

ESTRUCTURA Y ESTABILIDAD DE LAS PROTEÍNAS ¹

Ana Rosa Viguera

En el avance de la proteómica importará descubrir la relación entre las secuencias de aminoácidos, la estructura tridimensional y la función de las proteínas.

Las proteínas constituyen más de la mitad del peso seco de las células. Si el ADN almacena la información necesaria para construir una célula, corresponde a las proteínas proporcionar los instrumentos moleculares para su ejecución. La función biológica de una proteína viene determinada por su composición química y su estructura espacial.

Los aminoácidos constituyen los componentes elementales de las proteínas. Se unen entre sí mediante enlaces amida o peptídicos. Existen 20 aminoácidos diferentes. El tamaño de una proteína varía desde decenas a miles de aminoácidos; el promedio en *Escherichia coli* es de 317 aminoácidos y de 496 en levadura. Cada proteína tiene una conformación específica. Si quisiéramos detallar esta estructura habría que determinar la posición en el espacio de cada uno de sus átomos.

La cadena principal de la proteína se compone de la repetición del grupo peptídico, cuya estructura química le confiere una determinada tendencia. Dos ángulos diedros, phi (Φ) y psi (Ψ), por cada aminoácido definen a la cadena principal. Estos ángulos diedros toman valores que se hallan acotados entre ciertos límites. Pese a estas restricciones de orden estérico, muchos de los enlaces de una

cadena polipeptídica larga permiten la rotación libre de los átomos, razón de la sorprendente flexibilidad de la cadena principal.

Ahora bien, en condiciones fisiológicas la mayoría de las proteínas naturales se pliegan y forman una estructura compacta. Ello se debe a que las cadenas principales y laterales de los aminoácidos se asocian entre sí y con el disolvente para formar enlaces no covalentes. En las proteínas naturales, el plegamiento parece estar optimizado; sólo hay una conformación particularmente estable. No ocurre así con cualquier polipéptido. De las posibles secuencias de una determinada longitud (20^{100} en un polipéptido de 100 residuos), se estima que apenas un escaso 2 por ciento podría plegarse en una conformación regular. La mayoría engendraría especies con múltiples conformaciones o al azar, o especies agregadas.

En las conformaciones estables de proteínas solubles en disolución acuosa las cadenas laterales hidrofóbicas tienden a situarse en el interior de las moléculas, en tanto que las cadenas laterales polares prefieren alojarse cerca de la superficie. Este es el llamado efecto hidrofóbico que se considera responsable principal del colapso de la proteína en formas compactas.

La especificidad la proporcionan principalmente, sin embargo, los puentes de

¹Investigación y Ciencia, marzo 2003.70-77 págs.

hidrógeno entre grupos polares de la cadena principal. Estos y las tendencias Φ y Ψ se consideran parte esencial en la explicación de las estructuras de proteínas. La investigación sobre el plegamiento gira en torno al balance entre estos y otros tipos de fuerzas (puentes disulfuro, interacciones de van der Waals, etc.) en el establecimiento de la estructura nativa, su estabilidad y generación.

Estructura plegada y estructuras desplegadas

Entre el estado nativo y el desnaturalizado de una proteína existe una diferencia mínima en lo concerniente a la energía total. El estado desplegado tiene una entropía más favorable (lo componen múltiples conformaciones); la única conformación del estado plegado debe compensar dicha ventaja con diversas interacciones químicas. Por su parte, la estructura nativa posee una energía interna mucho menor que cada una de las conformaciones del estado desnaturalizado, pero éstas son más numerosas. La estabilidad observada de la estructura proteica es el resultado de una diferencia de 5-15 kcal/mol, entre las energías totales de los estados plegado y desplegado.

Se ha observado que existe degeneración en el código de plegamiento; a la misma estructura nativa puede llegarse desde secuencias diferentes. Durante el plegamiento se busca una estructura única nativa en el espacio conformacional para una secuencia dada, mientras que la evolución biológica busca en el espacio de secuencia una proteína con una función particular. Ambos espacios son inmensos, lo que dificulta su exploración y entendimiento.

Plegamiento *in vitro*

Desde el experimento clásico de Christian Anfinsen en 1961, se sabe que una proteína está capacitada para plegarse por sí misma en disolución. Durante los últimos años, se han registrado avances teóricos y técnicos en la dilucidación de las relaciones entre la secuencia de aminoácidos y la estructura terciaria. Se busca, ahora, un modelo que permita cuantificar las predicciones acerca de la secuencia de aminoácidos, la topología de la cadena, el pH, la concentración de sal y la temperatura, en la cinética y termodinámica del proceso de plegamiento.

El punto de partida de los experimentos de replegamiento es la proteína desnaturalizada en altas concentraciones de agente desnaturalizante (urea y cloruro de guanidinio son los más comunes). Los estados desnaturalizados son ensamblajes dinámicos de especies de energía similar que se interconvierten rápidamente.

A partir de medidas de volumen hidrodinámico y métodos espectroscópicos, se ha obtenido información de baja resolución acerca de las propiedades conformacionales de diversos estados desnaturalizados. Sabemos ya que puede persistir una estructura residual significativa en el estado desnaturalizado, estructura que depende de la forma de desnaturalización.

Los estados desnaturalizados son especies dinámicas. A pesar de encontrarse desplegados, sin estructura secundaria, retienen todavía significativas variaciones en las preferencias de los aminoácidos, que se expresan en la persistencia de interacciones locales nativas, con interacciones entre cadenas laterales e interacciones hidrofóbicas entre residuos.

Puesto que la cantidad y el tipo de estructura residual observada en los estados desnaturalizados dependen de las condiciones de la disolución, las propiedades del estado desnaturalizado subsiguientes a la dilución de desnaturalizante difieren de las anteriores. El estado desnaturalizado en ausencia de desnaturalizantes suele considerarse la referencia más adecuada.

Estados intermedios.

En 1968, Levinthal propuso que, si la proteína debía explorar al azar las conformaciones posibles hasta dar con el estado nativo, la búsqueda resultaría interminable. La evolución ha encontrado una solución eficaz. ¿Cuál? Creíase en un comienzo que las moléculas atravesarían estados parcialmente estructurados. Contando con nuevas vías específicas, bastaría el muestreo de una pequeña región del espacio conformacional.

Esta visión parecía corroborarse en experimentos que evidenciaban la existencia de intermediarios. A pesar de la rapidez y la cooperatividad del plegamiento, la mayoría de las proteínas engendraban intermediarios en una fase precoz del plegamiento. Estas especies variaban en sus propiedades conformacionales y estabildades.

Pero el mecanismo de formación de intermediarios no acaba de entenderse en la mayoría de las proteínas, habida cuenta de la notable rapidez con que se forman estas especies (2 – 5 milisegundos). Para las proteínas con intermediarios, la transición entre el intermediario y el estado nativo limita la velocidad de plegamiento. En algunas proteínas, los intermediarios parecen ser especies con estructura nativa. A esta clase de intermediarios pertenecen los glóbulos fundidos, en los que puede darse un carácter nativo desde el punto de vista de estructura secundaria y de morfología global; se aprecia, sin embargo, bastante desorden en las cadenas laterales.

La observación experimental de intermediarios parcialmente plegados dio lugar al modelo "andamio", según el cual el estado nativo se forma de una manera jerárquica en un camino lineal que incluye varios estados de transición consecutivos. En este modelo, se requieren intermediarios parcialmente plegados que dirijan la cadena al estado final.

En otras proteínas, se pueden acumular intermediarios plegados de forma incorrecta o atrapados. Abundan las pruebas de que los estados intermedios de proteínas plegadas con cinética de múltiples estados son trampas cinéticas. Por ejemplo, a pH mayor que 6 el plegamiento del citocromo c es lento, como resultado de la formación de una estructura plegada de forma incorrecta que incluye una interacción no nativa del grupo hemo.

A pesar de que el carácter de la estructura que se encuentra en muchos intermediarios sea esencialmente nativa, todavía pueden contener ciertos aspectos no nativos que sean eventualmente responsables de su acumulación, retardando la adquisición de la estructura final. Se ha demostrado que los intermediarios estables no son un prerrequisito para el plegamiento eficaz y rápido de cadenas polipeptídicas. Muchas proteínas pequeñas han mostrado plegarse con la cinética más simple de dos estados (ningún intermediario se acumula en el equilibrio o transitoriamente).

Las diferencias más significativas entre las proteínas que se pliegan a través de intermediarios y las que lo hacen directamente, son la longitud de la cadena y la estabilidad. En general, las proteínas que se pliegan por un mecanismo de dos estados constan de menos de 110 aminoácidos. De acuerdo con ciertos trabajos, puede provocarse un cambio desde una cinética de tres a dos estados si se desestabiliza el estado nativo. Asimismo se han encontrado cambios de mecanismos de dos a tres estados cuando se estabiliza o elonga una proteína pequeña.

El estado de transición

A menudo se compara el plegamiento proteico con una reacción química unimolecular, en la que el reactivo (la proteína desplegada) se convierte en el producto (la proteína plegada). Para las proteínas pequeñas monodominio, se

obtienen cinéticas de plegamiento monoexponenciales en la mayoría de los casos (las moléculas desnaturalizadas tienen la misma probabilidad de alcanzar el estado nativo), de forma análoga a las reacciones sencillas con estados de alta energía.

Las reacciones químicas unimoleculares están gobernadas por un paso simple limitante de la velocidad, cuando el sistema atraviesa el "estado de transición". En principio este concepto se introdujo para explicar por qué un determinado proceso no ocurre a la máxima velocidad. El estado de transición de una reacción química sencilla, es una única conformación con energía interna desfavorable, que representa la barrera principal entre reactivos y productos.

Para el plegamiento podría haber alguna conformación particular que actuase como "cuello de botella". Una conformación de esta índole tendría una energía libre interna alta y sería adoptada raramente por la molécula. Sin embargo, a diferencia de las reacciones químicas el plegamiento de proteínas está gobernado por la entropía: existen muchas conformaciones que se corresponden con el mismo grado de avance de la reacción.

A diferencia de una reacción química sencilla, en la que la barrera de energía libre representa la contribución de una única conformación, la barrera para el plegamiento es una barrera de energía libre total, dominada por la entropía conformacional. Esta barrera de origen entrópico aparece como resultado de la decreciente flexibilidad de la cadena peptídica según se va plegando.

Los estados de transición de plegamiento sólo pueden determinarse estudiando las cinéticas de plegamiento y desplegado. Normalmente el replegamiento se inicia por un cambio rápido del disolvente desnaturante (bajo o alto pH o alta concentración de urea) a condiciones nativas. El proceso de plegamiento subsiguiente se sigue por una

técnica espectroscópica o por resonancia magnética nuclear (RMN). Se ha utilizado también la espectrometría de masas y, en el caso de reacciones de plegamiento muy rápidas, la transferencia de electrones.

El estudio de la dependencia con la temperatura de las velocidades de plegamiento y desplegado aporta información acerca de la naturaleza del estado de transición. Mediante este método, podemos calcular los valores del cambio en entalpía (ΔH), entropía (ΔS) y capacidad calorífica (ΔC_p) entre el estado desnaturalizado, el estado de transición y el estado plegado. En general estos métodos limitan su eficacia a la determinación de propiedades promedio y no proporcionan información específica sobre la estructura del estado de transición. Con las técnicas de ingeniería de proteínas sí podemos hacernos una imagen detallada de la estructura y energía del estado de transición.

Los cambios ejercidos en las cinéticas de replegamiento y desplegado de una proteína tras una mutación sirven para definir un valor Φ como $(\Delta\Delta G_{U-\ddagger} / \Delta\Delta G_{U-F})$. El parámetro Φ_F es una medida de la fracción de energía libre que se pierde en el estado de transición con relación al estado nativo, cuando se realiza una mutación desestabilizante (o de la energía que se gana en una mutación que estabiliza la proteína).

Con el fin de obtener la máxima resolución, la mayoría de las mutaciones analizadas son individuales (sustitución de un aminoácido por otro). Los valores de Φ_F indican el grado de formación de estructura en el estado de transición; oscilan entre 0, que nos dice que existe poca estructura en la región que rodea al punto de la mutación en el estado de transición, y 1, cuando la región está altamente organizada: los valores intermedios revelan la formación de estructura proteica.

Se han analizado las propiedades conformacionales del estado de transición de varias proteínas. Los resultados obtenidos para más de 150 de mutantes del inhibidor de la quimotripsina C12, de 62 residuos, ha permitido proponer el mecanismo de

nucleación-condensación, que describe un conjunto de conformaciones en el estado de transición compacto y que contiene un núcleo deslocalizado y extendido estabilizado por numerosas interacciones parcialmente formadas. Se producen el desplome y la consolidación de la estructura después sobre el esqueleto proporcionado por el núcleo.

Este comportamiento de distribuciones deslocalizadas de valores de Φ_F intermedios se ha ratificado en otras proteínas que se pliegan por un mecanismo de dos estadios. Otras proteínas mayores, como la barnasa, de 110 aminoácidos, pueden tener interacciones totalmente formadas en el estado de transición, de acuerdo con modelos que contemplan un plegamiento secuencial por dominios.

De lo anterior se desprende que las proteínas difieren entre sí en el mecanismo exploratorio del paisaje de energía de plegamiento. Además, los cálculos teóricos y los datos experimentales indican que la forma del paisaje de plegamiento puede cambiar según las condiciones empleadas, la secuencia precisa de aminoácidos y la topología de la cadena.

Aproximaciones teóricas

Simular el plegamiento en el ordenador es tarea ardua y complicada. Exige tomar en cuenta varios supuestos: considerar un número inmenso de conformaciones, 2^N para una proteína de N aminoácidos (como mínimo dos variables por residuo, si atendemos a los ángulos diedros, que pueden tomar infinidad de valores); evaluar su energía, para lo que no existen datos contrastados experimentalmente; encontrar una coordenada de reacción válida (por ejemplo el radio hidrodinámico de giro); y ver cómo, tras conocer el estado nativo y proponer un potencial válido al que se sometan las moléculas desplegadas, intentar reproducir datos de plegamiento y valores de Φ_F experimentales. Aunque

existen simulaciones de dinámica molecular con representaciones de todos los átomos de una proteína, éstas pueden comportarse por tiempos irrelevantes comparados con los necesarios para evaluar el plegamiento.

Una alternativa a las descripciones atómicas es introducir representaciones simplificadas de la estructura de una proteína. En estos estudios, los aminoácidos se representan como una o varias unidades de interacción que pueden también tener algunos grados de libertad. Importa, sobre todo, determinar la función de energía apropiada.

A largo plazo debe comprobarse sistemáticamente cuál de los términos de la función resulta determinante para estabilizar los mapas nativos y sean, así, los de mínima energía.

Los modelos mecánico-estadísticos de plegamiento representan la cadena en sus múltiples conformaciones y sus interacciones microscópicas. En este nuevo planteamiento se han sustituido los diagramas de reacción sencillos, donde se representaban las energías libres de un número pequeño de macroestados, por los paisajes de energía, donde se representan las energías libres en función de los grados de libertad. Un microestado es una conformación particular de la cadena. El estado desnaturalizado, el estado de transición o los intermediarios de plegamiento son macroestados (colecciones de multitud de conformaciones individuales).

De acuerdo con los principios de la termodinámica, si un sistema tiene n grados de libertad $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_n]$, el estado estable del sistema puede encontrarse determinado por el conjunto de valores $\gamma^* = [\gamma_1^*, \gamma_2^*, \dots, \gamma_n^*]$ quedan los mínimos valores de la función de energía $F(\gamma) = F(\gamma_1, \gamma_2, \dots, \gamma_n)$, cuando se exploran todos los posibles valores de γ . Estas funciones $F(\gamma)$ se denominan paisajes de energía. La distinción entre los paisajes de energía y un diagrama de coordenada de reacción es una distinción entre la aproximación macroscópica y la microscópica.

Un microestado es un punto en el paisaje de energía $F_{\text{micro}} = F(\gamma)$, o energía libre interna. Un macroestado posee una energía $F_{\text{macro}} = F(\xi)$, donde ξ es una magnitud escalar, tal como una coordenada de reacción o una variable de progreso. Un valor dado de ξ representa un conjunto particular de conformaciones microscópicas. Hay muchas conformaciones de alta energía y muy pocas de baja energía, de ahí la forma de embudo de los paisajes de energía. La proteína desnaturalizada (en la zona superior del embudo) se parece a una cadena al azar, en la que las interacciones locales dominan el comportamiento conformacional.

Puesto que las proteínas tienen muchos grados de libertad, hay en principio numerosas coordenadas diferentes posibles para describir el progreso del plegamiento. Si se toma como variable de progreso de la reacción el número de contactos hidrofóbicos, la densidad de estado será una medida del número de microestados que componen un macroestado. El experimento de plegamiento simulado en el ordenador comienza con múltiples conformaciones con 0 contactos. Cuanto mayor es el número de contactos, menos son las conformaciones. Cuando se ha formado la estructura nativa, sólo una conformación contiene todos los contactos.

Para una determinada conformación, $F_{\text{micro}}(\xi)$ es una energía libre, pero no una energía libre total, sino la energía libre de una cadena, mientras que $F_{\text{macro}}(\xi)$ es la energía libre de un conjunto de conformaciones con ciertas características macroscópicas, como el estado desnaturalizado, intermediario, estado de transición. $F_{\text{macro}}(\xi)$, que incluye una entropía conformacional, es función de la variable ξ ; por lo tanto, se presta a una representación bidimensional de la energía libre de plegamiento frente a la coordenada de reacción ξ , que es el

diagrama tradicional. Sin embargo, $F_{\text{micro}}(\xi)$ es el paisaje de energía y depende de muchos gradientes de libertad.

Este tipo de aproximaciones teóricas al problema de plegamiento han permitido observar, por ejemplo, que el balance entre la energía interna y la entropía no es perfecto durante el plegamiento (la entropía se pierde antes) y los primeros pasos del plegamiento conllevarían una subida en energía libre total hasta un máximo que se identifica con el estado de transición. Esta barrera de potencial sería la responsable última de la velocidad de plegamiento. Se confirma también que la topología es un determinante importante de la vía de plegamiento. Asimismo, modelos teóricos sencillos han conseguido reproducir con relativa precisión parámetros experimentales.

Plegamiento en vivo y enfermedad

A pesar de que la mayoría de las cadenas proteicas muestran una elevada proporción de éxito plegamiento correcto dentro de las células, algunas proteínas se muestran proclives a la agregación; al eludir los mecanismos de plegamiento, provocan enfermedades.

Los estados intermedios de algunas cadenas exponen zonas hidrofóbicas, que favorecen la agregación entre cadenas idénticas o parecidas. En la célula, las chaperoninas contribuyen a prevenir la agregación de las proteínas recién sintetizadas. Se sabe que una parte notable de la cadena que se está sintetizando se asocia con miembros de la familia de Hsp70. Las chaperoninas cilíndricas actúan posteriormente y proporcionan un ambiente aislado para el plegamiento de una pequeña fracción (aproximadamente 10% de todas las proteínas citosólicas).

En el medio intracelular existe una elevada concentración macromolecular, que incide en el plegamiento de proteínas. El total de concentración de proteínas y ácidos nucleicos en el interior de la célula es tan alto (340 g/l

en el caso de *E. coli*), que cerca del 30% del volumen no está disponible para otras macromoléculas. La teoría predice dos consecuencias principales de tal grado de concentración: la difusión se verá reducida entre 3 y 10 veces y las constantes de asociación de las macromoléculas que interaccionan entre sí se incrementarán en uno o dos órdenes de magnitud. De ello se deriva que la tendencia a la agregación se reforzará con dicha concentración; dentro de este grupo se incluye las que dan lugar a enfermedades relacionadas con amiloides.

Patologías sin aparente relación —la enfermedad de Alzheimer, la fibrosis cística, la enfermedad de las vacas locas, una forma heredada de enfisema y numerosos tumores—resultan del plegamiento incorrecto de proteínas. El autoensamblaje anormal comporta deposición de material proteináceo en agregados insolubles ordenados y fibras de amiloide. Se han identificado cerca de veinte precursores amiloidogénicos. Se sospecha que en intermediarios desplegados o parcialmente plegados podría encerrarse la causa común de los desórdenes asociados a amiloides.

Estructura tridimensional

La identificación de la estructura primaria de una proteína a través de la secuenciación de ADN es inmediata (código genético). Pero la adquisición de datos tridimensionales resulta todavía lenta y se circunscribe a proteínas que cristalizan en una forma adecuada o pueden resolverse por RMN en disolución. Se requieren, pues, algoritmos que traduzcan la información lineal en espacial. La predicción de la estructura tridimensional de una proteína desde su estructura primaria abarca un doble aspecto: describir el estado nativo y conocer la trayectoria de plegamiento. Ambas tareas trascienden las posibilidades de los métodos actuales de computación.

Si podemos abordar la estructura proteica a partir de las energías de interacción para perturbaciones pequeñas de estructuras tridimensionales conocidas. Cuando conocemos con suficiente precisión una determinada estructura por cristalografía o por RMN, podemos entonces refinar las coordenadas atómicas mediante cálculos con funciones de energía; pero estos métodos no son aplicables si se carece de una estructura aproximada.

Bibliografía complementaria.

PROTEIN MISFOLDING, EVOLUTION AND DISEASE. C. M. Dobson, en *Trends Biochem. Sci.* vol. 24, págs. 329-332; 1999.

A SURPRISING SIMPLICITY TO PROTEIN FOLDING. D. Baker, en *Nature*, vol. 405, págs. 39-42; 2000.

CRITICAL ASSESSMENT OF METHODS OF PROTEIN STRUCTURE PREDICTION: (CASP): Round IV. J. Moult, K. Fidelis, A. Zemla y T. Hubbard, en *Proteins (Suppl.)*; vol. 45 (S5), 2001.