

Seminar 8

Evidentierea legaturii dintre gena si secventa proteica corespunzatoare (Identificarea genei GFP si determinarea mutatiilor 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

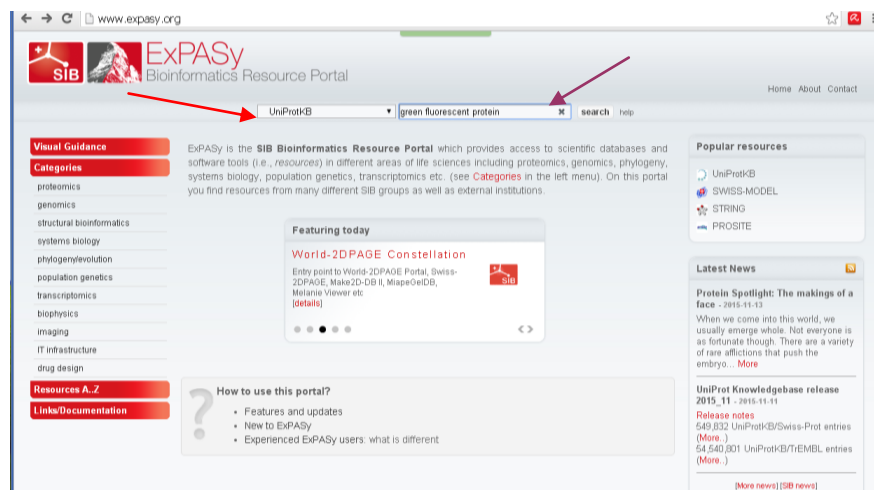
1) Determinarea secventei ADN a proteinei "Green Fluorescent Protein" (in baza de date UniProtKB)

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

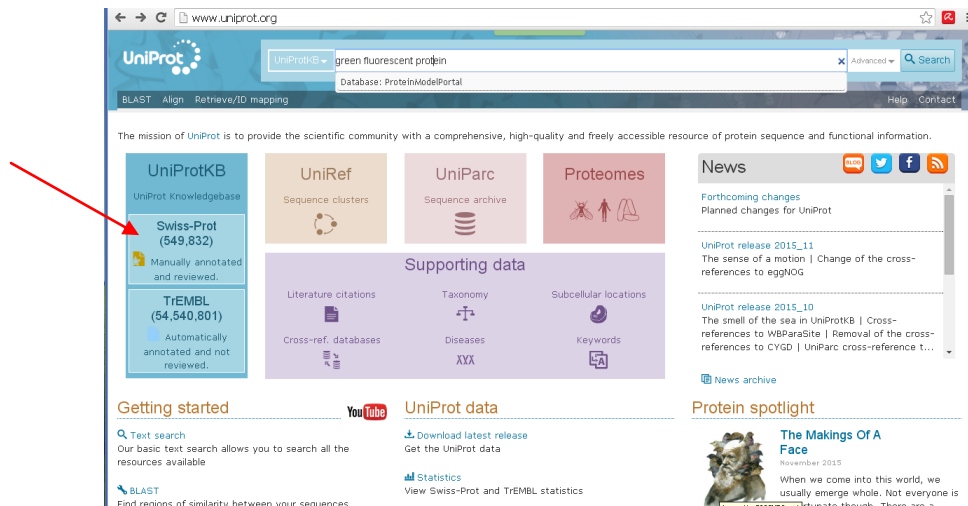
- organizeaza si noteaza secventele proteice.
- contine informatii importante pentru studierea relatiei dintre secventa proteica si functia proteinei.
- foloseste 2 tipuri de notatii:
 - una alocata manual (de cercetatori) bazata pe articolele publicate
 - alta alocata automat de un software sofisticat.

Pentru aceasta activitate practica vom folosi notatii alocate manual!

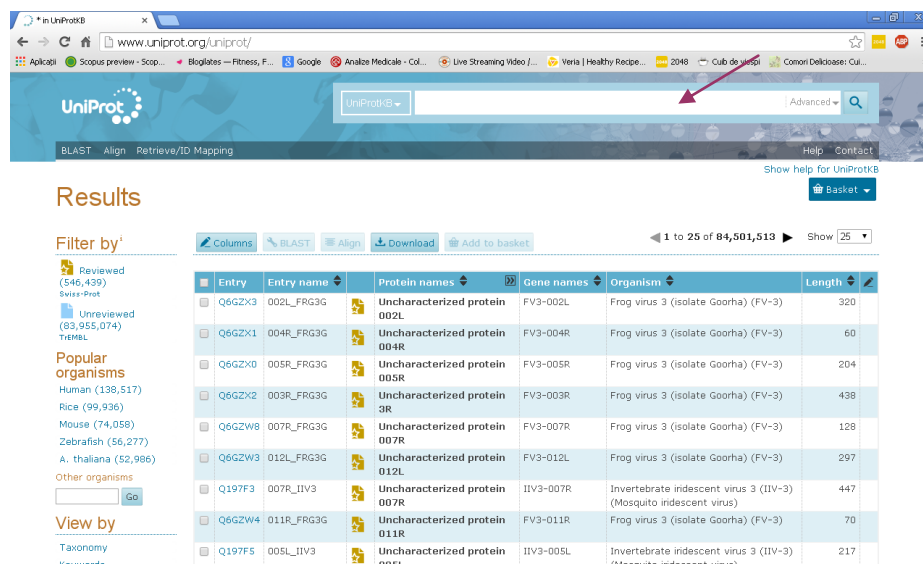
Daca ati folosit adresa "www.expasy.org/" se selecteaza "UniProtKB" in lista bazelor de date si se scrie in campul de cautare "Green Fluorescent Protein".



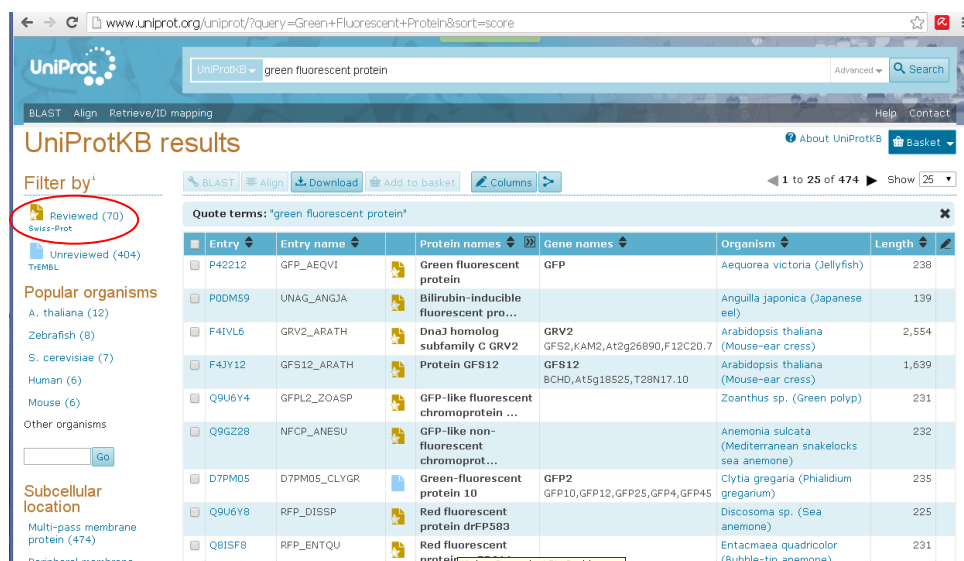
Daca ati folosit adresa "www.uniprot.org/" se selecteaza sectiunea "Swiss-Prot"



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



Rafinati cautarea facand click pe "reviewed"



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

Filter by:

- Reviewed (66)
- Popular organisms
 - A. thaliana (11)
 - S. cerevisiae (7)
 - Human (6)
 - Mouse (6)
 - Zebrafish (2)
- Other organisms

Search terms

Filter "fluorescent" as:

- protein name (11)
- gene ontology (35)
- keyword (27)
- location (22)
- protein family (7)
- protein name (38)

Filter "green" as:

- author (13)
- organism (2)
- protein name (3)
- taxonomy (2)

Entry	Entry name	Protein names	Gene names	Organism	Length
P42212	GFP_AEQVI	Green fluorescent protein	GFP	Aequorea victoria (Jellyfish)	238
P0DM59	UNAG_ANGJA	Bilirubin-inducible fluorescent pro...		Anguilla japonica (Japanese eel)	139
F41VL6	GRV2_ARATH	DnaJ homolog subfamily C GRV2	GRV2, GFS2, KAM2, At2g26890, F12C20.7	Arabidopsis thaliana (Mouse-ear cress)	2,554
Q9U6Y4	GFPL2_ZOASP	GFP-like fluorescent chromoprotein ...		Zoanthus sp. (Green polyp)	231
Q9GZ28	NFCP_ANESU	GFP-like non-fluorescent chromoprot...		Anemonia sulcata (Mediterranean snakelocks sea anemone) (Anemonia viridis)	232
Q9U6Y8	RFP_DISSP	Red fluorescent protein drFP583		Discosoma sp. (Sea anemone)	225
Q8ISF8	RFP_ENTQU	Red fluorescent protein eqFP611		Entacmaea quadricolor (Bubble-tip anemone) (Parasicyonis actinostoloides)	231
Q9U6Y6	GFPL_ANEMA	GFP-like fluorescent chromoprotein ...		Anemonia manjano (Sea anemone)	229
Q9U6Y3	GFPL_CLASP	GFP-like fluorescent chromoprotein ...		Clavularia sp. (Brown star polyp)	266
Q9U6Y5	GFPL1_ZOASP	GFP-like fluorescent chromoprotein ...		Zoanthus sp. (Green polyp)	231

Se obtine:

UniProtKB results

Filter by:

- Reviewed (4)
- Popular organisms
 - A. thaliana (2)
 - AEQVI (1)
 - ANGJA (1)
- Subcellular location
 - Multipass membrane protein
 - Peripheral membrane protein (1)
 - Protein storage vacuole
 - Single-pass membrane protein
- Search terms
 - Filter "green" as:
 - protein name
 - Filter "protein" as:
 - domain (1)
 - gene ontology (3)
 - keyword (3)
 - protein family (2)
- View by
 - Taxonomy
 - Keywords
 - Gene Ontology

Entry	Entry name	Protein names	Gene names	Organism	Length
P42212	GFP_AEQVI	Green fluorescent protein	GFP	Aequorea victoria (Jellyfish)	238
F41VL6	GRV2_ARATH	DnaJ homolog subfamily C GRV2	GRV2, GFS2, KAM2, At2g26890, F12C20.7	Arabidopsis thaliana (Mouse-ear cress)	2,554
F41VL2	GFS12_ARATH	Protein GFS12	GFS12, BCHD, A5G18525, T28N17.10	Arabidopsis thaliana (Mouse-ear cress)	1,639
P0DM59	UNAG_ANGJA	Bilirubin-inducible fluorescent pro...		Anguilla japonica (Japanese eel)	139

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

UniProt entry for P42212 (GFP)

Function:

Energy transfer acceptor. Its role is to transduce the blue chemiluminescence of the protein sequence into green fluorescent light by energy transfer. Fluorescence is only upon receiving energy from the GFP-activated photoexcited sequence.

Absorption:

Absorbs at 295 nm. Exhibits a smaller absorbance peak at 470 nm. The fluorescence emission spectrum peaks at 509 nm with a shoulder at 540 nm.

GO: Biological process:

- fluorescence @basic (link)
- generation of precursor metabolites and energy @basic (link)
- protein-chromophore linkage @basic (link)

Keywords:

- Molecular function: Fluorescence
- Biological process: Fluorescence
- Location: Cytoplasm
- Chromophore

Names & Taxonomy:

Protein names: Recommended name: Green fluorescent protein

Gene names: Name: GFP

Organism: Aequorea victoria (Jellyfish)

Taxonomy: Eukaryota

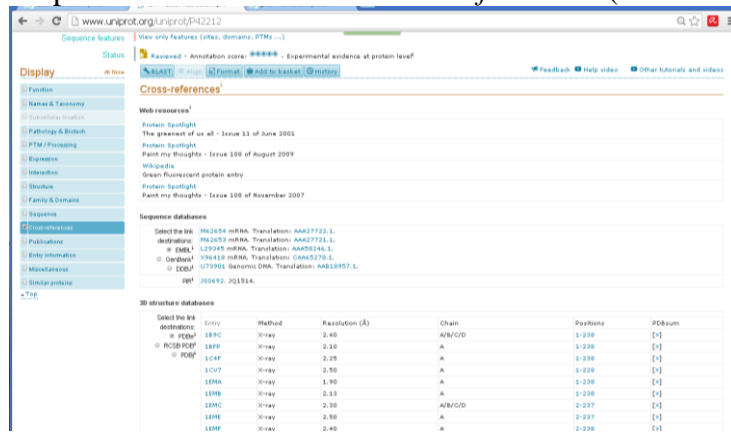
Taxonomic lineage: Eukaryota > Metazoa > Cnidaria > Hydrozoa > Hydrozoa > Leptothecata > Aequorea

Pathology & Biotech:

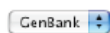
Biotechnological use:

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)

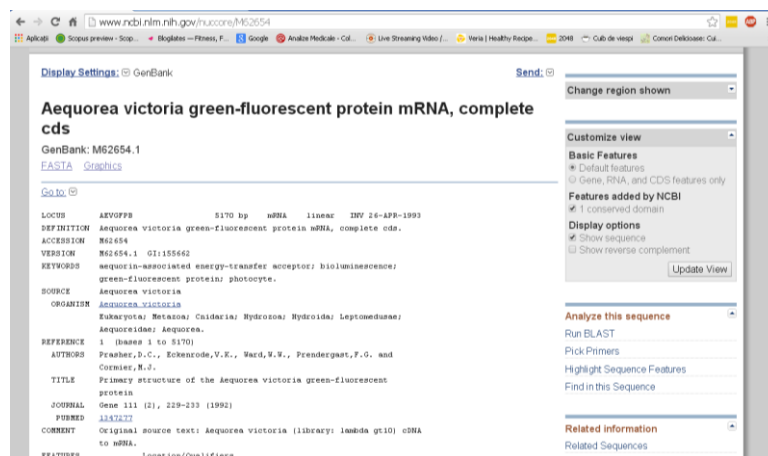


La capitolul "Sequence databases" se selecteaza *GenBank* , apoi se activeaza linkul "M62654":



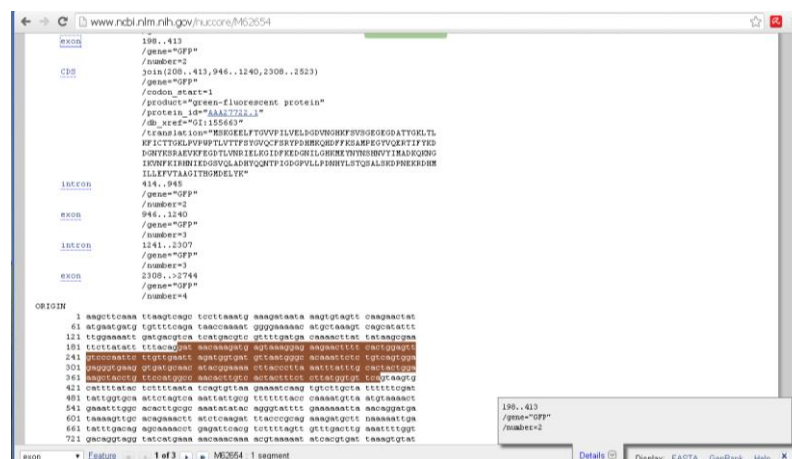
M62654 mRNA. Translation: AAA27722.1.
M62653 mRNA. Translation: AAA27721.1.
L29345 mRNA. Translation: AAA58246.1.
X96418 mRNA. Translation: CAA65278.1.
U73901 Genomic DNA. Translation: AAB18957.1.

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

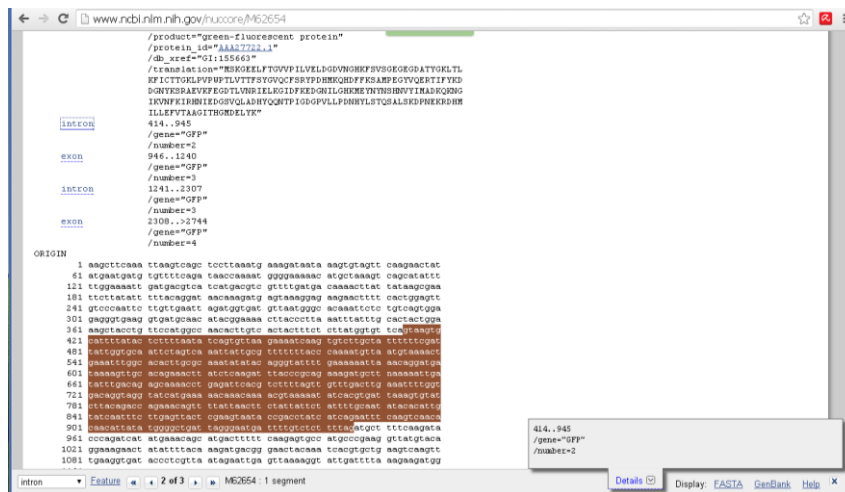


Secventa pre-ARNm obtinuta contine atat zone purtatoare de informatie genetica (exoni) cat si *zone nepurtatoare de informatie genetica (introni)*.

exon:

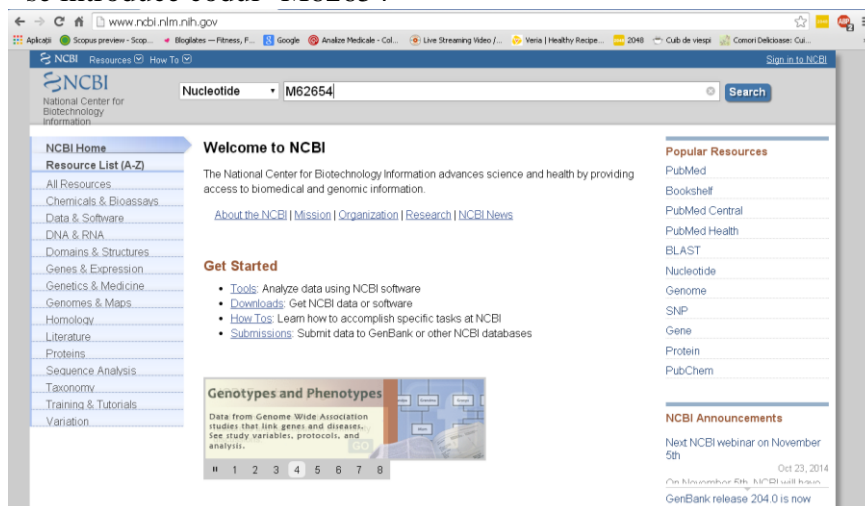


intron:



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 evidenta 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" join(198..413,946..1240,2308..>2744) /gene="GFP"
exon	198..413 /gene="GFP" /number=2 join(208..413,946..1240,2308..2523) /gene="GFP"
CDS	

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.

- secvențele "CDS" încep cu gruparea de nucleotide "ATG" și se termină cu una din grupările "TGA", "TAA" sau "TAG".

Pentru a vedea secvența completă a ADN trebuie vizualizată întreaga pagină.

Copiați cele 3 zone "mRNA" într-un fișier word pentru a obține secvența ADN cerută!

Se obține un sir de caractere care pe lângă literele ce indică bazele, conține și cifre, care nu fac parte din secvența ADN corespunzătoare moleculei ARNm:

```
gat aacaaagatg agtaaaggag aagaactttt cactggagtt 241 gtcccaattc ttgtgaatt agatgggatg gtaatgggc acaattctc
tgtcagtga 301 gaggtgaag gtgatgcaac atacggaaaa ctaccctta aattatttg cactactga 361 aagctacgtg ttccatggcc
aacactgtc actactttct cttatggtgt tca atgct ttcaagata 961 ccagatcat atgaacagc atgactttt caagagtgc
atgcccgaag gttatgtaca 1021 ggaaagaact atattttaca aagatgacgg gaactacaaa tcactgctg aagcaagtt 1081
tgaaggtgat accctcgta atagaattga gttaaagggt attgatttta aagaagatgg 1141 aaacattctt ggacacaaaa tggatacaa
ctataactca cacaatgtat acatcatggc 1201 agacaaacaa aagaatggaa tcaaagttaa ctcaaaatt aga cacaacattg
aagatggaag cgttcaacta 2341 gcagaccatt atcaacaaaa tactccaatt ggcgatggcc ctgtccttt accagacaaac 2401
cattacctgt ccacacaatc tgcctttcc aaagatccca acgaaaagag agatcacatg 2461 atcctctgt agtttgaac agctgctggg
attacacatg gcatggtatg actatacaaa 2521 taaatgtcca gacttccaat tgacactaaa gtgtccgaac aattactaaa atctcagggt
2581 tctggttaa attcaggctg agatattatt tatatattta tagattcatt aaaattttat 2641 gaataattta ttgatgttat taataggggt
tattttctta ttaaattaggc tactggagtg 2701 cattcctaatt tctatattaa ttacaatttg atttgacttg ctca
```

Deci nu așa se obține ceea ce dorim!!!!
Dacă nu așa atunci cum?

2) Traducerea secvenței ADN (obținerea secvenței proteice pornind de la ADN)

Pentru a obține secvența ADN corespunzătoare proteinei GFP se selectează "FASTA" (dreapta jos sau stanga sus)

Într-o fereastră nouă (new "Tab") deschideți pagina: <http://web.expasy.org/translate/>

Selectați secvența ADN (nu linia de comandă!) din pagina obținută prin activarea comenzii FASTA și copiați-o: "Copy" sau "Ctrl+c"

Puneți secvența ADN ("Paste" sau "Ctrl+v") în pagina: <http://web.expasy.org/translate/>

În secțiunea "Output format" selectați "Includes Nucleotide Sequences"

Click "Translate Sequences"

Se obțin **6 secvențe diferite** ce provin din schimbarea cadrului de citire (incepe cu prima, a doua sau a treia nucleotida) și a direcției de citire (5'3' sau 3'5').

Numai una din aceste secvențe este cea bună!

De obicei secvența bună (direcția și cadru de citire bun) este cea care produce **cea mai lungă translație**. (secvența neîntreruptă de aminoacizi, fără codoni "Stop" în interiorul secvenței proteice)

→ **Studiați cele 6 secvențe și încercați să găsiți secvența potrivită!**

Direcția 5'3' cadrul 2: Codonul Start (ATG) este destul de aproape de început, iar primul codon Stop (TAG) se află aproape de sfârșit!

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

În pagina obținută, este evidențiată **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 M I L L E F V T A A G I T H G M D E L Y K S t o p 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 S t o p D I I Y I F I D S L K F Y E S t o p F I D V I N R G Y F L I K S t o p A T G V H S S t o p F Y I N Y N L I S t o p L A

Click pe primul "M": Se obține secvența de aminoacizi din care este compusă proteina "Green Fluorescence Protein"

ExPASy
Bioinformatics Resource Portal

Translate

Home | Contact

ID VIRT20712 Unreviewed; 238 AA.
AC VIRT20712;
DE Translation of nucleotide sequence generated on ExPASy
DE on 26-Oct-2014 by 188.24.22.252.
CC -|- This virtual protein sequence will automatically be deleted
CC from the server after a few days.
DR SWISS-2DPAGE; VIRT20712; VIRTUAL
SQ SEQUENCE 238 AA; 0FECL982F082DFAP CRC64.
MSKGEELFTGVVILVELDGDVNGHFFSVS GEGEGDATYGLTLKFICTTGKLPVPWPTL
VTFFSYGVQC FSRYPDHMK HDFFKSAPE GYVQERTIFY KDDGNYKSA EVKFBGDTLV
NRLEKGIIDF KEDGNILGHK MEYFNYNHNV YIMADKQNG IKVNFKIRHN IEDGSVQLADH
HYQQTFTIGD GEVLLEPDNHY LSTQSALSKD PNEKRDMIL LEFVTAAGIT HGMDELYK

//

Sequence in [FASTA format](#)

Intr-o fereastră nouă (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:

ExPASy
Bioinformatics Resource Portal

Translate

Home | Contact

ID VIRT24501 Unreviewed; 238 AA.
AC VIRT24501;
DE Translation of nucleotide sequence generated on ExPASy
DE on 27-Oct-2014 by 188.24.22.252.
CC -|- This virtual protein sequence will automatically be deleted
CC from the server after a few days.
DR SWISS-2DPAGE; VIRT24501; VIRTUAL
SQ SEQUENCE 238 AA; 0FECL982F082DFAP CRC64.
MSKGEELFTGVVILVELDGDVNGHFFSVS GEGEGDATYGLTLKFICTTGKLPVPWPTL
VTFFSYGVQC FSRYPDHMK HDFFKSAPE GYVQERTIFY KDDGNYKSA EVKFBGDTLV
NRLEKGIIDF KEDGNILGHK MEYFNYNHNV YIMADKQNG IKVNFKIRHN IEDGSVQLADH
HYQQTFTIGD GEVLLEPDNHY LSTQSALSKD PNEKRDMIL LEFVTAAGIT HGMDELYK

//

Sequence in [FASTA format](#)

>VIRT24501

MSKGEELFTGVVILVELDGDVNGHFFSVS GEGEGDATYGLTLKFICTTGKLPVPWPTL
VTFFSYGVQC FSRYPDHMK HDFFKSAPE GYVQERTIFY KDDGNYKSA EVKFBGDTLV
NRLEKGIIDF KEDGNILGHK MEYFNYNHNV YIMADKQNG IKVNFKIRHN IEDGSVQLADH
HYQQTFTIGD GEVLLEPDNHY LSTQSALSKD PNEKRDMIL LEFVTAAGIT HGMDELYK

BLAST BLAST submission on ExPASy/SIB

Sequence analysis tools: ProtParam, ProtScale, Compute pI/Mw, PeptideMass, PeptideCutter.

ScanProsite

Direct Submission to SWISS-MODEL

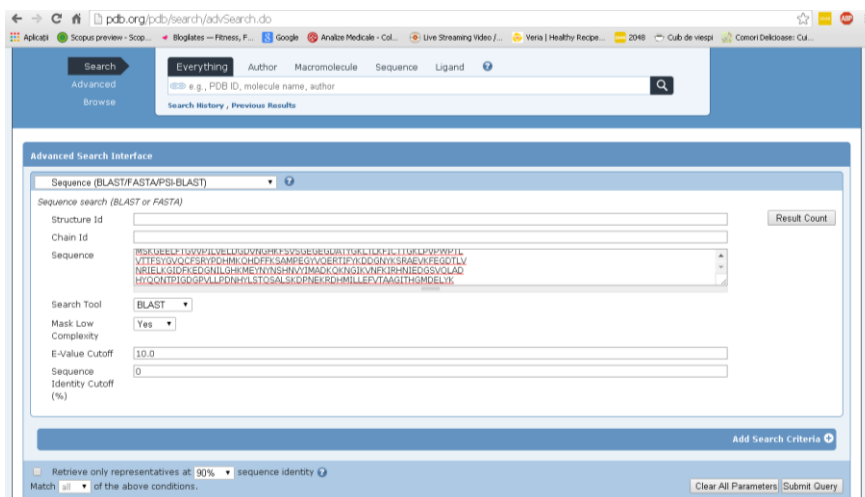
Comparati aceasta secventa de aminoacizi cu secvente proteice din PDB. Daca gasiti secvente similare, inseamna ca un cercetator a rezolvat structura 3D a acestei proteine!

Copiatii 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 fereastră "*Sequence*" puneti secventa copiată. Click "*Submit Query*"



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"

The screenshot shows the PDB results page for the query '1H3J'. The 'Alignment' section is expanded, showing the sequence alignment of the query (1H3J) with the target (1H3J). The alignment is shown as a table with columns for the query and target sequences. The alignment is as follows:

Query	Target
MSKLEELTGVVPLVLLDQWVWPKFVSUSEGURITURKLUKJLITLRLPVPWPL	MSKLEELTGVVPLVLLDQWVWPKFVSUSEGURITURKLUKJLITLRLPVPWPL
VTFTSDVQCFSEPCMRCHQFESAMEGIVQERTIPNLDQNTSRANKFEQGLTA	VTFTSDVQCFSEPCMRCHQFESAMEGIVQERTIPNLDQNTSRANKFEQGLTA
NRIELSDPEEDGNLGHMEYNHSHVYIMAAKNGKLVNFIHNIJEDGSSVLAD	NRIELSDPEEDGNLGHMEYNHSHVYIMAAKNGKLVNFIHNIJEDGSSVLAD
HTQNTPTGSGVLLPDNHLSTGSLKDFNEKRDHMLLEPVTAGTHGMDELTK	HTQNTPTGSGVLLPDNHLSTGSLKDFNEKRDHMLLEPVTAGTHGMDELTK

The alignment is 100% identical. The 'Display Full Alignment' button is visible at the bottom of the alignment section.

Se obtin proteine similare cu GFP si forme mutante!

→ Studiat proteina "1H3J"

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

The screenshot shows the PDB entry page for 1H3J. The 'Alignment' section is expanded, showing the sequence alignment of the query (1H3J) with the target (1H3J). The alignment is shown as a table with columns for the query and target sequences. The alignment is as follows:

Query	Target
MSKLEELTGVVPLVLLDQWVWPKFVSUSEGURITURKLUKJLITLRLPVPWPL	MSKLEELTGVVPLVLLDQWVWPKFVSUSEGURITURKLUKJLITLRLPVPWPL
VTFTSDVQCFSEPCMRCHQFESAMEGIVQERTIPNLDQNTSRANKFEQGLTA	VTFTSDVQCFSEPCMRCHQFESAMEGIVQERTIPNLDQNTSRANKFEQGLTA
NRIELSDPEEDGNLGHMEYNHSHVYIMAAKNGKLVNFIHNIJEDGSSVLAD	NRIELSDPEEDGNLGHMEYNHSHVYIMAAKNGKLVNFIHNIJEDGSSVLAD
HTQNTPTGSGVLLPDNHLSTGSLKDFNEKRDHMLLEPVTAGTHGMDELTK	HTQNTPTGSGVLLPDNHLSTGSLKDFNEKRDHMLLEPVTAGTHGMDELTK

The alignment is 100% identical. The 'Display Full Alignment' button is visible at the bottom of the alignment section.

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"

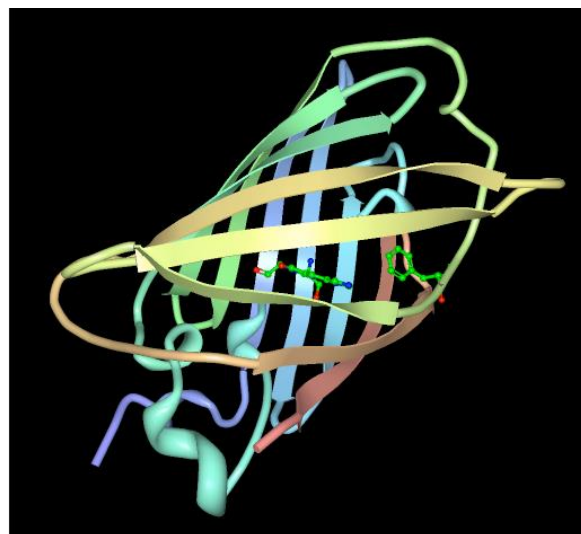
selectati aminoacidul "66"

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.