

Estudio estructural de proteínas - Visualizando Estructuras con Chimera

Lucía B. Chemes, Juliana Glavina

Introducción

Chimera, software de visualización de estructuras.

Para poder visualizar estructuras macromoleculares tales como proteínas globulares y ácidos nucleicos, se usará el software UCSF Chimera. Existen otros software similares pero Chimera tiene la ventaja de ser mantenido actualmente por sus desarrolladores.

Chimera es un programa disponible de manera gratuita, y está disponible para descargar e instalar en tu propia computadora en <http://www.cgl.ucsf.edu/chimera/>

Existen un alto número de guías y tutoriales disponibles online que pueden encontrarse en: <http://www.cgl.ucsf.edu/chimera/current/docs/UsersGuide/>

Antes de empezar, pueden responder: ¿Porqué es importante visualizar estructuras proteicas? ¿Qué información podríamos obtener de ellas? ¿Estas estructuras, son un objeto real o un modelo?

Bases de Datos Estructurales. ¿Dónde almacenan todas las estructuras? ¿Cómo accedemos a ellas?

La base de datos de proteínas (*Protein Data Bank*, PDB) almacena actualmente más de 150000 estructuras. Puedes acceder a ella aquí: <http://www.rcsb.org/pdb/home/home.do>

Existe también una versión europea de esta base de datos (*European Protein Data Bank*, PDBe). Puedes acceder a ella desde aquí: <https://www.ebi.ac.uk/pdbe/>

Las estructuras poseen un identificador de 4 caracteres alfanuméricos. Por ejemplo: **1GUX** es el identificador, o PDB ID, de la proteína retinoblastoma unida a un péptido de la proteína E7 de papillomavirus.

La búsqueda de estructuras puede realizarse utilizando palabras claves, por ejemplo, la palabra clave **retinoblastoma** devuelve un total de 173 estructuras, o por otras características como nombre de alguno de los autores que participó en el estudio de esa estructura, por ejemplo, **Rubin, S.M.** devuelve un total de 32 estructuras. Pueden explorar la base de datos RCSB PDB para familiarizarse con ella.

El archivo PDB. ¿Cómo están codificadas las estructuras?

Las estructuras tridimensionales de las proteínas pueden generarse por diferentes métodos (cristalografía de rayos X o XRay, resonancia magnética nuclear o RMN, criomicroscopía electrónica o CryoEM) y están codificadas en archivos pdb. Un archivo pdb está compuesto por múltiples líneas de registros, cada uno identificado por una etiqueta determinada incluidos dentro de distintas secciones. En la figura se muestra un fragmento de la sección de coordenadas que describe la estructura de la proteína dando las coordenadas x, y y z (azul claro) de cada uno de los átomos identificados.

Etiqueta ATOM identificando el registro, aminoácidos o nucleótidos.	HEADER				Coordenadas			
	...	Serie de números del átomo	Cadena		x y z			
	ATOM	442 N	SER A 30		-10.207	-6.447	0.056	...
	ATOM	443 CA	SER A 30		-11.223	-7.213	0.775	...
	ATOM	444 C	SER A 30		-10.690	-7.570	2.157	...
	ATOM	445 O	SER A 30		-9.499	-7.831	2.274	...
	ATOM	446 CB	SER A 30		-11.544	-8.487	-0.006	...
	ATOM	447 OG	SER A 30		-11.828	-8.169	-1.360	...
	ATOM	448 H	SER A 30		-9.869	-6.828	-0.814	...
	ATOM	449 HA	SER A 30		-12.124	-6.604	0.865	...
	ATOM	450 HB2	SER A 30		-10.689	-9.165	0.047	...
	ATOM	451 HB3	SER A 30		-12.399	-8.980	0.462	...
	ATOM	452 HG	SER A 30		-12.129	-8.971	-1.802	...
	ATOM	453 N	ALA A 31		-11.532	-7.606	3.195	...
	ATOM	454 CA	ALA A 31		-11.066	-7.597	4.584	...
	ATOM	455 C	ALA A 31		-9.986	-8.639	4.887	...
	ATOM	456 O	ALA A 31		-9.036	-8.345	5.615	...
	ATOM	457 CB	ALA A 31		-12.257	-7.802	5.523	...
	ATOM	458 H	ALA A 31		-12.513	-7.435	3.031	...
	ATOM	459 HA	ALA A 31		-10.618	-6.620	4.774	...
	ATOM	460 HB1	ALA A 31		-12.718	-8.774	5.329	...
	ATOM	461 HB2	ALA A 31		-11.907	-7.782	6.557	...
	ATOM	462 HB3	ALA A 31		-12.992	-7.012	5.379	...
	//							
	HETATM	1725 ZN	ZN A 57		7.486	7.761	5.097	...
	//							
	...	Nombre del átomo	Nombre del residuo	Número del residuo				
	END							

Ocupancia
Factor de Temperatura
Símbolo del átomo
Carga del átomo

En cada línea, además, se identifica si es un átomo (rojo) perteneciente a un aminoácido o nucleótido, o heteroátomo (azul oscuro), la numeración (verde), el nombre del átomo (naranja), el nombre del residuo en el que está incluido el átomo (violeta), la cadena a la que pertenece (negro), el número del residuo al que pertenece (verde). Este archivo puede incluir más columnas para cada átomo con datos relacionados con el espacio que ocupa el átomo, la movilidad del átomo (el factor de temperatura o B-factor), el símbolo que representa al átomo y la carga del mismo (señaladas con “...” en la figura). El encabezado o header del archivo PDB puede tener muchísima información no relacionada directamente con la conformación tridimensional de la proteína, sino con cómo se hizo el experimento, las publicaciones relacionadas y otros. La descripción del resto del contenido de las secciones del archivo pdb puede obtenerse en la sección documentación de <http://www.wwpdb.org/>.

Ejercicio. Busca, descarga y abre un archivo PDB. ¿Puedes identificar los distintos datos? ¿Puedes identificar qué otra información se encuentra? ¿Cómo está señalizada u organizada dicha información?

Los archivos PDB, fueron utilizados por décadas para describir la estructura de una macromolécula. Sin embargo, tiene un número limitado de líneas y átomos que se pueden

codificar y actualmente no puede manejar complejos macromoleculares muy grandes, partículas virales, etc.

Recientemente, se desarrolló un nuevo formato de archivo, MMCIF, que puede codificar estructuras de complejos macromoleculares mucho más grandes. Sin embargo, no todos los programas utilizados para visualizar estructuras soportan este formato.

Ejercicio. Busca, descarga y abre un archivo MMCIF. ¿Puedes identificar los distintos datos? ¿Puedes identificar qué otra información se encuentra? ¿Cómo está señalizada u organizada dicha información?

Guía de Ejercicios - Chimera

Chimera es un programa con muchas potencialidades y herramientas de análisis. En la siguiente guía, introduciremos el uso de las funciones más comunes de este programa, en relación al análisis estructura-función de proteínas

El objetivo principal de estos ejercicios iniciales es:

- Aprender a visualizar estructuras 3D de proteínas o complejos proteicos, incluyendo diferentes niveles de análisis
- Identificar superficies de interacción.
- Identificar diferentes tipos de unión molecular tales como puentes de hidrógeno o interacciones mediadas por carga y entender el rol de los mismos.
- Identificar interacciones hidrofóbicas y entender el rol de las mismas.
- Aprender a superponer estructuras relacionadas y analizar los resultados.

I- Iniciando Chimera

1. Localiza chimera en tu computadora y ábrelo.
2. Existen dos formas de cargar una estructura en Chimera:

Opción 1: Ve a **File**, selecciona **Fetch by ID**.

Ingresa el identificador del PDB deseado. Por ejemplo: **2AZE**.

Se descargará de manera remota el PDB 2AZE de la base de datos PDB.

Opción 2: Ve a **File**, selecciona **Open**.

Busca el archivo PDB en la carpeta donde lo hayas descargado previamente y selecciónalo. Todos los archivos se encuentran además en la carpeta **PDBs** del TP de la materia.

Archivos PDBs con los que trabajaremos:

PDB ID	Descripción
2AZE	Complejo creado a partir del ensamblado de fragmentos proteicos provenientes de pRB/E2F1/DP1
1YCR	Complejo de MDM2 y un péptido proveniente del dominio de transactivación de p53
2GDM	Estructura de la LegHemoglobina
3RGK	Structura de la Myoglobin humana
5LGY	4 Core Domain DE P53 unidos al BAX response element (DNA)

Ejercicio 1. Algunas definiciones antes de empezar a visualizar

- ¿Como se define una superficie de interacción?
- ¿Como se definen interacciones puentes de hidrógeno, interacciones mediadas por carga, hidrofóbicas?
- ¿Porqué sería útil superponer estructuras relacionadas?

Ejercicio 2. Familiarizándonos con Chimera

Existen diversas maneras de visualizar los complejos estructurales. Dependiendo cuál es el objetivo del análisis a veces es mejor utilizar distintos métodos de visualización como superficies, átomos o ribbons (cinta).

Chimera tiene un menú muy complejo. Pero solamente vamos a usar algunos comandos.

Primero carguemos la estructura correspondiente al **PDB 2AZE** y vayamos viendo como cambia la representación de la estructura.

MENÚ SELECT.

Este menú permite seleccionar y deseleccionar secciones de estructuras.

- Se pueden seleccionar cadenas (*chain*), o características estructurales (*structure*) como estructuras secundarias (*secondary structure*), determinado tipo de aminoácidos (*residue*) o aminoácidos agrupados por categorías (*amino acid category*) como ser hidrofóbicos (*hydrophobic*), cargados (*charged*), etc.
- Para deseleccionar se utiliza la opción *clear selection*.

MENU ACTIONS

Permite realizar determinadas acciones en la molécula, relacionadas con:

- Representación de la molécula.
 - Superficie (*Surface* → *Show*)

- Cintas (*Ribbon* → *Show*)
- Átomos/Uniones (*Atoms/Bonds* → *Show*)
- Apariencia de la representación
 - Superficie en forma sólida, como una red o como puntos (*Surface* → *Solid/Mesh/Dot*).
 - Cintas más planas, redondas o gruesas (*Ribbon* → *Flat, rounded, edge*)
 - Átomos como palillos (*sticks*), líneas (*wire*), esferas, etc.
- Color deseado (*Color*)
- Centrar la molécula (*Focus*)

Ejercicio para discutir. ¿Se te ocurre para qué puede ser útil cada forma de visualización?

Ejercicio 3. Analizando estructuras

En este ejercicio trabajaremos con el complejo formado entre las proteínas E2F y DP1 (factores de transcripción) y la región C-terminal del regulador del ciclo celular Retinoblastoma. El objetivo es familiarizarse con el uso de diferentes modos de visualización y análisis de interacciones dentro de cada cadena y entre dominios

Con el PDB 2AZE:

1. Selecciona cada una de las cadenas y asigne a cada cadena un color diferente para poder individualizarlas mejor.

MENU TOOLS

Abarca un conjunto variado de opciones.

- *Depiction*. Contiene opciones para modificar colores, etiquetas, entre otras cosas
- *Structure analysis*. Contiene distintos tipos de análisis comúnmente realizados sobre la estructura.
 - *FindHBond*. Permite encontrar los puentes de hidrógenos.

2. Selecciona la cadena A del PDB. Luego, ve a *Structure analysis, FindHBond*, y selecciona *Only Find H-Bonds with both ends selected*.

Asegúrate que la opción *Include Intra-molecule H-bonds* está seleccionada.

¿Qué se observa? ¿Entre qué grupos se forman los puentes de hidrógeno? ¿De qué tipo de estructura forman parte?

Para removerlos es necesario usar la línea de comandos que se encuentra en la parte inferior de la pantalla. Si no puedes visualizarla, ve a *Favorites* → *Command Line*.

En **Command**. Tipea: *~hbond*

3. Selecciona la cadena A del PDB. Luego, ve a *Structure analysis, FindHBond*, y selecciona *Only Find H-Bonds with both ends selected*.

Asegúrate que la opción *Include Intra-molecule H-bonds* **NO** está seleccionada.

¿Qué se observa? ¿Entre qué grupos se forman los puentes de hidrógeno? ¿De qué tipo de estructura forman parte?

Ejercicio 4. Estudiando interacciones intermoleculares

En este ejercicio vamos a trabajar con el complejo entre el dominio TAD de p53 y el dominio globular de la E3 ligasa MDM2, e interpretaremos el tipo de interacciones que estabilizan este complejo utilizando métodos para visualizar este tipo de interacción

1. Carga el pdb **1YCR**.
¿Cuántas cadenas observas? ¿Son dominios, y sino qué son?
2. En *Favorites* selecciona *Model Panel*. Allí podrás ver el identificador del pdb y algunos accesos rápidos a algunas de las acciones. En *select chain(s)*... Selecciona la **cadena A** y representala como superficie.
¿Qué observas? ¿Se ve bien la superficie haciéndolo de esta manera?
3. En la ventana *Model Panel* verás que aparece ahora también el elemento superficie (*MSMS main surface of 1ycr*)
Esconde la superficie. Puedes hacerlo desde *Model Panel* eliminando el tick al lado del elemento superficie en la columna S (o bien puedes hacerlo como antes: *Actions* → *Surface* → *Hide*).

Para poder representar toda la superficie de la cadena es necesario dividir a la estructura en elementos separados. Para eso:

En **Command** ingresa *split*. Este comando permite que todas las cadenas sean objetos individuales. En la ventana *Model Panel* aparecen ahora las dos cadenas por separado.

Selecciona nuevamente la **cadena A**, puedes hacerlo haciendo click sobre **1ycr A** y presionando *select*. Representala como superficie. Observa la diferencia: ¿se ve mejor?

Colorea la superficie por heteroátomos. Para esto: *Action* → *Color* → *by heteroatom*

¿A qué residuos corresponden los colores? ¿porqué es útil este tipo de representación? Pensar en las propiedades fisicoquímicas de los grupos que componen a las proteína, y sus tipos de interacciones

4. Selecciona ahora la **cadena B**, representala como **ribbon**.
Muestra las cadenas laterales de los aminoácidos. Para esto:
Action → *atoms/bonds* → *Show*
Aparecerán automáticamente representados como *Sticks*. Puedes cambiarlos a *ball & stick*
Action → *atoms/bonds* → *ball & stick*

Para visualizar la secuencia del péptido ve a: *Tools* → *Sequence* → *Sequence* y selecciona la **cadena B**.

¿Puedes correlacionar los residuos en la secuencia con la estructura?

5. Ahora representaremos algunos aminoácidos de distinta manera. En la ventana de la secuencia marca los residuos F19 y W23. Para seleccionar es necesario hacer click y

arrastrar levemente el mouse. Luego, manteniendo shift apretado selecciona el segundo residuo. Representa estos residuos como esferas.

- A. ¿Qué observas? ¿Qué información da esta representación?
- B. ¿Qué puedes decir de estos residuos?
- C. ¿Qué relación estructural encuentras entre estos residuos y el alineamiento realizado en la práctica de MSA? ¿De qué elemento se trata y qué aprendemos de cómo se relacionan estructura y secuencia?

Por último, Chimera tiene algunas representaciones ya preconfiguradas en el *Menú Presets*.

MENU PRESETS

Abarca un conjunto de opciones preconfiguradas muy útiles.

En *Presets* selecciona *Interactive 3 (hydrophobicity surface)*.

En *Model Panel* deselecciona la *MSMS main surface of 1ycr B*

De acuerdo a la ‘hydrophobicity surface, y comparando con la representación de heteroátomos: ¿qué significa la escala de color que se ve, y qué propiedad fisicoquímica tiene el área (parche de unión) donde se unen los residuos F19 y W23? ¿De qué tipo de interacciones se trata?

Ejercicio 5. Alineando estructuras de dos proteínas

En este ejercicio, compararemos dos estructuras de proteínas transportadoras de oxígeno: Mioglobina humana y Leghemoglobina de plantas. Carga los PDBs correspondientes: 3RGK y 2GDM utilizando la función “fetch”.

Explora las estructuras y describe lo que observas. ¿Qué compuesto químico se observa unido a las proteínas?

Para tener una noción de cuán similar es la estructura de dos proteínas, podemos realizar un **Alineamiento Estructural**, que consiste en superponer las estructuras de ambas proteínas en el espacio intentando alinear sus cadenas aminoacídicas. Alinear estructuras en chimera es muy fácil, sólo requiere un comando.

Ve a *Tools* → *Structure Comparison* → *MatchMaker*

Se abrirá una nueva ventana.

En *Reference structure* (el panel de la izquierda) puedes seleccionar una de las estructuras de referencia. Esta estructura es la que se mantendrá fija. (Ej. 2GDM)

En *Structure(s) to match* (el panel de la derecha) selecciona la estructura que será superpuesta y alineada con la que se eligió como referencia. (Ej. 3RGK)

Observar el resultado del alineamiento: ¿Son parecidas las estructuras? ¿En donde se observan las mayores diferencias?

Para ver cómo se corresponde el grado de similitud estructural con el grado de similitud en secuencia podemos realizar un alineamiento de ambas secuencias guiado por el alineamiento estructural. Para esto, ve a:

Tools → Structure comparison → "Match->Align"

Ahora, observando la estructura y el alineamiento responde:

- A. ¿Qué relación existe entre el alineamiento estructural y el alineamiento de secuencia?
- B. ¿Qué son las regiones marcadas en rosa en el alineamiento?
- C. ¿Este alineamiento: identifica las regiones que no alinean estructuralmente?
- D. ¿Qué elementos encuentra en el alineamiento de secuencia en las regiones que no alinean? ¿A qué estructuras corresponden?

Para cuantificar el alineamiento de secuencia obtenido, podemos calcular el % identidad de secuencia. Para ello, en la ventana del alineamiento de secuencias ve a: *Info → Percent identity*. Selecciona una estructura en *Compare* y la otra estructura en *with*. Presiona en Ok.

Prueba dividiendo por longer, shorter and non-gap columns.

¿Qué valores de identidad de secuencia obtienes? ¿Las sustituciones observadas en las secuencias son conservativas?

En la carpeta MSA del TP de la materia, pueden encontrar las secuencias FASTA de ambas proteínas. El archivo **P02240.fasta** corresponde a Leghemoglobina y el archivo **P02144.fasta** corresponde a la Mioglobina. Utilizando las herramientas ya aprendidas, realicen un alineamiento de a pares de las secuencias y respondan: Obtienen la misma identidad de secuencia que por el método usado recién en Chimera? Si hay diferencias, pueden interpretarlas?

Entonces: ¿Qué se puede concluir de la divergencia estructural en comparación a la divergencia de secuencia en proteínas transportadoras de oxígeno? ¿Qué implica esto a nivel evolutivo?

Ejercicio 6. Visualizando ácidos nucleicos

Aquí analizaremos el dominio de unión a DNA de p53 estudiado previamente. Carga el pdb **1YCR**.

¿Cuántas cadenas observas? ¿A qué corresponde cada cadena?

Colorea cada dominio proteico con un color diferente.

Ve a *Actions → Atoms/Bonds → Nucleotide objects → off*

Selecciona los ácidos nucleicos: *Select → Residue → Standard nucleic acids*

Encuentra los puentes de hidrógeno.

Tools → Structural Analysis → Find H Bonds

Selecciona:

Only find H-Bonds → with both ends selected

Y asegurate que NO estén seleccionados:

Include intra-molecule H-bonds

Include intra-residue H-bonds

Según los conocimientos que tienes de las uniones en el ADN ¿Se observa lo que esperabas?