

Dinámica Molecular de Biomoléculas

Molecular Dynamics of Biomolecules

ARTÍCULO

Juliana Palma

Universidad Nacional de Quilmes, Argentina. Contacto: juliana@unq.edu.ar

Gustavo Pierdominici-Sottile

Universidad Nacional de Quilmes, Argentina. Contacto: gsottile@unq.edu.ar

Recibido: abril de 2023

Aceptado: mayo de 2023

Resumen

Para cumplir con su función biológica, proteínas y ácidos nucleicos realizan precisos cambios conformacionales. Los mismos, son producidos por la coordinación de pequeñas fluctuaciones ejecutadas por sus átomos. Curiosamente, a simple vista, estos átomos parecen moverse de manera caótica. Las simulaciones de dinámica molecular (MD, por *Molecular Dynamics*) constituyen una herramienta particularmente versátil para investigar la conexión entre los cuasi-aleatorios movimientos atómicos y los cambios conformacionales. De esta manera, permiten construir puentes que van desde la secuencia a la estructura; y desde allí a la dinámica y la función biológica.

Los fundamentos de las simulaciones MD son simples e intuitivos. Sin embargo, al momento de implementar un estudio particular, es posible tomar muchas malas decisiones que no impedirán realizar el cálculo, sino que sorpresivamente harán que sus resultados sean carentes de cualquier valor. En este artículo, presentaremos los conceptos fundamentales de las simulaciones MD, resaltando aquellos aspectos que son críticos para asegurar su utilidad. Además, discutiremos sus fortalezas y limitaciones, destacando cómo las mismas fueron evolucionando con el tiempo. Asimismo, brindaremos ejemplos paradigmáticos que ilustran las capacidades actuales de esta valiosa herramienta computacional.

Palabras clave: Dinámica molecular; Simulación computacional de biomoléculas; Físicoquímica computacional; Dinámica de proteínas.

Abstract

To fulfill their biological function, proteins and nucleic acids undergo precise conformational changes. These changes are brought about by the coordination of small fluctuations executed by their atoms. Interestingly, at first glance, these atoms seem to move in a chaotic manner. Molecular dynamics simulations (MD) are a particularly versatile tool for investigating the connection between quasi-random atomic motions and conformational changes. In this way, they allow us to build bridges from sequence to structure, and from there to dynamics and biological function.

The fundamentals of MD simulations are simple and intuitive. However, when implementing a specific study, it is possible to make many poor decisions that will not prevent the calculation from being performed but will surreptitiously render its results devoid of any value. In this article, we will present the fundamental concepts of MD simulations, highlighting those aspects that are critical to ensuring their utility. Additionally, we will discuss their strengths and limitations, emphasizing how they have evolved over time. Furthermore, we will provide paradigmatic examples that illustrate the current capabilities of this valuable computational tool.

Keywords: Molecular dynamics, Computational simulations of biomolecules, Computational Physical Chemistry, Protein dynamics.

1. Introducción

Las simulaciones de Dinámica Molecular (MD) son un tipo particular de simulaciones. En ellas utilizamos computadoras para generar trayectorias que muestren, en tiempo real, los movimientos de las moléculas que forman un determinado sistema. Esta metodología ha contribuido enormemente al desarrollo de las ciencias de los materiales, en general. Pero uno de los campos donde su aplicación ha sido más prolífica y esclarecedora es en el estudio de las moléculas esenciales para la vida: proteínas y ácidos nucleicos. Esto se debe a que estas biomoléculas son entidades altamente dinámicas, cuyo funcionamiento depende de su capacidad para adoptar diferentes formas (conformaciones). Obviamente, existen también técnicas experimentales que

apuntan a caracterizar la dinámica de las biomoléculas [Gupta et al., 2022]. Entre ellas podemos mencionar a las espectroscopías de dicroísmo circular y de fluorescencia, la resonancia magnética nuclear (RMN), la dispersión de rayos X de ángulo pequeño (SAXS) y la resonancia de plasmón de superficie (SPR). Sin embargo, las mismas tienen una resolución temporal y espacial más limitada que las simulaciones MD. Por lo tanto, ambos enfoques se suelen complementar, a fin de producir una descripción más detallada y certera de las moléculas bajo estudio.

Como ocurre con todas las técnicas de investigación utilizadas en las ciencias naturales, las simulaciones MD de biomoléculas no son ni infalibles ni todo-poderosas. Por el contrario, pueden arrojar resultados completamente erróneos y existen sistemas y procesos que están más allá de su alcance. Por tanto, es fundamental que todo aquel que quiera utilizar esta metodología, tenga bien claro cuáles son esas falencias y limitaciones. En este artículo describiremos los fundamentos de las simulaciones MD, tratando de hacer foco en los aspectos conceptuales y reduciendo al mínimo el uso de ecuaciones. Por ello, no vamos a discutir las características técnicas de los algoritmos empleados, sino cuál es el objetivo de los mismos. Además, haremos una breve reseña histórica en la que describiremos cómo fueron evolucionando las capacidades de esta herramienta computacional, a fin de posibilitar la realización de simulaciones cada vez más extensas, sobre sistemas más grandes y procesos más complejos. Acompañaremos esta descripción con ejemplos ilustrativos. Cerraremos el artículo comentando las principales desafíos que las simulaciones MD encuentran en la actualidad.

2. ¿Qué es una simulación?

Muchas veces ocurre que conocemos las reglas que determinan el comportamiento de un sistema. Sin embargo, las mismas forman una red tan compleja e intrincada que se pierde de vista la conexión entre dichas reglas y el modo de funcionamiento del sistema. Desafortunadamente, es justamente ése el conocimiento que necesitamos para poder utilizar las leyes básicas de la ciencia para la resolución de problemas particulares. En circunstancias como

ésta, es útil recurrir a algún tipo de simulación. En ellas, podemos manipular a nuestro antojo las reglas que regulan el comportamiento del sistema, para luego observar cómo esos cambios repercuten su funcionamiento.

Para realizar una simulación uno debe primero crear un modelo. Con él, buscamos imitar el comportamiento del sistema bajo estudio. El modelo debería contener los elementos y las reglas que son necesarios para reproducir, fehacientemente, el funcionamiento del sistema o algún aspecto del mismo que queramos caracterizar. Luego, para una situación inicial dada, uno evoluciona el estado del sistema de acuerdo a las reglas establecidas y va registrando cómo van variando los parámetros que se quieren analizar. En esto consiste, justamente, la simulación. Finalmente, uno compara los datos obtenidos de la simulación con los del sistema real, con el fin de evaluar la capacidad predictiva del modelo. Si esta capacidad es elevada, el modelo puede ser utilizado para interpretar y racionalizar resultados experimentales; o incluso para hacer predicciones sobre situaciones que no llegan a observarse experimentalmente. Por otro lado, si la capacidad predictiva del modelo es baja, uno debe modificar sus elementos o a justar sus reglas hasta lograr un funcionamiento satisfactorio. Este proceso, que consiste en detectar qué característica del modelo es responsable de un resultado particular, resulta muy esclarecedor y constituye otra de las maneras mediante las cuales la combinación de “modelado y simulación” permite aumentar nuestro entendimiento sobre el modo de funcionamiento de un sistema.

2.1. Modelos de biomoléculas

Los modelos han tenido un rol muy destacado en el desarrollo de la ciencia. De hecho, casi cualquier persona que haya estudiado alguna ciencia natural ha tenido que lidiar con ellos. Como ejemplo, podemos mencionar a los diagramas de Lewis, para explicar las estructuras moleculares; y al gas ideal, para comprender el comportamiento de la fase gaseosa. La capacidad predictiva y explicativa de ambos modelos es asombrosamente alta. Sobre todo, si uno compara sus simples reglas con las leyes que verdaderamente rigen el comportamiento de esos sistemas (mecánica cuántica, en el caso de la estructura molecular y mecánica estadística, en el caso de la fase gaseosa). Los

ejemplos comentados demuestran que un modelo puede hacer predicciones valiosas aun aunque utilice reglas diferentes de las que verdaderamente gobiernan el sistema. Esta característica es resumida en una frase que se atribuye al matemático y estadístico británico George E. P. Box: “todos los modelos están equivocados, pero algunos son útiles” [Box, 1976].

La biología estructural de macromoléculas también se ha beneficiado enormemente por el uso criterioso de modelos simples (que obviamente estaban “equivocados”). Por ejemplo, Linus Pauling utilizó modelos de bolitas y palitos hechos en madera, para analizar cómo podían acomodarse los aminoácidos de las proteínas a fin de lograr una estructura estable. Así, encontró que las hélices- α y las láminas- β debían ser los principales elementos de la estructura secundaria de estas biomoléculas [Pauling et al., 1951]. Los modelos de Pauling se fundaron en dos componentes básicos: las geometrías de equilibrio de los aminoácidos constituyentes y la suposición de que los puentes de hidrógeno entre los grupos carbonilo y amida imponían restricciones a la distancia y el ángulo entre los enlaces C-O y N-H. Una estrategia similar siguieron Watson and Crick para postular la estructura de doble hélice del ADN [Watson and Crick, 1953]. Conocían las estructuras de las unidades (nucleótidos) que forman estas moléculas. Su trabajo con modelos consistió en ver cómo podían ensamblarse esas unidades para formar macromoléculas estables consistentes con los datos estructurales que Rosalind Franklin por un lado y Maurice Wilkins por el otro, habían obtenido por difracción de rayos-X. Tanto las investigaciones de Pauling sobre las proteínas, como las de Watson, Crick y Wilkins sobre el ADN, fueron galardonadas con el premio Nobel. Este hecho revela en qué medida la utilización de modelos contribuyó al desarrollo inicial de la biología molecular. Como una nota de color (oscuro) debemos agregar que Rosalind Franklin, la única mujer entre las personas involucradas en el trabajo y la que había producido los datos más significativos sobre la estructura del ADN, no fue galardonada con el Nobel. Están quienes dicen que esto se debe a que, al momento de otorgar ese galardón, Rosalind Franklin ya había fallecido¹ y el premio Nobel no se otorga *post-mortem*. Pero también están quienes sostienen que ese es solo un ejemplo más de cómo, durante muchos siglos,

¹Rosalind Franklin falleció en 1958, cuando solo tenía 38 años. El premio nobel a Watson, Crick y Wilkins fue otorgado en 1962.

las contribuciones de las mujeres al desarrollo de la ciencia fueron invisibilizadas.

3. Simulaciones de Dinámica Molecular

Si bien los modelos de “palitos y bolitas” fueron de gran utilidad en los albores de la biología estructural de macromoléculas, es obvio que los mismos tienen serias limitaciones. O como diría G.E.P. Box: fueron útiles pero estaban equivocados. Por ejemplo, sus enlaces no pueden vibrar ni torcerse. Por lo tanto, estos modelos no pueden ser utilizados para analizar los movimientos internos de las moléculas que a su vez son necesarios para que las mismas puedan cumplir su función.

Hay dos aspectos fundamentales en los que los modelos discutidos hasta el momento pueden ser mejorados. Uno de ellos consiste permitir variaciones en las distancias, ángulos y torsiones de enlace, ya que estos terminan generando los cambios conformacionales que se observan en las biomoléculas. Por otro lado, cualquier modelo molecular razonable debería permitir observar cómo ocurren esos cambios conformacionales en tiempo real. O sea, deberíamos poder apreciar la evolución temporal del sistema.

Para lograr el primer objetivo, se crearon los campos de fuerza de Mecánica molecular (MM). Los mismos serán descritos en la sección 3.2. Para el segundo objetivo, es necesario contar con ecuaciones de propagación que permitan predecir el estado del sistema en el futuro, a partir de un cierto estado inicial dado. En principio, estas reglas son perfectamente conocidas. Se trata de la ecuación de Schrödinger dependiente del tiempo, aplicada a todos los núcleos y electrones del sistema. Lamentablemente, este enfoque es inviable para casi la totalidad de los sistemas. Por lo tanto, se hace necesario introducir una serie de aproximaciones que convierten un cálculo exacto pero imposible, en uno con ciertas limitaciones pero susceptible de ser implementado. Estas aproximaciones serán presentadas en la sección siguiente, donde además discutiremos cómo las mismas pueden afectar el resultado de la simulación.

3.1. Transformando un cálculo “exacto” en una Dinámica Molecular clásica

Como dijimos en la sección precedente, las simulaciones MD no utilizan las ecuaciones de propagación exactas sino que introducen diversas aproximaciones. El número y las características de las mismas varía para cada caso particular. En los siguientes párrafos, vamos a presentar las aproximaciones más comúnmente utilizadas. Cada una de ellas limitará de alguna manera la clase de sistemas o de procesos que podemos estudiar. El tipo de simulación resultante se denomina “*all-atom classical MD simulation*”. Esta calificación indica que se trata de una simulación que emula la dinámica de un sistema formado por moléculas; que todos los átomos de esas moléculas serán explícitamente considerados; que los mismos interactúan como lo establece un campo de fuerza MM; y que su estado evoluciona de acuerdo con las leyes de Newton.

La Fig. 1 muestra un esquema de las simplificaciones que conducen desde un cálculo enteramente basado en primeros principios, también llamado cálculo “*ab-initio*” o “exacto”, hasta una simulación de dinámica molecular clásica de todos los átomos. Lo ideal, siempre que sea posible, es la opción (a)

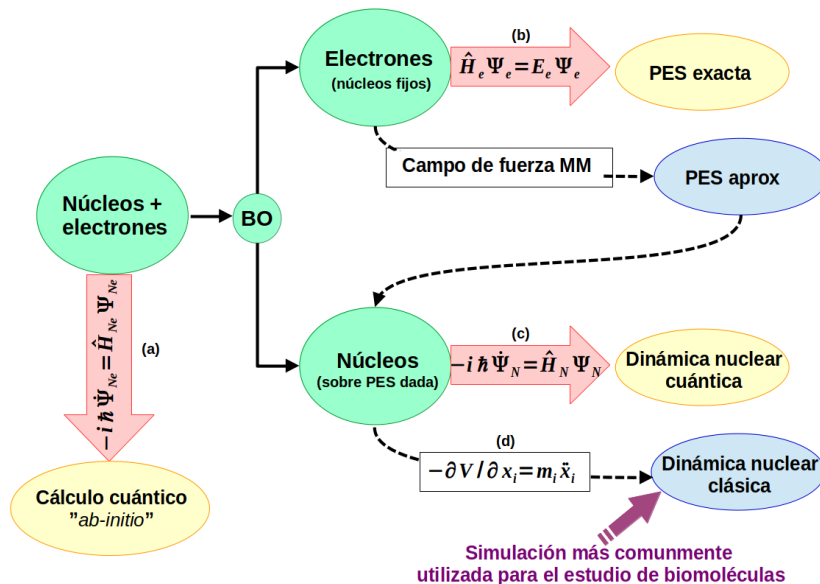


Figura 1: Alternativas existentes para estudiar computacionalmente un sistema molecular.

de dicha figura. Esta implica resolver la ecuación de Schrödinger dependiente del tiempo de manera conjunta, para todos los núcleos y electrones del sistema. Pero como ya anticipamos, incluso con la gran capacidad computacional de la que disponemos en estos momentos, esta estrategia es imposible de llevar a cabo (excepto para sistemas formados por muy poco átomos). Por lo tanto, para poder proceder, es necesario introducir simplificaciones y la primera que se hace consiste en desacoplar el movimiento electrónico del nuclear. Esta estrategia se conoce como “aproximación de Born y Oppenheimer” y se indica como BO en la Fig. 1. La aproximación BO se fundamenta en el hecho de que núcleos y electrones se mueven con velocidades muy distintas debido a sus diferentes masas. Por ello, los electrones se adaptan de manera casi instantánea a las posiciones nucleares. Eso sugiere que podemos, en primer lugar, resolver la ecuación de Schrödinger electrónica para diferentes posiciones (fijas) de los núcleos. De este modo sabemos cómo se distribuye la nube electrónica para cada configuración nuclear. Luego, con esa información, podemos calcular las fuerzas que sienten los núcleos en cada una de esas posiciones (estas son las fuerzas que debemos usar en la ecuación de Schrödinger nuclear). En otras palabras, la resolución del problema electrónico permite calcular la energía potencial de cada configuración nuclear. En muchos textos y artículos, esta energía potencial se denomina PES (por “*Potential Energy Surface*”). Sus derivadas parciales, respecto de las coordenadas nucleares, nos indican las fuerzas que actúan sobre los núcleos.

La aproximación BO supone que el sistema molecular siempre se encuentra en el mismo estado electrónico. Por lo tanto, la misma no puede ser invocada para estudiar los procesos de relajación que tienen lugar después de que una molécula ha absorbido un fotón de luz visible o ultravioleta. Tampoco se pueden estudiar procesos de transferencia de carga. Pero sí es posible analizar cambios conformacionales, la entrada y salida de sustratos al sitio activo de las enzimas, los mecanismos de unión proteína-proteína o proteína-ARN/ADN, la adhesión de moléculas sobre la superficie de una membrana o el pasaje de iones y pequeñas moléculas a través de canales transmembrana [Pierdominici-Sottile et al., 2016, Cossio-Pérez et al., 2019, Shi et al., 2008, Pierdominici-Sottile et al., 2019, Ormazábal et al., 2021]. En otras pa-

labras, la aproximación BO no representa en si misma una gran limitación, excepto porque nos impide estudiar procesos en los que cambia el estado electrónico (estos procesos reciben el nombre de no-adiabáticos).

Lamentablemente, aun después de aplicar la aproximación BO, nos enfrentamos con un cálculo de imposible realización. Esto se debe a que tanto la resolución de la ecuación de Schrödinger exclusivamente electrónica, como la exclusivamente nuclear, tienen un enorme costo computacional y resultan inabordable, excepto para moléculas con muy pocos átomos. Para el problema electrónico, la dificultad reside, en buena medida, en el tipo de cálculo utilizado. Los tratamientos que utilizan los llamados métodos cuánticos semiempíricos o los basados en la teoría de los funcionales de la densidad (DFT), pueden ser utilizados en sistemas de cien o doscientos átomos [Mlýnský et al., 2014]. Esto excluye a la gran mayoría de los sistemas de interés biológico. Y además, aun cuando resulten factibles, estos tratamientos cuánticos no pueden alcanzar los tiempos de simulación requeridos para muestrear fehacientemente el espacio conformacional de la gran mayoría de los sistemas de interés biológico. En otras palabras, todo lo que podríamos “ganar” por tener una PES más exacta, se pierde por un problema de muestreo insuficiente. Por lo tanto, la mayoría de los cálculos MD utilizan PES aproximadas, que evalúan las interacciones entre los núcleos mediante funciones empíricas. Estos son los campos de fuerza que discutiremos en la próxima sección.

Para abordar el problema nuclear, lo que se hace es pasar de una descripción cuántica a una clásica. O sea, en vez de resolver la ecuación de Schrödinger nuclear utilizando la PES disponible, utilizamos la misma PES para propagar las ecuaciones de Newton. Obviamente, este tratamiento impide la observación de cualquier efecto cuántico nuclear. Entre los más relevantes tenemos el efecto túnel, que juega un rol importante en ciertas reacciones de transferencia de hidrógeno. Además, en estos tratamientos clásicos, los movimientos internos de las moléculas no tienen la energía de punto cero. Esto hace que la distribución de la energía total entre los diferentes modos de movimiento difiera de la “verdadera” distribución. Esta discrepancia impacta de manera directa sobre la capacidad calorífica del sistema, ya que la misma depende de cómo está distribuida la energía. El impacto es mayor cuanto menor

es la temperatura, ya que en esas condiciones los efectos cuánticos se vuelven más prominentes. Afortunadamente, esta limitación no afecta la calidad de las simulaciones de biomoléculas, que por lo general se realizan a temperatura ambiente o fisiológica. En conclusión, el principal condicionamiento que surge de dar un tratamiento clásico al movimiento nuclear es que no se pueden estudiar procesos de transferencia de H, H⁺ o H⁻.

Podemos concluir, entonces, que la gran mayoría de las simulaciones de dinámica molecular que se realizan sobre macromoléculas invocan la aproximación BO para separar el movimiento electrónico del nuclear. Luego, modelan la PES sobre la que se mueven los núcleos usando un campo de fuerza empírico de mecánica molecular. Y finalmente, para obtener las trayectorias de los átomos con la PES dada, propagan las ecuaciones de Newton. Como mencionamos anteriormente, estas simulaciones reciben el nombre de “simulaciones de dinámica molecular clásica de todos los átomos”. Con esta denominación, las distinguimos de las simulaciones en las que las moléculas de solvente no son tratadas explícitamente (se dice que tienen el solvente implícito) [Kleinjung and Fraternali, 2014] y de las simulaciones de “grano grueso” (o *coarse graining*) [Joshi and Deshmukh, 2021], en las que las partículas simuladas son grupos de átomos en vez de los átomos individuales. Ambas aproximaciones tienen por objetivo lograr tiempos de simulación más largos. Pero para eso, deben resolver problemas distintos a los enfrentados por las simulaciones *all-atom*. La discusión de esos problemas está fuera del alcance de este artículo.

Por último, debemos mencionar que el tamaño de los modelos utilizados en las simulaciones MD es muchísimo menor que el tamaño de un sistema macroscópico. Esto se debe a que el número de interacciones a evaluar en cada paso de integración es del orden de $\sim N(N - 1)$, donde N es el número de átomos. Obviamente, este número N ha ido creciendo a medida que fue aumentando la velocidad y memoria de las computadoras. No obstante ello, aun en la actualidad, N es un número relativamente pequeño (ver Sección 4). Una consecuencia de esta restricción, es que la proporción entre los átomos que están en la superficie del sistema y los que están en su seno es artifi-

cialmente alta, produciendo efectos de borde.² Para evitar estos efectos, las simulaciones MD utilizan condiciones periódicas de contorno (o PBC por *periodic boundary conditions*) [Zhou and Liu, 2022]. Esta aproximación, consiste en simular una pequeña celda unitaria rodeada de celdas idénticas. La forma de la celda debe ser tal que, cuando la misma se repite en todas las direcciones, no quedan espacios vacíos. La Fig. 2 muestra esquemáticamente las PBC para el caso de un sistema bidimensional.

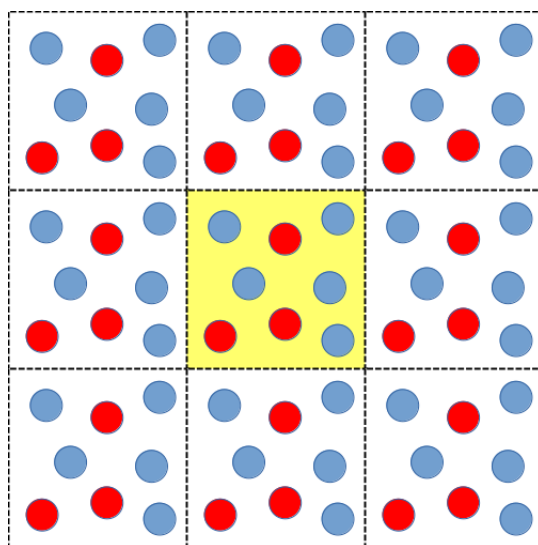


Figura 2: Condiciones periódicas de contorno en un sistema bidimensional. Las PBC son utilizadas para eliminar los efectos de borde causados por el tamaño relativamente pequeño de los sistemas simulados. Las ecuaciones de movimiento solo propagan la posición y velocidad de las partículas de la celda central. Pero las fuerzas que actúan sobre las mismas se evalúan haciendo interactuar a cada partícula de esta celda con sus vecinas más cercanas, ya sea de la misma celda o de una celda circundante.

Durante muchos años, las PBC fueron utilizadas rutinariamente sin tener mayores detractores. Sin embargo, en un artículo publicado en 2018, se reportó que las simulaciones de hemoglobina requerían celdas unitarias mucho mayores que las comúnmente utilizadas [El Hage et al., 2018]. De lo contrario, la forma T de esta proteína resultaba artificialmente inestable. El efecto fue atribuido a la incapacidad de las celdas pequeñas para dar cuenta del

²Los átomos de una interfase se comportan de manera distinta a los del seno de un fluido. En un sistema macroscópico, la proporción entre los átomos de la interfase y los del seno es pequeña. Por tanto, la mayoría de las propiedades que percibimos son las del seno del fluido. Si se aumenta artificialmente la proporción de la interfase el sistema se comportará de manera poco realista.

efecto hidrofóbico. Otros trabajos posteriores, pusieron en duda ese resultado [Gapsys and de Groot, 2019], por lo que el tema permanece aun abierto. En cualquier caso, es claro que estos posibles efectos del tamaño de la celda solo podrían llegar a aparecer en las simulaciones de grandes biomoléculas, en las que el efecto hidrofóbico juega un rol preponderante.

3.2. Los campos de fuerza de mecánica molecular

Los campos de fuerza MM son fórmulas empíricas que buscan emular las interacciones entre los núcleos de un sistema molecular. Al hablar de los mismos, debemos distinguir entre dos características fundamentales. Por un lado, tenemos la forma funcional del campo. Esto es, las ecuaciones que utiliza para calcular la energía potencial de cada configuración. Por otro lado, tenemos su parametrización. La misma está definida por el conjunto de valores asignados a las constantes (parámetros) que aparecen en su forma funcional.

La forma funcional del campo busca ser simple e intuitiva. La simpleza permite que la evaluación de la energía potencial y de las fuerzas derivadas del mismo, tenga un bajo costo computacional. De esta manera, los campos MM permiten alcanzar tiempos de simulación más largos que otras aproximaciones más rigurosas. Por otra parte, el carácter intuitivo de sus términos facilita el diseño de estrategias para la determinación de los parámetros. La ecuación 1 muestra la forma funcional utilizada por los campos de fuerza más comúnmente utilizados en el estudio de biomoléculas [Ponder and Case, 2003]. Además, en la figura 3, puede encontrarse una representación esquemática e ilustrativa de cada una de sus contribuciones.

$$\begin{aligned}
 V_{MM} = & \sum_{i=1}^{N_{en}} \frac{1}{2} k_{r,i} (r_i - r_{i,eq})^2 + \sum_{i=1}^{N_{ang}} \frac{1}{2} k_{\theta,i} (\theta_i - \theta_{i,eq})^2 \\
 & + \sum_{i=1}^{N_{tor}} \sum_n V_{i,n} (1 + \cos(n\omega_i - \delta_i)) \\
 & + \sum_{i=1}^{N_{at}-1} \sum_{j=1}^{N_{at}} 4\epsilon_{ij} \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}.
 \end{aligned} \tag{1}$$

A cada enlace covalente del sistema molecular (interacciones 1-2), el

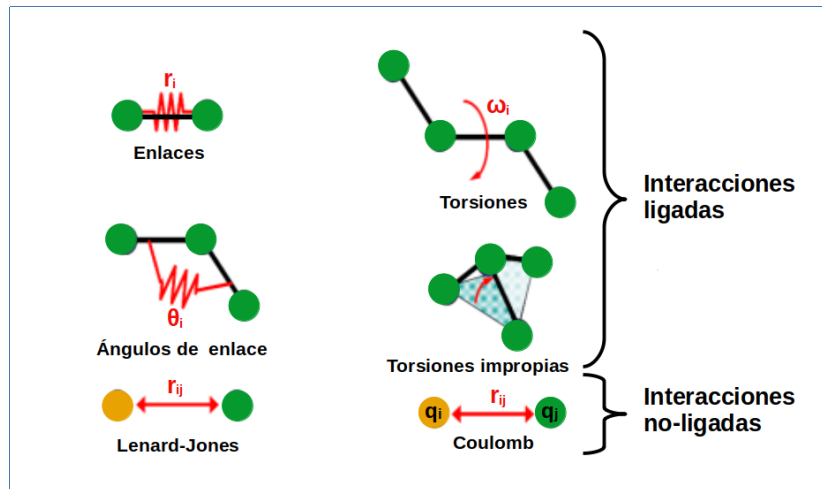


Figura 3: Términos que contribuyen al cálculo de la energía potencial con un campo de fuerza MM como el de la Eq.1.

campo de fuerza de la ecuación 1 le asigna un potencial armónico. Estos términos aparecen en la primer sumatoria de dicha ecuación. Asimismo, se asigna un potencial armónico a cada ángulo entre enlaces covalentes (interacciones 1-3). Estos términos aparecen en la segunda sumatoria de la ecuación 1. Las torsiones en torno a enlaces covalentes (interacciones 1-4), se tratan con series de Fourier. De esta manera, podemos dar cuenta de la periodicidad que tiene el potencial respecto de los ángulos diedros. La última sumatoria de la ecuación 1 describe las interacciones entre átomos que pertenecen a diferentes moléculas, o que pertenecen a la misma molécula pero tienen una separación mayor que 1-4. Estas interacciones reciben el nombre de “interacciones no-ligadas” aunque, como ya dijimos, también se aplican entre átomos pertenecientes a la misma molécula.

Como puede verse en la ecuación 1, las interacciones no-ligadas tienen tres contribuciones diferentes. Por un lado existen repulsiones muy fuertes, pero de muy corto alcance, que se manifiestan cuando los átomos están muy próximos entre sí (o sea, cuando sus nubes electrónicas se solapan). Por el otro tenemos las fuerzas atractivas de dispersión (también llamadas fuerzas de van der Waals), que se originan en las interacciones entre dipolos inducidos. En la ecuación 1, estas dos contribuciones aparecen encerradas entre corche-

tes, en lo que se conoce como un potencial de Lennard-Jones.³ Finalmente, la última de las interacciones no ligadas de la ecuación 1, son las interacciones coulómbicas entre las cargas parciales que se asignan a cada átomo.

En buena medida, la utilidad de una simulación MD depende de cuán bien el campo de fuerza da cuenta de las interacciones en el sistema simulado. Y para saber si un campo es adecuado o no, es preciso conocer cómo fue parametrizado. Por este motivo, en la próxima sección, discutimos las estrategias más comúnmente utilizadas para obtener los parámetros de los campos MM. Sin embargo, entendemos que esta discusión puede ser demasiado detallada para alguien sólo interesado en las aplicaciones de las simulaciones MD de biomoléculas. Quienes se reconozcan en ese grupo, pueden saltar la próxima sección y retomar en la 3.4.

3.3. La determinación de los parámetros

Los valores de los parámetros de los campos de fuerza MM se ajustan de manera tal de poder reproducir, mediante cálculos mecánicos o simulaciones MD, propiedades moleculares determinadas experimentalmente. Por eso decimos que estos campos son “empíricos”. Entre las interacciones ligadas, los parámetros más sencillos de determinar son las distancias y ángulos de los enlaces covalentes. Los mismos son característicos de cada par o terna de átomos involucrados y del orden de los enlaces presentes. Por ejemplo, la distancia de equilibrio entre el átomo de carbono y el de oxígeno en los grupos carbonilo (C=O) es ~ 122 pm, mientras que para un enlace simple entre los mismos átomos la distancia de equilibrio es de ~ 143 pm. Con estas ideas en mente, los valores de $r_{i,eq}$ y $\theta_{i,eq}$ de la ecuación 1 se asignan en concordancia con las distancias y ángulos observados al hacer cristalografía de rayos-X de pequeñas moléculas que contienen los enlaces requeridos. Por otra parte, las constantes de fuerza $k_{r,i}$ y $k_{\theta,i}$ se determinan a partir de las frecuencias de vibración observadas haciendo espectroscopía IR de pequeñas moléculas. Esta estrategia se basa en que dichas constantes determinan el espaciado entre los niveles cuánticos de vibración. Y justamente, en espectroscopía, lo

³Hay otras formas funcionales para las interacciones repulsivas, sin embargo el potencial de Lennard-Jones es el más utilizado por su simplicidad.

que medimos es el espaciado entre niveles.

Entre las interacciones no-ligadas, los parámetros más sencillos de determinar son los del potencial de Lennard Jones (LJ), ϵ_{ij} y r_{ij}^0 . Para comprender el fundamento del procedimiento es necesario conocer el sentido físico de esos parámetros. La figura 4 muestra una representación esquemática de un potencial LJ. Cuando $r_{ij} < R_{\min}$ el potencial es repulsivo. O sea, cuando los átomos están a distancias menores a R_{\min} experimentan fuerzas que tienden a separarlos. Por el contrario, para $r_{ij} > R_{\min}$, las fuerzas entre los átomos son atractivas. Finalmente, a grandes distancias, el potencial se vuelve plano indicando la ausencia de cualquier interacción entre los átomos (ver Fig. 4).

La Fig. 4 nos muestra que el valor de ϵ_{ij} mide la profundidad del pozo de potencial. Cuanto mayor es ϵ_{ij} , más fuerza hay que hacer para lograr separar el átomo i del átomo j . Por otra parte, el valor de r_{ij}^0 es el punto donde el potencial LJ se vuelve nulo. En base a lo discutido en el párrafo anterior, podemos comprobar que este valor es “casi” igual a la distancia de mínimo acercamiento. Decimos “casi” porque, en realidad, los átomos pueden alcanzar distancias un poquito menores a r_{ij}^0 . Pero si lo hacen, experimentan una fuerza repulsiva muy grande que les impide cualquier posterior acercamiento. Por eso, se considera que r_{ij}^0 es igual a la suma de los radios de los átomos involucrados. Así, el parámetro ϵ_{ij} es responsable del valor de las propiedades del sistema que tienen que ver con la cohesión entre sus átomos, tales como el calor de vaporización o la temperatura crítica, mientras que r_{ij}^0 determina el volumen ocupado por los mismos cuando están en contacto unos con otros. Por ende, la densidad o el volumen molar de un líquido, calculadas a partir de simulaciones MD, dependerán del valor asignado a r_{ij}^0 .

El último punto a considerar en la determinación de los parámetros LJ, es que necesitamos un parámetro para cada par de átomos i y j . Debido a que para N tipos de átomos distintos hay $N(N - 1)$ pares de átomos distintos (o sea hay muchos más pares de átomos que átomos), se decidió calcular los valores de r_{ij}^0 y ϵ_{ij} a partir de parámetros asignados a los átomos individuales. De esta manera, se reduce la cantidad de constantes a las que hay que asignar

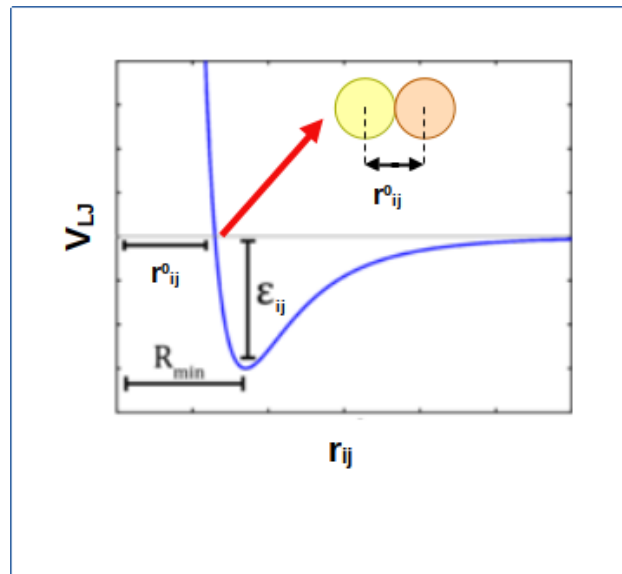


Figura 4: Potencial de Lennard-Jones. La pendiente (inclinación) de la curva en cada punto mide la fuerza experimentada por los átomos ($F = -\partial V_{LJ}/\partial r_{ij}$). Para $r_{ij} < R_{min}$ la pendiente es grande y negativa lo que hace que los átomos sientan una fuerte repulsión. Para $r_{ij} > R_{min}$ la fuerza es moderada pero atractiva. Finalmente para $r_{ij} \gg R_{min}$ el potencial se vuelve plano, lo que indica que la fuerza se hace cero.

un valor. Para ello se utilizan las reglas de Lorentz-Berthold,

$$r_{ij}^0 = (r_{ii}^0 + r_{jj}^0)/2,$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}.$$

En esta expresión r_{ii}^0 y ϵ_{ii} miden el tamaño y la cohesión entre los átomos del elemento i , respectivamente. Los mismos son ajustados de manera tal que simulaciones MD de esta especie en estado líquido puedan reproducir las propiedades experimentales mencionadas anteriormente. Lo mismo aplica a r_{jj}^0 y ϵ_{jj} .

Para la determinación de las interacciones electrostáticas entre moléculas, debemos asignar cargas parciales a cada átomo. Estos son los parámetros q_i y q_j que aparecen en la última sumatoria de la ecuación 1. Seguramente, la mayoría de los lectores de este artículo están familiarizados con la idea de las cargas parciales. Las mismas son ampliamente utilizadas en los cursos introductorios de Química, cuando se enseñan las propiedades del enlace co-

valente polar. No obstante, y más allá de cualquier familiaridad que el lector pueda tener, es importante destacar que las cargas parciales de los átomos son un concepto teórico que no tiene correlato experimental. O sea, no hay ninguna medición que podamos hacer, ni ningún cálculo basado en primeros principios, que no permita evaluar unívocamente estas cargas.

Una molécula puede ser considerada como un conjunto de cargas positivas puntuales (los núcleos), más una densidad de carga negativa dispersa (la nube electrónica). Y esta nube electrónica, no se distribuye de manera uniforme alrededor de los núcleos, sino que tiende a concentrarse sobre los átomos más electronegativos. Esto hace que se genere un campo (o potencial) eléctrico alrededor de la molécula, equivalente al que produciría un conjunto de pequeños dipolos de enlace. Este hecho, sugiere que podríamos determinar las cargas parciales de los átomos de manera tal que ajusten los dipolos de enlace. Pero el problema es que, cuando las moléculas interactúan entre sí, deforman sus nubes electrónicas. Debido a ello, los dipolos que podríamos observar en las moléculas aisladas son diferentes de los que operan cuando las mismas están interactuando.⁴ Este fenómeno de deformación de las nubes electrónicas se conoce con el nombre de polarización.

En un cálculo cuántico *ab-initio*, la polarización molecular es naturalmente tomada en cuenta, ya que los electrones son explícitamente considerados (ver Fig. 1). Sin embargo, en los campos de fuerza MM, el efecto de los electrones es incorporado de manera implícita, motivo por el cual estos campos son incapaces de describir la polarización molecular. En la actualidad, una parte importante de las investigaciones metodológicas sobre simulaciones MD están enfocadas en el desarrollo de campos de fuerza “polarizables” [Jing et al., 2019]. Hasta el momento, estos desarrollos han generado resultados promisorios en cuanto a la exactitud obtenida, pero poco alentadores respecto a los costos computacionales. Los largos tiempos de simulación requeridos para obtener resultados convergidos en simulaciones MD de macromoléculas no pueden ser aun alcanzados con estos campos.

Como las cargas parciales q_i y q_j de la ecuación 1 no pueden determinarse de manera unívoca, diferentes desarrolladores han adoptado diferentes

⁴Incluso para una molécula aislada, esos dipolos cambian al cambiar la conformación.

estrategias para su evaluación. Por ejemplo, los desarrolladores del campo de fuerza del paquete AMBER, han determinado las cargas parciales de los átomos de manera tal de reproducir el potencial electrostático en una grilla de puntos alrededor de un conjunto de moléculas [Wang et al., 2004a]. Por su parte, los desarrolladores del campo de fuerza CHARMM, han asignado las cargas parciales con el fin de reproducir la energía de interacción entre un conjunto de moléculas seleccionadas y una molécula de agua ubicada en diferentes puntos ubicados alrededor de las mismas [Zhu et al., 2012]. Ambas estrategias son razonables. Y por lo expuesto, ninguna es correcta o incorrecta.

Finalmente, quedan por asignar valores a los parámetros torsionales ($V_{i,n}$, ω_i y δ_i de la Eq. 1). Con ellos pasa algo similar a lo que ocurre con las cargas parciales: no se corresponden con ninguna cantidad que se pueda determinar unívocamente. Cuando dos grupos moleculares rotan en torno a un enlace covalente, se obtiene un perfil de energía potencial con picos y pozos, como el mostrado en la Fig. 5. La forma particular de la curva dependerá de la molécula y el enlace considerado, pero hay algunas características que son generales. Por ejemplo, los máximos se producen cuando grupos unidos a los átomos 2 y 3 de la torsión presentan una conformación “eclipsada”, dado que en esa situación hay una mayor repulsión entre sus nubes electrónicas. Cuanto más voluminosos sean los grupos que se encuentran eclipsados, mayor es la altura de la barrera. Por el contrario, cuando los grupos unidos a los átomos centrales están en una conformación alternada, esa tensión se libera y tenemos un mínimo de potencial. Desafortunadamente, los términos repulsivos del potencial de Lennard-Jones no pueden reproducir estas curvas porque fueron evaluados para átomos no enlazados. En las torsiones, una parte importante de la repulsión proviene de las nubes electrónicas de los enlaces covalentes. Por este motivo es que resulta necesario agregar términos específicos para dar cuenta de ella.

En una primera aproximación, los términos torsionales se determinan para que ajusten perfiles de energía como los de la Fig. 5, los cuales se obtienen mediante cálculos mecano-cuánticos realizados sobre moléculas seleccionadas. Sin embargo, se ha demostrado que esta estrategia no es suficiente para

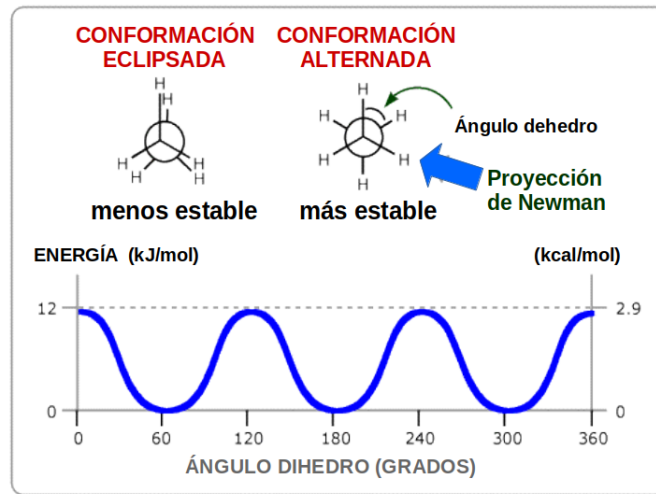


Figura 5: Perfil de energía para la torsión de los grupos metilo del etano. Los picos se producen porque cuando los enlaces C-H quedan eclipsados, debido a la repulsión de sus nubes electrónicas. En la conformación alternada, esta repulsión se reduce al mínimo.

las torsiones del *backbone* de proteínas y ácidos nucleicos, porque simulaciones MD que utilizan tales parámetros son incapaces de reproducir correctamente las estructuras 3D de dichas moléculas. Por ejemplo, la propensión de las proteínas para formar hélices- α o láminas- β no se corresponde con los datos experimentales. Lo mismo ocurre con los gráficos de Ramachandran. Por tal motivo, los desarrolladores de los campos de fuerza MM decidieron ajustar los parámetros de las torsiones del *backbone* para replicar, mediante simulaciones MD, las características estructurales de las diferentes biomoléculas.

Los primeros campos de fuerza de proteínas, calibraron sus parámetros torsionales con el fin de reproducir los datos provenientes de las pocas proteínas globulares cuya estructura era conocida en ese momento. Las mismas, habían sido determinadas mediante difracción de rayos-X, una técnica que requiere de bajas temperaturas para formar los cristales de la proteína. Con el transcurso del tiempo, se fueron dando a conocer más y más estructuras proteicas, muchas de las cuales fueron dilucidadas por RMN, una técnica que se aplica a las proteínas en solución. Al contrastar las características de estas nuevas estructuras, con los campos de fuerza existentes, se encontraron grandes discrepancias [Hornak et al., 2006]. Básicamente, las simulaciones

MD predecían proteínas más rígidas que las observadas experimentalmente. Esto se debía, simplemente, a que los campos de fuerza utilizados habían sido parametrizados para reproducir las rígidas estructuras determinadas por cristalografía de rayos X, en lugar de las correspondientes a proteínas en solución. Estas discrepancias, condujeron al desarrollo de nuevas versiones de los campos de fuerza, un fenómeno que continuó sucediendo a medida que nuevos experimentos fueron evidenciando la incapacidad de los campos de fuerza pre-existentes.

Por lo expuesto, podemos concluir con la afirmación de que los campos de fuerza evolucionan. Nunca son perfectos ni tienen validez general. Pero se pueden ir refinando, así como ampliando su campo de aplicación. En la actualidad, hay muchas investigaciones enfocadas en poner a prueba la exactitud de los campos MM disponibles, para predecir las características de las proteínas intrínsecamente desordenadas [Huang et al., 2017, Zerze et al., 2019]. Algunas mejoras en ese sentido ya se han logrado, pero las discusiones están lejos de estar saldadas.

3.4. Las limitaciones que impone el campo de fuerza

Los campos de fuerza MM, como el de la ecuación 1, permiten hacer simulaciones que jamás podríamos realizar si utilizáramos potenciales más exactos. Pero como discutimos en la sección anterior, los mismos están sujetos a decisiones arbitrarias referidas tanto a la estrategia usada en su parametrización como al conjunto de moléculas seleccionadas para llevar a cabo este proceso. Esto no significa que diferentes campos vayan a producir resultados muy diferentes, sino que alcanzarán resultados similares habiendo asignado un peso distinto a cada tipo de interacción. Por este motivo, los parámetros de los campos de fuerza no son transferibles. Por el contrario, cada campo tiene que ser considerado como un paquete cerrado. Obviamente, a veces sucede que queremos simular una molécula cuyos parámetros no están incluidos en el campo de fuerza. Por ejemplo, podríamos querer estudiar un nuevo ligando para un receptor. Para situaciones como ésta, los desarrolladores de los diferentes campos de fuerza han diseñado programas y procedimientos que permiten obtener los parámetros faltantes en concordancia con los ya existen-

tes [Wang et al., 2004b, Vanommeslaeghe et al., 2010].

Es también importante destacar que existen diferentes campos para diferentes tipos de biomoléculas. Hay para proteínas globulares, para proteínas intrínsecamente desordenadas, para ADN, para ARN, para azúcares, etc. [Pastor and MacKerell Jr, 2011, Galindo-Murillo et al., 2016, Wang et al., 2004a, Huang et al., 2017]. Esta falta de generalidad es el precio a pagar para obtener mejores resultados con cada tipo. Asimismo, existen varios campos de fuerza para representar las interacciones del agua [Zielkiewicz, 2005, Price and Brooks III, 2004]. Al momento de hacer una simulación MD, tenemos que saber qué campo de fuerza de agua es compatible con el elegido para la macromolécula. Esta concordancia es fundamental ya que las conformaciones que puede alcanzar una macromolécula en solución dependen del balance entre las interacciones soluto-soluto, solvente-solvente y soluto-solvente. Típicamente, en las páginas web de los desarrolladores de los campos de fuerza se indica qué parametrización del agua debe ser utilizada.

En cuanto a las limitaciones que son propias de la forma funcional utilizada, una de las más obvias es el hecho de considerar a los enlaces covalentes como resortes armónicos. Esta representación funciona bien (en general) porque a temperatura ambiente las moléculas suelen estar en el nivel cuántico vibracional fundamental. Sin embargo, los enlaces armónicos no se pueden romper ya que cuanto mayor es su estiramiento más alta es la energía potencial. Por este motivo, los campos de fuerza como el de la ecuación 1 son incapaces de describir la ruptura y formación de enlaces. Así, si se quiere estudiar una reacción química en un sistema biológico, hay que recurrir a campos más sofisticados que reciben el nombre de QM/MM [Senn and Thiel, 2009]. En ellos, las interacciones de una pequeña parte del sistema se obtiene mediante cálculos mecanocuánticos. Para el resto del sistema se utiliza un potencial MM como el de la ecuación 1.

Finalmente, la otra gran limitación de los campos de fuerza MM como el de la ecuación 1, es su incapacidad para describir la polarización molecular. Ya discutimos en la sección anterior, el origen de este problema. Por lo tanto, en este punto, sólo cabe destacar en qué situaciones esta característica tendrá efectos más notorios. Una molécula es más polarizable cuanto más extendida

es su nube electrónica. Así, los elementos del tercer período de la tabla periódica son más polarizables que los del segundo; al igual que los aniones lo son respecto a los átomos neutros, que a su vez son más polarizables que los cationes. Por otra parte, cuánto más elevada y más concentrada es una carga puntual, mayor es su capacidad para polarizar una nube electrónica. Así, los cationes son más polarizantes que los aniones (porque son más pequeños) y los cationes de carga +2 son más polarizantes que los de carga +1. Por lo expuesto, es claro que los campos de fuerza para ADN y ARN, cuyos *backbones* contienen a grupos fosfato de muy alta polarizabilidad en interacción con iones Mg^{+2} , representan el mayor desafío para las simulaciones de biomoléculas con campos de fuerza no polarizables. Este problema, es actualmente el foco de numerosas investigaciones [Lemkul and MacKerell Jr, 2018, Zhang et al., 2018]. Algunas de ellas intentan optimizar los resultados que pueden obtenerse con la forma funcional de la ecuación 1. Otras, sostienen que la polarización es demasiado importante para este tipo de moléculas y que por lo tanto, la única alternativa es utilizar campos polarizables. Con esta perspectiva intentan desarrollar algoritmos que aceleren la aplicación de los potenciales polarizables [Huang et al., 2018]. Por último, hay varios intentos de emplear algoritmos de *machine learning* para determinar estos potenciales [Fröhking et al., 2020].

4. Hitos en las simulaciones de biomoléculas

Las primeras simulaciones MD de fases condensadas se llevaron a cabo a mediados de los años 50 del siglo pasado [Alder and Wainwright, 1959]. Debido a las limitaciones computacionales de esa época, estas primeras aplicaciones se enfocaron en sistemas muy sencillos. En particular se simularon líquidos atómicos, tales como Ne y Ar [Rahman, 1964]. En esos sistemas, los únicos términos de la Ec. 1 que sobreviven son los del potencial de Lennard-Jones. Esta esta reducción en el número de términos, acelera enormemente los cálculos.

Por lo expuesto, no es de extrañar que debieran pasar más de 20 años antes de que se realizaran las primeras simulaciones MD de proteínas. Alcanzar ese hito requirió sobreponerse a numerosas dificultades, tanto tecnológi-

cas como teóricas. Entre los obstáculos teóricos a superar, uno de los principales fue la obtención de campos de fuerza confiables. Como ya discutimos, la parametrización de los mismos requiere conocer las estructuras 3D de un conjunto de proteínas. Y tal conjunto, no estuvo disponible hasta la segunda mitad de los años 70.⁵ Vencido ese obstáculo, se pudo concretar la primera simulación MD de una proteína en solución. El trabajo, fue realizado por McCammon, Gelin y Karplus y se publicó en *Nature* en 1977. El sistema simulado fue una proteína globular pequeña, el inhibidor de la tripsina pancreática bovina o BPTI, por sus siglas en inglés [McCammon et al., 1977]. El solvente no fue explícitamente incluido en la simulación, sino que fue tratado como un dieléctrico continuo, a excepción de cuatro moléculas de agua que estaban presentes en la estructura cristalina de BPTI. Además, el tiempo simulado fue relativamente corto. Por este motivo, los resultados no permitieron esclarecer ningún aspecto de relevancia biológica referido al funcionamiento de BPTI. Sin embargo, tuvieron la virtud de demostrar que las simulaciones de proteínas eran factibles y constituían un campo fértil para nuevas investigaciones.

En paralelo a la determinación de nuevas estructuras proteicas, se desarrollaron numerosos algoritmos que mejoraron la eficiencia y precisión de los cálculos MD. Entre ellos podemos mencionar:

- algoritmos como los de Verlet o *leap-frog*, que hacen posible la propagación estable y eficiente de las ecuaciones de movimiento [Verlet, 1967];
- algoritmos como los de Andersen, que permitieron por primera vez simular sistemas a T y P constante (en vez de E y V constante como surge naturalmente de la propagación de las ecuaciones de Newton)[Andersen, 2008];
- algoritmos como *SHAKE* o *RATTLE*, que fijan la distancia de los enlaces covalentes que involucran átomos de hidrógeno. De esta manera, los movimientos de más alta frecuencia del sistema quedan eliminados, lo que permite utilizar pasos de integración más grandes [Andersen, 1983].

Aun con los avances recién mencionados, las simulaciones MD estándar suelen tener deficiencias que se originan en un muestreo insuficiente del espa-

⁵La primera proteína cuya estructura pudo ser resuelta fue la hemoglobina, que se dio a conocer en 1958. A principio de los años 70, el número de estructuras conocidas era inferior a 20.

cio configuracional. Esta limitación tiene particular impacto cuando se quieren estudiar procesos activados, ya que las configuraciones correspondientes a los puntos más altos de la barrera de potencial pueden no ser muestreadas en la debida proporción o, directamente, no ser muestreadas. Para salvar este inconveniente, se han desarrollado numerosos métodos de “muestreo-mejorado”. Entre los más comúnmente utilizados podemos mencionar:

- Metadinámica [Bussi and Laio, 2020, Barducci et al., 2011];
- *Umbrella Sampling* [Kästner, 2011];
- Dinámica molecular de intercambio de réplicas [Bernardi et al., 2015];
- Dinámica molecular acelerada [Wang et al., 2021]

La implementación de estas técnicas ha hecho posible el estudio de reacciones enzimáticas, pasaje de iones o moléculas por un poro y salida del sustrato desde el sitio activo de una enzima, mediante simulaciones MD [Pierdominici-Sottile et al., 2019, Pierdominici-Sottile et al., 2011, Pierdominici-Sottile and Roitberg, 2011, Cossio-Pérez et al., 2019, Pierdominici-Sottile et al., 2014b, Pierdominici-Sottile et al., 2014a].

De la publicación inicial de Karplus y colaboradores a la fecha, el campo de las simulaciones MD de biomoléculas ha crecido enormemente. Tan es así, que hoy en día esta herramienta se utiliza rutinariamente en la gran mayoría de los laboratorios. Una prueba de este uso generalizado es que hay más de mil publicaciones anuales con esta temática en revistas de alto impacto [Hollingsworth and Dror, 2018] y casi veinte mil en el total de las revistas científicas [Sinha et al., 2022]. En parte, esta “popularidad” se debe a la existencia de paquetes y programas que permiten llevar a cabo las simulaciones sin necesidad de tener que programar los algoritmos (que son relativamente complejos). Esos programas, pueden ser descargados sin ningún costo, si van a ser usados dentro del ámbito académico. En la tabla 1 presentamos los sitios web de los programas más comúnmente utilizados para hacer simulaciones MD, así como también de programas que sirven para visualizar las trayectorias.

La Fig. 6 muestra un conjunto de investigaciones que constituyeron hitos en el desarrollo de las simulaciones MD de biomoléculas. Seleccionamos

Cuadro 1: Programas más comúnmente utilizados para realizar simulaciones MD y programas de visualización de biomoléculas.

Programa de simulación	Sitio web
CHARMM	https://academiccharmm.org/
GROMACS	https://www.gromacs.org/
AMBER	https://ambermd.org/index.php
NAMD	http://www.ks.uiuc.edu/Research/namd/
Programa de visualización	Sitio web
VMD	http://www.ks.uiuc.edu/Research/vmd/
PyMOL	https://pymol.org/2/
Chimera	https://www.cgl.ucsf.edu/chimera/index.html

estos trabajos con el propósito de ilustrar cómo esta herramienta ha servido para dilucidar el mecanismo de procesos moleculares relevantes. Asimismo, queríamos resaltar los avances producidos a lo largo del tiempo en relación con el tamaño de los sistemas abordados y los tiempos de simulación alcanzados.

Para el análisis, tomaremos como referencia el trabajo seminal sobre la BPTI de Karplus y colaboradores, de 1977 [McCammon et al., 1977]. El sistema estudiado en ese trabajo, tenía tan solo 58 aminoácidos y el tiempo de simulado fue de escasos 10.0 ps. Gracias a los avances producidos desde entonces hasta la fecha, tanto en el *hardware* como en el *software*, hoy se pueden realizar simulaciones del orden de los milisegundos [Shaw et al., 2010, Turoňová et al., 2020]. O sea, el tiempo simulado se ha incrementado en ocho órdenes de magnitud. Si nos enfocamos en el tamaño de los sistemas, debemos mencionar que se han reportado simulaciones de estructuras supra-moleculares como nucleosomas, ribosomas y hasta capsides virales [Sanbonmatsu et al., 2005, Zhao et al., 2013], los cuales contienen cientos o miles de millones de átomos.

No sólo el tiempo y el tamaño de las simulaciones MD se incrementaron en varios órdenes de magnitud con el transcurso de los años, sino que también aumentó la complejidad de los procesos estudiados. Como ejemplo, podemos mencionar diversas simulaciones enfocadas sobre el plegamiento de proteínas [Simmerling et al., 2002, Lindorff-Larsen et al., 2011], en sus cambios confor-

macionales [Grant et al., 2009], así como en interacciones proteína-ligando [Liu et al., 2017]. Asimismo, cálculos MD han ayudado a dilucidar los mecanismos de reacciones enzimáticas [Villa et al., 2000, Warshel et al., 2006], de procesos de traducción celular [Bock et al., 2013] y de transporte a través de la membrana celular [Törnroth-Horsefield et al., 2006], entre otros. Por último, debemos destacar la contribución de las simulaciones MD al diseño de fármacos, ya que en numerosas ocasiones ellas permitieron identificar bolsillos proteicos que no habían sido detectados mediante estudios estructurales experimentales (como por ejemplo de rayos-X). Un ejemplo de este tipo puede hallarse en el primer antirretroviral aprobado en Estados Unidos para inhibidores de la integrasa del VIH, cuyo descubrimiento se debe a investigaciones de esta naturaleza [Schames et al., 2004].

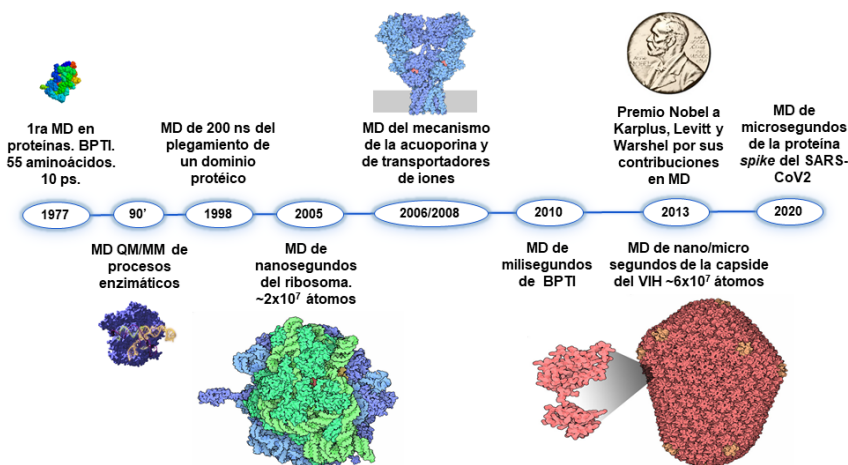


Figura 6: Línea temporal que describe eventos destacados en simulaciones de MD en sistemas biológicos. Las referencias bibliográficas corresponden a [McCammon et al., 1977] (1977), [Villa et al., 2000, Warshel et al., 2006] (90'), [Duan et al., 1998] (1998), [Sanbonmatsu et al., 2005] (2005), [Törnroth-Horsefield et al., 2006, Shi et al., 2008] (2006/2008), [Shaw et al., 2010] (2010), [Zhao et al., 2013] (2015) y [Rath and Kumar, 2020] (2020).

Para concluir esta sección debemos mencionar que el premio Nobel de Química del año 2013 fue otorgado conjuntamente a Martin Karplus, Michael Levitt and Arieh Warshel “por el desarrollo de modelos multi-escala para sistemas químicos complejos”. Con esas palabras, el comité Nobel hacía referencia a las contribuciones de estos investigadores a la obtención de métodos

QM/MM para permitir el estudio de reacciones en enzimas mediante simulaciones de dinámica molecular.

5. Perspectivas

A pesar de los notables avances que tuvieron las simulaciones MD de biomoléculas desde el año 1977 a la fecha, quedan aun muchos aspectos por mejorar/obstáculos por vencer. Por este motivo, en la actualidad, diversas líneas de trabajo son el foco intensas investigaciones. Algunas de ellas ya han sido mencionadas en este texto.

Tenemos, por un lado, el mejoramiento del cálculo de las interacciones no ligadas. Como dijimos, una vía para alcanzar estas mejoras es desarrollar campos de fuerza polarizables [Jing et al., 2019]. En este caso, el desafío actual se basa en lograr que la propagación de las ecuaciones de movimiento, usando estos campos, sea más eficiente [Huang et al., 2018]. Otra alternativa sobre la que se está trabajando es el desarrollo de PESs calculadas mediante algoritmos de *machine learning*. Estos algoritmos permiten evaluar el potencial mucho más rápidamente que cuando se utilizan campos de fuerza como el de la ecuación 1. Con este incentivo, lo que se hace es generar una red neuronal y se la entrena para que pueda reproducir las energías de interacción de diversos complejos moleculares [Behler and Parrinello, 2007]. Las energías de estos complejos son previamente evaluadas mediante cómputos cuánticos de gran exactitud. Alternativamente, se ha propuesto entrenar redes neuronales para que reproduzcan, mediante simulaciones MD, diversas propiedades de los sistemas determinadas experimentalmente [Fröhlking et al., 2020]. A diferencia de las redes que buscan reproducir energías, esta aproximación enfrenta el desafío de generar simulaciones MD muy extensas, a fin de asegurar que no haya problemas de muestreo. En otras palabras, es necesario garantizar que las discrepancias entre la propiedad calculada y la determinada experimentalmente sólo se origine en limitaciones de la PES y no, en un muestreo insuficiente.

Otro punto sobre el que se está trabajando intensamente es en el desarrollo de algoritmos de muestreo mejorado. Ocurre que los métodos men-

cionados en la sección 4, especialmente *Metadynamics* y *Umbrella Sampling*, tienen la limitación que requerir un conocimiento previo del proceso estudiado a fin de funcionar eficientemente. Típicamente, lo que se requiere es conocer cuales son los movimientos de relevancia biológica que están impedidos por altas barreras de potencial. En cambio, las investigaciones actuales apuntan a desarrollar algoritmos que puedan ser usados como una “caja negra”. O sea, métodos que permitan estudiar procesos de baja probabilidad de ocurrencia sin que tengamos que decirle al programa cuáles son los modos de movimiento que activamente participan ese proceso. Entre los métodos más prometedores podemos mencionar: *transition-path sampling* [Dellago et al., 2006], *forward flux sampling* [Hussain and Haji-Akbari, 2020] y *milestoning* [Elber, 2020].

Por último, un área de intensa investigación en la actualidad es el análisis de los datos generados por MD. Ocurre que los largos tiempos de simulación accesibles, sumados al gran tamaño de los sistemas examinados, conducen a que una simulación típica genere enormes archivos de datos (del orden de 0.1 a 100 Gigabites). Generalmente, estos archivos reportan las posiciones y/o velocidades de los átomos para cada momento en el cual se tomó una muestra. El desafío de los algoritmos de análisis consiste en poder encontrar patrones y regularidades en esa inmensa cantidad de datos, a fin de relacionar tales patrones con su funcionalidad biológica. Entre los problemas que se intenta resolver tenemos el de la *clusterización*, que busca agrupar las diferentes conformaciones moleculares de acuerdo a su cercanía estructural o cinética [Sittel and Stock, 2016]; el de la identificación de los modos de movimiento relevantes y la reducción de la dimensionalidad [Sittel and Stock, 2018] y el de la determinación de la conectividad entre los distintos conjuntos estructurales [Jain and Stock, 2012]. Cabe destacar que casi todos estos algoritmos hacen uso del Análisis de Componentes Principales como un primer paso para la identificación de los grados de libertad relevantes [Palma and Pierdominici-Sottile, 2023]. Recientemente también se han propuesto métodos de *machine learning* para alcanzar este fin [Chen and Ferguson, 2018, Noé et al., 2019].

En conclusión, podemos decir que las simulaciones MD han demostrado, hasta el momento, ser una herramienta valiosa para dilucidar los mecanismos

detallados del funcionamiento de las biomoléculas. Sin embargo, también han mostrado tener deficiencias y sabemos que enfrentan ciertas limitaciones. No obstante, la gran cantidad de investigaciones que actualmente buscan superar estas limitaciones, sumadas a los constantes desarrollos del *hardware* auguran que, para las simulaciones MD, lo mejor está aun por venir.

Referencias

- [Alder and Wainwright, 1959] Alder, B. J. and Wainwright, T. E. (1959). Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466.
- [Andersen, 1983] Andersen, H. C. (1983). Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52(1):24–34.
- [Andersen, 2008] Andersen, H. C. (2008). Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393.
- [Barducci et al., 2011] Barducci, A., Bonomi, M., and Parrinello, M. (2011). Metadynamics. *WIREs Computational Molecular Science*, 1(5):826–843.
- [Behler and Parrinello, 2007] Behler, J. and Parrinello, M. (2007). Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401.
- [Bernardi et al., 2015] Bernardi, R. C., Melo, M. C., and Schulten, K. (2015). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1850(5):872–877. Recent developments of molecular dynamics.
- [Bock et al., 2013] Bock, L. V., Blau, C., Schröder, G. F., Davydov, I. I., Fischer, N., Stark, H., Rodnina, M. V., Vaiana, A. C., and Grubmüller, H. (2013). Energy barriers and driving forces in trna translocation through the ribosome. *Nature structural & molecular biology*, 20(12):1390–1396.

- [Box, 1976] Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799.
- [Bussi and Laio, 2020] Bussi, G. and Laio, A. (2020). Using metadynamics to explore complex free-energy landscapes. *Nat Rev Phys*, 2:200–212.
- [Chen and Ferguson, 2018] Chen, W. and Ferguson, A. L. (2018). Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *Journal of Computational Chemistry*, 39(25):2079–2102.
- [Cossio-Pérez et al., 2019] Cossio-Pérez, R., Pierdominici-Sottile, G., Sobrado, P., and Palma, J. (2019). Molecular dynamics simulations of substrate release from trypanosoma cruzi udp-galactopyranose mutase. *Journal of chemical information and modeling*, 59(2):809–817.
- [Dellago et al., 2006] Dellago, C., Bolhuis, P., and Geissler, P. (2006). *Transition Path Sampling Methods*, pages 349–391. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Duan et al., 1998] Duan, Y., Wang, L., and Kollman, P. A. (1998). The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. *Proceedings of the National Academy of Sciences*, 95(17):9897–9902.
- [El Hage et al., 2018] El Hage, K., Hédin, F., Gupta, P. K., Meuwly, M., and Karpplus, M. (2018). Valid molecular dynamics simulations of human hemoglobin require a surprisingly large box size. *eLife*, 7:e35560.
- [Elber, 2020] Elber, R. (2020). Milestoning: An efficient approach for atomically detailed simulations of kinetics in biophysics. *Annual Review of Biophysics*, 49(1):69–85.
- [Fröhking et al., 2020] Fröhking, T., Bernetti, M., Calonaci, N., and Bussi, G. (2020). Toward empirical force fields that match experimental observables. *The Journal of Chemical Physics*, 152(23).
- [Galindo-Murillo et al., 2016] Galindo-Murillo, R., Robertson, J. C., Zgarbova, M., Spomer, J., Otyepka, M., Jurecka, P., and Cheatham III, T. E. (2016). As-

sessing the current state of amber force field modifications for dna. *Journal of chemical theory and computation*, 12(8):4114–4127.

[Gapsys and de Groot, 2019] Gapsys, V. and de Groot, B. L. (2019). Comment on 'valid molecular dynamics simulations of human hemoglobin require a surprisingly large box size'. *eLife*, 8:e44718.

[Grant et al., 2009] Grant, B. J., Gorfe, A. A., and McCammon, J. A. (2009). Ras conformational switching: simulating nucleotide-dependent conformational transitions with accelerated molecular dynamics. *PLoS computational biology*, 5(3):e1000325.

[Gupta et al., 2022] Gupta, A., Singh, A., Ahmad, N., Singh, T. P., Sharma, S., and Sharma, P. (2022). Chapter 12 - experimental techniques to study protein dynamics and conformations. In Tripathi, T. and Dubey, V. K., editors, *Advances in Protein Molecular and Structural Biology Methods*, pages 181–197. Academic Press.

[Hollingsworth and Dror, 2018] Hollingsworth, S. A. and Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143.

[Hornak et al., 2006] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725.

[Huang et al., 2018] Huang, J., Lemkul, J. A., Eastman, P. K., and MacKerell Jr., A. D. (2018). Molecular dynamics simulations using the drude polarizable force field on gpus with openmm: Implementation, validation, and benchmarks. *Journal of Computational Chemistry*, 39(21):1682–1689.

[Huang et al., 2017] Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., Grubmüller, H., and MacKerell Jr, A. D. (2017). Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods*, 14(1):71–73.

[Hussain and Haji-Akbari, 2020] Hussain, S. and Haji-Akbari, A. (2020). Studying rare events using forward-flux sampling: Recent breakthroughs and future outlook. *The Journal of Chemical Physics*, 152(6).

- [Jain and Stock, 2012] Jain, A. and Stock, G. (2012). Identifying metastable states of folding proteins. *Journal of Chemical Theory and Computation*, 8(10):3810–3819.
- [Jing et al., 2019] Jing, Z., Liu, C., Cheng, S. Y., Qi, R., Walker, B. D., Piquemal, J.-P., and Ren, P. (2019). Polarizable force fields for biomolecular simulations: Recent advances and applications. *Annual Review of Biophysics*, 48(1):371–394.
- [Joshi and Deshmukh, 2021] Joshi, S. Y. and Deshmukh, S. A. (2021). A review of advancements in coarse-grained molecular dynamics simulations. *Molecular Simulation*, 47(10-11):786–803.
- [Kleinjung and Fraternali, 2014] Kleinjung, J. and Fraternali, F. (2014). Design and application of implicit solvent models in biomolecular simulations. *Current Opinion in Structural Biology*, 25:126–134.
- [Kästner, 2011] Kästner, J. (2011). Umbrella sampling. *WIREs Computational Molecular Science*, 1(6):932–942.
- [Lemkul and MacKerell Jr, 2018] Lemkul, J. A. and MacKerell Jr, A. D. (2018). Polarizable force field for rna based on the classical drude oscillator. *Journal of computational chemistry*, 39(32):2624–2646.
- [Lindorff-Larsen et al., 2011] Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011). How fast-folding proteins fold. *Science*, 334(6055):517–520.
- [Liu et al., 2017] Liu, K., Watanabe, E., and Kokubo, H. (2017). Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. *Journal of computer-aided molecular design*, 31:201–211.
- [McCammon et al., 1977] McCammon, J. A., Gelin, B. R., and Karplus, M. (1977). Dynamics of folded proteins. *nature*, 267(5612):585–590.
- [Mlýnský et al., 2014] Mlýnský, V., Banás, P., Sponer, J., van der Kamp, M. W., Mulholland, A. J., and Otyepka, M. (2014). Comparison of ab initio, dft,

and semiempirical qm/mm approaches for description of catalytic mechanism of hairpin ribozyme. *Journal of Chemical Theory and Computation*, 10(4):1608–1622.

- [Noé et al., 2019] Noé, F., Olsson, S., Köhler, J., and Wu, H. (2019). Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147.
- [Ormazábal et al., 2021] Ormazábal, A., Palma, J., and Pierdominici-Sottile, G. (2021). Molecular dynamics simulations unveil the basis of the sequential binding of rsme to the noncoding rna rsmz. *The Journal of Physical Chemistry B*, 125(12):3045–3056.
- [Palma and Pierdominici-Sottile, 2023] Palma, J. and Pierdominici-Sottile, G. (2023). On the uses of pca to characterise molecular dynamics simulations of biological macromolecules: Basics and tips for an effective use. *ChemPhysChem*, 24(2):e202200491.
- [Pastor and MacKerell Jr, 2011] Pastor, R. and MacKerell Jr, A. (2011). Development of the charmm force field for lipids. *The journal of physical chemistry letters*, 2(13):1526–1532.
- [Pauling et al., 1951] Pauling, L., Corey, R. B., and Branson, H. R. (1951). The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences*, 37(4):205–211.
- [Pierdominici-Sottile et al., 2014a] Pierdominici-Sottile, G., Cossio Pérez, R., Galindo, J. F., and Palma, J. (2014a). Qm/mm molecular dynamics study of the galactopyranose→ galactofuranose reaction catalysed by trypanosoma cruzi udp-galactopyranose mutase. *PLoS One*, 9(10):e109559.
- [Pierdominici-Sottile et al., 2011] Pierdominici-Sottile, G., Horenstein, N. A., and Roitberg, A. E. (2011). Free energy study of the catalytic mechanism of trypanosoma cruzi trans-sialidase. from the michaelis complex to the covalent intermediate. *Biochemistry*, 50(46):10150–10158.

- [Pierdominici-Sottile et al., 2016] Pierdominici-Sottile, G., Moffatt, L., and Palma, J. (2016). The dynamic behavior of the p2x4 ion channel in the closed conformation. *Biophysical journal*, 111(12):2642–2650.
- [Pierdominici-Sottile et al., 2014b] Pierdominici-Sottile, G., Palma, J., and Roitberg, A. E. (2014b). Free-energy computations identify the mutations required to confer trans-sialidase activity into trypanosoma rangeli sialidase. *Proteins: Structure, Function, and Bioinformatics*, 82(3):424–435.
- [Pierdominici-Sottile et al., 2019] Pierdominici-Sottile, G., Racigh, V., Ormazábal, A., and Palma, J. (2019). Charge discrimination in p2x4 receptors occurs in two consecutive stages. *The Journal of Physical Chemistry B*, 123(5):1017–1025.
- [Pierdominici-Sottile and Roitberg, 2011] Pierdominici-Sottile, G. and Roitberg, A. E. (2011). Proton transfer facilitated by ligand binding. an energetic analysis of the catalytic mechanism of trypanosoma cruzi trans-sialidase. *Biochemistry*, 50(5):836–842.
- [Ponder and Case, 2003] Ponder, J. W. and Case, D. A. (2003). Force fields for protein simulations. In *Protein Simulations*, volume 66 of *Advances in Protein Chemistry*, pages 27–85. Academic Press.
- [Price and Brooks III, 2004] Price, D. J. and Brooks III, C. L. (2004). A modified tip3p water potential for simulation with ewald summation. *The Journal of chemical physics*, 121(20):10096–10103.
- [Rahman, 1964] Rahman, A. (1964). Correlations in the motion of atoms in liquid argon. *Physical review*, 136(2A):A405.
- [Rath and Kumar, 2020] Rath, S. L. and Kumar, K. (2020). Investigation of the effect of temperature on the structure of sars-cov-2 spike protein by molecular dynamics simulations. *Frontiers in molecular biosciences*, 7:583523.
- [Sanbonmatsu et al., 2005] Sanbonmatsu, K. Y., Joseph, S., and Tung, C.-S. (2005). Simulating movement of trna into the ribosome during decoding. *Proceedings of the National Academy of Sciences*, 102(44):15854–15859.

- [Schames et al., 2004] Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., and McCammon, J. A. (2004). Discovery of a novel binding trench in hiv integrase. *Journal of medicinal chemistry*, 47(8):1879–1881.
- [Senn and Thiel, 2009] Senn, H. M. and Thiel, W. (2009). Qm/mm methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229.
- [Shaw et al., 2010] Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., et al. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346.
- [Shi et al., 2008] Shi, L., Quick, M., Zhao, Y., Weinstein, H., and Javitch, J. A. (2008). The mechanism of a neurotransmitter: sodium symporter—inward release of na⁺ and substrate is triggered by substrate in a second binding site. *Molecular cell*, 30(6):667–677.
- [Simmerling et al., 2002] Simmerling, C., Strockbine, B., and Roitberg, A. E. (2002). All-atom structure prediction and folding simulations of a stable protein. *Journal of the American Chemical Society*, 124(38):11258–11259.
- [Sinha et al., 2022] Sinha, S., Tam, B., and Wang, S. M. (2022). Applications of molecular dynamics simulation in protein study. *Membranes*, 12(9):844.
- [Sittel and Stock, 2016] Sittel, F. and Stock, G. (2016). Robust density-based clustering to identify metastable conformational states of proteins. *Journal of Chemical Theory and Computation*, 12(5):2426–2435.
- [Sittel and Stock, 2018] Sittel, F. and Stock, G. (2018). Perspective: Identification of collective variables and metastable states of protein dynamics. *The Journal of Chemical Physics*, 149(15).
- [Törnroth-Horsefield et al., 2006] Törnroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R., and Kjellbom, P. (2006). Structural mechanism of plant aquaporin gating. *Nature*, 439(7077):688–694.

- [Turoňová et al., 2020] Turoňová, B., Sikora, M., Schürmann, C., Hagen, W. J., Welsch, S., Blanc, F. E., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., et al. (2020). In situ structural analysis of sars-cov-2 spike reveals flexibility mediated by three hinges. *Science*, 370(6513):203–208.
- [Vanommeslaeghe et al., 2010] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell Jr., A. D. (2010). Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):671–690.
- [Verlet, 1967] Verlet, L. (1967). Computer experiments on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103.
- [Villa et al., 2000] Villa, J., Štrajbl, M., Glennon, T., Sham, Y., Chu, Z., and Warshel, A. (2000). How important are entropic contributions to enzyme catalysis? *Proceedings of the National Academy of Sciences*, 97(22):11899–11904.
- [Wang et al., 2021] Wang, J., Arantes, P. R., Bhattarai, A., Hsu, R. V., Pawnikar, S., Huang, Y.-m. M., Palermo, G., and Miao, Y. (2021). Gaussian accelerated molecular dynamics: Principles and applications. *WIREs Computational Molecular Science*, 11(5):e1521.
- [Wang et al., 2004a] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004a). Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174.
- [Wang et al., 2004b] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004b). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174.
- [Warshel et al., 2006] Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H., and Olsson, M. H. (2006). Electrostatic basis for enzyme catalysis. *Chemical reviews*, 106(8):3210–3235.

- [Watson and Crick, 1953] Watson, J. and Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738.
- [Zerze et al., 2019] Zerze, G. H., Zheng, W., Best, R. B., and Mittal, J. (2019). Evolution of all-atom protein force fields to improve local and global properties. *The Journal of Physical Chemistry Letters*, 10(9):2227–2234.
- [Zhang et al., 2018] Zhang, C., Lu, C., Jing, Z., Wu, C., Piquemal, J.-P., Ponder, J. W., and Ren, P. (2018). Amoeba polarizable atomic multipole force field for nucleic acids. *Journal of chemical theory and computation*, 14(4):2084–2108.
- [Zhao et al., 2013] Zhao, G., Perilla, J. R., Yufenyuy, E. L., Meng, X., Chen, B., Ning, J., Ahn, J., Gronenborn, A. M., Schulten, K., Aiken, C., et al. (2013). Mature hiv-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497(7451):643–646.
- [Zhou and Liu, 2022] Zhou, K. and Liu, B. (2022). *Molecular Dynamics Simulation: Fundamentals and Applications*. Elsevier Science.
- [Zhu et al., 2012] Zhu, X., Lopes, P. E., and MacKerell Jr, A. D. (2012). Recent developments and applications of the charmm force fields. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1):167–185.
- [Zielkiewicz, 2005] Zielkiewicz, J. (2005). Structural properties of water: Comparison of the spc, spce, tip4p, and tip5p models of water. *The Journal of chemical physics*, 123(10):104501.