Seminar 8

Evidentierea legaturii dintre gena si secventa proteica corespunzatoare

(Identificarea genei GFP si determinarea mutaţiilor importante)

Pentru a demonstra realizarea activitatilor urmatoare, creati un fisier Word (nume_seminar8.doc) in care sa salvati ("print screen" si apoi "paste") imaginile create!

1) Determinarea secventei ADN a proteinei "Green Fluorescent Protein" (in baza de date UniProtKB)
2) Translatia secventei AND
3) Compararea secventei de aminoacizi obtinuta cu secvente similare din baza de date PDB
4) Vizualizarea mutatiilor genetice

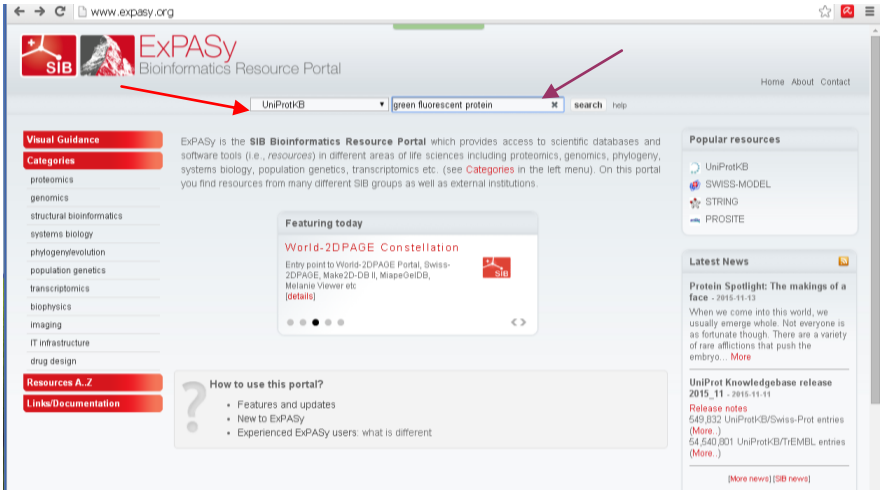
1) Determinarea secventei ADN a proteinei "Green Fluorescent Protein" (in baza de date UniProtKB)

Baza de date **UniProtKB** (www.uniprot.org/ www.expasy.org/):

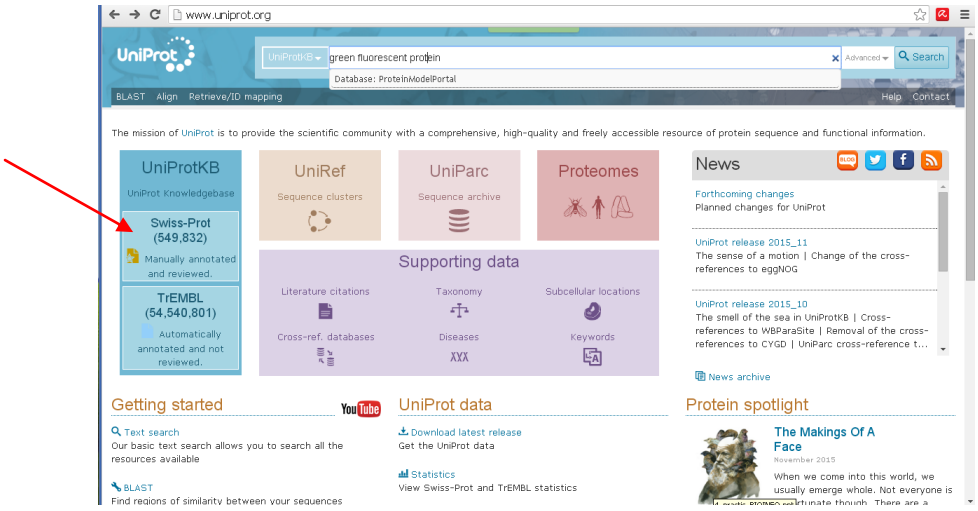
- organizeaza si noteaza secventele proteice.
- contine informatii importante pentru studierea relatiei dintre secventa proteica si functia proteinei.
- foloseste 2 tipuri de notatii:
  - -- una alocata manual (de cercetatori) bazata pe articolele publicate
  - -- alta alocata automat de un software sofisticat.

Pentru aceasta activitate practica vom folosi notatii alocate manual!

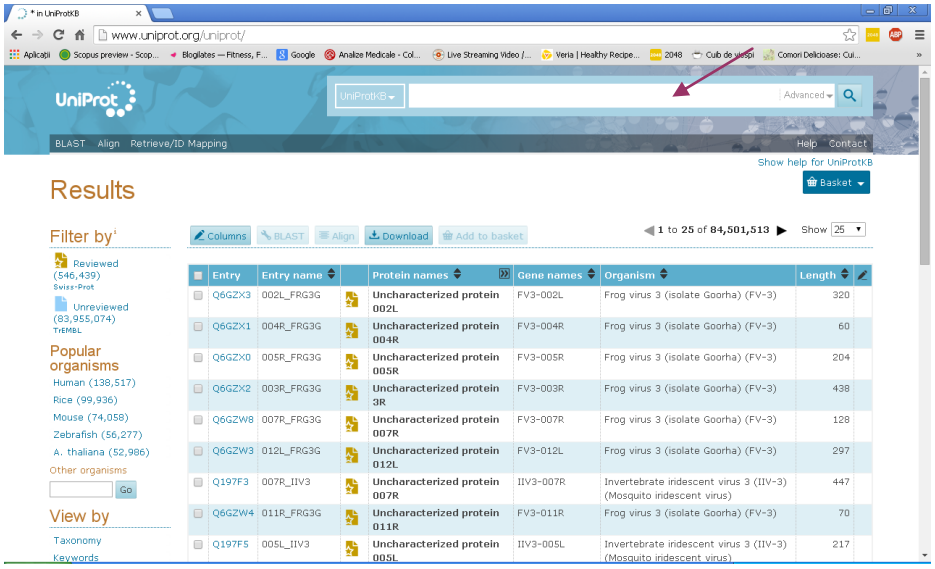
Daca ati folosit adresa "www.expasy.org/" se selecteaza "UniProtKB" in lista bazelor de date si se scrie in campul de cautare "Green Fluorescent Protein".



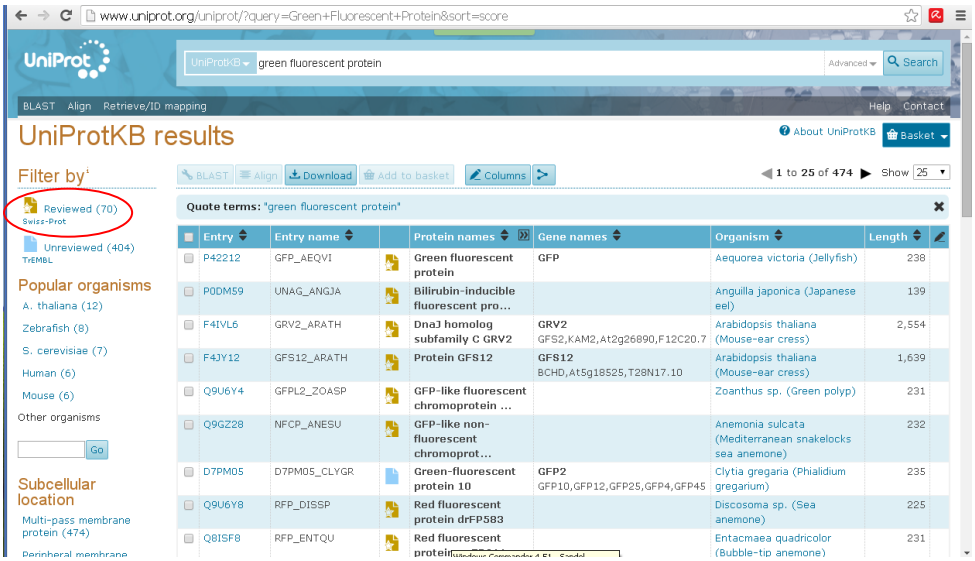
Daca ati folosit adresa "www.uniprot.org/" se selecteaza sectiunea "Swiss-Prot"



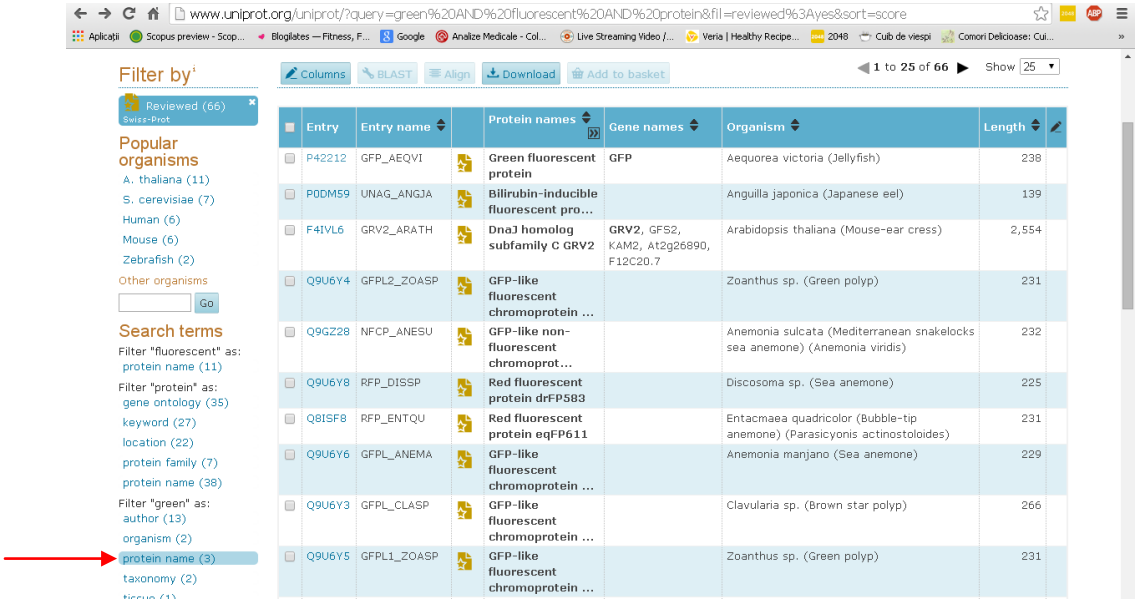
si se scrie in campul de cautare "Green Fluorescent Protein".



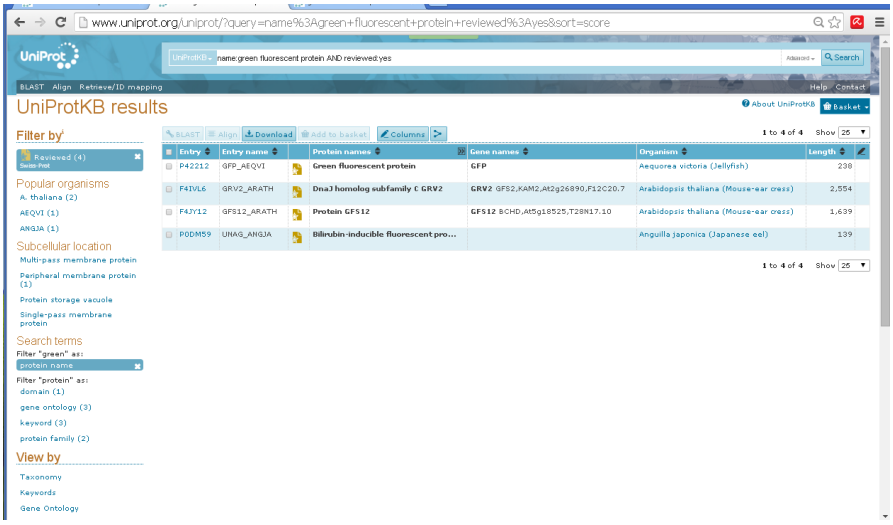
Rafinati cautarea facand click pe "*reviewed*"



Selectati "*protein name*" in sectiunea Filter "green" as



Se obtine:

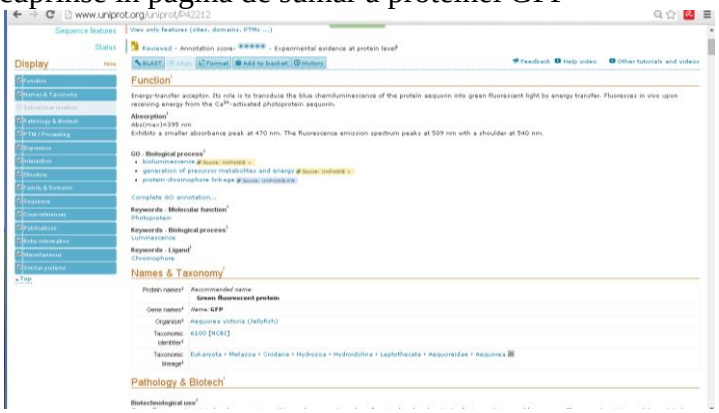


Se selecteaza codul **P42212** (corespunde proteinei GFP extrasa din Aequorea victoria (Jellyfish))

(Codul P42212 se afla si in pagina dedicata proteinei 1EMA in baza de date PDB!).

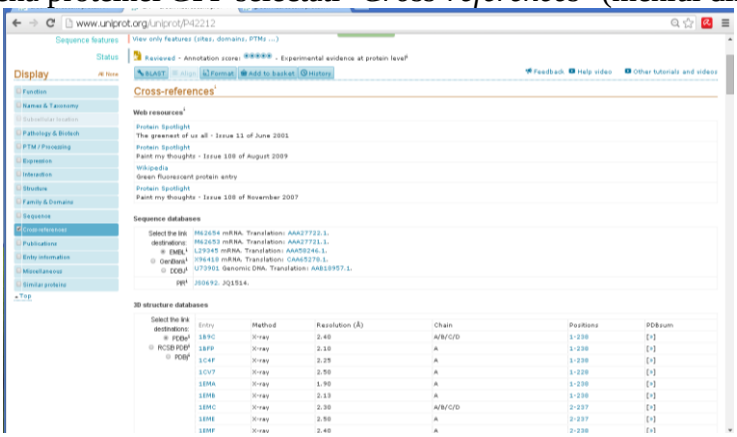
Verificati unde se afla acest cod!!!

Studiati informatiile cuprinse in pagina de sumar a proteinei GFP



Folosind codul "P42212" se poate ajunge direct la aceasta pagina (fara pasii anteriori)!

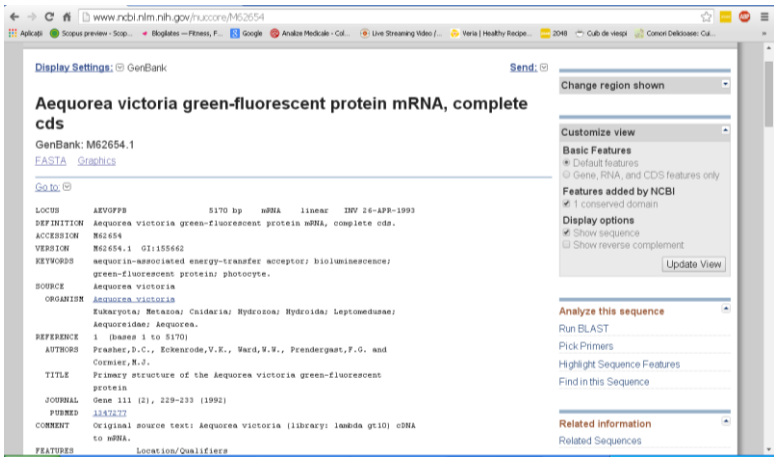
► Pentru a studia gena proteinei GFP selectati "*Cross-references*" (meniul din stanga)



La capitolul “Sequence databases” se selecteaza *GenBank* , apoi se activeaza linkul ”*M62654*”:

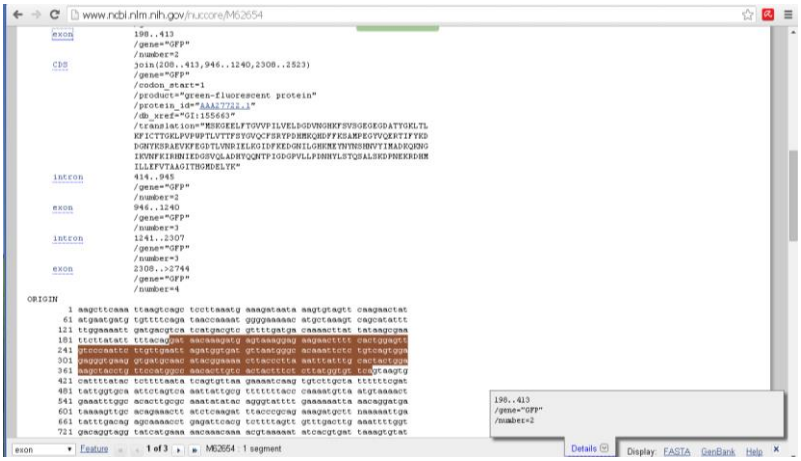
M62654 mRNA. Translation: AAA27722.1.
M62653 mRNA. Translation: AAA27721.1.
L29345 mRNA. Translation: AAA58246.1.
X96418 mRNA. Translation: CAA65278.1.
U73901 Genomic DNA. Translation: AAB18957.1.

Pagina "http://www.ncbi.nlm.nih.gov/nuccore/M62654" afiseaza structura pre-ARNm pentru molecula GFP (ARN mesager precursor - inainte de splicing si translatie)!

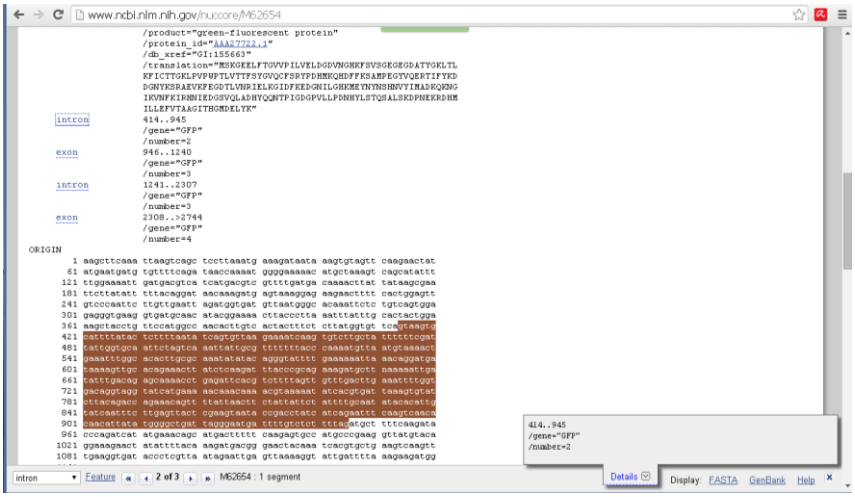


Secventa pre-ARNm obtinuta contine atat zone purtatoare de informatie genetica (exoni) cat si zone nepurtatoare de informatie genetica (introni).

exon:

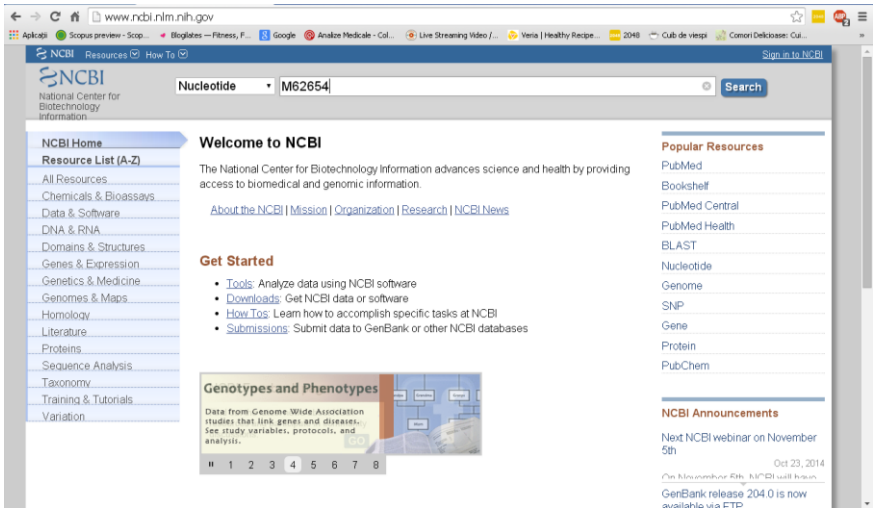


intron:



Se poate ajunge direct la informatia anterioara prin acesarea paginii http://www.ncbi.nlm.nih.gov/

- se selecteaza baza de date "*Nucleotide*"
- se introduce codul "M62654"



Pentru a evidentia secventa ADN corespunzatoare moleculei ARNm ("matrita" ce serveste ca model pentru translatie) se activeaza “mRNA”.

```
FEATURES             Location/Qualifiers
     source          1..5170
                     /organism="Aequorea victoria"
                     /mol_type="mRNA"
                     /db_xref="taxon:6100"
                     /tissue_lib="lambda gt10"
     gene            1..>5170
                     /gene="GFP"
     intron          <1..197
                     /gene="GFP"
                     /number=1
     misc_feature    193..201
                     /gene="GFP"
                     /experiment="experimental evidence, no additional details
                     recorded"
                     /note="fluorescent chromophore"
     mRNA            join(198..413,946..1240,2308..>2744)
                     /gene="GFP"
     exon            198..413
                     /gene="GFP"
                     /number=2
     CDS             join(208..413,946..1240,2308..2523)
                     /gene="GFP"
```

Obs:

- secventele ADN obtinute prin activarea "*mRNA*" difera de secventele ADN obtinute prin activarea "*CDS*" (CoDing Sequence), deoarece secventele CDS contin doar regiuni definite de codonii start si stop.

- secventele "CDS" incep cu gruparea de nucleotide "ATG" si se termina cu una din gruparile "TGA", "TAA" sau "TAG".

Pentru a vedea secventa completa a ADN trebuie vizualizata intreaga pagina.
Copiati cele 3 zone "mRNA" intr-un fisier word pentru a obtine secventa ADN ceruta!

Se obtine un sir de caractere care pe langa literele ce indica bazele, contine si cifre, care nu fac parte din secventa ADN corespunzatoare moleculei ARNm:

gat aacaaagatg agtaaaggag aagaactttt cactggagtt 241 gtcccaattc ttgttgaatt agatggtgat gttaatgggc acaaattctc tgtcagtgga 301 gagggtgaag gtgatgcaac atacggaaaa cttaccctta aatttatttg cactactgga 361 aagctacctg ttccatggcc aacacttgtc actactttct cttatggtgt tca atgct tttcaagata 961 cccagatcat atgaaacagc atgacttttt caagagtgcc atgcccgaag gttatgtaca 1021 ggaaagaact atattttaca aagatgacgg gaactacaaa tcacgtgctg aagtcaagtt 1081 tgaaggtgat accctcgtta atagaattga gttaaaaggt attgatttta aagaagatgg 1141 aaacattctt ggacacaaaa tggaatacaa ctataactca cacaatgtat acatcatggc 1201 agacaaacaa aagaatggaa tcaaagttaa cttcaaaatt aga cacaacattg aagatggaag cgttcaacta 2341 gcagaccatt atcaacaaaa tactccaatt ggcgatggcc ctgtcctttt accagacaac 2401 cattacctgt ccacacaatc tgccctttcc aaagatccca acgaaaagag agatcacatg 2461 atccttcttg agtttgtaac agctgctggg attacacatg gcatggatga actatacaaa 2521 taaatgtcca gacttccaat tgacactaaa gtgtccgaac aattactaaa atctcagggt 2581 tcctggttaa attcaggctg agatattatt tatatattta tagattcatt aaaattttat 2641 gaataattta ttgatgttat taatagggggt tattttctta ttaaataggc tactggagtg 2701 cattcctaat tctatattaa ttacaatttg atttgacttg ctca

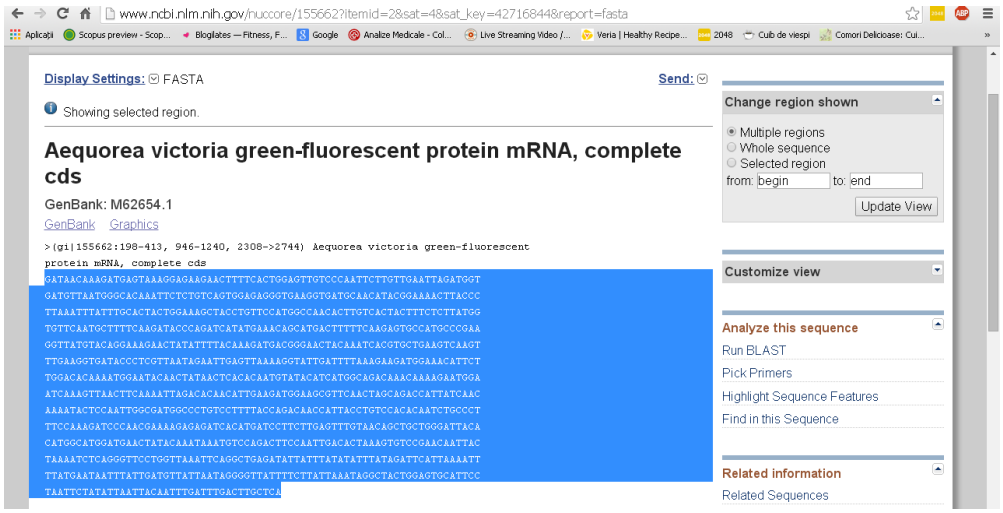
Deci nu asa se obtine ceea ce dorim!!!!!
Daca nu asa atunci cum?

2) Translatia secventei ADN (*obtinerea secventei proteice pornind de la ADN*)

Pentru a obtine secventa ADN corespunzatoare proteinei GFP se selecteaza “FASTA” (dreapta jos sau stanga sus)

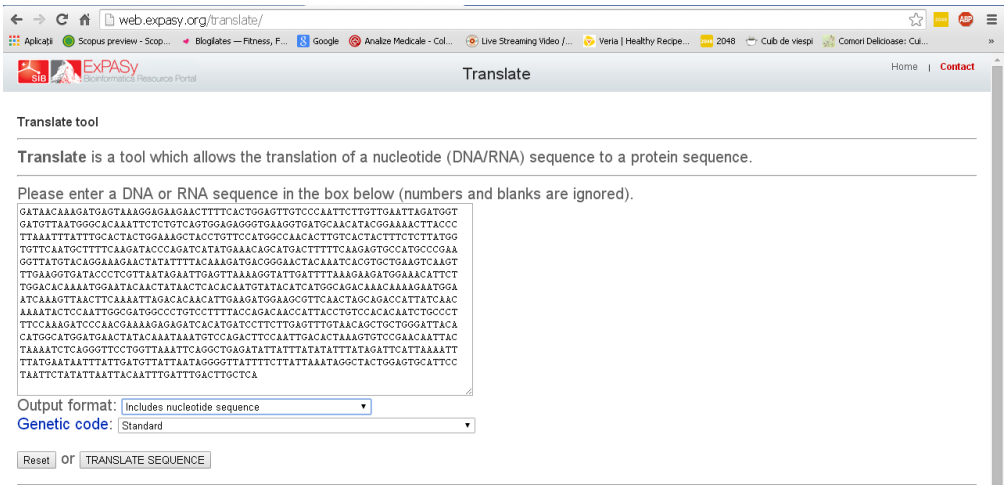
Intr-o fereastra noua (new "Tab") deschideti pagina: http://web.expasy.org/translate/

Selectati secventa ADN (nu linia de comanda!) din pagina obtinuta prin activarea comenzii FASTA si copiati-o: "Copy" sau "Ctrl+c"



Puneti secventa ADN ("*Paste*" sau "Ctrl+v") in pagina: http://web.expasy.org/translate/

In sectiunea "Output format" selectati "*Includes Nucleotide Sequences*"



Click "Translate Sequences"

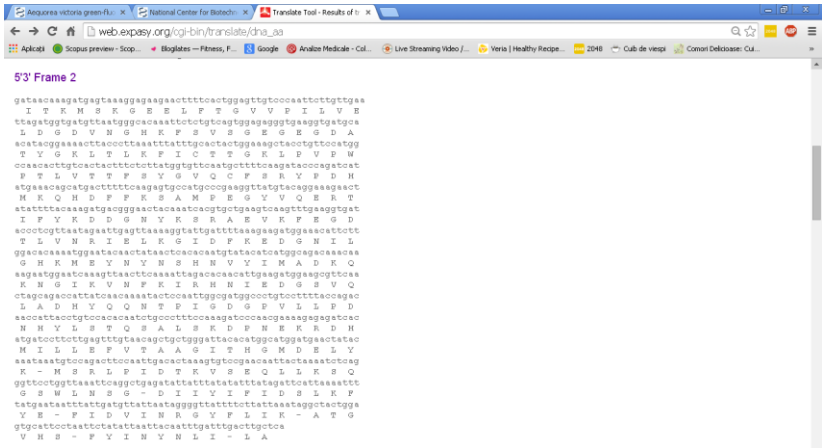
Se obtin ***6 secvente diferite*** ce provin din schimbarea cadrului de citire (incepe cu prima, a doua sau a treia nucleotida) si a directiei de citire (5'3' sau 3'5').

Numai una din aceste secvente este cea buna!

De obicei secventa buna (directia si cadru de citire bun) este cea care produce **cea mai lunga translatie**. (secventa neintrerupta de aminoacizi, fara codoni "Stop" in interiorul secventei proteice)

→ Studiati cele 6 secvente si incercati sa gasit secventa potrivita!

Directia 5'3' cadrul 2: Codonul Start (ATG) este destul de aproape de inceput, iar primul codon Stop (TAG) se afla aproape de sfarsit!

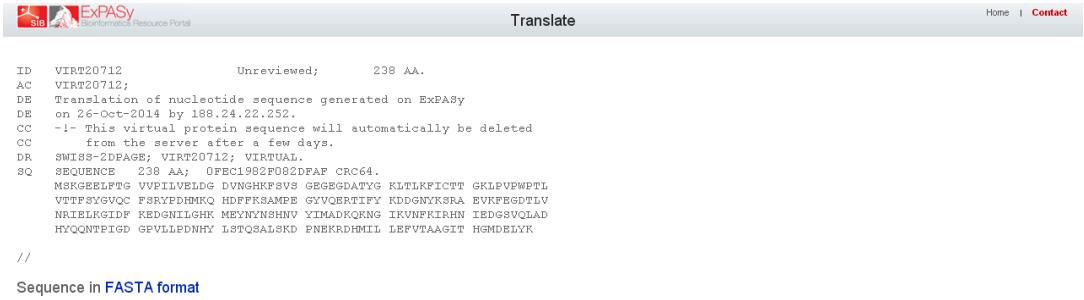


Click pe link-ul "*5'3' Frame 2* "

In pagina obtinuta, este evidentiata metionina (M → codon ATG):

I T K M S K G E E L F T G V V P I L V E L D G D V N G H K F S V S G E G E G D A T Y G K L T L
K F I C T T G K L P V P W P T L V T T F S Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G
Y V Q E R T I F Y K D D G N Y K S R A E V K F E G D T L V N R I E L K G I D F K E D G N I L G
H K M E Y N Y N S H N V Y I M A D K Q K N G I K V N F K I R H N I E D G S V Q L A D H Y Q Q
N T P I G D G P V L L P D N H Y L S T Q S A L S K D P N E K R D H MI L L E F V T A A G I T H
G M D E L Y K **Stop** M S R L P I D T K V S E Q L L K S Q G S W L N S G **Stop** D I I Y I F I D S L K
F Y E **Stop** F I D V I N R G Y F L I K **Stop** A T G V H S **Stop** F Y I N Y N L I **Stop** L A

Click pe primul "M": Se obtine secventa de aminoacizi din care este compusa proteina "Green Fluorescence Protein"

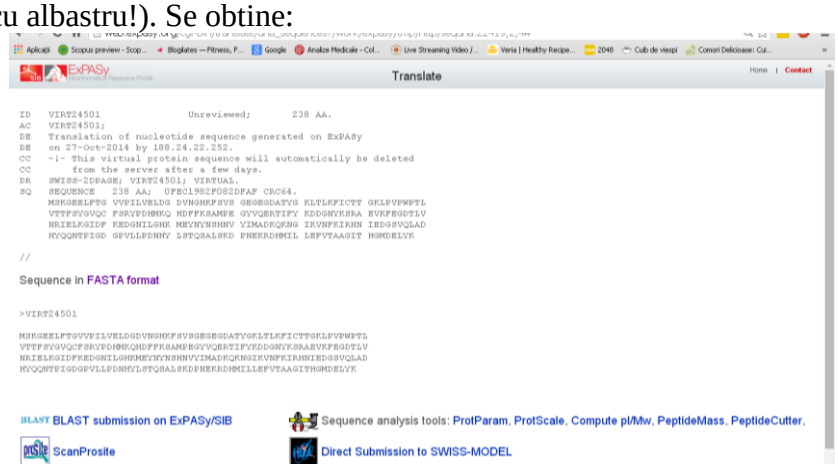


Intr-o fereastra noua (new "Tab") mergeti la pagina proteinei "Green Fluorescence Protein" (1EMA) din baza de date **pdb.org** si activati "*Sequence*". Ar trebui sa obtineti secventa de aminoacizi urmatoare:

Comparati cele 2 secvente de aminoacizi?

3) Compararea secventei de aminoacizi obtinuta cu secvente similare din baza de date PDB

In pagina rezultata dupa translatarea ADN in secventa de aminoacizi se selecteaza "*Fasta Format*" (scris cu albastru!). Se obtine:



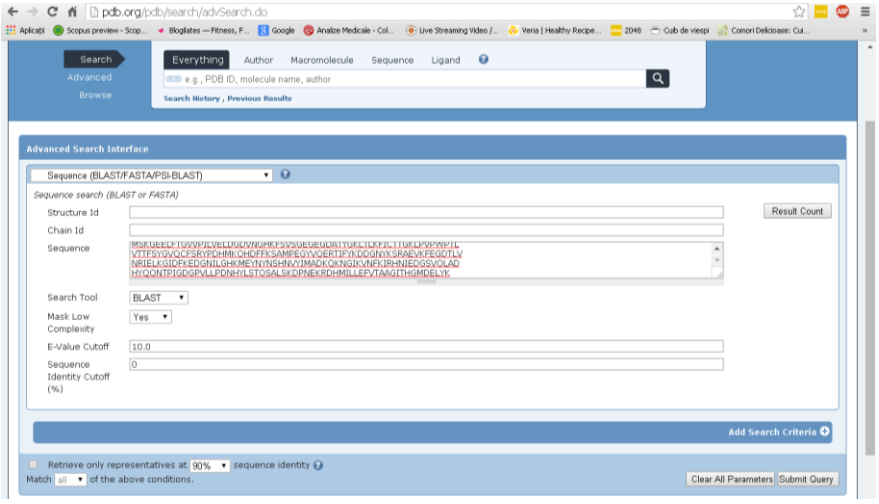
Comparati aceasta secventa de aminoacizi cu secvente proteice din PDB. Daca gasiti secvente similare, inseamna ca un cercetator a rezolvat structura 3D a acestei proteine!

Copiati secventa proteica obtinuta in format Fasta.

Deschideti pagina http://www.pdb.org/ si click "*Advanced Search*"

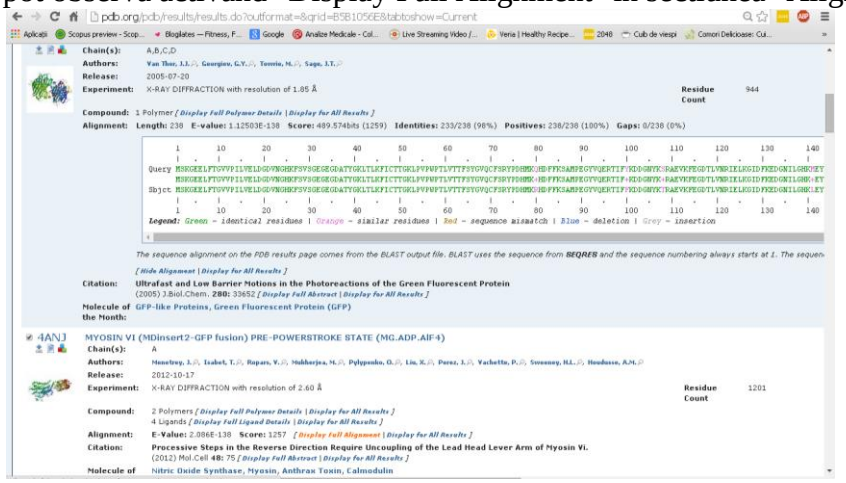
Din meniul "*Chose a Query Type*" selectati "*Sequence (Blast/Fasta)*"

In fereastra "*Sequence*" puneti secventa copiata. Click "*Submit Querry*"



Pagina rezultata contine o lista de proteine a caror secventa de aminoacizi este apropiata de secventa data.

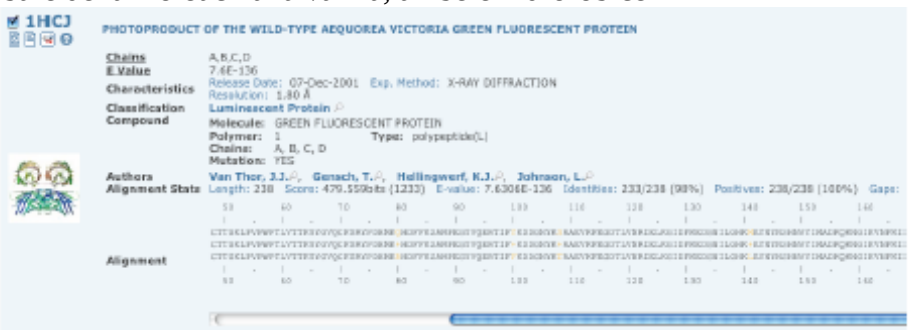
Similaritatile se pot observa activand "Display Full Alignment" in sectiunea "Alignment"



Se obtin proteine similare cu GFP si forme mutante!

→ Studiati proteina “1HCJ”

- numai 5 aminoacizi diferiti
- modificarile sunt conservative: aminoacizii din cele 2 forme sunt similari
- ex: schimbare de la izoleucina la valina, ambele hidrofobice



4) Vizualizarea mutatiilor genetice

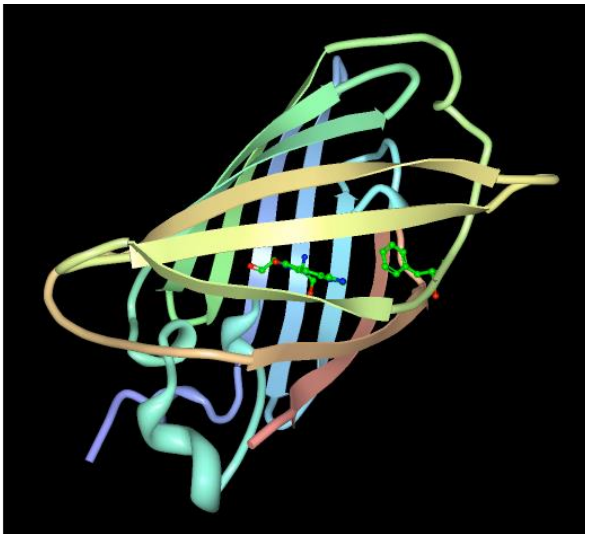
Cautati in pagina de pe site-ul PDB la care ati ajuns in partea precedenta codul **1BFP**.
(o proteina mutanta creata de cercetatori, care are emisia de fluorescenta albastru in loc de verde)
→ Determinati diferenta dintre structurile primareale celor doua proteine.

In tabelul urmator sunt indicate mutatiile care duc la modificarea emisiei de fluorescenta a proteinei GFP

Green Fluorescent	No mutation
Yellow Fluorescent	S65G, S72A, T203F
Cyan Fluorescent	Y66W
Blue Fluorescent	Y66H, Y145F

"S65G" inseamna ca in pozitia 65 serina a fost inlocuita cu glicina
→ Identificati mutatiile Y66H si Y145F (in structural alignment)
Pentru vizualizarea mutatiilor ce apar in 1BFP se va folosi "Protein workshop"

selectati "Visibility"
selectati "Atoms and Bonds"
selectati aminoacidul "145"
selectati aminoacidul "66"



Mutatii

- tirozina (Y) este inlocuita de histidina (H) in pozitia 66 ,
- tirozina (Y) este inlocuita de fenilalanina (F) in pozitia 145

In pozitia 66 era situat cromoforul proteinei "Green fluorescent"
→ Comparati cu proteina 1 EMA!

Tema 8:

1. Evidentiati mutatiile ce apar la proteinele: 1GFL, 1EMG
2. Indicati primele 12 nucleotide ale secventei codante ARNm pentru GFP.
3. Indicati o referinta pentru articolul in care s-a raportat secventa codanta a GFP.
4. Explorati moleculele din pagina interactiva Molecular Machinery poster (http://mm.rcsb.org/). Click pe molecula care de interes activeaza vizualizarea interactive.