



LIBIPMET



## Práctica 2 | SOFTWARE DE ANÁLISIS DE SECUENCIAS Y ESTRUCTURAS

Docente: Dr. Jesus Alvarado

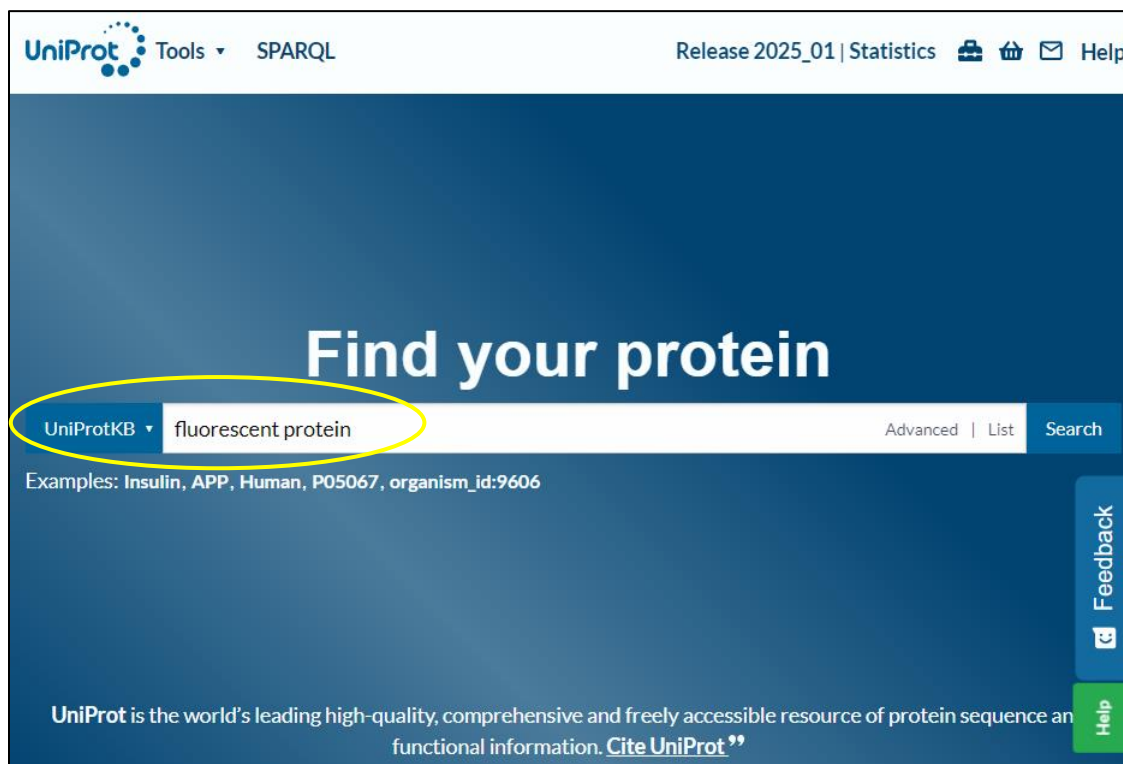
---

En esta práctica dirigida empleamos UniProt, Clustal y PyMOL para el análisis de secuencias y estructuras moleculares. De acuerdo con las siguientes instrucciones:

1. Buscar en UniProt proteínas fluorescentes (al menos 06 proteínas distintas) y descarga sus secuencias para alinearlas.

Ingresa a : <https://www.uniprot.org/>

En el campo señalado (en amarillo) en la siguiente imagen, ingresa “fluorescent protein” (sin comillas) y luego “Search”.

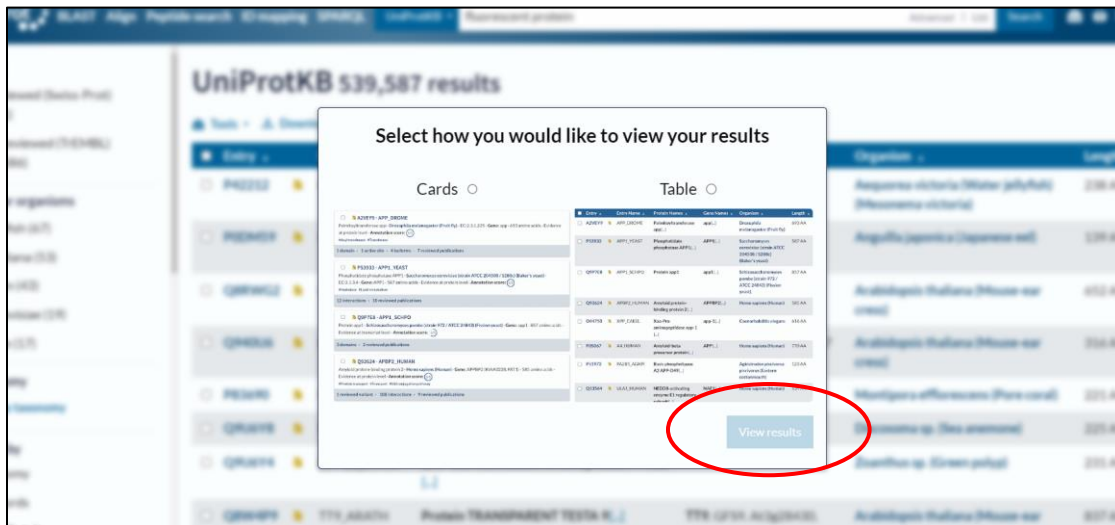




LIBIPMET



Selecciona Tabla y luego “ver resultados”:



¿Cuántos resultados se obtuvieron? Puedes seleccionar diferentes filtros (línea morada), por ejemplo, por organismo, taxonomía, estructura 3D, etc. Seleccionaremos GFP (línea verde).

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
P42212	GFP_AEQVI	Green fluorescent protein	GFP	Aequorea victoria (Water jellyfish) (Mesonema victoria)	238 AA
P0DM59	UNAG_ANGJA	Bilirubin-inducible fluorescent protein UnaG[...]		Anguilla japonica (Japanese eel)	139 AA
Q8RWG2	CF107_ARATH	Protein high chlorophyll fluorescent 107	HCF107, MBB1, At3g17040, K14A17.11	Arabidopsis thaliana (Mouse-ear cress)	652 AA
Q940U6	FLU_ARATH	Protein FLUORESCENT IN BLUE LIGHT, chloroplastic	FLU, At3g14110, MAG2.7	Arabidopsis thaliana (Mouse-ear cress)	316 AA
P83690	NFCP_MONEF	GFP-like non-fluorescent chromoprotein[...]		Montipora efflorescens (Pore coral)	221 AA
Q9U6Y8	RFP_DISSP	Red fluorescent protein drFP583[...]		Discosoma sp. (Sea anemone)	225 AA
Q9U6Y4	GFPL2_ZOASP	GFP-like fluorescent chromoprotein FP538 [...]		Zoanthus sp. (Green polyp)	231 AA
Q8W4P9	TT9_ARATH	Protein TRANSPARENT TESTA 9[...]	TT9, GFS9, At3g28430, MFJ20.13	Arabidopsis thaliana (Mouse-ear cress)	837 AA
Q81SER	RFP_ENTOU	Red fluorescent protein eoFP611[...]		Entacmaea quadricolor (Bubble-tin)	231 AA

Seleccionamos “Sequence” (línea celeste):



**UniProt** BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced

**P42212 · GFP\_AEQVI**

Protein<sup>i</sup> | Green fluorescent protein | Amino acids | 238 (go to sequence)

Gene<sup>i</sup> | GFP | Protein existence<sup>1</sup> | Evidence at protein level

Status<sup>i</sup> | UniProtKB reviewed (Swiss-Prot) | Annotation score<sup>2</sup> | 5/5

Organism<sup>i</sup> | *Aequorea victoria* (Water jellyfish) (*Mesonema victoria*)

Entry Variant viewer<sup>3</sup> Feature viewer Genomic coordinates Publications External links History

Tools<sup>4</sup> Download<sup>5</sup> Add<sup>6</sup> Add a publication<sup>7</sup> Entry feedback<sup>8</sup>

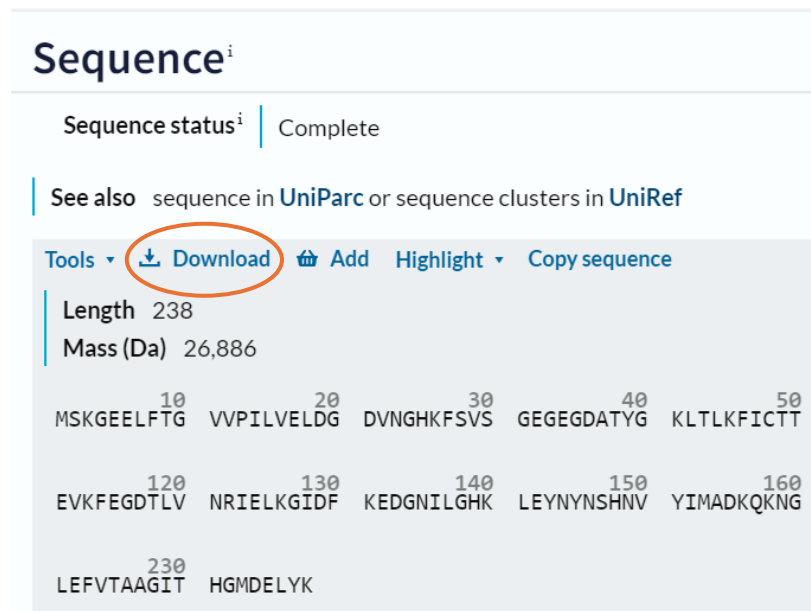
**Function<sup>1</sup>**

Energy-transfer acceptor. Its role is to transduce the blue chemiluminescence of the protein aequorin into green fluorescent light by *in vivo* upon receiving energy from the  $\text{Ca}^{2+}$ -activated photoprotein aequorin.

**Biotechnology<sup>1</sup>**

Green fluorescent protein has been engineered to produce a vast number of variously colored mutants, fusion proteins, and biosensors can be mutated to emit at different wavelengths such as blue for BFP (when Tyr-66 is replaced by His), cyan for CFP (when Tyr-66 is replaced by Ser).

Procedemos a descargar la secuencia (línea anaranjada):



**Sequence<sup>1</sup>**

Sequence status<sup>1</sup> | Complete

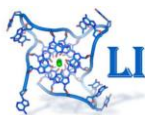
See also sequence in UniParc or sequence clusters in UniRef

Tools<sup>2</sup> Download<sup>3</sup> Add<sup>4</sup> Highlight<sup>5</sup> Copy sequence<sup>6</sup>

Length 238  
Mass (Da) 26,886

MSKGEELFTG<sup>10</sup> VVPILVELDG<sup>20</sup> DVNGHKFSVS<sup>30</sup> GEGEGDATYG<sup>40</sup> KLTLKFICTT<sup>50</sup>  
EVKFEGDTLV<sup>120</sup> NRIELKGIDF<sup>130</sup> KEDGNILGHK<sup>140</sup> LEYNYNSHNV<sup>150</sup> YIMADKQKNG<sup>160</sup>  
LEFVTAAGIT<sup>230</sup> HGMDELYK

3. Realizar el mismo alineamiento usando la herramienta online de UniProt.

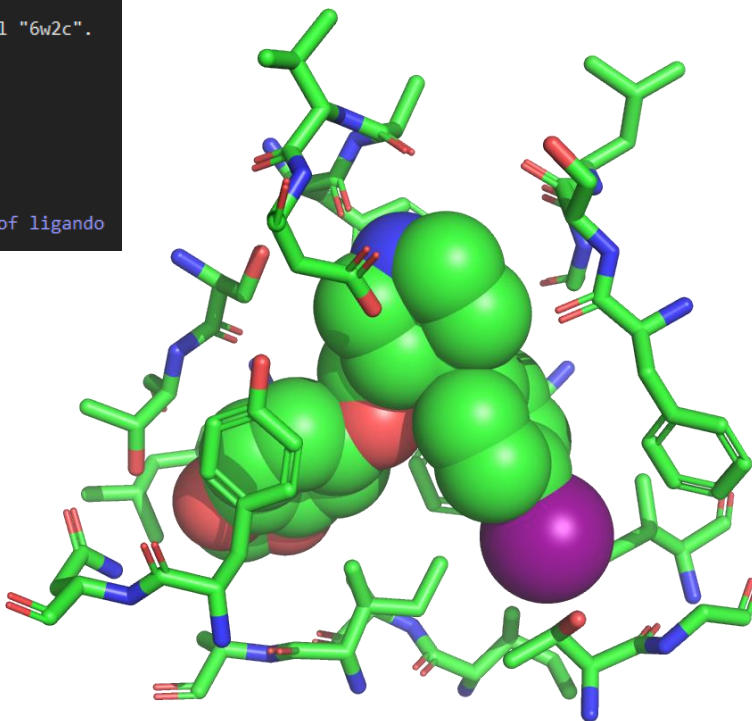


4. Disponer de Pymol (ver sugerencias de instalación al final de la guía) en la PC e identificar las siguientes operaciones básicas. Anotar los comandos indicados por el docente en la siguiente tabla.

Operación	Comando
Descargar la proteína 6W2C y abrirla en PyMOL	
Cambiar el fondo a blanco	
Seleccionar las cadenas B y C, eliminarlas	
Centrar la visualización de la cadena A	
Seleccionar al ligando, mostrarlo como esferas	
Crear un objeto llamado “ligando”	
Colorear la proteína de blanco y el ligando en rojo	
Mostrar el ligando en rojo rodeado de los aminoácidos en un radio de 5 Å.	

En el caso de ligando-AND puedes usar (recuerda previamente haber creado el objeto “ligando”):  
select around\_ligand, byres (all within 5 of ligand)  
show sticks, around\_ligand

```
PyMOL>fetch 6w2c
TITLE      Anomalous iodine signal reveals the posi
ExecutiveLoad-Detail: Detected mmCIF
CmdLoad: ".\6w2c.cif" loaded as "6w2c".
PyMOL>remove chain B chain C
Remove: eliminated 3316 atoms in model "6w2c".
PyMOL>show spheres, organic
PyMOL>zoom
PyMOL>create ligando, organic
Selector: found 24 atoms.
Executive: object "ligando" created.
PyMOL>zoom
PyMOL>hide cartoon
PyMOL>show sticks, byres all within 5 of ligando
```





LIBIPMET



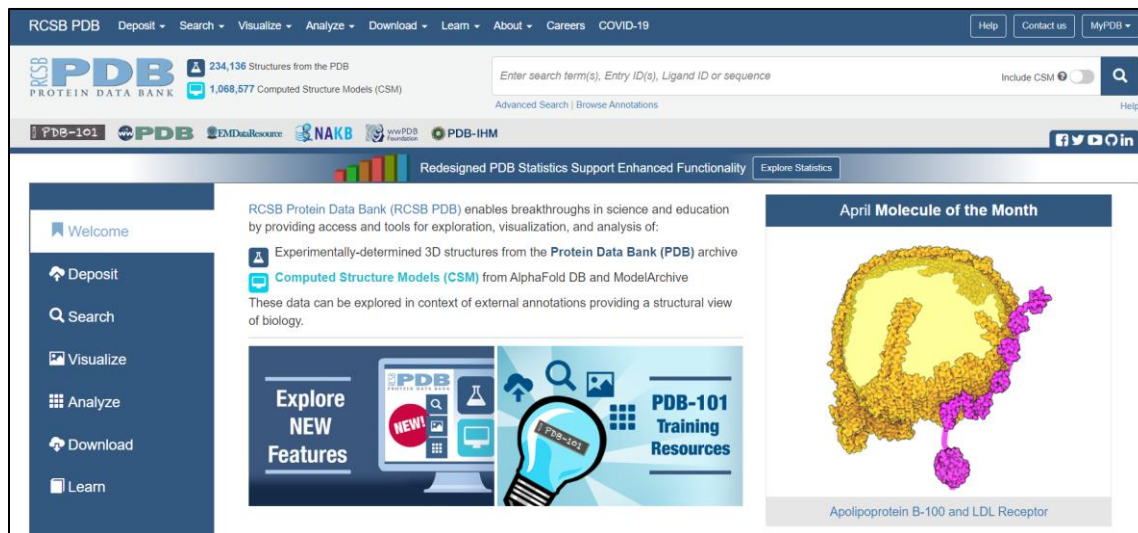
5. Desde Pymol descargar las siguientes estructuras de la CRP de *Escherichia coli* en sus tres estados:

5.1 Proteína sola (PDB ID: 3HIF)

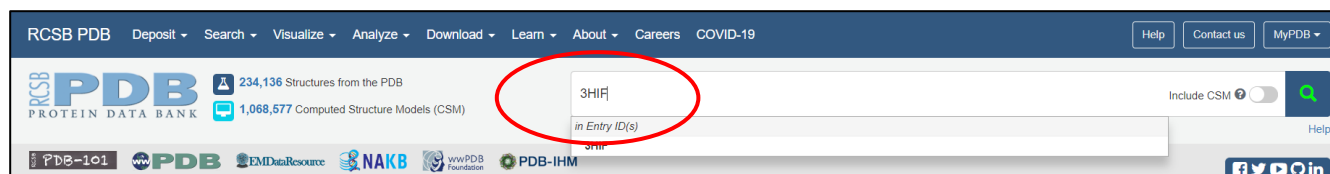
5.2 En complejo con cAMP (PDB ID: 1I5Z)

5.3 En complejo con caMP y ADN (PDB ID: 1ZRD)

Para ello ingresa a: <https://www.rcsb.org/>



Ingresa el código PDB en el campo disponible, como se visualiza a continuación y luego haz clic en la lupa:



Clic en “Download Files” y selecciona PDB Format para obtener la estructura.





6. Analizar similitudes y diferencias de las 03 estructuras descargadas. Puedes emplear el comando *align* para superponer las estructuras.

7. ¿Qué cambios le haría a la proteína si quiere inhibir la unión al ADN?

8. Desafío: “Encaje” una nueva pequeña molécula orgánica en lugar de los ligandos co-cristalizados.

## Referencias:

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000;28(1):235–42.
2. Meyer EE. The first years of the Protein Data Bank. *Protein Sci.* 1997 Jul;6(7):1591–7.
3. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015 Jan 28;43(D1):D204–12.
4. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Res.* 2014;42(D1):1091–7.
5. Hu T, Sprague ER, Fodor M, Stams T, Clark KL, Cowan-Jacob SW. The impact of structural biology in medicine illustrated with four case studies. *J Mol Med.* 2018 Jan;96(1):9–19.
6. Comandos básicos para la representación molecular con PyMOL desde la terminal:
  - [https://williams.chemistry.gatech.edu/course\\_Information/py\\_script\\_coords/pymol\\_commands.txt](https://williams.chemistry.gatech.edu/course_Information/py_script_coords/pymol_commands.txt)
  - <https://medium.com/@snippetsbio/visualizing-protein-protein-docking-using-pymol-cc49494e54bb>
  - [https://pymolwiki.org/index.php/Selection\\_Algebra](https://pymolwiki.org/index.php/Selection_Algebra)

## Recomendaciones para la instalación de PyMOL:

- A. Descargue el ejecutable en: <https://www.pymol.org/>
- B. Ejecute el archivo EXE descargado, no requiere permisos de administrador.
- C. En cuanto a formatos de archivos, puede continuar con lo recomendado por defecto.

