

# M todos num ricos en din mica molecular

Christophe Chipot

*Equipe de dynamique des assemblages membranaires, Unit  Mixte de Recherche CNRS/UHP 7565, Institut Nanc ien de Chimie Mol culaire, Universit  Henri Poincar , B.P. 239, 54506 Vand uvre-l s-Nancy, France  
Christophe.Chipot@edam.uhp-nancy.fr*

(Version: September 11, 2007)

## Table of contents

<b>1. Introducci�n</b>	<b>2</b>
1.1. Conectando el microsc�pico al meso- y macrosc�pico . . . . .	2
1.2. Legitimidad de las simulaciones de din�mica molecular . . . . .	4
<b>2. Las ecuaciones de la din�mica molecular</b>	<b>5</b>
2.1. La funci�n de energ�a potencial . . . . .	7
2.2. Los propagadores de din�mica molecular . . . . .	12
2.3. Din�mica molecular a temperatura constante . . . . .	14
2.4. Din�mica molecular a presi�n constante . . . . .	16
2.5. Interacciones electrost�ticas: M�s all� de las aproximaciones usuales . . . . .	18
2.6. Algunas propiedades accesibles a partir de la trayectoria . . . . .	22
<b>3. Conclusion</b>	<b>25</b>

## 1. Introducción

El conocimiento a nivel atómico de los aspectos estructurales y dinámicos de los sistemas organizados es de primer importancia para aumentar nuestra comprensión de la función de dichos complejos ensamblajes moleculares. En varios casos, alcanzar el detalle microscópico a través de técnicas experimentales convencionales es imposible. Sin embargo, el crecimiento masivo de los recursos informáticos iniciado hace unos diez años, junto con el desarrollo de algoritmos efectivos han permitido la investigación de ensamblajes de complejidad creciente utilizando herramientas teóricas.

El objetivo principal de este curso es un examen de la faceta de la Química Teórica que constituyen las simulaciones moleculares de la mecánica estadística. Su papel es acceder al detalle atómico en fases condensadas a través de experimentos informáticos *in silico*. Varios métodos están actualmente disponibles, incluyendo simulaciones de dinámica molecular (MD), de dinámica estocástica (SC) y sus casos especiales — por ejemplo, dinámica browniana y dinámica de Langevin — o de Monte Carlo (MC). Esos diferentes enfoques teóricos se pueden ver como un puente conectando observaciones experimentales a nivel macroscópico con el mundo microscópico. En lo que sigue, nos centraremos en la dinámica molecular sola.<sup>1</sup>

### 1.1. Conectando el microscópico al meso- y macroscópico

Es el uso de las simulaciones moleculares para modelar fases condensadas legítimo? En un sentido estricto, el estudio completo de un sistema químico tan complejo como un líquido molecular necesita resolver la ecuación de Schrödinger dependiente del tiempo para un sistema grande de {electrones + núcleos}. De hecho, este tipo de enfoque no se puede aplicar directamente, a pesar de progresos recientes en el campo de los cálculos de tipo lineal-creciente, así nos constreñiendo a limitarnos a una descripción clásica del sistema. Incluso en esta descripción, por óbvias razones computacionales, las simulaciones moleculares están generalmente restringidas a un número de partículas entre unas centenas y unos miles de átomos, bastante inferior al número de Avogadro.

Para correlacionar las propiedades del sistema microscópico a las que describen la fase macroscópica, es muy importante que los efectos de la frontera sean eliminados. En la práctica, se usan las condiciones periódicas de contorno (PBC), que consisten en una replicación del ensamble finito de partículas confinadas en la caja, a menudo paralelepípedica, según las tres direcciones del espa-

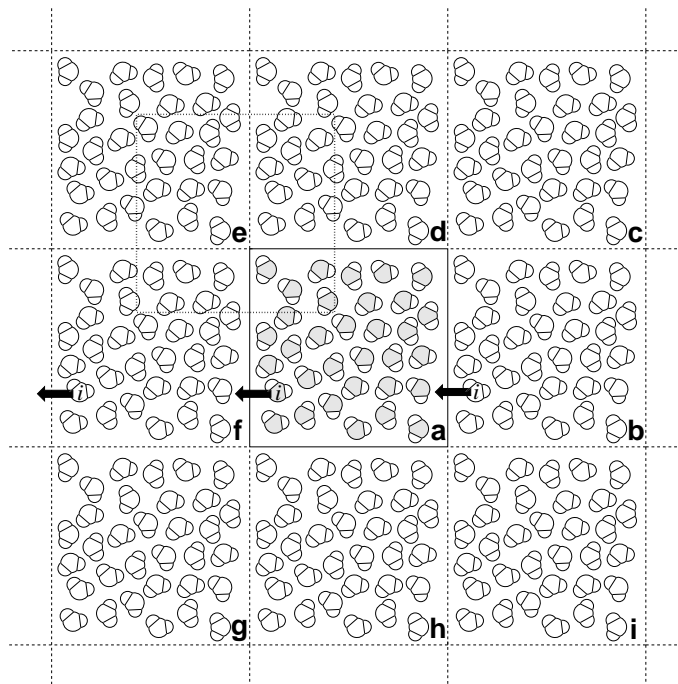


Figure 1: Vistazo bidimensional de la caja de simulación replicada en las tres dimensiones del espacio cartesiano. Empleando condiciones periódicas al borde (PBC), cuando una molécula  $i$  se vaya de la caja central **a**, sus imágenes en las cajas fantasmas vecinas se mueven de una manera similar. La caja depictada con trazo discontinuo, superpuesta con las cajas **a**, **d**, **e** y **f**, simboliza la famosa convención de la “imagen minimal”.

cio cartesiano (Figura 1). Se ha observado que una reproducción fiel y precisa de cantidades termodinámicas a partir de estos modelos de tamaño reducido justifica *a posteriori* la utilización de ese enfoque. La naturaleza pseudo–infinita del sistema así generado implica la necesidad de unas aproximaciones para tratar interacciones intermoleculares.<sup>2</sup> En particular, la aproximación de la “imagen minimal” supone que cada partícula  $i$  de la caja central interactúe con la imagen más cercana de todas las otras partículas  $j$ . Además, la introducción de una esfera de corte, o de *cut-off*, puede usarse para ignorar interacciones más lejos que una distancia arbitraria, igual o menos que la mitad de la dimensión más pequeña de la caja de simulación (Figura 2). Claramente, la validez de esas aproximaciones está condicionada por el rango de las interacciones intermoleculares consideradas. Si las interacciones de dispersión y de repulsión de corto alcance no necesitan ninguna atención específica, no se puede decir lo mismo en el caso de las interacciones electrostáticas. El tamaño del sistema evolucionando según  $r^3$ , es usualmente aceptado que las interacciones de tipo  $1/r^n$  — donde  $n < 3$  — no se pueden tratar correctamente con una truncatura esférica. Para salvar esta dificultad, está recomendado considerar un método adaptado, como los basados en sumas de red, *e.g.* Ewald–Kornfeld o Ladd,<sup>1,2</sup> que consisten en una evaluación de las interacciones de una partícula con todas las otras contenidas en la caja central, así como en todas las cajas fantasmas adyacentes. Adoptar este tipo

de enfoque aumenta, sin embargo, considerablemente el costo computacional, pero es indispensable para una descripción rigurosa y correcta de las interacciones de rango largo.

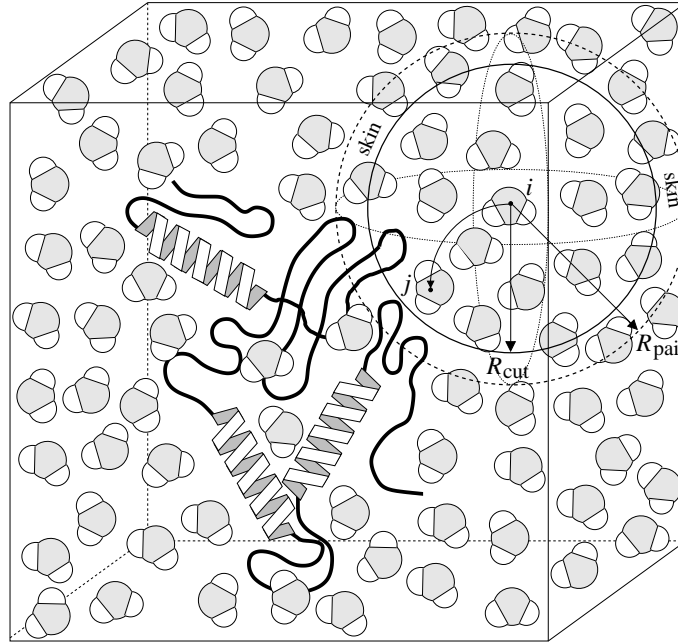


Figure 2: Uso de una esfera de corte de radio  $R_{cut}$  para limitar el cálculo de las interacciones de la partícula  $i$  con sus vecinos en la convención de la “imagen minimal”. Una esfera de radio  $R_{pair}$ , superior a  $R_{cut}$ , se utiliza para construir la lista de todos los vecinos de  $i$ . Esta lista de todas las pares  $\{i, j\}$  se actualiza periódicamente.

## 1.2. Legitimidad de las simulaciones de dinámica molecular

El principio de MD, particularmente sencillo, consiste en la generación de trayectorias para un colectivo finito de partículas integrando numéricamente las ecuaciones clásicas del movimiento. Este enfoque, *a priori* discutible, tiene su justificación en dos hechos remarcables: (i) la aproximación de Born–Oppenheimer, que permite separar el movimiento de los núcleos y el de los electrones, y, (ii) la longitud de onda de de Broglie es mucho más corta que las distancias intermoleculares típicas en la mayor parte de los casos. Con ello los efectos cuánticos se pueden despreciar con seguridad. Las trayectorias calculadas por este enfoque permiten determinar propiedades estáticas y dinámicas en forma de promedios temporales, coincidentes con los promedios estadísticos para sistemas *ergodicos*:

$$\lim_{t \rightarrow \infty} \overline{\mathcal{A}}_t = \langle \mathcal{A} \rangle \quad (1)$$

donde  $\mathcal{A}$  represente cualquiera propiedad observable,  $\overline{\mathcal{A}}_t$  es su promedio temporal y  $\langle \mathcal{A} \rangle$ , su promedio estadístico. En la práctica, la hipótesis de ergodicidad se cumple, por lo menos en el caso de líquidos simples.

## 2. Las ecuaciones de la dinámica molecular

En MD clásica,<sup>1-3</sup> la trayectoria de las varias componentes del sistema se genera por integración de las ecuaciones de Newton del movimiento, que, para cada partícula  $i$ , se escriben:

$$\begin{cases} m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \mathbf{f}_i(t) \\ \mathbf{f}_i(t) = -\frac{\partial \mathcal{V}(\mathbf{r}^N)}{\partial \mathbf{r}_i(t)} \end{cases} \quad (2)$$

$\mathcal{V}(\mathbf{r}^N)$  es la función de energía potencial del sistema de  $N$  partículas, que depende solamente de las coordenadas cartesianas  $\{\mathbf{r}_i\}$ . Las ecuaciones (2) se integran numéricamente utilizando un paso de tiempo infinitesimal,  $\delta t$ , para garantizar la conservación de la energía total del sistema — *viz.* típicamente 1–2 fs.

Generar trayectorias *exactas* para tiempos largos es, sin embargo, ilusorio, dado que las ecuaciones de Newton del movimiento se resuelven numéricamente, con un paso de tiempo finito. En cambio, la exactitud de la solución de las ecuaciones (2) no es tan crucial como parece *a priori*. Lo que importa más, en realidad, es el comportamiento estadístico correcto de la trayectoria para asegurar que las propiedades termodinámicas y dinámicas del sistema sean reproducidas con precisión suficiente. Esta condición *sine qua non* se cumple solo si el propagador del movimiento empleado posee la propiedad de simplecticidad.<sup>4</sup> Un propagador *simplectico* conserva la métrica invariante del espacio de las fases,  $\Gamma$ . Más aún el error asociado al propagador está acotado:

$$\lim_{n_{\text{paso}} \rightarrow \infty} \left( \frac{1}{n_{\text{paso}}} \right) \sum_{k=1}^{n_{\text{paso}}} \left| \frac{\mathcal{E}(k\delta t) - \mathcal{E}(0)}{\mathcal{E}(0)} \right| \leq \varepsilon_{\text{MD}} \quad (3)$$

Aquí,  $n_{\text{step}}$  es el número de pasos de la simulación,  $\mathcal{E}(0) \equiv \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N; 0)$ , la energía total inicial del sistema en equilibrio, y  $\varepsilon_{\text{MD}}$ , el límite superior que asegura la conservación de la energía — *viz.*  $10^{-4}$  constituye un valor aceptable. Considerando que el paso de tiempo está limitado, la integración de las

ecuaciones del movimiento no conduce a un crecimiento errático del error asociado a la conservación de la energía total, lo que podría afectar significativamente al comportamiento estadístico de MD a tiempos largos. Hay que destacar, para sistemas hamiltonianos, que la propiedad de simplecticidad implica que el jacobiano:

$$\mathbf{J}(\mathbf{\Gamma}_{\delta t}, \mathbf{\Gamma}_0) = \frac{\partial(\mathbf{\Gamma}_{\delta t}^1, \dots, \mathbf{\Gamma}_{\delta t}^N)}{\partial(\mathbf{\Gamma}_0^1, \dots, \mathbf{\Gamma}_0^N)} \quad (4)$$

sea unitario.  $\mathbf{\Gamma}_0$  represente el vector inicial del espacio de fases de  $N$  dimensiones, que contiene todas las variables de posición,  $\mathbf{r}$ , y de impulsión,  $\mathbf{p}$ , que describen el sistema.

Como ya se ha indicado, la naturaleza de rango largo de las interacciones carga–dipolo, *i.e.*  $1/r^2$ , y, *a fortiori*, carga–carga, *i.e.*  $1/r$ , impone el uso de algoritmos bien adaptados para tratar dichas contribuciones, lo que puede aumentar apreciablemente el coste computacional de la simulación. Rescribiendo las ecuaciones (2) de una manera más formal:

$$\mathbf{\Gamma}_t = e^{i\mathcal{L}t} \mathbf{\Gamma}_0 \quad (5)$$

donde  $\mathcal{L}$  es el operador de Liouville que genera una distribución  $\varrho(\mathbf{\Gamma}, t)$  para un ensamble termodinámico dado, según:

$$\frac{\partial \varrho(\mathbf{\Gamma}, t)}{\partial t} = -i\mathcal{L}\varrho(\mathbf{\Gamma}, t) \quad (6)$$

y aplicando la factorización de Trotter:

$$e^{i\mathcal{L}\Delta t} = e^{i\mathcal{L}_1 \frac{\Delta t}{2}} e^{i\mathcal{L}_2 \Delta t} e^{i\mathcal{L}_1 \frac{\Delta t}{2}} + \mathcal{O}(\Delta t^3) \quad (7)$$

con la cual  $i\mathcal{L} = i\mathcal{L}_1 + i\mathcal{L}_2$ , la deconvolución de las contribuciones de rango corto y de rango largo es directa. En función de la naturaleza de cada interacción, pasos de tiempo distintos se pueden utilizar para propagar el movimiento. Cortando, por ejemplo, el hamiltoniano total,  $\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)$ , que gobierna el sistema en contribuciones cinética,  $\mathcal{T}(\mathbf{p}^N)$ , de valencia,  $\mathcal{V}_{\text{valencia}}(\mathbf{r}^N)$ , de rango corto,  $\mathcal{V}_{\text{corto}}(\mathbf{r}^N)$ , y de rango largo,  $\mathcal{V}_{\text{largo}}(\mathbf{r}^N)$ , sigue que:

$$e^{i\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)\Delta t} = e^{i\mathcal{V}_{\text{largo}}(\mathbf{r}^N) \frac{\Delta t}{2}} \left\{ e^{i\mathcal{V}_{\text{corto}}(\mathbf{r}^N) \frac{\Delta t}{2n}} \left[ e^{i\mathcal{V}_{\text{valencia}}(\mathbf{r}^N) \frac{\Delta t}{2pn}} e^{i\mathcal{T}(\mathbf{p}^N) \frac{\Delta t}{pn}} \right] \right\} \quad (8)$$

$$\times \left. e^{i\mathcal{V}_{\text{valencia}}(\mathbf{r}^N) \frac{\Delta t}{2pn}} \right] \left. e^{i\mathcal{V}_{\text{corto}}(\mathbf{r}^N) \frac{\Delta t}{2n}} \right\}^n e^{i\mathcal{V}_{\text{largo}}(\mathbf{r}^N) \frac{\Delta t}{2}}$$

Esta partición en varias contribuciones de  $\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)$  constituye la idea central de los llamados algoritmos a pasos de tiempo multiples, como RESPA.<sup>5</sup> En ello se destaca claramente el uso de pasos de tiempo distintos para actualizar esas contribuciones, disminuyendo apreciablemente, por consecuencia, el coste informático de la simulación estadística.

## 2.1. La función de energía potencial

Esta función constituye la piedra angular de todos calculos de mecánica molecular, porque tiene que reproducir las interacciones intra- y intermoleculares del sistema lo más fiel y preciso que se pueda. En principio, dado la naturaleza de varios cuerpos del problema, esta funcional debe escribirse como una sumatoria de  $N$  términos:

$$\mathcal{V}(\mathbf{r}^N) = \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j>i} v_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots \quad (9)$$

donde  $v_1(\mathbf{r}_i)$ ,  $v_2(\mathbf{r}_i, \mathbf{r}_j)$ ,  $\dots$  representa el potencial intermolecular, el potencial de interacción de pares,  $\dots$   $\mathcal{V}(\mathbf{r}^N)$  es, por consecuencia, característico de un problema de  $N$  cuerpos, aunque se pueda argüar que  $v_2(\mathbf{r}_i, \mathbf{r}_j)$  probablemente constituye el término dominante de la contribución intermolecular.<sup>2</sup> Este punto de vista es el origen de la famosa aproximación de pares, en la cual los efectos de orden mayor están incluidos parcialmente en un potencial efectivo:

$$\mathcal{V}(\mathbf{r}^N) \simeq \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2^{\text{efectivo}}(r_{ij}) \quad (10)$$

Esta aproximación es utilizada en la mayor parte de los campos de fuerzas comerciales, en particular los diseñados para estudiar sistemas macromoleculares, para los cuales el costo computacional está íntimamente relacionado a la complejidad de  $\mathcal{V}(\mathbf{r}^N)$ . En la plétora de funciones de energía potencial disponibles, una de las más minimalistas descripciones del sistema es la del campo de fuerzas AMBER.<sup>6,7</sup>

$$\begin{aligned}
\mathcal{V}(\mathbf{r}) &= \sum_{\text{enlaces}} k_r (r - r_0)^2 + \sum_{\text{valencia}} k_\theta (\theta - \theta_0)^2 \\
&+ \sum_{\text{torsiones}} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\
&+ \frac{1}{k_{\text{vdW}}^{1-4}} \sum_{\substack{i < j \\ \{i,j\} \in 1-4}} \varepsilon_{ij} \left[ \left( \frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^*}{r_{ij}} \right)^6 \right] + \frac{1}{k_{\text{Coulomb}}^{1-4}} \sum_{\substack{i < j \\ \{i,j\} \in 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \\
&+ \sum_{\substack{i < j \\ \{i,j\} > 1-4}} \varepsilon_{ij} \left[ \left( \frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^*}{r_{ij}} \right)^6 \right] + \sum_{\substack{i < j \\ \{i,j\} > 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}}
\end{aligned} \tag{11}$$

en la cual,  $k_r$  y  $r_0$  representen, respectivamente, la constante de fuerza del enlace químico y su longitud de equilibrio,  $k_\theta$  y  $\theta_0$ , la constante de fuerza y su valor de equilibrio, y  $V_n/2$ ,  $n$  y  $\gamma$ , la barrera torsional, su periodicidad y su fase.  $\epsilon_0$  y  $\epsilon_1$  son, respectivamente, las permitividades dielectricas del vacío y relativa.  $q_i$  es la carga parcial llevada por el átomo  $i$ .  $R_{ij}^*$  y  $\varepsilon_{ij}$  corresponden a los parámetros de van der Waals para el par de átomos  $\{ij\}$ , obtenidos a partir de las reglas de combinación de Lorentz–Berthelot:

$$\begin{cases} \varepsilon_{ij} &= \sqrt{\varepsilon_i \varepsilon_j} \\ R_{ij}^* &= R_i^* + R_j^* \end{cases} \tag{12}$$

Considerando que, en el curso de su parametrización, generalmente utilizando calculos cuanticos sofisticados, los términos diédricos ya incluyen una componente electrostatica y de van der Waals, la mayor parte de los campos de fuerza comerciales distinguen entre aquellas interacciones de átomos separados por exactamente tres enlaces químicos (*i.e.* los llamados términos “1–4”), y todos los otros, dado que  $\{i, j\}$  no estén separados por una o dos enlaces químicos (Figura 3). El peso de las contribuciones “1–4” de van der Waals y de Coulomb se reduce usualmente por los factores  $1/k_{\text{vdW}}^{1-4}$  y  $1/k_{\text{Coulomb}}^{1-4}$  que aparecen en la expresión (11).

La descripción precisa de los ángulos de torsión constituye un aspecto particularmente crucial en el desarrollo de funciones de energía potencial. El comportamiento del potencial torsional,  $\mathcal{V}(\phi)$ , es a menudo excesivamente complejo para estar mimido correctamente por un término único de la serie de Fourier en la ecuación (11). El caso de fosfolípidos represente un excelente ejemplo ilustrativo de esa dificultad. La clave hacía una reproducción fiel de los parametros de orden,  $S_{\text{CD}}$ , de las



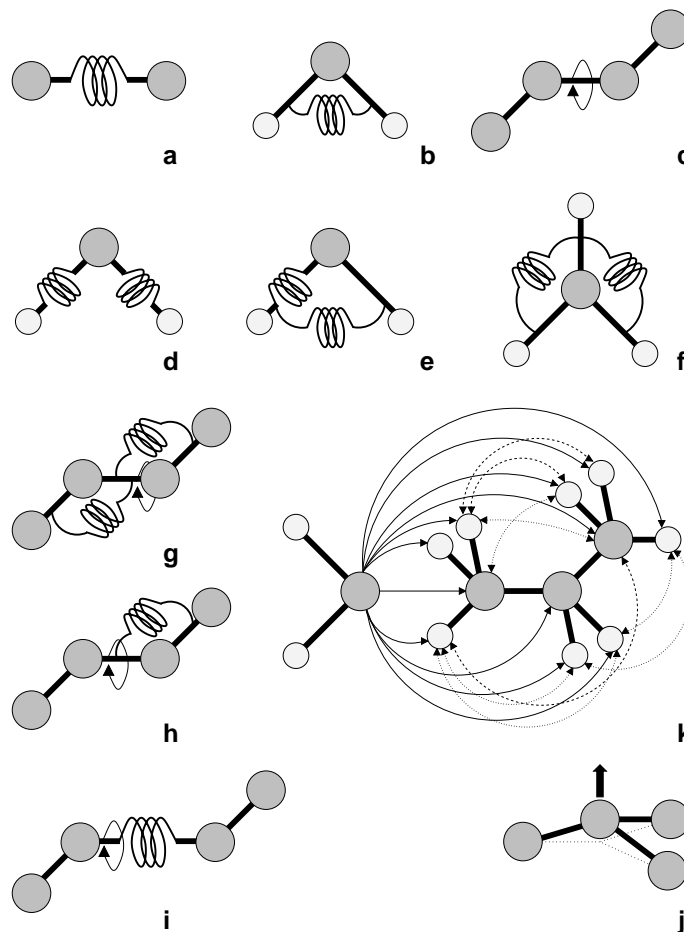


Figure 3: Ilustración de los varios términos incluidos en una función de energía potencial empírica. Las contribuciones **a-j** representan el campo de fuerzas de valencia, entre las cuales **d-j** son los términos de acoplamiento. **j** es el término de *out-of-plane*, que garantiza que el átomo central se quede en el plano formado por sus tres vecinos, a los cuales está conectado químicamente. **k** caracteriza las interacciones de Coulomb y de van der Waals de los átomos que no están conectados químicamente: Interacciones intermoleculares (trazo continuo), interacciones intermoleculares “1-4” (trazo punteado), y interacciones intramoleculares  $> 1-4$  (trazo discontinuo).

cadenas alifáticas de hecho reside en una descripción correcta de los ángulos diédricos en el curso de la simulación estadística. Solo un potencial que incluye varios términos puede dar cuenta del sutil equilibrio *trans-gauche* al largo de las cadenas lipídicas. El potencial de Ryckaert y Bellemans,<sup>8</sup>

$$\mathcal{V}(\phi) = \sum_{i=1}^6 a_i \cos^i \phi \quad (13)$$

en el cual los coeficientes  $a_i$  han sido optimizados sobre la base de la rotación interna del  $n$ -butano, es empleado a menudo para simular bicapas fosfolipídicas.

Igualmente sutil es el equilibrio entre las varias contribuciones del campo de fuerzas. La componente

torsional representa solamente una pequeña faceta de este equilibrio. La elección de los parámetros de Lennard–Jones,  $R_{ij}^*$  y  $\varepsilon_{ij}$ , y de las cargas parciales,  $q_i$ , es igualmente crítica, considerando la precisión de las cantidades termodinámicas y dinámicas calculadas. Una función de energía potencial es una compleja construcción, cuyos ladrillos de base han sido calibradas para una reproducción global de propiedades claves físicas y químicas, aunque sin la necesidad de que esos elementos tengan un sentido físico y químico.

Un de los enfoques de moda para desarrollar modelos de cargas puntuales consiste en su ajuste al potencial electrostático, la verdadera huella de la molécula. Limitando el desarrollo del potencial electrostático a nivel de monopolo, el juego optimal de las  $N_{\text{átomos}}$  cargas atómicas netas se deriva por minimización del funcional siguiente:<sup>9</sup>

$$f(\{q_k\}) = \sum_{i=1}^{N_{\text{puntos}}} \left[ V^{\text{referencia}}(\mathbf{r}_i) - \sum_{j=1}^{N_{\text{átomos}}} \frac{q_j}{r_{ij}} \right]^2 \quad (14)$$

donde  $V^{\text{referencia}}(\mathbf{r}_i)$  es el potencial electrostático evaluado en el punto  $\mathbf{r}_i$  de una rejilla de  $N_{\text{puntos}}$  alrededor de la molécula.  $V^{\text{referencia}}(\mathbf{r}_i)$  es obtenido a partir de cálculos cuánticos, *i.e.* el promedio  $\langle \Psi | 1/|\mathbf{r}_i| | \Psi \rangle$ , usualmente a un nivel sofisticado de teoría.

La determinación de los parámetros de Lennard–Jones constituye a menudo una tarea mucho más ardua. Una ruta posible para alcanzar este blanco consiste en un ajuste de las contribuciones de dispersión y de repulsión a las energías de interacción obtenidas para un número importante de cálculos cuánticos de nivel alto, realizados con distintas configuraciones. Este enfoque es generalmente aplicable a sistemas químicos de tamaño reducido, *e.g.* el heterodímero formamida–agua, en el que se busca el potencial de interacción de van der Waals. Un enfoque más heurístico, como alternativa a numerosos cálculos cuánticos, se basa sobre simulaciones estadísticas de fases condensadas. Empezando con un conjunto dado de parámetros de Lennard–Jones,  $\{R_{ij}^*, \varepsilon_{ij}\}$ , podemos reproducir con precisión cantidades termodinámicas fundamentales de un líquido puro, como su densidad,  $\rho$ , su entalpía de vaporización,  $\Delta H_{\text{vap}}$ , y posiblemente propiedades dinámicas como su coeficiente de auto-difusión,  $D$  ?

En algunos casos, la descripción minimalista impuesta por el funcional (11) puede revelarse inadecuada. En la mayoría de los campos de fuerzas comerciales, ese funcional es polivalente, aunque se ha

diseñado originalmente para estudiar biopolímeros, y específicamente proteínas y ácidos nucleicos. La investigación más fina de pequeñas moléculas orgánicas, a menudo necesaria en el campo emergente de la *de novo* de drogas, necesita un nivel de aproximación que va más allá de las hipótesis subyacentes simplísticas escondidas en la expresión (11). Entre otras, la sustitución del término armónico que describe la deformación de los enlaces químicos (Figura 3 **a**) por un potencial disociativo de Morse constituye un ejemplo interesante:

$$\mathcal{V}(r) = D_0 \left[ e^{-\alpha(r-r_0)} - 1 \right]^2 \quad (15)$$

donde  $D_0$  es precisamente la energía de disociación y  $r_0$ , la longitud de equilibrio del enlace. La descripción armónica puede revelarse inadaptada cuando los efectos de anarmonicidad no son despreciables. En los dos casos de la deformación de los enlaces químicos y de los ángulos de valencia, correcciones cúbicas y cuárticas son quizás necesarias:

$$\mathcal{V}(r) = k_r (r - r_0)^2 \left[ 1 - k'_r (r - r_0) + k''_r (r - r_0)^2 \right] \quad (16)$$

La abertura de un ángulo de valencia causa el acortamiento de los enlaces químicos asociados — un efecto claramente ausente en la ecuación (11). Para corregir esta deficiencia, la función energía potencial puede incluir de términos de acoplamiento, como:

$$\mathcal{V}(r-\theta) = k_{r\theta} (r - r_0) (\theta - \theta_0) \quad (17)$$

Además del término de *stretch–bend*, depictado en la expresión precedente, es posible introducir otros términos de acoplamientos en  $\mathcal{V}(\mathbf{r}^N)$ , como se puede ver en la Figura 3, **e–i**. A parte de la evaluación más pesada de la función de energía potencial para cada paso de tiempo,  $\delta t$ , cuando se integra las ecuaciones (2), la inclusión de los términos de acoplamiento implica un coste suplementario para parametrizar estas contribuciones.

El tratamiento de la parte electrostática, sin duda, es lo más problemático por varias razones. Primero, porque la aproximación en la ecuación (14) no es necesariamente suficiente para una descripción fiel de cualquiera molécula. Así, puede ser deseable incluir dipolos permanentes en los modelos simples de cargas atómicas netas. Más cruciales son los posibles efectos de inducción, obviamente ausentes en la aproximación aditiva de pares (10). Para paliar esta dificultad, se limita el número de soluciones

disponibles. El enfoque lo más barato computacionalmente hablando, todavía ampliamente utilizado en simulaciones numéricas de sistemas macromoleculares, consiste en inflar artificialmente la cargas puntuales, de tal manera que reproduzcan el momento dipolar característico de la fase condensada más bien que el de la fase gaseosa de baja presión. En esta representación, los efectos de polarización están tenidos en cuenta en un sentido promedio. La introducción explícita de contribuciones de polarizabilidad en el campo de fuerzas constituye la más rigurosa solución y, al mismo tiempo, la más cara. En este caso, el término electrostático completo se escribe:

$$\mathcal{V}_{\text{elec}}(\mathbf{r}^N) = \frac{1}{2} \sum_i q_i V_i \quad (18)$$

Aquí, el potencial electrostático,  $V_i$ , viene de dos fuentes distintas — por una parte, el conjunto de cargas puntuales del sistema; por otra parte, los momentos multipolares inducidos al sitio  $i$ . Nos limitando al solo dipolo inducido,  $\boldsymbol{\mu}_i$ , que está linealmente relacionado al campo eléctrico,  $\mathbf{E}_i$ , creado en este punto por todos los otros sitios polarizables  $j \neq i$ , o sea,  $\boldsymbol{\mu}_i = \alpha_i \mathbf{E}_i$ , el potencial se puede expresar como:

$$V_i = \sum_{j \neq i} \left[ \frac{q_j}{4\pi\epsilon_0 r_{ij}} + \frac{\mathbf{r}_{ij} \cdot \boldsymbol{\mu}_j}{4\pi\epsilon_0 r_{ij}^3} \right] \quad (19)$$

Mismo cuando se utilice como un punto de partida al tiempo  $t + \delta t$ , los momentos inducidos,  $\boldsymbol{\mu}_i$ , obtenidos al tiempo  $t$ , convergencia de esos en una simulación MD de un líquido polarizable aumenta el coste informático por un factor de *ca.* 3–4 respecto a una simulación incluyendo la aproximación aditiva de pares.

## 2.2. Los propagadores de dinámica molecular

Varios enfoques para integrar numéricamente las ecuaciones de Newton del movimiento (2) están actualmente disponibles. De ellos, tres se detallarán aquí. Incontestablemente lo más simple, el algoritmo de Verlet está basado sobre el conocimiento del triplet  $\{\mathbf{r}_i(t), \mathbf{r}_i(t - \delta t), \mathbf{a}_i(t)\}$ , donde  $\mathbf{a}_i(t) = \ddot{\mathbf{r}}_i(t) = d^2\mathbf{r}_i(t)/dt^2 = \mathbf{f}_i(t)/m_i$  es la aceleración de la partícula  $i$ .<sup>10</sup> La modificación de las posiciones de las partículas está realizado a través de un desarrollo de Taylor de la posición al tiempo  $t - \delta t$  y al tiempo  $t + \delta t$ , nos conduciendo a:

$$\mathbf{r}_i(t + \delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t) + \mathbf{a}_i(t) \delta t^2 + \mathcal{O}(\delta t^4) \quad (20)$$

Hay que destacar que las velocidades,  $\mathbf{v}_i(t) = \dot{\mathbf{r}}_i(t) = d\mathbf{r}_i(t)/dt$ , no aparecen explícitamente en este esquema. Desaparecen en el desarrollo de Taylor de  $\mathbf{r}_i(t + \delta t)$  y  $\mathbf{r}_i(t - \delta t)$ . Aunque no sean necesarias para describir la trayectoria, su evaluación es un paso obligatorio para calcular la energía cinética,  $\mathcal{T}(\mathbf{p})$ , que depende solo de las variables de impulsión,  $\mathbf{p}$ , y, por consecuencia, la energía total del sistema,  $\mathcal{E} \equiv \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)$ , según:

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \delta t) - \mathbf{r}_i(t - \delta t)}{2 \delta t} \quad (21)$$

A cada paso de tiempo, el error asociado es en  $\mathcal{O}(\delta t^2)$ .

El algoritmo *leap-frog*, derivado del precedente, utiliza el triplet  $\{\mathbf{r}_i(t), \mathbf{v}_i(t - \delta t/2), \mathbf{a}_i(t)\}$ . El origen de su nombre aparece claramente en la escritura del algoritmo:

$$\begin{cases} \mathbf{r}_i(t + \delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t + \frac{\delta t}{2}) \delta t \\ \mathbf{v}_i(t + \frac{\delta t}{2}) &= \mathbf{v}_i(t - \frac{\delta t}{2}) + \mathbf{a}_i(t) \delta t \end{cases} \quad (22)$$

En la práctica, el primer paso es el calculo de  $\mathbf{v}_i(t + \delta t/2)$ , con el cual  $\mathbf{v}_i(t)$  se deduce — una etapa obligatoria para evaluar el término cinético,  $\mathcal{T}(\mathbf{p})$ , según:

$$\mathbf{v}_i(t) = \frac{\mathbf{v}_i(t + \frac{\delta t}{2}) + \mathbf{v}_i(t - \frac{\delta t}{2})}{2} \quad (23)$$

Por fin, la forma velocidad del algoritmo de Verlet corrige el defecto principal del esquema estandar de Verlet o del *leap-frog*, o sea la definición de las velocidades — cuyo error asociado varia como  $\mathcal{O}(\delta t^2)$ . La incorporación explícita de las velocidades en el algoritmo de Verlet se puede expresar como:

$$\begin{cases} \mathbf{r}_i(t + \delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t) \delta t + \frac{1}{2} \mathbf{a}_i(t) \delta t^2 \\ \mathbf{v}_i(t + \delta t) &= \mathbf{v}_i(t) + \frac{\mathbf{a}_i(t) + \mathbf{a}_i(t + \delta t)}{2} \delta t \end{cases} \quad (24)$$

Este esquema incluye los dos pasos siguientes:

$$\mathbf{v}_i(t + \frac{\delta t}{2}) = \mathbf{v}_i(t) + \frac{1}{2} \mathbf{a}_i(t) \delta t \quad (25)$$

a partir del cual las fuerzas termodinámicas,  $\mathbf{f}_i$ , y las aceleraciones,  $\mathbf{a}_i$ , a tiempo  $t + \delta t$  se pueden evaluar. Sigue que:

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}_i(t + \frac{\delta t}{2}) + \frac{1}{2} \mathbf{a}_i(t + \delta t) \delta t \quad (26)$$

La energía cinética se puede deducir al tiempo  $t + \delta t$ , mientras que la energía potencial,  $\mathcal{V}(\mathbf{r}^N)$ , está calculada en los bucles de fuerzas.

En el curso de la integración simultanea de las ecuaciones de Newton del movimiento, la energía total del sistema tiene que ser conservada. Manteniendo el volumen a un valor constante, la simulación generara un ensamble estadístico microcanónico, *i.e.*  $(N, V, \mathcal{E})$ . Se habla a menudo para este tipo de simulación de MD newtoniana. Esta situación simplística, sin embargo, no es siempre satisfactoria. Puede ser deseable realizar simulaciones estadísticas, en las cuales la temperatura o la presión sean cantidades independientes en lugar de propiedades derivadas.

### 2.3. Dinámica molecular a temperatura constante

Un número de esquemas, más o menos sofisticados, para realizar simulaciones MD isotérmicas han sido propuestos. Quizas lo más simple consiste en actualizar periódicamente las velocidades con un factor  $\sqrt{T/T_{\mathcal{T}}}$ , donde  $T_{\mathcal{T}}$  es la temperatura cinética instantanea — *viz.*  $2\mathcal{T}(\mathbf{p})/3Nk_B$  — y  $T$ , la temperatura deseada. La aplicación de este factor correctivo, sin embargo, no proporciona una MD newtoniana. La mecánica newtoniana implica que la energía y el momento estén conservados. La dinámica molecular con temperatura cinética constante necesita resolver las ecuaciones del movimiento constreñidas:<sup>2</sup>

$$\begin{cases} \dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i = \mathbf{f}_i - \xi(\mathbf{r}; \mathbf{p}) \mathbf{p} \end{cases} \quad (27)$$

en las cuales  $\xi(\mathbf{r}; \mathbf{p})$  se puede relacionar con un término de fricción que garantiza  $\dot{T}_{\mathcal{J}} = 0$ . Esta restricción es elegida de tal manera que la perturbación de la trayectoria newtoniana sea mínima:

$$\xi(\mathbf{r}; \mathbf{p}) = \frac{\sum_i \mathbf{p}_i \cdot \mathbf{f}_i}{\sum_i |\mathbf{p}_i|^2} \quad (28)$$

Nota: Sustituyendo  $\xi(\mathbf{r}; \mathbf{p}) \mathbf{p}$  por  $\xi(\mathbf{r}; \mathbf{p}) \mathbf{p}_i$  en la ecuación (27) correspondería a un termostato riguroso, que lleva a la distribución canónica esperada.<sup>2</sup> Un segundo enfoque, más riguroso, consiste en introducir en las ecuaciones del movimiento un grado de libertad suplementario,  $s$ . Así se puede escribir la velocidad de la partícula  $i$  como  $\mathbf{v}_i = s \dot{\mathbf{r}}_i = \mathbf{p}_i/m_i s$ . Los términos de energía potencial y cinética están asociados al grado de libertad  $s$ , que se puede relacionar con el termostato del sistema como sigue:

$$\begin{cases} \mathcal{V}_s = (f + 1) k_B T \ln s \\ \mathcal{T}_s = \frac{1}{2} \mathcal{Q} \dot{s}^2 \end{cases} \quad (29)$$

donde  $\mathcal{Q}$  represente el parámetro de inercia termal, que regula la fluctuaciones de la temperatura, y  $f$ , el número de grados de libertad en el sistema —  $3N - 3$  si la impulsión total,  $\mathbf{p}$ , es constante. Dicho enfoque está conocido como el lagrangiano extendido, dado al hecho que este se puede expresar como  $\mathcal{L}_s(\mathbf{r}; \mathbf{p}) = \mathcal{T}(\mathbf{p}) + \mathcal{T}_s(\mathbf{p}) - \mathcal{V}(\mathbf{r}) - \mathcal{V}_s(\mathbf{r})$ . Las ecuaciones del movimiento se pueden restablecer según:

$$\begin{cases} \ddot{\mathbf{r}}_i = \frac{\mathbf{f}_i}{m_i s^2} - 2 \frac{\dot{s} \dot{\mathbf{r}}_i}{s} \\ \mathcal{Q} \ddot{s} = \sum_i m_i \dot{\mathbf{r}}_i^2 s - \frac{(f + 1) k_B T}{s} \end{cases} \quad (30)$$

Este formalismo, ideado por Nosé,<sup>11</sup> ha sido revisado por Hoover,<sup>12,13</sup> suprimiendo el tiempo-dependiente del parámetro  $s$ . En la ecuaciones constreñidas del movimiento (27), el término de fricción está ahora proporcionado por una ecuación diferencial de primer orden:

$$\dot{\xi} = \frac{f}{\mathcal{Q}} k_B (T_{\mathcal{J}} - T) \quad (31)$$

La cantidad conservada aquí es el hamiltoniano total, *i.e.* él del sistema químico más el termostato,  $\mathcal{H}_s(\mathbf{r}; \mathbf{p}) = \mathcal{T}(\mathbf{p}) + \mathcal{T}_s(\mathbf{p}) + \mathcal{V}(\mathbf{r}) + \mathcal{V}_s(\mathbf{r})$ .

El último enfoque, llamado acoplamiento débil o *weak coupling*,<sup>14</sup> consiste en dejar la temperatura cinética instantánea,  $T_{\mathcal{T}}(t)$ , “relajar” hacia la temperatura de referencia deseada,  $T$ , según:

$$\frac{dT_{\mathcal{T}}(t)}{dt} = \frac{T - T_{\mathcal{T}}(t)}{\tau_T} \quad (32)$$

donde  $\tau_T$  represente precisamente el tiempo de relajación asociado a las fluctuaciones de la temperatura. La energía cinética es modificada por una cantidad  $\Delta\mathcal{T}$ , definida como:

$$\Delta\mathcal{T} = \frac{1}{2} (\chi^2 - 1) N k_B T_{\mathcal{T}}(t) \quad (33)$$

durante un paso de tiempo,  $\delta t$ , reactualizando las velocidades por un factor  $\chi$ :

$$\chi = \left[ 1 + \frac{\delta t}{\tau_T} \left( \frac{T}{T_{\mathcal{T}}(t)} - 1 \right) \right]^{1/2} \quad (34)$$

Este acoplamiento aperiódico a una reserva de calor, gracias a un proceso de primer orden, no lleva a respuestas oscilantes a los cambios de temperatura. No obstante, como se ha demostrado, no proporciona la distribución canónica correcta, al contrario del enfoque de Nosé y Hoover.

## 2.4. Dinámica molecular a presión constante

Aquí, de nuevo, varios métodos alternativos, más o menos sofisticados, se pueden adoptar para mantener la caja de simulación a una presión constante en el curso del tiempo. Sería efectivamente deseable, en un número de circunstancias, generar trayectorias en un ensamble termodinámico isobarico–isotermico,  $(N, P, T)$ .

Como para mantener la temperatura a un valor constante, el formalismo del lagrangiano extendido se puede aplicar a la presión. Inicialmente imaginado por Andersen,<sup>15</sup> este enfoque implica que el sistema esté acoplado a una variable externa,  $V$ , caracterizando el volumen de la caja de simulación. Este acoplamiento simboliza la acción que un piston ejercería sobre el sistema, al cual un término potencial y cinético está asociado:



$$\begin{cases} \mathcal{V}_V = \frac{1}{2} m_P \dot{V}^2 \\ \mathcal{T}_V = P V \end{cases} \quad (35)$$

donde  $m_P$  se puede ver como la masa del piston y  $P$  represente la presión deseada. Actualizando las variables de posición,  $\mathbf{r}$ , y de velocidad,  $\mathbf{v}$ , bajo la forma  $\mathbf{s} = \mathbf{r}/V^{1/3}$  y  $\dot{\mathbf{s}} = \mathbf{v}/V^{1/3}$ , se puede reescribir las energías potenciales y cinéticas como  $\mathcal{V}(\mathbf{r}) \equiv \mathcal{V}(V^{1/3}\mathbf{s})$  y  $\mathcal{T}(\mathbf{p}) = \frac{1}{2} m V^{2/3} \sum_i \dot{s}_i^2$ . Sigue que a partir del lagrangiano,  $\mathcal{L}_V(\mathbf{r}; \mathbf{p}) = \mathcal{T}(\mathbf{p}) + \mathcal{T}_V(\mathbf{p}) - \mathcal{V}(\mathbf{r}) - \mathcal{V}_V(\mathbf{r})$ , se puede establecer las ecuaciones del movimiento:

$$\begin{cases} \ddot{\mathbf{s}}_i = \frac{\mathbf{f}_i}{m_i V^{1/3}} - \frac{2}{3} \frac{\dot{\mathbf{s}}_i \dot{V}}{V} \\ \ddot{V} = \frac{P_{\mathcal{D}} - P}{m_P} \end{cases} \quad (36)$$

donde la fuerza,  $\mathbf{f}_i$ , y la presión instantanea,  $P_{\mathcal{D}}$  — derivada del virial,  $P_{\mathcal{D}} = \frac{1}{V} (Nk_B T - \frac{1}{2} \sum_i \mathbf{r}_i \cdot \mathbf{f}_i)$  — están evaluadas a partir las coordenades cartesianas y impulsiones antes su actualización. Aquí, la cantidad conservada es el hamiltoniano del sistema extendido,  $\mathcal{H}_V(\mathbf{r}; \mathbf{p}) = \mathcal{T}(\mathbf{p}) + \mathcal{T}_V(\mathbf{p}) + \mathcal{V}(\mathbf{r}) + \mathcal{V}_V(\mathbf{r})$ , o sea su entalpía, a la cual se añade la contribución cinética de  $\frac{1}{2}k_B T$  que viene de las fluctuaciones del volumen de la caja de simulación. Hay que destacar que, formalmente, este algoritmo genera una distribución isobarica–isoentalpica,  $(N, P, H)$ . Su acoplamiento a un termostato, como él gobernado por la ecuación (30), proporciona la verdadera distribución isobarica–isotermica.

Ha estado observado que el esquema precedente podría llevar a oscilaciones de  $P_{\mathcal{D}}$ , en función de la masa del piston,  $m_P$ . Feller *et al.* han desarrollado una alternativa que suprime este efecto indeseable, amortiguando el grado de libertad del piston a través de una ecuación de Langevin. Restableciendo la ecuación (36), sigue que:

$$\begin{cases} \ddot{\mathbf{s}}_i = \frac{\mathbf{f}_i}{m_i V^{1/3}} - \frac{2}{3} \frac{\dot{\mathbf{s}}_i \dot{V}}{V} \\ \ddot{V} = \frac{P_{\mathcal{D}} - P}{m_P} - \gamma \dot{V} + R(t) \end{cases} \quad (37)$$

donde  $\gamma$  es la frecuencia de colisión y  $R(t)$ , una fuerza aleatoria tomada de una distribución gaussiana de promedio cero y de variancia  $\langle R(0)R(t) \rangle = 2\gamma k_B T \delta(t)/m_P$ . Es interesante notar que  $R(t)$

satisface la relación de fluctuación–disipación, *i.e.*  $\langle R(t_1)R(t_2) \rangle = \frac{1}{m_P} 2k_B T \kappa(t_1 - t_2)$ , en la cual  $\kappa(t)$  represente el factor de amortización.

Un otro enfoque, propuesto por Berendsen *et al.*,<sup>14</sup> es una extensión a las simulaciones con presión constante del algoritmo de acoplamiento débil descrito antes. Así como el esquema de temperatura constante, las ecuaciones del movimiento están modificadas en respuesta de la relajación de la presión instantánea,  $P_{\mathcal{P}}(t)$ , hacía su valor de referencia,  $P$ , según:

$$\frac{dP_{\mathcal{P}}(t)}{dt} = \frac{P - P_{\mathcal{P}}(t)}{\tau_P} \quad (38)$$

Aquí,  $\tau_P$  es el tiempo de relajación asociado a las fluctuaciones de la presión. Reactualizando las coordenadas atómicas del sistema y el tamaño de la caja por un factor  $\varsigma$ , el volumen total está modificado por  $\Delta V = (\varsigma^3 - 1) V$ , llevando naturalmente a una variación de la presión, lo que se puede expresar como:

$$\Delta P = \frac{\Delta V}{\beta_{\mathcal{P}} V} \quad (39)$$

en el cual  $\beta_{\mathcal{P}}$  represente la compresibilidad isothermal. Solucionando ecuaciones (38) y (39) por un valor dado de  $\varsigma$ , sigue que:

$$\varsigma = \left[ 1 - \beta_{\mathcal{P}} \delta t \frac{P - P_{\mathcal{P}}(t)}{\tau_P} \right]^{1/3} \quad (40)$$

Por razones similares al caso de la temperatura constante, este algoritmo no proporciona un ensamble termodinámico bien definido.

## 2.5. Interacciones electrostaticas: Más allá de las aproximaciones usuales

Un aspecto particularmente crítico de la simulaciones MD se encuentra en el tratamiento apropiado de las interacciones electrostaticas. Por obvias razones de costo informático y eficiencia, la truncatura esférica, mencionada antes, aún sigue siendo ampliamente utilizada, especialmente cuando escalas de tiempo significativas, más allá que el rango del nanosegundo, están exploradas — *e.g.* simulaciones de bicapas totalmente hidratadas de fosfolípidos. Si la naturaleza de rango largo de las interacciones

dipolo–dipolo en  $1/r^3$  está suficientemente limitada para garantizar una reproducción satisfactoria de propiedades estructurales y de promedios estadísticos de ensamble del sistema, la influencia de esas interacciones sobre procesos físicos y químicos complejos, como el plegamiento de proteínas, necesitaría un análisis detallado. La presencia de especies iónicas es claramente problemática, considerando que el uso de una esfera de corte induce numerosos artefactos que evidentemente distorsionan los resultados de la simulación. Además, hay que destacar que, nos limitando a especies dipolares, emplear una esfera de corte causa singularidades en las derivadas de la energía potencial a la frontera de *cut-off*. Los efectos deletéreos se pueden reducir añadiendo una función de *switching*,<sup>2</sup> que sustituye al comportamiento abrupto de tipo Heaviside un decrecimiento suave en la cercanía de la frontera esférica. Este método sin embargo no palia las dificultades naciendo de la presencia de iones en el sistema. Un enfoque más riguroso para tratar interacciones de tipo carga–carga y carga–dipolos, aparentemente barato, está basado sobre la teoría de Debye–Hückel y la solución de la ecuación linealizada de Poisson–Boltzmann.<sup>16</sup> Como está ilustrado en la Figura 5, esta alternativa, bautizada *campo de reacción generalizado*, no elimina por completo los artefactos cuando  $r \rightarrow R_{\text{cut-off}}$ , el radio de la esfera de corte. Tal vez el enfoque lo más riguroso es el propuesto por Ewald. Admitiendo que la sumatoria de Coulomb,

$$\mathcal{V}_{\text{Coulomb}}(\mathbf{r}) = \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \quad (41)$$

sobre la caja de simulación y sus imágenes vecinas no converge formalmente, la idea central del método, conocido como *sumas de Ewald*, consiste en un recorte de la expresión (41) en dos sumatorias evaluadas, respectivamente, en los espacios directo y recíproco:

$$\sum_{\mathbf{n}} \frac{1}{|\mathbf{n}|} \mathcal{F}(\mathbf{n}) + \sum_{\mathbf{m}} \frac{1}{|\mathbf{m}|} [1 - \mathcal{F}(\mathbf{n})] \quad (42)$$

Rodeando cada carga puntual del sistema por una distribución gaussiana de carga:

$$\varrho_i(\mathbf{r}) = q_i \alpha^3 \frac{\exp(-\alpha^2 r^2)}{\sqrt{\pi^3}} \quad (43)$$

donde  $\alpha$  es un parámetro positivo caracterizando la anchura de la distribución gaussiana, la primera suma converge rápidamente cuando  $\mathbf{n} \rightarrow \infty$ , porque  $\mathcal{F}(\mathbf{n})$  decrece rápidamente. La contribución del espacio directo es de rango corto — Figura (4). La segunda sumatoria evaluada en el espacio

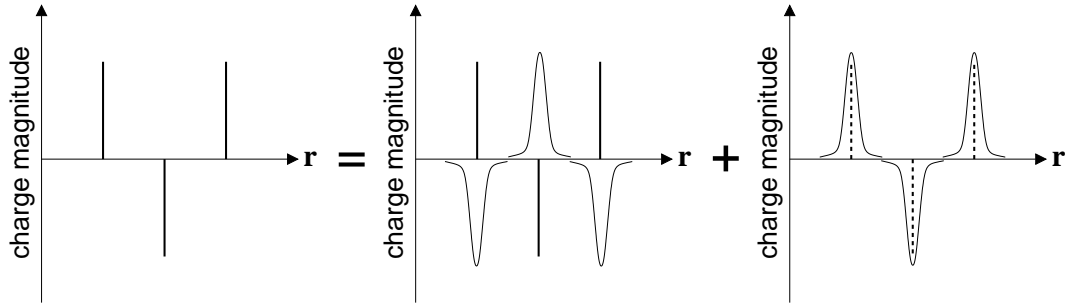


Figure 4: Componentes de una suma de Ewald en un sistema unidimensional de cargas puntuales. En el espacio directo, cada carga está rodeada por una distribución de carga gaussiana,  $\rho_i(\mathbf{r})$ , de igual amplitud, pero de signo inverso. Esta contribución está contrabalanceada en el espacio recíproco por una distribución gaussiana,  $\rho_j(\mathbf{r})$ , de signo inverso.

recíproco utiliza una transformada de Fourier para solucionar la ecuación de Poisson, *i.e.*  $\nabla^2 \mathcal{V}_i(\mathbf{r}) = -4\pi \rho_i(\mathbf{r})$ . La transformada decrece rápidamente y la suma converge igualmente rápidamente.<sup>17</sup>

Implementación de esas sumas de red según el esquema imaginado por Ewald en un código de MD que usa un campo de fuerzas macromolecular, como él descrito en la ecuación (11), se puede resumir como sigue:

$$\begin{aligned}
 \mathcal{V}_{\text{Ewald}}(\mathbf{r}) &= \frac{1}{2V\epsilon_0} \sum_{\mathbf{k} \neq \mathbf{0}} \frac{\exp(-k^2/4\alpha^2)}{k^2} \left[ \sum_j q_j \exp(-i\mathbf{k} \cdot \mathbf{r}_j) \right] \left[ \sum_j q_j \exp(i\mathbf{k} \cdot \mathbf{r}_j) \right] \\
 &+ \frac{1}{4\pi\epsilon_0} \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}} \text{erfc}(\alpha r_{ij}) - \frac{\alpha}{4\pi^{3/2}\epsilon_0} \sum_i q_i^2 \\
 &+ \frac{1}{4\pi\epsilon_0} \sum_i \sum_{\substack{j \text{ conectado a } i \\ j>i}} \frac{q_i q_j}{r_{ij}}
 \end{aligned} \tag{44}$$

El primero término corresponde a la sumatoria en el espacio recíproco sobre vectores  $\mathbf{k}$ . El segundo término es la sumatoria del espacio directo;  $\text{erfc}(x)$  es la función error complementaria, *i.e.*  $1 - \text{erf}(x)$ , que explica la naturaleza de rango corto de este término convergiendo rápidamente. Las tercera y cuarta contribuciones son correcciones, dado que la suma en el espacio recíproco está realizada sobre *todas* pares atómicas  $\{i, j\}$ , que necesariamente incluye términos de *self*, 1-2 y 1-3, ausentes en la expresión (11).

Formalmente, el costo informático implicado en una suma de red clásica de Ewald es  $\mathcal{O}(N^2)$ , donde  $N$  es el número total de partículas del sistema. Como lo ha enseñado Perram *et al.*, por una parte, y Fincham, por otra parte, este costo se puede disminuir a  $\mathcal{O}(N^{3/2})$  eligiendo juiciosamente la anchura

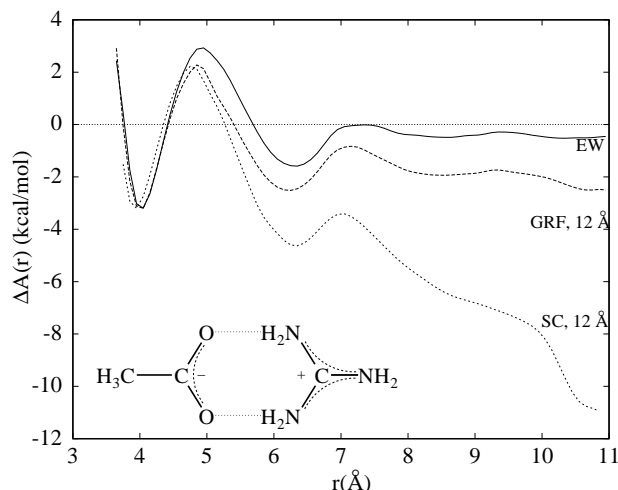


Figure 5: Perfil de energía libre depictingo la asociación reversible de un cation guanidinium con un anion acetata en la geometría  $\mathcal{C}_{2v}$ , en un entorno acuoso.<sup>18</sup> El trazo continuo corresponde a una simulación con sumas de Ewald (EW) —  $\alpha = 0.3 \text{ \AA}^{-1}$ ; El trazo discontinuo, a una simulación con el campo de reacción generalizado (GRF) —  $R_{\text{cut-off}} = 12 \text{ \AA}$ ; El trazo punteado, a una simulación con una esfera de corte (SC) para tratar las interacciones intermoleculares —  $R_{\text{cut-off}} = 12 \text{ \AA}$ .

de la distribución gaussiana,  $\alpha$ , el número de vectores  $\mathbf{k}$ , y la truncatura de las interacciones de todas pares en el espacio directo. Esta generalmente aconsejado que el tiempo CPU invertido en las sumas de los espacios directo y recíproco estén bien balanceadas para alcanzar el crecimiento  $N^{3/2}$  deseado.

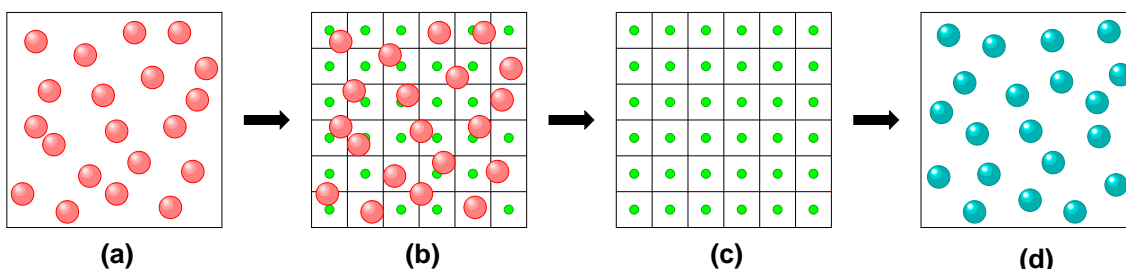


Figure 6: Un esquema *particle-mesh* sobre una red bidimensional. **(a)** Un sistema de partículas cargadas. **(b)** Las cargas están interpoladas sobre la rejilla bidimensional. **(c)** Utilizando una transformada de Fourier rápida (FFT), el potencial y las fuerzas están evaluadas en cada punto de la rejilla. **(d)** Las fuerzas están interpoladas hacia las partículas, cuyas posiciones están actualizadas.

Alternativas menos caras que la suma estandar de Ewald se apoyan sobre una evaluación de la contribución del espacio recíproco usando una transformada de Fourier rápida (FFT). Se construye una rejilla tridimensional llenando el espacio cartesiano en el cual la simulación de MD está realizada. Las cargas puntuales llevadas por las partículas del sistema están interpoladas sobre la rejilla, y la distribución de carga correspondiente,  $\varrho(\mathbf{r})$ , está calculada. Utilizando un técnico FFT, la transformada de la distribución de carga,  $\hat{\varrho}(\mathbf{k})$ , está determinada sobre la base de los vectores  $\mathbf{k}$  del espacio

recíproco. Luego, la contribución de rango largo del potencial electrostático está evaluada según  $\hat{V}_{\text{long}}(\mathbf{k}) = \hat{\mathcal{G}}(\mathbf{k})\hat{\rho}(\mathbf{k})$ , donde  $\hat{\mathcal{G}}(\mathbf{k})$  es la función de *influencia*, definida por  $\hat{\mathcal{G}}(\mathbf{k}) = \hat{\lambda}(\mathbf{k})/\epsilon_0 k^2$ , en la cual  $\lambda(\mathbf{r})$  es una distribución que depende de las solas características de la caja de simulación. La contribución  $V_{\text{long}}(\mathbf{r})$  está estimada en los varios puntos de la rejilla tridimensional gracias a una transformada inversa. Luego, las fuerzas electrostáticas están determinadas por derivación numérica del potencial. Por último, el campo eléctrico y el potencial están interpolados desde la rejilla hacia las posiciones de las partículas — Figura (6). Este esquema constituye la idea central de todos algoritmos de tipo *particle–mesh*,<sup>19,20</sup> e.g. *particle–mesh Ewald* (PME) o *particle–particle–particle–mesh*<sup>21</sup> (P<sup>3</sup>M), cuyo costo CPU evoluciona formalmente en  $\mathcal{O}(N \ln N)$ .

## 2.6. Algunas propiedades accesibles a partir de la trayectoria

Entre las propiedades estáticas que se pueden extraer de las trayectorias generadas en el curso de una simulación MD, las cantidades estructurales son particularmente interesantes, porque dan cuenta del orden local en el sistema molecular. Las funciones de distribución radiales (RDF),  $g(r)$ , indican la probabilidad de encontrar un par de átomos separados por una distancia  $r$ , respecto a la probabilidad esperada para una distribución totalmente aleatoria, con la misma densidad.<sup>2,22,23</sup> La definición de  $g(r)$  necesita la integración de la función de distribución configuracional sobre todas las posiciones atómicas, excepto la de las dos partículas marcadas:

$$g(\mathbf{r}_1; \mathbf{r}_2) = \frac{N(N-1)}{\rho^2 \int e^{-\mathcal{V}(\mathbf{r}_1, \dots, \mathbf{r}_N)/k_B T} d\mathbf{r}_1 \dots d\mathbf{r}_N} \int e^{-\mathcal{V}(\mathbf{r}_1, \dots, \mathbf{r}_N)/k_B T} d\mathbf{r}_3 \dots d\mathbf{r}_N \quad (45)$$

donde  $\rho$  represente la densidad del líquido. Para un sistema en el cual todos los átomos serían idénticos, la RDF se reduciría a un simple promedio estadístico sobre pares de átomos:

$$g(r) = \frac{V}{N^2} \left\langle \sum_i \sum_{j \neq i} \delta(\mathbf{r} - \mathbf{r}_{ij}) \right\rangle \quad (46)$$

Para un par dado de átomos,  $\{i, j\}$ , la RDF se puede evaluar inmediatamente a través de la relación siguiente:

$$g_{ij}(r) = \frac{\langle n_j(r + \delta r) \rangle}{4\pi\rho_j \int r^2 dr} \quad (47)$$

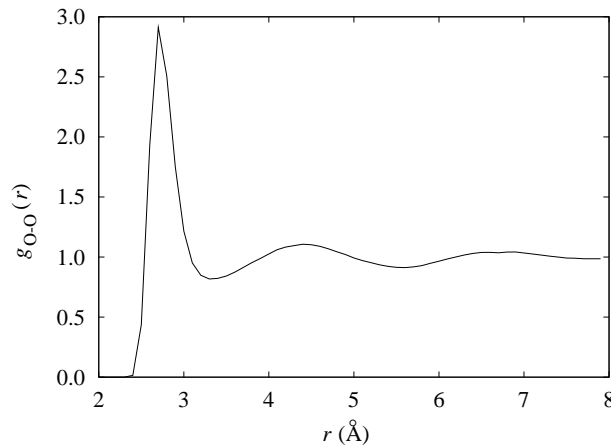


Figure 7: Función de distribución radial (RDF) del modelo TIP4P del agua, obtenida a partir de una simulación de dinámica molecular a 300 K.

donde  $\langle n_j(r + \delta r) \rangle$  corresponde al número promedio de sitios  $j$ , por los cuales la distancia  $i$  está comprendida entre  $r$  y  $r + \delta r$ ;  $\rho_j$  es una densidad promedio de sitios  $j$  en la muestra de interés.

Una segunda cantidad, igualmente pertinente para apreciar el orden en un líquido molecular, es el distancia-dependiente factor de Kirkwood,  $G_K(R)$ .<sup>24</sup> Esta cantidad caracteriza la correlación entre el momento dipolar,  $\mu_i$ , llevado por una molecular  $i$  dada y el llevado por moléculas vecinas  $j$ , contenidas en una esfera de radio  $R$  y centrado al sitio  $i$ :

$$G_K(R) = \frac{\left\langle \sum_{i,j;r_{ij}<R} \mu_i \cdot \mu_j \right\rangle}{N\mu^2} \quad (48)$$

El  $R$ -dependiente factor de Kirkwood es particularmente sensible al tratamiento de las interacciones electrostáticas en la simulación molecular. Utilizando un radio de corte causa varios artefactes que deterioran la correlación dipolo-dipolo cerca del borde de la esfera, y luego la estimación de la permitividad dieléctrica que se puede inferir a partir de  $G_K(R)$ .

MD, como lo sugere su nombre, puede también proporcionar informaciones preciosas sobre las propiedades dinámicas del sistema investigado. Entre éstas, las funciones de correlación temporal son especialmente informativas a propósito de los tiempos de relajación para los varios grados de libertad del sistema.<sup>1,2</sup> La correlación entre dos cantidades observables,  $\mathcal{B}$  y  $\mathcal{B}'$ , se puede expresar por:

$$c_{\mathcal{B}\mathcal{B}'} = \frac{\langle \delta\mathcal{B} \delta\mathcal{B}' \rangle}{\sigma(\mathcal{B}) \sigma(\mathcal{B}')} \quad (49)$$

donde  $\delta\mathcal{B} = \mathcal{B} - \langle \mathcal{B} \rangle$ , con  $\langle \mathcal{B} \rangle$ , es el promedio del ensamble estadístico de la cantidad  $\mathcal{B}$ ;  $\sigma(\mathcal{B}) = \sqrt{\langle \mathcal{B}^2 \rangle - \langle \mathcal{B} \rangle^2}$ .  $c_{\mathcal{B}\mathcal{B}'}$  varía entre 0 y 1, 0 corresponde a una ausencia de correlación. Esta formulación se puede generalizar en el caso donde  $\mathcal{B}$  y  $\mathcal{B}'$  están evaluadas a tiempos distintos. Sigue que la correlación ahora se escribe:

$$c_{\mathcal{B}\mathcal{B}'}(t) = \frac{\langle \delta\mathcal{B}(t) \delta\mathcal{B}'(0) \rangle}{\sigma(\mathcal{B}) \sigma(\mathcal{B}')} \quad (50)$$

El promedio temporal al numerator está realizado sobre todos orígenes del tiempo. En la eventualidad que  $\mathcal{B}' \equiv \mathcal{B}$ , la función calculada se llama *función de autocorrelación*, definida por:

$$c_{\mathcal{B}\mathcal{B}}(t) = \frac{\langle \delta\mathcal{B}(t) \delta\mathcal{B}(0) \rangle}{\langle \delta\mathcal{B}(0) \delta\mathcal{B}(0) \rangle} \quad (51)$$

Integración de la función de correlación entre  $t = 0$  y  $t = \infty$  entonces lleva al tiempo de correlación. Como una ilustración, consideremos el tiempo de correlación reorientacional del benceno en agua. El tiempo de simulación es necesariamente más largo que el para un líquido puro, dado que el promedio estadístico está estimado a partir de una sola molécula.

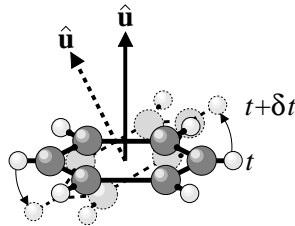


Figure 8: Reorientación de la molécula de benceno en función del tiempo.  $\hat{u}$  es el vector unidad llevado por el eje  $C_6$  de la molécula.

La cantidad dinámica evaluada aquí es el coeficiente de autocorrelación  $c_{\hat{u}\hat{u}}(t) = \langle \hat{u}(t) \hat{u}(0) \rangle$ , donde  $\hat{u}$  es el vector unidad llevado por el eje  $C_6$  o  $C'_2$  del benceno. Mirando al tensor de inercía de la molécula, se puede anticipar que el movimiento rotacional alrededor de  $C'_2$  se decorrelacione más



rápido que el alrededor de  $C_6$ .

A partir del conocimiento de las funciones de correlación para las derivadas,  $\dot{\mathcal{B}}$ , de la observable,  $\mathcal{B}$ , en lugar de la observable misma, las simulaciones numéricas permiten acceder a los coeficientes de transporte. MD al equilibrio ofrece la posibilidad de estimar coeficientes de difusión gracias a la integral siguiente:<sup>2</sup>

$$D = \frac{1}{3} \int_0^\infty \langle \dot{\mathbf{r}}_i(t) \cdot \dot{\mathbf{r}}_i(0) \rangle dt \quad (52)$$

Aquí,  $\dot{\mathbf{r}}_i(t) \equiv \mathbf{v}_i(t)$  es la velocidad del centro de masa de la molécula. Sobre tiempos suficiente largos,  $D$  se puede obtener utilizando la relación de Einstein:

$$D = \frac{1}{3} \frac{\langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \rangle}{2t} \quad (53)$$

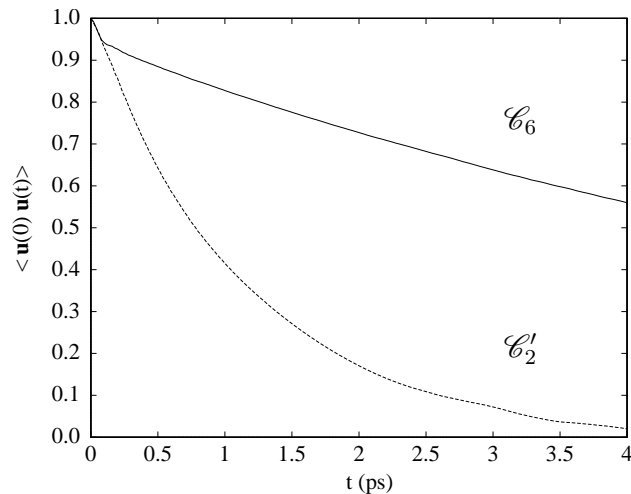


Figure 9: Funciones de autocorrelación reorientacional del benceno en agua líquido, a 300 K. Integración de los perfiles proporciona los tiempos de correlación  $\tau(C_6) = 2.4$  ps y  $\tau(C'_2) = 0.5$  ps.

### 3. Conclusion

Las pocas ilustraciones presentadas aquí ofrecen un vistazo dentro del increíble campo de investigación accesible a las simulaciones numéricas para abordar problemas encontrados cada día por el modelizador. Este campo de investigación crece progresivamente a medida que decrece la razón precio/potencia de computadores modernos. Además, el acceso a las arquitecturas masivamente par-

alleles, todavía algo desigual en función del país, abre la vía hacia aplicaciones complejas y realistas, que se pueden tratar por simulaciones MD. Estudios *in silico* de sistemas macromoleculares de interés físico, químico y biológico, sobre escalas de tiempo significativas, son ejemplos patentes de los progresos hechos en los años recientes, no solo sobre el frente del *hardware*, sino también el del *software*. No obstante, estos progresos sugieren un número de comentarios. Primero, la calidad de los cálculos realizados está fuertemente condicionada por el protocolo de simulación elegido. En función del problema abordado, hay que desterrar varias aproximaciones y se debe recurrir a métodos más rigurosos, que a menudo son más costosos. Es el caso, por ejemplo, del tratamiento de las interacciones electrostáticas, que, en su forma lo más simple, está limitado a una aproximación aditiva de pares y una truncatura esférica de las fuerzas de rango largo, pero una descripción más sofisticada puede incluir sumas de red y tal vez modelos de polarizabilidades distribuidas. Si hoy en día la metodología está bien establecida y es robusta, la simulación de *todo*, utilizando MD, no es todavía posible, dado la integración numérica de las ecuaciones de Newton del movimiento con pasos de tiempo infinitesimales, *ca.* 1 to 2 fs, que se deben comparar con las escalas de tiempo exploradas, conjuntamente con el tamaño del sistema examinado. Éste puede constituir un insalvable obstáculo al uso cotidiano de dichos cálculos de escalas largas. Además, es tentador usar sistemáticamente simulaciones numéricas mientras que el modelaje molecular tiene tendencia a proliferar en un creciente número de campos de la física, de la química y de la biología, en la vasta mayoría de los casos, códigos de MD no se deben considerar como *cajas negras*. Solo un examen prudente de los resultados obtenidos, un análisis detallado de las trayectorias generadas y una comprobación exhaustiva de las propiedades termodinámicas derivadas confirmarían la integridad y el sentido químico y físico de los cálculos realizados.<sup>25</sup>

## Acknowledgments

Agradezco al Centre Charles Hermite (CCH) por haberme ofrecido una cantidad generosa de tiempo CPU sobre su arquitectura paralela SGI Origin 3000. Expreso también mi agradecimiento a François Dehez, Jérôme Delhommelle, Mounir Tarek, José-Luis y Michael A. Wilson por haber aceptado leer este manuscrito y por sus comentarios inspirados y sugerencias útiles para mejorar la calidad y la legibilidad del texto.

## References

- [1] Frenkel, D.; Smit, B., *Understanding molecular simulations: From algorithms to applications*, Academic Press: San Diego, 1996.
- [2] Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*, Clarendon Press: Oxford, 1987.
- [3] van Gunsteren, W. F.; Berendsen, H. J. C., Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry, *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992–1023.
- [4] Tuckerman, M. E.; Martyna, G. J., Understanding modern molecular dynamics: Techniques and applications, *J. Phys. Chem. B* **2000**, *104*, 159–178.
- [5] Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L., Explicit reversible integrators for extended systems dynamics, *Mol. Phys.* **1996**, *87*, 1117–1128.
- [6] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. C.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- [7] Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. The development/application of a “minimalist” force field using a combination of ab initio and experimental data. in *Computer simulation of biomolecular systems: Theoretical and experimental applications*, Van Gunsteren, W. F.; Weiner, P. K., Eds. Escom, The Netherlands, 1997, pp. 83–96.
- [8] Ryckaert, J.; Bellemans, A., Molecular dynamics of liquid alkanes, *Chem. Soc. Faraday Discuss.* **1978**, *66*, 95–106.
- [9] Cornell, W. D.; Chipot, C. Alternative approaches to charge distribution calculations. in *Encyclopedia of computational chemistry*, Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F. and ; Schreiner, P. R., Eds., vol. 1. Wiley and Sons, Chichester, 1998, pp. 258–263.
- [10] Verlet, L., Computer “experiments” on classical fluids. I. Thermodynamical properties of Lennard–Jones molecules, *Phys. Rev.* **1967**, *159*, 98–103.

- [11] Nosé, S., A molecular dynamics method for simulations in the canonical ensemble, *Mol. Phys.* **1984**, *52*, 255–268.
- [12] Hoover, W. G., Nonequilibrium molecular dynamics, *A. Rev. Phys. Chem.* **1983**, *34*, 103–127.
- [13] Hoover, W. G., Canonical dynamics: Equilibrium phase–space distributions, *Phys. Rev.* **1985**, *A31*, 1695–1697.
- [14] Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- [15] Andersen, H. C., Molecular dynamics simulations at constant pressure and/or temperature, *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- [16] Barker, J. A., Reaction field, screening, and long–range interactions in simulations of ionic and dipolar systems, *Mol. Phys.* **1994**, *83*, 1057–1064.
- [17] Toukmaji, A. Y.; Board Jr., J. A., Ewald summation techniques in perspective: A survey, *Comput. Phys. Comm.* **1996**, *95*, 73–92.
- [18] Rozanska, X.; Chipot, C., Modeling ion–ion interaction in proteins: A molecular dynamics free energy calculation of the guanidinium–acetate association, *J. Chem. Phys.* **2000**, *112*, 9691–9694.
- [19] Darden, T. A.; York, D. M.; Pedersen, L. G., Particle mesh Ewald: An  $N\log N$  method for ewald sums in large systems, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- [20] Essman, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method, *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- [21] Hockney, R. W.; Eastwood, J. W., *Computer simulation using particles*, IOP Publishing Ltd.: Bristol, England, 1988.
- [22] McQuarrie, D. A., *Statistical mechanics*, Harper and Row: New York, 1976.
- [23] Chandler, D., *Introduction to modern statistical mechanics*, Oxford University Press, 1987.
- [24] Neumann, M.; Steinhauser, O., The influence of boundary conditions used in machine simulations on the structure of polar systems, *Mol. Phys.* **1980**, *39*, 437–54.

- [25] van Gunsteren, W. F.; Mark, A. E., Validation of molecular dynamics simulation, *J. Chem. Phys.* **1998**, *108*, 6109–6116.