

CAPÍTULO 5

CÁLCULOS COMPUTACIONALES EN MACROMOLÉCULAS BIOLÓGICAS

Marcelo D. Costabel y Fernando Zamarreño

Introducción

Las macromoléculas biológicas constituyen sistemas complejos que pueden adoptar diferentes estados y funciones, sosteniendo un comportamiento dinámico irregular. Para establecer modelos computacionales que respondan a estos requerimientos, debe plantearse la utilización de programas basados en leyes y teorías con fundamentos matemáticos, físicos y químicos, lo que representa un ejemplo de problema multidisciplinario por excelencia.

Al mismo tiempo, el modelado molecular contiene una dosis interesante de arte; donde podemos encontrarnos con un boceto inicial de líneas simples o una figura detallada de pintura real, todo dependiendo del conocimiento previo que tengamos sobre el sistema bajo estudio; y como dijo Pablo Picasso: “el Arte es una mentira que nos ayuda a comprender la realidad”.

5.1. Orígenes

Desde siempre, el hombre se interesó en la descripción de los fenómenos naturales y el consiguiente entendimiento de las estructuras subyacentes que son partícipes de los mismos. No obstante, los comienzos del modelado molecular como hoy lo conocemos, se remontan a comienzos del siglo XX, donde la cristalografía de Rayos X emergente en esos días, resultó una herramienta crucial para desentrañar la estructura molecular de los sistemas biológicos.

Sin embargo la descripción inicial no resultaba sencilla y como el trazado en papel sólo permitía una figura bidimensional, el único camino posible para el armado 3D provino del diseño de mapas de densidad electrónica con materiales como plastilina y madera balsa y la utilización de kits para el armado de modelos de esferas y palos que permitieron una figura más realista de las estructuras en estudio (Figura 5.1.).



© A. Barrington Brown/Photo Researchers, Inc.

Figura 5.1. *Watson y Crick mostrando el modelo estructural de la doble hélice de DNA, armado con alambre.*

Seguramente el trabajo de Watson, Crick y Franklin en la resolución de la estructura del DNA por técnicas cristalográficas (Watson, 1953) es un hito en la historia del modelado molecular de macromoléculas. A partir de allí, la cristalografía y el modelado molecular crecieron en simultáneo.

El aporte siguiente, sin duda, lo brindó la computación. El advenimiento de la informática y el continuo desarrollo de las computadoras, permitió el planteo de modelos cada vez más precisos y la posibilidad de una visualización de los mismos en graficas de tres dimensiones. Pero la computación no sólo permitió el armado de un modelo y la figura correspondiente, también aportó el entorno de trabajo para aplicar el cálculo a sistemas con decenas, centenares o miles de átomos y de esa forma considerar que tipo de interacciones involucran esos

sistemas, dando lugar a descripciones estáticas y dinámicas que dieron origen a lo que hoy conocemos como simulación computacional de macromoléculas y nos permiten inferir o describir el complejo entramado en la relación estructura-función de moléculas biológicas.

5.2. Modelo Inicial

El punto de inicio para el estudio de una macromolécula biológica por técnicas computacionales, es generar ese modelo que permita definir las posiciones de los átomos en el espacio respecto a un sistema de coordenadas cartesianas. Este modelo se obtiene a partir de técnicas experimentales como son la Resonancia Magnética Nuclear (RMN) o la cristalografía de Rayos X. De todas formas, cabe aclarar que, en aquellos casos donde la posibilidad del experimento se ve frustrada, es factible la realización de un modelo que permita, a partir de una estructura conocida, armar el conjunto de coordenadas de una nueva molécula por medio de un trabajo basado en la homología de las secuencias de aminoácidos que las componen.

Dada la importancia de contar con estas estructuras fue necesario almacenarlas en un sitio de fácil acceso. Por esto, las estructuras resueltas experimentalmente se encuentran depositadas en una base de datos denominada Protein Data Bank (PDB) (Berman, 2000). PDB fue creado en 1971 y contó en sus inicios con 7 estructuras de moléculas depositadas y ha crecido hasta contar con más de 70000 estructuras en la actualidad; lo cual la constituye en el punto de referencia por excelencia dentro de la biología molecular estructural.

5.3. Estudios Computacionales de la Estructura Biomolecular

Conocida la estructura, el paso siguiente es desentrañar el funcionamiento de la misma dentro del ámbito fisiológico en que se instala. La descripción de

estos sistemas biológicos a nivel molecular ha sido durante mucho tiempo ámbito exclusivo de diferentes técnicas experimentales, que permitieron dilucidar los mecanismos involucrados. El inconveniente suele ser que estos experimentos requieren, en muchos casos, mucho tiempo de implementación y un análisis previo minucioso para garantizar la optimización de los recursos que se aplican para su realización.

En ese contexto, el surgimiento de la informática durante el siglo XX, dio lugar, entre una innumerable variedad de utilidades, a la aplicación de la misma en el análisis de complejos biomoleculares.

Durante las primeras décadas este análisis estuvo orientado a la verificación de lo observado de forma experimental; pero en estos momentos se ha convertido en una herramienta predictiva, orientadora del trabajo *in vitro* y ha trasladado el paradigma a una nueva biología *in silico*.

Dado que la posibilidad del cálculo computacional crece día a día gracias al progreso continuo en el hardware, los programas aplicados al análisis de sistemas biológicos, en los cuales se considera un número elevado de átomos, es cada vez más variado. En este capítulo presentamos algunos de los estudios computacionales involucrados en el modelado molecular y que pueden aportar información en la relación estructura-función de biomoléculas.

5.4. Modelo y Energía

Contar con la estructura tridimensional de una molécula es el paso inicial necesario para un estudio computacional. A partir de allí, conocer la Energía del sistema nos da la posibilidad de evaluar el funcionamiento bajo distintas condiciones. En este sentido, la formulación adecuada de la función Energía potencial resulta fundamental para que el modelo planteado sea efectivo. A su vez, la combinación entre el modelo planteado y los algoritmos empleados es fundamental para solucionar los problemas de sistemas a gran escala.

Cuando se plantea entonces la Energía, una serie de cuestiones deben ser tenidas en cuenta. En primer lugar, debe considerarse el espacio de

configuraciones. Este espacio de configuraciones para un sistema molecular es de $3N-6$ grados de libertad, (donde el valor $3N$ corresponde a los desplazamientos espaciales posibles para cada uno de los N átomos que componen la molécula, y las 6 restricciones se deben a los movimientos de traslación y rotación conjunta que no implican movimientos relativos de los átomos) lo cual, para una proteína es del rango de 10^4 , y mayor aún al considerar hidratación explícita dentro del modelo. Es por esto que, dentro del modelo, se aplican restricciones específicas que dan lugar a la reducción de variables que permiten plantear la forma funcional de la Energía en un espacio cartesiano.

La Energía potencial dentro de un modelo molecular es la combinación de varios términos:

$$E = E_{dist} + E_{ang} + E_{tor} + E_{LJ} + E_{coul} \quad \text{Ec. 5.1.}$$

$$E_{dist} = \sum_{ij} S_{ij} \left(r_{ij} - \bar{r}_{ij} \right)^2 \quad \text{Ec. 5.2.}$$

$$E_{ang} = \sum_{ijk} K_{ijk} \left(\cos \theta_{ijk} - \cos \bar{\theta}_{ijk} \right)^2 \quad \text{Ec. 5.3.}$$

$$E_{LJ} = \sum_{ij} \left(\frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \right) \quad \text{Ec. 5.4.}$$

$$E_{coul} = \sum_{ij} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \quad \text{Ec. 5.5.}$$

Los dos primeros términos corresponden a la Energía de enlace interatómica (longitud y ángulo), el tercero es un potencial torsional, luego tenemos un potencial de Lennard –Jones que modela la repulsión a distancias cortas y la

atracción a distancias grandes y un potencial Coulombiano entre las partículas cargadas del sistema.

Si bien estas relaciones representan aproximaciones, las mismas son razonables en el caso de las macromoléculas.

5.5. Electroestática de Macromoléculas

En los últimos años ha habido un interés renovado por la Electroestática en la biología, debido a tres factores básicamente:

1. El rol muy importante que juegan dichas interacciones en la caracterización de las propiedades estructurales y funcionales de las macromoléculas,
2. Los avances en la resolución numérica de las ecuaciones básicas que gobiernan estos fenómenos, y
3. El incremento del poder de cálculo que hizo factible crear herramientas computacionales que implementen los algoritmos numéricos en tiempos de cálculo razonables.

La electrostática clásica está regida por la Ecuación de Poisson:

$$-\nabla[\varepsilon(r)\nabla\phi(r)]=\rho(r) \quad \text{Ec. 5.6.}$$

Donde $\varepsilon(r)$ es la permitividad del medio, $\phi(r)$ es el potencial electrostático y $\rho(r)$ es la densidad de carga.

En los casos particulares en los cuales podemos suponer que la permitividad es constante e igual a 1, la solución de la ecuación para este potencial es:

$$\phi(r) = \int \rho(r') \frac{1}{|r - r'|} dr' \quad \text{Ec. 5.7.}$$

Que es la Ley de Coulomb y donde

$$\rho(r) = \sum q\delta(r - r'), \quad \text{Ec. 5.8.}$$

es la densidad de carga correspondiente a un conjunto de N cargas puntuales ubicadas en los sitios atómicos.

La solución del sistema para macromoléculas con forma arbitraria, implica la utilización de métodos numéricos complejos. En estos métodos el espacio continuo se transforma en una matriz discreta compuesta de puntos equiespaciados. El dominio del problema entonces se divide en dos regiones. El volumen asignado al solvente se define como la unión de los volúmenes accesibles a esferas del tamaño de una molécula de agua de radio 0,14 nm que no ingresan al espacio reservado a los átomos de las macromoléculas. Al volumen correspondiente al solvente se le asigna una constante dieléctrica de 78,54 y al volumen correspondiente a las moléculas se les asigna una constante igual a 2.

Cuando las macromoléculas se encuentran en solución, debemos considerar no sólo la presencia de agua, sino también los iones inmersos en la misma y su distribución en la solución. La descripción de un sistema de estas características corresponde a la Ecuación de Poisson-Boltzmann

$$-\nabla[\varepsilon(r)\nabla\phi(r)] = \rho_m(r) + \rho_s(r) \quad \text{Ec. 5.9.}$$

Suponiendo que tenemos una solución salina con cationes y aniones, la densidad de carga en la solución la escribimos como

$$\rho_s(r) = -2ex \sinh\left[\frac{e\phi(r)}{kT}\right] \quad \text{Ec. 5.10.}$$

Con lo cual obtenemos la Ecuación de Poisson – Boltzmann no linealizada:

$$-\nabla[\varepsilon(r)\nabla\phi(r)] = \rho_m(r) - 2ex \sinh\left[\frac{e\phi(r)}{kT}\right] \quad \text{Ec. 5.11.}$$

La cual, con alguna aproximación la podemos escribir como

$$-\nabla[\varepsilon(r)\nabla\phi(r)] = \rho_m(r) - \varepsilon(r)\kappa^2\phi(r), \text{ donde} \quad \text{Ec. 5.12.}$$

$$\kappa^2 = \frac{2Ie^2}{\varepsilon(r)kT} \quad , \text{ con una Fuerza Iónica } I = \frac{1}{2} \sum n \left(\frac{q}{e} \right)^2 .$$

A partir de este modelo y especificados los parámetros adecuados se resuelve el sistema discreto y se asignan los valores de potencial a cada uno de los puntos del reticulado, lo cual permite calcular la Energía Electroestática del sistema como

$$G_E = \frac{1}{2} \sum q_i \phi_i \quad \text{Ec. 5.13.}$$

La relevancia de las interacciones electrostáticas es particularmente notable para las moléculas con carga neta no nula. En el caso de las proteínas han sido reconocidos, por ejemplo, efectos funcionales de importancia producidos por los dipolos correlacionados de las hélices alfa. Además, en problemas que involucran la presencia de una membrana biológica muchas de las propiedades estructurales de las mismas (rigidez, estabilidad, transición lateral de fase y dinámica) dependen sustancialmente de las interacciones electrostáticas.

Dada la naturaleza ubicua de estas interacciones electrostáticas en los sistemas biomacromoleculares, se ha trabajado fuertemente en el desarrollo de métodos computacionales para el esclarecimiento de los mecanismos y procesos en los cuales subyace una componente electrostática.

Varios programas han sido desarrollados para el cálculo de interacción electrostática en moléculas de interés biológico. Uno de ellos es llamado APBS (Adaptive Poisson Boltzmann Solver) (Baker, 2001).

APBS resuelve la ecuación de Poisson-Boltzmann a partir del método de elementos finitos, donde una porción del espacio continuo se transforma en una matriz discreta para permitir las aproximaciones correspondientes, estableciéndose condiciones de contorno determinadas.

Una vez especificados todos los parámetros se emplea para la obtención de la solución un algoritmo iterativo de enfoque. Refinando el cálculo en cada ciclo y utilizando la solución anterior como punto de partida. La salida de este proceso

consiste en la obtención de los valores de potencial para cada uno de los puntos de la red propuesta.

Un caso de aplicación del cálculo de la componente electrostática de la Energía en macromoléculas es el estudio de la interacción entre la proteína que une ácidos grasos (FABP por sus siglas en inglés) y una membrana biológica y, a partir de los resultados obtenidos, se puede establecer la clasificación de las FABPs de acuerdo a su mecanismo de interacción con la membrana (Zamarreño, 2012) (Figura 5.2).

5.6. Dinámica de Macromoléculas

Así como el cálculo de la componente electrostática de la Energía nos da una visión estática de las posibles configuraciones energéticamente más favorables para un sistema biomolecular, la Dinámica Molecular nos muestra un sistema en movimiento generando la posición y la velocidad de cada partícula del sistema en cada instante de tiempo y nos permite observar el espacio de configuraciones posibles en su totalidad. Además, mediante la Mecánica Estadística podemos pasar de esta información microscópica a la obtención de magnitudes macroscópicas observables.

Una descripción rápida del modelo nos muestra que, dado un sistema compuesto por N partículas, dentro de un volumen V y con una energía fija E , las posiciones y velocidades definen un espacio de fases de $6N$ dimensiones. Obtener la posición y la velocidad de cada una de las partículas, en cada instante, significa obtener la trayectoria de un punto en el espacio de las fases en función del tiempo.

Por medio del formalismo de la simulación lo que se hace es generar una sucesión de diferentes estados (puntos) del espacio de fases, compatibles con las condiciones externas macroscópicas (N , V , E) en este caso, sobre los cuales se toman los promedios.

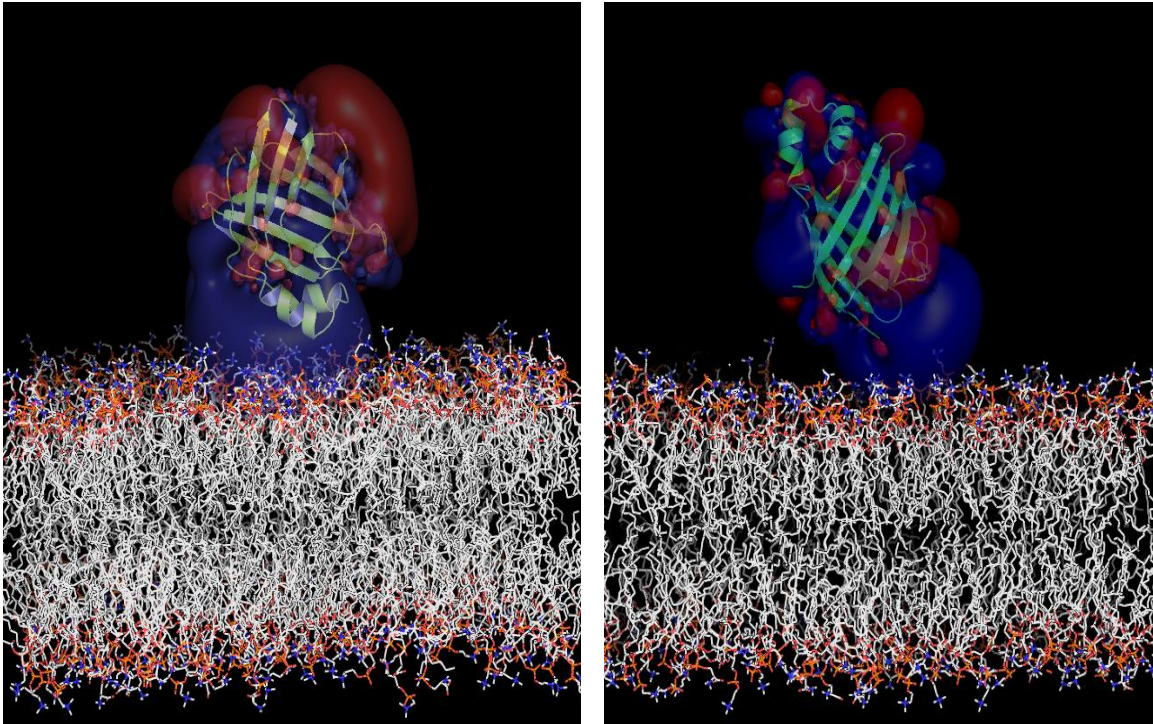


Figura 5.2. Enfoque electrostático inicial de dos diferentes FABP (IFABP y LFABP; izquierda y derecha respectivamente), mostrando como cada una de ellas inicia su interacción con la membrana de manera distinta.

La elección del conjunto bajo el cual llevar a cabo la simulación está dictada fundamentalmente por el tipo de problema a tratar. Los promedios estadísticos pueden llevar a pequeñas diferencias en los diferentes conjuntos, pero éstas desaparecen en el límite termodinámico, que se alcanza incluso con unos pocos cientos de partículas. Sin embargo la elección del conjunto sí influye al momento de calcular las fluctuaciones cuadráticas medias de las magnitudes termodinámicas. Estas permiten calcular, por ejemplo, la capacidad calorífica o el módulo de elasticidad.

Los experimentos típicos de Dinámica Molecular se realizan en el conjunto microcanónico (N, V, E). Mediante algoritmos que acoplan adecuadamente el sistema a un baño térmico o a un “baño” de presión, se pueden realizar simulaciones en el canónico (N, V, T) o en el conjunto isotérmico-isobárico (N, P, T) a presión y temperatura constante, respectivamente.

Típicamente el estudio mediante DM tiene tres fases:

FASE I: Inicialización: Se especifican las condiciones iniciales, la topología, condiciones de borde, las condiciones periódicas de contorno y se describe el sistema a simular.

FASE II: Ejecución de la Dinámica Molecular: Se integran numéricamente las ecuaciones de movimiento que gobiernan el sistema.

Veamos la forma general de un algoritmo de Dinámica Molecular:

1. Se leen los parámetros que especifican las condiciones de la corrida: Temperatura inicial, el número de partículas, la densidad, el intervalo de integración, tiempo total de simulación.

2. Se inicializa el sistema, esto es, se asignan las posiciones y las velocidades iniciales de todas las partículas del sistema. Se leen los parámetros del potencial $V(r)$.

Deben repetirse los siguientes pasos, hasta llegar al último:

3. Se calculan las fuerzas sobre todos los átomos no ligados, más las fuerzas debidas a todas las interacciones ligadas, que pueden depender de 1, 2, 3 o 4 átomos. Se calculan las restricciones, o fuerzas externas si las hubiera. Se calculan las energías cinéticas y el tensor de presiones.

4. El movimiento en el instante de cada átomo se obtiene resolviendo numéricamente las ecuaciones de Newton.

5. Si corresponde se imprimen las posiciones, velocidades, fuerzas, energías, presiones.

Un paso fundamental es la evaluación de la Energía Potencial que resulta la parte más costosa en tiempos de computación.

FASE III: Análisis de los resultados: Se evalúan las propiedades físicas tomando como dato la información de la trayectoria del sistema, que es la *salida* de la etapa anterior. El análisis se realiza comúnmente tomando promedios temporales sobre las diferentes configuraciones. Los valores medios así obtenidos, considerados durante un tiempo suficientemente largo, representan promedios termodinámicos suponiendo un comportamiento ergódico del sistema, donde consideramos que el promedio temporal es igual al promedio espacial.

Las propiedades microscópicas de interés que se estudian a partir de un cálculo de Dinámica Molecular son: distancia entre átomos, superficie expuesta al solvente, desvíos cuadráticos medios, Radios de Giro y Estudio de Puentes de Hidrógeno.

Con todos estos datos podemos, entonces evaluar las características del sistema y obtener un conocimiento más detallado de su comportamiento dinámico.

Siguiendo con el ejemplo anterior, podemos analizar el mecanismo de interacción de FABP con una membrana biológica, observando en este caso la dinámica en la aproximación y los movimientos acoplados en los componentes estructurales involucrados en la interacción (Zamarreño, 2012).

Al igual que para el cálculo electrostático, en el caso de la Dinámica Molecular, podemos encontrar distintos programas para analizar el problema.

En este caso se utilizó el paquete GROMACS (Berendsen, 1995), de amplia distribución y reconocida eficiencia. Como resultado se puede observar una aproximación diferenciada para diferentes FABPs y la formación de un complejo energéticamente estable (Figura 5.3).

Conclusiones

En este capítulo se presentó un conjunto no exhaustivo de técnicas computacionales aplicadas a sistemas moleculares de interés biológico. El alcance y objetivo no ha sido el de ahondar en las técnicas, ni dar un detalle acabado de cálculos posibles, sino presentar posibles alternativas y referenciar ejemplos de resultados obtenidos con este tipo de metodologías.

Seguramente el avance incesante de la informática permitirá el desarrollo de programas cada vez más precisos que nos aproximen a un modelo de características cada vez más cercanas a la realidad, en un continuo acercamiento a la descripción de la relación estructura-función en complejos moleculares de interés biológico.

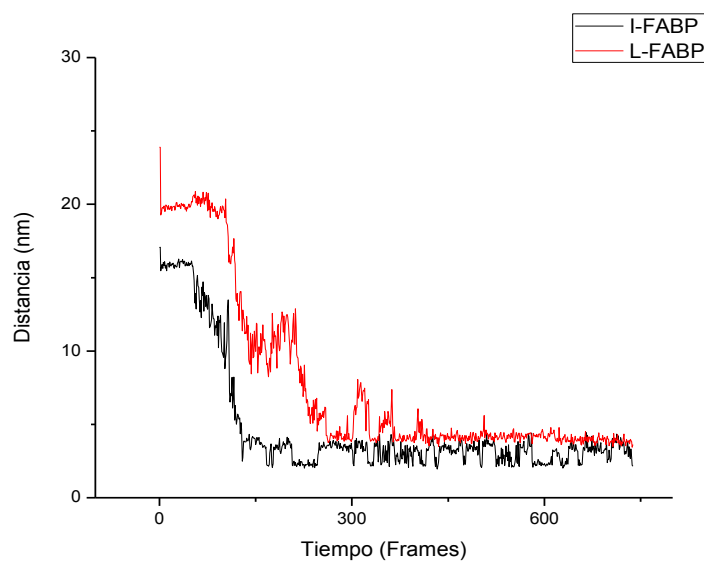
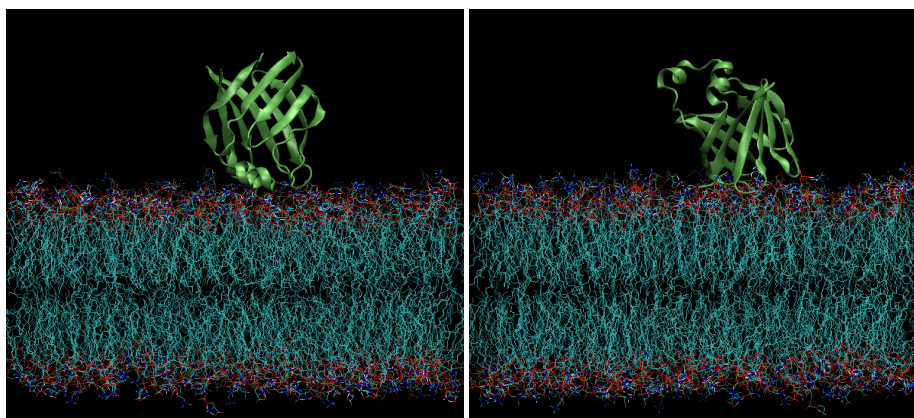


Figura 5.3. Posición final de la interacción entre IFABP (izquierda) y LFABP (derecha), y una membrana biológica modelada a partir de una simulación realizada por Dinámica Molecular. El esquema inferior muestra las curvas de aproximación que indican la distancia entre un átomo de la molécula y uno de la membrana, en función del tiempo de simulación, para cada uno de los casos.

En las referencias bibliográficas que se dan a continuación, podrán encontrar enumerados libros con información detallada que permitirán completar lo desarrollado en el presente capítulo (Leach, 2001; Hinchliffe, 2008; Schlik, 2010).

Bibliografía

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98, 10037-10041, 2001. <http://www.poissonboltzmann.org/apbs/>

Berendsen, et al. GROMACS: A message-passing parallel molecular dynamics implementation *Comp. Phys. Comm.* **91**: 43-56, 1995 <http://www.gromacs.org/>

Berman HM, Westbrook J, Feng Z., Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Research*, 28: 235-242, 2000. www.pdb.org

Hinchliffe A. *Molecular Modeling for Beginners*. Second Edition. Wiley and Sons Ltd, United Kingdom, 2008.

Leach A. *Molecular Modeling Principles and applications*. Second Edition Glaxo Wellcome Research and Development. Prentice Hall- Pearson Education Limited. Edinburgh, 2001.

Schlick T. *Molecular Modeling and Simulation: an interdisciplinary guide*. Second Edition. Springer, New York, USA, 2010.

Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171 (4356):737-738, 1953.

Zamarreño F, Herrera F, Córscico B, Costabel MD. Similar structures but different mechanisms. Prediction of FABPs-Membrane interaction by electrostatic calculation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818(7):1691-1697, 2012.