

Predicción de la estructura TriDimensional de la Proteína

Víctor Alfonso Rivera Zúñiga

October 15, 2008

Introducción

Dogma de la Biología Molecular
Proteína

Planteamiento del Problema

Cálculo energético de la estructura de la proteína
Mecánica Molecular

Búsqueda del espacio conformacional de la proteína
Dinámica Molecular
Método Monte Carlo
Algoritmos Genéticos

Métodos Conocidos

El dogma de la Biología Molecular describe el proceso en dos pasos (transcripción y traducción) por la cual la información genética lleva a la producción de proteínas.



Recordemos que el ADN es una molécula en donde, en forma química, se encuentra información que se hereda. Está constituido por un polímero lineal a partir de 4 monómeros:

- ▶ Adenina (A)
- ▶ Timidina (T)
- ▶ Guanina (G)
- ▶ Citosina (C)

Descripción

► Proteína:

La proteína es un polímero de amino ácidos. (Alfabeto de 20 amino ácidos) El polímero se pliega en el espacio formando una estructura tridimensional par. Cumple funciones. Su función está dada por la complejidad de su estructura.

Descripción

▶ Proteína:

La proteína es un polímero de amino ácidos. (Alfabeto de 20 amino ácidos) El polímero se pliega en el espacio formando una estructura tridimensional par. Cumple funciones. Su función está dada por la complejidad de su estructura.

▶ Cadenas de plipéptidos:

Los amino ácidos se combinan por medio de químicos unos con otros. El enlace covalente que une dos amino ácidos se denomina peptídico. Cuando dos amino ácidos se combinan, forman un dipéptido y una cade más larga recibe el nombre de polipéptido.

Descripción

- ▶ Estructura de la proteína:
Una proteína está formada por cadenas de polipéptidos, estas cadenas se encuentran enrolladas o plegadas de modo que forman una macromolécula con una conformación específica tridimensional. Se tiene 4 niveles de organización (estructura) en una proteína:
 - ▶ Primaria
 - ▶ Secundaria
 - ▶ Terciaria
 - ▶ Cuaternaria

La función bioquímica de las proteínas está determinada por la complejidad de la estructura de la proteína. La estructura de la proteína es principalmente detectada por la secuencia lineal específica de sus aminoácidos, esto fue demostrado por Anfinsen en un experimento histórico (...).

La estructura de una proteína puede ser determinada experimentalmente usando técnicas como cristalografía de rayos X o resonancia magnética nuclear. Sin embargo, estas técnicas consumen mucho tiempo y toman mucho labor. Además, no todas las proteínas del universo pueden ser experimentalmente determinadas.

Problema: Inferir la estructura tridimensional de una proteína, dado la secuencia de aminoácidos de dicha proteína.

Cálculo de la Energía

La observación que una proteína, en condiciones apropiadas, se pliegue en la misma estructura estable implica que su estructura nativa esté libre de energía (o sea mínima) entre los estados que la proteína puede explorar.

Una proteína gana una estructura estable si la energía perdida mientras se pliega es compensada por la energía ganada en su estado de plegado.

- ▶ Durante el plegado, una proteína pierde entropía porque la entropía es relativa al número de estados disponibles en el sistema
- ▶ La cadena de la proteína, pierde energía asociada con sus enlaces de hidrógeno debido a la atracción entre sus átomos polares y el disolvente en el que se encuentra.
- ▶ Una vez la proteína se encuentra plegada, esta gana entropía ya que encierra el lado hidrófobo y deja en el exterior el lado hidrófilo

Para estimar el valor de la menor energía con una precisión razonable, debemos estimar la energía del estado no plegado y plegado de la proteína con una precisión que no es actualmente alcanzable. Calcular la energía de cada posible conformación de una cadena de proteína para seleccionar el estado de más baja energía es imposible, no solo por el enorme número de conformaciones sino también, por la carencia de precisión en los cálculos. Sin embargo, existen métodos para obtener una aproximación estimada de la energía asociada a una determinada conformación de una proteína. Aunque dichos métodos también carecen de precisión para poder predecir una estructura, podrían ser de mucho uso en muchas aplicaciones.

El problema de explorar el espacio conformacional es muy complejo. Estadísticamente, el tiempo promedio del proceso de plegado debería ser comparable con el tiempo que necesita intentar cada posible conformación. Un simple cálculo muestra que si esto fuera cierto, el tiempo de vida del universo no sería suficiente para que una proteína pueda plegarse.

Paradoja de Levinthal!

Lo primero que debemos hacer es entender, cómo una proteína puede alcanzar su estructura en cuestión de segundos a milisegundos.

Una aproximación a dicho entendimiento es encontrar las energías del sistema haciendo uso de la mecánica molecular.

- ▶ Dos átomos enlazados covalentemente pueden ser vistos como dos clásicas partículas, con una masa, unidas por un material elástico el cual tiene una constante de elasticidad que es proporcional a la fuerza de ese enlace.

$$F_r = K_r(r - r_0)$$

- ▶ Para estimar la energía asociada con otros enlaces que tengan un ángulo θ en cada posible tripleta de átomos o para ángulos dihedrales ϕ entre posibles cuádrupletas, comúnmente se usa

$$K(\theta - \theta_2)^2 \text{ y } \frac{1}{2}K_\phi(1 + \cos n\phi)$$

- ▶ Para interacciones no enlazadas el modelo es un poco más complicado. Puede usarse la aproximación de VanderWaals:

$$A_{ij}r_{ij}^{-12} - B_{ij}r_{ij}^{-6}$$

donde r_{ij} es la distancia entre los átomos.

- ▶ Para interacciones electrostáticas en los enlaces de hidrógeno se usa:

$$C_{ij}r_{ij}^{-12} - D_{ij}r_{ij}^{-10}$$

- En general, la fórmula que se usa para calcular la energía dado un secuencia de átomos en una proteína es:

$$\begin{aligned}
 E = & \sum_{\text{bonds}} K_r(r - r_0) + \sum_{\text{angles}} K(\theta - \theta_2)^2 + \\
 & \sum_{\text{dihedralangles}} \int \frac{1}{2} K_\phi (1 + \cos n\phi) + \sum_i \sum_{j < i} A_{ij} r_{ij}^{-12} - B_{ij} r_{ij}^{-6} + \\
 & \sum_i \sum_{j < i} C_{ij} r_{ij}^{-12} - D_{ij} r_{ij}^{-10}
 \end{aligned}$$

Si la anterior ecuación fuera lo suficientemente precisa para evaluar correctamente la energía de una determinada estructura de una proteína, podríamos buscar la conformación de proteínas que minimiza el valor de la energía de esta conformación y esta conformación debería corresponder a la estructura de la proteína nativa. Sin embargo, el cálculo de la energía no es preciso debido a la inevitable introducción de aproximaciones de un problema mecánico. Los valores obtenidos de energía por las ecuaciones de este tipo sólo puede ser utilizado para distinguir entre arreglos correctos y arreglos incorrectos de proteínas, cuando el arreglo incorrecto es muy diferente del correcto. La ecuación anterior se puede utilizar en combinación con técnicas de optimización para la búsqueda de espacio de conformaciones de manera más eficiente. Algoritmos como Monte Carlo, simulación Annealing, dinámica molecular, algoritmos genéticos que han sido aplicados en combinación con diferentes métodos y también métodos heurísticos.

Dada una función de energía, ya sea por estadística o por mecánica molecular, se debe encontrar el espacio conformacional de una proteína para encontrar el mínimo de la función de energía. Un simple algoritmo de minimización cambia las coordenadas de los átomos y, generalmente, una nueva conformación es aceptada solo si su energía es menor a la anterior.

Métodos estocástico para la exploración del espacio conformacional no intenta simular un proceso físico, sino más bien estudiar al azar las conformaciones del espacio. el método Monte Carlo usa el métodos de movidas aleatorias; Por ejemplo, un ángulo en la estructura de proteínas es cambiado al azar y la energía resultante de la conformación se calcula. Si el nuevo estado tiene una menor energía que el anterior, la movida se mantiene. Sin embargo, las conformaciones que dan una energía más alta no son necesariamente descartadas. Ellas pueden ser aceptadas con una probabilidad que depende de un parámetro preseleccionado, llamado temperatura, y sobre la diferencia entre su energía y la energía del estado anterior. Cuanto menor sea el aumento de la energía, más alta es la probabilidad de aceptar la movida.

$$P = \min\left(1, \exp\frac{-(E_{new} - E_{previous})}{kT}\right)$$

Los algoritmos genéticos son algoritmos de optimización basados en los conceptos de evolución. Comienzan a partir de un conjunto de posibles soluciones, como un conjunto de la conformación de una estructura de la proteína. De manera similar a lo que ocurre en la evolución molecular, cada elemento del conjunto tiene una cierta probabilidad de ser mutado (por ejemplo, al azar un ángulo diédrico cambió) o fue "Cruzado" con otro elemento del conjunto (por ejemplo, la adhesión a la primera parte de una conformación con el segundo de otro, y viceversa). Un valor de fitness se evalúa para cada uno de los miembros de la población, y un nuevo conjunto (generación) se construye. La distribución de las posibles soluciones en este nuevo conjunto depende de la aptitud de cada uno de los miembros. Los miembros con mayor aptitud física será más representados que aquellos con menor aptitud, y el proceso se repetirá. La función objetivo puede ser la energía mecánica molecular de cada conformación

Ninguno de los métodos descritos anteriormente es capaz de deducir la estructura tridimensional de una proteína a partir de su secuencia, esto se debe a las limitaciones del cálculo aproximado que se hace (su poca precisión) y porque la capacidad de explorar el espacio conformacional es insuficiente. Sin embargo, el gran conjunto de estructuras de proteínas disponibles nos da una manera de encontrar una solución al problema sobre la base de reglas heurísticas que podemos aprender de la serie de ejemplos resueltos disponibles (es decir, de las proteínas que se conoce tanto la secuencia y la estructura). Si pudiéramos utilizar estos métodos heurísticos para predecir con precisión satisfactoria la estructura de cada proteína en el universo, no sería necesario comprender la naturaleza del proceso de plegado.