLHLALCWIVRKYDIQ-ATDNE-P-VEMLHSGTLVPSRELPIAFCQR
P19099.3   RWLDIRGSG----------RNFHHVPFGFGMRQCLGRRLAEAEMLLLLHHVLKHFLVETLTQ-E-D-IKMVYSFILRPGTSPLLTFRAIN
P15538.5   RWLDIRGSG----------RNFYHVPFGFGMRQCLGRRLAEAEMLLLLHHVLKHLQVETLTQ-E-D-IKMVYSFILRPSMFPLLTFRAIN
P78329.1   RFDPENIKER---------SPLAFIPFSAGPRNCIGQTFAMAEMKVVLALTLLLRFRVLP-DHT--E-PRRKPELVLRAEGGLWLRVEPLS
Q08477.2   RFDPKNIKER---------SPLAFIPFSAGPRNCIGQAFAMAEMKVVLGLTLLLRFRVLP-DHT--E-PRRKPELVLRAEGGLWLRVEPLS
Q9HBI6.3   RFDQENIKER---------SPLAFIPFSAGPRNCIGQAFAMAEMKVVLALTLLLHFRILP-THT--E-PRRKPELILRAEGGLWLRVEPLGANSQ
P98187.1   RFDPENAQKR---------SPMAFIPFSAGPRNCIGQKFAMAEMKVVLALTLLLRFRILP-DHR--E-PRRTPEIVLRAEDGLWLRVEPLG
Q6ZWL3.2   RFFPENAQGR---------HPYAYVPFSAGPRNCIGQKFAVMEEKTILSCILRHFWIES-NQKREE-LGLEGQLILRPSNGIWIKLKRRNADER
P08684.4   RFSKKNKDNI---------DPYIYTPFGSGPRNCIGMRFALMNMKLALIRVLQNFSFKPCKETQIP-LKLSLGGLLQPEKPVVLKVESRDGTVSGA
Q9Y6A2.1   RFGPGAPKPR----------FTYFPFSLGHRSCIGQQFAQMEVKVVMAKLLQRLEFRLVPGQR---FGLQEQATLKPLDPVLCTLRPRGWQPAPPPPPC
P22680.2   RYLDENGKTKTTFYCNGLKLKYYYMPFGSGATICPGRLFAIHEIKQFLILMLSYFELELIEG------QAKCPPLDQSRAGLGILPPLNDIEFKYKFKHL
Q9NR63.1   RFSQARSEDKDG-------RFHYLPFGSGVRTCLGKHLAKLFLKVLAVELASTSRFELATR---TFPRITLYPVLHPVDGLSVKFFGLDSNQNEILPETEAMLSATV
O43174.2   RFMLPHPEDAS---------RFSFIPFGGGLRSCVGKEFAKILLKIFTVELARHCDWQLLNG---P-PTMKTSPTYYPVDNLPARFTHFHGEI
Consensus  rfl.........     .......PFg.G.R.C.G..lA..e....l...l..f....................l.p........................
```

Putative heme binding cysteine with consensus pattern from Prosite database:

[FW]-[SGNH]-x-[GD]-{F}-[RKHPT]-{P}-C-[LIVMFAP]-[GAD]

CYTOCHROME_P450, PS00086; Cytochrome P450 cysteine heme-iron ligand signature (PATTERN)

- Consensus pattern:
  [FW]-[SGNH]-x-[GD]-{F}-[RKHPT]-{P}-C-[LIVMFAP]-[GAD]
  C is the heme iron ligand
- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 1137
    - detected by PS00086: 1046 (true positives)
    - undetected by PS00086: 91 (81 false negatives and 10 'partials')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS00086:
  46 false positives.
- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:
  Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00086
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00086
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS00086
- View ligand binding statistics of PS00086
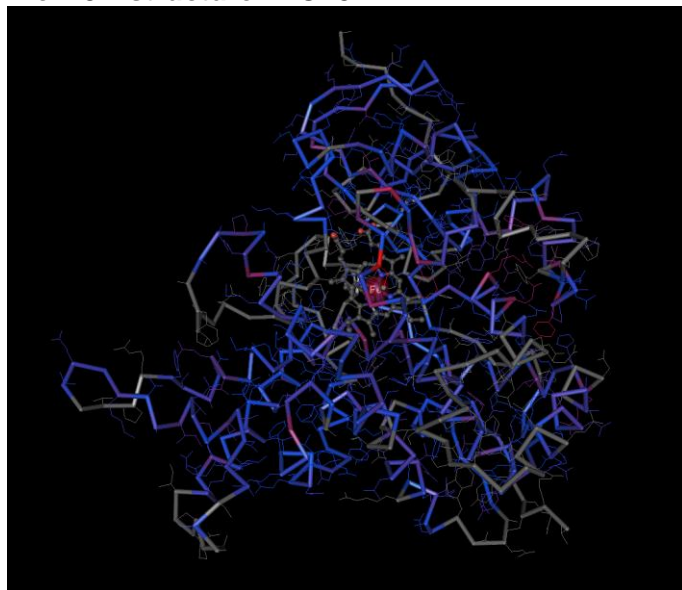- Matching PDB structures: 1AKD 1BU7 1BVY 1C8J ... [ALL]

Corresponding Pfam ID:
PF00067

Mouse (*Mus musculus*) sequences in this family (use species tree from Pfam db):
155

View 3D structure in Cn3D

Yellow is the conserved C residue for heme binding: