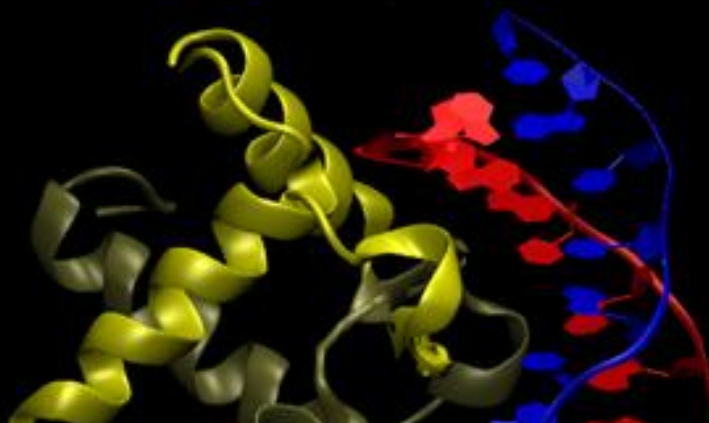
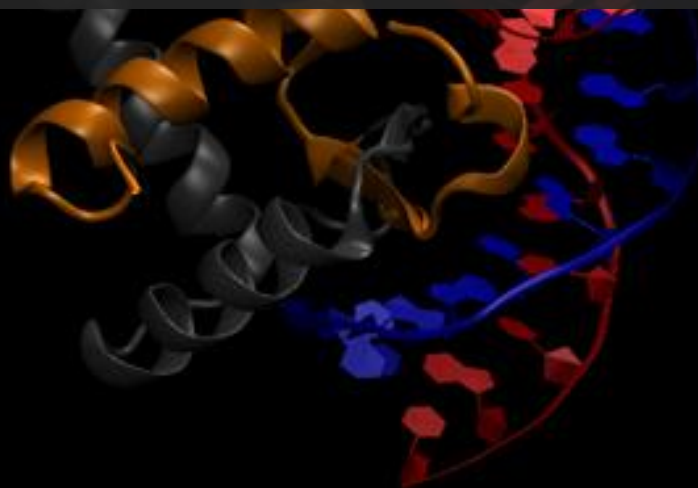


24 de septiembre de 2012



RSG
ARGENTINA

Workshop Pre-Congreso 3CAB2C 2012
DINÁMICA MOLECULAR DE PROTEÍNAS



Bustamante JP, Dumas VG, Gauto DF

Dinámica Molecular de Proteínas

Objetivos

El presente trabajo práctico se plantea como objetivo analizar simulaciones de dinámica molecular utilizando como herramienta el software VMD (Visual Molecular Dynamics), para responder preguntas concretas de un sistema biológico. De esta manera, se espera que los participantes se lleven una visión general de algunas herramientas de análisis utilizadas de manera cotidiana como punto de partida para plantear otras estrategias de estudio más avanzadas.

Introducción

Los métodos computacionales han surgido en las últimas décadas como herramientas muy poderosas para responder preguntas científicas. Entre las numerosas estrategias de cómputo, la dinámica molecular es hoy en día una de las principales herramientas utilizadas en el estudio de macromoléculas de interés biológico. Este método, que calcula el comportamiento dependiente del tiempo de un sistema molecular, se emplea ya de forma rutinaria para investigar la estructura, así como para proporcionar información detallada sobre la dinámica (fluctuaciones y cambios conformacionales) y la termodinámica de macromoléculas biológicas. Dentro de éstas encontramos proteínas, polisacáridos y ácidos nucleicos, así como sus complejos entre sí o con pequeñas moléculas que pueden ser cofactores, sustratos o ligandos en general.

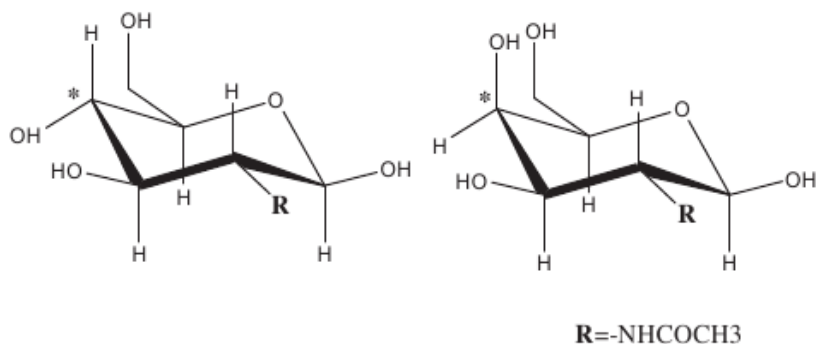
Una breve descripción del problema a resolver

Un ejemplo concreto lo vemos en la lectina del hongo *Agaricus Bisopus* (ABL) la cual une hidratos de carbono en dos sitios de unión distintos, los llamados Sitios de Reconocimiento de Carbohidratos (CBS). Una característica distintiva de estos CBS es que pueden reconocer cambios sutiles en los azúcares, dando esto como resultado la unión o la no unión de los mismos.

Por ejemplo, N-Acetil-Galactosomina (GalNAc) y N-Acetil-Glucosamina (GlcNAc) sólo difieren en la orientación del oxidrilo 4 (Esquema 1), y pueden ser reconocidos selectivamente por cada sitio de unión de ABL, como es posible observar a partir de la evidencia experimental aportada por la cristalografía de rayos X.¹ No obstante, a pesar de estos resultados, resulta difícil poder dar una explicación detallada del fenómeno de selectividad epimérica en ambos CBSs. Con el objetivo de poder encontrar una explicación atomística y dinámica de la selectividad, recurriremos a las

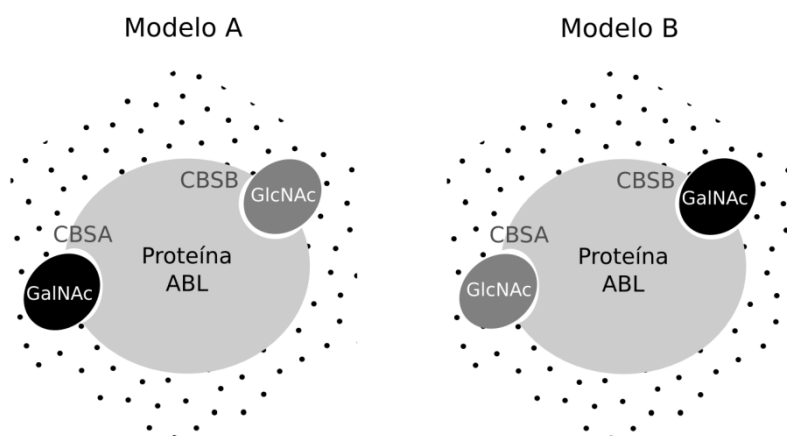
herramientas aportadas por la simulación computacional. De este modo hemos simulado dos modelos, que llamaremos a partir de ahora, modelo A y modelo B.

El modelo A es una simulación por dinámica molecular a presión constante (1 bar) y temperatura constante (298 K) de la proteína ABL inmersa en una caja octaédrica de moléculas de agua, con GalNAc y GlcNAc unidos a los sitios CBSA y CBSB respectivamente (como se observa experimentalmente).²



Esquema 1. Estructura de N-acetilglucosamina y N-acetilgalactosamina.

Por otro lado, el modelo B es una simulación en las mismas condiciones de presión y temperatura que el modelo A pero con la diferencia que en CBSA hemos unido GlcNAc y en CBSB GalNAc, es decir, las condiciones de unión de ligandos que no se observa experimentalmente (Esquema 2) con el objetivo de buscar poner en evidencia la diferencia en la interacción de los ligandos con la proteína en los CBSs.



Esquema 2. Representación esquemática de los dos modelos simulados

Práctica

En la carpeta **WorkshopDM** encontrará tres subcarpetas: **modeloA**, **modeloB** y **modeloC** (este último modelo es una simulación por dinámica molecular de la apo-proteína, es decir, el receptor sin sus ligandos en las mismas condiciones que modelo A y modelo B) y en cada una de ellas habrá tres archivos:

Un archivo de extensión **.pdb**, el cual contiene la información tridimensional estática de los complejos. Un archivo de trayectoria de extensión **.binpos** en el que se han grabado fotos de la evolución del sistema en función del tiempo cada 200 ps. Y por último un archivo que contiene información física de los tipos de átomo de los complejos (carga, radio, etc.) que nos será útil para hacer cálculos de energía de interacción de los azúcares con la proteína.

En este trabajo práctico, nos propondremos dilucidar las razones de la especificidad de ABL para sendos hidratos de carbono en sus respectivos sitios, para ello les sugerimos una lista de actividades que lo ayudarán a analizar los resultados de las simulaciones por dinámica molecular.

1) Visualización

Lo primero que haremos será cargar en VMD la estructura de ambos complejos de la siguiente manera:

Menú **File** → **New molecule**

Cargamos el archivo **modeloA.pdb**. Como puede observarse, se trata de una vista estática del sistema. A continuación haremos clic derecho sobre el nombre del sistema cargado anteriormente y en el menú desplegable seleccionaremos la opción **Load data into molecule**, esto nos permitirá cargar la trayectoria de la dinámica molecular (archivo **modeloA.binpos**).

Una vez realizado esto, veremos en la ventana principal de VMD que se han cargado 750 frames de dinámica.

Repetiremos el paso 1 para el Modelo B.

Dado que por el momento no nos interesa ver la caja de aguas, solo seleccionaremos la proteína y el ligando. Para ello iremos al menú **Graphics** → **Representations** y en el desplegable **Drawing method** elegiremos **New Cartoon**. Esta representación simplifica mucho la visualización de la estructura de la proteína. El desplegable **Coloring Method** mediante la opción **ColorID** nos permite definir un color para identificar cada modelo. Para visualizar el ligando haremos clic en **Create rep** y en el campo **Selected Atoms** escribiremos **resname OVA ROH** para GalNAc, y **resname OYA ROH**

para GlcNAc. Una representación adecuada para el ligando se consigue eligiendo como Drawing method **Licorice** o **CPK**.

Como vemos, las dos proteínas no están en la misma posición, lo cual dificulta el análisis. Es por ello que deberemos alinear ambas estructuras del siguiente modo:

Iremos al menú **Extensions** → **Analysis** → **RMSD Trajectory Tool**

Allí seleccionaremos la opción **backbone**, y luego haremos clic en el botón **Align**. Desde esta misma ventana podemos graficar el RMSD de la selección tildando la opción **plot** y haciendo clic en el botón **RMSD**. Esta función, que se detallará más adelante, nos será de utilidad cuando hagamos el análisis estructural.

Ahora que ya tenemos las estructuras bien alineadas podemos comenzar con el análisis.

2) Análisis Estructural

Vamos a posicionarnos en la primera foto de la dinámica y correr la película para ambos complejos. Observar el grado de movilidad de los ligandos. ¿Cuál está más retenido por la proteína? ¿Alguno de ellos se va del CBS en el lapso de la dinámica?

A) Ahora que ha visto de manera global el comportamiento dinámico de los sistemas, establezca cual es la interacción proteína-carbohidrato que es más susceptible al cambio de ligando. Para ello grafique la distancia de los OH de los ligandos con aquellos átomos de la proteína con los que establecen puentes de hidrógeno (para ello tenga en cuenta los átomos de O y N de los aminoácidos).

B) Una manera muy utilizada de estimar la movilidad de las moléculas en sistemas estudiados por dinámica molecular es el cálculo de la raíz cuadrada de la desviación cuadrática media, RMSD (Root Mean Square Deviation), que es una medida de la movilidad media de todos los átomos involucrados en el cálculo.

- i) Calcule el RMSD vs. Tiempo para la proteína,
- ii) Repita el cálculo pero para los ligandos

¿Qué conclusión puede sacar al respecto?

3) Análisis Energético

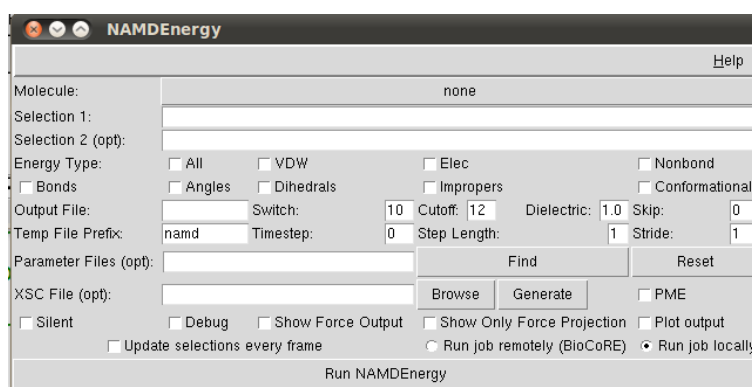
Una manera de estimar la factibilidad de que dos moléculas se unan y/o hacer un análisis comparativo de que ligando se une más fuertemente a un dado receptor, es necesario hacer el cálculo de la energía libre de unión (ΔG_B) entre ambas biomoléculas, en donde se incluye la influencia de todo el sistema y el entorno.³ Si bien existen numerosos modelos teóricos que nos

permiten estimar el ΔG_B a partir de métodos computacionales, los mismos escapan al objetivo de este curso.

No obstante, nos podemos proponer en este punto calcular la energía de interacción de los ligandos con la proteína, con el objetivo de poder observar, al menos de manera aproximada, si el ligando que se observa experimentalmente tiene una mejor energía de interacción con el CBSA (entenderemos que mejor energía de interacción es energía más negativa).

Para ello, iremos a la solapa **Extensions** → **NAMD Energy**

En donde se observará la siguiente ventana:



En la primera línea (**Molecule**), podrá seleccionar el sistema al cual le quiera computar la energía. En la segunda y la tercera (**selection1** y **selection2** respectivamente) se seleccionarán aquellas partes del sistema que se quieren evaluar. Por ejemplo, si queremos calcular la energía de interacción de la N-Acetil-Galactosamina con la proteína, donde dice **molecule**, seleccionamos **modeloA** y donde dice **selection** escribimos lo siguiente:

Selection1: rename 0VA ROH

(con esto le estamos diciendo que seleccione a la N-Acetil-Galactosamina)

Selection2: protein

(y con esto último le decimos que seleccione únicamente a la proteína sin las aguas que la rodean y sin el ligando)

Luego en la línea donde dice **Energy Type**, elegiremos los términos que queremos que se tengan en cuenta para el cálculo de la energía.

Para este caso, nosotros podemos tildar **All** (donde se calculará la energía total), **VDW** (es el cálculo de la energía de Van der Waals) y **Elec** (que es la energía de interacción electrostática entre la porción del sistema que elegimos en **selection 1** y **selection2**).

A continuación, seleccione **plot output** (abajo a la izquierda encontrará la solapa) para que le imprima un gráfico de los distintos términos de la energía que ha seleccionado en función del número de fotos y finalmente presione en la solapa **RUN NAMD Energy** para que comience a realizar el cálculo.

a) Calcule la energía de interacción de VDW, Electrostática y Total para GalNAc y GlcNAc en el CBSA tanto en el modeloA como en el modeloB.

¿Qué conclusión puede sacar al respecto?

4) Análisis de las aguas

Es sabido que las moléculas de agua juegan un rol fundamental en el reconocimiento entre biomoléculas, dado que pueden puentear la interacción entre las mismas o bien ejercer un efecto sobre la energía de unión global por el desplazamiento de las mismas desde el sitio de reconocimiento de ligando del receptor hacia el seno del solvente.⁴

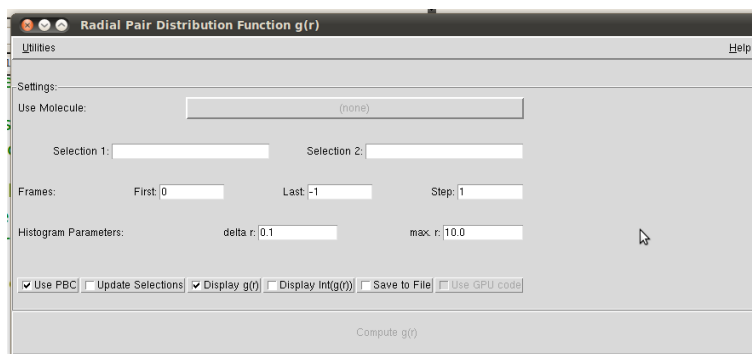
En este trabajo práctico, nos concentramos en estudiar las moléculas de agua que están cercanas al OH4 de los azúcares para poder dilucidar si las mismas juegan o no un rol crucial en la especificidad epimérica que muestra ABL.

Para ello, estudiaremos la solvatación de los residuos cercanos al OH4 de GalNAc (átomo de O carbonílico de la Gly 48 y N ϵ de la histidina 71) de dos maneras distintas:

a) Visualmente con el VMD, representaremos los las moléculas de agua a menos de 4 Å de los mismos. ¿Observa alguna molécula de agua que le llame la atención?

b) Calcularemos la función de distribución radial entre los átomos de O del solvente y el O carbonílico de la Gly 48 por un lado y el N ϵ de la His 71 por el otro, en los modelos A, B y C con el objetivo de tener una idea de la probabilidad de encontrar moléculas de agua en torno a estos residuos.

Para ellos vaya a: **Analysis** → **Extensions** → **Radial Pair Distribution Function g(r)** y en **selection1** escriba el residuo clave de la proteína (**resid 48 and name O**) y en **selection2**, seleccione todos los átomos de oxígeno de las moléculas de agua circundantes (**resname WAT and name O**).



- 1) Anote la distancia más probable a la que se encuentran presentes moléculas de aguas para los 3 modelos y para los dos residuos (Gly48 e His71).
- 2) ¿Hay moléculas de agua que se desplazan selectivamente en el proceso de unión de ligandos? Puede visualizarlas en el VMD.
- 3) ¿Hay moléculas de agua que son claves para puentear la interacción entre ambas biomoléculas, en el modelo A que no es posible observarlas en el modelo B?

Referencias

1. Carrizo, M.E. *et al.* The antineoplastic lectin of the common edible mushroom (*Agaricus bisporus*) has two binding sites, each specific for a different configuration at a single epimeric hydroxyl. *The Journal of biological chemistry* **280**, 10614-23 (2005).
2. Gauto, D.F., Di Lella, S., Estrin, D. a, Monaco, H.L. & Martí, M. a Structural basis for ligand recognition in a mushroom lectin: solvent structure as specificity predictor. *Carbohydrate research* **346**, 939-48 (2011).
3. Gilson, M.K., Given, J. a, Bush, B.L. & McCammon, J. a The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical journal* **72**, 1047-69 (1997).
4. Gauto, D.F., Di Lella, S., Guardia, C.M. a, Estrin, D. a & Martí, M. a Carbohydrate-binding proteins: Dissecting ligand structures through solvent environment occupancy. *The journal of physical chemistry. B* **113**, 8717-24 (2009).