



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE COMPUTACIÓN

Herramienta web para análisis de Dinámica molecular (DM): Modelo de electroporación de membranas celulares

Tesis presentada para optar al título de
Licenciada en Ciencias de la Computación

Karina Giselle Borgna

Director: Dra. María Laura Fernández

Codirector: Dr. Marcelo Risk

Buenos Aires, 2015

HERRAMIENTA WEB PARA ANÁLISIS DE DINÁMICA MOLECULAR (DM): MODELO DE ELECTROPORACIÓN DE MEMBRANAS CELULARES

La presente tesis es el desarrollo de una aplicación web interactiva, que permite analizar resultados obtenidos por simulaciones de dinámica molecular. Una herramienta de estas características puede no solo personalizar el análisis de los datos sino además representarlos en un entorno gráfico amigable para el usuario. Para desarrollar esta aplicación se utilizaron datos de simulación computacional de membranas celulares, de manera de conformar una base de datos accesible en un entorno web. Para verificar el funcionamiento de la aplicación se seleccionó una simulación de una bicapa lipídica expuesta a un campo eléctrico, ya que requiere un alto costo computacional en el manejo de los datos. El sitio web se realizó con el framework Django basado en Python, mientras que la base de datos se implementó en PostgreSQL. La aplicación obtenida permite cargar datos de una trayectoria, luego seguir cada átomo o molécula para generar resultados tanto individuales como por grupos. Los resultados se pueden analizar estadísticamente en módulos personalizados generando también gráficos y reportes. Además se desarrolló un módulo para la caracterización tridimensional de estructuras biológicas, en particular en aplicaciones de electropermeabilización de membranas celulares. Una de las más importantes aplicaciones clínicas de la electropermeabilización es la electroquimioterapia para el tratamiento de tumores.

Palabras claves: simulación; aplicación web; análisis de datos; dinámica molecular; electroporación.

WEB ANALYSIS TOOL FOR MOLECULAR DYNAMICS (MD): ELECTROPORATION OF CELL MEMBRANES MODEL

This work concerns the development of an interactive web application that allows to analyze results obtained by molecular dynamics simulations. A tool of these characteristics not only allows to personalize the analysis of the obtained data, but also can represent it in a windows system in a graphical and friendly way. For this study data generated by computational simulation of cell membranes were used to be incorporated into a database accessible in a web environment. To test the application functionality a simulation of a lipid bilayer exposed to an electric field was selected because of the great computational cost that this implies to manage the data. The web site was developed with the Django framework based in Python, and the database was implemented using PostgreSQL. The obtained application allows to load information about a trajectory and then can follow each atom or molecule to generate results individually and also by groups. The results can be analyzed statistically in custom modules and generating graphs and reports . Additionally, it was developed a module for tridimensional characterization of biological structures especially developed for applications of electroporation in cell membranes.

Keywords: Simulation; web application; data analysis; molecular dynamics; electroporation.

Índice general

| | |
|---|----|
| Lista de figuras | IV |
| 1.. Introducción | 1 |
| 2.. Objetivos | 8 |
| 3.. Detalles técnicos de la simulación computacional | 9 |
| 3.1. Simulaciones en membranas | 9 |
| 4.. Desarrollo de la Herramienta Web para Post-análisis de Simulaciones de DM | 12 |
| 4.1. Descripción de la herramienta y el modelo de datos | 12 |
| 4.2. Instalación de la aplicación | 16 |
| 4.3. Ejecución de la aplicación | 17 |
| 4.4. Descripción de la interfaz | 17 |
| 4.5. Detalles de implementación | 41 |
| 4.6. Optimizaciones realizadas a la aplicación | 43 |
| 4.7. Decisiones tomadas | 44 |
| 5.. Verificación del Desarrollo de la Herramienta Web | 48 |
| 6.. Conclusiones | 65 |
| 6.1. Alcance de los resultados obtenidos | 65 |
| 6.2. Discusión | 65 |
| 6.3. Futuros trabajos | 66 |
| Bibliografía | 68 |

Índice de figuras

| | |
|---|----|
| 3.1. Se muestra la caja a la que se hace referencia durante este trabajo. | 10 |
| 3.2. Se muestra la estructura celular | 10 |
| 3.3. Poro en la célula al aplicar un campo eléctrico | 11 |
| 3.4. Poro en la célula al aplicar un campo eléctrico | 11 |
| 4.1. Esquema de la herramienta web. | 14 |
| 4.2. Página principal. | 18 |
| 4.3. Formulario para crear una trayectoria. | 18 |
| 4.4. Selección de tipo de análisis de trayectoria. | 19 |
| 4.5. Configuración del análisis de trayectoria para un átomo. | 20 |
| 4.6. Posiciones en X de un átomo determinado. | 20 |
| 4.7. Posiciones en Y de un átomo determinado. | 21 |
| 4.8. Posiciones en Z de un átomo determinado. | 22 |
| 4.9. Velocidades en X de un átomo determinado. | 23 |
| 4.10. Velocidades en Y de un átomo determinado. | 24 |
| 4.11. Velocidades en Z de un átomo determinado. | 25 |
| 4.12. Configuración del análisis de trayectoria para más de un átomo. | 26 |
| 4.13. Gráfico con las velocidades en X del grupo de átomos seleccionado y sus valores estadísticos. | 28 |
| 4.14. Gráfico con las velocidades en X del grupo de átomos seleccionado y la media móvil. | 29 |
| 4.15. Gráfico con las velocidades en X del grupo de átomos seleccionado y sus curvas envolventes. | 30 |
| 4.16. Gráfico con los valores estadísticos de las velocidades en X del grupo de átomos seleccionado. | 31 |
| 4.17. Gráfico de velocidades por posiciones - plano X. | 32 |
| 4.18. Gráfico de la velocidades por posiciones - plano Y. | 33 |

| | |
|--|----|
| 4.19. Gráfico de la velocidades por posiciones - plano Z. | 34 |
| 4.20. Gráfico de la componente X de las velocidades por posiciones - plano X. . . | 35 |
| 4.21. Gráfico de la componente Y de las velocidades por posiciones - plano Y. . . | 36 |
| 4.22. Gráfico de la componente Z de las velocidades por posiciones - plano Z. . . | 37 |
| 4.23. Tabla de datos estadísticos. | 38 |
| 4.24. Configuración para obtener el volumen del poro formado. | 38 |
| 4.25. Gráfico que muestra el volumen del poro a lo largo de la trayectoria, con un valor establecido por defecto de los límites dentro de la membrana. . . . | 39 |
| 4.26. Gráfico que muestra el volumen del poro a lo largo de la trayectoria perso- nalizando los límites dentro de la membrana. | 40 |
| 4.27. Se visualizan las velocidades por posiciones de los átomos de P8 con los cuales se delimita la membrana, en el tiempo inicial cuando aún no hay un poro formado. | 42 |
| 4.28. Se visualizan las velocidades por posiciones de los átomos de P8 con los cuales se delimita la membrana, en el tiempo final cuando hay un poro formado. | 42 |
| 4.29. Se divide el poro en elipsoides de igual espesor. | 43 |
| 4.30. Gráfico de las velocidades por posiciones que muestra el poro formado en la membrana, descentrado. | 45 |
| 4.31. Representación de un poro formado en la membrana, mostrando el caso ideal en que está centrado. | 45 |
| 4.32. Representación de un poro formado en la membrana, partido en dos. | 46 |
| 4.33. Representación de un poro formado en la membrana, ubicado en las esquinas de la caja. | 46 |
| 5.1. Coordenada X de las posiciones del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su posición en el eje X en cada tiempo. | 49 |

| | |
|--|----|
| 5.2. Coordenada Y de las posiciones del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su posición en el eje Y en cada tiempo. | 50 |
| 5.3. Coordenada Z de las posiciones del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su posición en el eje Z en cada tiempo. | 51 |
| 5.4. Coordenada X de las velocidades del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su velocidad en el eje X en cada tiempo. | 52 |
| 5.5. Coordenada Y de las velocidades del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su velocidad en el eje Y en cada tiempo. | 53 |
| 5.6. Coordenada Z de las velocidades del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su velocidad en el eje Z en cada tiempo. | 54 |
| 5.7. Velocidad absoluta en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje X, viendo las posiciones en Z e Y. | 55 |
| 5.8. Velocidad absoluta en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje Y, viendo las posiciones en X y Z. | 55 |
| 5.9. Velocidad absoluta en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje Z, viendo las posiciones en X e Y. | 56 |
| 5.10. Coordenada X de la velocidad en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje X, viendo las posiciones en Z e Y. | 56 |
| 5.11. Coordenada Y de la velocidad en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje X, viendo las posiciones en X y Z. | 57 |

| | |
|--|----|
| 5.12. Coordenada Z de la velocidad en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje X, viendo las posiciones en X e Y. | 58 |
| 5.13. Desvío estándar, la media, el mínimo y el máximo de la coordenada X de las velocidades en cada tiempo, del grupo de átomos seleccionado para analizar a lo largo de toda la trayectoria. | 59 |
| 5.14. Desvío estándar, la media, el mínimo y el máximo de la coordenada X de las velocidades en cada tiempo, del grupo de átomos seleccionado para analizar a lo largo de toda la trayectoria. Además, los ítems muestran las velocidades en X para cada átomo del grupo. | 60 |
| 5.15. Media móvil, las curvas envolventes, el desvío estándar, la media, el mínimo y el máximo de la coordenada X de las velocidades en cada tiempo, del grupo de átomos seleccionado para analizar a lo largo de toda la trayectoria. La media móvil se calculó promediando 5 puntos. | 61 |
| 5.16. Media móvil, el desvío estándar, la media, el mínimo y el máximo de la coordenada X de las velocidades en cada tiempo, del grupo de átomos seleccionado para analizar a lo largo de toda la trayectoria. La media móvil se calculó promediando 5 puntos. | 62 |
| 5.17. Volúmen que alcanza el poro formado en la membrana en cada momento de la trayectoria. En este caso, se establecieron los límites de la membrana en 3,4 nm y 5,8 nm. | 63 |
| 5.18. Velocidad absoluta en el tiempo 15 ns, definida por el color del ítem, de cada átomo perteneciente a moléculas de lípidos. Se puede ver la forma del poro desde el eje Z, viendo las posiciones en X e Y. | 64 |

1. INTRODUCCIÓN

La simulación computacional de moléculas o estructuras biológicas es una poderosa herramienta que permite modelar aquellos fenómenos producidos en la experimentación de laboratorio. Las simulaciones *in silico* son un complemento para analizar los resultados obtenidos por otros métodos ya sea porque permiten simular escenarios que por su escala nanométrica son difíciles de analizar en el laboratorio o porque permiten una variedad de condiciones que por su costo no serían viables a la hora de desarrollar un protocolo experimental, lo que agrega una nueva dimensión en la producción de resultados. Una de las estrategias más utilizadas para realizar modelos y ponerlos a prueba es mediante la técnica de dinámica molecular (DM). Para tomar dimensión de la importancia de la DM se cita el siguiente artículo *Significance of Molecular Dynamics Simulations for Life Sciences* publicado por Martin Karplus y Richard Lavery (Martin Karplus obtuvo junto a Michael Levitt y Arieh Warshel el Premio Nobel de Química en el año 2013) donde se muestra con diferentes ejemplos cómo la DM es capaz de aportar significativamente a las ciencias de la vida. [Karplus and Lavery, 2014]. Técnicamente la DM resuelve de manera numérica las leyes de movimiento de Newton para un sistema de N partículas que interaccionan, describiendo de manera temporal y en detalle el movimiento y la trayectoria de éstas (átomos, conjunto de átomos, moléculas o conjuntos de moléculas), así como sus interacciones luego de fijadas las condiciones experimentales para la simulación. De esta manera provee el movimiento de partículas individuales en función del tiempo siguiendo las leyes de la física clásica.

Existen una variedad de desarrollos computacionales utilizados para simulación que incorporan cálculos de mecánica molecular clásica como DM. En este trabajo se utilizaron resultados obtenidos por simulación y modelado molecular generados por Gromacs, que es una herramienta que realiza DM y minimización de la energía [Abraham, van der Spoel, Lindahl, Hess, and the GROMACS development team, 2014]. Esta herramienta es de uso libre y es una de los más utilizados para la simulación de membranas mediante DM,

ya que admite paralelizar los cálculos lo que permite llevar a cabo simulaciones para un número muy elevado de partículas. ([Berendsen, van der Spoel, and van Drunen, 1995], [Hess, Kutzner, van der Spoel, and Lindahl, 2008], [Pronk, Páll, Schulz, Larsson, Bjelkmar, Apostolov, Shirts, Smith, Kasson, van der Spoel, Hess, and Lindah, 2013], [Spoel, Lindahl, Hess, Groenhof, Mark, and Berendsen, 2005]).

Para simular moléculas por DM se utiliza una forma funcional y un conjunto de parámetros obtenidos a partir de datos experimentales o de mecánica cuántica, todos contenidos en campos de fuerzas, de modo de obtener la energía potencial. En el campo de fuerzas entonces se concentran todos los datos necesarios para describir una partícula (átomo, conjunto de átomos) ya sea mediante sus propiedades o la relación de la misma con sus vecinos según tengan o no conectividad (longitudes de enlace, ángulos, etc.). De este modo permite el análisis de la trayectoria de dichas partículas y moléculas así como su interacción de manera temporal.

Hay varios tipos de campos de fuerzas con distinto grado de detalle: *all atom* (todos los átomos), *united atom* (átomos unidos) y *coarse grained* (grano grueso) [Wang and Deserno, 2010]. Así se puede elegir simular moléculas con campos de fuerzas que describen todos los átomos, por ejemplo para proteínas o con campos de fuerzas de átomos unidos, en el caso de lípidos, representando cada carbono alifático y sus hidrógenos asociados como una sola partícula (o pseudo átomo), lo que reduce el costo computacional en simulaciones de elevado número de átomos como ocurre con las membranas biológicas, permitiendo analizar fenómenos que ocurren en tiempos más largos de simulación.

El modelado de membranas mediante DM permite describir los fenómenos que ocurren a escala molecular siguiendo una trayectoria a lo largo del tiempo. Una trayectoria refleja la posición y velocidad de cada átomo o partícula en cada instante de tiempo correspondiente al intervalo de muestreo. Para este trabajo se ha elegido un modelo de membrana con aplicación de un campo eléctrico, simulando el tiempo suficiente como para que en esta membrana se genere un poco. Este fenómeno, denominado electroporación o electroporación de la membrana celular, es de suma importancia ya que está siendo utilizado en medicina para el tratamiento de ciertos tumores. Mediante esta técnica expe-

rimental se aplica un campo eléctrico externo que permite el ingreso de drogas citotóxicas que normalmente serían impermeables a la membrana afectando solamente a las células que se encuentran entre los electrodos, pudiendo elegir selectivamente el campo de acción para la droga, ya que la disposición de los electrodos es periférica al tumor [Belehradek, Domenge, Luboinski, Orłowski, Behlradek, and Mir, 1993], [Orłowski, Behlradek, Paoletti, and Mir, 1988]. A escala molecular se ha predicho que el ingreso de la droga se logra mediante poros generados en la membrana de las células durante la aplicación del pulso eléctrico, sin embargo los eventos producidos a esta escala sólo se han podido dilucidar mediante DM, demostrando la potencia de esta técnica ([Böckmann, de Groot, Kakorin, Neumann, and Grubmüller, 2008], [Fernández, Marshall, Sagués, and Reigada, 2010], [Fernández, Risk, Reigada, and Vernier, 2012], [Tieleman, 2004], [Vernier, Ziegler, Y Sun, and Tieleman, 2006], [Ziegler and Vernier, 2008]). En estos trabajos se sugiere que el inicio del poro está liderado por las aguas que se orientan debido a la aplicación de un campo eléctrico y logran introducirse y atravesar la porción hidrofóbica de la membrana, situación que en ausencia de campo eléctrico sería energéticamente desfavorable.

Luego la porción hidrofílica de los lípidos se orienta espacialmente para acompañar el movimiento de las aguas dando por resultado un poro que atraviesa de lado a lado la membrana, mientras dure el campo eléctrico ([Tokman, Lee, Levine, Ho, Colvin, and Vernier, 2013], [Ziegler and Vernier, 2008]). Esta situación a escala de membrana celular está esquematizada en las figuras 3.1, 3.2, 3.3 y 3.4 .

Dado el costo computacional que significa seguir a lo largo de una trayectoria más de 27000 partículas durante 25 ns, se ha elegido un sistema de membrana sometida a campo eléctrico para desarrollar una aplicación web que permita realizar un post-análisis de las simulaciones hechas en este caso con el paquete Gromacs. Se eligió un sistema biológico de alta demanda de cómputo de modo de poner a prueba la potencia de la aplicación desarrollada. En este trabajo se implementa una interfaz web que permite el seguimiento de cada partícula o conjunto de partículas a lo largo de una trayectoria realizada con Gromacs.

La enorme importancia de este desarrollo se ve plasmado en la generación en los últimos

años de varias herramientas web diseñadas simplemente para analizar estructuras biológicas o bien para ayudar al usuario inexperto en el diseño de configuraciones de partida o para realizar o analizar simulaciones. Así se pueden citar algunos ejemplos: Parmodel que se utiliza para el modelado comparativo y evaluación de estructuras de proteínas [Brandao, Eberhart, Freitas, Camera, Canduri, and Filgueira, 2004]. MDWeb y MDMoby son plataformas integradas basadas en la web para simulaciones de DM, éstas ayudan a generar los archivos de inicio y los envían para realizar la simulación ya sea mediante Amber, NAMD o Gromacs. [Hospital, Andrio, Fenollosa, Cicin-Sain, Orozco, and Gelpí, 2012]. OGBMX que es un generador de topologías basado en la WEB para Gromacs utilizando el Campo de Fuerzas Universal. [Garberoglio, 2012] MemBuilder que es un generador de estructuras de partida para simular membranas en Gromacs.[Ghahremanpour, Arab, Aghazadeh, Zhang, and van der Spoel, 2014] Vienna-PTM que es un generador de modificaciones post-traduccionales en proteínas y su posterior minimización para ser simuladas en Gromacs. [Margreitter, Petrov, and Zagrovic, 2013] ST-Analyzer: que es una interfase web para el análisis de las trayectorias obtenidas por simulación, en donde se han implementado varios módulos de análisis estándar para membranas como perfiles de densidad, área por lípido, grosor de la membrana, etc. [Jeong, Jo, Wu, Qi, Monje-Galvan, Yeom, Gorenstein, Chen, Klauda, and Im, 2014]. Esta “explosión” en el diseño de herramientas web, muchas de ellas realizadas exclusivamente para Gromacs, orientadas a usuarios no expertos, capaces de centralizar y estandarizar la información para poder realizar simulaciones, se debe a la gran demanda que genera el uso de la DM, es por eso que este trabajo se centra en el desarrollo de una herramienta web que permita realizar análisis no convencionales y personalizados sobre trayectorias realizadas en membranas y membranas expuestas a campos eléctricos externos.

Se debe tener en cuenta entonces, que existen campos de investigación en los cuales el tamaño de la información con la cual se necesita trabajar es de gran escala. Para lograr esto, se utilizan bases de datos en cuales se puede guardar toda la información necesaria para trabajar, pudiendo consultarla fácilmente.

Si bien los programas existentes de administración de bases de datos como pgAdmin

[Postgres core team, 2002-2010] permiten realizar todo tipo de consultas y cálculos sobre la información, un sistema hecho a medida es más flexible y se puede adaptar a las necesidades de cada usuario en particular. Asimismo, contar con un sistema personalizado permite que un usuario sin conocimientos sobre bases de datos pueda utilizarlo con facilidad.

Por otra parte, no todos los sistemas están preparados para procesar un gran caudal de datos. Se debe tener en cuenta la velocidad de procesamiento ya que no sería de utilidad un programa que demorara una cantidad considerable de tiempo. Para saber si es posible utilizar el sistema es necesario conocer los requerimientos mínimos de hardware para que pueda funcionar. A la hora de crear una aplicación nueva, se deben contemplar estas particularidades para ver si es factible resolver un problema de la vida cotidiana en un programa de software.

Al momento de comenzar un proyecto se deben tomar una serie de decisiones respecto a su implementación. Tales decisiones son:

1. qué tipo de interfaz gráfica de usuario o GUI (por sus siglas en inglés *Graphical User Interface*) es la más adecuada para el tipo de problema.
2. el lenguaje de programación que se utilizará
3. el motor de base de datos
4. el *framework* que facilitará la programación de la aplicación entre otras cosas.

Una GUI es un programa que actúa como intermediario entre el usuario y el sistema para el cual fue definida. Con formas gráficas (botones, íconos, etc.) e imágenes, facilita al usuario el uso de las diferentes funcionalidades que proporciona la aplicación para la cual actúa como interfaz. Algunos de los elementos que suelen utilizarse en las interfaces son formularios con diferentes tipos de campos para ingresar datos, por ejemplo cuadros de texto, botones, casilleros de verificación, barras de estado, ventanas, vínculos, cuadros de diálogo, etc.

La GUI puede ser aplicada a un sistema de uso local o bien en una página web. A esta última se la llama Web Based GUI. Una Web Based GUI resulta adecuada para utilizarse

en ocasiones en las que es necesario o conveniente compartir la aplicación (y los datos guardados en ella) con otro grupo de personas. Si bien una aplicación local permite compartir información con otras personas (se puede descargar una base de datos y compartirla por un medio de almacenamiento), el hecho de que la información esté disponible constantemente facilita esta condición. Además, los sistemas web guardan y procesan la información en un servidor, por lo que disminuye notablemente el costo de almacenamiento y procesamiento de parte de la computadora del usuario. Dado que en ocasiones un usuario promedio no cuenta con una computadora con los requerimientos mínimos necesarios, resulta una muy buena opción las Web Based GUI.

Una vez elegida el tipo de GUI, se debe seleccionar qué *framework* se utilizará. Un *framework* o marco de trabajo es un conjunto de modelos y métodos que facilitan la programación de distintas aplicaciones. Existen *frameworks* para diferentes tipos de aplicaciones. Algunos tienen orientación web, otros se concentran en una funcionalidad en particular o cálculos específicos (puede ser un *framework* para realizar todo tipo de estadísticas matemáticas), etc. Un *framework* con orientación web le otorga al programador las herramientas básicas para que pueda poner en funcionamiento la aplicación web. Si bien un programador podría generar el código para su aplicación desde el inicio, utilizar un esquema ya corregido, y con diferentes herramientas disponibles (que lleva un tiempo considerado de desarrollo y para el cual trabajan muchas personas en el proyecto), permite que utilice este tiempo en mejorar las funcionalidades propias de su aplicación.

Un ejemplo de *framework* con orientación web es Django, que tiene como base el lenguaje Python. Django permite dividir el código de nuestra aplicación en MVC (modelo, vista, controlador), así un MVC es un patrón de ingeniería de software (un modo de estructurar el código) que separa los controladores (las funciones propias de la aplicación), de las vistas (la GUI), y los modelos de datos (los objetos de estudio del problema para el cual se lleva a cabo la aplicación).

Luego de elegir el *framework* a utilizar, en algunas ocasiones es posible que no se disponga de diferentes motores de bases de datos para seleccionar ya que algunos *frameworks* se limitan a trabajar con uno en particular. En el caso de Django permite utilizar

PostgreSQL, SQLite, MySQL y Oracle. De acuerdo a la documentación de Django, recomiendan utilizar PostgreSQL pero sólo por una cuestión de preferencia personal.

2. OBJETIVOS

El objetivo es generar una aplicación web con interfaz gráfica para simulaciones de DM obtenidas por Gromacs que sea fácil de utilizar aún por usuarios no expertos en programación. También se espera que el número de partículas no sea una limitación utilizando simulaciones de biomoléculas. En particular se busca desarrollar una aplicación de post-análisis orientada a la implementación de algoritmos personalizados que no son los que brinda Gromacs como herramientas post análisis estándar. Se busca incorporar análisis de estadística descriptiva o la aplicación de algoritmos personalizados sobre los resultados de las trayectorias de las moléculas de interés; generando representaciones gráficas de dichos análisis para facilitar la inmediata visualización de resultados, evitando el proceso de descarga de archivos de datos y la utilización de otra aplicación ya sea para gráficos como para visualizaciones.

3. DETALLES TÉCNICOS DE LA SIMULACIÓN COMPUTACIONAL

3.1. Simulaciones en membranas

A continuación se detalla la generación de las simulaciones de membrana: Se utilizó Gromacs que es una herramienta de código abierto que permite la simulación de DM para sistemas de entre cientos y miles de partículas. Se lo reconoce por su alta capacidad de procesamiento, velocidad y por la gran cantidad de datos que pueden extraerse. Se la considera una de las más veloces debido al uso de precisión simple para el punto flotante aritmético y porque las rutinas más demandantes han sido escritas en *assembly*. [Garberoglio, 2012] [Pronk, Páll, Schulz, Larsson, Bjelkmar, Apostolov, Shirts, Smith, Kasson, van der Spoel, Hess, and Lindah, 2013]

Para este trabajo se utilizaron bicapas lipídicas de dioleoil fosfatidil colina (DOPC) con colesterol (Fig. 3.1 y 3.2) realizadas en Gromacs en un ensamble NPT con una temperatura de 310 K y presión 1 atm, se utilizaron condiciones periódicas de contorno. Se aplicó un campo eléctrico externo de 750 mV/nm con el objetivo de obtener poros (Fig. 3.3 y 3.4) en un tiempo de simulación razonable.

La limitación que presenta Gromacs es que no permite de una manera sencilla modificar sus módulos de análisis de las trayectorias, sino accediendo al código fuente. Esto implica que se necesitan de conocimientos de programación, uso de comandos por consola, así como un enorme cuidado a la hora de hacer modificaciones sobre el mismo. Por otro lado no permite ejecutar de manera amigable y a través de ventanas sus módulos estándar de análisis, sino que el acceso debe ser mediante comandos por consola o a través de herramientas de ejecución desarrollados recientemente que sólo permiten un acceso muy limitado a análisis de rutina. Es por eso que se planteó el presente desarrollo de la herramienta web.

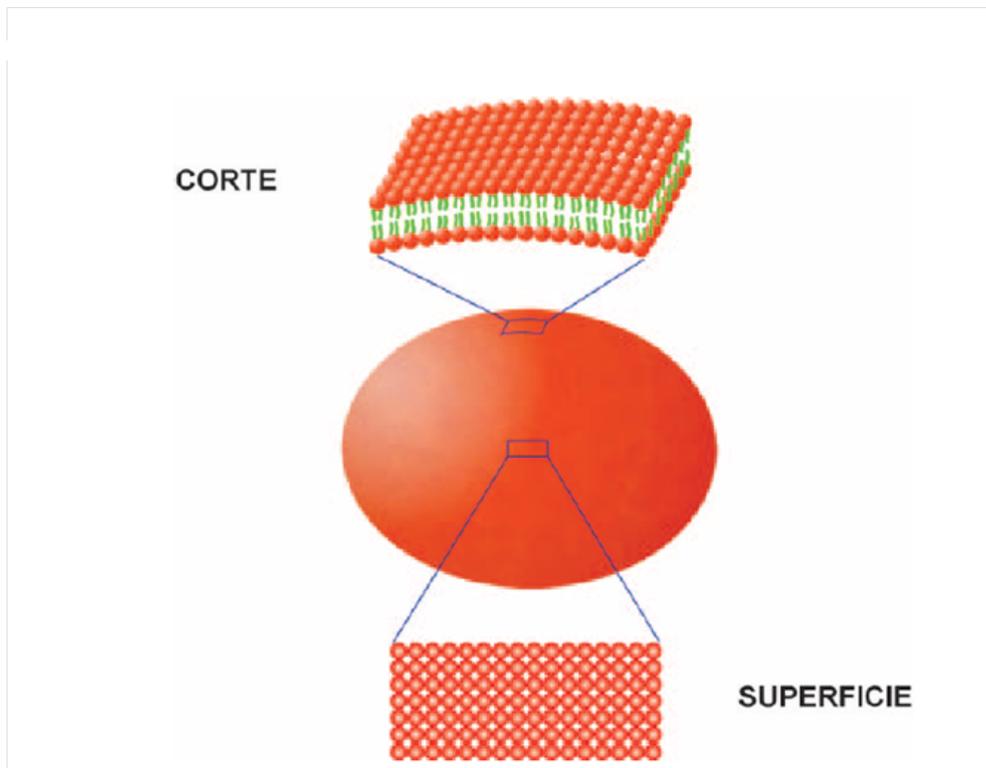


Fig. 3.1: Se muestra la caja a la que se hace referencia durante este trabajo.

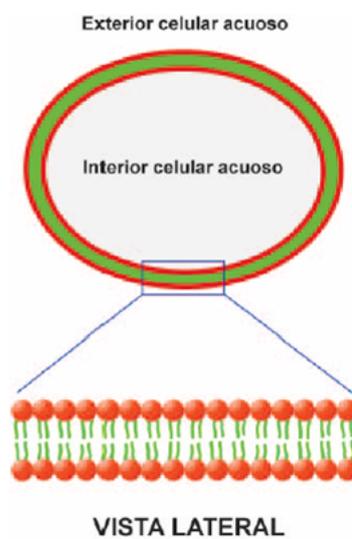


Fig. 3.2: Se muestra la estructura celular

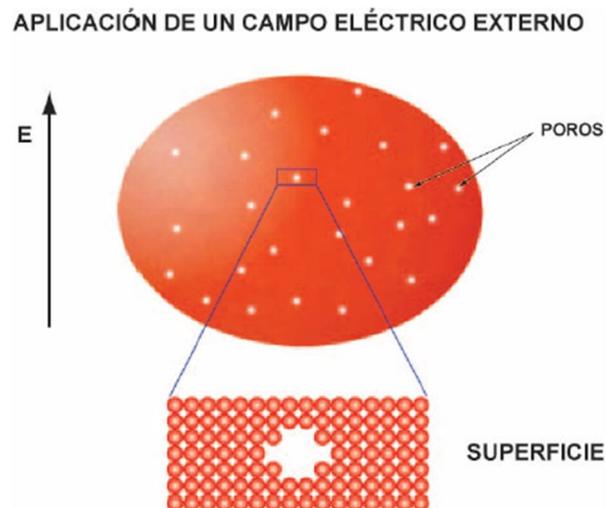


Fig. 3.3: Poro en la célula al aplicar un campo eléctrico

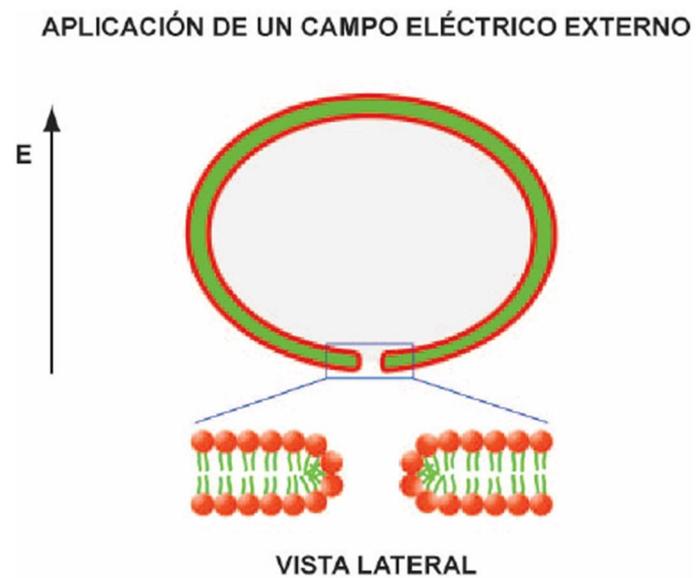


Fig. 3.4: Poro en la célula al aplicar un campo eléctrico

4. DESARROLLO DE LA HERRAMIENTA WEB PARA POST-ANÁLISIS DE SIMULACIONES DE DM

Para el desarrollo de la herramienta y puesta a punto de la misma se utilizó una trayectoria de una bicapa lipídica en presencia de un campo eléctrico externo cuya apertura máxima de poro se produce a los 9ns.

4.1. Descripción de la herramienta y el modelo de datos

Descripción de la herramienta

El sitio web [Borgna, Fernández, and Risk, 2013] se desarrolló con el *framework* Django, en su versión 1.4 [Django core team, 2012], el cual utiliza Python 2.6 [van Rossum, 1997], [Watters, van Rossum, and Ahlstrom, 1996]. La base de datos se implementó en PostgreSQL 8.4 [PostgreSQL Global Development Team, 1996-2015]. Para facilitar los cambios en la base de datos según el modelo de datos se utilizó South [Torchbox, 2008].

La visualización de la información se programó en Python con la biblioteca matplotlib [Hunter, 2007]. Los requerimientos del sistema incluyen un mínimo de 4 GB de memoria RAM (testado en una máquina virtual dentro de una PC Intel i7 con 8 GB de RAM, designando a la máquina virtual 6 GB).

La Fig. 4.1 muestra el esquema de la herramienta web desarrollada. El primer paso es generar o crear una nueva trayectoria presionando el botón *Nueva trayectoria*. La creación de una trayectoria se realiza mediante la selección de varios archivos de texto obtenidos de la trayectoria generada por Gromacs cuya extensión es gro, los cuales son procesados y cargados en la base de datos. Cada uno de estos archivos representa un momento de la trayectoria y contiene los siguientes datos: Coordenadas de la caja que se está mostrando en ese momento, el momento de la trayectoria que está representando, cuántos átomos son, y para cada uno de ellos una línea especificada de la siguiente manera:

```
1DOPC C1 1 4.681 3.825 2.430 -0.0688 -0.1726 -0.2565
```

Estos datos son: nombre de la molécula que indica el tipo de molécula, nombre del átomo que luego se separa en tipo y subtipo de átomo, número de átomo, y las seis coordenadas de posición y velocidad.

Luego de tener en el sistema al menos una trayectoria, se habilitan dos nuevas opciones: borrar trayectorias y seleccionar una trayectoria para analizar.

En el primer caso, se borran todos los datos de cualquier archivo gro cargado previamente en la base de datos. Si se selecciona la segunda opción, la aplicación abre otra ventana donde: 1- Se puede seleccionar el análisis de un único átomo perteneciente a esta trayectoria donde el resultado final será un gráfico y la exportación de los datos a un archivo de texto CSV. 2- Se puede seleccionar un conjunto de átomos de la misma donde se permitirá generar diferentes tipos de gráficos y datos estadísticos sobre el conjunto de datos seleccionado. 3- Y en una tercera opción más específica se puede calcular el volumen del poro si lo que se quiere analizar es una membrana expuesta a un campo eléctrico externo para seguir la dinámica de la electroporación.

El objetivo de obtener un archivo CSV es tener un formato universal que pueda ser leído fácilmente por otro software, por ejemplo, por el lenguaje R para un posterior procesamiento estadístico [R Development Core Team, 2010].

Del archivo de base gro elegido, se selecciona: el nombre de la molécula (POPC, DOPC, CHOL, SOL, etc. ¹), el número, el tipo de átomo perteneciente a cada molécula (C, N, O, H, etc. ²), el subtipo de átomo (1, 2, W1, etc.), el tiempo de la muestra que representa el momento en la trayectoria, las coordenadas espaciales de cada átomo descriptos como X, Y, Z, las velocidades V_x , V_y , V_z y finalmente el tamaño de la caja de simulación para cada tiempo determinado.

Modelo de datos El *framework* Django permite definir diferentes clases o modelos de datos, vinculadas con la base Postgresql. Posee funciones que facilitan el trabajo con los datos, y optimizan las consultas a la base. Para realizar esta aplicación se dividió el problema en los siguientes modelos:

¹ DOPC: Dioleoylphosphatidyl-Choline, POPC: 1-Palmitoyl-2-oleoylphosphatidylcholine. CHOL: colesterol, SOL: solvente.

² C: carbono, N: nitrógeno, O: oxígeno, H: hidrógeno

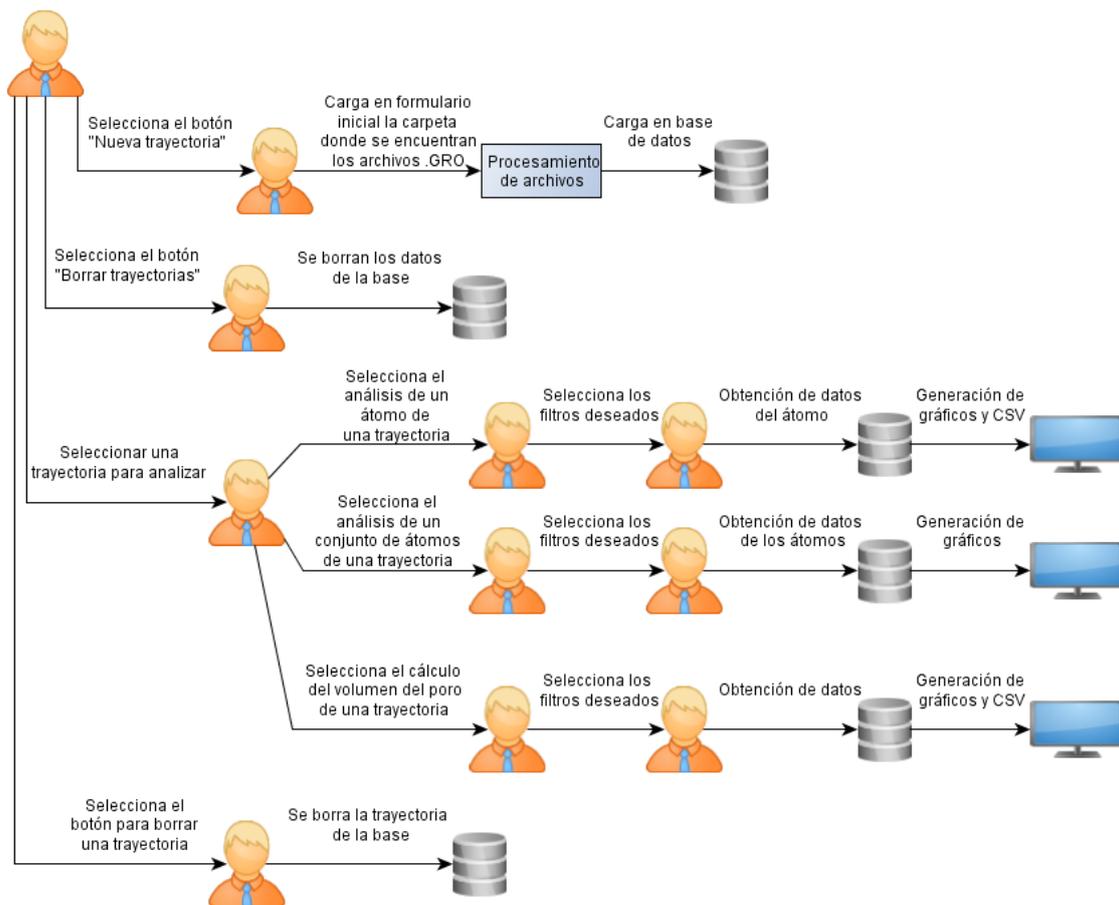


Fig. 4.1: Esquema de la herramienta web.

- **Trayectoria:** Se define con un nombre de longitud máxima 200 caracteres. También cuenta con un identificador numérico que se establece automáticamente al momento de crearla.
- **TipoAtomo:** Como su nombre lo indica, establece un tipo de átomo. Se define con un nombre de longitud máxima 200 caracteres. También cuenta con un identificador numérico que se establece automáticamente al momento de crearlo. Los tipos de átomos se crean al momento de recorrer los archivos de la trayectoria.
- **SubtipoAtomo:** Se define con un nombre de longitud máxima 200 caracteres. También cuenta con un identificador numérico que se establece automáticamente al momento de crearlo. Los subtipos de átomos se crean al momento de recorrer los archivos de la trayectoria.
- **Molécula:** Se define con un nombre de longitud máxima 200 caracteres y un número de molécula dado por los archivos de la trayectoria que se está creando. También cuenta con un identificador numérico que se establece automáticamente al momento de crearla. Tiene asociada la trayectoria desde la cual se creó.
- **Átomo:** Se define con un número de átomo dado por los archivos de la trayectoria que se está creando. También cuenta con un identificador numérico que se establece automáticamente al momento de crearlo. Tiene asociada la molécula a la cual pertenece, un tipo de átomo y subtipo de átomo.
- **Muestra:** Se define con un número t que indica el tiempo de la trayectoria que representa la muestra. También cuenta con un identificador numérico que se establece automáticamente al momento de crearlo. Tiene asociada la trayectoria de la que forma parte y tres coordenadas: X, Y, Z que definen los límites de la caja que se está analizando.
- **Posición:** Cuenta con un identificador numérico que se establece automáticamente al momento de crearla. Tiene asociado el átomo sobre el cual define la posición, las coordenadas x, y, z que definen la posición, y el tiempo en el cual el átomo toma

estas coordenadas.

- **Velocidad:** Cuenta con un identificador numérico que se establece automáticamente al momento de crearla. Tiene asociado el átomo sobre el cual define la velocidad, las coordenadas x, y, z que definen la velocidad en cada uno de los ejes, y el tiempo en el cual el átomo tiene esa velocidad.

4.2. Instalación de la aplicación

Para instalar la aplicación es necesario instalar los siguientes paquetes en el sistema:

- matplotlib-1.2.1
- Python 2.6.6
- Django 1.4
- South 0.7.6
- psycopg2 2.2.1
- pycrypto 2.1.0

Una vez instalados en el sistema estos paquetes, se deben seguir los siguientes pasos:

- Modificar el archivo de opciones de Python settings.py que se encuentra dentro de la carpeta del proyecto, teniendo en cuenta los siguientes ítems:
 - Configuración de la base de datos
 - Carpeta en la que se desean guardar los archivos de procesamiento generados por la aplicación, llamada carpeta_output
 - Carpeta en la que se desean guardar los gráficos y archivos csv, llamada CARPETA_OUTPUT_GRAFICOS_Y_CSV
 - Cantidad de puntos a promediar al realizar el cálculo de media móvil, cuya variable se llama CANTIDAD_DE_PUNTOS_A_PROMEDIAR_MEDIA_MOVIL

- Rango, que indica el valor por el que se dividirá al tiempo registrado en cada archivo gro (Si se desea convertir la unidad ajusta por un factor de 1000 dependiendo de la unidad de tiempo original que es en picosegundos (ps))
 - DEBUG, que indica si el modo está activado o no. En este caso se debe tener en cuenta que Django guarda en memoria cache todas las consultas realizadas a la base de datos si se tiene el modo DEBUG activado, por lo que el rendimiento disminuye considerablemente.
- Crear la base de datos en postgres con la misma configuración que se indicó en el archivo settings.py.
 - Para crear las tablas en la base de datos, se deben cargar los archivos de migración de South que se encuentran en la carpeta correspondiente *Migrations*, dentro del proyecto ejecutando el comando `python manage.py migrate AnalizadorDeTrayectoriasApp`. Cada vez que se realice un cambio al archivo models.py (módulo de software en Python que define los modelos de datos y sus funciones asociadas), se deberá ejecutar el comando `python manage.py schemamigration -auto AnalizadorDeTrayectoriasApp`. Este comando realizará automáticamente las modificaciones correspondientes en la base de datos.

4.3. Ejecución de la aplicación

Para ingresar a la aplicación se puede optar por configurar un virtualhost de apache, o bien utilizar el servidor que ofrece Django. En el primer caso, se configura la dirección a la que se desea acceder en el virtual host. En el segundo caso, se debe ejecutar el comando “python manage.py runserver” y luego en la consola aparecerá la dirección IP mediante la cual se debe ingresar en la web a la aplicación.

4.4. Descripción de la interfaz

Página principal



Fig. 4.2: Página principal.

En la página principal de la aplicación (Fig. 4.2) se muestra la lista de trayectorias creadas en el sistema, con dos opciones posibles: borrar la trayectoria o analizarla.

Además el usuario puede optar por crear una nueva trayectoria o bien borrar toda la información guardada en la base de datos, es decir, todas las trayectorias previamente creadas.

Nueva trayectoria

The screenshot shows the 'Formulario para crear una trayectoria' (Form to create a trajectory). It has the same header and navigation bar as the main page. The form contains three input fields: 'Carpeta' with the value '/home/trayectoria1/', 'Nombre trayectoria' with the value 'Trayectoria 1', and 'Incluir hidrógenos de aguas' with a checked checkbox. Below the fields is a blue 'Crear trayectoria' button. At the bottom right, there is a footer that reads 'Facultad de Ciencias Exactas y Naturales - U.B.A.'.

Fig. 4.3: Formulario para crear una trayectoria.

Para crear una nueva trayectoria (Fig. 4.3) se deben procesar los archivos obtenidos a partir de Gromacs para luego insertar la información en la base de datos. Al llegar a esta opción se deriva al usuario a un campo de texto en el cual debe ingresar la dirección de la

carpeta en la que se encuentran los archivos con extensión gro que se desea analizar. La aplicación toma de la carpeta seleccionada sólo los archivos con dicha extensión, los procesa e inserta los datos en la base. También se pide el nombre que se le desea asignar a esta trayectoria para poder referenciarla en la aplicación y decidir si se incluirá la información sobre los hidrógenos de aguas o no. Esta última opción se agregó como parte de una de las optimizaciones que se realizaron a la aplicación (ver sección *Optimizaciones realizadas a la aplicación*). Mientras se cargan los archivos, se muestra un avance del procesamiento en la consola. Se indica por separado el procesamiento de los archivos iniciales, donde se dividen las columnas y se analiza la información en el formato de archivo de texto. Por otro lado se muestra el avance del ingreso a la base de datos de los datos obtenidos de estos archivos.

Analizar datos de una trayectoria

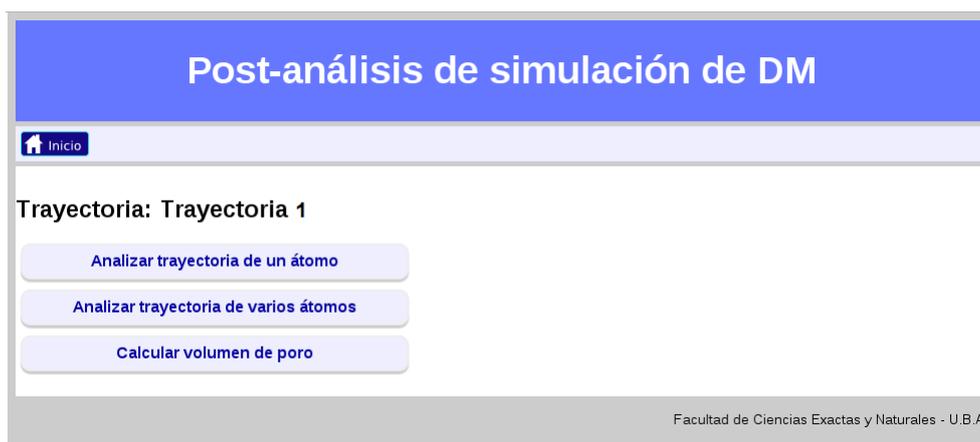


Fig. 4.4: Selección de tipo de análisis de trayectoria.

En esta ventana (Fig. 4.4) abren los tres módulos de análisis posibles donde se puede optar: 1- analizar los datos a un único átomo 2- analizar los datos de un grupo de átomos 3- calcular el volumen del poro formado

Analizar datos de un átomo en una trayectoria

Se muestra un formulario en el que se debe seleccionar de una lista desplegable, el número de átomo que se desea analizar y el período de tiempo de la trayectoria del cual se desea obtener la información (Fig. 4.5). Luego se mostrarán y se guardarán 3 gráficos

Post-análisis de simulación de DM

[Inicio](#)

Trayectoria: Trayectoria 1

Atomo:

Período:

Tiempo inicial: Tiempo final:

Facultad de Ciencias Exactas y Naturales - U.B.A.

Fig. 4.5: Configuración del análisis de trayectoria para un átomo.

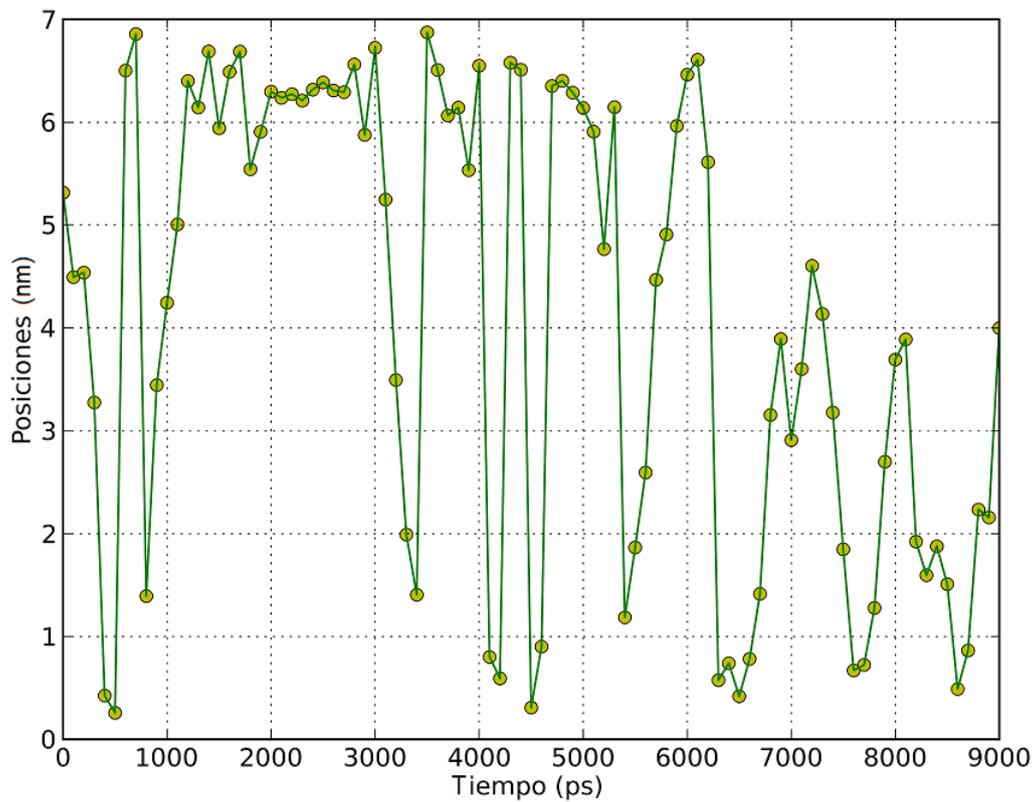


Fig. 4.6: Posiciones en X de un átomo determinado.

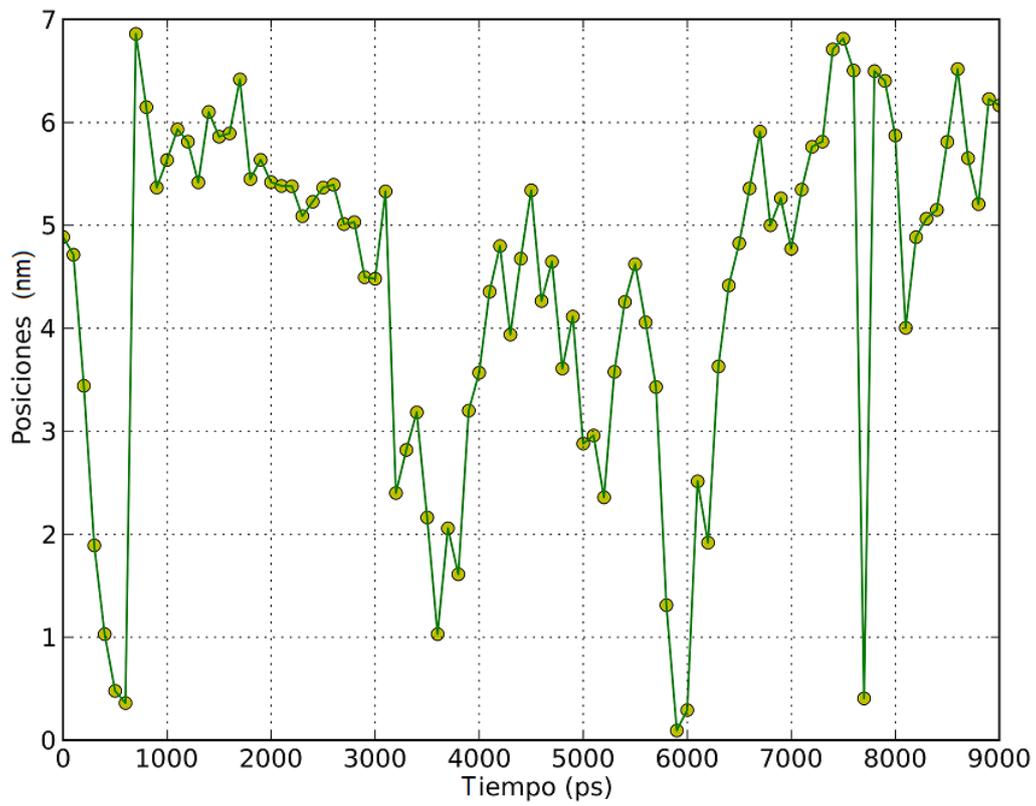


Fig. 4.7: Posiciones en Y de un átomo determinado.

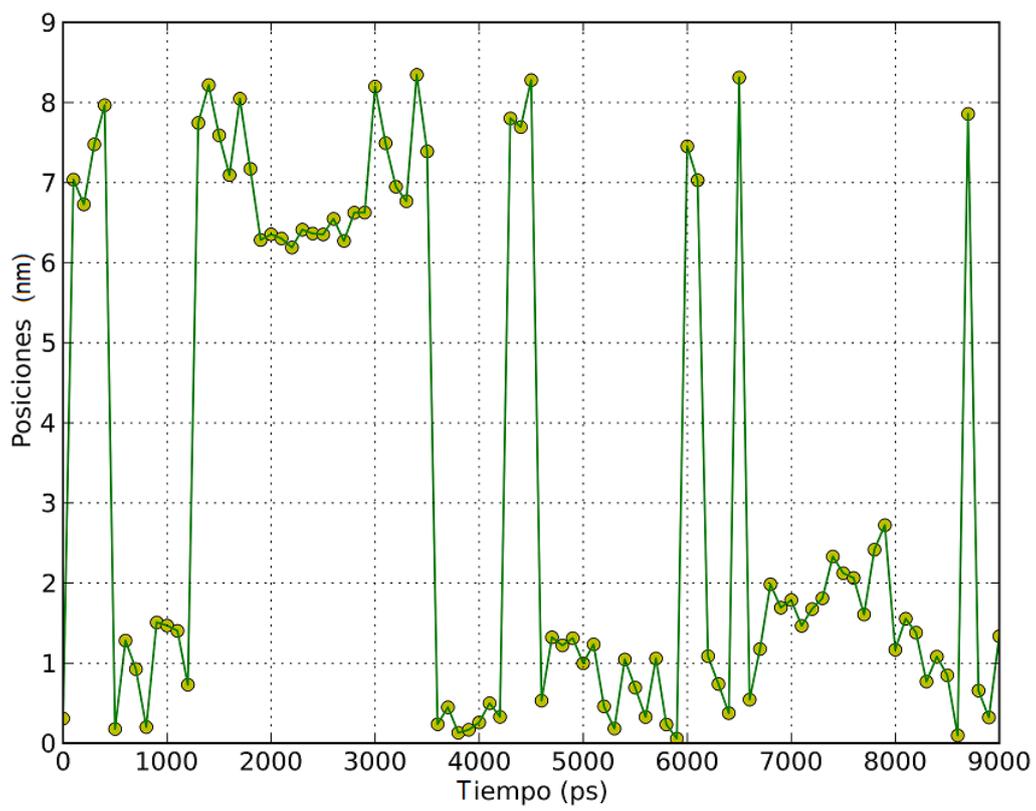


Fig. 4.8: Posiciones en Z de un átomo determinado.

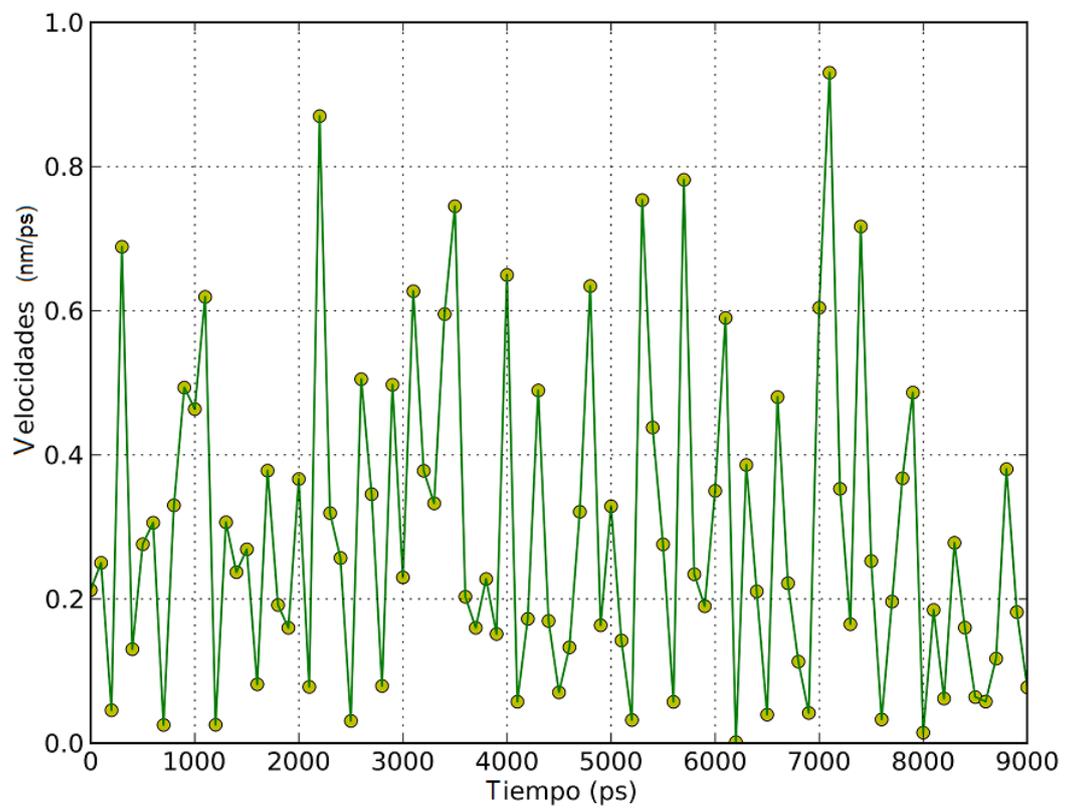


Fig. 4.9: Velocidades en X de un átomo determinado.

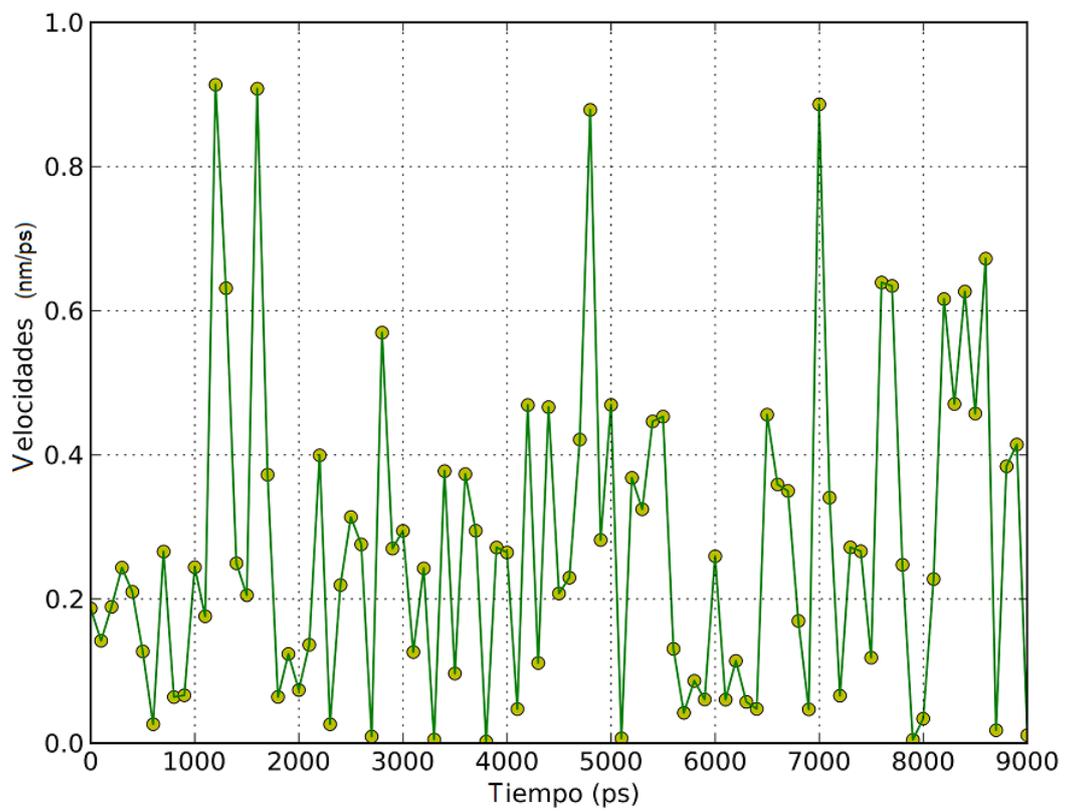


Fig. 4.10: Velocidades en Y de un átomo determinado.

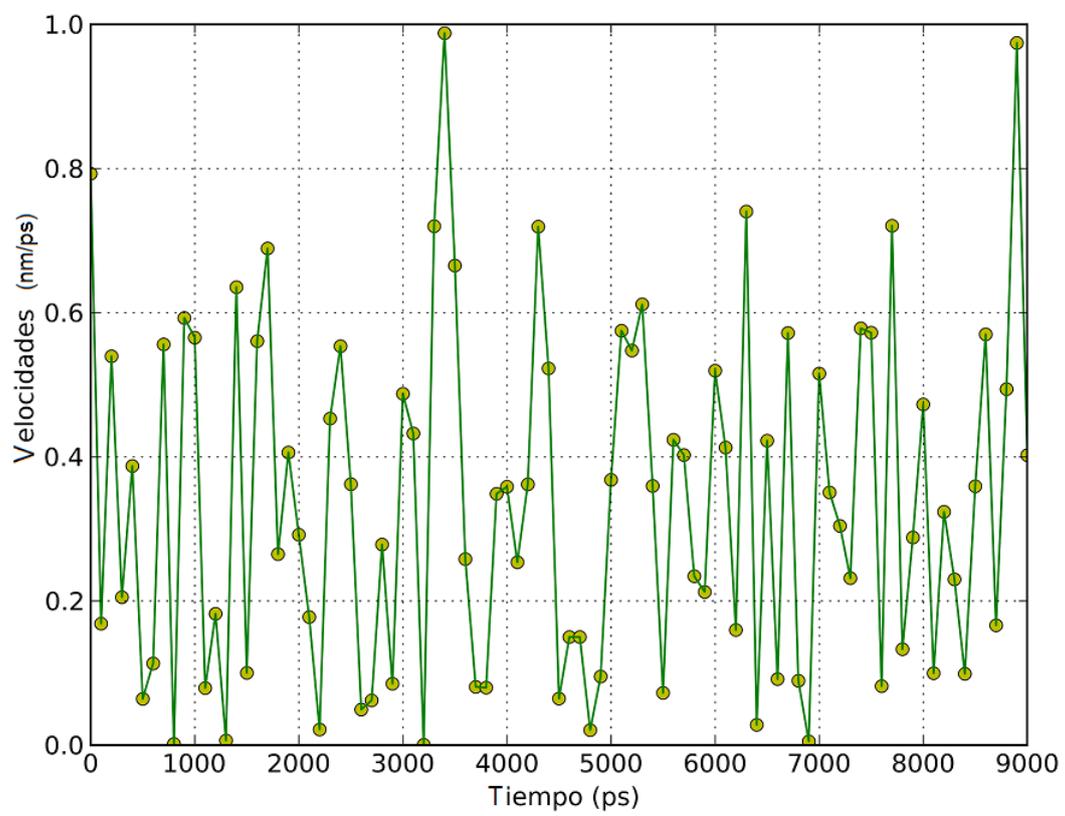


Fig. 4.11: Velocidades en Z de un átomo determinado.

correspondientes a las posiciones X (Fig. 4.6), Y (Fig. 4.7) y Z (Fig. 4.8) y 3 gráficos correspondientes a las velocidades V_x (Fig. 4.9), V_y (Fig. 4.10), V_z (Fig. 4.11) (en caso de que el archivo sea con velocidades) respecto a toda la trayectoria del átomo analizado, indicando las variaciones que realiza el átomo en cuanto a su posición y velocidad.

Analizar datos de más de un átomo en una trayectoria

Fig. 4.12: Configuración del análisis de trayectoria para más de un átomo.

Se muestra un formulario en el que se permite obtener la información de un grupo de átomos en particular, seleccionando los siguientes filtros, los cuales permiten seleccionar una porción del volumen total de simulación y un lapso dentro del tiempo total. (Fig. 4.12). Se pueden explicitar las siguientes opciones:

- Lista de átomos: Números de átomos separados por comas. Se definen los átomos que queremos seleccionar para analizar.
- Tipo átomo: Depende de la trayectoria cargada, se mostrarán los tipos de átomos

disponibles. Ejemplos: C, N, O.

- Subtipo átomo: Al igual que el filtro anterior, depende de la trayectoria cargada. Ejemplos: 4, 6, W.
- Molécula: Se filtran los átomos pertenecientes a una única molécula, seleccionada en esta opción.
- Tipo molécula: También depende de la trayectoria cargada. Ejemplos: SOL, POPC.
- Posición en X, Y, Z (min y max): Se definen los límites de posición de los átomos en cada eje. Se debe tener en cuenta que este límite está dado durante toda la trayectoria.
- Velocidad en X, Y, Z (min y max): Similar al campo anterior, se limitan las velocidades de los átomos a lo largo de toda la trayectoria.
- Tiempo: Limita el período de tiempo de la trayectoria a analizar.
- Realizar gráficos estadísticos: Se muestra en la página un cuadro de doble entrada mostrando el mínimo, máximo, media, desvío estándar y varianza de las velocidades V_x , V_y , V_z . Se grafican estos mismos datos para cada eje y se agregan los valores envolventes. Se generan varios gráficos: uno de los gráficos contiene los datos estadísticos y los datos crudos de todas las velocidades (Fig. 4.13), otro agrega la media móvil central, la cual se calculó promediando 5 puntos (Fig. 4.14), otro agrega la envolvente, es decir los mínimos y máximos valores en cada tiempo (Fig. 4.15), mientras que el último sólo muestra los cálculos estadísticos (Fig. 4.16).

En todos los casos se muestra una simulación donde el poro ya está completamente abierto y se grafica un grupo seleccionado de aguas.

- Realizar gráficos de velocidades: También se ve la información separada en diferentes gráficos según el eje que se está analizando. La información de la velocidad se puede ver por el color de los ítems (que representan los átomos) en el plano 2D. Se generan diferentes gráficos según la velocidad y el plano de la membrana que se está considerando. Tomando los valores de velocidades absolutas (vectoriales) y el eje X (Fig.

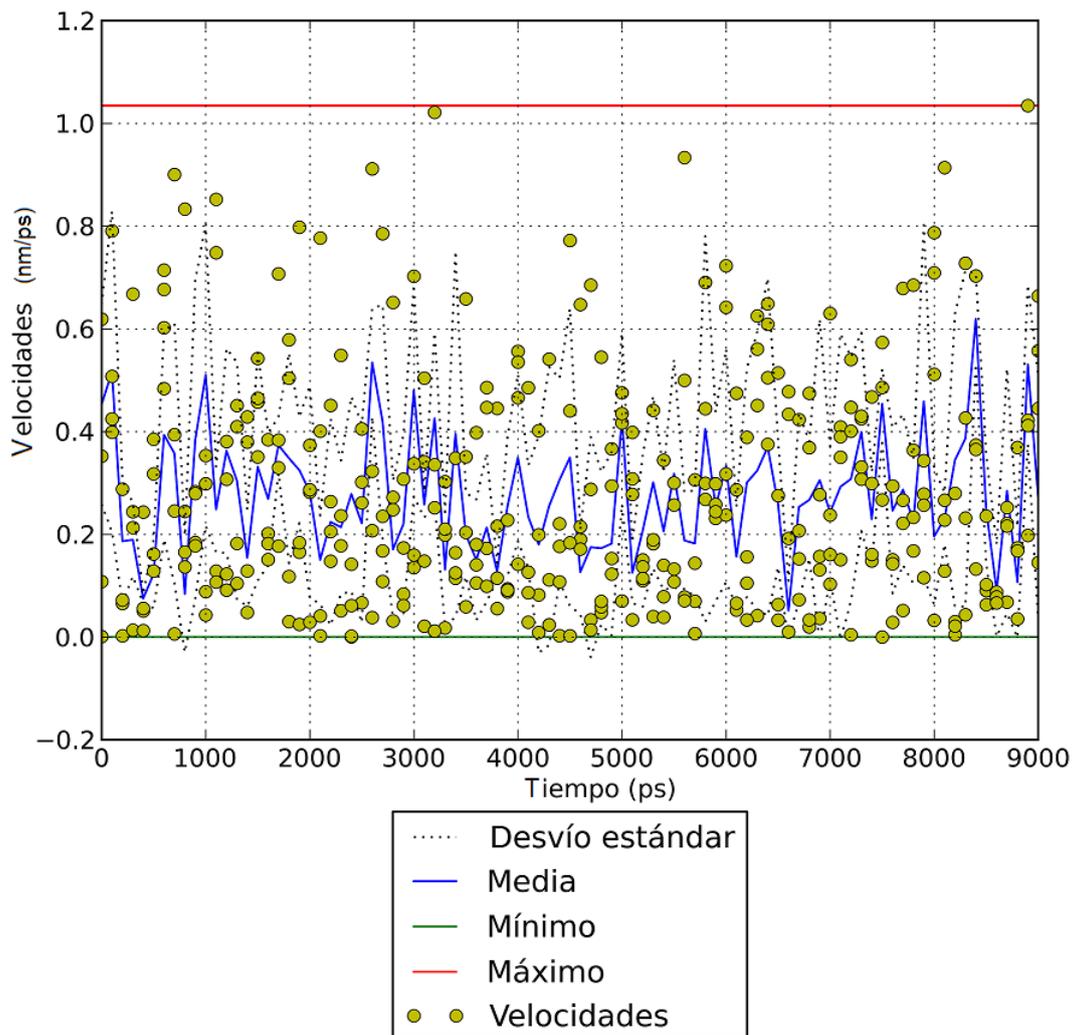


Fig. 4.13: Gráfico con las velocidades en X del grupo de átomos seleccionado y sus valores estadísticos.

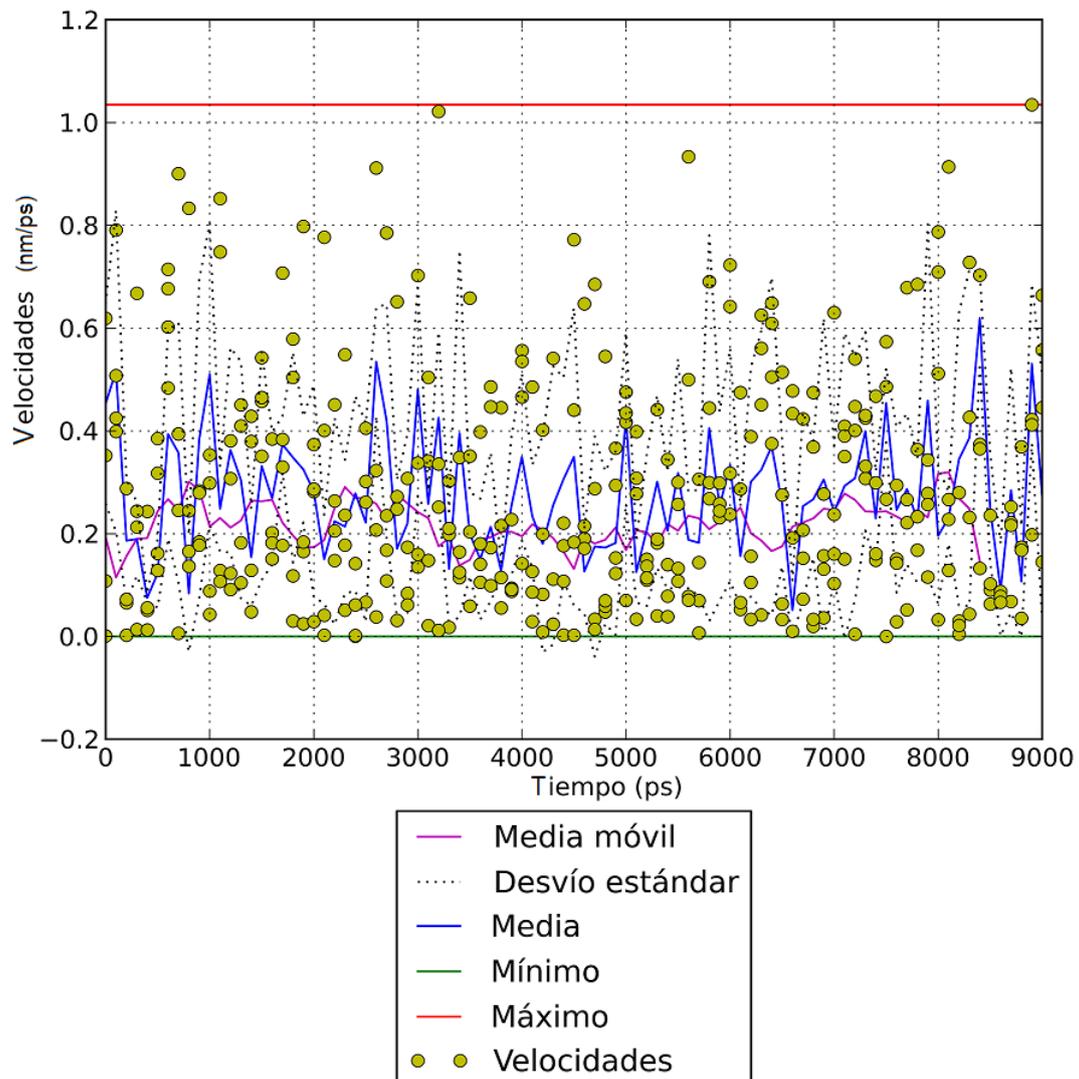


Fig. 4.14: Gráfico con las velocidades en X del grupo de átomos seleccionado y la media móvil.

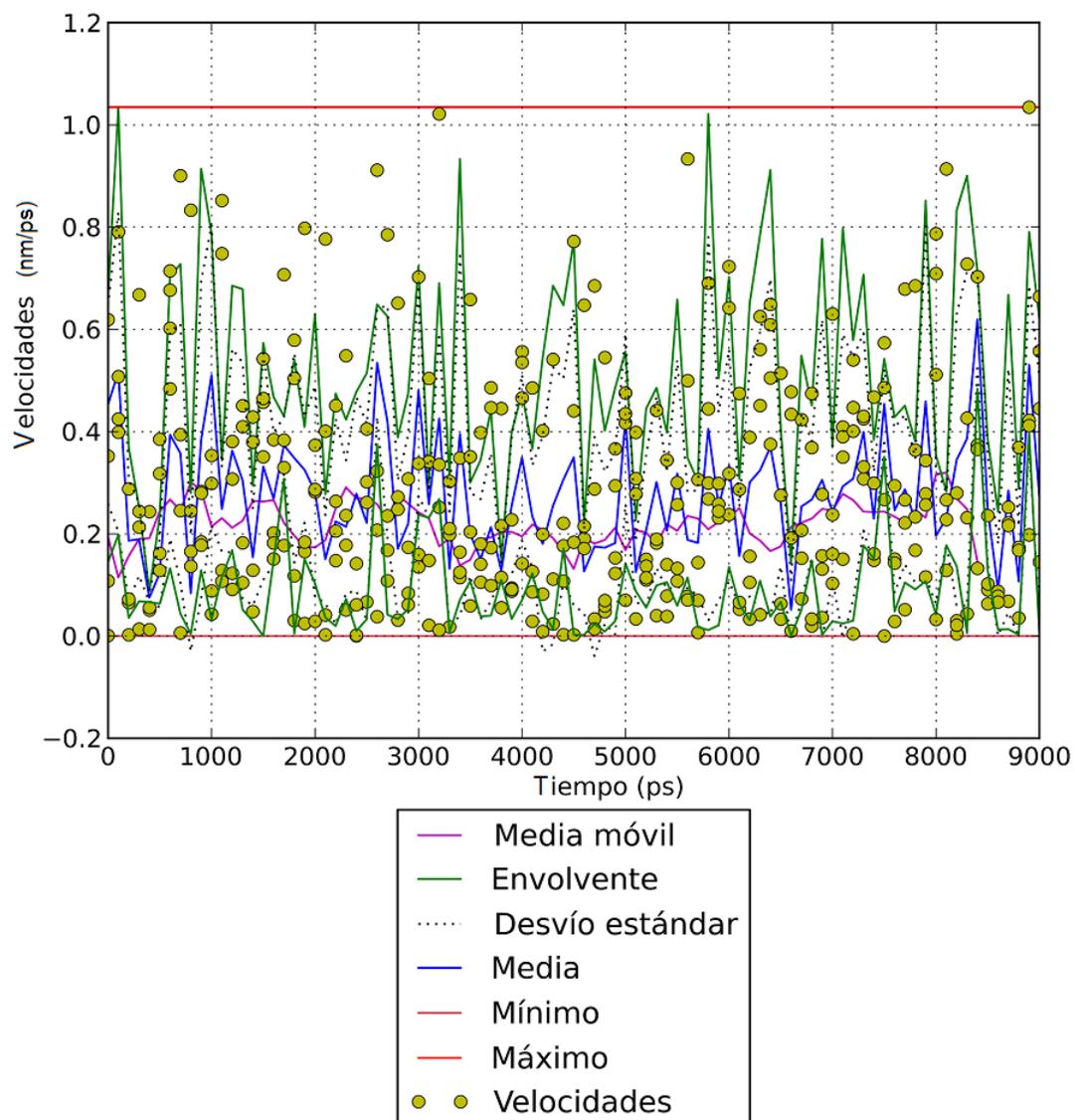


Fig. 4.15: Gráfico con las velocidades en X del grupo de átomos seleccionado y sus curvas envolventes.

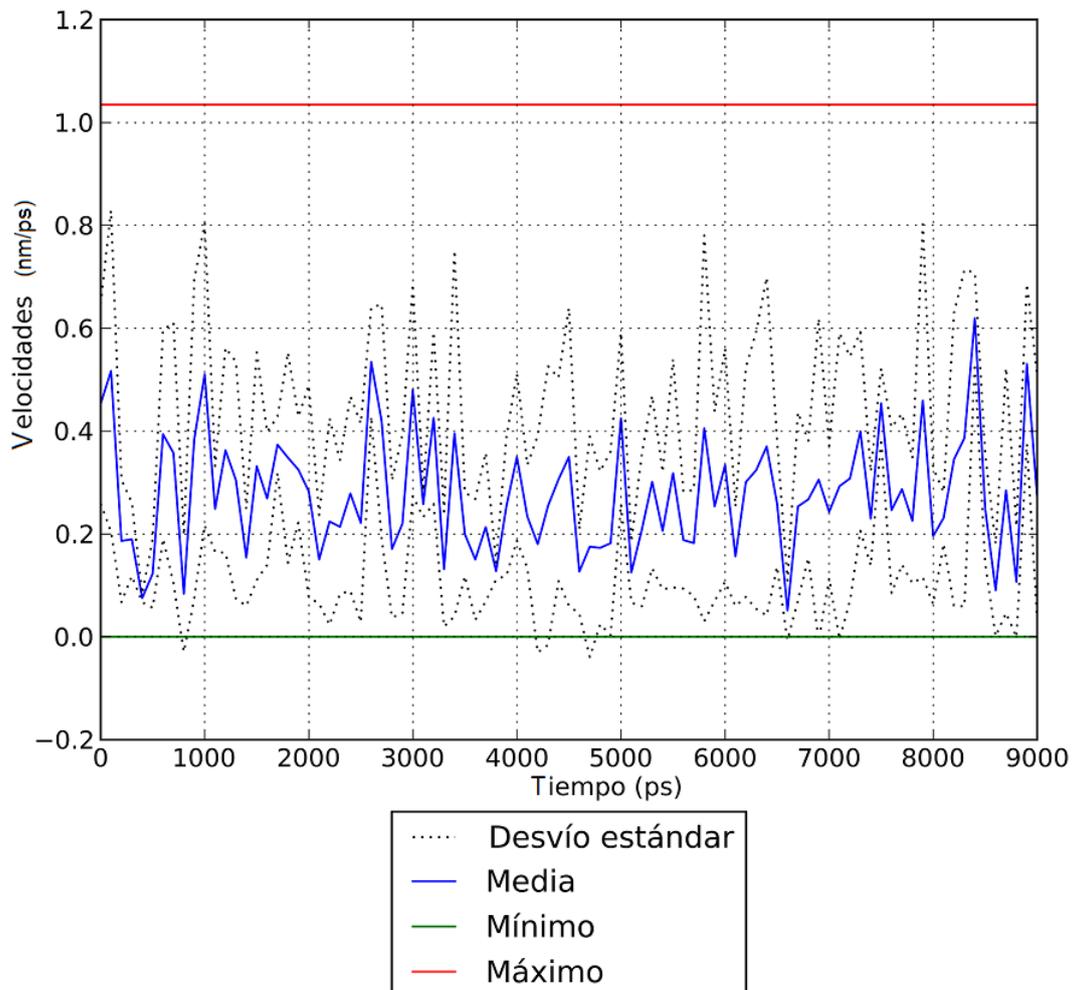


Fig. 4.16: Gráfico con los valores estadísticos de las velocidades en X del grupo de átomos seleccionado.

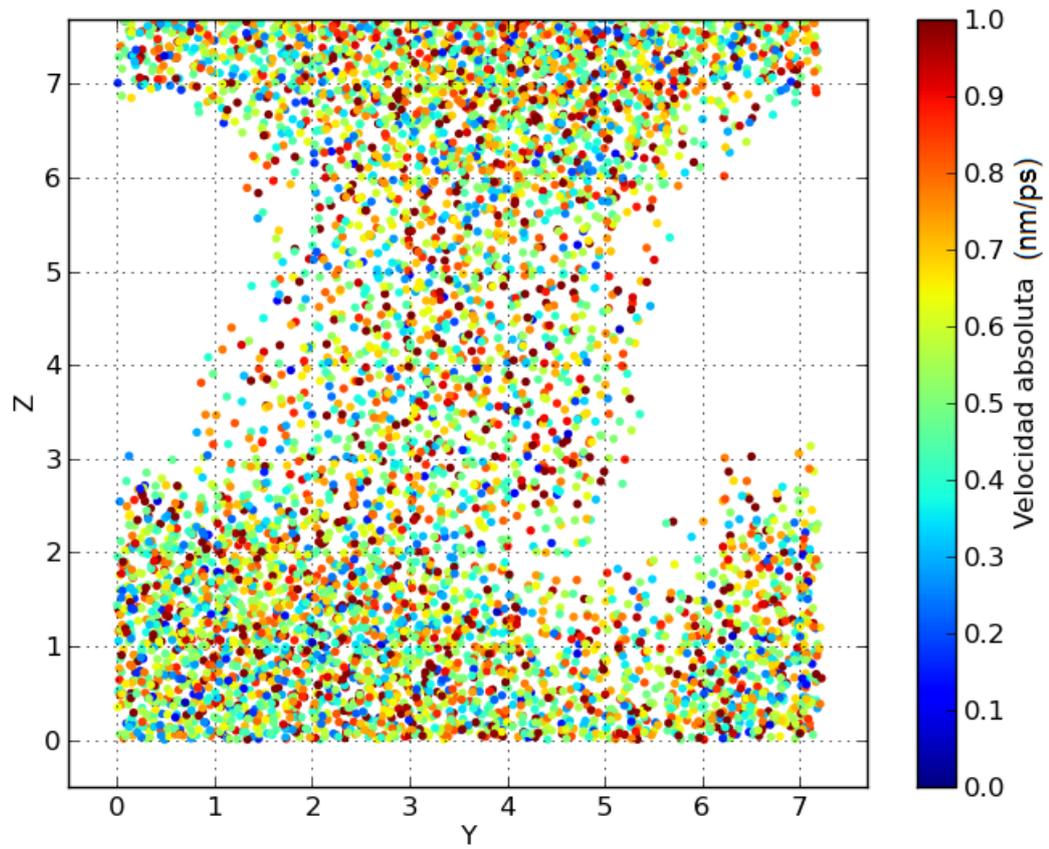


Fig. 4.17: Gráfico de velocidades por posiciones - plano X.

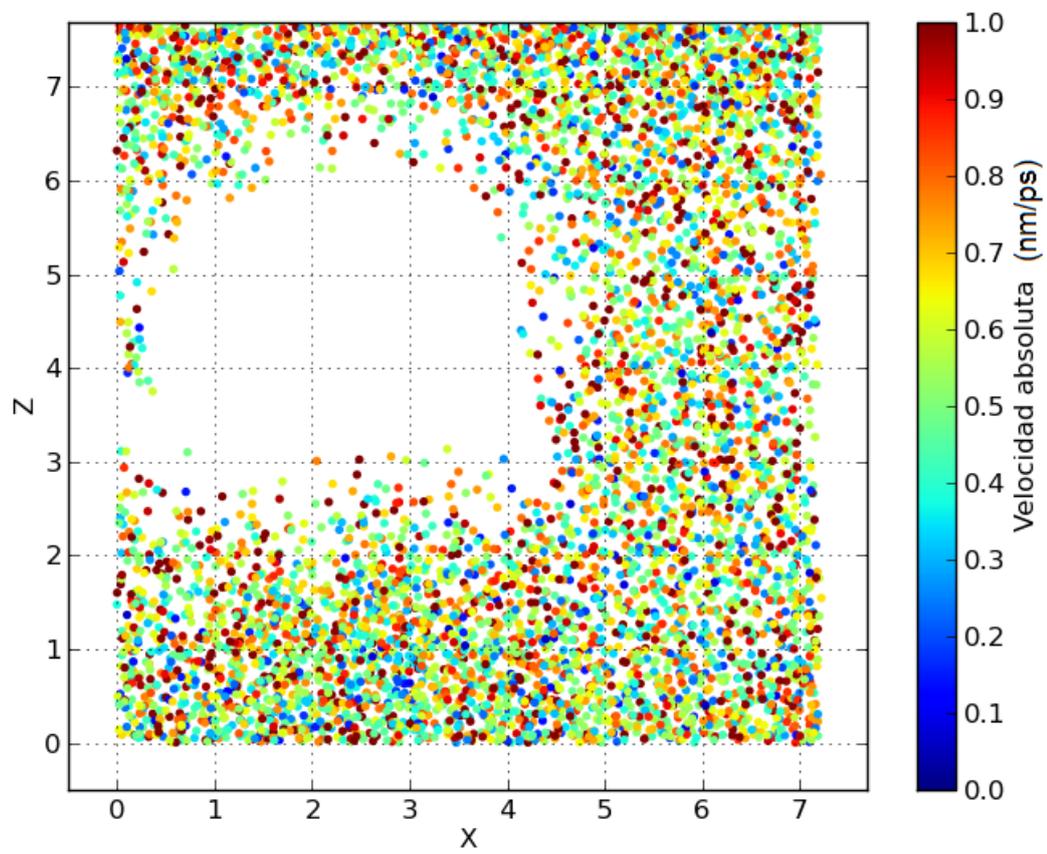


Fig. 4.18: Gráfico de la velocidades por posiciones - plano Y.

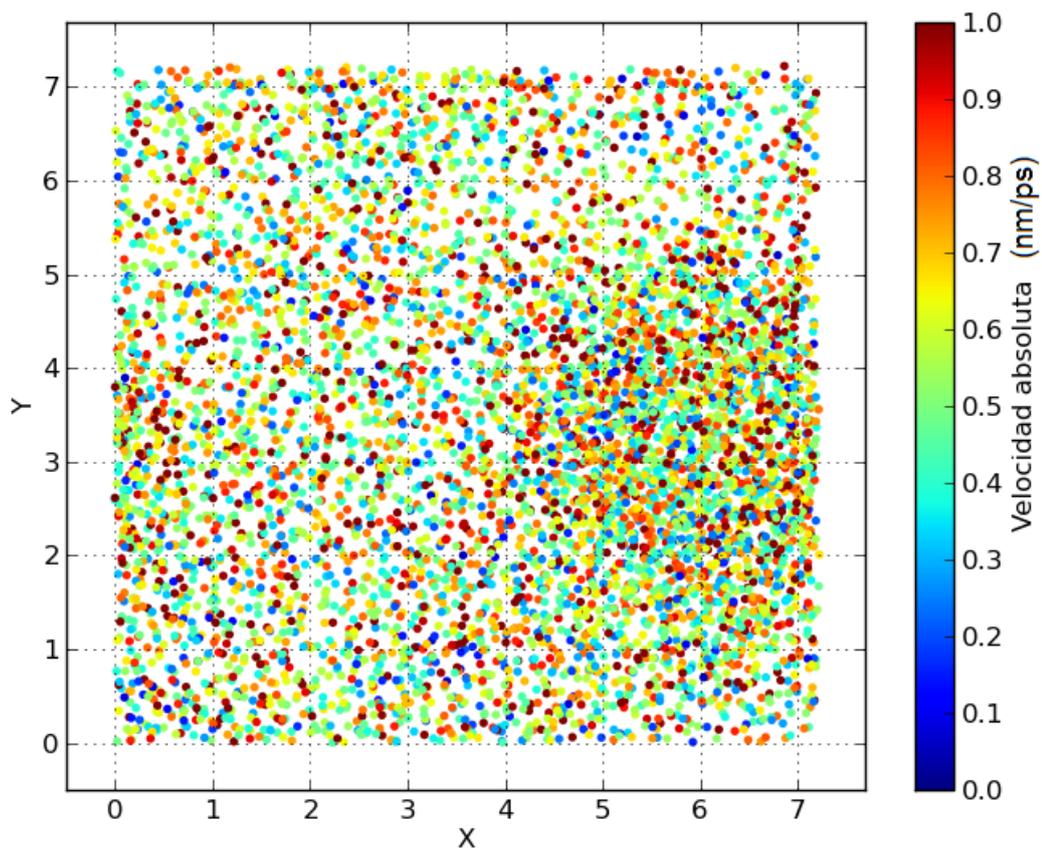


Fig. 4.19: Gráfico de la velocidades por posiciones - plano Z.

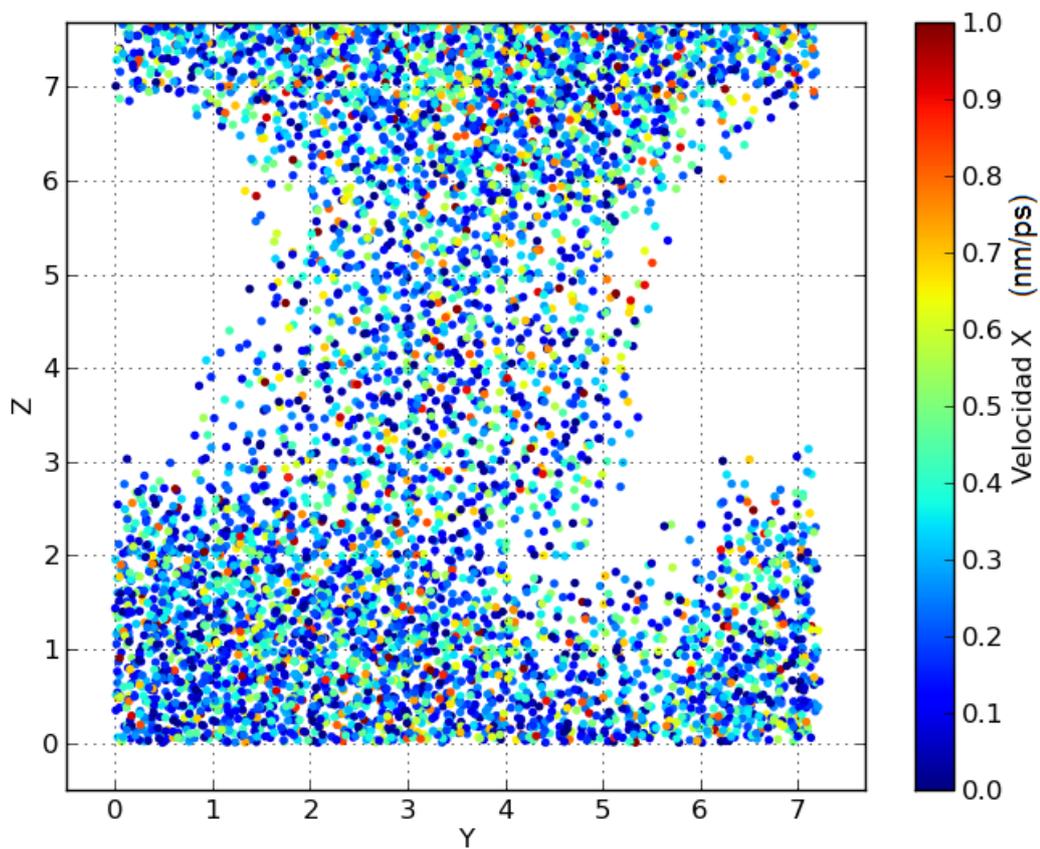


Fig. 4.20: Gráfico de la componente X de las velocidades por posiciones - plano X.

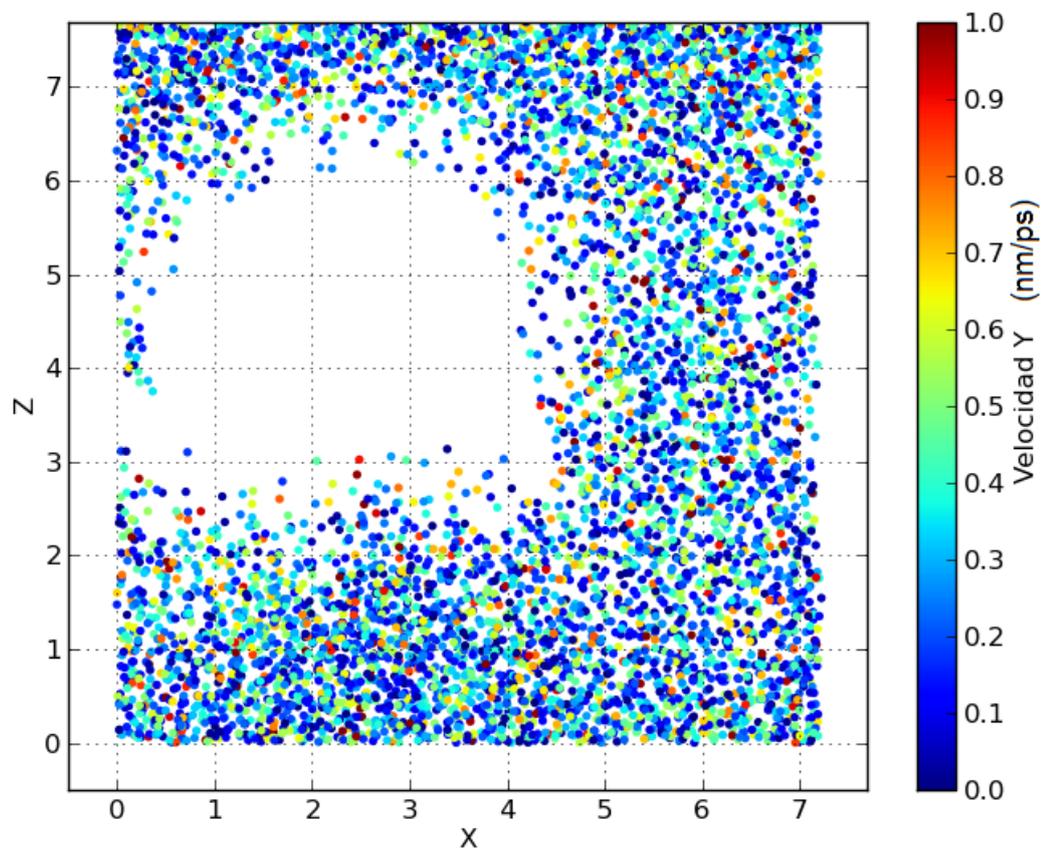


Fig. 4.21: Gráfico de la componente Y de las velocidades por posiciones - plano Y.

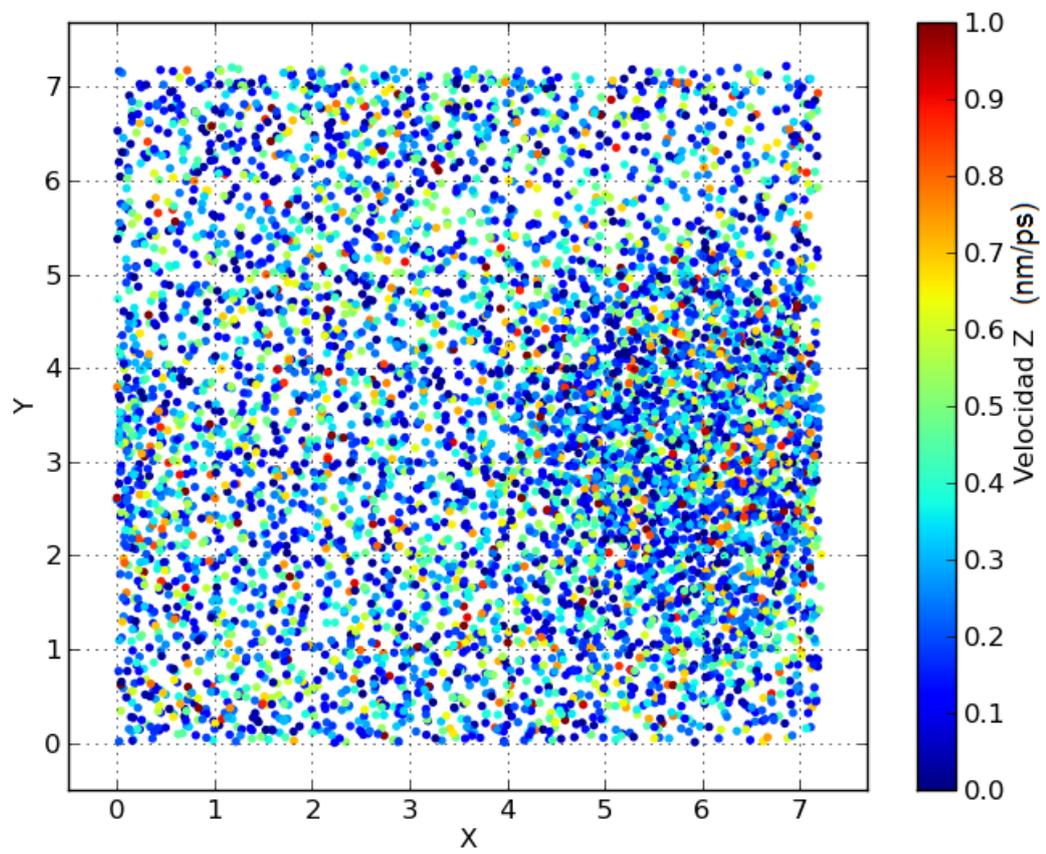


Fig. 4.22: Gráfico de la componente Z de las velocidades por posiciones - plano Z.

4.17), Y (Fig. 4.18) o Z (Fig. 4.19). En estos gráficos se representa la velocidad de cada átomo de oxígeno en cada plano lo que permite revelar la la forma del poro desde distintos ejes en el tiempo de 9 nanosegundos (ns).

Luego de seleccionar las coordenadas y los límites de la caja dentro del total de la simulación, se mostrarán y se guardarán los gráficos correspondientes respecto al período seleccionado de la trayectoria (si no se selecciona ningún período se toma la trayectoria completa).

En todos los casos se muestran gráficos de una simulación donde el poro ya está completamente abierto y se representan los datos de todas las aguas. Aquí hay que tener en cuenta que el resultado de la simulación puede ser un poro no centrado en algún eje (X o Y), lo que se tomará en consideración más adelante para calcular el volumen del poro.

| Velocidades | | | |
|-----------------|-----------------|-----------------|-----------------|
| | X | Y | Z |
| Mínimo | 0,0 | 0,0 | 0,0 |
| Máximo | 1,411 | 1,5957 | 1,7249 |
| Media | 0,30709138377 | 0,312940704817 | 0,310818331717 |
| Desvío estándar | 0,233234907744 | 0,238448763073 | 0,239141720999 |
| Varianza | 0,0543985221906 | 0,0568578126108 | 0,0571887627225 |

Fig. 4.23: Tabla de datos estadísticos.

Calcular volumen de poro

Fig. 4.24: Configuración para obtener el volumen del poro formado.

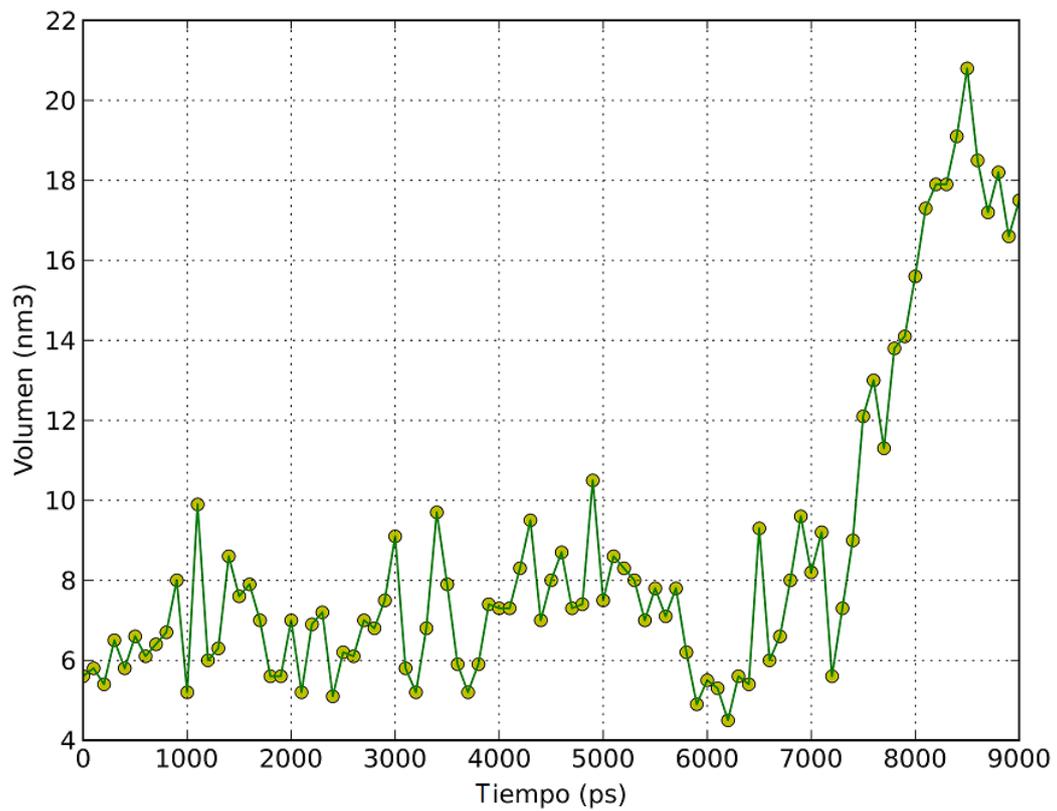


Fig. 4.25: Gráfico que muestra el volumen del poro a lo largo de la trayectoria, con un valor establecido por defecto de los límites dentro de la membrana.

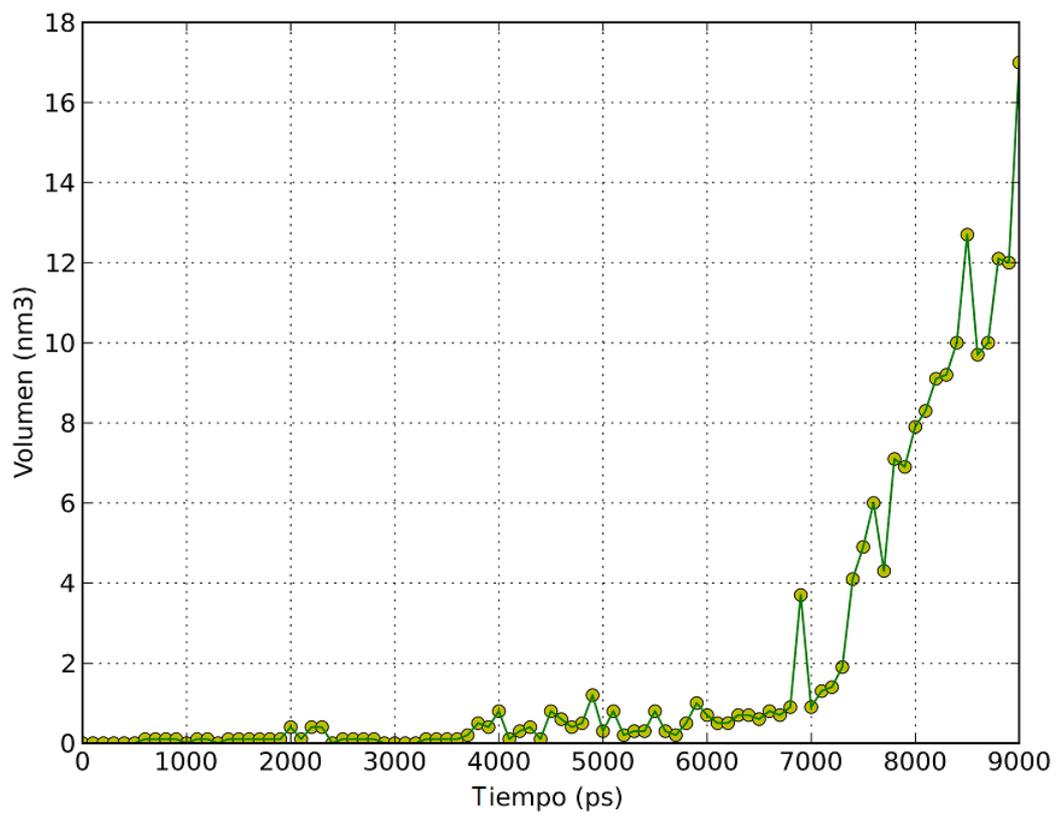


Fig. 4.26: Gráfico que muestra el volumen del poro a lo largo de la trayectoria personalizando los límites dentro de la membrana.

Para hacer este cálculo se utilizó el algoritmo de estimación del volumen mediante elipsoides [Fernández, Risk, Reigada, and Vernier, 2012]. De elegir esta opción se genera una ventana (Fig. 4.24) donde se debe completar el formulario eligiendo los límites del tiempo en los que se quiere obtener información dándose la posibilidad de personalizar los límites dentro la membrana para calcular el volumen. Si no se establece ningún valor para la membrana, se toma un valor por defecto establecido en 2,5 y 4,5 obteniendo un resultado que puede diferir de la realidad (Fig. 4.25). Si se establece el valor conocido de los límites, el resultado será más cercano al verdadero (Fig. 4.26). El por qué de esta situación se detallará en la sección *Decisiones tomadas*.

4.5. Detalles de implementación

Para analizar datos obtenidos de los archivos gro (aproximadamente 1 MB por archivo) fue necesario dividir las columnas de estos archivos para agrupar los átomos de acuerdo a varios criterios (tipo de átomo, subtipo de átomo, número de molécula, etc.).

Luego de generar este nuevo conjunto de archivos procesados, se procede a insertar la información en la base de datos. Este paso demora aproximadamente media hora cada diez archivos. Una vez insertados los datos en la base no es necesario consultar los archivos gro.

Cálculo del volumen del poro

En la opción para el cálculo del volumen se debe seleccionar el intervalo de tiempo deseado. Para cada uno de estos tiempos (o muestras) se aplica el siguiente algoritmo [Fernández, Risk, Reigada, and Vernier, 2012]:

En primer lugar, se ubica al poro en el centro de la caja de simulación (dado que lo más común es que los poros no aparezcan centrados en la caja, como ya se dijo previamente). Luego, se divide la caja en tres partes horizontalmente tomando como división la línea de átomos de fósforo, que representan las cabezas de los lípidos y marcan los límites de la membrana (Fig. 4.27) (cuando ya está creado el poro estos límites delinearán su forma (Fig. 4.28)). Luego tomando estas posiciones más y menos dos desvíos estándar correspondientes a cada una de las dos líneas (arriba y abajo) se generan los límites.

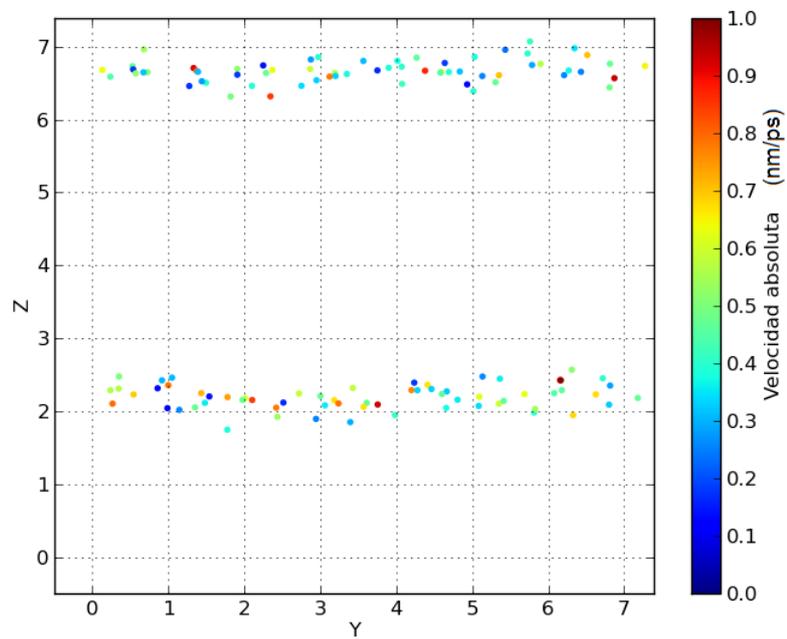


Fig. 4.27: Se visualizan las velocidades por posiciones de los átomos de P8 con los cuales se delimita la membrana, en el tiempo inicial cuando aún no hay un poro formado.

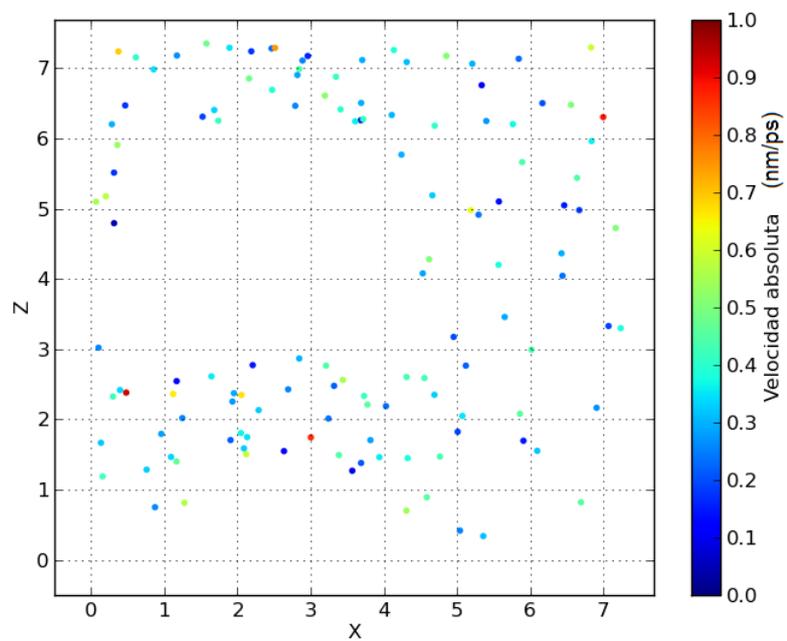


Fig. 4.28: Se visualizan las velocidades por posiciones de los átomos de P8 con los cuales se delimita la membrana, en el tiempo final cuando hay un poro formado.

Una vez que se tienen calculados los límites de la membrana, se procede a dividir este sector interno de la membrana en elipsoides con radios X e Y (Fig. 4.29). Estos elipsoides tienen un espesor de 0,15 nm.

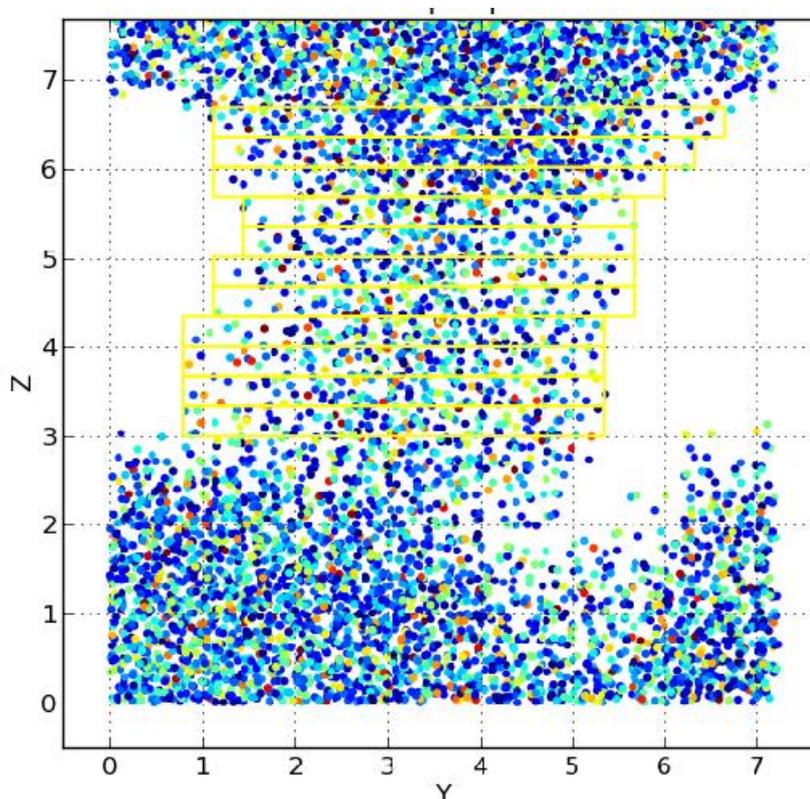


Fig. 4.29: Se divide el poro en elipsoides de igual espesor.

Para finalizar el cálculo del volumen del poro, se suman los volúmenes de todos los elipsoides.

4.6. Optimizaciones realizadas a la aplicación

Una molécula de agua está formada por dos átomos de hidrógeno y uno de oxígeno. Para analizar el movimiento de estas moléculas alcanza con seguir la trayectoria solo del átomo de oxígeno. Teniendo en cuenta esta condición, se da la opción al usuario de incluir o no los hidrógenos de agua al momento de cargar la trayectoria en la base de datos. Si el usuario decide no incluir los hidrógenos, la cantidad de datos se reduce considerablemente, mejorando de esta forma también el rendimiento de la aplicación.

Dado el gran volumen de datos con el que se debe trabajar, se utilizaron generadores de python para iterar sobre conjuntos de elementos. De esta manera no se ocupa la memoria con todos los datos sino sólo con los que se están analizando.

4.7. Decisiones tomadas

Condiciones de contorno

Al realizar una simulación se utilizan condiciones periódicas de contorno, es decir que el sistema se replica en total 26 veces en los 3 ejes de modo que si una molécula sale de la caja por algún extremo las condiciones de contorno lo corrigen compensando el ingreso de la misma por el extremo opuesto.

Como puede observarse en el gráfico de posiciones de X (Fig. 4.6) parecería haber variaciones muy grandes en instantes de la trayectoria. Esto es un artificio de la simulación debido a las condiciones periódicas de contorno, situación que es detectable al analizar los gráficos.

Para subsanar estas condiciones, se decidió transformar estas posiciones en valores positivos. La forma de transformarlas fue sumarle el valor en el eje correspondiente del tamaño de la caja en esa muestra. En el caso de las velocidades también se transformaron a valores positivos, en este caso porque la dirección no era un objeto de interés para esta aplicación. Se tomaron sus valores absolutos.

Posición del poro y la caja Existe una condición que se da al generarse un poro en la membrana luego de aplicar un campo eléctrico donde el poro no necesariamente se genera en el centro de la membrana, sino que se forma en cualquier posición en X o en Y quedando descentrado como se muestra en la siguiente figura (Fig. 4.30). Aquí se representaron todos los átomos de los lípidos a los 9 ns lo cual permite ver el poro como un agujero ya que las aguas no fueron incluidas.

En las siguientes figuras se muestran tres esquemas esperables:

- un caso ideal con el poro centrado en X e Y (Fig. 4.31)
- un poro descentrado en X o Y, en este caso se eligió en X (Fig. 4.32)

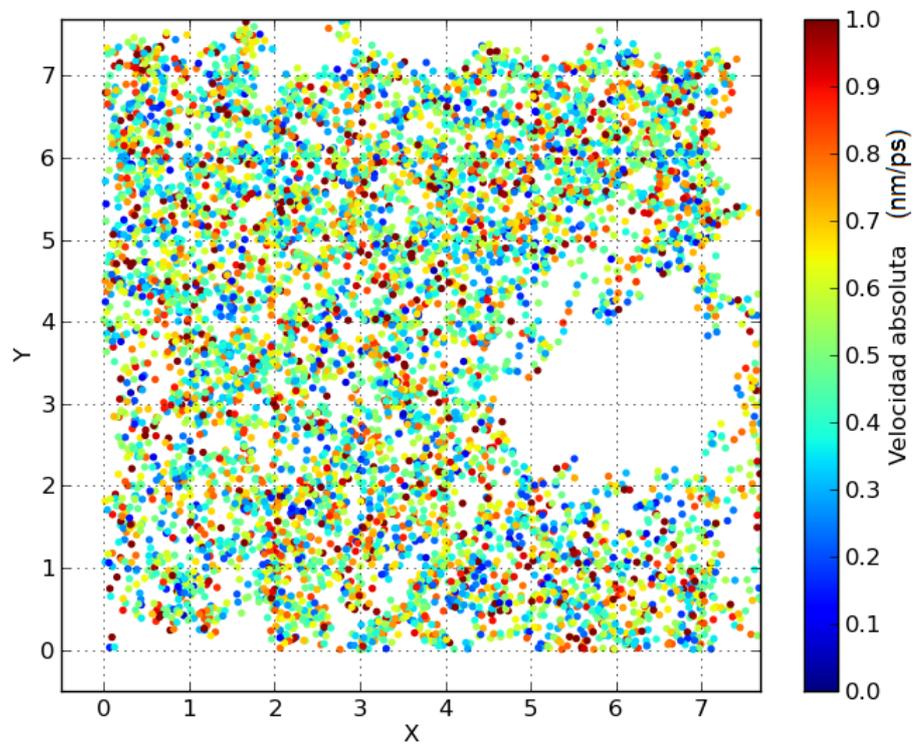


Fig. 4.30: Gráfico de las velocidades por posiciones que muestra el poro formado en la membrana, descentrado.

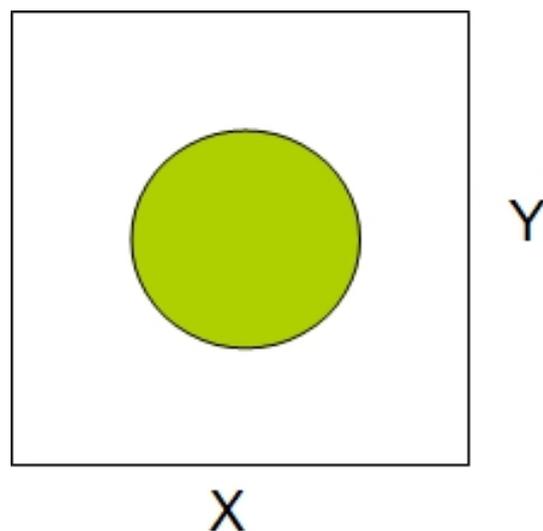


Fig. 4.31: Representación de un poro formado en la membrana, mostrando el caso ideal en que está centrado.

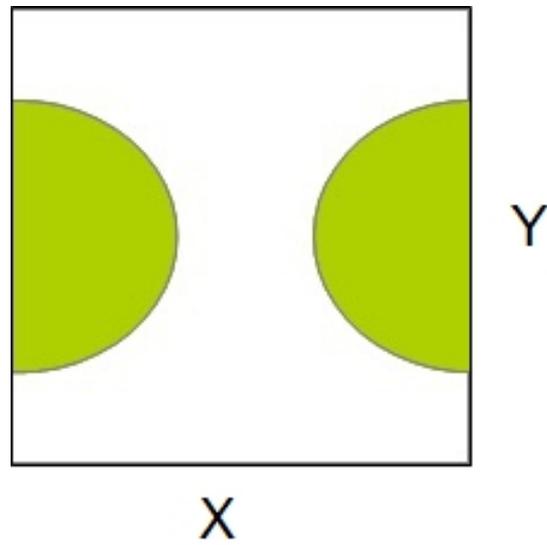


Fig. 4.32: Representación de un poro formado en la membrana, partido en dos.

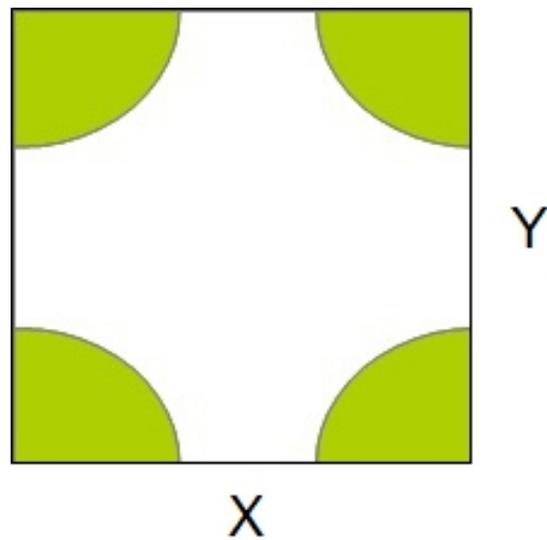


Fig. 4.33: Representación de un poro formado en la membrana, ubicado en las esquinas de la caja.

- un poro descentrado en X e Y (Fig. 4.33)

5. VERIFICACIÓN DEL DESARROLLO DE LA HERRAMIENTA WEB

La verificación del desarrollo de la herramienta web se logró a través de un segundo conjunto de simulación, de forma que se pudieron calcular todos los parámetros como en el conjunto que se utilizó para el desarrollo, y dichos resultados se verificaron gráficamente. Este conjunto es una réplica en las mismas condiciones experimentales de presión, temperatura y estructura inicial. La réplica la genera Gromacs al inicio con un generador de velocidades al azar que se asignan a la estructura inicial. En esta trayectoria la apertura máxima de poro se produce a los 15ns. El objetivo es introducir otra simulación para validar los resultados anteriores.

Las imágenes se pueden ver en las secciones anteriores, donde se muestra la interfaz de la aplicación.

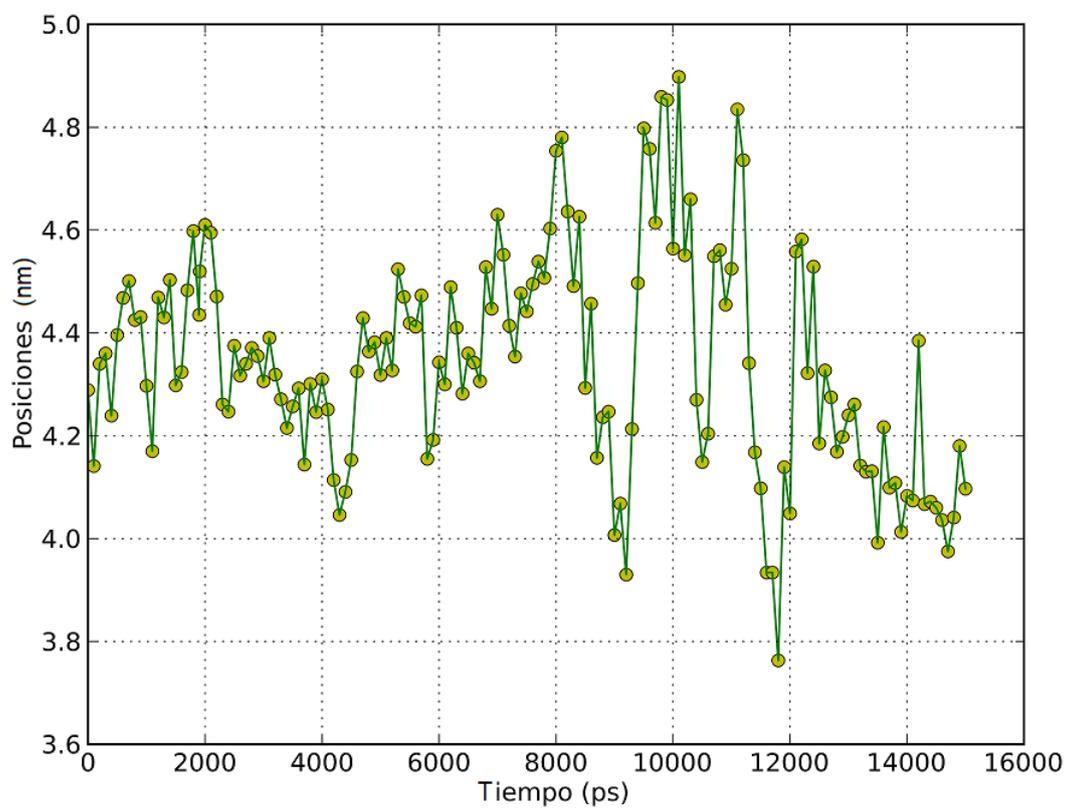


Fig. 5.1: Coordenada X de las posiciones del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su posición en el eje X en cada tiempo.

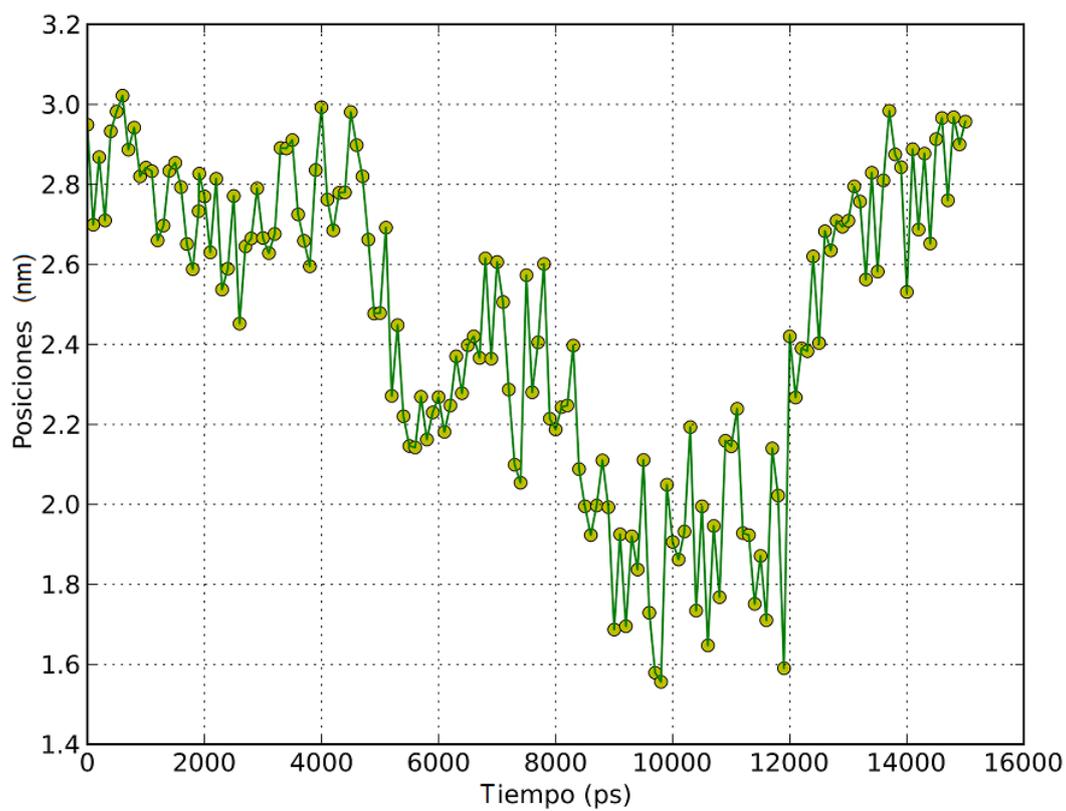


Fig. 5.2: Coordenada Y de las posiciones del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su posición en el eje Y en cada tiempo.

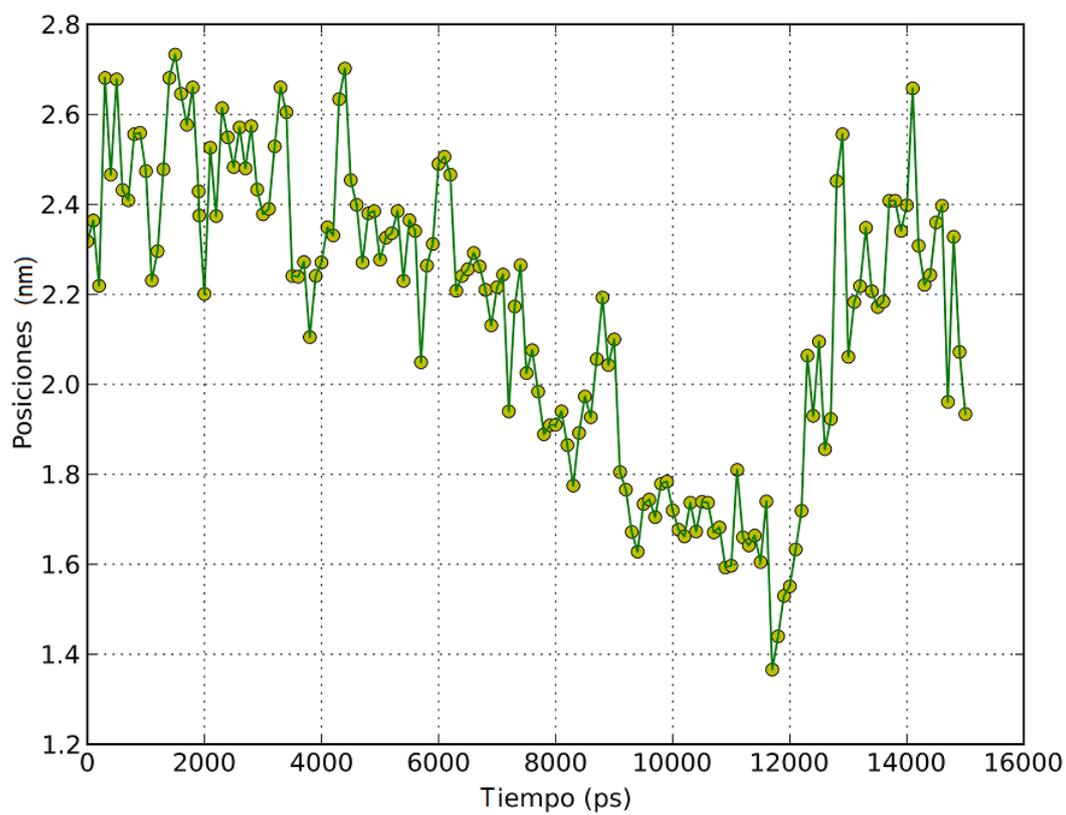


Fig. 5.3: Coordenada Z de las posiciones del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su posición en el eje Z en cada tiempo.

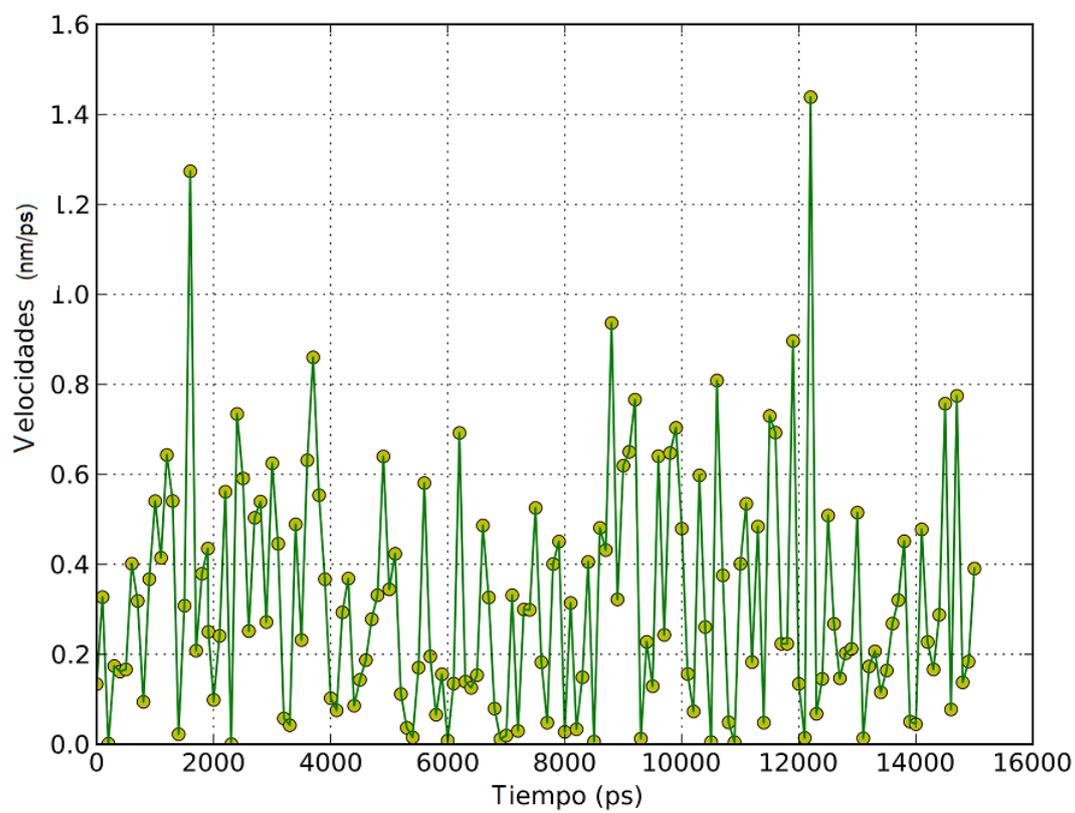


Fig. 5.4: Coordenada X de las velocidades del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su velocidad en el eje X en cada tiempo.

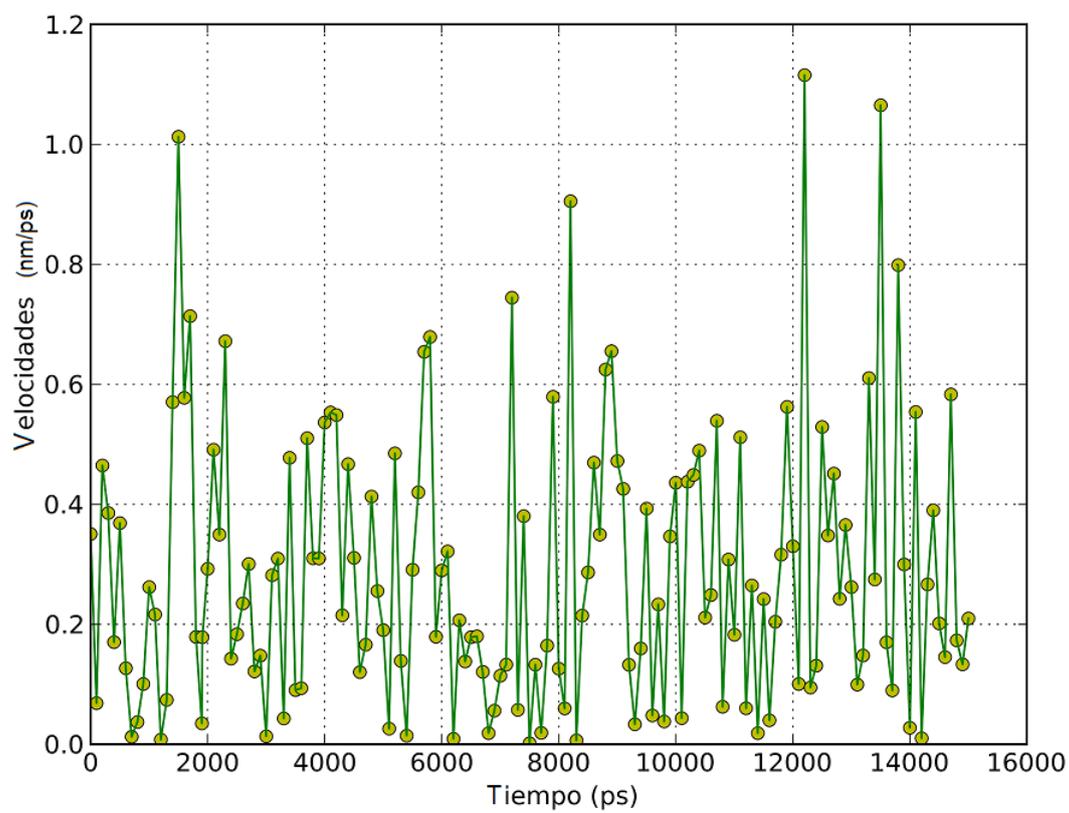


Fig. 5.5: Coordenada Y de las velocidades del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su velocidad en el eje Y en cada tiempo.

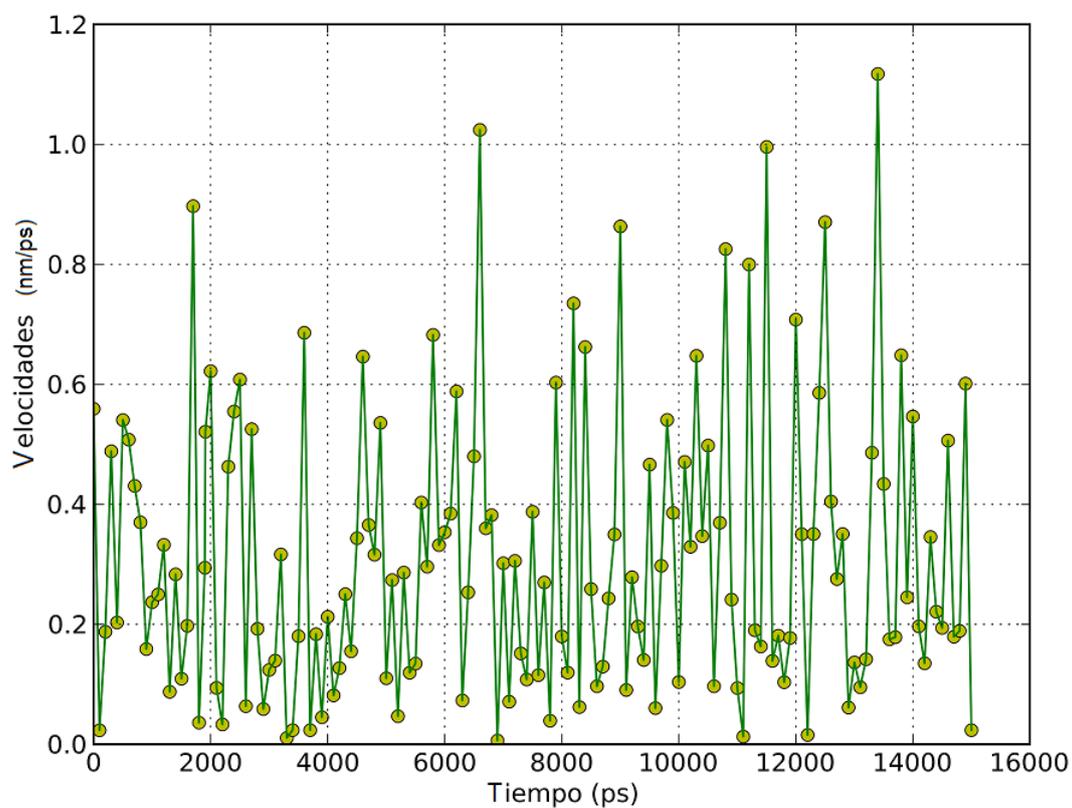


Fig. 5.6: Coordenada Z de las velocidades del átomo seleccionado a lo largo de toda la trayectoria. Se ve como va cambiando su velocidad en el eje Z en cada tiempo.

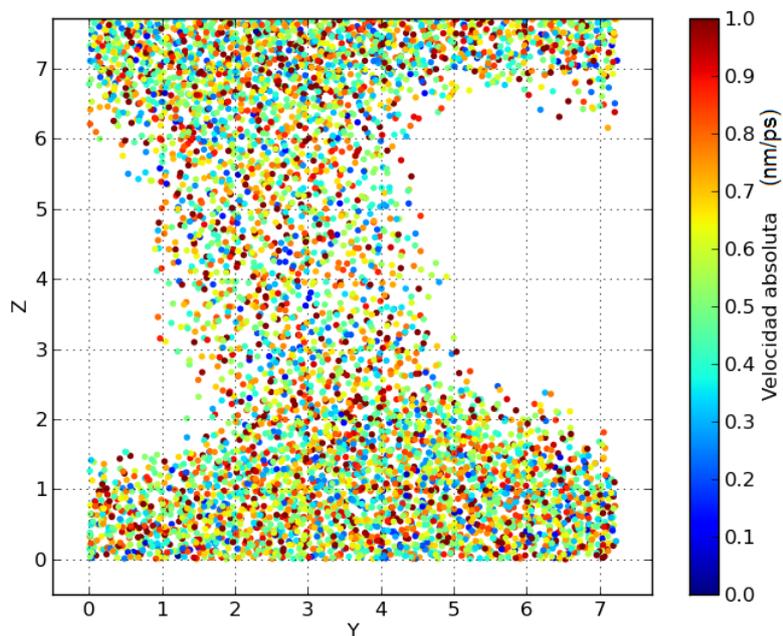


Fig. 5.7: Velocidad absoluta en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje X, viendo las posiciones en Z e Y.

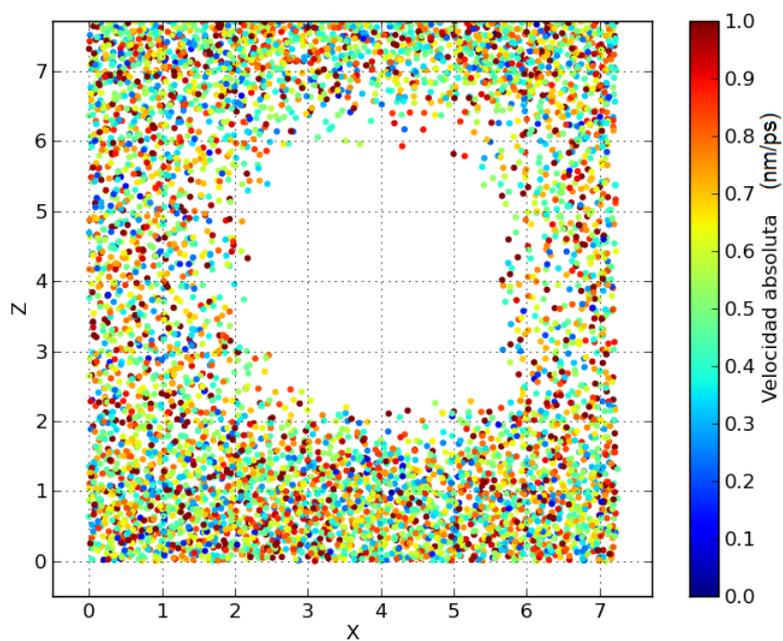


Fig. 5.8: Velocidad absoluta en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje Y, viendo las posiciones en X y Z.

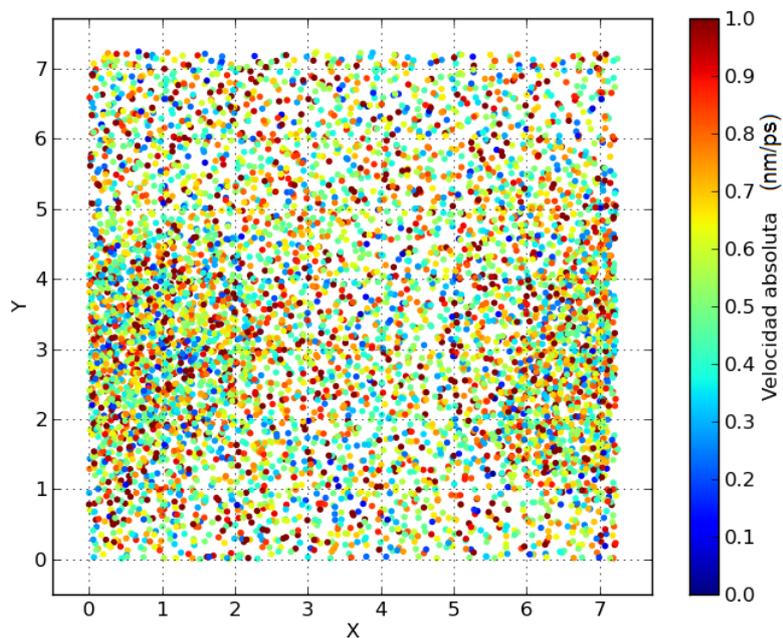


Fig. 5.9: Velocidad absoluta en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje Z, viendo las posiciones en X e Y.

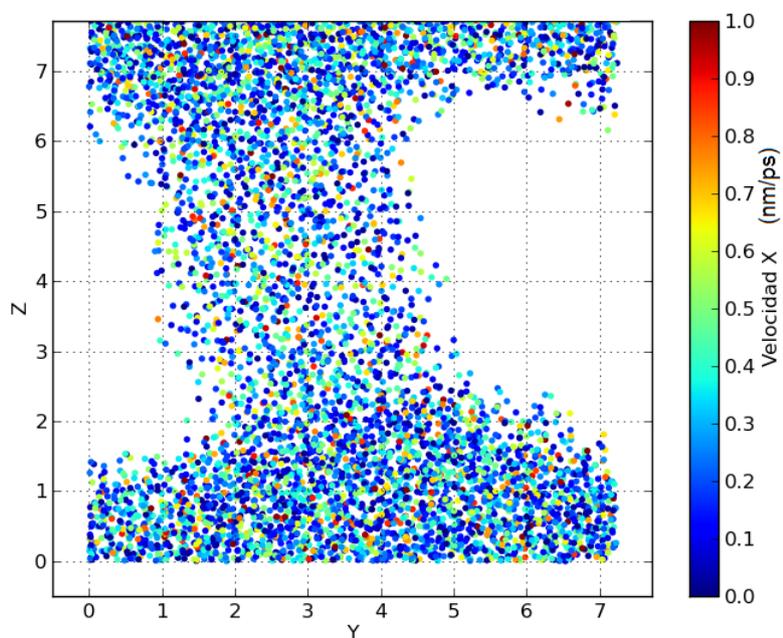


Fig. 5.10: Coordenada X de la velocidad en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje X, viendo las posiciones en Z e Y.

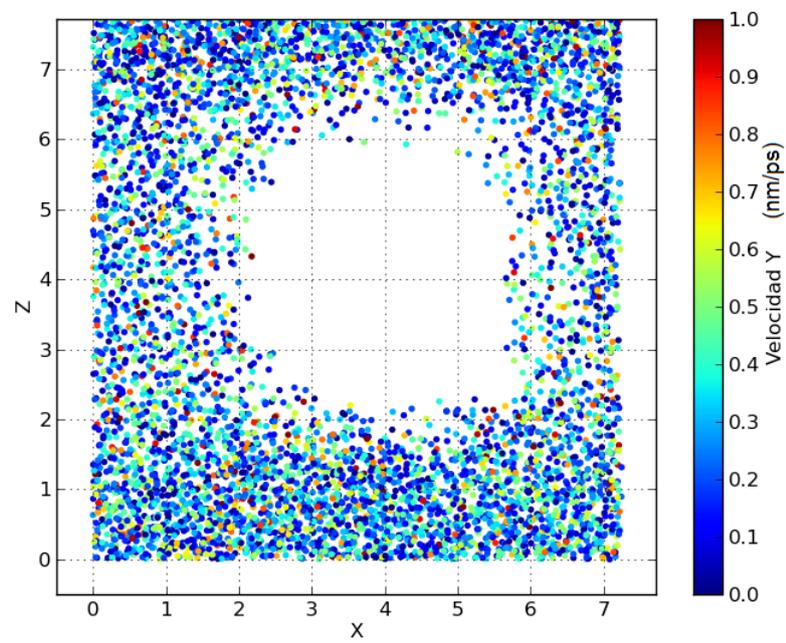


Fig. 5.11: Coordenada Y de la velocidad en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje X, viendo las posiciones en X y Z.

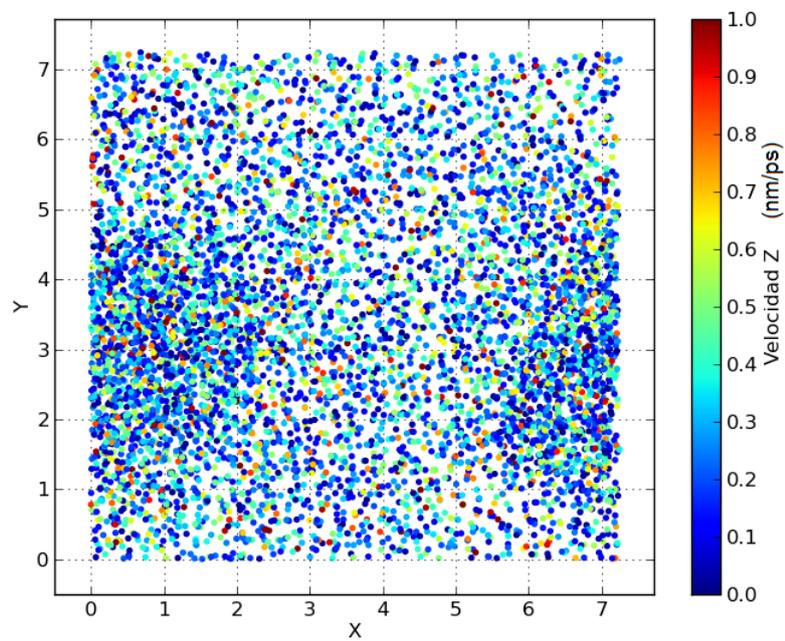


Fig. 5.12: Coordenada Z de la velocidad en el tiempo 15 ns de cada átomo de oxígeno de agua, definida por el color del ítem. Se puede ver la forma del poro desde el eje X, viendo las posiciones en X e Y.

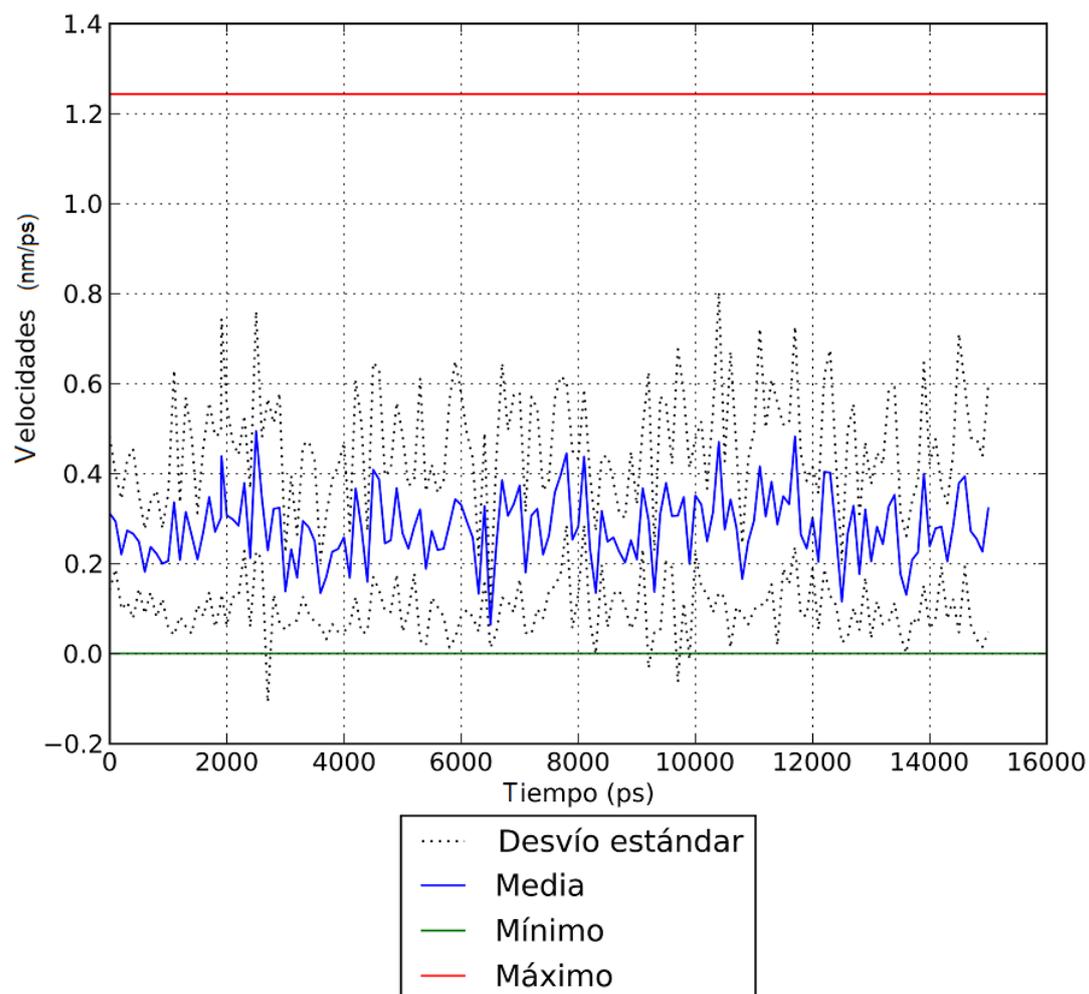


Fig. 5.13: Desvío estándar, la media, el mínimo y el máximo de la coordenada X de las velocidades en cada tiempo, del grupo de átomos seleccionado para analizar a lo largo de toda la trayectoria.

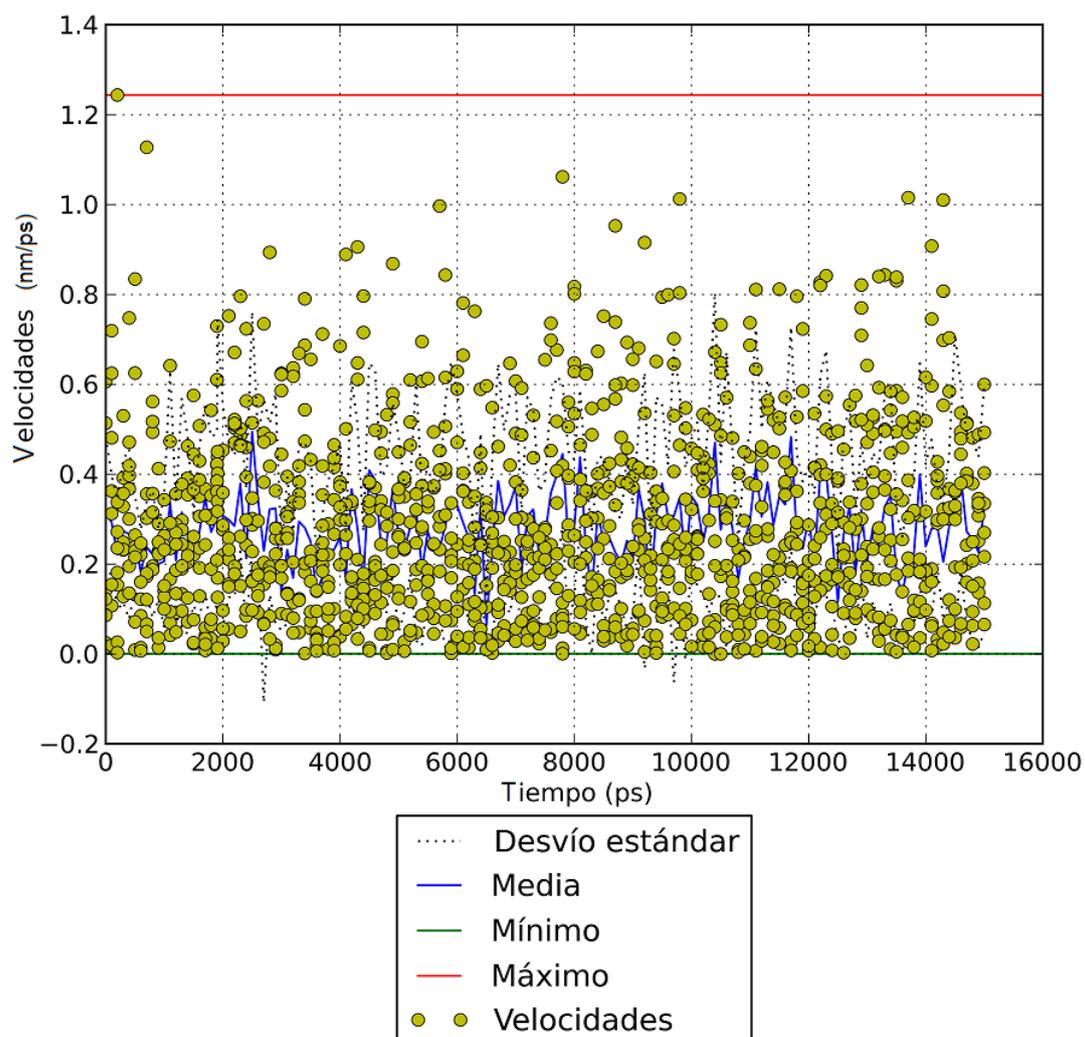


Fig. 5.14: Desvío estándar, la media, el mínimo y el máximo de la coordenada X de las velocidades en cada tiempo, del grupo de átomos seleccionado para analizar a lo largo de toda la trayectoria. Además, los ítems muestran las velocidades en X para cada átomo del grupo.

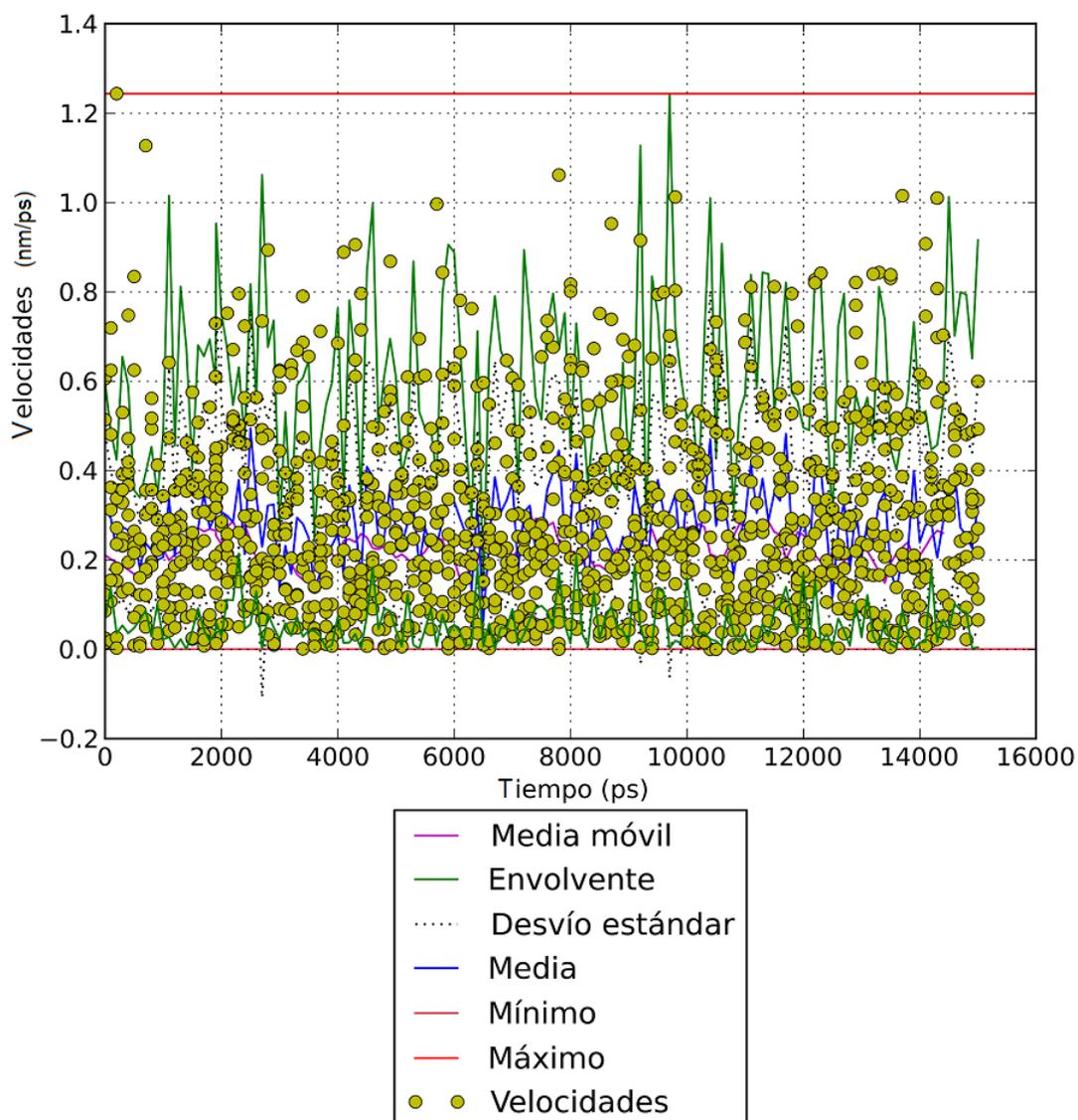


Fig. 5.15: Media móvil, las curvas envolventes, el desvío estándar, la media, el mínimo y el máximo de la coordenada X de las velocidades en cada tiempo, del grupo de átomos seleccionado para analizar a lo largo de toda la trayectoria. La media móvil se calculó promediando 5 puntos.

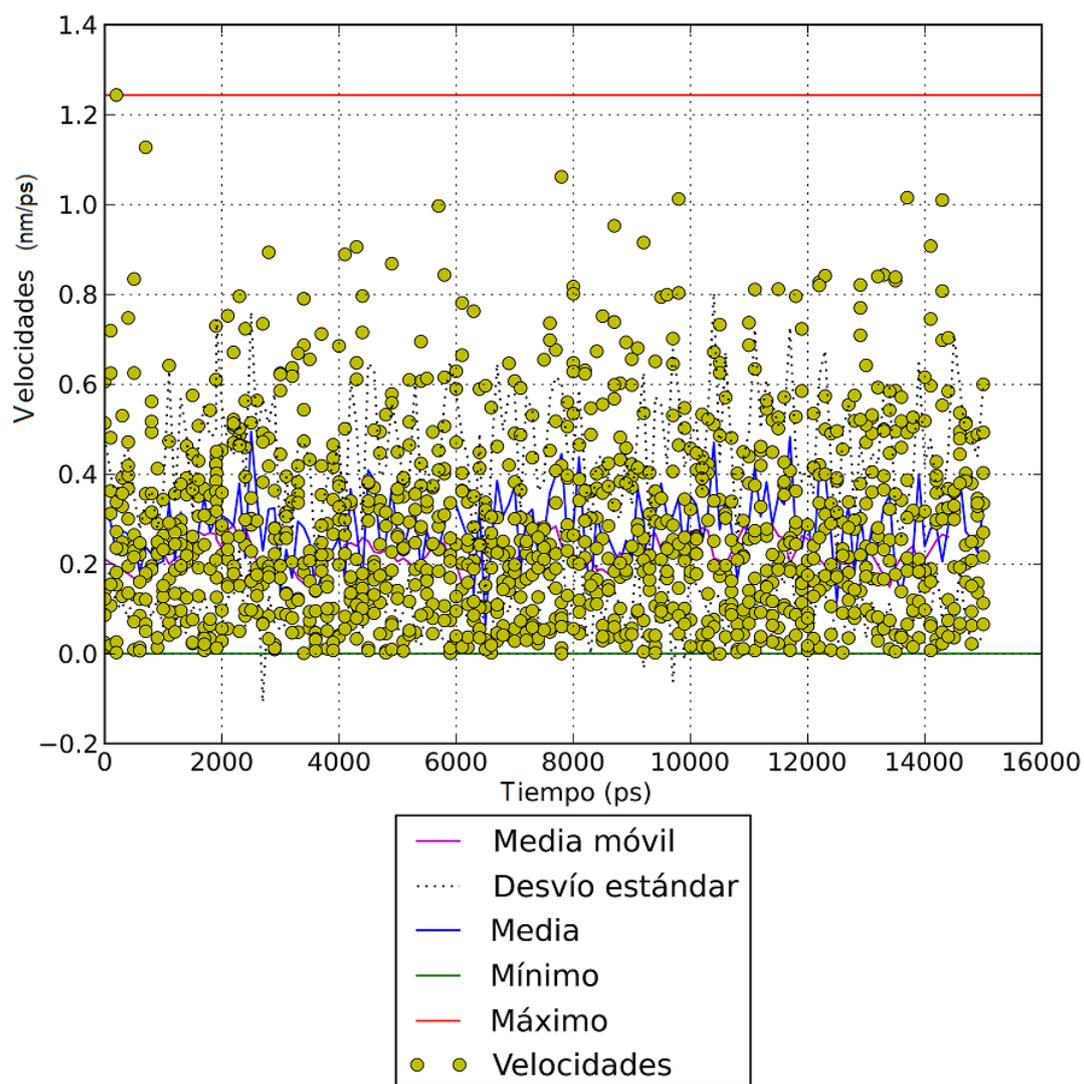


Fig. 5.16: Media móvil, el desvío estándar, la media, el mínimo y el máximo de la coordenada X de las velocidades en cada tiempo, del grupo de átomos seleccionado para analizar a lo largo de toda la trayectoria. La media móvil se calculó promediando 5 puntos.

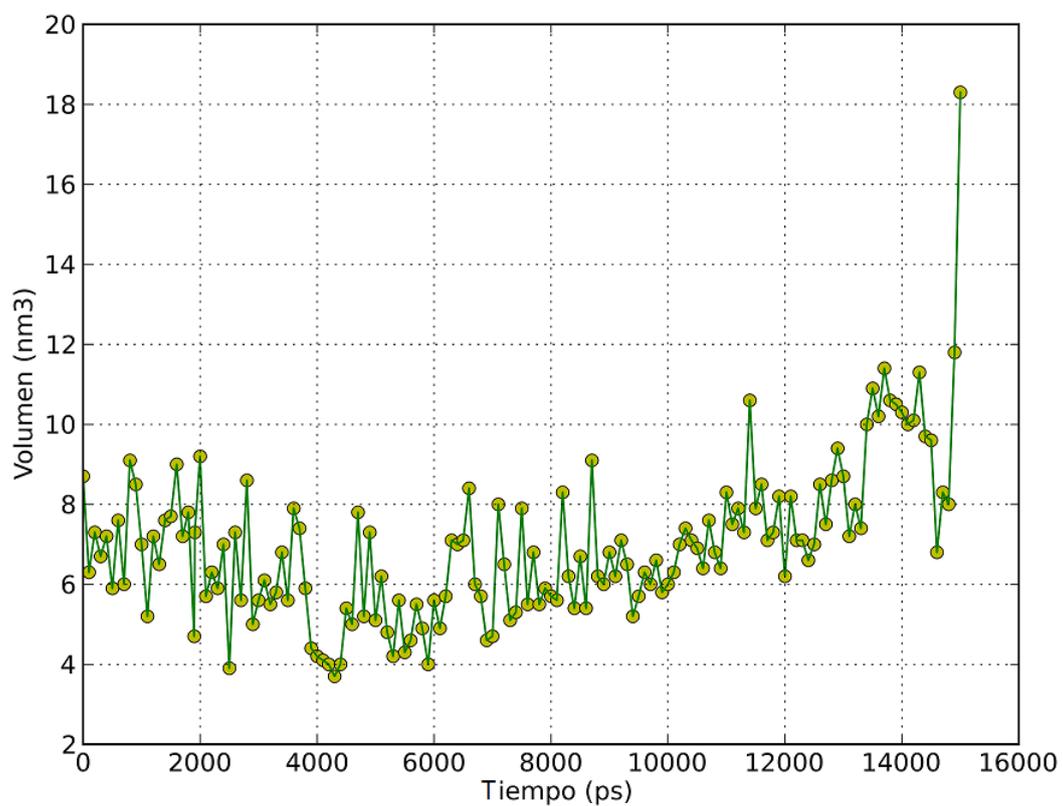


Fig. 5.17: Volúmen que alcanza el poro formado en la membrana en cada momento de la trayectoria. En este caso, se establecieron los límites de la membrana en 3,4 nm y 5,8 nm.

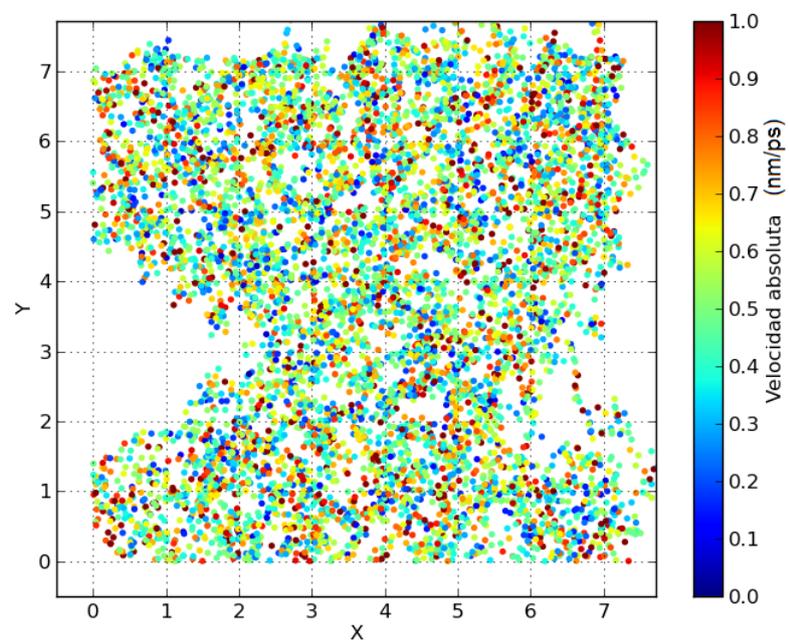


Fig. 5.18: Velocidad absoluta en el tiempo 15 ns, definida por el color del ítem, de cada átomo perteneciente a moléculas de lípidos. Se puede ver la forma del poro desde el eje Z, viendo las posiciones en X e Y.

6. CONCLUSIONES

6.1. Alcance de los resultados obtenidos

Luego de analizar los resultados exhibidos anteriormente, se puede afirmar que la aplicación realizada para analizar trayectorias de un átomo o grupo de átomos funciona de acuerdo a lo planteado en los objetivos de esta trabajo. Fue posible analizar trayectorias con gran volumen de datos y generar gráficos con información estadística. Si bien los parámetros estadísticos calculados son básicos el objetivo, que era su implementación como muestra de la potencia de la incorporación de nuevos módulos de cálculo, se ha cumplido, generando además una fuente importante de información que actualmente es difícil de obtener debido al volumen de datos con el que se trabaja.

La herramienta permite analizar una gran cantidad de datos con módulos de cálculo personalizados, sin embargo el desempeño computacional aún no es el óptimo. Esto se debe en parte a la gran cantidad de archivos y la extensión de los mismos que conforman una trayectoria.

6.2. Discusión

El presente trabajo se dividió en dos etapas, la primera fue el desarrollo de la aplicación web y la segunda su verificación. Para esto se utilizaron dos réplicas de simulaciones de bicapas lipídicas (sólo difieren en el tiempo de apertura de poro). Esta aplicación es capaz de analizar trayectorias con gran número de datos para generar gráficos y reportes con información estadística. Además muestra la flexibilidad de la herramienta al momento de generar módulos de análisis personalizados.

La importancia propuesta en este trabajo sobre el desarrollo de una herramienta web para el análisis de simulaciones de membranas [Borgna, Fernández, and Risk, 2013], ha sido corroborada con la aparición del ST-Analyzer [Jeong, Jo, Wu, Qi, Monje-Galvan, Yeom, Gorenstein, Chen, Klauda, and Im, 2014] que es una herramienta generada en el

mismo entorno por un grupo de varios autores de varias universidades. Si bien el presente desarrollo tiene menor alcance en cuanto a la posibilidad de incorporar archivos de diferente origen de los datos (archivos que no provengan de Gromacs), su flexibilidad radica en la generación de módulos personalizados con implementación de análisis no estándar. La fortaleza de esta aplicación es la posibilidad de personalizar el análisis y es en lo que se ha fundado este trabajo. Para ello se generaron módulos de estadística descriptiva y de análisis de volumen de poro, cumpliendo de esta manera los objetivos planteados con gran éxito.

Limitaciones de la herramienta

Al comienzo de este trabajo la versión de Django 1.4 aún estaba siendo utilizada. Actualmente se realizaron actualizaciones al *framework* y la última versión es la 1.8. Sin embargo, no se realizó esta actualización en esta aplicación ya que la versión utilizada no presenta dificultades ni limitaciones para realizar las diferentes funcionalidades con las que cuenta.

Actualmente el sistema sólo permite archivos de entrada con extensión gro provenientes de Gromacs. Dado que el análisis de los datos se hace a partir de una base de datos, se podría modificar la carga de archivos para que acepte otras extensiones.

Por otro lado para utilizar el sistema no es necesario tener un usuario, cualquier persona puede cargar trayectorias y analizarlas. Dada esta cualidad de la aplicación, cualquier usuario podría eliminar una trayectoria ajena. Además, puede resultar incómodo el uso ya que se mostrarían trayectorias que no son de interés para la persona que lo está utilizando.

6.3. Futuros trabajos

Futuras mejoras a la herramienta son:

- Optimización de la base de datos
- Actualización de versiones de paquetes utilizados.
- Utilizar una aplicación para Django (django-dowser y Heapy son algunas de las opciones) que ayude a controlar los problemas de memoria para minimizarlos

-
- Mejorar la interfaz y el análisis y representación de datos tales como gráficos en 3D, modelos 3D de estructuras simuladas, análisis estadísticos avanzados, barras de progreso en los procesamientos que demoran un tiempo considerable y menús en distintos idiomas (inglés y español)
 - Universalizar la herramienta permitiendo que la aplicación pueda leer y procesar cualquier formato de datos obtenido de una simulación independientemente del paquete que le dio origen
 - Login para permitir ingreso de varias personas y mostrar diferentes trayectorias según usuario.
 - Calcular dinámicamente los límites de la membrana para calcular el volumen del poro
 - Comparar entre diferentes trayectorias los resultados
 - Permitir al usuario guardar sus resultados en su computadora
 - Paralelizar el procesamiento que así lo permita para agilizar el uso del sistema

Bibliografía

- MJ Abraham, D van der Spoel, E Lindahl, B Hess, and the GROMACS development team. Gromacs user manual version 5.0.4, 2014.
- M Belehradek, C Domenge, B Luboinski, S Orłowski, J Belehradek, and LM Mir. Electrochemotherapy, a new antitumor treatment. *Cancer*, 72:3694–3700, 1993.
- HJC Berendsen, D van der Spoel, and R van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91:43–56, 1995.
- KG Borgna, ML Fernández, and M Risk. Herramienta web para post-análisis de simulaciones de dinámica molecular. *Revista Argentina de Bioingeniería ISSN 0329-5257*, 19: 24–28, 2013.
- H Uchoa Brandao, JG Eberhart, NJ Da Silveira Freitas, Jr JC Camera, F Canduri, and De Azevedo Jr. W. Filgueira. Parmodel: a web server for automated comparative modeling of proteins. *Biochem Biophys Res Comm*, 325:1481–1486, 2004.
- RA Böckmann, BL de Groot, S Kakorin, E Neumann, and H Grubmüller. Kinetics, statistics, and energetics of lipid membrane electroporation studied by molecular dynamics simulations. *Biophys. J.*, 95:1837–1850, 2008.
- Django core team. Django: A web framework for the python programming language. <http://www.djangoproject.com/>, 2012.
- ML Fernández, G Marshall, F Sagués, and R Reigada. Structural and kinetic molecular dynamics study of electroporation in cholesterol-containing bilayers. *J Phys Chem B.*, 114:6855–6865, 2010.
- ML Fernández, M Risk, R Reigada, and PT Vernier. Size-controlled nanopores in lipid

- membranes with stabilizing electric fields. *Biochem Biophys Res Commun*, 423:325–330, 2012.
- G Garberoglio. Obgmx: A web-based generator of gromacs topologies for molecular and periodic systems using the universal force field. *J Comp Chem*, 33:2204–2208, 2012.
- MM Ghahremanpour, SS Arab, SB Aghazadeh, J Zhang, and D van der Spoel. Membuilder: a web-based graphical interface to build heterogeneously mixed membrane bilayers for the gromacs biomolecular simulation program. *Bioinformatics*, 30(3):439–441, 2014.
- B Hess, C Kutzner, D van der Spoel, and E Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435–447, 2008.
- A Hospital, P Andrio, C Fenollosa, D Cicin-Sain, M Orozco, and JL Gelpí. Mdweb and mdmoby: an integrated web-based platform for molecular dynamics simulations. *Bioinformatics*, 28(9):1278–1279, 2012.
- J.D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science and Engineering*, 9:90–95, 2007.
- JC Jeong, S Jo, EL Wu, Y Qi, V Monje-Galvan, MS Yeom, L Gorenstein, F Chen, JB Klau-
da, and W Im. St analyzer: A web based user interface for simulation trajectory analysis. *Journal of Computational Chemistry*, 35:957–963, 2014.
- M Karplus and R Lavery. Significance of molecular dynamics simulations for life sciences. *Isr J Chem*, 54:1042–1051, 2014.
- C Margreitter, D Petrov, and B Zagrovic. Vienna-ptm web server: a toolkit for md simulations of protein post-translational modifications. *Nucleic Acids Res*, 41:W422–W426, 2013.
- S Orłowski, J Jr Belehradek, C Paoletti, and LM Mir. Transient electropermeabilization of cells in culture: increase of the cytotoxicity of anticancer drugs. *Biochem Pharmacol*, 37:4727–4733, 1988.

- Postgres core team. pgadmin3. <http://www.pgadmin.org/>, 2002-2010.
- PostgreSQL Global Development Team. Postgresql. <http://www.postgresql.org/docs/8.4/static/>, 1996-2015.
- S Pronk, S Páll, R Schulz, P Larsson, P Bjelkmar, R Apostolov, MR Shirts, JC Smith, PM Kasson, D van der Spoel, B Hess, and E Lindah. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29:845–854, 2013.
- R Development Core Team. R: A language and environment for statistical computing. <http://www.R-project.org/>, 2010.
- D Van Der Spoel, E Lindahl, B Hess, G Groenhof, AE Mark, and HJC Berendsen. Gromacs: Fast, flexible, and free. *J. Comput. Chem.*, 26:1701–1718, 2005.
- DP Tieleman. The molecular basis of electroporation. *BMC Biochemistry*, 5:10, 2004.
- M Tokman, JH Lee, ZA Levine, MC Ho, ME Colvin, and PT Vernier. Electric field-driven water dipoles: nanoscale architecture of electroporation. *PLoS One*, 8:e61111, 2013.
- Torchbox. South. <http://south.aeracode.org/>, 2008.
- G van Rossum. Scripting the web with python. *Scripting Languages: Automating the Web*, *World Wide Web Journal*, 2, 1997.
- PT Vernier, MJ Ziegler, MA Gundersen Y Sun, and DP Tieleman. Nanopore-facilitated, voltage-driven phosphatidylserine translocation in lipid bilayers—in cells and in silico. *Phys. Biol.*, 3:233–247, 2006.
- Z Wang and M Deserno. Systematic implicit solvent coarse-graining of bilayer membranes: lipid and phase transferability of the force field. *New Journal of Physics. The open access journal for physics*, 12:095004, 2010.
- A Watters, G van Rossum, and JC Ahlstrom. Internet programming with python. 1996.

MJ Ziegler and PT Vernier. Interface water dynamics and porating electric fields for phospholipid bilayers. *J. Phys. Chem. B*, 112:13588–13596, 2008.