

12 DE FEBRERO DE 2001

## La clave para la predicción del plegamiento de proteínas estaría en Rosetta

Un método computacional desarrollado por el investigador del Instituto Médico Howard Hughes, David A. Baker y sus colegas, ha resultado ser muy exitoso en la predicción de la estructura tridimensional de una proteína plegada, partiendo de su secuencia lineal de aminoácidos.

Rosetta, el nombre de la técnica computacional desarrollada por Baker y sus colegas en la Universidad de Washington, tuvo un éxito llamativo al predecir la estructura tridimensional de proteínas durante la cuarta Evaluación Crítica de Técnicas para la Predicción de la Estructura de Proteínas (CASP4, por sus siglas en inglés).

---

"El poder de estos métodos es que, dado que no se necesita otra información que no sea la secuencia de aminoácidos, es concebible pensar en ir a lo largo del genoma, generando estructuras y posiblemente obteniendo información sobre la función de cada proteína."

— **Tania A. Baker**

---

Para realizar el experimento CASP4, que comenzó en abril de 2000, más de 100 grupos de investigación generaron las estructuras tridimensionales de 40 proteínas candidatas. Una proteína candidata, o diana, era considerada elegible para CASP4 si su estructura tridimensional había sido deducida a través de análisis estructural, pero sólo si todavía no había sido publicada por los investigadores o hecha pública en una base de datos de estructura de proteínas. A cada grupo de investigación le dieron la secuencia de aminoácidos de una proteína diana, y les hicieron desarrollar modelos tridimensionales de las proteínas plegadas. Los resultados de CASP4 fueron presentados y discutidos en una conferencia en Asilomar, en California, al inicio de diciembre.

Incluso hace unos pocos años, dice Baker, el éxito en la predicción de cómo las proteínas asumen sus complejas formas tridimensionales era considerado altamente improbable si no se conocía una proteína de estructura relacionada. Para aquellas proteínas cuya secuencia se asemeja a la de una proteína de estructura conocida, la estructura tridimensional de la proteína conocida se puede utilizar como "templado" para deducir la estructura desconocida de la proteína. Sin embargo, cerca del 60 por ciento de las secuencias de proteínas que surgen de los proyectos de secuenciación genómica, no tiene ningún homólogo de estructura conocida.

A pesar de la falta de éxito, los investigadores han continuado tratando de resolver el problema de predecir la estructura tridimensional de las proteínas, partiendo sólo de la secuencia de aminoácidos-llamada predicción *ab initio* -dado que es una problemática central de la biología molecular computacional. Recientemente, el problema ha adquirido más importancia a medida que el trabajo de secuenciación del genoma humano ha provisto a los investigadores de cantidades masivas de datos de secuencias génicas sin analizar.

"Uno de los problemas que tiene la predicción de la estructura es que es muy fácil producir un programa que prediga correctamente la estructura de una proteína, sólo si se conoce la estructura correcta de antemano", dijo Baker. "Al desafiar a los investigadores a que produzcan modelos antes de conocer la respuesta correcta, los experimentos CASP han generado un avance inestimable en el campo".

El algoritmo computacional Rosetta para la predicción del plegamiento de proteínas utiliza los estudios experimentales de plegamiento de proteínas, del laboratorio de Baker y muchos otros. "Durante el plegamiento, cada segmento local de la cadena oscila entre un subconjunto distinto de conformaciones locales", dijo Baker. "El plegamiento de la estructura nativa tiene lugar cuando las conformaciones adoptadas por los segmentos locales y sus orientaciones relativas permiten el entierro de los residuos hidrofóbicos, el apareamiento de las láminas beta, y otras características de baja energía de las estructuras nativas de proteínas. En el algoritmo Rosetta, la distribución de las conformaciones observadas para cada segmento corto de secuencias, presentes en estructuras proteicas conocidas, se toma como una aproximación del conjunto de conformaciones locales que el segmento de la secuencia experimentaría durante el plegamiento. Entonces, el programa busca la combinación de estas conformaciones locales que tiene la energía total más baja".

Los resultados usando Rosetta, publicados en la reunión CASP4, revelaron que se han realizado enormes progresos en la predicción *ab initio* de estructura, dijo Baker. Por ejemplo, hace cuatro años, en la reunión CASP2, había pocas predicciones razonables *ab initio* de estructura, dijo. "En cambio, en el experimento CASP4, el análisis de las estructuras predichas demostró que para la mayoría de las proteínas sin homología con proteínas de estructura conocida, habíamos producido modelos razonables de baja resolución para grandes fragmentos, de hasta cerca de 90 aminoácidos.

"Fue interesante que algunas de las estructuras predichas eran bastante similares a las estructuras de proteínas que ya se habían resuelto, y que resultaron tener funciones similares a las proteínas dianas, aunque no había semejanza significativa a nivel de la secuencia. De esta manera, nuestras estructuras predichas proporcionaron pistas sobre la función, lo que no se podía lograr por métodos tradicionales de comparación de secuencias", dijo Baker.

Peter Kollman, experto en modelado molecular computacional, en la Universidad de California, en San Francisco, quien participó en el experimento CASP4, ofreció una perspectiva adicional: "Los evaluadores de las estructuras para las predicciones *ab initio* le daban dos puntos a una estructura que estaba 'entre las mejores', un punto a una estructura que era 'bastante buena' y cero si la estructura estaba muy alejada de la correcta".

"Fue asombroso que el grupo de David Baker obtuviera 31 puntos y que el grupo que le seguía sólo tuviera 8 puntos. Es como lo que sucedió en béisbol en 1927, cuando Babe Ruth hizo 60 homeruns y el segundo jugador 14, y algunos equipos completos no realizaron tantos como él.

"No obstante, todavía existe una forma de predecir estas estructuras con exactitud experimental", dijo Kollman, "pero todos estamos esperanzados en que esto también avanzará".

Baker coincide: "Aunque estas estructuras tridimensionales no son lo suficientemente detalladas, por ejemplo, para el diseño de drogas basado en estructuras, pueden producir inestimables avances en la comprensión de la función de proteínas desconocidas", dijo Baker. "Así que nuestro objetivo es utilizar el método de predicción *ab initio* de estructuras para producir los modelos tridimensionales de proteínas cuya función es desconocida. Y usando esos modelos, podremos buscar en las bases de datos de estructuras de proteínas para determinar si son similares a proteínas de la función conocida. A partir de esta semejanza sería posible realizar inferencias funcionales sobre lo que lo hacen esas proteínas.

"Estamos muy excitados con la idea de intentar hacer esto a gran escala, para hacer inferencias funcionales sobre la gran porción de proteínas acerca de las que no se puede decir nada en la actualidad", dijo Baker. "El poder de estos métodos es que, dado que no se necesita otra información que no sea la secuencia de aminoácidos, es concebible pensar en ir a lo largo del genoma, generando estructuras y posiblemente obteniendo información sobre la función de cada proteína".