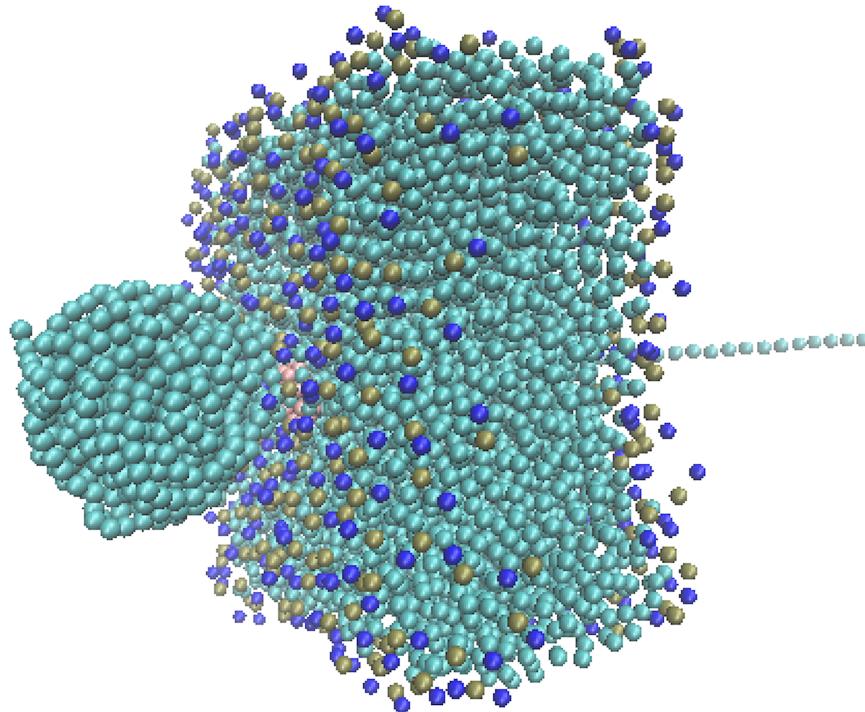


Trabajo de fin de máster

# TRASLOCACIÓN DE POLÍMEROS A TRAVÉS DE NANOPOROS

Guillermo Díez Señorans



Dirigido por Dr. Fernando Falo Forniés y Dr. Alessandro Fiasconaro.

Departamento de Física de la Materia Condensada

Universidad de Zaragoza

# Índice

<b>1. Introducción y objetivos</b>	<b>1</b>
<b>2. Física de la traslocación de polímeros</b>	<b>3</b>
<b>3. Máquina de simulación química <i>Gromacs</i> [14]</b>	<b>4</b>
3.1. Entorno de simulación . . . . .	4
3.2. Interacciones en <i>Gromacs</i> . . . . .	5
3.2.1. Interacciones <i>non-bonded</i> . . . . .	5
3.2.2. Interacciones <i>bonded</i> . . . . .	5
3.2.3. Esquema <i>group</i> . . . . .	6
3.2.4. Esquema <i>verlet</i> . . . . .	6
3.3. Algoritmos de simulación . . . . .	7
3.3.1. Minimización energética: <i>steepest-descendent</i> . . . . .	7
3.3.2. Integración de las ecuaciones del movimiento: <i>leap-frog</i> estocástico . . . . .	7
3.4. Condiciones iniciales de la simulación . . . . .	8
<b>4. Simulaciones <i>coarse graining</i>: campo de fuerzas <i>MARTINI</i></b>	<b>9</b>
4.1. Descripción del modelo . . . . .	10
4.2. Interacciones <i>bonded</i> y <i>non-bonded</i> . . . . .	10
4.3. Parámetros constantes de la simulación con <i>MARTINI</i> . . . . .	11
<b>5. Simulación de experimentos de traslocación de biopolímeros a través de nanoporos en la mesoescala</b>	<b>12</b>
5.1. Diseño del sistema . . . . .	12
5.1.1. Bicapa lipídica . . . . .	12
5.1.2. Poro . . . . .	13
5.1.3. Polímero . . . . .	15
5.2. Análisis estructural del nanoporo . . . . .	17
5.3. Etapas de la simulación . . . . .	19
<b>6. Resultados</b>	<b>21</b>
6.1. Verificación del modelo: ley de escala del tiempo de traslocación con el tamaño del polímero . . . . .	23
6.2. Curvas de fuerza y posición durante la traslocación: condiciones estándar . . . . .	23
6.3. Influencia de la velocidad de tracción y la fricción en las trayectorias . . . . .	26
6.4. Mapa del trabajo de traslocación frente a las propiedades mecánicas del poro . . . . .	26
<b>7. Resumen del trabajo realizado y conclusiones</b>	<b>28</b>

## 1. Introducción y objetivos

El fenómeno de la traslocación de polímeros, esto es, el desplazamiento de una cadena polimérica entre dos regiones del espacio (llamadas *cis* → *trans*, fig. 1) separadas por una membrana, es un proceso ubicuo en la biología molecular. La necesidad de una homeostasis efectiva para mantener a la célula aislada del medio limita la composición (y por tanto las propiedades mecánicas, en particular la permeabilidad) de la membrana lipídica. Por lo tanto la traslocación a través de nanoporos (*i.e* poros de dimensiones nanométricas) en dicha membrana aparece, en última instancia, como el mecanismo definitivo capaz de garantizar el transporte de información y energía entre el interior y exterior celular, e incluso entre regiones aisladas en el interior de las propias células (como es el caso del núcleo en células eucariotas, donde se almacena el material genético). Así por tanto no es sorprendente que [la traslocación de polímeros] constituya la etapa inicial de un enorme número de fenómenos trascendentales para la vida celular (y por tanto orgánica), entre los cuales se puede destacar el transporte de proteínas (cadenas polipeptídicas) a través de la membrana, o la inyección de material genético desde la cápside vírica hacia el interior celular (durante una infección vírica); de manera muy destacada es también un mecanismo esencial para el proceso de la expresión genética, ya que permite el transporte de un gen transcrita en un polímero nucleico de ARN<sub>m</sub> desde el núcleo celular (que almacena la información genética) hacia los ribosomas (orgánulos encargados de traducir la información contenida en la copia de ARN<sub>m</sub> en forma de una cadena polipeptídica, la estructura primaria de una proteína), e incluso el propio proceso de traducción se identifica con un fenómeno de traslocación. Adicionalmente existen poros con una altísima especificidad que actúan no solo permitiendo el paso de grandes bioestructuras, si no actuando como verdaderos separadores químicos [1].

Esta lista de lugares comunes a la traslocación bastaría para justificar el estudio dinámico del proceso desde la perspectiva de la física básica, incluso sin reparar en las aplicaciones biomédicas que se desprenden en potencia del dominio de estos mecanismos, como podrían ser el diseño de fármacos antivirales (o por el contrario la optimización de vectores para terapia génica), la sensibilización de células como diana en tratamientos hormonales o anticancerígenos, etc. En los últimos años el diseño conceptual de secuenciadores de ADN basados en la traslocación forzada de polímeros de material genético a través de matrices de nanoporos se ha postulado como la aplicación potencialmente más relevante de esta tecnología [2].

Se denomina *secuenciación* del ADN a los procesos tecnológicos por los que se extrae la información contenida en el genoma (conjunto de genes de un individuo, codificados en tripletes de nucleótidos), abarcando una variedad de métodos desarrollados desde la década de 1970. Un punto de inflexión en este campo fue la culminación del proyecto *Genoma Humano* en el año 2003, que después de trece años y casi tres mil millones de dólares concluyó con la secuenciación del genoma completo de una muestra heterogénea de individuos. La técnica de secuenciación empleada (*Chain-Termination Sequencing Method*, que le valió a Frederick Sanger el premio Nobel de química en 1977) está basada en métodos químicos y requiere de la fragmentación del ADN para reintegrar la información posteriormente, disparando la complejidad del proceso [3]. La optimización de esta técnica permite fijar entorno al año 2010 en (aproximadamente) una semana y diez mil dólares el coste de secuenciación del genoma humano. En el año 2012 comienza a cristalizar comercialmente una línea de investigación iniciada por Kasianowicz *et al.* en 1996, quienes lograron detectar la presencia de un polímero de ARN en un canal de *α-hemolisina* por medios eléctricos [4]. En este nuevo método se dispone de una solución iónica separada por una membrana (biológica o de estado sólido) con un poro nanométrico, de forma que al establecer una diferencia de potencial entre ambos lados de

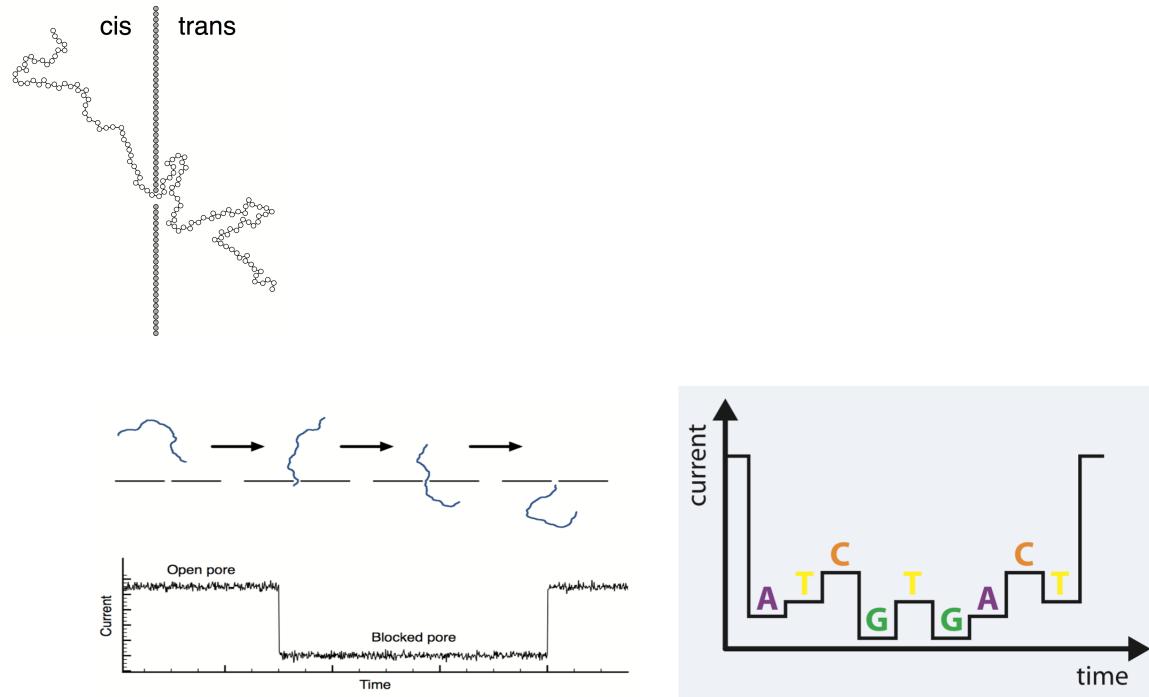


Figura 1: Arriba: Esquema de las regiones *cis-trans*. Abajo: señal eléctrica medida durante la traslocación de ADN (izda), esquema de la relación entre un perfil de corriente y una secuencia de nucleótidos (dcha.). [5]

la membrana se crea una corriente iónica medible a través del canal. Si el canal es bloqueado por algún nucleótido se modificará la corriente que atraviesa la membrana (*i.e.* se reducirá la conductividad de la membrana), y debido a las diferencias químicas y estructurales de cada base nitrogenada (biocompuesto característico de cada nucleótido) cabe esperar que exista una relación unívoca entre la pérdida de conductividad y el nucleótido que instantáneamente atraviesa el canal. Por lo tanto, mediante la traslocación (forzada externamente o a través del propio campo eléctrico) de una hebra de ADN a través del poro puede obtenerse un perfil de corriente directamente relacionado con la secuencia de bases de la hebra (fig. 1).

Este método de secuenciación, fácilmente escalable mediante la paralelización del proceso, hace factible pensar en la secuenciación del genoma de un ser humano en el orden de los quince minutos y cien dólares [3]. Este margen de costes elevaría la secuenciación genética al paradigma moderno de los análisis sanguíneos, favoreciendo la explosión de aplicaciones como las mencionadas anteriormente. Por otro lado, el método señalado es, de forma abstracta, la descripción de un analizador químico, y su aplicación a otra clase de moléculas permitiría una mejora en las prestaciones de técnicas contemporáneas, permitiendo por ejemplo el análisis bioquímico sin actuar sobre la estructura del objeto a medir, o (en combinación con un control adecuada de la composición química del poro) la segregación de compuestos en disolución<sup>1</sup> [6].

A pesar de la simplicidad del principio que sustenta esta tecnología y de la actividad investigadora floreciente en el campo de la traslocación de polímeros, lo cierto es que existen dificultades técnicas y aún una comprensión incompleta de los fundamentos físicos que rigen la traslocación forzada

<sup>1</sup>Resulta notable cómo la propia naturaleza es pionera en la explotación de esta técnica, *e.g.* canales transmembrana altamente específicos sensibles a iones tan parecidos químicamente como  $\text{Na}^+$  y  $\text{K}^+$ .

externamente de polímeros a través de nanoporos. Es en esta dirección que el presente trabajo aspira a colaborar, introduciendo un sistema inédito en la literatura especializada como es la traslocación de polímeros en presencia de un poro y una membrana flexible.

En investigaciones previas del grupo de Física Estadística y No Lineal de la Universidad de Zaragoza sobre la traslocación polimérica, se han encontrado acoplos resonantes entre la dinámica del polímero y la acción de un campo externo dependiente del tiempo aplicado sobre el poro [7], [8]; motivados por este descubrimiento se busca investigar en qué medida dicho efecto es reproducido por un poro flexible sometido a fluctuaciones térmicas (de tal modo que actúe efectivamente sobre el polímero traslocante como un campo externo dependiente del tiempo).

En este contexto, el objetivo principal de este trabajo será diseñar un protocolo para la simulación de un sistema novedoso: la traslocación forzada (mediante tracción externa) de un polímero a través de un poro flexible, insertado en una membrana también flexible. Un sistema semejante presentará un elevado número de grados de libertad, y consecuentemente un alto coste computacional asociado a resolver su dinámica, por lo que se acudirá a técnicas de cálculo no convencionales como la simulación acelerada mediante procesadores gráficos (*GPU, graphic processor unit*) en combinación con cálculo clásico en *CPU*; para ello se aprovecharán las capacidades implementadas en el entorno de simulación *Gromacs*, originalmente diseñado para efectuar simulaciones de dinámica molecular a todos los átomos y que deberá ser adaptado al uso dado en este trabajo (en particular, a la escala mesoscópica y a la geometría diseñada para las simulaciones de traslocación, tal y como se detallará en las siguientes secciones).

## 2. Física de la traslocación de polímeros

Las primeras aproximaciones teóricas a la traslocación de polímeros tuvieron lugar a finales de los años noventa, (Sung y Park en 1996, Muthukumar en 1999 [9]), modelando el proceso como un fenómeno de escape en un pozo potencial unidimensional, y utilizando como coordenada la longitud de la cadena en la región *trans* (coordenada  $s$ ). El resultado central de estos estudios es la determinación del tiempo de traslocación del polímero en función del número de monómeros, encontrándose una ley de escala  $t_{tras} \sim N^\alpha$ , con  $\alpha = 2$  si el término de fricción dominante sobre el polímero se debe a la viscosidad del medio acuoso en que ocurre el proceso, y  $\alpha = 1$  en caso de que domine la fricción debida a la interacción poro-polímero. Posteriormente se refinó el cálculo para incluir el efecto de estabilización del polímero, que tiende a contraerse en un espacio del tamaño de su *radio de giro* (ver apéndice A)  $R_g \sim N^{\nu_F}$ , donde  $\nu_F$  es el *exponente de Flory*, dando lugar a la ley de escala  $t_{tras} \sim N^{1+\nu_F}$  [10, 11].

En la práctica los experimentos de traslocación (y en consecuencia sus potenciales aplicaciones) deben realizarse forzando externamente el paso del polímero a través del canal (fuerza externa), de forma que este ha sido el caso más estudiado desde la perspectiva teórica. El formalismo más extendido para el estudio dinámico de la variable  $s$  es la *teoría de propagación de la tensión* [12], en el que la fuerza que experimenta cada monómero se propaga desde la partícula sobre la que se actúa externamente hacia el extremo *cis* del polímero en forma de un frente de fuerzas. En este sentido se pueden distinguir varios regímenes en función de la severidad de la tracción (fig. 2): (i) *strong stretching (SS)*, en el que la fuerza de arrastre es grande comparada con las fluctuaciones térmicas y el polímero se encuentra completamente estirado una vez es alcanzado por el frente de propagación. (ii) *stem-flower (SF)*, en el que fuerzas de arrastre más débiles resultan insuficientes para mantener estirada la cadena durante el proceso, y (iii) *trumpet-regime (TR)*, donde la fuerza es insuficiente para estirar siquiera una pequeña fracción de la cadena [13].

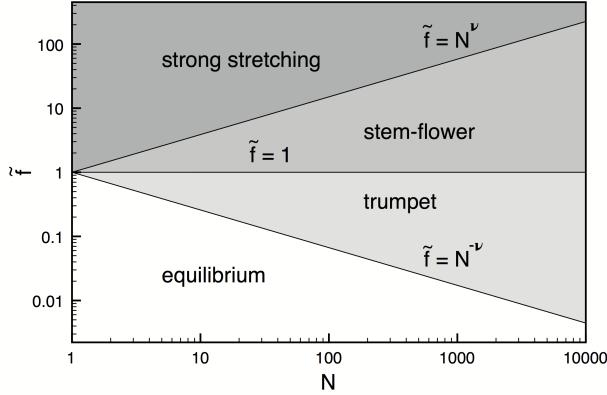


Figura 2: Mapa de los varios regímenes de traslocación en función de la fuerza aplicada y el número de monómeros. La magnitud  $\tilde{f} = f^{b_0/k_B T}$ , con  $f$  la fuerza de tracción y  $b_0$  la distancia de equilibrio entre monómeros [5].

En tanto que el régimen de fuerzas altas *strong stretching* produce una dinámica muy separada del equilibrio (y por tanto unos resultados menos influenciados por las fuerzas entrópicas del polímero), así como tiempos de traslocación menores (permitiendo mayor estadística), se adoptará este régimen en el desarrollo del presente trabajo.

### 3. Máquina de simulación química *Gromacs* [14]

El paquete de simulación *Gromacs* consta de una serie de herramientas informáticas diseñadas para llevar a cabo ciertas tareas en el ámbito de la química computacional y el modelado molecular, entre las que destacan la minimización energética de estructuras, la simulación de dinámica molecular y el análisis de propiedades mecanoestadísticas dinámicas y en equilibrio. El sistema de unidades del paquete está adaptado para producir valores numéricos cercanos a la unidad en la escala de la biología molecular:  $[L, T, \theta, Q, M] = [nm, ps, K, e, uma]$

A continuación se detallan los aspectos técnicos de *Gromacs* más relevantes que fueron explotados para la realización de este trabajo:

\*Nota sobre la nomenclatura: en esta sección y siguientes aparecerán en *cursiva* los nombres propios de *Gromacs* y en **palo seco** los parámetros seleccionables del módulo de simulación.

#### 3.1. Entorno de simulación

El entorno o caja de simulación será el entorno virtual definido por las distancias máximas a las que dos partículas simuladas pueden situarse en las tres dimensiones del espacio. La restricción de un espacio de simulación limitado es antinatural para la mayoría de experimentos a simular, de modo que *Gromacs* aplica condiciones periódicas de contorno (*PBC*) junto un convenio de mínima imagen (dada una partícula, esta solo verá una copia de cada una de las restantes partículas del sistema: aquella que se encuentre a menor distancia) para minimizar los efectos de borde; esto introduce, sin embargo, un artificio debido a la periodicidad del sistema, que en cualquier caso se prevé menos grave que una frontera con el vacío.

*Gromacs* permite la elección de diversas geometrías para la caja de simulación; en el caso de este trabajo se ha utilizado un entorno de simulación ortoédrico para ajustarse a la morfología

membranosa del sistema.

### 3.2. Interacciones en *Gromacs*

*Gromacs* clasifica las funciones potenciales de las que se derivan las fuerzas en tres clases: (i) *non-bonded*, (ii) *bonded* y (iii) *restraints* (ligaduras, no utilizadas para la realización de este trabajo).

#### 3.2.1. Interacciones *non-bonded*

Potenciales de interacción entre dos partículas  $i, j$  del sistema, de Coulomb:

$$V_C(r_{ij}) = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_i q_j}{r_{ij}} \quad (1)$$

y Lennard-Jones:

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2)$$

que dan cuenta de las interacciones electrostáticas y de Van der Waals entre los diferentes elementos del sistema. Las cargas eléctricas de cada partícula y los elementos de las matrices de parámetros de LJ ( $C_{ij}$ ) dependen de cada partícula o tipo de átomo (de hecho este conjunto de valores junto con la masa constituyen en gran medida la información que *Gromacs* engloba bajo la palabra «átomo»; el conjunto de átomos que participan en una simulación deben ser provistos con anterioridad en el fichero *campo de fuerzas*, que detallaremos en las próximas secciones). En aras de la eficiencia computacional, las interacciones de Van der Waals y Coulomb se truncarán artificialmente (*cutoff*) a un rango máximo de acción que se definirá en los parámetros *rvdw* y *rcoulomb* respectivamente. En adelante se empleará la notación  $r_{nb}$  para referirnos indistintamente al *cutoff* de las interacciones *non-bonded*.

#### 3.2.2. Interacciones *bonded*

Las interacciones *bonded* modelan el enlace químico entre átomos de una misma molécula. Las partículas ligadas mediante estos potenciales deberán aparecer listadas en ficheros que leerá *Gromacs* antes de una simulación, y permanecerán invariantes durante todo el desarrollo de la misma (por lo tanto no es posible observar reacciones químicas). *Gromacs* contiene programados diferentes potenciales de esta clase a dos (*bond*) tres (*angle*) o cuatro (*dihedrals*) partículas; los utilizados en la realización de este trabajo para el modelo del poro han sido:

- Potencial armónico (dos cuerpos):

$$V_b(r_{ij}) = \frac{1}{2} k_b^{ij} (r_{ij} - b_{ij})^2 \quad (3)$$

- Potencial armónico angular (tres cuerpos):

$$V_a(\theta^{ijk}) = \frac{1}{2} k_\theta^{ijk} (\theta^{ijk} - \theta_0^{ijk})^2 \quad (4)$$

Donde  $\theta^{ijk} \equiv \arccos \left( \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{kj}}{r_{ij} r_{kj}} \right)$

A pesar de distinguirse en las formas de los potenciales y de modelar fuerzas de naturaleza diferente, la principal distinción entre interacciones *bonded* y *non-bonded* radica en la manera en que *Gromacs* etiqueta las partículas involucradas en cada interacción concreta; las partículas ligadas mediante interacciones *bonded* deberán aparecer especificadas en parejas o tríos en uno de los ficheros que se incluirán como condición inicial de la simulación (llamado fichero de *topología*), de tal suerte que estas configuran una lista fija, en el sentido de que no es modificada durante el desarrollo de la simulación.

Por otro lado, para evitar una gasto de recursos innecesario, las partículas que participan en una interacción *non-bonded* respecto de una [partícula] dada se encuentran limitadas a una distancia máxima, a partir de la cual la ganancia en exactitud se puede considerar despreciable en comparación con la pérdida en velocidad de computación. Los pares de partículas susceptibles de participar en una interacción de esta clase se encuentran especificados en listas de vecinos dinámicas, que son actualizadas cada `nstlist` iteraciones por el módulo de simulación (típicamente  $nstlist \approx 10$ ). Esta lista dinámica contiene en cada instante y para cada partícula del sistema: el índice entero que la identifica en el sistema ( $i$ ), un radiovector que especifica su posición ( $\mathbf{r}_i$ ) y los índices de todas las partículas que se encuentran dentro de la bola de centro  $\mathbf{r}_i$  y radio (*cutoff*) `rlist`,  $\{j \text{ tales que } |\mathbf{r}_j - \mathbf{r}_i| < \text{rlist}\}$ .

*Gromacs* implementa dos esquemas para gestionar las listas dinámicas y determinar las partículas que efectivamente participarán en una interacción *non-bonded*: *group* y *verlet*.

### 3.2.3. Esquema *group*

En la aplicación de este esquema, la lista de vecinos estará formada por pares de *grupos de carga* (de al menos una partícula), que deberán ser predefinidos en el fichero de topología. El criterio para incluir a un par de grupos en la lista de vecinos será que sus respectivos centros de geometría se encuentre dentro de la distancia de *cutoff* de la interacción ( $r_{nb}$ ). Una vez un par de grupos ha sido seleccionado como interactuantes (en un instante concreto de la simulación), la lista de vecinos se rellena con todas las parejas que se pueden formar entre partículas de uno y otro grupo.

### 3.2.4. Esquema *verlet*

En este caso el listado de vecinos se realiza atendiendo a grupos de cuatro u ocho partículas, que al contrario que en el esquema *group* se constituyen dinámicamente entre las partículas más cercanas. Durante la búsqueda de vecinos bajo el esquema *verlet* se define una distancia *buffer* ( $r_b$ ) mayor que el rango de acción de las interacciones *non-bonded*, que actuará efectivamente como *cutoff*; de esta manera se incluyen partículas dentro de una bola mayor que la generada por el propio alcance de la interacción (radio  $r_{nb}$ ), *i.e.* se incluyen las partículas pertenecientes a una corteza esférica  $r_{nb} < r < r_b$  que contribuirán nulamente a la fuerza total sobre la partícula de referencia. Dado que la actualización de las listas de vecinos ocurre en intervalos constantes de `nstlist` iteraciones, este método garantiza que entre dos pasos de actualización prácticamente todas las partículas que se encuentren dentro del rango de acción de la interacción sean calculadas, ya que se evita falsear el cálculo con partículas a distancias  $\sim r_{nb}$  (que podrían entrar y salir del *cutoff non-bonded* en períodos inferiores a  $nstlist \cdot \Delta t$ , donde  $\Delta t$  es el intervalo temporal de la simulación). Existirá no obstante una pequeña probabilidad de que, entre dos pasos de actualización, alguna partícula escape de (o entre en) la bola de radio  $r_b$ , de forma que la interacción de este par de partículas sea calculada/obviada erróneamente (ver fig. 3). Esta probabilidad se puede traducir en un error por exceso/defecto de la energía potencial de una partícula, que puede a su vez calcularse suponiendo un sistema en equilibrio a temperatura finita. El módulo de simulación permite fijar la tolerancia máxima a este error en la energía (parámetro *verlet-buffer-tolerance*); este valor es utilizado para definir la cantidad  $r_l$  y permanece constante durante la simulación (en este caso el parámetro `rlist` es ignorado).

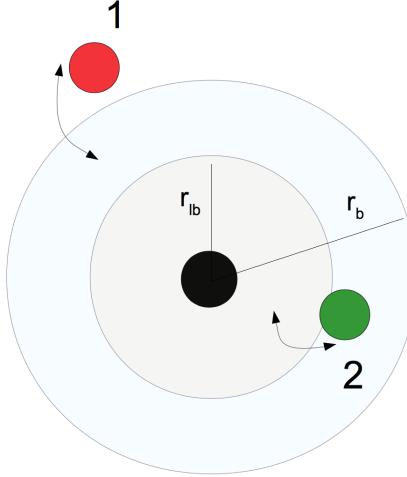


Figura 3: Rango de acción de interacciones de largo alcance y *buffer* en el esquema de Verlet. La partícula (1) puede dar lugar a errores si penetra hasta la bola  $r_{lb}$  antes de que la lista se actualice, pero esta técnica permite evitar el error debido a la cinemática de (2).

### 3.3. Algoritmos de simulación

Los módulos de minimización energética y simulación molecular empleados en este trabajo implementan los siguientes algoritmos:

#### 3.3.1. Minimización energética: *steepest-descendent*

En este algoritmo se realizan cambios tentativos en las posiciones de cada partícula, de tal modo que el cambio se acepta si reduce la energía potencial total, y se rechaza si esta se incrementa.

Dados un sistema de  $N$  partículas, un incremento inicial  $h_0$  y los vectores de 3-N componentes  $\mathbf{r}$  (posición) y  $\mathbf{F} = -\nabla V$  (fuerzas), la rutina de minimización del sistema entre dos iteraciones  $n$  y  $n + 1$  será:

$$\begin{aligned}
 1^0) \quad & \mathbf{r}'_{n+1} = \mathbf{r}_n + \frac{\mathbf{F}_n}{\max(|\mathbf{F}_n|)} h_n \\
 2^0) \quad & \text{Cálculo de } V_{n+1} \left\{ \begin{array}{l} \text{si } V_{n+1} < V_n \longrightarrow \left\{ \begin{array}{l} \mathbf{r}_{n+1} = \mathbf{r}'_{n+1} \\ h_{n+1} = 1,2h_n \end{array} \right. \\ \text{si } V_{n+1} > V_n \longrightarrow \left\{ \begin{array}{l} \mathbf{r}_{n+1} = \mathbf{r}_n \\ h_{n+1} = 0,2h_n \end{array} \right. \end{array} \right.
 \end{aligned}$$

Donde  $\max(|\mathbf{F}_n|)$  representa el mayor valor absoluto de las componentes de la fuerza. El algoritmo se detiene una vez esta cantidad ha alcanzado un valor de *tolerancia* ( $\epsilon$ ).

#### 3.3.2. Integración de las ecuaciones del movimiento: *leap-frog* estocástico

En un sistema de  $N$  partículas, la resolución de las ecuaciones del movimiento para cada una de las  $3N$  coordenadas conducirá a un conjuntos estadístico  $NVE$ , donde se conservarán el número de partículas, el volumen y la energía. Para extraer información realista de un sistema simulado deberá

utilizarse un conjunto estadístico que conserve la temperatura (*NVT*), *i.e* el conjunto canónico. *Gromacs* proporciona varias vías para lograr esto en forma de distintas modificaciones a las ecuaciones del movimiento (*termostatos*), de manera que la distribución final de velocidades satisfaga las condiciones del conjunto.

En el desarrollo de este trabajo se ha utilizado el módulo [de *Gromacs*] de *dinámica estocástica* (*SD* para acoplar el sistema a un baño térmico, el cual introduce unos términos de disipación viscosa y ruido térmico en las ecuaciones de Newton de la dinámica molecular, dando lugar a un sistema de ecuaciones de Langevin que determinarán la evolución del sistema:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = -m_i \tau_i^{-1} \frac{d\mathbf{r}_i}{dt} + \mathbf{F}_i(\mathbf{r}) + \dot{\mathbf{r}}_i, \quad \langle \dot{\mathbf{r}}_i(t) \dot{\mathbf{r}}_j(t') \rangle = 2m_i \gamma_i k_B T \delta(t - t') \delta_{i,j}$$

Donde  $\tau$  es la constante característica de relajación térmica,  $\mathbf{F}$  la fuerza neta ejercida sobre cada partícula por el sistema y  $\dot{\mathbf{r}}$  un ruido blanco que modela el ruido térmico.

El programa de simulación integra estas ecuaciones mediante un algoritmo *leap-frog* modificado para incluir los efectos de la fluctuación/disipación, que se implementa en forma de una fuerza impulsiva actuando sobre cada partícula al finalizar un paso temporal  $(t + \Delta t)$ .

<i>leap-frog</i> determinista	<i>leap-frog</i> estocástico
$\mathbf{v}' \equiv \mathbf{v}(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m} \mathbf{F}(t)$	$\mathbf{v}' \equiv \mathbf{v}(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m} \mathbf{F}(t)$
$\Delta \mathbf{v} = 0$	$\Delta \mathbf{v} = \mathbf{v}'(e^{-\gamma \Delta t} - 1) + \sqrt{\frac{k_B T}{m}(1 - \alpha^2)} \mathbf{r}^G$
$\mathbf{v}(t + \frac{1}{2}\Delta t) = \mathbf{v}'$	$\mathbf{v}(t + \frac{1}{2}\Delta t) = \mathbf{v}' + \Delta \mathbf{v}$
$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}' \Delta t$	$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \left( \mathbf{v}' + \frac{1}{2} \Delta \mathbf{v} \right) \Delta t$

donde  $\mathbf{r}^G$  es una variable aleatoria gausiana de media  $\mu = 0$  y varianza  $\sigma = 1$ , y  $\alpha \equiv 1 - e^{-\gamma \Delta t}$

### 3.4. Condiciones iniciales de la simulación

El módulo de simulación *mdrun* requiere una serie de ficheros con información relativa a los parámetros de la simulación, posiciones (y opcionalmente velocidades) iniciales de todas las partículas del *sistema*, listas fijas y una descripción de cada partícula que el módulo podrá encontrar: nombre que distingue a cada *átomo*<sup>2</sup>, masa, carga eléctrica, forma funcional de los potenciales *non-bonded* y parámetros  $C^{(6,12)}$  del potencial de Lennard-Jones. Una agrupación de *átomos* entre las que existen interacciones *bonded* define una *molécula* del sistema (esto es, una partícula  $i$  pertenecerá a una molécula  $I$  si, partiendo de  $i$ , pueden recorrerse todas las partículas de  $I$  «saltando» entre enlaces *bonded*). El listado de los nombres y parámetros atómicos junto con [las listas fijas de interacciones *bonded* de] una selección de moléculas frecuentes en la biología molecular constituyen un *campo de fuerzas*. La información del *campo de fuerzas* y el número de *moléculas* de cada tipo presentes en una simulación concreta (que deberá coincidir con el contenido del fichero de posiciones iniciales) constituyen el contenido del fichero de *topología*.

<sup>2</sup>el término *átomo* es la unidad mínima particular de *Gromacs*, pero esta no debe coincidir necesariamente con un verdadero *átomo* natural. En simulaciones atomísticas es frecuente que el término incluya *iones* o *moléculas diatómicas*.

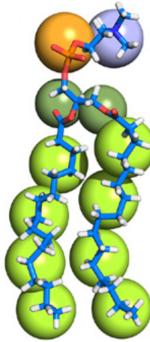


Figura 4: Representación *coarse grained* de un fosfolípido superpuesta a su representación atomista [15]

#### 4. Simulaciones *coarse graining*: campo de fuerzas *MARTINI*

Un modelo de química computacional se dice *coarse grained* (*CG*) cuando los grados de libertad que representan un sistema se construyen mediante la integración metódica y sistemática de varios ( $\sim 3$  ó  $4$ ) grados de libertad de un modelo atomista (en contraposición a la escala atómica se dirá que los modelos *CG* operan en la mesoscala); para ello habitualmente se substituyen pequeños grupos de *átomos* por partículas individuales, y las interacciones elementales (Coulomb y Lennard-Jones) son reemplazadas por interacciones efectivas entre estos grupos. En tanto que se trata de modelos aproximados, estos se diseñan para reproducir alguna cualidad específica de un modelo de referencia a todos los átomos (o de información experimental), *e.g* buen ajuste en la distribución de energía libre, de propiedades espaciales (ángulos promedio, elongaciones, etc) o fuerzas.

Debido al incremento de las escalas espacio-temporales en los modelos *CG* (y dado las limitaciones en el tiempo de computación y la acumulación de errores numéricos), estos [modelos] permiten explorar tamaños mayores y en particular intervalos de tiempo sensiblemente superiores a aquellos accesibles mediante la simulación a todos los átomos; es en esta tarea donde los modelos en la mesoscala han encontrado una mayor aplicación, particularmente aplicados a sistemas lipídicos [16].

Para la realización de este trabajo se ha empleado una adaptación (que se detallará en siguientes secciones) del campo de fuerzas en la mesoscala *MARTINI*; este modelo fue desarrollado por Siewert Marrink *et al.* en la Universidad de Groningen en el año 2004 para la simulación de sistemas de lípidos, y fue posteriormente reparametrizado y ampliado para permitir la reproducción de una mayor gama de biomoléculas [17].

Los parámetros de este campo han sido ajustados contrastando los valores de solubilidad (coeficiente de partición) en una mezcla en equilibrio de agua y disolvente orgánico (*i.e* aceite), para una gran variedad de biomoléculas *coarse grained* frente a datos experimentales [16, 17]. En comparación con otros campos *CG* orientados a buenos ajustes de propiedades estructurales bajo condiciones muy específicas, la reproducción de propiedades termodinámicas garantiza un amplio rango de aplicaciones: ha sido empleado con éxito en el estudio de la formación y fusión de vesículas, dinámica de proteínas de membrana, monocapas y bicapas lipídicas; no obstante debe tenerse en cuenta que se han reportado comportamientos patológicos en las interfaces de sistemas líquido-vapor y líquido-sólido [16].

Un aspecto relevante en los modelos mesoscópicos es la manera en que debe interpretarse la escala temporal. Debido al proceso *CG* la fricción de los grados de libertad atómicos es desestimada, lo que conduce a un paisaje de energía libre suavizado, y por tanto a una dinámica acelerada respecto de

modelos atomistas. Mediante comparaciones de la constante de difusión, se han encontrado tiempos efectivos de simulación entre dos y diez veces superiores en sistemas parametrizados con el campo *MARTINI* que en simulaciones a todos los átomos [16]. De forma estándar se toma:

$$\text{tiempo físico simulado} = 4 \times \text{tiempo de simulación}$$

#### 4.1. Descripción del modelo

El campo de fuerzas *MARTINI* se basa en un mapeado «4 : 1» (cuatro-a-uno), *i.e* en promedio cuatro *átomos* son representados por una única partícula (centro de interacción); esta regla general sin embargo no es de estricto cumplimiento, así las estructuras en forma de anillo se mapean con tantas partículas *CG* como sean necesarias para preservar la morfología (lo que resulta habitualmente en 2 : 1 o 3 : 1).

En este campo se definen principalmente cuatro tipos diferentes de centros de interacción, denominados: *polar* (*P*), *no-polar* (*N*), *apolar* (*C*) y *cargado* (*Q*); el campo asigna a estas partículas la misma masa estándar  $m = 72 \text{ umu}$  (correspondiente a cuatro moléculas de agua) A su vez, cada una de estas clases se divide en una serie de subtipos:

- para los tipos *Q* y *N*: *d* (dador), *a* (aceptor), *da* (ambos), *0* (ninguno) en referencia a su capacidad para ceder un electrón a un puente de hidrógeno.
- para los tipos *P* y *C*: 1, 2, 3, 4, 5, para indicar su grado de polarizabilidad (1 mínima, 5 máxima).

También existe un modificador de la partícula introducido específicamente para formar parte de anillos, y etiquetado mediante *S* («small») que puede ser aplicado a cualquier clase de centro de interacción (*SP*, *SN*, *SC*, *SQ*), de tal forma que varíen los parámetros de interacción *non-bonded* entre partículas modificadas (en particular la masa se ve reducida  $m = 45 \text{ umu}$ ).

Por último cabe mencionar la partícula *P*<sub>4</sub> por ser la parametrización del agua en el campo de fuerzas *MARTINI* (equivalente a cuatro molécula de agua, sin embargo no debe confundirse con el subíndice del grado de polarizabilidad). Estas partículas tienen un punto de congelación dirigido mediante nucleación entorno a los 300 K, claramente superior al agua en condiciones termodinámicas normales; si bien el proceso de nucleación es de naturaleza estocástica y las simulaciones muestran vidas medias tan largas como  $\sim$  microsegundos (lo que en la práctica es a menudo mayor que la propia duración de la simulación), lo cierto es que estructuras superficiales sólidas o incluso lipídicas pueden actuar como centros de nucleación *per se*, convirtiendo en un problema la congelación del agua incluso a temperaturas moderadas. Para evitar esto se introduce *ad hoc* una partícula especial *BP*<sub>4</sub> (Big*P*<sub>4</sub>), que interactúa con el agua intensamente a través de un radio de Van der Waals ampliado. La perturbación introducida evita la cristalización del agua en concentraciones  $\sim 10\%$  de *BP*<sub>4</sub> respecto de *P*<sub>4</sub>.

#### 4.2. Interacciones *bonded* y *non-bonded*

Los términos *bonded* y *non-bonded* en *MARTINI* remiten exactamente a las definiciones dadas en *Gromacs*; las interacciones entre partículas ligadas mediante enlace químico son aquellas descritas en las expresiones (3) y (4) de la sección previa, con unos valores estándar de las constantes  $k_b = 1250 \text{ kJ/nm}^2 \text{ mol}^{-1}$ ,  $b_0 = 0,47 \text{ nm}$ ,  $k_\theta = 25 \text{ kJ mol}^{-1}$  y  $\theta_0 = 180^\circ$  (sin perjuicio de que puedan ser modificadas para biomoléculas concretas en el ficheros de *topología*).

Las interacciones de largo alcance toman así mismo las formas funcionales (2) y (1), si bien en este caso *MARTINI* define diez niveles de interacción (*i.e* diez diferentes parejas de constantes  $\epsilon$ ,

Nomenclatura	<i>O</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>V</i>	<i>VI</i>	<i>VII</i>	<i>VIII</i>	<i>IX</i>
$\epsilon \text{ kJ/mol}$	5,6	5,0	4,5	4,0	3,5	3,1	2,7	2,3	2,0	2,0
$\sigma \text{ nm}$	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,62

Cuadro 1: Nomenclatura y parámetros de las distintas interacciones *non-bonded* previstas en *MARTINI*.

		Q				P				N				C						
	sub	da	d	a	0	5	4	3	2	1	da	d	a	0	5	4	3	2	1	
Q	da	O	O	O	II	O	O	O	I	I	I	I	I	IV	V	VI	VII	IX	IX	
	d	O	I	O	II	O	O	O	I	I	I	III	I	IV	V	VI	VII	IX	IX	
	a	O	O	I	II	O	O	O	I	I	I	III	IV	V	VI	VII	IX	IX	IX	
	0	II	II	II	IV	I	O	I	II	III	III	III	IV	V	VI	VII	IX	IX	IX	
	5	O	O	O	I	O	O	O	O	O	I	I	I	IV	V	VI	VI	VII	VIII	
P	4	O	O	O	O	O	I	I	II	II	III	III	III	IV	V	VI	VI	VII	VIII	
	3	O	O	O	I	O	I	I	II	II	II	II	II	IV	V	V	VI	VII	VII	
	2	I	I	I	II	O	II	II	II	II	II	II	II	III	IV	IV	V	VI	VII	
	1	I	I	I	III	O	II	II	II	II	II	II	II	IV	IV	IV	IV	V	V	
	0	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	V	V	
N	5	V	V	V	V	V	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	V	V	
	4	VI	VI	VI	VI	VI	VI	V	IV	IV	V	V	V	IV	IV	IV	IV	V	V	
	3	VII	VII	VII	VII	VI	VI	V	IV	VI	VI	VI	VI	IV	IV	IV	IV	IV	IV	
	2	IX	IX	IX	IX	VII	VII	VI	VI	V	VI	VI	VI	V	V	V	IV	IV	IV	
	1	IX	IX	IX	IX	VIII	VIII	VII	VII	VI	VI	VI	VI	V	V	IV	IV	IV		

Figura 5: Relación entre las posibles parejas de centros de interacción contempladas en *MARTINI* y el nivel de sus interacciones[17].

$\sigma$ ), etiquetadas como: *O* (sólidos a temperatura ambiente), *I* (interacción polar fuerte, *e.g* agua), *II* y *III* (líquidos volátiles como etanol o acetona), *IV* (interacciones no polares en cadenas alifáticas), *V* a *VIII* (repulsión hidrofóbica entre fases polar/apolar), *IX* (repulsión entre partículas cargadas y un medio muy apolar).

Junto con la propia definición de los sitios de interacción, el aspecto principal del campo de fuerzas es la asignación de uno de estos niveles a cada posible pareja de entre las partículas del campo (fig. 5).

#### 4.3. Parámetros constantes de la simulación con *MARTINI*

El estándar del campo de fuerzas *MARTINI* incluye la predefinición de algunos parámetros de la simulación que, en principio, *Gromacs* permite modificar. Estos valores deben entenderse como parte del campo, y por tanto no podrán ser alterados sin comprometer la fidelidad del modelo que subyace a *MARTINI*.

Algunos de estos parámetros incluyen el intervalo temporal  $dt \leq 3 \text{ ps}$ , constante dieléctrica relativa del medio  $\text{epsilon\_r} = 15$  (utilizando un modelo de agua  $P_4$ ; existe una parametrización polarizada de agua que requiere reducir  $\text{epsilon\_r} = 2,5$ ), esquema para la elaboración de listas dinámicas  $\text{cutoff-scheme} = \text{verlet}$ , distancia de *cutoff* para las interacciones de largo alcance  $\text{rcoulomb/rvdw} = 1,1 \text{ nm}$  y tolerancia al error en la energía de las interacciones *non-bonded*  $\text{verlet\_buffer\_tolerance} = 0,005 \text{ kJ/ps mol}^{-1}$ .

## 5. Simulación de experimentos de traslocación de biopolímeros a través de nanoporos en la mesoescala

En este trabajo se pretende introducir un entorno poco explorado para la simulación numérica de experimentos de traslocación en la mesoescala, a saber la simulación de un sistema tridimensional con una membrana y un poro activos (en el sentido de interactuantes con el traslocando), particularmente orientado al análisis del acople entre las propiedades mecánicas del poro (condensadas estas en su espectro de modos normales) y el fenómeno de la traslocación. Debido al uso intensivo de recursos computacionales que supone la integración de la dinámica en sistemas con elevado número de grados de libertad (en el caso de simulación explícita de solvente puede ascender a  $\sim 10^5$  partículas), se ha optado por el uso del paquete de simulación *Gromacs*, que implementa un potente motor de paralelización en *CPU* y *GPU* (ver sección 1). Como campo de fuerzas se ha utilizado *MARTINI* por su orientación a la mesoescala (tal y como se detalló en la sección previa), así como por estar especialmente optimizado para la simulación de lípidos, que constituyen la biomolécula más numerosa en este sistema de traslocación; pese a esto, debido a (i) el interés que despierta el fenómenos de la traslocación polimérica en sistemas nanotecnológicos y no estrictamente biológicos [18] y (ii) la relación no unívoca entre biomoléculas y partículas del campo (necesario debido al carácter *coarse graining* de este), se han introducido importantes modificaciones sobre los parámetros del campo en lo que se refiere al diseño del poro y las constantes características del polímero.

Para comunicarse con el *cluster* donde se han ejecutado los programas, automatizar el lanzamiento de simulaciones en *Gromacs* (incluyendo la gestión de los ficheros de estructuras y topología) y efectuar el análisis de los datos extraídos de aquellas se ha utilizado el programa *script* descrito en el apéndice B.

### 5.1. Diseño del sistema

El sistema que se presenta consta de tres estructuras diferenciadas (clasificación heredada en el fichero de topología, donde cada estructura es independiente de las demás): **membrana lipídica, polímero y canal (o poro)**. Tanto el polímero como el canal constituyen una única *molécula*, mientras que la membrana consiste en una bicapa de fosfolípidos de la familia de las fosfatodilcolinas (molécula parametrizada en *MARTINI* y que presenta buenas propiedades de autoensamblaje). Las dimensiones espaciales estarán caracterizadas respectivamente por la superficie de la membrana ( $nm^2$ ), la longitud del polímero ( $nm$ ) y el radio del canal ( $nm$ ). Antes de iniciar las simulaciones se construye un fichero de posiciones iniciales en el que cada estructura ocupa un lugar adecuado para una termalización efectiva, esto es, el eje del canal se alinearán con el eje del polímero a lo largo de la dirección  $z$ , mientras la membrana se extiende en las direcciones  $xy$ ; la membrana definirá dos regiones del espacio de simulación (*cis* y *trans*). Ambas regiones estarán interconectadas a través del poro, que se situará en perpendicular a la membrana y ocupando su centro geométrico. En cuanto a la posición del polímero nos adherimos a los argumentos dados en [3] (el proceso por el que el polímero accede a la entrada del poro es diferente de su traslocación, por lo que ambos fenómenos pueden estudiarse independientemente) para situarlo parcialmente dentro del canal (ocupando la mitad de la longitud del canal, *fig. 6*)

#### 5.1.1. Bicapa lipídica

La doble capa lipídica constituye la base de la estructura membranosa que rodea a una célula; en disolución acuosa esta estructura aparece de forma espontánea mediante el ensamblaje

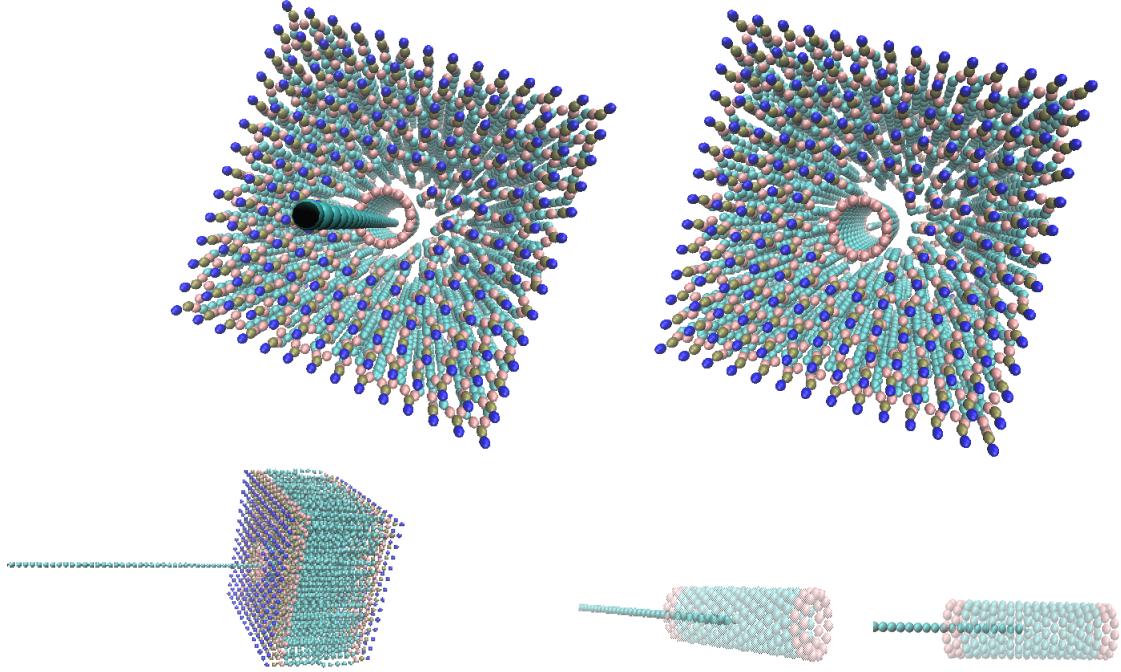


Figura 6: Configuraciones iniciales (previo a termalizar) propuestas en este trabajo para simulaciones de traslocación. Arriba izda: sistema desde *cis*. Arriba dcha: sistema desde *trans*. Abajo izda: visión global del sistema. Abajo dcha: poro y polímero.

auto-organizado de biomoléculas anfipáticas, *i.e* que presentan simultáneamente carácter hidrofílico/fóbico. Las moléculas que mayoritariamente presentan este comportamiento en sistemas biológicos son los *fosfolípidos*, formados por una cabeza hidrofílica unidas mediante un grupo fosfato a dos cadenas alifáticas hidrofóbicas (fig. 7).

Dado que *MARTINI* se desarrolló inicialmente para simulaciones de lípidos, existe un amplio repertorio de fosfolípidos de la familia de las fosfatodilcolinas parametrizados y ampliamente testadas. Para la construcción de la membrana se opta por el *fosfatidilcolina distearol* (*DSPC*, fig. 8) debido a la tendencia que presenta a organizarse en forma de bicapas (enfrentando entre sí las colas alifáticas para reducir el contacto con el solvente, y exponiendo de este modo las cabezas hidrofílicas), así como por la estabilidad de la estructura.

A pesar de esta capacidad para conformarse en una sección de membrana a partir de una suspensión, con el fin de reducir el tiempo necesario para termalizar se el programa *script* construye un sistema inicial preconfigurado en forma de bicapa a partir de una medida de la superficie de la membrana ( $S_m$ ) (introducida inicialmente como parámetro), separando cada molécula una distancia igual al radio de Van der Waals  $\sigma = 0,47\text{ nm}$ . Seguidamente se crea el hueco donde se acomodará el poro, creando una perforación en la membrana con origen en su centro y radio igual al radio del canal,  $r_c$  (visto desde la alzada de la bicapa). Con todo esto, la membrana contendrá  $2\frac{S_m - \pi r_c^2}{\sigma^2}$  fosfolípidos.

### 5.1.2. Poro

El poro diseñado para estas simulaciones ha sido parametrizado *ad hoc* imitando los aspectos esenciales de las proteínas transmembrana que se encargan de regular el transporte de grandes moléculas entre los espacios *cis* y *trans* de la membrana celular (en particular la traslocación poli-

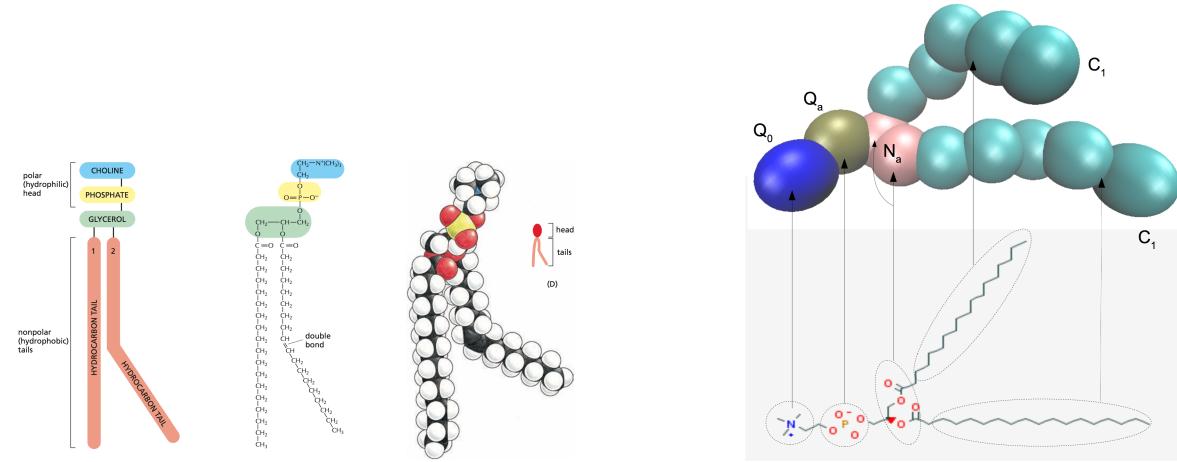


Figura 7: Representación de la molécula de fosfatidilcolina (*izda*). Esquema de la parametrización *coarse graining* del fosfolípido (*DSPC*) empleado para construir la membrana (*dcha*)

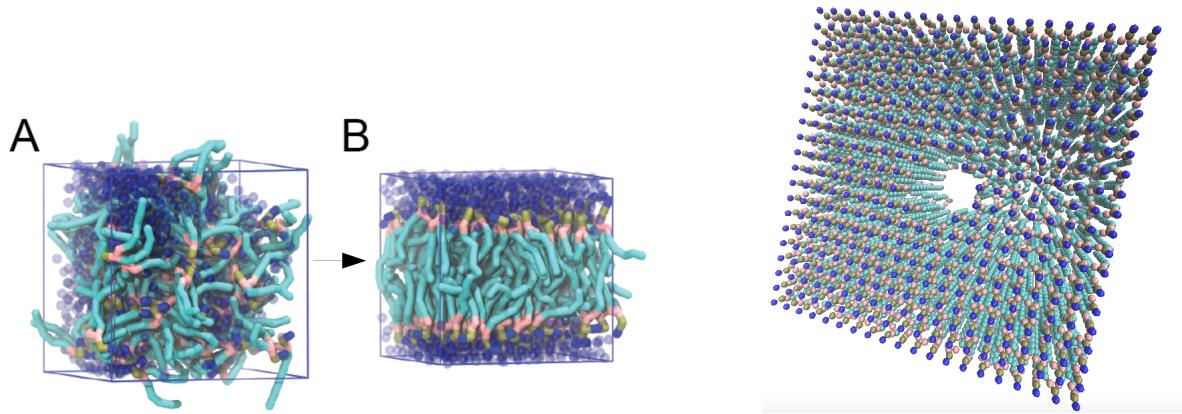


Figura 8: Autoensamblaje de la membrana a partir de un coloide de fosfolípidos (*izda*). Sección de membrana celular tal y como es producida por el programa *script* (*dcha*).

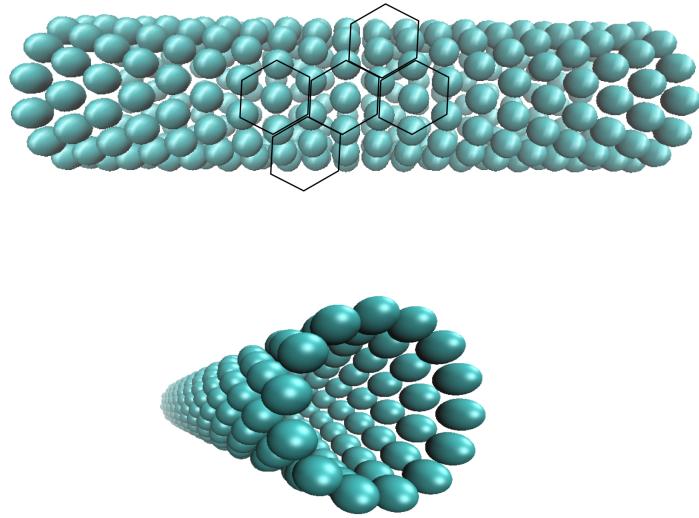


Figura 9: Morfología del canal. En la vista de perfil se evidencia la estructura formada por el colapso de un plano cristalino hexagonal.

mérica). Está formado por una serie de partículas dispuestas en geometría cilíndrica formando sobre su superficie una red hexagonal (originalmente se experimentó con redes cuadradas -más sencillas de implementar-, pero esta disposición conducía a canales absurdamente inestables).

Los enlaces químicos del canal se simulan con interacciones *bonded* entre vecinos de cada celda unidad: cada vértice de la red se enlaza con su centro mediante potenciales de tipo *bonding* (armónico en la distancia, ec. 3; representados como flechas en la fig. 10), y cada par de átomos opuestos se ligan mutuamente vía el átomo central con potenciales *bending* (armónicos en el ángulo relativo, ec. 4; representados en el panel izquierdo de susodicha figura con pares de flechas de distintos colores).

El canal se alojará en el hueco abierto en la membrana (fig 6) de tal manera que ocupe todo el ancho de aquella (en el caso de una membrana formada por *DSPC* se fija un poro de longitud  $l_c = 5,4 \text{ nm}$  de longitud). En cuanto a la química del canal, se pueden distinguir dos regiones: una *central* en contacto con las cadenas alifáticas de los fosfolípidos, que estará formada por partículas  $C_1$  (dando lugar a un nivel *IV* de interacción atractivo, ver fig. 5), y dos regiones en los *bordes* en contacto con las cabezas polares de la membrana, caracterizados en este caso por centros de interacción con carácter polar  $P_3$  [1, 19]. Esta configuración en dos regiones imita (de forma simplificada) la organización de proteínas transmembrana en sistemas biológicos, en los que un canal formado por una o varias *hélices- $\alpha$*  se encuentra aislado del cuerpo de la membrana por una estructura envolvente de aminoácidos hidrófobos.

Para abarcar un amplio margen de propiedades mecánicas las constantes características de las interacciones *bond* del poro se variarán entre grandes cotas durante la explotación del modelo; sin embargo cabe mencionar una suerte de «valores estándar» (basados en la parametrización de *MARTINI* para aminoácidos frecuentes como la alanina) de las constantes (ver sección 3.2.2)  $k_b \sim 7500 \text{ kJ/nm}^2 \text{ mol}^{-1}$ ,  $b_0 \sim 0,33 \text{ nm}$ ,  $k_\theta \sim 50 \text{ kJ mol}^{-1}$ . Esta será la notación usada en adelante para las constantes estructurales del poro.

### 5.1.3. Polímero

El modelo de polímero utilizado en el desarrollo de este trabajo está formado por una cadena sencilla unidimensional, parcialmente basada en la parametrización de polietilenglicol (*PEG*) dis-

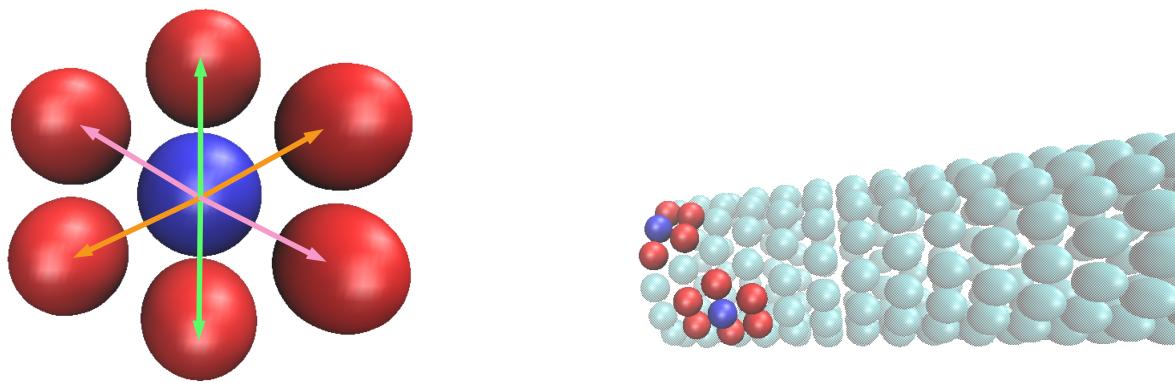


Figura 10: *Izda*: esquema de uniones *bonded* de cada partícula del canal. *Dcha*: segmento del poro donde se han destacados las partículas *bonded*-interactuantes (rojo) con dos elementos del poro (azul).

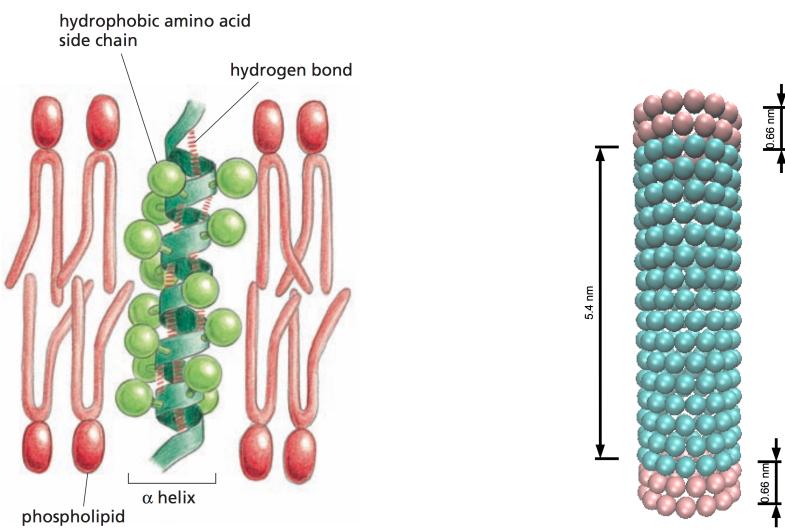


Figura 11: *Izda*: ilustración de un canal simple (una  $\alpha$ -hélice) embebido en una sección de membrana celular. *Dcha*: simulación de proteína transmembranosa mediante un canal con una región central apolar (azul) y extremos polares (rojo) [1]

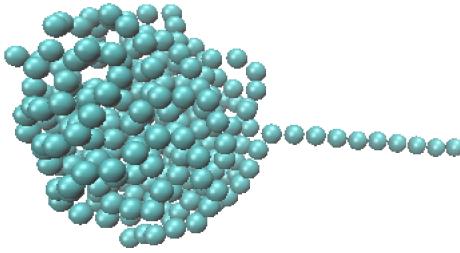


Figura 12: Polímero equilibrado preparado para la simulación. El extremo estirado comienza insertado en el poro.

ponible para *MARTINI*. Está constituido en su mayor parte por centros de interacción  $N_0$  (frente a  $SN_0$  de la parametrización original) que interactúan de forma débilmente repulsiva con la zona intermedia del poro, y atractiva con los bordes; las partículas de los extremos del polímero están formadas por *átomos*  $P_2$  (heredado del *PEG*, donde representan grupos *OH* terminales).

La topología es simple, encontrándose unidas mediante interacciones *bond* las partículas adyacentes, y *bending* cada grupo de tres *átomos* contiguos; así mismo, se conserva la constante *bond* del modelo original ( $k_{b,pol} = 1,7 \times 10^4 \text{ kJ/nm}^2\text{mol}^{-1}$ ), pero se reduce  $k_{\theta,pol}$  y se incrementa el ángulo de equilibrio  $\theta_{0,pol}$  para facilitar el plegamiento del polímero (fig. 12) durante la fase de termalización ( $k_{\theta,pol} = 10 \text{ kJmol}^{-1}$ ,  $\theta_{0,pol} = 180^\circ$ ). En relación al polímero se empleará la notación  $l_{\text{pol}}$  para especificar su longitud,  $b_{0,\text{pol}}$  para la constante de equilibrio *bond* y  $N_{\text{pol}}$  para el número de partículas que lo forman (relacionadas mediante  $l_{\text{pol}} = N_{\text{pol}} \cdot b_{0,\text{pol}}$ )

## 5.2. Análisis estructural del nanoporo

Dado que en este trabajo se busca investigar la contribución de un poro con estructura a la fenomenología de la traslocación, se ha prestado especial atención al análisis de las propiedades mecánicas derivadas de las interacciones *bonded* del canal, así como a su dependencia con estas. Para ello (y debido a la dificultad de manipular las constantes elásticas en un sistema discreto), se utilizará como observable propio de «las propiedades mecánicas» el espectro de modos normales del poro (para unos valores  $k_b$ ,  $k_\theta$  determinados, ver ecs. 3 y 4).

El análisis de los modos normales del canal en aproximación armónica (esto es, pequeñas oscilaciones entorno al equilibrio, *i.e* baja temperatura) se llevará a cabo utilizando el módulo específico para tal efecto del paquete *Gromacs* mediante la diagonalización de la matriz hessiana,  $\hat{\mathbf{H}} \equiv H_{ij} = \frac{\partial^2 V(x_i, x_j)}{\partial x_i \partial x_j}$ . *Gromacs* enumera los modos mediante un índice  $n \in [1, 3N]$  en el sentido de frecuencias crecientes; los modos 1-6 están reservados para los movimientos de frecuencia nula: traslación en las tres direcciones del espacio (1-3), y rotación entorno a los ejes  $\hat{x}$ ,  $\hat{y}$ ,  $\hat{z}$  (4-6). Del conjunto de  $3N$  modos normales se seleccionan aquellos que podrían llamarse «modos de respiración» del poro, movimientos colectivos de gran amplitud de onda (y baja frecuencia) que puedan influir en la dinámica del polímero a lo largo del canal en las escalas de tiempo de la traslocación ( $\nu \sim 0,1 \text{ ps}^{-1}$ ); se tomará como paradigma de esta dinámica de respiración el modo nº 15, que presenta la forma expuesta en fig. 13.

A continuación se obtienen sistemáticamente los mapas de frecuencias características de estos modos en un espacio  $k_b \times k_\theta \in [10^3, 2 \cdot 10^5] \times [25, 1200]$ , para canales de diferente radio (0,50 y 0,70 nm); una vez más se reproducen aquí únicamente los mapas obtenidos para el modo de referencia (fig. 14)

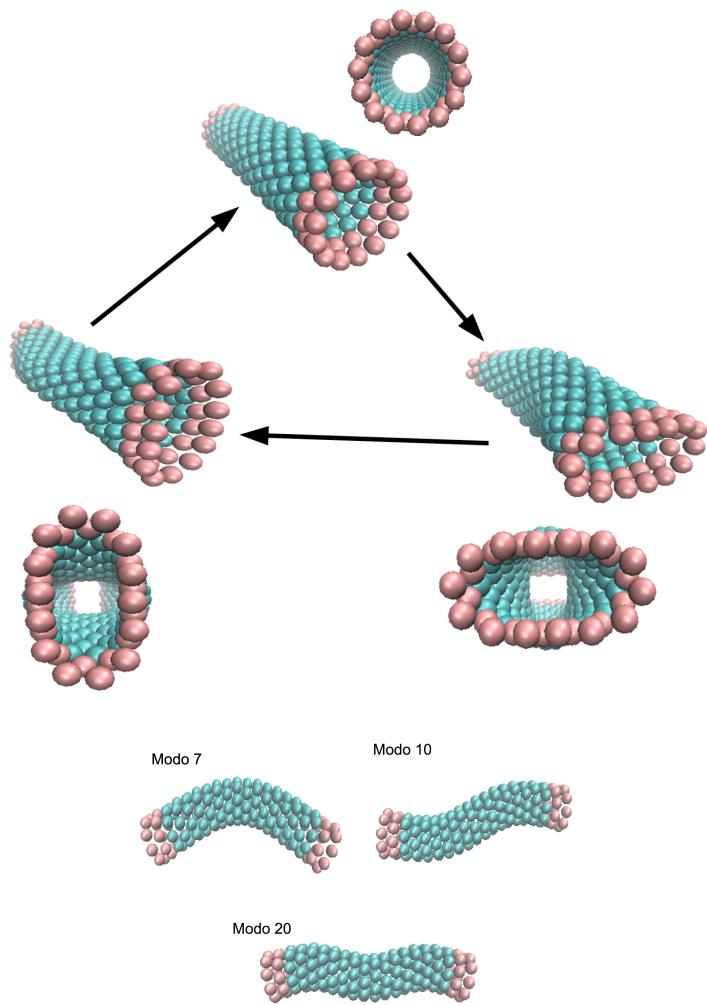


Figura 13: *Arriba*: Configuraciones (amplitudes máximas) del canal oscilando en el decimoquinto modo normal (tomado como referencia de modo de respiración). *Abajo*: ejemplo de otros modos normales (séptimo, décimo y vigésimo).

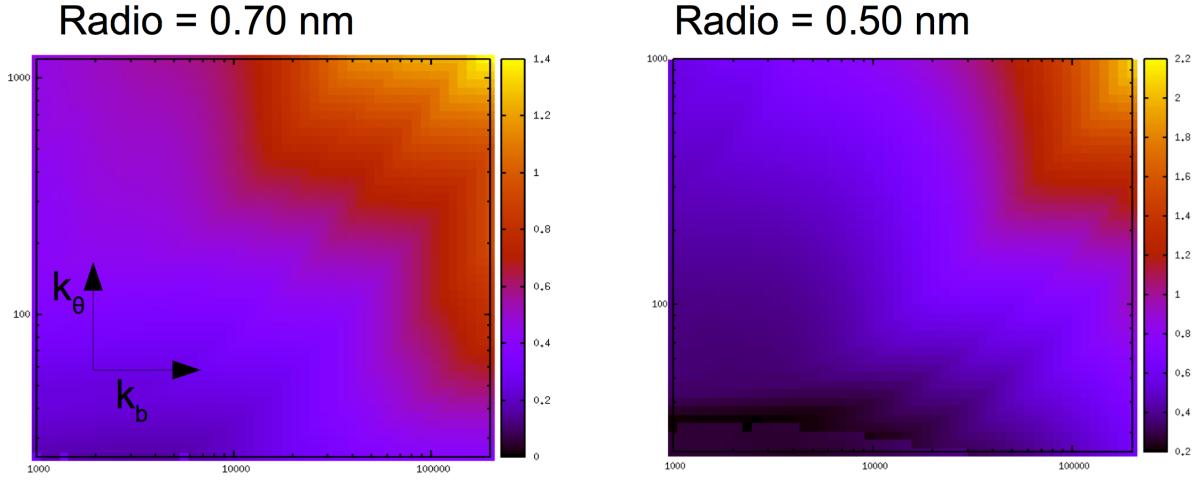


Figura 14: Mapa de las frecuencias características para modos de gran amplitud, en función de las constantes estructurales del poro.

### 5.3. Etapas de la simulación

El programa *script* está diseñado para barrer una línea recta del mapa de frecuencias (fig. 14), produciendo en cada punto una trayectoria promediada sobre varias simulaciones ( $\sim 24$  repeticiones, aunque este número puede variar en función de las condiciones de ruido de la trayectoria, particularmente dependiente de la velocidad de arrastre y la constante de fricción viscosa). Cada simulación utiliza como condiciones iniciales las estructuras descritas en **5.1**, y se compone de varios pasos entre los que se puede distinguir:

**1) Diseño del entorno de simulación** Paso preliminar a cualquier simulación con *Gromacs*; aquí se define la *caja de simulación* (*i.e.* dimensiones de la celda unidad sobre cuyas paredes se aplicarán las condiciones periódicas de contorno), y se centra el sistema en dicho entorno.

Para este trabajo se han utilizado celdas ortoédricas de tal modo que el plano  $XY$  de la caja sea paralelo a la membrana y normal a la dirección de traslocación ( $\hat{z}$ ). Las dimensiones de la caja en el plano de la membrana se calculan como:  $L_{x,y} = \sqrt{S_m} + 2 \times \sigma$ , donde  $L$  denota la longitud de la caja,  $S_m$  la superficie de la membrana y  $\sigma$  el radio de Van der Waals ( $= 0,47\text{ nm}$ ). Esta disposición con el borde de la membrana cercano a la pared, junto con el parámetro de *Gromacs* `periodic - molecules = yes` (que desactiva el criterio de la mínima imagen por el que cada *átomo* solo interactúa con el vecino más cercano a pesar de las condiciones periódicas de contorno), permiten que la membrana se extienda virtualmente más allá del entorno de simulación redundando en su estabilidad, sin aumentar realmente el número de partículas del sistema (*y* por tanto el gasto computacional). En cuanto a la dimensión longitudinal  $L_z = 2 \times l_{pol}$ , siendo  $l_{pol}$  la longitud del polímero completamente estirado.

**2) Primera minimización** Se lleva a cabo utilizando el algoritmo *steepest-descendent* descrito en **3.3.1**, y busca garantizar la estabilidad numérica del módulo *mdrun* durante la simulación de la dinámica (específicamente, la simulación terminará con un error si el movimiento de un *átomo* en un solo intervalo temporal expulsa a este de su dominio de paralelización [20]; para evitar esto deben minimizarse las tensiones como paso previo).

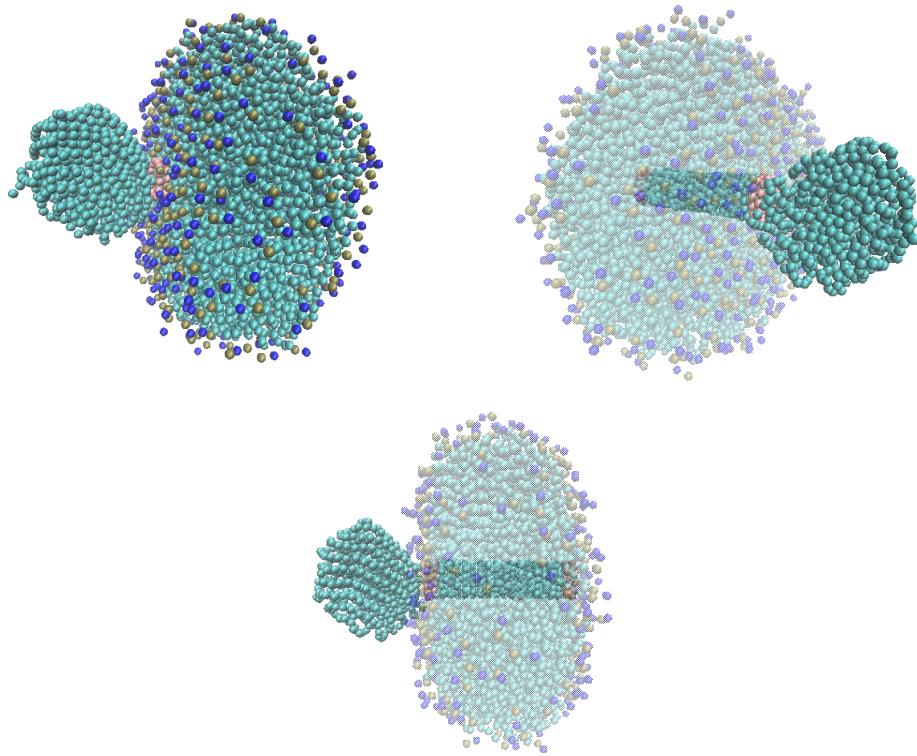


Figura 15: Sistema membrana-poro-polímero termalizado. Si difuminan las partículas que forman parte de la membrana para visualizar la situación del canal transmembrana

**3) Termalización** Orientado a alcanzar el equilibrio térmico del sistema con el baño térmico. La mayor parte del tiempo es consumido en termalizar el polímero, definiendo como tal la situación en la que el *radio de giro* (ver apéndice A) alcanza un estado estacionario; *script* implementa esta condición, sin embargo debido al volumen de cálculo implicado en esta tarea en la práctica se ha utilizado una base de datos de polímeros pre-termalizados. La dinámica del polímero durante la termalización tiende a modificar la porción pre-insertada en el poro: si el canal es grande parte del polímero se puede estabilizar dentro, y si es pequeño puede verse extraído hacia la región *cis*; para evitar esto se bloquea la dinámica de dicho segmento polimérico durante esta fase (fig. 12).

**4) Introducción del solvente en el sistema:** En caso de realizar una simulación que incluya explícitamente la presencia de agua parametrizada, en esta etapa se llena la caja de simulación (utilizando el módulo *solvate*). *Script* modifica el fichero de *topología* de la forma adecuada para incluir las nuevas moléculas.

**5) Segunda minimización:** Simulación idéntica a (2), motivada por la inclusión del solvente recién añadido.

**6) Equilibrado:** Nueva termalización, orientada en este caso a la estabilización de las estructuras del sistema con el agua; se simulan espacios temporales breves (en general no mayores que la mitad del tiempo de termalización), y de hecho esta etapa es totalmente prescindible si el sistema no incluye solvente explícito.

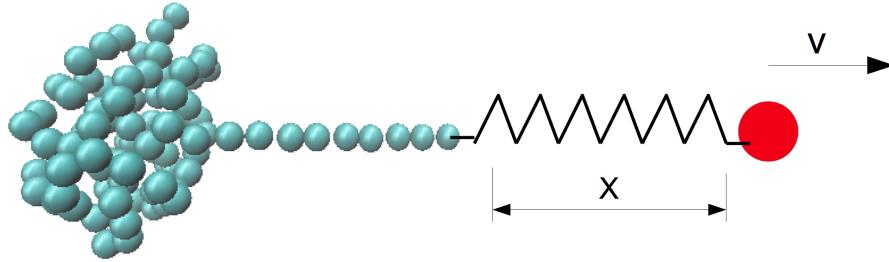


Figura 16: Esquema de tracción; (i) régimen de *fuerza constante*:  $X = \text{cte}$ ,  $v = \text{variable dinámica}$ . (ii) *velocidad constante*:  $X = \text{variable dinámica}$ ,  $v = \text{cte}$

**7) Traslocación:** La traslocación se realiza mediante la simulación de una partícula (virtual) tractora, que ejerce una fuerza armónica sobre el extremo del polímero inicialmente insertado en el poro (en adelante extremo *trans*, y por extensión partícula *trans*). Se distinguen dos regímenes de tracción en función del comportamiento de esta partícula (fig. 16): *fuerza constante*, en el que la partícula *trans* sufre una fuerza constante y predefinida en la dirección de traslocación (por lo cual la velocidad del centro de tracción es incontrolable y debe ser calculada dinámicamente) y *velocidad constante*, que de forma recíproca define la velocidad de la partícula virtual, dejando por tanto a merced de las ecuaciones del movimiento la distancia respecto de *trans* (*i.e* la fuerza ejercida sobre el extremo).

A través de *script* se pueden fijar una gran cantidad de parámetros previo a la simulación; aquellos más característicos para la realización de este trabajo han sido: el radio del poro ( $0,50 \leq r_c \leq 0,70 \text{ nm}$ ), la constante de fricción ( $1,0 \leq \tau^{-1} \leq 2,0 \text{ ps}^{-1}$ ), velocidad de tracción (modo *velocidad*,  $0,01 \leq v \leq 0,5 \text{ ps}^{-1}$ <sup>3</sup>), fuerza de tracción (modo *fuerza*,  $200 \leq F \leq 400 \text{ kJ/nm mol}^{-1}$ ), constante de *bonding* ( $100 \leq k_b \leq 10^5 \text{ kJ/nm}^2 \text{ mol}^{-1}$ ) y flexión (*bending*,  $10 \leq k_\theta \leq 800 \text{ kJ mol}^{-1}$ ) de las partículas del poro (ver fig. 17), y longitud del polímero ( $15 \leq l_{pol} \leq 200 \text{ nm}$ ). La duración total de la simulación ( $t_{sim}$ ) se calcula: (i) para el modo *fuerza*,  $t_{sim} = \tau^{-1} M l_c / F$ , donde  $M \equiv$  masa total del polímero (*i.e*  $M = 72 \times N_{pol} \text{ uma}$ ); (ii) para el modo *velocidad*,  $t_{sim} = N_{pol} / v$ .

Durante la simulación se registran: la posición de la partícula tractora, la fuerza realizada sobre la partícula *trans*, la posición de esta y el anillo del poro en la zona *trans*.

**8) Análisis de resultados:** Se dirá que el polímero ha traslocado cuando el último de sus centros de interacción ha atravesado la membrana, esto es ha penetrado en la región *trans*; la información bruta producida en una sola simulación se materializa en forma de tres ficheros: curva de posición de la partícula tractora frente al tiempo, curva de la fuerza ejercida por esta (sobre el polímero) frente al tiempo, y tiempo de traslocación ( $t_{tras}$ ). El tiempo de traslocación es calculado por *script* a partir de la trayectoria producida por el módulo de simulación *mdrun* (fig. 18).

Una vez se han completado todas las simulaciones a lo largo de una línea en el mapa  $k_b, k_\theta$  se promedian las repeticiones realizadas sobre cada punto.

## 6. Resultados

A continuación se exponen los resultados de las simulaciones llevadas a cabo utilizando la herramienta expuesta en las secciones previas; debido a las exigencias en tiempo de cálculo las simulaciones

<sup>3</sup>En unidades de la distancia de equilibrio para la interacción *bond* de las partículas del polímero,  $b_{0,pol} = 0,33 \text{ nm}$ . Esto es, la velocidad de la tracción será  $0,33 \cdot v \text{ nm/ps}$

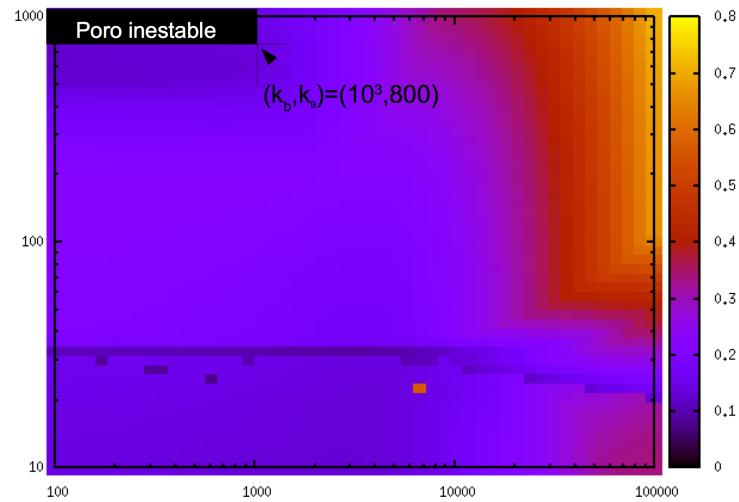


Figura 17: Para valores pequeños de la constante de *bond*  $k_b$  y grandes de *bending*  $k_\theta$ , la atracción del canal con las cadenas alifáticas de la bicapa cae bruscamente, propiciando el arrastre del poro fuera de la membrana durante la traslocación; por ello, se evitará simular en la región sombreada.

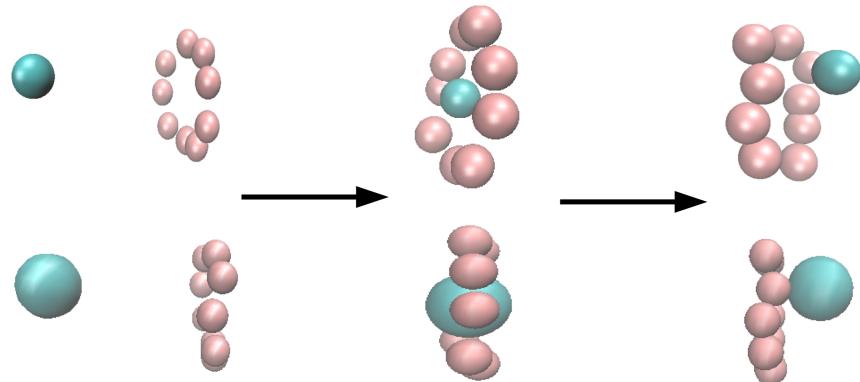


Figura 18: Extracción del tiempo de traslocación a partir de la trayectoria: *script* detecta el fotograma en el que la última partícula del polímero atraviesa el último anillo del poro (entendiendo por tal el más cercano al lado *trans*)

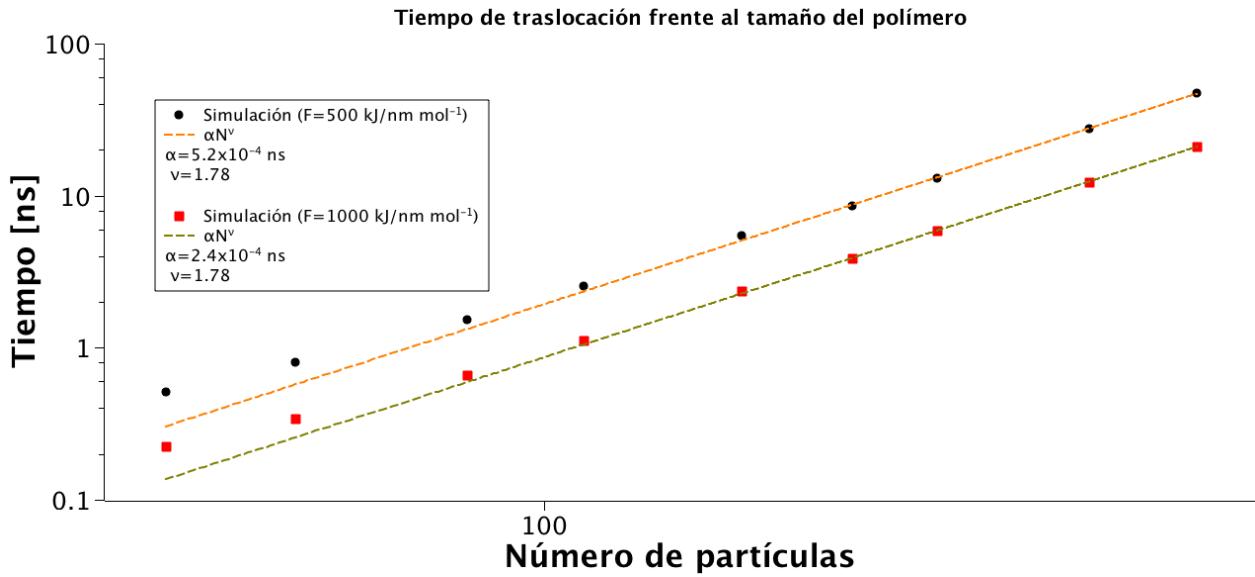


Figura 19: Escalado del tiempo de traslocación con el tamaño del polímero para dos fuerzas diferentes (escala logarítmica)

por defecto han sido realizadas sin incluir agua de forma explícita, para lo cual se implementa una versión (estándar) del campo *MARTINI* adaptado a simulaciones con solvente implícito (*dry\_martini*). Los ejes y magnitudes temporales que aparecen en esta sección son consistentes con las unidades físicas a que acompañan, *i.e* están reescalados en un factor cuatro respecto de los valores numéricos arrojados por las simulaciones (ver sección 4). Así mismo se mantendrá una temperatura constante  $T = 320\text{ K}$ , radio del poro  $r_p = 0,5\text{ nm}$  y longitud del polímero  $l_{\text{pol}} = 30\text{ nm}$  (salvo especificación en contrario).

### 6.1. Verificación del modelo: ley de escala del tiempo de traslocación con el tamaño del polímero

Para comenzar esta sección se reproducirán algunos resultados conocidos a propósito del tiempo de traslocación en función del número de partículas del polímero bajo la acción de una fuerza constante (fig. 19) [10, 11]. Para la siguiente simulación se han empleado valores estándar de rigidez del poro ( $k_b = 7500\text{ kJ/nm}^2\text{mol}^{-1}$ ,  $k_\theta = 50\text{ kJmol}^{-1}$ ), y una constante de fricción  $\tau^{-1} = 1,0\text{ ps}^{-1}$ :

Tal y como cabía esperar se encuentra una dependencia  $t \sim \alpha(F)N^\nu$  para el tiempo de traslocación; de acuerdo con la ley de escala  $t \sim N^{1+\nu_F}$ , se encuentra para el polímero utilizado en este trabajo  $\nu_F = 0,78$ , en buen acuerdo con el exponente de Flory para un polímero tridimensional con volumen excluido e interacción *bending* ( $0,6 \leq \nu_{F_0} \leq 1,0$ , ver sección 2).

### 6.2. Curvas de fuerza y posición durante la traslocación: condiciones estándar

Introducimos las curvas de fuerza  $F(t)$  y posición  $X(t)$  frente al tiempo durante la traslocación. Estas simulaciones han sido realizadas con parámetros estándar ( $k_b = 7500\text{ kJ/nm}^2\text{mol}^{-1}$ ,  $k_\theta = 50\text{ kJmol}^{-1}$ ,  $l_{\text{pol}} = 30\text{ nm}$ ) a velocidad  $v = 0,1\text{ ps}^{-1}$  (en régimen de *velocidad constante*) y fuerzas  $F = 200; 400\text{ kJ/nm mol}^{-1}$  (régimen de *fuerza constante*).

El aspecto más destacable de estas curvas es la evidencia de una barrera de potencial que se opone a la traslocación del polímero en su etapa final: suponiendo que (en el régimen de velocidad

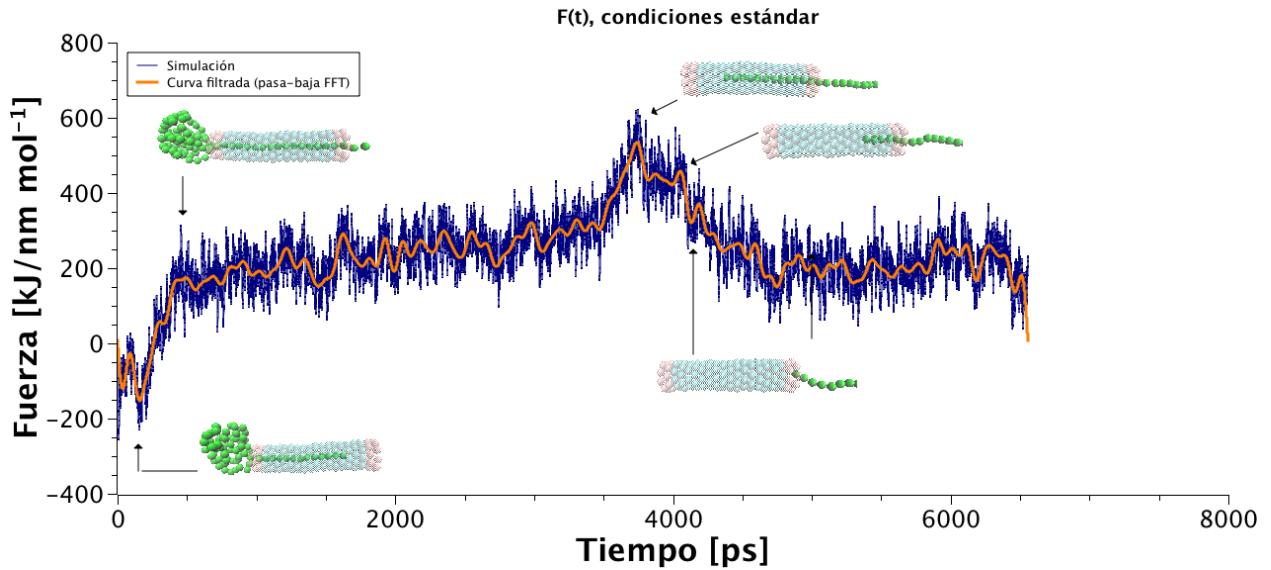


Figura 20: Fuerza frente al tiempo en una simulación estándar a velocidad de tracción constante, donde se han destacado los eventos relevantes de la trayectoria. En naranja se muestra la curva suavizada mediante *transformada rápida de Fourier*

constante) el muelle tractor es lo bastante rígido como para que el polímero se desplace a velocidad  $v$ , y la constante de fricción es lo bastante elevada como para que el polímero se mueva constantemente en velocidad límite, entonces la fuerza neta que este experimenta será:  $F_{\text{neta}} = vM/\tau$ , donde  $M = 72N_{\text{pol}}\text{ una}$ ; dados los valores de estas simulaciones esto implica  $F_{\text{neta}} = 213,84\text{ kJ/nm mol}^{-1}$ . De esta forma se puede expresar la fuerza ejercida por el poro como

$$F_{\text{neta}} = F + f_{\text{poro}} \quad (5)$$

donde  $F$  denota la fuerza registrada en las curvas, *i.e* la fuerza ejercida sobre el polímero *por* la partícula tractora. A partir de las curvas de trabajo realizado sobre el polímero:

$$W(x) = \sum vF(x/v)\Delta t \rightarrow \int F(x) dx$$

y de la expresión 5, puede deducirse el perfil de potencial (fig. 21) como:

$$\begin{aligned} \int F_{\text{neta}} dx &= 213,84\text{ kJ/nm mol}^{-1} \cdot x = W + \int f_{\text{poro}} dx = W - V_{\text{poro}}(x) \\ V_{\text{poro}}(x) &= W(x) - 213,84\text{ kJ/nm mol}^{-1} \cdot x \end{aligned} \quad (6)$$

Bajo la suposición de velocidades de arrastre suficientemente pequeñas, el sistema podría considerarse en condiciones de cuasiequilibrio e identificar la figura 21 con el perfil de energía libre,  $V_{\text{poro}} \simeq \Delta F_{\text{poro}}$ <sup>4</sup> [21].

Por último resulta interesante notar cómo se manifiesta la barrera de energía en las simulaciones a *fuerza constante*, en forma de un descenso de velocidad y en la existencia de una fuerza umbral mínima para superar la traslocación.

<sup>4</sup>Como ya se expresó en la introducción, la dinámica en un régimen *strong stretching* se encuentra lejos del equilibrio. Por lo tanto, este resultado de hecho *no* representa el perfil de energía libre del poro que nos ocupa, y el perfil de potencial obtenido no es extrapolable más allá de las condiciones de la simulación (suficiente sin embargo para ilustrar la presencia de la barrera).

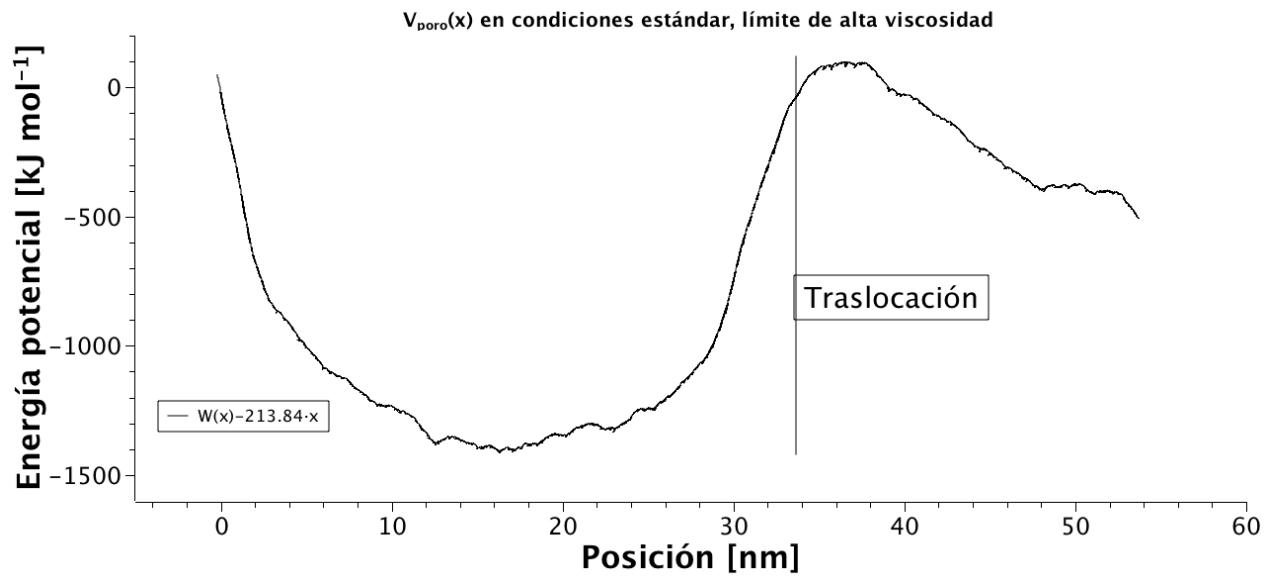


Figura 21: Perfil de potencial del poro, donde se aprecia la barrera responsable del aumento de fuerza antes de la traslocación.

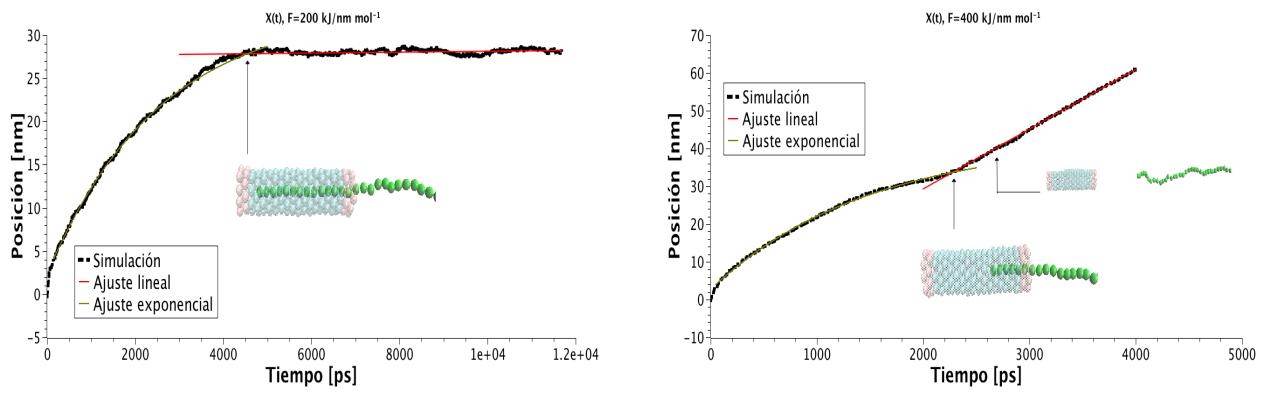


Figura 22: Curvas de posición frente al tiempo para dos simulaciones a *fuerza constante*. Destaca como en la curva superior, la fuerza es insuficiente para vencer la barrera de potencial

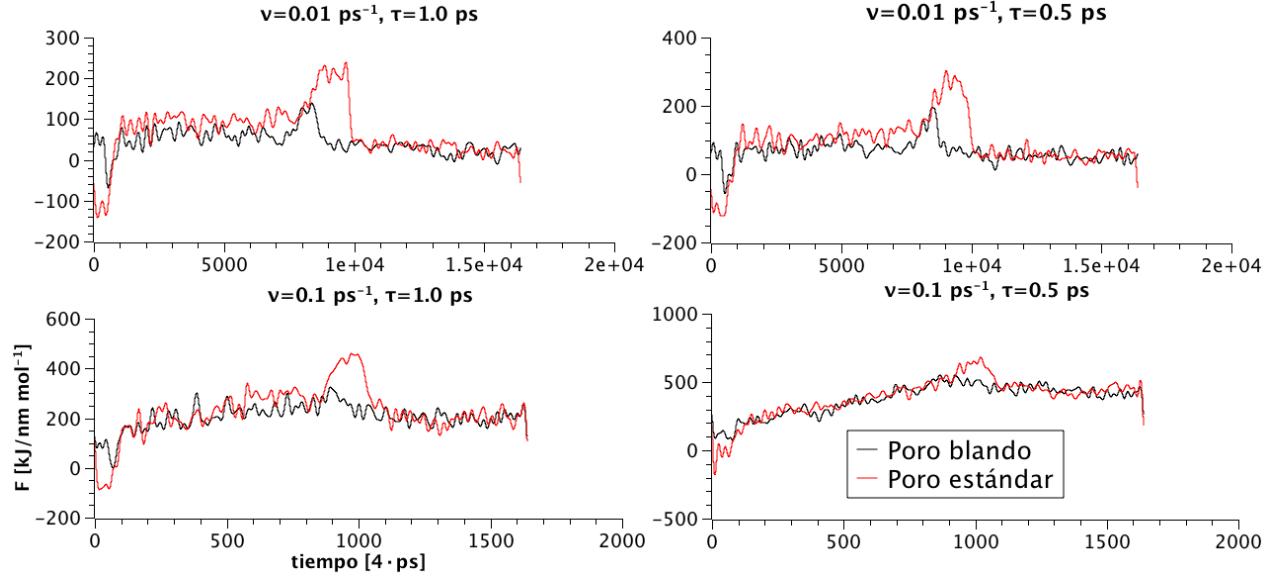


Figura 23: Simulación de traslocación para dos clases de poros, a diferentes velocidades y en medios de distinta fricción.

### 6.3. Influencia de la velocidad de tracción y la fricción en las trayectorias

Debido a la alta dimensionalidad del espacio de parámetros y con el fin de justificar la restricción sobre algunos de ellos en siguientes simulaciones, se muestra una comparativa de varias curvas  $F(t)$  obtenidas mediante traslocación a través de un poro «blando» ( $k_b = 100 \text{ kJ/nm}^2 \text{ mol}^{-1}$ ,  $k_\theta = 10 \text{ kJ mol}^{-1}$ ) y estándar, a velocidades  $v = 0,1 \text{ ps}^{-1}$  y  $v = 0,01 \text{ ps}^{-1}$ , y en medios de diferente viscosidad  $\tau^{-1} = 1 \text{ ps}^{-1}$ ,  $\tau^{-1} = 2 \text{ ps}^{-1}$  (fig. 23).

En todos los casos se observa una reducción de la fuerza necesaria para efectuar la traslocación cuando esta tiene lugar a través de un poro blando frente a uno rígido, tal y como cabría esperar en sistemas biológicos (donde la flexibilidad del poro es importante).

En virtud de estos resultados y teniendo en cuenta el coste computacional (que escala inversamente con la velocidad de arrastre), se decide utilizar por defecto los valores de velocidad y fricción  $v = 0,1 \text{ ps}^{-1}$ ,  $\tau^{-1} = 1 \text{ ps}^{-1}$ .

### 6.4. Mapa del trabajo de traslocación frente a las propiedades mecánicas del poro

El resultado principal de este trabajo es un barrido del espacio de parámetros  $k_b, k_\theta$  (fig. 24), manteniendo fijos otros tal y como se ha justificado previamente (en particular, *velocidad constante*  $v = 0,1 \text{ ps}^{-1}$ ,  $\tau^{-1} = 0,1 \text{ ps}^{-1}$ ,  $l_{\text{pol}} = 30 \text{ nm}$ ,  $T = 320 \text{ K}$ ). Debido a la imposibilidad de mostrar todas las trayectorias obtenidas se utilizará el trabajo total realizado sobre el polímero como observable para caracterizar la dinámica de traslocación.

En la figura se observa una tendencia clara hacia la reducción del trabajo necesario para la traslocación conforme mayor sea el parámetro  $k_\theta$ . Como se puede ver en la ampliación de la fig. 24, esta apreciación se extiende a la constante  $k_b$  en el rango de parámetros adaptado a los sistemas biológicos<sup>5</sup> (en favor de la hipótesis «la flexibilidad del poro favorece el fenómeno de la traslocación»); sin embargo resulta llamativo cómo el valor de  $k_b$  modula este comportamiento en el espacio

<sup>5</sup>Valores moderados de las constantes estructurales.

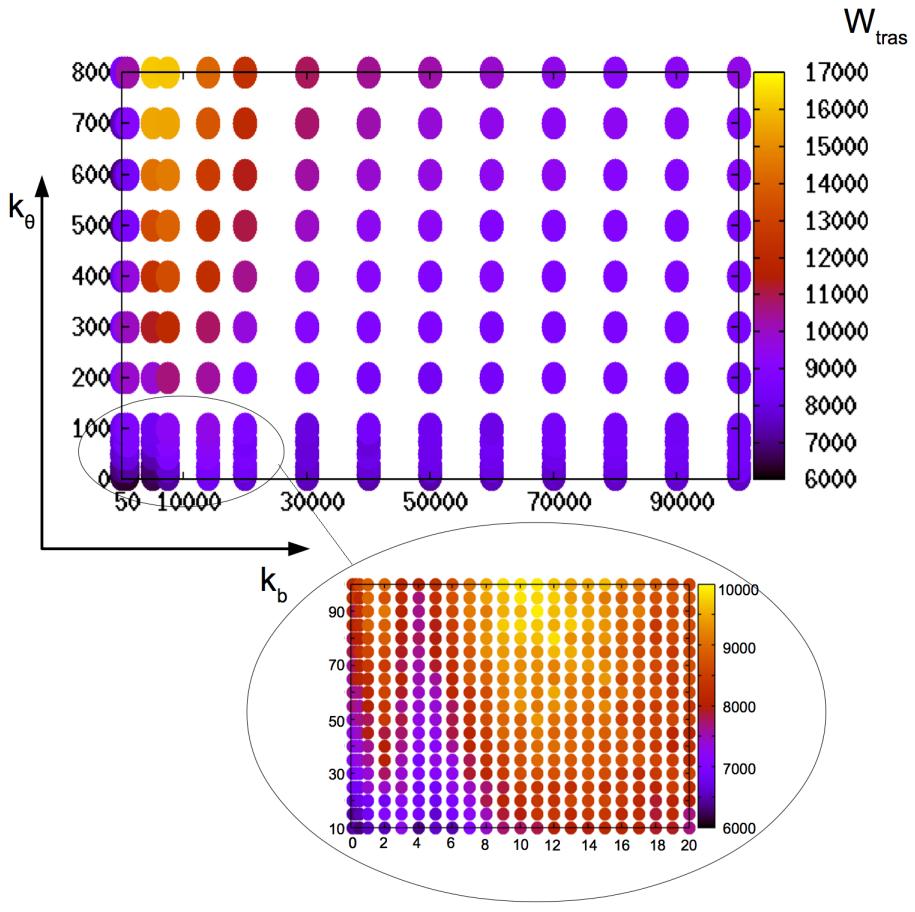


Figura 24: Mapa del trabajo invertido en la traslocación a  $v = cte$  en el espacio  $k_b$  [ $\text{kJ}/\text{nm}^2 \text{mol}^{-1}$ ]  $\times$   $k_\theta$  [ $\text{kJ mol}^{-1}$ ]. El eje de abscisas en la ampliación está en unidades de centenas de  $\text{kJ}/\text{nm}^2 \text{mol}^{-1}$

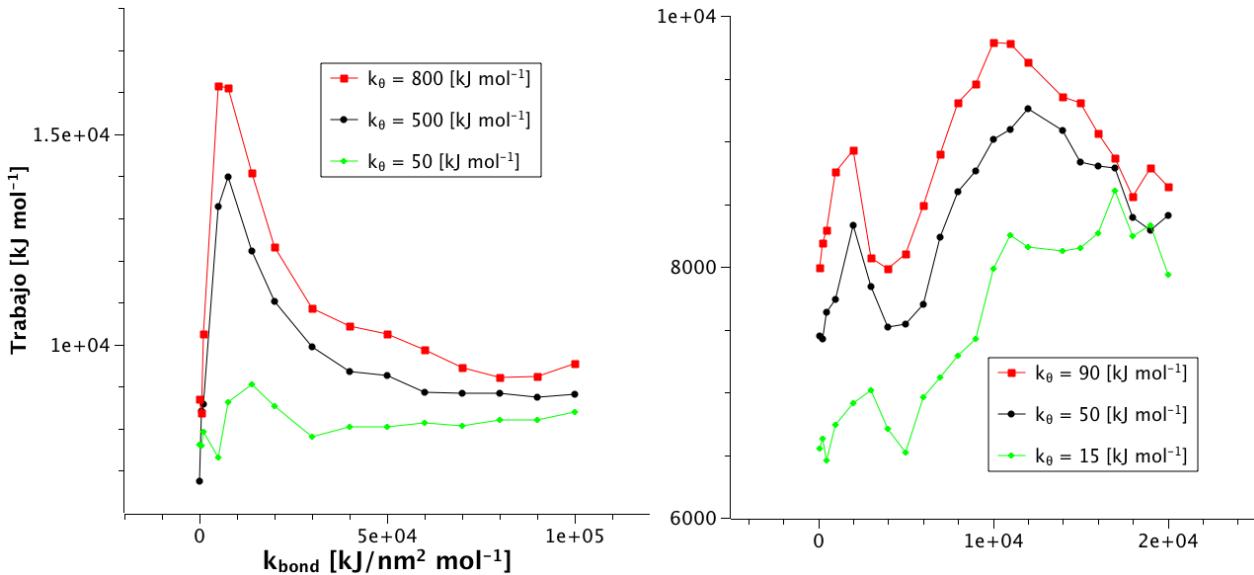


Figura 25: Curvas del trabajo invertido en la traslocación en función del parámetro  $k_b$  del poro, para varios  $k_\theta$ .

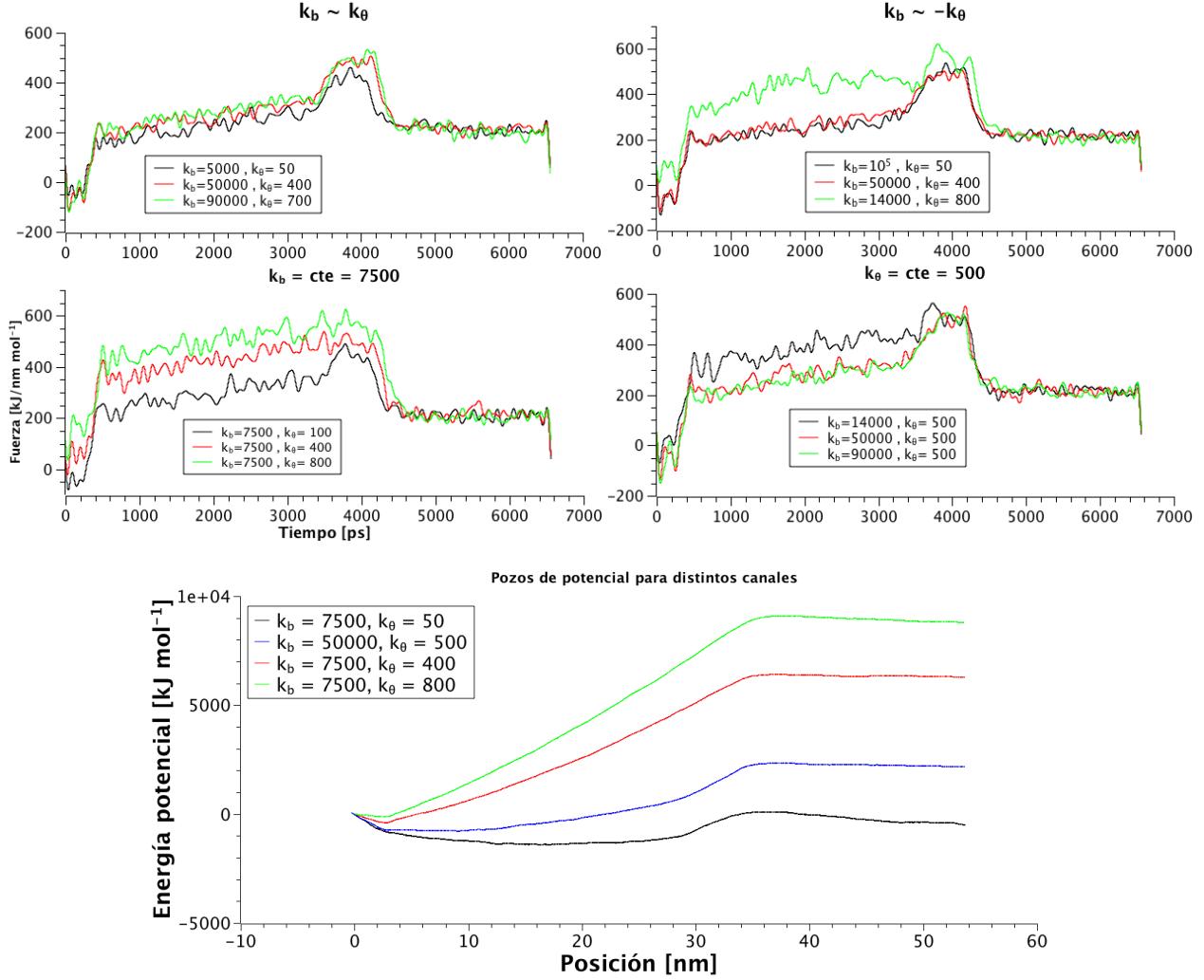


Figura 26: Arriba: curvas  $F(t)$  a lo largo de varias líneas en el espacio  $k_b \times k_\theta$ . Abajo: comparación del perfil de potencial para varios poros.

$k_b, k_\theta$  ampliado, resultando en un poro localmente favorable a la traslocación en la línea  $k_b \propto k_\theta$  (donde «localmente» hace referencia al hecho de que, en términos absolutos, los menores trabajos encontrados pertenecen efectivamente al rango llamado «biológico», tal y como cabría esperar de sistemas altamente optimizados).

En la figura (26) se han representado varias curva para distintos valores de las constantes estructurales del poro, así como distintos perfiles de energía potencial; destaca el mecanismo por el cual valores de flexión ( $k_\theta$ ) elevados se oponen a la traslocación, vía la profundización de la barrera de potencial.

## 7. Resumen del trabajo realizado y conclusiones

En este trabajo se ha diseñado una configuración adecuada para la simulación de la traslocación de polímeros en condiciones inéditas, como es un sistema poro-membrana flexible e interactuante, que ha sido posteriormente implementado en Gromacs vía un campo de fuerzas mesoscópico (*MARTINI*).

Para ello se ha creado un programa auxiliar (*script*, ver apéndice **B**) que actúa como interfaz, gestionando todas las etapas de una simulación desde la elección de parámetros; este programa emplea una librería, creada específicamente, que contiene todas las funciones para:

- Inicializar de forma apropiada el programa de simulación Gromacs a partir de una lista de parámetros de entrada: magnitudes físicas y geométricas del sistema membrana-poro-polímero como la composición química, las constantes de interacciones *bonded*, la longitud del polímero o el radio del canal; y cantidades relacionadas con la gestión de la simulación, *e.g* cantidad de estadística, número de procesadores que emplear en la paralelización, etc. El sistema biofísico *generado* consta de (*i*) una *membrana*, constituida por una bicapa lipídica de moléculas *DSPC* cuyas colas hidrofóbica se disponen enfrentadas, formando una geometría plana en el eje XY. Para ello *script* lee la estructura general de un único fosfolípido *DSPC* a partir de un fichero externo, calcula su centro de masas y su orientación en la dirección longitudinal de las colas alifáticas, y manipula estas dos cantidades para disponer el número adecuado de moléculas en la posición correcta de la membrana, lo cual incluye evitar una pequeña región perpendicular al plano de la bicapa donde se ubicará el canal que forma el poro. (*ii*) Este *poro* será una estructura de geometría cilíndrica y hueca, que se construye a partir del plegamiento de una red hexagonal de átomos; los parámetros de rigidez del poro se han modelado mediante una interacción a dos cuerpos ( $k_b$ ) entre primeros vecinos, y una interacción de flexión ( $k_\theta$ ) entre tríos de partículas adyacentes; además *Script* distingue dos regiones en el canal para emular la distribución de aminoácidos polares/hidrófobos en las proteínas transmembrana: una zona central, y dos extremos. La química de ambas regiones debe ser pasada a *script* como entrada (sin restricciones, aunque la estructura propuesta y explorada en este trabajo consta de centro hidrófobo y extremos polares). Una vez almacenadas las posiciones relativas del poro en memoria, el programa repite la operación llevada a cabo con los fosfolípidos para calcular el centro de masas y la orientación del canal, e insertarlo en el hueco de la membrana. Y (*iii*) un *polímero* que se construye linealmente a partir de un número de subunidades especificado en la entrada de *script* y es posteriormente insertado en el canal.
- A partir de la información de entrada sobre número de puntos y número de realizaciones a simular, *script* crea una serie de subprocessos y asigna el lanzamiento de cada uno de ellos según la máquina en que se ejecutara: (*i*) en un *cluster* multinodo, se asignará a una *CPU* concreta dentro de un núcleo específico, de forma que dentro de un mismo nodo solo se simulan repeticiones del mismo punto  $k_b, k_\theta$  para obtener estadística (en paralelo); (*ii*) en una máquina de *GPUs* se crea un subprocesso por cada procesador gráfico y se reparten de forma equilibrada las *CPU* entre el número total de estos procesos, de manera que cada uno de ellos realice la simulación de un único punto del mapa, y la estadística se obtiene mediante la repetición en serie en el mismo nodo; y (*iii*) en un único nodo se comporta como un caso particular de (*i*).
- Una vez han concluido la integración de la dinámica, *script* extrae el tiempo de traslocación de cada repetición utilizando para ello el fichero de trayectorias producido por Gromacs, de donde calcula en cada instante las posiciones relativas del último monómero que entra en el poro y el último anillo del canal, extrayendo el instante en el que ambas estructuras se cruzan (*i.e* se ha producido la traslocación). Con esta información y las curvas fuerza-tiempo (en simulaciones  $v = cte$ ) o posición-tiempo ( $F = cte$ ), *script* promedia las repeticiones de cada punto para generar un único valor (promedio) del tiempo de traslocación y una única curva (promedio), por cada punto del mapa de parámetros  $k_b, k_\theta$ .

En cuanto a la exploración del sistema propuesto, se ha encontrado una dependencia de la traslocación del polímero con las propiedades mecánicas del poro, entendidas como tal la intensidad del

enlace químico entre sus componentes ( $k_b$ ), así como su resistencia a ser doblado (rigidez,  $k_\theta$ ). Se observa una tendencia clara hacia el incremento del trabajo requerido, y la presencia de una fuerza umbral creciente, conforme aumenta el parámetro de rigidez (modulado en cualquier caso por la intensidad del enlace entre las partículas del poro); recíprocamente se aprecia una facilidad para la traslocación en valores propios de un poro blando (que coincide con el rango de valores biológicos, de acuerdo a las parametrizaciones del campo de fuerzas *MARTINI*). También se puede destacar la verificación de las estructuras empleadas mediante el ajuste de las leyes de escala [del tiempo de traslocación en función de la longitud del polímero] del tipo ley de potencia,  $t_{tras} = N^{1+\nu}$ , donde  $\nu$  es el exponente de Flory, cuyo valor está sustancialmente confirmado  $\nu = 0,78$ . Las simulaciones se han ejecutado con polímeros de grandes dimensiones, hasta 600 monómeros. Otro de los cálculos realizados ha sido estimar la barrera de potencial responsable de la dinámica de traslocación.

En relación a las limitaciones del trabajo, se pueden destacar principalmente dos que han dificultado la explotación del protocolo: en primer lugar, la vasta dimensión del espacio de parámetros a explorar, que ha impuesto la restricción (desde etapas tempranas del estudio) sobre ciertas cantidades, *e.g* coeficiente de fricción, química del poro y del polímero, constantes estructurales del polímero, temperatura, etc; y por otro lado (aunque en relación con lo anterior), las altas exigencias de cálculo que han cerrado algunas vías de exploración, como por ejemplo la simulación con solvente explícito.

A pesar de estas dificultades, se han podido extraer algunas conclusiones cuantitativas. En concreto se ha calculado la fuerza media, trabajo medio, y posición en función del tiempo, en cuyas trayectorias se pueden destacar las distintas fases de progresión del polímero (traslocación de la partícula tractora, despliegue del polímero en equilibrio, su total entrada en el poro, salida parcial y finalmente, escape del potencial del canal). El estudio de estas magnitudes en función de los parámetros de flexibilidad del poro ha mostrado una interesante rica fenomenología entre la cual se puede destacar la presencia abundante de extremos relativos en las curvas de trabajo y fuerza aplicadas, enseñando también condiciones óptimas.

La confirmación de que el fenómeno de la traslocación es sensible a las condiciones expuestas abre la puerta al campo de la espectroscopía de fuerzas y al uso potencial de este sistema como secuenciador/analizador de polímeros.

## Referencias

- [1] Alberts, Bray, Hopkin, Johnson, Lewis, Raff, Roberts, Walter; *Essential cell biology*, Garland Science, (2010)
- [2] Meni Wanunu, *Nanopores: A journey towards DNA sequencing*, Phys. Life Rev. **9**, 125 - 158, (2012)
- [3] Michel G. Gauthier; *Simulation of polymer translocation through small channels*, Universidad de Ottawa (tesis doctoral), (2003)
- [4] John J. Kasianowicz *et. al*, *Characterization of individual polynucleotide molecules using a membrane channel*, Proc. Natl. Acad. Sci. **93**, 13770–13773, (1996)
- [5] Timo Ikonen, *Dynamics of Escape and Translocation of Biopolymers* , Universidad de Aalto (tesis doctoral), (2012)
- [6] Murugappan Muthukumar, Calin Plesa, y Cees Dekker, *Single-molecule sensing with nanopores*, Physics Today **68**, 10.1063/PT.3.2881, (2015)

- [7] A. Fiasconaro, J.J. Mazo, F. Falo, *Active polymer translocation in the three-dimensional domain*, Phys. Rev E **91**, 10.1103/PhysRevE.91.022113, (2015)
- [8] A. Fiasconaro, J.J. Mazo, F. Falo, *Translocation of a polymer chain driven by dichotomous noise*, J. Stat. Mech. Theor. Exp, 10.1088/1742-5468/2011/11/P11002, (2011)
- [9] W. Sung, P. J. Park, *Polymer Translocation through a Pore in a Membrane*, Phys. Rev. Lett. **77**, 783 - 786, (1996)
- [10] Payam Rowghanian, Alexander Y. Grosberg, *Force-Driven Polymer Translocation through a Nanopore: An Old Problem Revisited*, J. Phys. Chem. B, dx.doi.org/10.1021/jp204014r, (2011)
- [11] Kaifu Luo, Santtu T. T. Ollila, Ilkka Huopaniemi *et. al*, *Dynamical scaling exponents for polymer translocation through a nanopore*, Phys. Rev. E **78**, 10.1103/PhysRevE.78.050901, (2008)
- [12] Jalal Sarabadani, Bappa Ghosh, Srabanti Chaudhury, Tapio Ala-Nissila, *Dynamics of end-pulled polymer translocation through a nanopore*, arXiv:1708.09184 [cond-mat.soft], (2017)
- [13] T. Ikonen, A. Bhattacharya, T. Ala-Nissila, W. Sung, *Unifying model of driven polymer translocation*, Phys. Rev. E **85**, 10.1103/PhysRevE.85.051803, (2012)
- [14] M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess *et.al*, *GROMACS User Manual version 5.1.4*, (2016)
- [15] Kristyna Pluhackova y Rainer A. Böckmann, *J. Phys. Condens. Matter* **27**, 10.1088/0953-8984/27/32/323103, (2015)
- [16] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, Alex H. de Vries, *The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations*, J. Phys. Chem **111**, 7812-7824, (2007)
- [17] SieWert J. Marrink y D.Peter Tieleman, *Perspective on the Martini model*, Chem. Soc. Rev **42**, 10.1039/c3cs60093a, (2013)
- [18] Ulrich F. Keyser, *Nanopores – mission accomplished and what next? Comment on “Nanopores: A journey towards DNA sequencing” by M. Wanunu*, Phys. Life Rev. **9**, 170 - 171, (2012)
- [19] Tom A. Rapoport, Long Li, Eunyong Park, *Structural and Mechanistic Insights into Protein Translocation*, Annu. Rev. Cell Dev. Biol, <https://doi.org/10.1146/annurev-cellbio-100616-060439>, (2017)
- [20] Berk Hess, Carsten Kutzner, David van der Spoel, Erik Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, JCTC, 10.1021/ct700301q, (2008)
- [21] Christopher Jarzynski, *Equalities and Inequalities: Irreversibility and the Second Law of Thermodynamics at the Nanoscale*, Annu. Rev. Condens. Matter Phys. **2**, 329-351, (2011)