

SEBBM DIVULGACIÓN

LA CIENCIA AL ALCANCE DE LA MANO

La supercomputación: una poderosa herramienta para la investigación en bioquímica y biología molecular

Juan Fernández-Recio
Centro Nacional de Supercomputación de Barcelona



Biografía

Doctor en Ciencias Químicas por la Universidad de Zaragoza (1999) y Premio Extraordinario de Doctorado. Desde 1999 a 2003 realiza una estancia postdoctoral en el Scripps Research Institute (San Diego, USA), y desde 2003 a 2004 realiza una segunda estancia postdoctoral en la Universidad de Cambridge (Reino Unido) con una beca Marie Curie. En 2004 se incorpora al Parc Científic de Barcelona mediante el programa Ramón y Cajal, para dirigir su propio grupo de investigación. En 2007 obtiene una plaza de investigador principal en el Barcelona Supercomputing Center - Centro Nacional de Supercomputación. Su principal interés es el estudio teórico del reconocimiento molecular en proteínas, campo en el que ha publicado más de 100 artículos. Sus principales contribuciones han sido el desarrollo de herramientas computacionales innovadoras para el modelado estructural de interacciones entre proteínas y la caracterización de superficies de unión. Ha llevado a cabo numerosos proyectos de colaboración para el estudio de diversos sistemas de interés biomédico.

<http://www.sebbm.es/>

HEMEROTECA:

http://www.sebbm.es/ES/divulgacion-ciencia-para-todos_10/la-ciencia-al-alcance-de-la-mano-articulos-de-divulgacion_29

Resumen

La supercomputación se ha convertido en una herramienta fundamental para la investigación en bioquímica y biología molecular, tanto para almacenar y analizar las ingentes cantidades generadas de datos genómicos y estructurales, como para la aplicación de modelos computacionales para entender y predecir el comportamiento de grandes biomoléculas y complejos moleculares.

Summary

Supercomputing is becoming a fundamental tool for research in biochemistry and molecular biology, not only to help storing and analyzing the huge amount of currently generated genomic and structural data, but also to efficiently apply computer models to understand and predict the behavior of large biomolecules and molecular complexes.

En pocos campos es tan evidente el progreso tecnológico como en la informática. Los ordenadores son cada vez más rápidos, siguiendo la inexorable ley de Moore, por la que cada 2 años se duplica la potencia de cálculo de los procesadores que aparecen en el mercado. En este contexto aparece el concepto de supercomputación, que empieza a ser familiar no sólo en la mayor parte de campos científicos, sino incluso en la vida diaria. La supercomputación hace referencia a los grandes ordenadores

construidos para situarse en la frontera de la computación de altas prestaciones. Los primeros superordenadores se construyeron en los años 60 del siglo pasado por Seymour Cray, gracias a la sustitución de las válvulas de vacío por transistores, lo que supuso un salto significativo sobre la capacidad de computación existente, llegando a alcanzar velocidades en la escala del megaflop (un millón de operaciones de coma flotante por segundo). Posteriormente fueron incorporándose diferentes avances tecnológicos, como los circuitos integrados y la paralelización masiva. Desde los años 90, la velocidad de cálculo de los superordenadores ha ido escalando de forma exponencial, aumentando por 1000 aproximadamente cada 10 años. Los superordenadores actuales más potentes se componen de millones de procesadores en paralelo, lo que permite alcanzar velocidades de petaflops, y el almacenamiento de petabytes de información en sus discos [1]. En pocos años, se espera alcanzar la velocidad de exaflop y la capacidad de almacenar exabytes (computación a exaescala).

Desde sus inicios, la supercomputación se ha aplicado a la investigación científica, especialmente en el campo de la Física fundamental. En los años ochenta, la supercomputación empieza a aplicarse de forma general a otras ramas de la ciencia, incluyendo las ciencias de la vida. Es en los últimos años cuando el uso de la supercomputación en ciencias de la vida está siendo especialmente relevante, debido sobre todo a la explosión en información producida por las llamadas ciencias

"ómicas" (genómica, epigenómica, transcriptómica, proteómica y metabolómica), y a los avances en los métodos de modelización molecular.

Más concretamente, la supercomputación se ha convertido en una herramienta importante en bioquímica y biología molecular, donde la determinación, caracterización y modelización de la estructura de las biomoléculas es esencial para entender el mecanismo y función de procesos biológicos a nivel molecular. En primer lugar, la gran cantidad de información estructural generada sobre biomoléculas necesita el almacenaje y gestión en superordenadores, como ocurre con la base de datos de estructuras de proteínas (PDB) [2], que se encuentra asociada a centros de supercomputación desde sus inicios. Además, las técnicas clásicas de determinación estructural, cristalografía de rayos X y resonancia magnética nuclear (RMN), aplican cada vez más procedimientos automáticos con un fuerte componente computacional, y por ejemplo, la criomicroscopía electrónica necesita capturar y analizar cientos de miles de imágenes para ajustar computacionalmente las estructuras. Por otro lado, la modelización molecular permite estudiar el comportamiento de biomoléculas en condiciones realistas. El premio Nobel de Química de 2013 fue galardonado a Michael Levitt, Martin Karplus y Arieh Warshel, por sus trabajos pioneros en los años 70 en el desarrollo de métodos computacionales de modelado molecular para entender y

predecir sistemas químicos complejos. Desde entonces, los algoritmos de modelado molecular y su aplicación a sistemas bioquímicos han evolucionado en paralelo a la capacidad de cálculo de los ordenadores [3]. De esta forma, la implementación de métodos de mecánica cuántica en programas altamente eficaces para su uso en supercomputación, como CPMD [4], y su extensión a la metodología híbrida mecánica cuántica/mecánica molecular (QM/MM), ha permitido la aplicación de dichas metodologías a sistemas biológicos, así como su integración con métodos de acoplamiento molecular (*docking*) y cribado virtual (VLS por sus siglas en inglés), para el diseño de nuevos fármacos. Los algoritmos de simulación de dinámica molecular más populares, como GROMACS, NAMD o AMBER, se han adaptado a los requerimientos de la supercomputación, pudiendo usar simultáneamente miles de procesadores en paralelo. Esto ha hecho posible la simulación de sistemas de millones de átomos como las cápsidas virales, el modelado a alta resolución de canales iónicos, o la descripción del complejo del poro nuclear, uno de los ensamblados proteicos más grandes de la célula eucariota (65 MDa en levadura). Gracias a superordenadores especialmente dedicados a las simulaciones biomoleculares se están consiguiendo tiempos de simulación que llegan a la escala del milisegundo, y permiten describir procesos relevantes a nivel biológico, como el plegamiento de proteínas, la formación

de complejos, transiciones alostéricas, o la unión a sustratos, entre otros [5].

Uno de los retos futuros para la supercomputación de exaescala será la interpretación de datos genómicos a nivel funcional en un contexto de medicina personalizada [6], lo que requerirá su integración con datos bioquímicos, estructurales y biofísicos, junto con el modelado de biomoléculas y sus interacciones. Será esencial completar el mapa de interacciones entre biomoléculas y modelar estructuralmente cada una de las cientos de miles de interacciones que se forman en la célula (interactómica estructural), lo que servirá para generar modelos cada vez más complejos y pasar del estudio a nivel molecular a un nivel superior (biología de sistemas).

En definitiva, la supercomputación está ya cambiando nuestra forma de entender la investigación científica, permitiendo desarrollar de forma eficaz modelos teóricos complejos para testar hipótesis de mecanismos biológicos a nivel molecular, lo que tendrá un impacto evidente en todas las áreas biológicas, incluyendo la bioquímica y biología molecular.

Referencias

- [1] <http://www.top500.org/>
- [2] <http://www.rcsb.org/pdb>
- [3] Schlick et al. (2011), "Biomolecular modeling and simulation: a field coming of age" *Q. Rev.Biophys.* 44, 191-228. DOI: 10.1017/S0033583510000284 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3700731/>
- [4] <http://www.cpmd.org>
- [5] Piana et al. (2013) "Atomistic description of the folding of a dimeric protein". *J.Phys.Chem B* 117, 12935-12942. [dx.doi.org/10.1021/jp4020993](https://doi.org/10.1021/jp4020993) <http://pubs.acs.org/doi/abs/10.1021/jp4020993>
- [6] Rincón del profesor: SEBBM. Abril 2015. "El papel clave de la bioinformática en el tratamiento personalizado del cáncer." Fátima Al-Shahrour. http://www.sebbm.es/ES/divulgacion-ciencia-para-todos_10/el-papel-clave-de-la-bioinformatica-en-el-tratamiento-personalizado-del-cancer--_1132

Figura. "El superordenador Mare Nostrum (BSC), con casi 50.000 procesadores y 1 petaflop de velocidad, se encuentra actualmente entre los 100 ordenadores más potentes del mundo".

