



Bioinformática UNQ

Trabajo Práctico N° 7: Estructura de proteínas

Objetivo: El objetivo de este TP es familiarizarnos con distintas herramientas para estudiar proteínas cristalizadas usando programas de visualización molecular y análisis de datos.

Muchos de estos ejercicios se llevan a cabo usando el programa **Pymol**. Este programa sirve para visualizar estructuras de macromoléculas, como proteínas y ácidos nucleicos. Pymol es uno de los programas más utilizados en la comunidad científico-académica para manipulación de estructuras y, por lo tanto cuenta con una basta documentación de uso y tutoriales en su [wiki](https://pymolwiki.org/): <https://pymolwiki.org/>

¿Como instalar Pymol?

Para su [instalación](#) debemos ir a la página oficial del [Software](#), y hacer click en el ícono que se corresponda con el sistema operativo de su máquina.

- En el caso de **windows** se puede obtener una licencia educativa registrándose en el siguiente enlace: <https://pymol.org/edu/?q=educational/>. Luego la descarga e instalación es igual a la de cualquier otro programa de windows.
- En el caso de **linux** (Distribuciones basadas en Ubuntu) se puede instalar la última versión a partir del repositorio mediante el siguiente comando: `sudo apt install pymol`

El programa es gratuito para uso académico, pero para otras versiones más actualizadas que requieren una [licencia](#) se pueden descargar en forma gratuita de <https://pymol.org/2/> e instalar tanto en PC, Mac como linux.

Uso del Programa

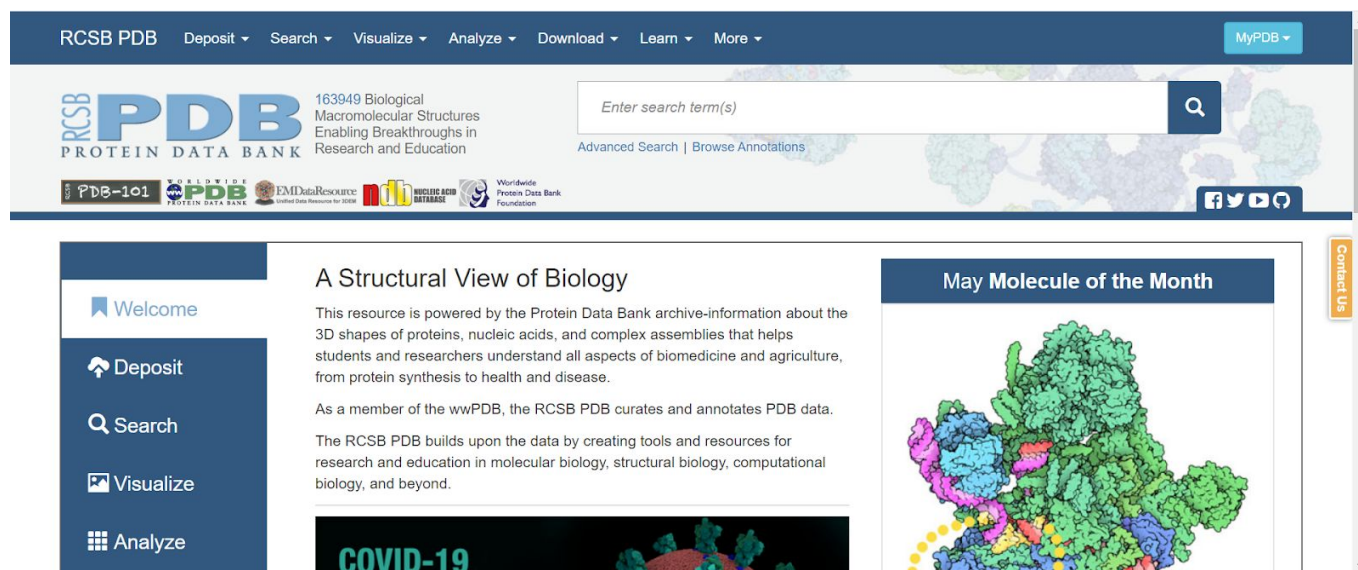
Como *input* el programa usa archivos que contienen las coordenadas xyz de los átomos que componen la macromolécula (en este caso de cada aminoácido de la proteína en estudio). Un archivo que contenga esta información se ve de la siguiente manera:

ATOM	1	N	MET	A	1	24.887	99.728	-28.491	1.00	24.53	N
ATOM	2	CA	MET	A	1	24.678	98.599	-27.540	1.00	24.09	C
ATOM	3	C	MET	A	1	25.436	98.859	-26.235	1.00	20.23	C
ATOM	4	O	MET	A	1	26.450	99.538	-26.234	1.00	21.53	O
ATOM	5	CB	MET	A	1	25.248	97.307	-28.116	1.00	25.14	C
ATOM	6	CG	MET	A	1	24.269	96.455	-28.908	1.00	37.63	C
ATOM	7	SD	MET	A	1	25.198	95.089	-29.651	1.00	38.25	S
ATOM	8	CE	MET	A	1	24.748	95.349	-31.373	1.00	36.67	C
ATOM	9	N	LYS	A	2	24.905	98.265	-25.144	1.00	22.07	N
ATOM	10	CA	LYS	A	2	25.574	98.441	-23.875	1.00	17.60	C
ATOM	11	C	LYS	A	2	26.691	97.384	-23.824	1.00	21.48	C
ATOM	12	O	LYS	A	2	26.699	96.508	-24.678	1.00	23.47	O
ATOM	13	CB	LYS	A	2	24.649	98.288	-22.672	1.00	21.64	C

Las proteínas con estructura conocida se pueden descargar de una base de datos que se denomina PDB.



Bioinformática UNQ



El uso del **Pymol** es bastante intuitivo. Los archivos se abren usando el menú *File*, o alternatively utilizando el comando 'fetch' seguido del código correspondiente a la proteína en la base de datos del protein data bank PDB. Los códigos PDB están constituidos en general por 4 caracteres alfanuméricos (ej. 1LXA). El programa Pymol posee en la barra lateral derecha, panel superior (ver figura), las opciones más importantes en los botones **A S H L C**, correspondientes a **Action, Show, Hide, Label, Color**. Ahí podemos encontrar distintas formas para ver los átomos, la macromolécula (DNA o proteínas) y en algunos casos superficies. En las distintas prácticas se destacan algunos comandos especiales requeridos para un objetivo determinado.

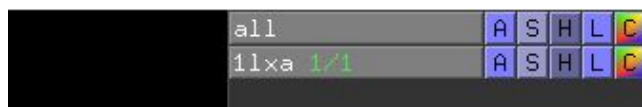
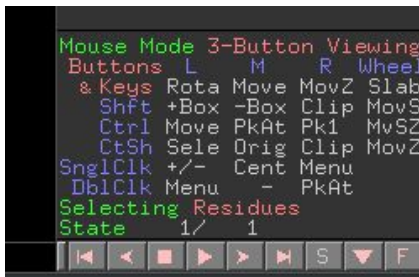


Fig. Interfaz del programa Pymol.
=< Barra lateral derecha panel superior

Bioinformática UNQ

Captura de pantalla de la Interfaz completa



En la barra lateral derecha, panel inferior, se pueden seleccionar diferentes modalidades de control de la visualización. En este panel nos interesa principalmente el botón S, que nos va a permitir abrir la barra de secuencia, que aparecerá en la parte superior. La barra de secuencia permite seleccionar individual o grupalmente diferentes aminoácidos. El botón F es para visualizar en la pantalla completa. El programa Pymol permite realizar, además, muchas acciones mediante comandos ingresados por línea de comandos.

Actividades

1. ¿Cómo describirías la estructura de esta proteína?

Las estructuras de las proteínas pueden ser muy diversas. Se las puede clasificar por su contenido en estructuras secundarias y según su cantidad de dominios. ¿Cómo podrías describir la estructura de esta proteína? Realizá la misma descripción pero para la proteína 1THJ, 3OGB,

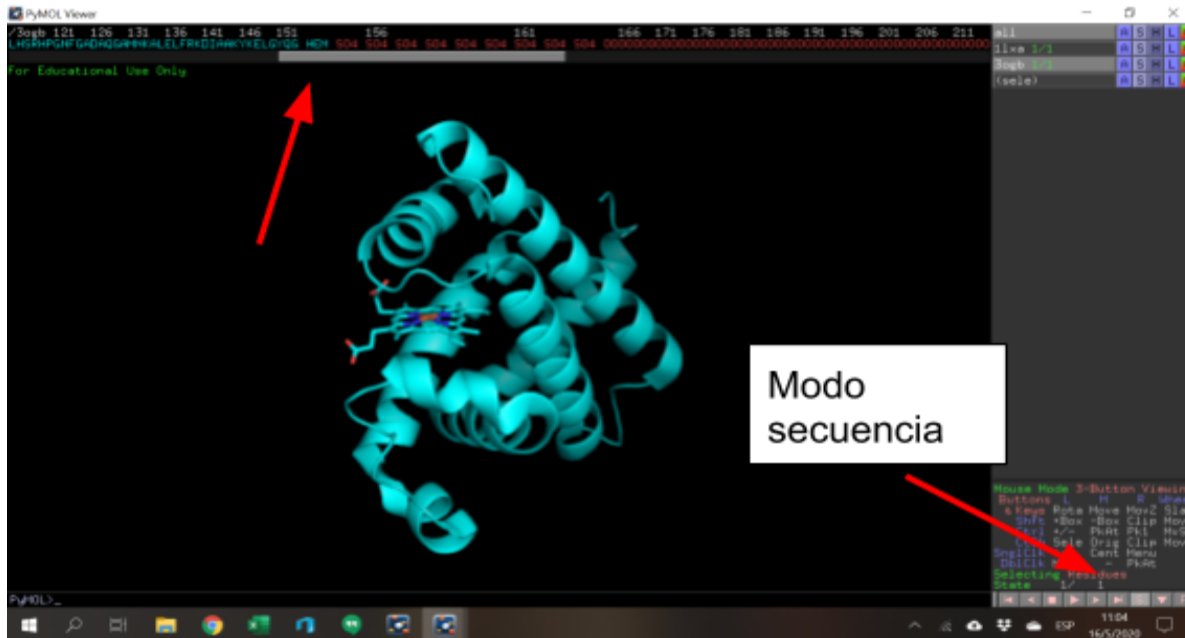
2. Proteínas con estructura cuaternaria

Utilizando la estructura 1THJ vamos a usar un comando para ver las distintas cadenas que la componen. En la línea de comandos, escribí “split_chains”. Inmediatamente verás en el panel que la estructura 1THJ se ha subdividido en sus cadenas que la componen (A, B y C usualmente). Este comando también es útil para separar las cadenas cuando la estructura derive de un estudio NMR.

Bioinformática UNQ

3. Hetero átomos en las estructuras

Utilizando la proteína 3OGB activá el panel “Sequence” y seleccione el grupo “HEM”



Una vez seleccionado el grupo Hemo, visualizá el mismo en formato “lines” o “stick”. Esta forma nos permite visualizar grupos que no están formados por aminoácidos.

4. Un mar de conformaciones...

Estudiá la proteína 2CPE. ¿Qué tipo de proteína es? ¿Cómo la describirías? Utilizá el comando “set all_states, on” para ver todos los estados conformacionales estimados para esa estructura. Compará el espacio conformacional de la 2CPE con la de la estructura de la mioglobina (1MYF). Utilizando el modo secuencia, seleccioná la HIS 64. Mostrala en formato stick o lines. ¿Qué función podría llegar a tener? ¿Y la HIS 93?

RETO I: Estas estructuras difieren de las estructuras sobre las que venimos trabajando en su determinación. Como habrás notado estas fueron obtenidas mediante la técnica de MNR, ¿Pero en qué consiste esta técnica?

Te dejamos un video para que indagues más sobre esta técnica y sus aplicaciones:
<https://www.youtube.com/watch?v=H-SQFSynKOk>

5. Cálculo de la distancia promedio de un puente de hidrógeno.

Como hemos visto las distintas propiedades de los aminoácidos y su orden determinarán la estructura de la proteína toda. Mediante interacciones de sus residuos ésta irá adoptando un plegamiento en el espacio.



Bioinformática UNQ

RETO II: Investigá en qué consisten las interacciones puentes de hidrógeno, π - π y π -catión y qué aminoácidos podrían intervenir en dichas interacciones.

Nos enfocaremos ahora en un tipo de interacción, los puentes de hidrógeno, y para ello se usará la proteína UDP-N-acetiltransferasa de *Escherichia coli* (Código PDB 1lxa). Para visualizar los puentes de hidrógeno, lo primero que hay que hacer es agregar los átomos de hidrógeno a la molécula, cosa que se hace desde el menú **Action -> Hydrogens -> add**, o ingresando el comando `h_add` en la consola. Para cambiar el modo de visualización de esquema (cartoon) a bolas y palitos, **A -> preset -> balls and sticks**.

Explorá la estructura y usando el zoom centrar un grupo de residuos entre los cuales exista un puente de hidrógeno. Para realizar mediciones en el modelo estructural, es necesario activar el modo de medición (measurement) en el menú principal, pestaña **Wizard -> Measurement**. Seleccioná distancias (**distances**), y luego hacer click en los átomos entre los cuales se quiere determinar la distancia (ej. un átomo de hidrógeno y un átomo aceptor). Repetí esta operación varias veces con distintos grupos atómicos y calculá el promedio de la distancia.

6. Identificación de interacciones π - π y π -catión

Estudiaremos más de cerca estas interacciones, utilizando la proteína miohemeritina (hacé **fetch** del PDB 1A7E). Una vez que la proteína se cargó:

- Ocultá las moléculas de agua (**H -> waters**),
- Seleccioná los residuos aromáticos (PHE/F, TYR/Y, TRP/W) manualmente en la barra de secuencia o mediante el siguiente comando: **select aro, resn PHE+TYR+TRP**.
El comando select se puede utilizar para seleccionar cualquier átomo, residuo, cadena, etc. Para más información sobre el comando select visite [Select help](#). Más detalles en el ítem [Selection Algebra](#). [`select selection_name, type XXX+XXX+...`]
- Una vez seleccionados los residuos aromáticos, mostralos como bolas y palitos (**A -> preset -> balls and sticks**) y localizá anillos aromáticos que participen de interacciones π - π . Una vez identificadas medí la distancia entre los anillos, para ello, en el panel de medidas (measurements) seleccioná **Measurement mode -> distance to rings**.

Usando las estructuras 1A7E y la 1EJ1 y los protocolos descriptos en la parte anterior, explorá la proteína en busca de interacciones π -catión. Para ello seleccioná los aminoácidos con carga positiva y mostralos como bolas y palitos. Medí la distancia entre los grupos intervinientes.

Chequeá tus resultados utilizando el siguiente servidor: <http://capture.caltech.edu/> (Capture program).

7. Sitio activo de la anhidrasa carbónica

Utilizado la estructura de la anhidrasa carbónica 1THJ y la opción **Select** en el Menú **Edit**:

- ¿Cuántas subunidades (cadenas) tiene la estructura nativa?
- ¿Tiene heteroátomos esta molécula? ¿Cuáles?
- Identificá los residuos que unen el Zinc. Para ello utilizá los siguientes comandos:
 - **select zincs, metals**
 - **select nearzincs, zincs around 6**

También se puede usar el comando: **select all within 6 of metals**.

Cambiá la visualización de la selección a bolas y palillos.

Bioinformática UNQ

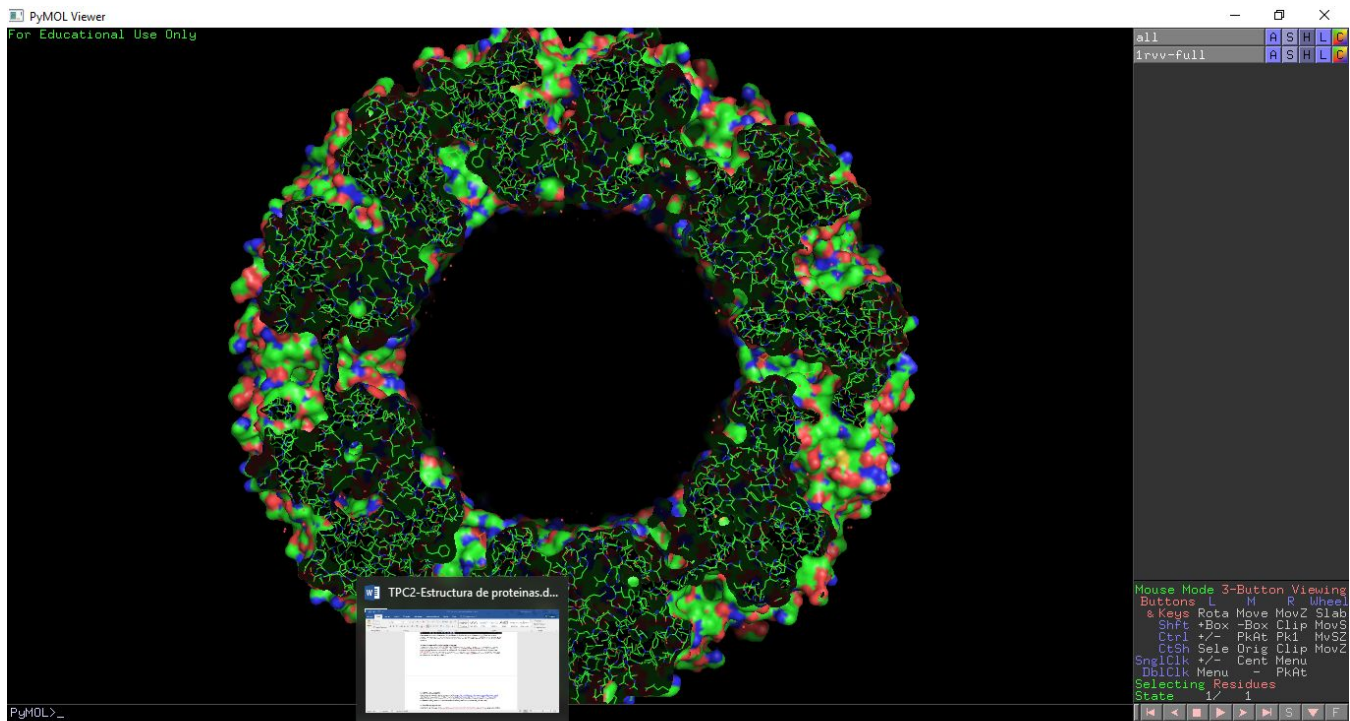
- d. ¿Los residuos que unen el Zinc pertenecen a la misma subunidad?
- e. Determiná los contactos proteína-proteína en la estructura cuaternaria de la anhidrasa carbónica. En este punto usaremos distintos programas para establecer qué aminoácidos participan en contactos proteína-proteína. Para ello accedé al servidor Protein-Protein interaction server InterProSurf (<http://curie.utmb.edu/pdbcomplex.html>) y cargá el código pdb 1THJ. Registrá el número de los residuos que mantiene contactos entre subunidades. Visualizá los mismos en Pymol. Utilizá la barra de secuencia o los siguientes comandos:
`- select interactingA, chain A & resi x+x+x+x... donde x es el número de los residuos que se desea seleccionar (Obtenidos de InterProSurf). Realizá la misma operación para las cadenas B y C.`

8. Uso de superficies y slab mode en Pymol

Cargá la estructura 1RVV biological assembly 2 a partir del archivo 1RVV_full.pdb, o utilizando el comando:

```
fetch 1RVV, type=pdb2, multiplex=1
```

Esta proteína es la lumazina sintasa y está formada por 62 monómeros formando una esfera. Utilizando **S** → **Surface** determiná la superficie de la proteína. Ahora usando la rueda del mouse utilizá el slab mode para recorrer en profundidad la esfera que forma la proteína.



Bioinformática UNQ

9. Estudio de la porina de *N. gonorrhoeae*

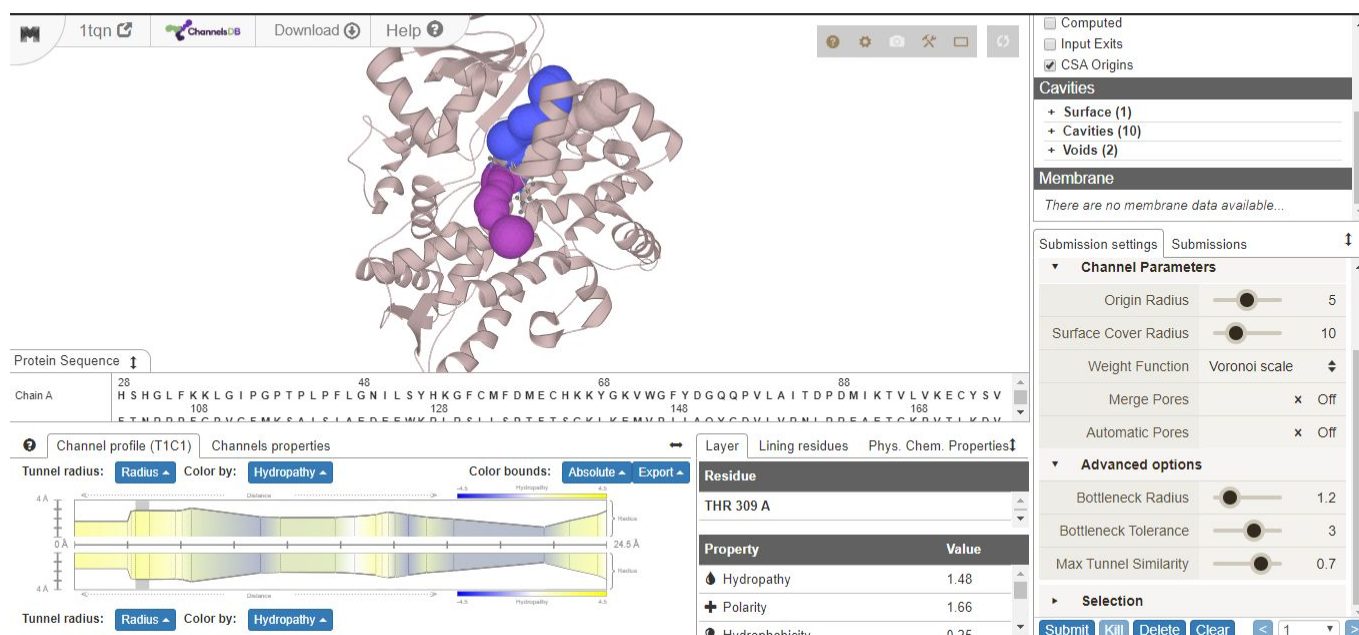
Utilizando el programa **Pymol** estudiá la estructura terciaria y cuaternaria de la proteína 4AUI. Determiná la superficie, localizá ligandos e identificalos. Si querés obtener más información visitá el sitio <https://www.rcsb.org/> y buscá por el ID: 4AUI.

Para seleccionar los ligandos podés usar el comando: `select ligandos, organic`

Para seleccionar grupos inorgánicos podés usar el comando: `select inorganicos, inorganic`

10. Estudio de túneles

Utilizando el programa (o servidor <https://mole.upol.cz/>) MOLE estudiá los túneles asociados al sitio activo de las estructuras 1THJ y 1F90. Este programa te pedirá que indiques a partir de qué regiones de la proteína debe comenzar la exploración en búsqueda de túneles. El punto de inicio puede ser a partir de cavidades, átomos o residuos seleccionados manualmente, o a partir de la base de datos CSA ([Catalytic site atlas](#)).



* Seleccionar los puntos de inicio y agregarlos pulsando **add**

* Se pueden ajustar los criterios de búsqueda como el diámetro mínimo del canal, etc en las secciones **Cavity parameters** y **Channel parameters**.

11. Estudio de pockets en EGFR

El EGFR es uno de los principales marcadores de cáncer de pulmón. Utilizando la sesión de Pymol estudiá los pockets. 1M14 es un conformero activo y 3W32 uno inactivo. Compará el sitio activo de ambos conformeros así como también los tamaños de los pockets.

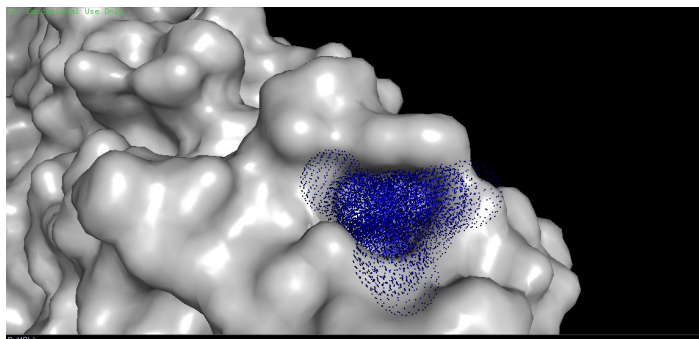
La determinación de las cavidades se puede realizar mediante el servidor fpocket: <https://mobyli.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::fpocket>

Desactivar la opción de demo: "Test the service with server sample data (input parameters will be discarded)".

PyMol visualization script [PML] [SH] -- all-in-one [TGZ]
 VMD visualization script [ICL] [SH] -- all-in-one [TGZ]
 Detected pockets alpha spheres [PDB]
 Alpha spheres centers and radii [PQR]

Bioinformática UNQ

👁 El servidor es un poco lento.. momento de arreglar el mate!!



Una vez que termina podés bajarte el **Pymol visualization script [TGZ]** para ver los resultados. Descomprimir (con 7zip u otro) y abrir el archivo .pml con Pymol....

Lo que vas a ver es para cada cavidad un conjunto de esferas (fig. siguiente)

Primero para la molécula **H** → **cartoon**, **H** → **spheres**, **S** → **Surface**

Luego para cada 'pocket' **S** → **dots**!!

12. Túneles y diseño de fármacos

El estudio estructural de proteínas nos proporciona múltiples campos de aplicación, uno de los más explorados en la actualidad es el diseño racional de fármacos. Si se conoce la base biológica de una enfermedad, es decir se conocen las moléculas implicadas, es posible diseñar medicamentos que interactúen con la molécula responsable, de tal forma que la modifique y se modifique el cuadro patológico. En otras palabras, el diseño racional de fármacos consiste en la aplicación del conocimiento biológico y estructural de los receptores (proteínas involucradas en una dada enfermedad) para diseñar moléculas que interactúen sólo con estos... dentro de lo posible!

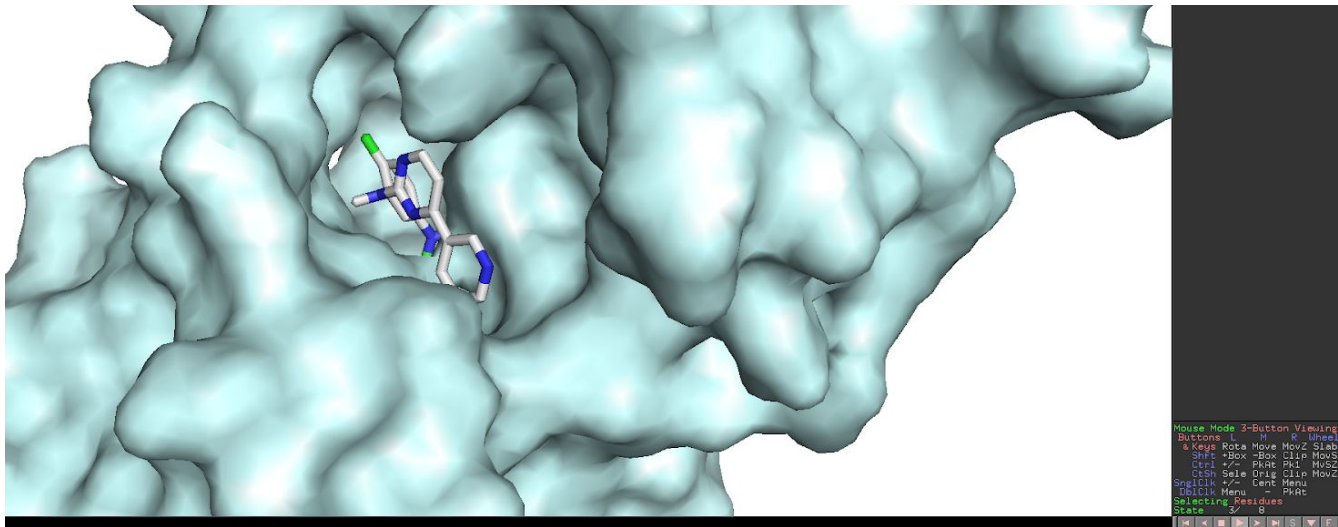
Veamos el ejemplo de la proteína Acetil-CoA sintetasa de humano, involucrada en cáncer de mamas. Como bien sabemos para que una droga pueda unirse a la proteína y modificar su función debe tener dónde hacerlo. Tomando como base la estructura PDB 3GPC, estudié sus túneles utilizando el Fpocket ¿Encontrás algún/os túnel/es en la proteína donde pueda unirse algún fármaco? ¿Qué residuos se encuentran en dicha cavidad?

Ahora analicemos los resultados de los investigadores de la UBA y UNQ (<https://doi.org/10.1007/s00018-020-03679-5>), que mediante una técnica de acoplamiento molecular (docking) diseñaron un inhibidor para esta proteína, capaz de unirse a la misma y reducir entonces la proliferación de células tumorales.

RETO III: Investigá en qué consiste el docking, en qué ideas basa su funcionamiento.

[Descargá](#) la estructura de la proteína y la del inhibidor y visualizalas en Pymol (abrí ambos archivos en la misma sesión) ¿Dónde se une el inhibidor? ¿Coincide con el túnel que propusiste anteriormente? ¿Dada esta inspección ocular cómo creés que actúa el inhibidor? Usando el play de la botonera, observá las distintas conformaciones del ligando o inhibidor.

Bioinformática UNQ

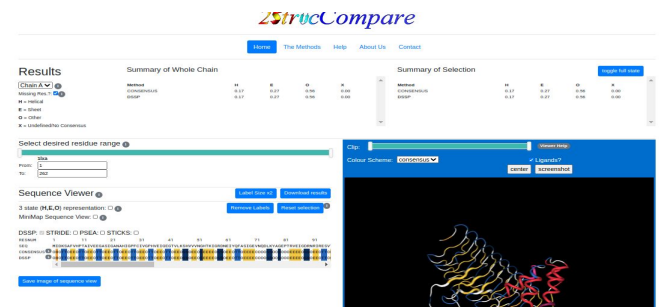


13. Asignación de estructura secundaria basado [DSSP](#)

Para inferir estructura secundaria utilizaremos el servidor 2StrucCompare con la Endolisina de bacteriófago en dos variantes (3F8V y 4LMZ).

- 1) Primero determinará la estructura secundaria de ambas conformaciones por separado usando el servidor <https://2struccompare.cryst.bbk.ac.uk/index.php>

- a) Corré el análisis por separado de cada estructura (recuadro derecho), cargando su código PDB y haciendo click en submit
- b) Observá las distintas regiones de estructura secundaria, ¿todas son de igual cantidad de residuos? ¿Se requieren más o menos residuos para formar una alfa hélice que un loop? ¿Por qué?



- 2) Luego podemos comparar ambas conformaciones utilizando el mismo servidor <https://2struccompare.cryst.bbk.ac.uk/index.php>. Esta vez cargaremos las estructuras a la par (recuadro izquierdo) y el servidor nos mostrará las diferencias

- a) Utilizando el botón Colour Scheme especificá las zonas de mayor movimiento de la proteína

