



# Instituto Politécnico Nacional Escuela Superior de Cómputo



Bioinformatics

Lab Session 2. PDB

González Bocio Erik Alexander 2020630163

Jorge Luis Rosas Trigueros

Fecha de realización: 21 de febrero de 2022

Fecha de entrega: 28 de febrero de 2022

## Marco Teórico:

El Protein Data Bank (PDB) es una base de datos donde se almacenan las estructuras cuya estructura tridimensional (es decir, sus coordenadas atómicas) ha sido resuelta.

Básicamente, hay tres métodos experimentales que permiten determinar la estructura tridimensional de una molécula:

- A partir de un cristal, por difracción de Rayos-X
- En disolución, por RMN
- A partir de micrografías electrónicas, mediante programas de reconstrucción de imágenes

La base de datos de la AP se actualiza semanalmente (UTC +0 miércoles), junto con su lista de existencias. A 1 de abril de 2020, la PDB estaba compuesta como se puede ver en la Fig. 1

Método experimental	Proteínas	Ácidos nucleicos	Proteína/Ácidos nucleicos complejos	Otros	Total
Cristalografía de rayos X	135170	2097	6945	4	144216
Espectroscopia de resonancia magnética nuclear	11337	1325	264	8	12934
Microscopía electrónica	3475	35	1136	0	4646
Híbridos	155	5	3	1	164
Otros	286	4	6	13	309
Total:	150423	3466	8354	26	162269

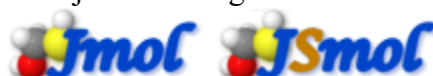
*Fig. 1. Estructura de los registros del Protein Data Bank, en el podemos ver el total de existencias hasta el 2020, teniendo el total de proteínas, ácidos nucleicos, entre otros*

Los registros del PDB contienen gran cantidad de información. Esta información se distribuye en 12 secciones y cada sección comprende distintos apartados, tal y como se indica en la fig. 2.

SECTION	DESCRIPTION	RECORD TYPE
Title	Summary descriptive remarks	HEADER, OBSLTE, TITLE, CAVEAT, COMPND, SOURCE, KEYWDS, EXPDTA, AUTHOR, REVDAT, SPRSDE, JRNL
Remark	Bibliography, refinement	REMARKs 1, 2, 3 & annotations
Primary structure	Peptide and/or nucleotide sequence and the relationship between the PDB sequence and that found in the sequence database(s)	DBREF, SEQADV, SEQRES MODRES
Heterogen	Description of non-standard groups	HET, HETNAM, HETSYN, FORMUL
Secondary structure	Description of secondary structure	HELIX, SHEET, TURN
Connectivity annotation	Chemical connectivity	SSBOND, LINK, CISPEP
Miscellaneous features	Features within the macromolecule	SITE
Crystallographic	Description of the crystallographic cell	CRYST1
Coordinate transformation	Coordinate transformation operators	ORIGXn, SCALEn, MTRIXn, TVECT
Coordinate	Atomic coordinate data	MODEL, ATOM, SIGATM, ANISOU, SIGUIJ, TER, HETATM, ENDMDL
Connectivity	Chemical connectivity	CONECT
Bookkeeping	Summary information, end-of-file marker	MASTER, END

*Fig. 2. Estructura de los registros del Protein Data Bank, podemos ver la distribución mencionada y sus apartados.*

Por otra parte, tenemos Jmol el cuál es un visor Java de código abierto para estructuras químicas en 3D con características para productos químicos, cristales, materiales y biomoléculas. JSmol es la modalidad HTML5 de Jmol, capaz de ser incrustada en páginas web. Toda la funcionalidad de Jmol (como una aplicación independiente) también está presente en JSmol. JSmol es un objeto de navegador web interactivo.



*Fig. 3. Logotipos de Jmol y JSmol respectivamente.*

Jmol es un visor gratuito y de código abierto de estructuras moleculares útil para estudiantes, educadores e investigadores en química, bioquímica y otros campos relacionados con la estructura molecular. Es multiplataforma, se ejecuta en sistemas Windows, Mac OS X y Linux/Unix.

- La aplicación Jmol es una aplicación Java independiente que se ejecuta en el escritorio.
- El JSmol es un objeto que se puede integrar en páginas web. No requiere Java, ya que se ejecuta utilizando solo los motores HTML5 y JavaScript del navegador.
- El JmolViewer es un kit de herramientas de desarrollo que se puede integrar en otras aplicaciones Java.

Material y equipo:

- Computadora

- Página de PDB ([www.rcsb.org](http://www.rcsb.org))
- Página de introducción a Jmol ([http://earth.callutheran.edu/Academic\\_Programs/Departments/BioDev/omm/jsmol/scripting/molmast.htm](http://earth.callutheran.edu/Academic_Programs/Departments/BioDev/omm/jsmol/scripting/molmast.htm))
- Red Wifi

## Desarrollo de la práctica:

### Instrucciones:

1. Go to the PDB at [www.rcsb.org](http://www.rcsb.org) and find the entry 1UBQ.

Al entrar a la página del PDB encontramos un montón de noticias sobre descubrimientos y otras cosas (Fig. 4).



Fig. 4. Página principal del PDB, podemos ver una infinidad de noticias sobre bioinformática y descubrimientos hechos en la plataforma.

Luego, se buscó 1UBQ resultando ser la proteína Ubiquitina como se muestra en la fig. 5.

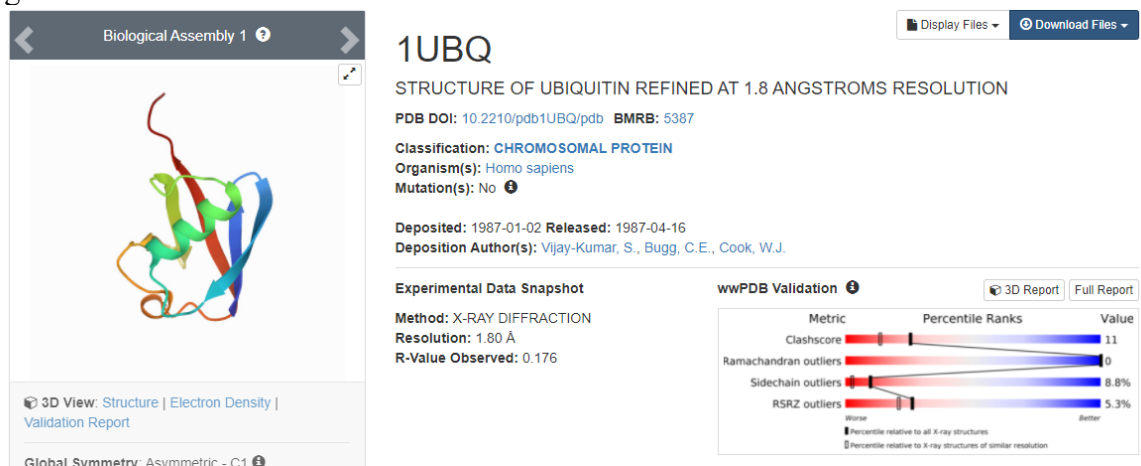


Fig. 5. Ubiquitina, se puede observar una descripción de la proteína y opciones para trabajar con ella como son los Display Files

Se dio una introducción acerca de la página, lo que se puede hacer, un poco sobre lo que es la ubiquitina y lo que se muestra en la página, se dio una introducción a la parte 3d donde se nos muestra la proteína y finalmente como trabajar con Jmol desde esta herramienta (Fig. 6).



Fig. 6. Vista 3D de la Ubiquitina. Esta es la interfaz con la que se trabaja el Jmol

2. Follow the "Introduction to Jmol scripting tutorial" at [http://earth.callutheran.edu/Academic\\_Programs/Departments/BioDev/omm/jsmol/scripting/molmast.htm](http://earth.callutheran.edu/Academic_Programs/Departments/BioDev/omm/jsmol/scripting/molmast.htm).

En este link se dan los conocimientos básicos sobre los comandos que pueden usarse en Jmol con ayuda de una interfaz web muy amigable.

Fig. 7. Introducción a Jmol, podemos ver una figura 3d que cambia dependiendo de las opciones que elijamos, con el podemos ver que hace cada comando.

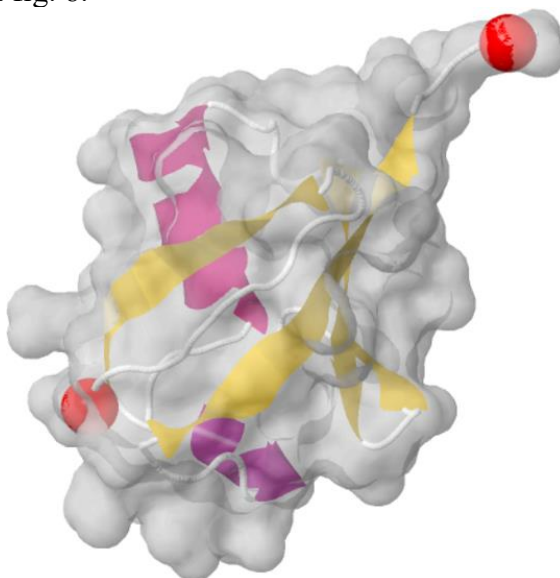
La introducción no quedo allí ya que también la página de PDB nos da más links de ayuda para trabajar con Jmol, como lo son:

- <http://wiki.jmol.org/index.php/Rendering> el cual nos ayuda con la renderización
- <http://wiki.jmol.org/index.php/Selecting> que nos ayuda con la selección
- <http://jmol.sourceforge.net/docs/> y <https://chemapps.stolaf.edu/jmol/docs/> que nos da la documentación de Jmol y JSmol
- y finalmente <http://jmol.sourceforge.net/docs/surface/> que nos ayuda con las superficies

Con ayuda de estas llevamos a cabo el siguiente punto.

3. Obtain the representation depicted in Figure 1 for ubiquitin.

Luego de ver la introducción tenemos la siguiente imagen, la cual es la ubiquitina y se nos muestra en la fig. 8.



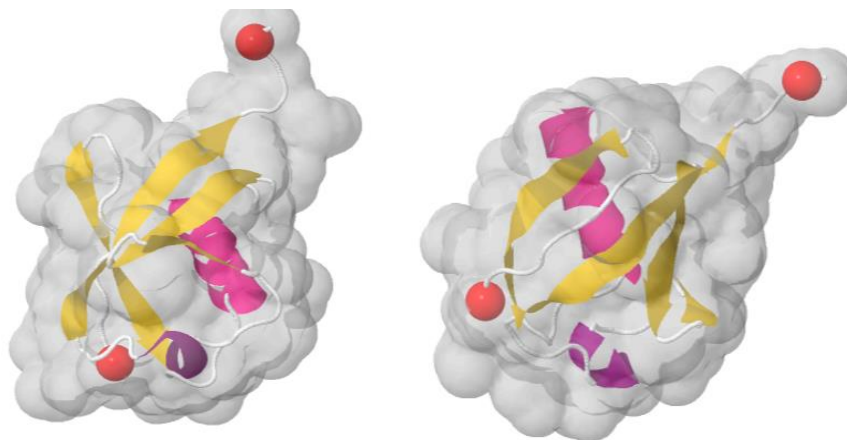
*Fig. 8. Modelo 3D que se desea obtener, este es accesible al solvente con superficie translúcida y su primer y último aminoácidos con carbono alfa en esfera roja.*

Primero se colocó como “Solvent Accessible” en la parte de Surface en Select Options. Luego tenemos los comandos:

- [isosurface ignore\(solvent\) colorscheme sets translucent 0.7 white; select \(1,76\) and \\*.CA; spacefill; color red;](#)

Primero tenemos el comando: `isosurface ignore(solvent) colorscheme sets translucent 0.7 white;` el cual colocará el “Solvente accesible” de color blanco translucido en 0.7, luego los comandos: Y conseguimos el resultado de la fig. 9.



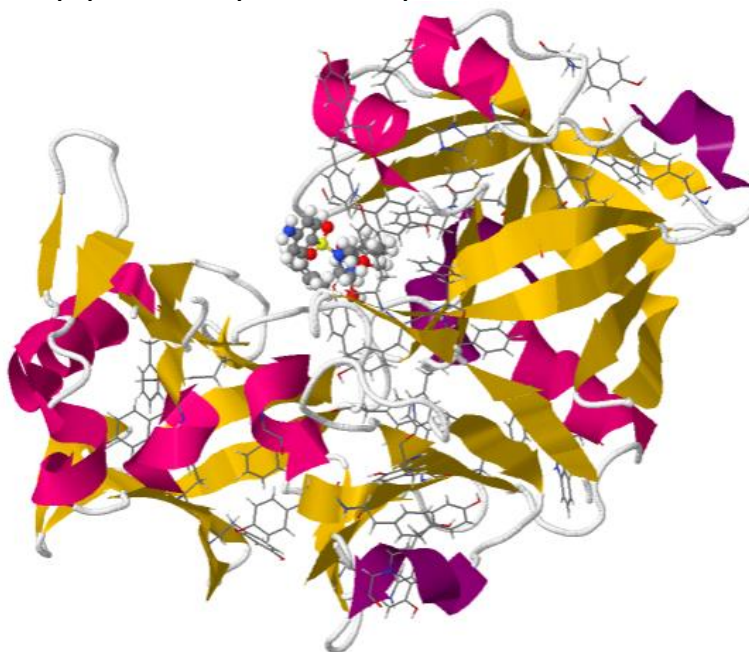


*Fig. 9. Resultado de la ubiquitina luego de colocarle comandos.*

4. Find another protein in the PDB and obtain a representation that shows the aromatic residues in the protein in wireframe and cpk coloring.

La proteína seleccionada es:

6XCT: Porcine pepsin in complex with amprenavir



Y usamos los siguientes comandos para llevar a cabo la representación:

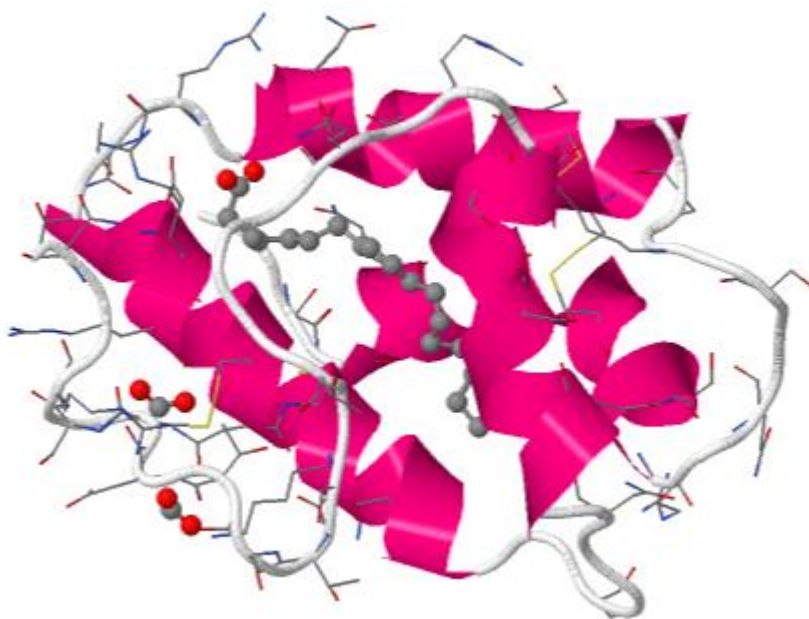
- [select aromatic; wireframe;color cpk;](#)

El comando select aromatic; muestra los residuos aromáticos, mientras wireframe muestra el wireframe o una estructura inalámbrica y por último el color cpk que le da el color dependiendo del átomo.

5. Find yet another protein in the PDB and obtain a representation that shows the polar residues in the protein in wireframe and cpk coloring.

La proteína seleccionada es:

## 1FK6: STRUCTURAL BASIS OF NON-SPECIFIC LIPID BINDING IN MAIZE LIPID-TRANSFER PROTEIN COMPLEXES WITH ALPHA-LINOLENIC ACID REVEALED BY HIGH-RESOLUTION X-RAY CRYSTALLOGRAPHY



Para llevar a cabo la representación de la proteína usamos los siguientes comandos:

- `select polar; wireframe;color cpk;`

El comando `select polar;` muestra los residuos polares, mientras `wireframe` muestra el wireframe o una estructura inalámbrica y por último el `color cpk` que le da el color dependiendo del átomo

### Conclusiones y recomendaciones:

La página de PDB muestra una base de datos con una gran cantidad de proteínas, que es con lo que trabajamos en esta práctica, en ella podemos hacer uso de un visor de JAVA con el que podemos hacer representaciones de estas proteínas, además de que nos da una descripción de estas.

Me pareció una buena práctica ya que la herramienta utilizada me pareció muy interesante y divertido hacerles cambios a las proteínas y ver que les pasaba con cada comando, luego hacer con los comandos algo en específico y trabajar con varias proteínas, haciéndonos conocer más proteínas.

### Bibliografía:

- *Jmol: an open-source Java viewer for chemical structures in 3D*. Jmol.sourceforge.net. (2022). Retrieved 26 February 2022, from <http://jmol.sourceforge.net/>.
- *Protein Data Bank*. Ehu.eus. (2022). Retrieved 26 February 2022, from <http://www.ehu.eus/biofisica/juanma/rasmol/pdb.htm>.



- Silva, N., & Marcey, D. (2022). *Intro to Jmol Scripting*. Earth.callutheran.edu. Retrieved 26 February 2022, from [http://earth.callutheran.edu/Academic\\_Programs/Departments/BioDev/omm/jsmol/scripting/molmast.htm](http://earth.callutheran.edu/Academic_Programs/Departments/BioDev/omm/jsmol/scripting/molmast.htm).