

16 DE SEPTIEMBRE DE 2005

## Computadoras hacen grandes avances hacia la predicción de la estructura proteica

Las computadoras pueden predecir la estructura de proteínas pequeñas casi tan bien como los métodos experimentales, al menos algunas veces, según indican nuevos estudios realizados por investigadores del Instituto Médico Howard Hughes.

Los resultados, que fueron publicados en el número del 16 de septiembre de 2005, de la revista *Science*, hacen que se pueda esperar que los científicos eventualmente sean capaces de determinar la estructura de proteínas a partir de sus secuencias genómicas, problema que hasta el momento parece insuperable.

---

"La razón por la que esta investigación es interesante es que estamos avanzando hacia la predicción de estructuras en base a secuencias. No es que el problema esté resuelto, sino que hay esperanza."

— **David Baker**

---

“Por más de 40 años, se ha sabido que la secuencia de aminoácidos de una proteína especifica su estructura tridimensional, pero nadie ha podido traducir la secuencia a una estructura exacta”, dijo el autor senior David Baker, investigador del HHMI en la Universidad de Washington. “La razón por la que esta investigación es interesante es que estamos avanzando hacia la predicción de estructuras en base a secuencias. No es que el problema esté resuelto, sino que hay esperanza”.

Las proteínas son máquinas biológicas y los científicos necesitan determinar sus estructuras para entender la forma en la que funcionan. Actualmente, los científicos determinan las estructuras exclusivamente al medir las características atómicas de las proteínas en el laboratorio. En cambio, “en este caso, nunca tocamos un tubo de ensayo”, dijo Baker. “Se lo dimos a una computadora y dijimos, ‘adelante’”.

En el estudio, un programa de computadora sofisticado plegó 17 cadenas cortas de aminoácidos en 100.000 variaciones posibles. Cuando los investigadores compararon las mejores predicciones con las estructuras reales resueltas anteriormente por otros científicos que usaron técnicas experimentales, obtuvieron un índice de éxito comparable al de los mejores bateadores de la liga mayor de béisbol.

“Para cerca de un tercio del grupo de proteínas pequeñas que utilizamos como ejemplo, generamos predicciones de estructuras de relativamente alta resolución, y partes de las mismas se predijeron con resolución cercana a la atómica”, dijo el primer autor Philip Bradley, estudiante postdoctoral del laboratorio de Baker. “Para nosotros, es un verdadero paso adelante para lograr estructuras que son de cierta forma comparables a lo que se puede obtener mediante experimentos”.

Los resultados alentadores provienen del refinamiento de un programa de modelado computacional sofisticado llamado Rosetta, desarrollado por primera vez hace varios años en el laboratorio de Baker. El programa funciona con la premisa de que las proteínas se colapsan en su estado de menor energía, como una bola que rueda hacia abajo por una colina hasta que se detiene a nivel del piso. Se computan las energías de centenares de miles de formas posibles generadas por la computadora y la forma de energía más baja se selecciona como predicción.

El proceso de predicción tiene lugar en dos pasos, dijo Bradley. La primera etapa utiliza un modelado aproximado que permite el cálculo rápido de la energía y, por lo tanto, se puede realizar rápidamente, mientras que el segundo utiliza un modelado muy detallado para el cual los cálculos de energía llevan más tiempo pero son mucho más exactos. Una búsqueda a gran escala a través de posibles estructuras se realiza en la primera etapa y las localizaciones prometedoras son entonces exploradas detalladamente en la segunda etapa.

La primera etapa aprovecha el hecho de que todos los aminoácidos tienen secciones idénticas, que forman la columna de la proteína. La computadora agrega una imagen borrosa de las cadenas laterales que sobresalen y que dan a cada uno de los aminoácidos su identidad única. La secuencia de cadenas laterales da, en última instancia, a cada una de las proteínas su forma característica en base al entorno y a los vecinos que prefieren.

Entonces la computadora aleatoriamente tuerce, enrolla y dobla cada secuencia de aminoácidos en 100.000 formas distintas que están basadas en la localización preferida de los aminoácidos. Algunos aminoácidos tienden a zambullirse hacia el mundo acuoso de la superficie de la proteína mientras que otros se ponen a cubierto dentro de la misma. La computadora también explica los hábitos sociales de los 20 aminoácidos; algunos desean estar cerca uno de otro y a otros les gusta estar distanciados.

En la segunda etapa, Rosetta substituye la imagen borrosa de las cadenas laterales por modelos detallados y físicamente realistas con todos los átomos representados. A partir de las posiciones de los átomos en las cadenas

laterales y de la columna de la proteína, luego la computadora utiliza un detallado campo de fuerza fisicoquímico que favorece el empaquetamiento compacto de los átomos y la unión de hidrógeno para computar de forma más precisa la energía de la estructura.

“Lo que parece ser crítico es el empaquetamiento de la molécula”, dijo Baker. “La proteína encaja perfectamente sin agujeros en el centro y sin átomos encimados. Está tan densamente empaquetada como es posible. Es como un rompecabezas tridimensional”.

Los investigadores aumentaron sus probabilidades de encontrar la contraparte correcta al repetir el proceso de dos etapas con 50 homólogos de las proteínas pertenecientes a otras genomas, tales como de ratón o mosca. Se analizó el protocolo primero en una prueba de predicción ciega que es considerada como el estándar más alto para eliminar las tendencias de los modelos de predicción a determinadas estructuras proteicas.

“No podemos computar las energías perfectamente, pero el problema más grande es la búsqueda a través de las formas posibles”, dijo Baker. “Cuando no podíamos obtener la respuesta correcta en la computadora, lo que casi siempre sucedía era que la estructura real tenía la energía más baja, así que hubiéramos acertado si hubiéramos explorado esta parte del espacio”.

En un artículo relacionado publicado en el número de agosto de la revista *Proteins*, Baker y sus colegas publicaron que se pueden utilizar metodologías similares para predecir las estructuras de los complejos de proteínas. “Por primera vez, los métodos computacionales pueden, para un subconjunto de casos, producir modelos realmente exactos”, dijo.

Baker compara las simulaciones computacionales de las proteínas con el problema de intentar encontrar el punto más bajo en la superficie de la Tierra por primera vez. Una forma simple de encontrar el lugar más bajo en el planeta es enviar a tantos exploradores como sea posible. Cuanto más exploradores haya, más probable será que uno de ellos tropiece con la costa del Mar Muerto - punto más bajo de la Tierra en tierra que no está cubierto por agua-. Cada una de las miles de simulaciones computacionales es como un explorador.

Aunque el índice de éxito del 33 por ciento publicado en el artículo de *Science* podría ser lo suficientemente bueno como para asegurarle un lugar en el salón de la fama a un jugador de béisbol, Baker señala rápidamente que todavía no es lo suficientemente confiable para la biología. Modelos mejores dependerán de estrategias más inteligentes de exploración y de más poder computacional. “Si los métodos siguieran siendo iguales, no solucionaríamos el problema”, dijo Baker. “Por otra parte, nos iría mejor con 10 veces más de tiempo computacional”.

Lleva menos de un minuto que una proteína se pliegue en su forma correcta en las células, pero una estimación repetida varias veces predice que llevaría más que la edad del universo para que una computadora muestree todas las confirmaciones posibles de una proteína plegada. El laboratorio de Baker ya

recibe ayuda de grandes centros de supercomputadoras en San Diego e Illinois.

Pronto habrá más ayuda gracias a varios de los 5.000 estudiantes de primer año que ingresan este otoño a la Universidad de Washington. Utilizando el software desarrollado para ayudar al proyecto de Búsqueda de Inteligencia Extraterrestre (SETI, por sus siglas en inglés), los estudiantes pueden poner sus computadoras a trabajar durante la noche mientras ellos están durmiendo para buscar el paisaje atómico de la estructura de energía más baja de proteínas.

Para mejorar la predicción de la estructura de proteínas aún más, el grupo de Baker también ha comenzado un proyecto computacional que esperan sea ayudado por miembros del público general. El proyecto, llamado Rosetta@home, es un proyecto de investigación científica que utiliza computadoras conectadas por Internet para predecir y diseñar estructuras de proteínas, e interacciones entre proteínas y entre proteínas y ligandos. El objetivo es desarrollar métodos que predigan y diseñen exactamente estructuras y complejos proteicos, esfuerzo que podría en última instancia ayudar a los investigadores a desarrollar curaciones para enfermedades humanas tales como el cáncer, el VIH/SIDA y la malaria. Hay más información disponible en Internet en <http://boinc.bakerlab.org/rosetta> .