

#*Danio rerio* proteome filtering according to UGL enzyme features
#(as in Y.akada, T. Noguchi, Biochem. J., 1986, 235, 391–397)

```
#Download D. rerio and H. sapiens proteomes from Uniprot (one protein sequence per gene)
Url='https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/'
wget -q -O - $Url"Eukaryota/UP000000437/UP000000437_7955.fasta.gz" | gunzip > Dr.fasta
wget -q -O - $Url"Eukaryota/UP000005640/UP000005640_9606.fasta.gz" | gunzip > Hs.fasta
grep -c '>' ??*.fasta
```

```
## Dr.fasta:20358
## Hs.fasta:20607
```

```
#Danio proteins with Mw = 64000 +/- 10%
pepstats Dr.fasta -auto -stdout | perl -ne '$name=$1 if (/STATS of (\S+)/);if (/ weight = (\S+)/)
{printf "$name\n" if ($1>(64000-64000*0.1) and $1<(64000+64000*0.1) )}' | tee Dr_64kDa.acc | wc -l
```

```
## 2043
```

```
#Danio proteins with PTS1**
fuzzpro -pattern "[SAC][KRHSN][LM]>" Dr.fasta -auto -stdout|
perl -lane 'print $F[2] if (/Sequence:\/)'|tee Dr_PTS1.acc | wc -l
```

```
## 175
```

```
#Mw and PTS1 intersection
grep -f Dr_PTS1.acc Dr_64kDa.acc | tee Dr_PTS1_64kDa.acc | wc -l
```

```
## 25
```

```
#Make Blast database
makeblastdb -in Hs.fasta -dbtype prot -parse_seqids -logfile logfile
makeblastdb -in Dr.fasta -dbtype prot -parse_seqids -logfile logfile
```

```
#UGL candidates (in *Danio* and not in *Homo*):
blastdbcmd -db Dr.fasta -entry_batch Dr_PTS1_64kDa.acc > Dr_PTS1_64kDa.fasta
blastp -db Hs.fasta -query Dr_PTS1_64kDa.fasta -evalue 1e-3 -num_threads 4|
tee Dr_PTS1_64kDa.blast | grep -B6 "No hits found" |grep "Query="
```

```
## Query= AOA8M1P3F0 Malate synthase OS=Danio rerio OX=7955 GN=mls1 PE=4 SV=1
## Query= I6V1W0 Protein brambleberry OS=Danio rerio OX=7955 GN=bmb PE=2 SV=1
```