

Capítulo 11

Modelos simplificados de plegamiento de proteínas

Alberto Pascual-García

11.1. Introducción

En la presente sección vamos a presentar algunas aproximaciones computacionales orientadas a comprender los fundamentos del plegamiento de proteínas. *Por plegamiento entendemos el proceso espontáneo por el cual la proteína alcanza su conformación funcional*, es decir, su conformación nativa. Este proceso es espontáneo, por tanto termodinámicamente favorable, si se dan las condiciones apropiadas en el ambiente en el que está inmersa la proteína, tales como la concentración de sales o la temperatura. Además hay que tener en cuenta que en el medio natural puede ser necesaria la mediación de chaperonas, que evitan agregaciones o conformaciones subóptimas. También existe un creciente interés por regiones de las proteínas, que pueden incluso representar incluso una fracción muy significativa de la longitud total, cuya estructura no es posible resolver experimentalmente pero que pueden tener una relevancia funcional (las llamadas regiones intrínsecamente desordenadas) que serán tratadas en el Capítulo 13.

Desde el punto de vista biológico la importancia de conocer el plegamiento de una proteína reside fundamentalmente en la necesidad de conocer o profundizar en su función, si bien es un proceso interesante desde otros puntos de vista como por ejemplo el físico, ya que es un claro ejemplo de proceso autoorganizativo. Existen esencialmente *dos problemas de plegamiento*: 1) el que se centra en *entender el proceso* de plegamiento en sí, tanto dinámica como termodinámicamente y 2) el orientado a la *predicción de la estructura* de las proteínas. Nosotros discutiremos aquí el primer punto, dentro de una línea argumental que dará especial énfasis a los aspectos evolutivos, lo que nos permitirá ver la relación entre los dos problemas.

La literatura alrededor del problema de plegamiento es ingente y, dada la limitación de espacio de la que disponemos, es inevitable centrarse en únicamente algún tipo de modelos. Nosotros vamos a considerar una representación muy simplificada de las proteínas (ver la Figura 11.1, representación más simplificada). La justificación de esta simplificación tan fuerte reside en que podremos tratar el problema del plegamiento desde una perspectiva que nos permitirá sacar algunas conclusiones sobre el rol de la evolución (que se ampliará en el Capítulo 12) y con la que habremos establecido las bases para comprender también el problema de la modelización de estructuras de proteínas. Por tanto, los

modelos de plegamiento más detallado no serán tratados exhaustivamente en el libro, si bien creemos que el lector interesado habrá obtenido las bases necesarias para una aproximación autónoma a dichos problemas si combina los resultados de esta sección con los que puede encontrar en el capítulo de dinámica molecular y modelización de estructura de proteínas. En particular, dentro de los modelos simplificados que vamos a discutir, hemos seleccionado modelos que nos permitirán alcanzar una serie de objetivos, alrededor de los cuales está organizado el capítulo y que explicamos a continuación.

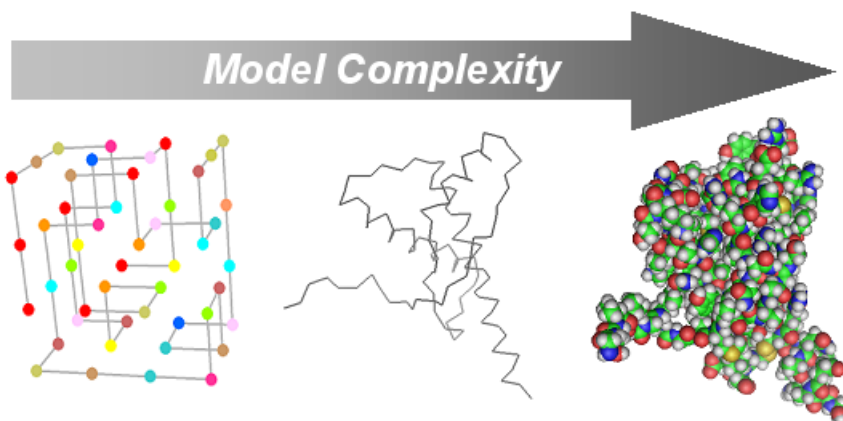


Figura 11.1: Complejidad de la representación en modelos de estructura de proteínas. Un nivel mayor de detalle como en los modelos que consideran todos los átomos en dinámica molecular (derecha) nos permite responder a preguntas concretas pero el coste computacional es muy elevado. A medida que vamos reduciendo la complejidad de la representación, como en los modelos que consideran carbonos alpha fuera de la red en la dinámica molecular de Langevin (centro) o simplemente aminoácidos en un retículo (izquierda), el coste computacional se reduce y podemos afrontar preguntas más generales como los principios básicos del plegamiento de proteínas. Esta figura ha sido gentilmente cedida por la Dra. Patrícia Faísca ¹.

En la *primera sección*, hacemos una *introducción a resultados e ideas históricamente relevantes como la paradoja de Levinthal, el experimento de Anfinsen o el paisaje energético*. Los dos primeros puntos se refieren al problema del plegamiento en general, y será con el concepto de paisaje energético y el principio de mínima frustración cuando comencemos a introducir los modelos específicos que utilizaremos.

En la *segunda sección* presentaremos un *tratamiento analítico de dichos modelos simplificados de plegamiento*. Estos modelos nos permitirán hacer cierto tratamiento analítico sencillo en un problema que es muy complejo de atacar desde el punto de vista teórico. Se mostrarán algunos resultados teóricos que, aun siendo sencillos, serán accesibles a alumnos con cierto nivel en matemáticas y física, y serán probablemente difíciles para aquéllos que no dispongan de esta formación, que podrán prescindir de esta sección. El objetivo es que los alumnos interesados en la formalización tengan la oportunidad de encontrar desarrollos orientados a tratar conceptos relacionados con el problema de plegamiento, difíciles de encontrar de manera asequible para estudiantes de grado o máster en la literatura. Esto nos ha llevado a seleccionar modelos que se considerarían antiguos en este área. Este criterio de selección tiene como parte negativa que no permite conocer el estado actual del arte. Pero tiene como valor añadido el comenzar a aprender los conceptos en el orden en el que históricamente han sido explorados, lo cual permite fundamentar unas bases mínimas para continuar autónomamente el aprendizaje.

La *tercera sección* trata *ejemplos concretos* utilizando el mismo tipo de modelos. Los ejemplos que se discuten han sido seleccionados siguiendo dos criterios. El primero es que están basados en artículos

¹*Protein Folding Group*, Universidad de Lisboa. [http://www.ciul.ul.pt/~sim\\$patnev/index.html](http://www.ciul.ul.pt/~sim$patnev/index.html)

científicos asequibles, lo cual creemos que es importante pues existe típicamente una falta de acercamiento de los estudiantes a la literatura científica. Por otra parte son ejemplos que tratan los conceptos discutidos anteriormente y hacen especial énfasis alrededor de argumentos evolutivos, lo cual permitirá al lector relacionar conceptos con otros capítulos del libro.

11.2. Conceptos básicos

11.2.1. La paradoja de Levinthal y el experimento de Anfinsen

Para comenzar, vamos a intentar evaluar la complejidad del problema que estamos afrontando. *Imaginemos que la conformación en la cual la proteína es funcional es única*, es decir, hay un único conjunto de coordenadas describiendo las posiciones de los átomos de la proteína en el espacio. Esta asunción es muy fuerte, porque obviamente la proteína estará en movimiento por su interacción con el medio. Pero sí es realista pensar que *la proteína se encuentra estadísticamente en un conjunto de conformaciones*, y que estas conformaciones son esencialmente las mismas si utilizamos medidas globales de “grano grueso” como el número de contactos (ver Capítulo 10). Estaríamos promediando (e incluso en algunos casos despreciando) magnitudes más detalladas como la posición de las cadenas laterales o movimientos de cuerpo rígido de elementos de estructura secundaria. No tenemos que perder de vista en ningún momento que, el no considerar *a priori* estas diferencias “en el detalle”, no implica que no puedan tener importancia en el proceso analizado. Así que debemos considerar este hecho cuando extraigamos conclusiones de cualquier análisis obtenido únicamente considerando una *descripción de “grano grueso”*. Sin embargo vamos a ver que, incluso considerando una descripción simplificada, la complejidad sigue siendo enorme.

Si consideráramos una proteína de unos 100 aminoácidos y cada aminoácido tuviera, por ejemplo, dos conformaciones posibles, explorar todas las conformaciones implicaría visitar $2^{100} \sim 10^{30}$ estados posibles. Las proteínas muestrean estados alrededor de su estado nativo a un ritmo cuyo orden de magnitud oscila entre $0,1ps$ y $10ns$, pero ni siquiera esa velocidad es suficiente para explorar todas las conformaciones en un tiempo biológicamente relevante. Por ejemplo, para poner en un contexto computacional el problema, supongamos que somos capaces de explorar del orden de 10^6 conformaciones en un día. Esto es el número de conformaciones que se explora en una dinámica molecular utilizando 24 procesadores estándar y un paso de integración típico de $0,01ps$ para generar $1ns$ de trayectoria. *Necesitaríamos entonces 10^{24} días en explorar todas las conformaciones*, ¡lo cual es bastante más que la edad del universo!

Estábamos asumiendo en el cálculo anterior que la configuración en la que la proteína es funcional es única (salvo fluctuaciones), o si existen varias configuraciones son muy pocas en relación al número total de configuraciones posibles. Esta asunción se fundamentó en los años 60 gracias a los experimentos que llevó a cabo Anfinsen y sus colaboradores [1] sobre la cinética de la ribonucleasa y se ha convenido en llamar la *hipótesis termodinámica*. En dichos experimentos monitorizaban las propiedades estructurales en un proceso en el que, partiendo de unas condiciones estables para el plegamiento, llevaban a la proteína hasta la desnaturalización para volver después de nuevo a las condiciones iniciales. Su observación fue que, tras la desnaturalización, las propiedades estructurales no habían cambiado, lo que parecía sugerir que la proteína recuperaba de nuevo su estado nativo. La importancia de estos experimentos, por los cuales recibió junto a Moore y Stein (quienes determinaron la secuencia de la enzima) el Premio Nobel de Química, reside en que implica que *toda la información necesaria para que la proteína se pliegue en su estado nativo se encuentra codificada en la secuencia de aminoácidos*.

Este hecho motivó rápidamente la pregunta de si sería entonces posible entonces predecir la estructura

nativa de una proteína a partir de su secuencia. Pero unos pocos años más tarde, en 1969, Levinthal [17] planteó un importante problema conceptual que hemos ya ilustrado en el pequeño ejercicio anterior: si el considerar los grados de libertad de la proteína nos lleva a que el número de configuraciones posibles es astronómicamente grande, *¿cómo puede una proteína plegarse en una escala de tiempo biológicamente relevante?* Como hemos visto, aun asumiendo que somos capaces de muestrear las configuraciones en un tiempo muy pequeño, encontrar el estado nativo con una búsqueda exhaustiva o al azar sería inviable. Si la velocidad de plegamiento típica es del orden de milisegundos e incluso del microsegundo [16], *¿cómo se resuelve la paradoja de Levinthal?*

Por un lado, la *paradoja de Levinthal* lo que nos dice es que, *en un escenario en el que todas las conformaciones son equiprobables una búsqueda al azar es simplemente inviable*. Si los estados fueran efectivamente equiprobables, implicaría que encontrar la estructura nativa requeriría que existieran algunos caminos muy concretos que guiaran a la proteína hacia ése estado nativo, es decir, que existiera un *control cinético* muy estricto. Por el otro, los experimentos de Anfinsen señalan que es suficiente con que la proteína esté bajo *control termodinámico*, es decir, el estado plegado representa el mínimo absoluto de la superficie de energía libre. La realidad es mucho más compleja y en el plegamiento tanto el control cinético como el termodinámico son relevantes [6]. Cuánta relevancia tiene el control cinético o el termodinámico en el plegamiento de una proteína dependerá para comenzar de la naturaleza de la proteína y de su contexto ambiental. Nuestro objetivo será intentar dar una aproximación a esta pregunta desde el punto de vista físico y evolutivo.

11.2.2. El paisaje energético y el principio de mínima frustración

Desde el punto de vista de la paradoja de Levinthal, lo que plantea es un conflicto entre que la proteína encuentre un *mínimo de energía* (es decir, está bajo *control termodinámico*) y el que lo alcance *suficientemente rápido* (lo que implicaría algún tipo de *control cinético*). Lo primero que tenemos que observar es que este conflicto surge de la asunción de que el estado nativo de una proteína es igual de probable que cualquier otro. Esto sería como intentar meter una canica de un milímetro cuadrado de radio en el agujero de un campo totalmente plano y del tamaño de dos veces la superficie de España. Y en esta imagen lo principal es darse cuenta de que la canica difícilmente encontrará el agujero mientras el campo sea plano, pero la situación puede cambiar mucho si la superficie tuviera pendiente. Si consideráramos la superficie de energía libre de una proteína, el que la superficie no sea plana significaría que las distintas configuraciones de la proteína no son equiprobables. Ahora bien, *si exigimos únicamente control cinético podríamos tener una superficie de energía en el que hay algún tipo de “canalización” que nos lleva del estado desnaturalizado al plegado*, como se muestra en la Figura 11.2.

Si la proteína al desnaturalizarse se encuentra en el punto *A*, encontrará la canalización que le lleva hasta el estado nativo, pero este escenario es entonces fuertemente dependiente de las condiciones iniciales. La desnaturalización de la proteína (y, en general, las conformaciones intermedias) no deben considerarse como una conformaciones únicas sino como conjuntos estadísticos (en inglés *ensemble*) repartidos por la superficie de energía para determinados valores de la misma. De este modo, la paradoja de Levinthal sería más bien un artefacto que surge al considerar un escenario como el que acabamos de describir. A principios de los ochenta, surgió una visión en la que lo que se considera es que *la superficie de la energía libre no es plana, sino que más bien tendría una forma de embudo* (del inglés *funnel*) como en la Figura 11.2. Esto implicaría que *no existen caminos concretos por los cuales la proteína deba pasar, y por tanto con un número pequeño de conformaciones intermedias*. Sino que más bien existirían conjuntos estadísticos suficientemente grandes para que el control cinético tenga lugar, por los cuales la proteína puede alcanzar el mínimo de energía, haciendo compatible el control cinético y termodinámico.

¿Qué particularidades podemos esperar de esta superficie? Por un lado, esperaríamos que tuviera

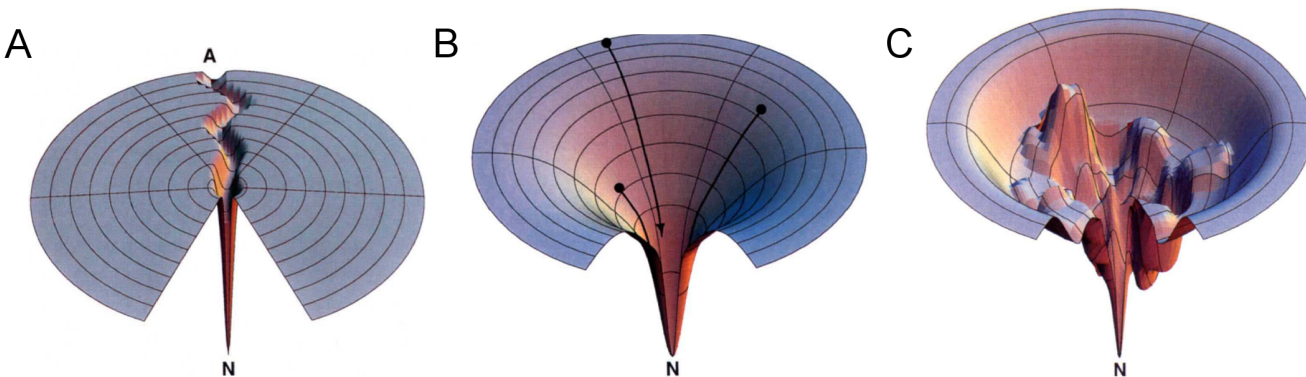


Figura 11.2: Ilustración de posibles paisajes energéticos. En el eje vertical se representa la energía libre sin considerar la entropía conformacional, una “energía libre interna”. Un punto en la superficie representaría una conformación, que a su vez contiene un número elevado de grados de libertad. Frente a un escenario en el que el paisaje energético es plano excepto para el estado nativo, como se asume en la paradoja de Levinthal que implicaría una búsqueda al azar entre conformaciones equiprobables, el estado nativo (N) podría ser alcanzado más rápidamente desde un estado desnaturalizado (A) si existiera un camino en la superficie energética preferente (A). Un paisaje en forma de embudo (B) permitiría considerar conjuntos estadísticos de conformaciones energéticamente más favorables. De este modo la proteína alcanzaría el mínimo de energía (control termodinámico) pero habría múltiples caminos. En un paisaje más realista (C) hay que considerar que la superficie no es suave porque existen fuentes de frustración (ver texto). Figuras extraídas de [8] con permiso del editor.

cierta forma suave de embudo, lo que permitiría “guiar” a la proteína hacia el estado nativo en una escala de tiempo razonable. Esto no significa que sea sencillo porque, al descender por el embudo, el espacio conformacional sigue siendo grande y encontraremos aún cuellos de botella en la cinética del plegamiento en regiones energéticamente cercanas al estado nativo precisamente porque el número de conformaciones accesibles es aún elevado, como veremos en uno de los ejemplos más adelante (una buena aclaración al respecto la encontramos en [14]).

Además, la complejidad de las interacciones de la proteína que se encuentra sometida a las fluctuaciones del medio nos lleva a pensar que habrá *múltiples mínimos locales debidos a conformaciones no nativas*, pero con cierta globularidad, que se forman durante el proceso y que darían lugar a un paisaje energético más bien rugoso como muestra la Figura 11.2. Estos mínimos locales serían “trampas” en las que la proteína se podría quedar atrapada si se dan las condiciones ambientales apropiadas.

Es lógico postular entonces que el paisaje energético de una proteína real sea algo intermedio entre una superficie suave y en forma de embudo y una superficie rugosa sin una pendiente global determinada. Este fue el razonamiento seguido por Bryngelson y Wolynes [4] para abrir una vía a modelos simplificados de la superficie energética de proteínas, basado en la bien conocida teoría física de los vidrios de espín. Según su planteamiento, un polímero con una secuencia al azar debería de presentar una superficie rugosa debido a que tendrá una fuente importante de *frustración*. La frustración debemos entenderla como un conflicto irresoluble entre distintas tendencias que tienen lugar en el sistema que es sometido a nuestro análisis, donde tendencia se puede entender en un sentido muy general. Por ejemplo, en cualquier clasificación diremos que está frustrada si no es posible realizarla sin ambigüedad. Veremos un ejemplo al respecto en el capítulo de evolución de estructura de proteínas, al intentar clasificar estructuras (Subsección 12.3.1). En el caso que nos ocupa, la frustración surge de que existen interacciones que entran en conflicto tanto por su naturaleza (determinada por la secuencia de la proteína) como por las restricciones de carácter geométrico. Si imaginamos por ejemplo una secuencia

al azar en donde consideramos únicamente una propiedad, por ejemplo tenemos residuos hidrofílicos e hidrofóbicos, no será posible encontrar una configuración en la que todos los residuos hidrofóbicos estén en contacto formando un núcleo compacto alrededor del cual se encuentren los residuos hidrofílicos, que sería la configuración más estable en medio acuoso. Así que en una secuencia al azar, esperaremos que la superficie de energía libre sea rugosa, y que haya *numerosos mínimos locales* en donde la proteína se queda atrapada en *configuraciones semicompactas*.

Por otro lado, los paisajes energéticos suaves a pesar de no tener trampas locales, parten de un gran número de estados de energía elevada para ir reduciéndose a unos pocos de baja energía. Diremos entonces que existe un “ *cuello de botella*” si existe alguna región de la superficie de energía libre en donde esta reducción es muy abrupta, y que podremos estimar con la disminución relativa del número de estados. Si este cuello de botella no es muy dramático, la proteína encontrará el estado nativo en un tiempo biológicamente relevante, como veremos en uno de los ejemplos discutidos. Este tipo de paisajes (superficie no muy rugosa y cuello de botella suficientemente ancho) será esperado entonces en sistemas en donde exista una alta *cooperatividad*, ya que implica una coordinación que reduce las posibles fuentes de frustración. La última característica que queremos resaltar de los sistemas que poseen un paisaje energético suave es que sus fases termodinámicas están determinadas por la temperatura, como veremos más adelante (ver Figura 11.7).

Bryngelson y Wolynes observaron que, si bien en proteínas reales debe existir cierta frustración, como se encuentra en polímeros cuya secuencia es al azar, esta frustración debería de ser mínima para que la cinética de la proteína sea lo suficientemente rápida para plegarse en el medio celular. Además, las estructuras globulares son muy compactas, lo que implica que la estructura secundaria y terciaria no entran en conflicto. Y además implicaría que, desde el punto de vista evolutivo, las secuencias de proteínas reales deberían haber sido seleccionadas para cumplir los requerimientos termodinámicos y cinéticos compatibles con su plegamiento. Llegamos entonces a un escenario en el que *las proteínas reales son un subconjunto de todas las cadenas polipeptídicas al azar que podríamos construir*, como ilustramos en la Figura 11.3.

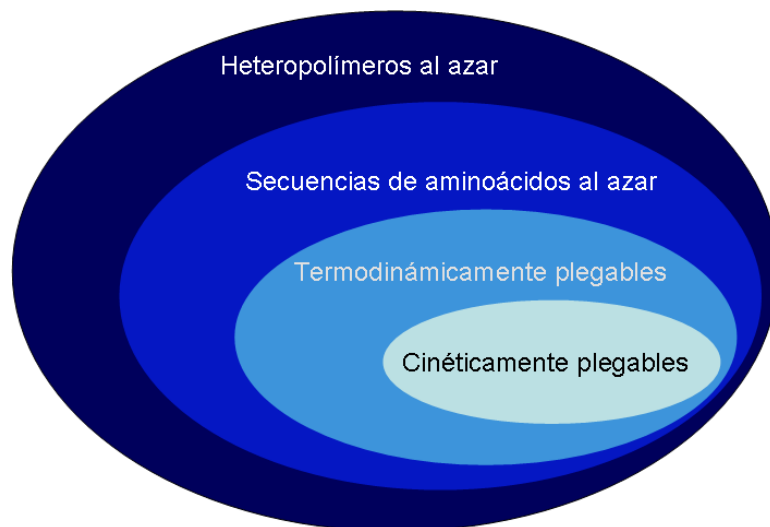


Figura 11.3: Partición del conjunto de heteropolímeros al azar en subgrupos en función de las restricciones que imponemos. El conjunto de polímeros basados en los aminoácidos encontrados en la naturaleza es un primer subgrupo dentro del cual encontraremos aquéllos termodinámicamente plegables, y a su vez aquéllos cinéticamente plegables a una temperatura relevante para la función biológica. Téngase en cuenta que las superficies relativas de los conjuntos no son realistas, ya que las diferencias entre cada grupo y el siguiente serían mucho mayores. Figura adaptada de [22].

11.3. Fundamentos teóricos

En la presente sección, vamos a presentar algunos *fundamentos teóricos* que, de manera natural, nos llevan a comprender formalmente los posibles ingredientes necesarios *para describir una superficie de energía* razonablemente realista. Si bien sería conveniente ciertos conocimientos en mecánica estadística para comprender esta sección no es necesario comprender los detalles matemáticos alrededor de los resultados, si bien mostramos algunos por completitud. Para los lectores que lo requieran, una primera introducción a la física estadística la encontrarán en [31] y un buen resumen pedagógico de la aplicación de la mecánica estadística a los modelos simples de plegamiento se encontrará en [24]. Es en la última sección en donde, dotados de los conocimientos adquiridos, abordaremos ejemplos computacionales concretos que ayudarán a afianzar los contenidos.

11.3.1. El paisaje energético del modelo de energía al azar

Es conveniente comenzar entendiendo qué propiedades esperaríamos si tuviéramos el modelo más simple posible. En general, el modelo más simple es aquél que explica un fenómeno observado como consecuencia del azar. Si introduyéramos nuestro modelo en el contexto estadístico de un *contraste de hipótesis*, podríamos de hecho considerar dicho modelo como aquél asociado a una *hipótesis nula* (modelo nulo). Este ejercicio es importante realizarlo siempre que se intente introducir un nuevo formalismo, ya que será la descripción más económica del proceso que estemos observando. Sus resultados servirán como referencia para evaluar la significatividad de los patrones que encontramos experimentalmente al contrastarlos con el modelo nulo, patrones que aceptaremos como *hipótesis alternativa* si conseguimos falsar el modelo nulo. La hipótesis alternativa contendrá algún ingrediente adicional al efecto del azar (cuya evaluación es probablemente la que motiva nuestro análisis).

En el caso que nos ocupa, podríamos considerar como aproximación de orden cero la aplicación del

modelo de energía al azar propuesto por Derrida [7] y aplicado por vez primera en polímeros por Bryngelson y Wolynes [4]. Vamos a examinar con este modelo qué superficie de energía obtenemos, para lo cual vamos a realizar asunciones muy fuertes. La superficie de energía la entendemos en este ejercicio como el espectro energético de los distintos estados en los que podemos encontrar a nuestra proteína problema.

Dada una cadena de monómeros, consideremos un *conjunto de microestados* (conformaciones estructurales), m_1, m_2, \dots, m_M con sus respectivas energías E_1, E_2, \dots, E_M . El número M de microestados es enorme como ya hemos visto. En particular, lo podemos estimar como $M = \gamma^N$, donde N es el *número de residuos* de la proteína y γ es el *número de configuraciones por residuo*. El valor de γ será menor que 5 cuando hagamos representaciones reducidas de la proteína, como las que hemos explicado, pero puede llegar a ser del orden de 10 si se consideran las cadenas laterales. En cualquiera de los casos, se irá reduciendo a medida que la proteína se vaya compactando.

¿Cómo estimaremos las energías? En primer lugar hay que tener en cuenta que, para una conformación m_i determinada su energía depende, no solamente de la estructura que la define, sino de lo que formalmente llamaremos una *realización del desorden*. En nuestro caso, una realización del desorden será una secuencia de aminoácidos que generamos al azar (de aquí el nombre de modelo de energía al azar). Lo que pretendemos al considerar un número grande de secuencias (realizaciones), es asignar a cada conformación todas las energías que se obtendrían ajustando cada una de las secuencias a dicha conformación. Esto podríamos hacerlo en la práctica asumiendo algún modelo sencillo para representar las estructuras (como veremos en la siguiente sección) y ajustando cada secuencia a las estructuras, para lo cual calcularemos la energía asociada al par secuencia-estructura con alguna definición que estimemos oportuna.

De momento vamos a asumir simplemente que, para cada conformación m , las realizaciones de desorden (secuencias) ajustadas sobre dicha conformación nos dan un *conjunto de energías* que siguen una cierta *distribución de probabilidad*:

$$P(E) = \frac{1}{\sqrt{2\pi\Delta E^2}} \exp\left(-\frac{E^2}{2(\Delta E)^2}\right) \quad (11.1)$$

Donde estamos diciendo que estas energías son simplemente variables aleatorias que siguen una distribución Gaussiana con ancho ΔE . Vamos a hacer además una asunción aún más fuerte. Asumamos que, dada una conformación m_i , y otra conformación con energía m_j , sus distribuciones de energías no están correlacionadas, es decir: $P(E_i, E_j) = P(E_i)P(E_j)$. Esta es una asunción de *independencia entre las conformaciones* que es obviamente falsa, porque habrá muchas conformaciones tan parecidas entre sí que las correspondientes energías que calculemos al generar secuencias en ambas sí estén correlacionadas. Podrían ser incluso exactamente la misma distribución aun siendo conformaciones distintas, basta imaginar dos conformaciones que sean la una imagen especular de la otra. Pero esta asunción va a simplificar mucho nuestro cometido, pues ahora queremos saber cómo se distribuyen los microestados (conformaciones) a lo largo de los posibles valores de la energía. Es decir, queremos calcular $n(E)$, que es el número de estados comprendidos entre $(E, E + dE)$. Como tenemos el número estimado de configuraciones podemos calcular fácilmente el valor esperado del *número de estados* $\langle n(E) \rangle$, que resulta:

$$\langle n(E) \rangle = \gamma^N P(E) \sim \exp\left(N \ln \gamma - \left(\frac{E}{\Delta E}\right)^2\right) \quad (11.2)$$

De esta expresión vemos que, como el número de conformaciones γ^N es muy grande, la densidad de estados también lo será siempre y cuando $P(E)$ no sea muy pequeña. Si $P(E)$ no es pequeña, quiere decir que nos encontraremos un espectro continuo de energías. Y en ese continuo no seremos capaces de distinguir los estados independientemente de la realización de secuencia que tengamos. Pero si la

probabilidad disminuye lo suficiente la densidad de estados cambia significativamente. Siguiendo la Ecuación 11.2, el número de estados y la probabilidad son del mismo orden ($\gamma^N P(E) \sim 1$) para una determinada *energía crítica* E_0 :

$$E_0 \approx -\Delta E \sqrt{\ln \gamma} \quad (11.3)$$

tal que si $|E| > E_0$, la densidad de estados es menor que 1. Y este valor dependerá, a su vez, de ΔE , que podemos interpretar como la rugosidad de la superficie de energía, de modo que disminuyendo la rugosidad podemos alcanzar antes el valor crítico, facilitando el plegamiento. Hay que observar además que, si tenemos menos de un estado en el intervalo de energía $(E, E + dE)$, esto significa que habrá regiones de dicho intervalo en donde no encontramos ningún estado. Tenemos por tanto que la distribución de energías por debajo de la energía crítica se *discretiza*, es decir, habrá determinados intervalos en los que encontramos algún estado y otros en los que no encontraremos nada.

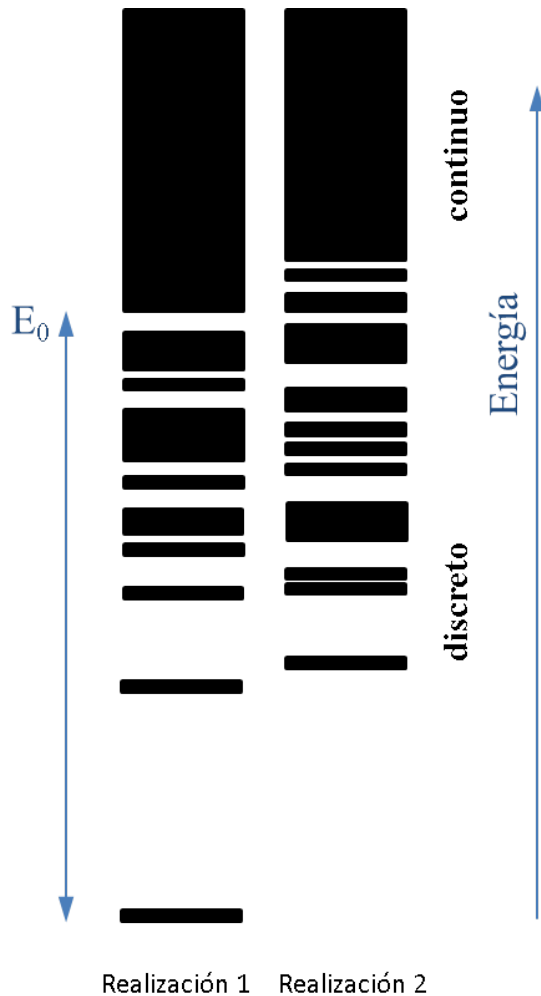


Figura 11.4: Espectros de energía típicos del modelo de energía al azar. A diferencia de lo que ocurre en la región continua del espectro que es indistinguible para las distintas realizaciones del desorden, a partir de una energía crítica E_0 encontramos una discretización que sí es específica para cada realización del desorden considerado.

¿De qué dependerá que encontremos o no un microestado en un determinado intervalo y cuál encontraremos? Pues dependerá precisamente de la secuencia con la que estemos trabajando, porque dicha secuencia se encontrará para cada microestado con una determinada probabilidad asociada en su

correspondiente Gaussiana. Para la conformación o conformaciones en las que dicha secuencia tiene probabilidades muy bajas de la energía obtendremos una contribución al espectro de energía. Será entonces en esta región discretizada en donde podremos distinguir las realizaciones, y podríamos considerar el espectro como una “*huella de identidad*” para cada realización (ver Figura 11.4). Este resultado tan sencillo es relevante porque nos abre la puerta a plantearnos si esa huella de identidad podría ser distinta cuando tratamos con secuencias al azar o cuando tratamos con secuencias reales, ya que estas últimas han superado un proceso de purificación a través de la selección natural. Veremos más adelante que parece que sí podemos pensar en un determinado rasgo característico de los espectros de proteínas reales respecto de aquéllas que no lo son.

Para continuar con nuestro ejemplo, podemos preguntarnos ahora cuál es el *estado termodinámicamente más probable para una realización*. Para ello multiplicamos la densidad de estados de energía E por el factor de Boltzmann, con lo que obtenemos:

$$p(E) = \frac{1}{Z} \gamma^N P(E) \exp(-E/k_B T) \quad (11.4)$$

Donde Z es la función de partición. Como hemos visto, el número de estados (y por tanto la entropía) se reduce drásticamente por debajo de un cierto valor mínimo. Si maximizamos la Ecuación 11.4 con respecto de la energía, obtendremos el número de estados ocupados a una temperatura dada [3]:

$$\gamma^N P(E_{m.p.}) = \exp\left(\frac{S}{k_B} - \frac{\Delta E^2}{2(k_B T)^2}\right) \quad (11.5)$$

Donde S es la entropía y el subíndice *m.p.* nos indica que es la energía más probable. De este ejercicio obtenemos algunos resultados también interesantes para nuestra comprensión del proceso de plegamiento. La dependencia con la temperatura nos lleva a observar que, al reducirse, entraremos en la región en donde el número de estados tiende a cero. La transición de una a otra región tendrá lugar para un valor crítico de la entropía S_0 , y a la temperatura asociada le llamaremos temperatura de transición vítrea², T_G . Este valor lo obtenemos cuando en la Ecuación 11.5 el argumento de la exponencial tiene un valor en torno a cero, que implica que encontramos la región discreta:

$$T_G = \left(\frac{\Delta E^2}{2k_B S_0}\right)^{1/2} \quad (11.6)$$

A esta temperatura el sistema sufre una *transición de fases* y se “congela” accediendo a unos pocos estados (ver Figura 11.5). Más adelante veremos cómo podemos incluir más propiedades de la superficie de energía para derivar un mapa de fases en el que aparecerá una nueva temperatura relevante, la *temperatura de plegamiento* T_F , que podremos relacionar con la *temperatura de vitrificación* T_G . Y veremos que la relación entre ambas es una de las características que nos permitirán diferenciar a las proteínas reales de las generadas por azar.

11.3.2. Un paisaje energético un poco más realista

Como explicamos anteriormente, podemos esperar que en proteínas reales los contactos nativos sean más favorables energéticamente de lo esperado por azar, como resultado del proceso evolutivo. Un modo de implementar computacionalmente estas ideas fue aplicar un tipo de modelo como el propuesto por Gō [29]. En este tipo de modelos se intenta *estudiar la dinámica y termodinámica de plegamiento de la*

²El subíndice G lo utilizamos para referirnos a la notación que se utiliza típicamente en inglés, donde se denomina *glass transition*. Del mismo modo utilizaremos el subíndice F para referirnos a la temperatura de plegamiento, del inglés *foldng*.

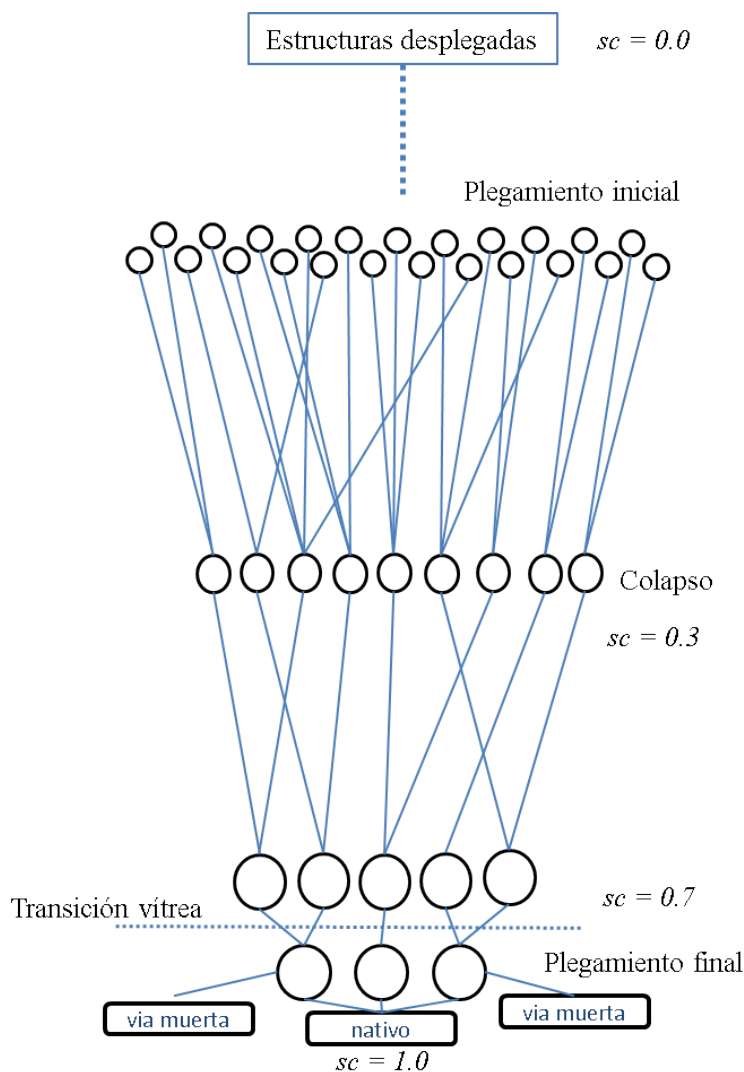


Figura 11.5: Representación esquemática de los caminos en el plegamiento de la proteína. Hay distintos caminos posibles para cada conformación, lo que hace que el estado nativo pueda ser cinéticamente accesible. Se describen los distintos estadios en función del parámetro de orden sc (solapamiento de contactos). Tras la transición vítrea hay pocos caminos posibles y algunos de ellos pueden llevar a conformaciones mal plegadas (vías muertas). Figura adaptada de [22]

proteína una vez es conocido su estado nativo. Como en el modelo de energía al azar, el valor de estos modelos residen en su capacidad de aumentar nuestra comprensión del problema (porque obviamente no nos va a servir para plegar una proteína cuando las condiciones del modelo implican que ya la conocemos). De modo que se construye un potencial en el que, si la interacción que tiene lugar entre dos residuos es la misma que la que obtienen en su estado nativo será energéticamente más favorable que en otra configuración de contactos, en donde será equivalente a lo esperado en una interacción al azar. Esto inducirá un *embudo* dentro de la superficie rugosa que construimos considerando interacciones al azar, como vimos en la Figura 11.2.

Para monitorizar la dinámica a lo largo de la superficie, introduciremos un *parámetro de orden*, es decir, una cantidad que nos indique cómo de cerca estamos del estado nativo y por tanto nuestra posición en el embudo energético. Hay varias opciones posibles como el radio de giro o el número de ángulos que

coinciden con los del estado nativo. Nosotros, por simplicidad conceptual y por lo extendido de esta elección, vamos a trabajar con el *número de contactos compartidos* o *solapamiento de contactos* (sc), cuya definición se puede encontrar la Sección 10.3. Dada una conformación de la proteína a lo largo de la reacción de plegamiento, definimos el solapamiento de contactos simplemente como el número de contactos compartidos entre la conformación considerada y el estado nativo.

El tratamiento analítico es prácticamente el mismo que en el modelo de energía completamente al azar. La principal diferencia es que ahora necesitamos *obtener las variables termodinámicas* como la energía media $\bar{E}(sc)$, rugosidad $\sqrt{\Delta E^2(sc)}$, entropía $S(E, sc)$ y temperatura de transición $T_G(sc)$, *en función del parámetro de orden elegido* (sc). Cuando el parámetro de orden $sc \approx 0$, tendremos un conjunto estadístico en el que las transiciones dependen simplemente de la rugosidad, como en el modelo de energía al azar, mientras que cuando el parámetro de orden tome valores elevados (del orden del 70 % de solapamiento) habrá pocos estados y estaremos interesados en conocer en detalle la cinética de transición entre los mismos.

Asumiremos ahora una distribución de probabilidad, para cualquier estrato del embudo energético, de que una estructura con solapamiento de contactos sc con el estado nativo tenga una energía E :

$$P(sc, E) = \frac{1}{\sqrt{2\pi\Delta E^2(sc)}} \exp\left(-\frac{[E - \bar{E}(sc)]^2}{2\Delta E^2(sc)}\right) \quad (11.7)$$

El desarrollo formal a partir de esta distribución, asumiendo la hipótesis de independencia entre estados de energía del modelo de energía al azar, es completamente análogo al caso anterior por lo que lo dejamos como ejercicio para el lector, que podrá encontrar resuelto en la referencia [3]. Vamos a interpretar directamente el resultado que obtenemos para la *energía libre*, que se obtiene como el logaritmo de los estados pesados termodinámicamente para una similitud dada [5]:

$$\begin{aligned} F(sc) &= E_{m.p.}(sc) - TS(E_{m.p.}(sc), sc) \\ &= \bar{E}(sc) - \frac{\Delta E^2(sc)}{2k_B T} - TS_0(sc) \end{aligned} \quad (11.8)$$

donde $E_{m.p.}$ es la configuración con energía más probable dado un solapamiento de contactos. La temperatura determina el balance entre la componente energética y la entrópica, pero además podemos ver qué valores del parámetro de orden se corresponden con el dominio de uno u otro término. Cuando la temperatura es alta la energía libre tendrá un único mínimo. Este mínimo puede corresponder a dos posibles conjuntos, el de estados completamente desplegados y el de un conjunto en el que la proteína colapsa pero de manera desordenada denominado *glóbulo fundido* (del inglés *molten globule* [28]). Con nuestra descripción no sería posible separar los dos conjuntos, y necesitaríamos un segundo parámetro de orden como el radio de giro [9]. Para energías bajas tendremos otro mínimo en el que son configuraciones de estados plegados las dominantes pero también tendremos que ir a modelos más complejos para poder diferenciar entre los estados que se encuentran en una fase vítrea (compactos pero no similares al nativo) y los propiamente similares al nativo. A temperaturas intermedias ambos mínimos estarán poblados pero con una ocupación mínima entre ambos, indicando la existencia de una barrera de energía.

11.3.3. Cinética del plegamiento y *gap* de estabilidad

Como en el modelo de energía al azar sin parámetro de orden, podemos encontrar la *temperatura de vitrificación*, cuya expresión es la misma que la Ecuación 11.6, pero donde ahora la entropía será una función del parámetro de orden. Por debajo de esta temperatura, el sistema “expulsa” la entropía ($S = 0$) y queda atrapado en un *microestado vítreo*. Aunque aún podría pasar a algún otro estado

compacto y haya menos estados en esta región, la búsqueda del estado nativo no es posible en una escala de tiempo razonable y volveríamos a un resultado como el que pretendemos resolver, a saber, la paradoja de Levinthal.

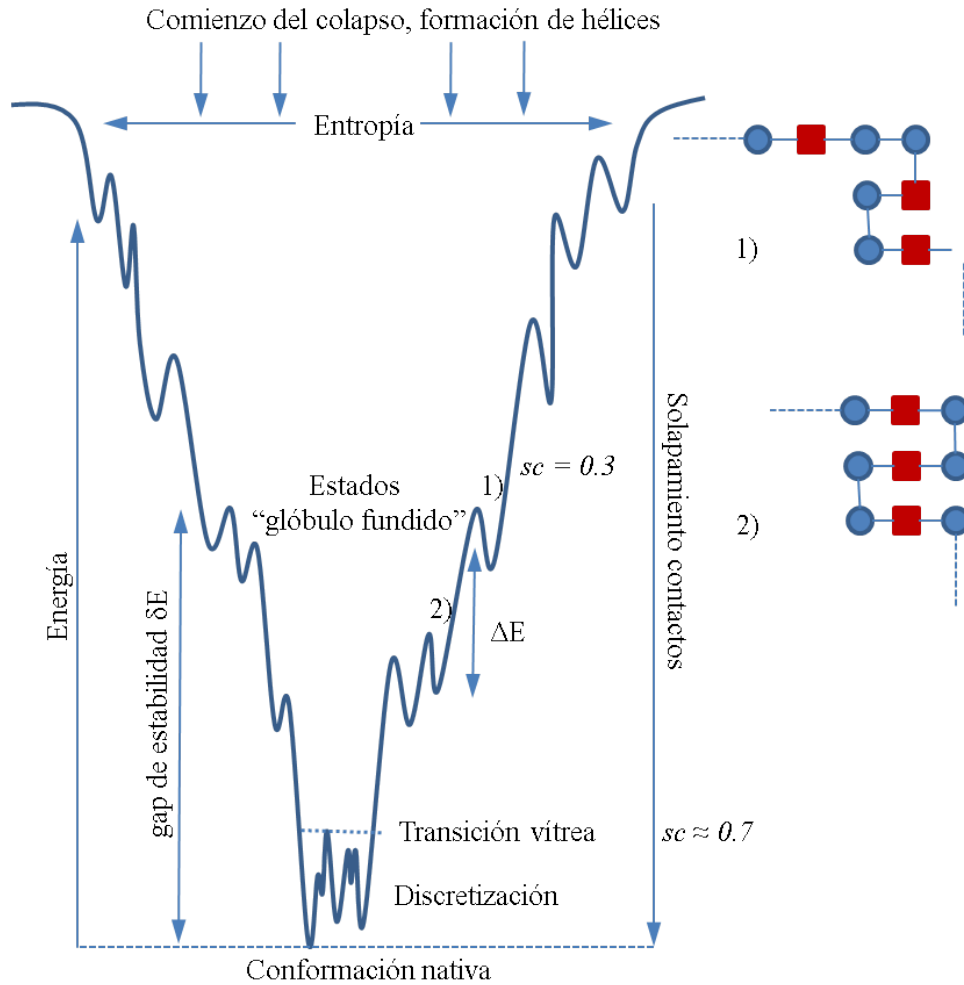


Figura 11.6: Ilustración de un paisaje energético en forma de embudo para una proteína pequeña (todo α), en función de los parámetros de orden sc (solapamiento de contactos) y E (la energía promediando el disolvente). Se representa la rugosidad ΔE relacionándolo con un cambio de conformación local ilustrado en los puntos 1 y 2, y el *gap* de estabilidad δE . Figura adaptada de [22].

Para resolver este problema, debemos introducir correlaciones en los estados de energía de modo que estén dinámicamente conectados [3, 4, 5]. El introducir correlaciones implica que existen combinaciones de aminoácidos que son favorables para la cinética de plegamiento, lo que nos lleva a buscar entender la relación entre la superficie de energía y la secuencia. Utilizando modelos más complejos [12] podemos estimar la temperatura de plegamiento T_f en función de parámetros relacionados con la superficie de energía:

$$T_f = \frac{\delta E + (\delta E^2 - 2S_0 \Delta E^2)^{\frac{1}{2}}}{2S_0} \quad (11.9)$$

En esta expresión el nuevo parámetro que aparece, δE , se denomina *gap de energía* o *gap de estabilidad*³

³Se ha preferido no traducir aquí el término *gap*, porque el referente al que alude en inglés creemos que contiene

y es un parámetro con el que vamos a poder caracterizar diferencias entre proteínas reales y polímeros al azar. En la Figura 11.6 se ilustran los ingredientes introducidos hasta el momento en el paisaje energético. Existen varias definiciones de *gap* de energía más o menos restrictivas pero que siguen la misma idea. Por ejemplo, se puede definir como la diferencia de energía entre la media de energías de los estados desplegados y la media de los estados plegados. A continuación relacionamos la temperatura de plegamiento con la temperatura de vitrificación [22], obteniendo:

$$\frac{T_f}{T_g} \approx \frac{\delta E}{\Delta E} \sqrt{\frac{2k_B}{S_0}} \quad (11.10)$$

Tanto numérica como experimentalmente se ha observado que el cociente T_f/T_g debe ser mayor que 1 para que el estado nativo sea cinéticamente accesible. Así que nos acercamos a entender qué esperaríamos de la superficie de energía en proteínas reales, porque observamos que el aumentar la relación entre el *gap* de estabilidad y la rugosidad $\delta E/\Delta E$ mejorará la *accesibilidad del estado nativo*, y por tanto las posibilidades de acceder al mismo en un tiempo biológicamente relevante. Hay que notar que tanto la energía como la rugosidad y la entropía son variables que dependen linealmente de la longitud, por lo que esta medida es independiente de la longitud de la proteína. En las secciones siguientes ampliaremos nuestro estudio alrededor del *gap* de energía, lo que nos conducirá progresivamente a dejar más de lado la física para introducirnos en la evolución. Vamos a profundizar un poco más en la relación entre los estados plegados compactos y los estados vítreos a través de un diagrama de fases.

11.3.4. Diagrama de fases

Las expresiones típicas que encontramos en la literatura para la energía libre en función de un parámetro de orden como el *sc* tienen uno o dos mínimos de energía en estos modelos formales sencillos [28]. Si tuviéramos un único mínimo cercano al escenario nativo, estaríamos ante un escenario (Tipo 0 en la Figura 11.7) que se denomina “colina abajo”, del inglés *downhill*, porque, si pensamos en el paisaje de la energía libre, no encontraríamos ninguna barrera y “bajaríamos” directamente al estado nativo [21]. Este escenario se ha observado en algunos casos experimentalmente [11] y podría ser común en condiciones fuertemente propicias para el plegamiento, principalmente para proteínas pequeñas. En este escenario que llamamos de *Tipo 0* (ver la Figura 11.7) *la fase desordenada es inestable*, y la proteína se pliega espontáneamente en un proceso cooperativo dirigiéndose hacia el mínimo de energía, como muestra en la figura el perfil de energía libre.

Existen además sistemas en los que *la fase desordenada es estable* (*Tipos I y II*), que son los que vamos a representar en un diagrama de fases como el que mostramos en la Figura 11.7. A estos sistemas les llamaremos *biestables*. En la figura tenemos una región en donde las conformaciones son equivalentes a las de un polímero aleatorio, otra región correspondiente al estado de glóbulo fundido (que es contigua al estado vítreo) en el que la cadena colapsa, y por último una región en la que encontraríamos propiamente el plegamiento nativo. Las transiciones entre las distintas fases vienen señaladas por líneas continuas (si son de primer orden) o discontinuas (de segundo orden). La transición del estado desplegado al plegado tiene lugar a través de una *barrera de energía* (ver las representaciones en la parte inferior de la Figura 11.7). Esta barrera de energía supone un cuello de botella en el número de caminos posibles de la proteína hacia el plegamiento, que influirá en el tiempo de plegamiento.

Las dos variables que determinan el diagrama de fases nos permiten entender bien en qué región nos encontraríamos en función de las propiedades que hemos ido mostrando. Vemos que un aumento

implícitamente más información que las posibles traducciones al español. Por un lado indica un salto en los valores de energía pero, como veremos, existe una discretización en el espectro con regiones vacías. Creemos que el término *gap* contiene ambos referentes, salto y vacío.

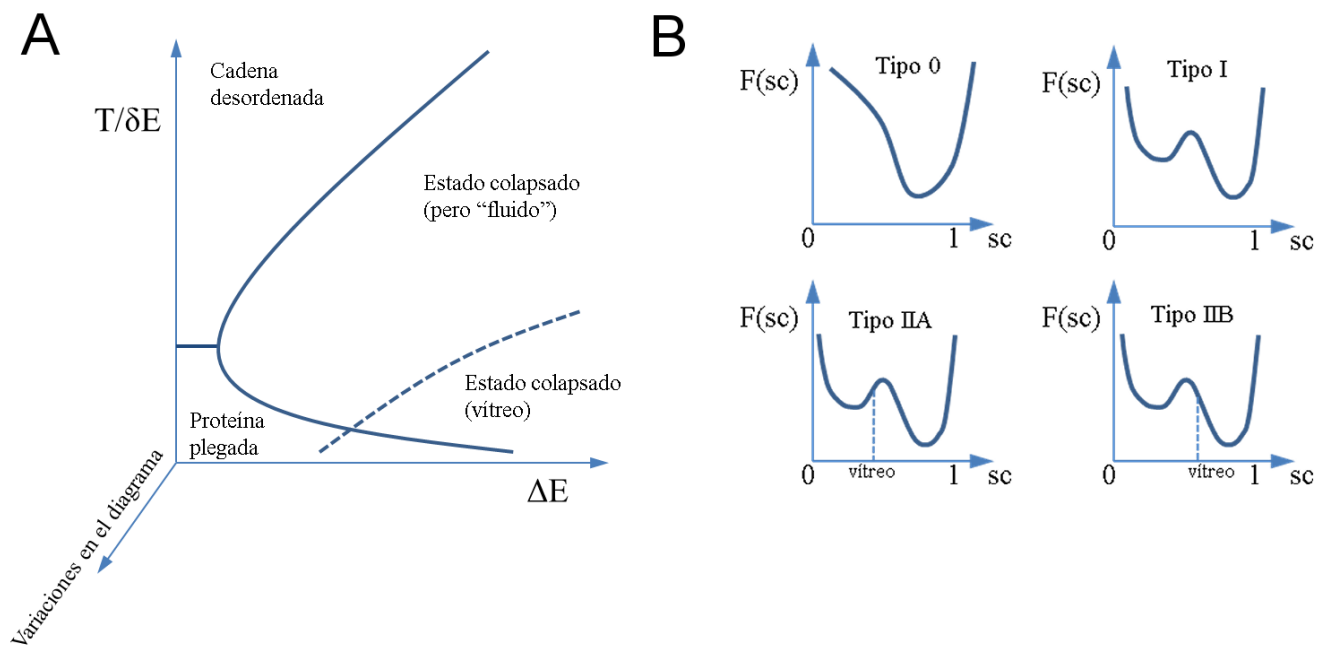


Figura 11.7: A. Diagrama de fases de una proteína plegable biestable. En el eje horizontal se representa la rugosidad energética y en el eje vertical la temperatura dividida por el *gap* de estabilidad. Las distintas fases estables se identifican en las distintas regiones. B. Distintos escenarios de plegamiento en los que se representa cualitativamente el perfil de energía libre (eje vertical) frente a un parámetro de orden como el solapamiento de contactos. Ver el texto principal para la explicación de las figuras. Figura adaptada de [3].

de la temperatura nos lleva a regiones en donde el polímero se encuentra en conformaciones más extendidas, pero esta variable se ve modificada por una propiedad determinada por la secuencia, el *gap* de estabilidad. Un *gap* de estabilidad mayor hará más probable mantener la proteína en su estado nativo. Por otro lado nos encontramos en el eje horizontal la segunda variable que es determinada por la secuencia, ΔE , que sabemos está relacionada con la rugosidad de la superficie de energía. Esperamos que una mayor rugosidad incremente la probabilidad de que la proteína encuentre un estado vítreo en el que colapsa para ciertas temperaturas, en vez del estado nativo. Entrar en la fase vítrea implica que el fenómeno es menos cooperativo, lo cual hace que sea más lento y que sea necesaria la búsqueda de caminos hacia el estado nativo, es decir, es un proceso más exigente cinéticamente. Experimentalmente se deberían de observar entonces estados intermedios en los que la proteína se queda atrapada. Es importante señalar también que este diagrama de fases variaría en función de otros parámetros que no estamos teniendo en cuenta al simplificar nuestro modelo. Por un lado la temperatura afecta a la propia rugosidad y al *gap* de estabilidad. Y más en general, cualquier modificación que afecte a la hidrofobicidad modificará el diagrama, con lo que tenemos que considerarlo como una “rebanada” de un diagrama de fases mucho más complejo.

Podemos imaginar distintos tipos de plegamiento en función de las características de la proteína que observemos. Si estuviéramos a la izquierda del todo, la proteína pasaría de un estado desordenado directamente al nativo al reducir la temperatura. Este tipo de escenario lo denominamos de *Tipo I*, y la accesibilidad cinética al estado nativo viene determinada por el número relativo de estados a los que la proteína puede acceder en el cuello de botella. En este caso el plegamiento es altamente cooperativo porque no existe vitrificación, y a pesar de que existe un cuello de botella encontraremos la proteína podrá acceder a un número importante de los estados que existen en este máximo de la energía libre. La velocidad de plegamiento se podría entonces estimar a través de la fracción de estados accesibles. Es importante señalar por último que, el que la proteína pueda visitar un número importante de estados en el cuello de botella del plegamiento la hará más robusta frente a mutaciones puntuales. En el capítulo de evolución de estructura de proteínas veremos que es de esperar que una única mutación no tenga un efecto muy dramático en la estructura de la proteína. Esto lo explicaríamos diciendo que una única mutación no varía muy significativamente el número de estados accesibles en el cuello de botella. Sin embargo, si el número de estados accesibles es muy pequeño como veremos a continuación, una mutación podría tener un efecto dramático si eliminara algunos de los pocos caminos accesibles.

Según nos movemos hacia la derecha en el diagrama de fases aumenta la rugosidad de la superficie de energía, y encontraríamos una fase vítrea. En función de la rugosidad del paisaje energético (más o menos a la derecha), encontraríamos en primer lugar el cuello de botella del plegamiento antes que la fase vítrea (*Tipo IIA*) o viceversa (*Tipo IIB*), lo que determinaría la accesibilidad del estado nativo y por tanto la velocidad de plegamiento. Si se encuentra primero el cuello de botella, la proteína tendrá, como en el caso anterior, un buen número de conformaciones que le llevan hacia el estado nativo. Pero tras una relativamente rápida nucleación entrará en la fase vítrea, lo que hará que los estadios finales del plegamiento sean lentos, pues deberá encontrar caminos específicos y encontraremos conformaciones intermedias. El último escenario (Tipo IIB) es el más lento, porque encuentra la fase vítrea antes de entrar en el cuello de botella. Esto condicionará fuertemente los posibles caminos a los que la proteína podrá acceder en el cuello de botella, y el tiempo de plegamiento dependerá de lo que haya sucedido en la fase vítrea, por lo que su estimación es complicada. El proceso será poco cooperativo y dependerá por tanto fuertemente de las condiciones iniciales. Por este mismo motivo, cualquier pequeño cambio en la secuencia de la proteína también podría influir negativamente en su tiempo de plegamiento, por lo que sería poco robusta por ejemplo frente a mutaciones puntuales. Vamos a pasar a continuación de la perspectiva más formal a algunas aplicaciones computacionales más concretas que tratan los conceptos que hemos ido presentando.

11.4. Algunos ejemplos computacionales sencillos

11.4.1. Modelos de grano grueso.

Una característica común de los modelos de sistemas con una elevada complejidad es que el primer paso siempre consiste en reducir el problema. La motivación para hacer esta reducción es variada, a veces proviene de la necesidad de hacer el problema más tratable o simplemente por intentar un abordaje basado en primeros principios. Existe además siempre un balance entre el nivel de detalle o la generalidad de la pregunta que queremos responder. La complejidad matemática aumenta tanto si queremos describir muy en detalle un proceso como si queremos que la pregunta sea muy general, por tanto el rol de la reducción de la descripción del sistema es fundamental. En la Figura 11.1 ya mostramos un ejemplo de cómo cambia la representación en función de la complejidad del modelo.

En primer lugar vamos a realizar reducciones muy importantes en la dimensionalidad del sistema a través de los llamados modelos de *grano grueso* (en inglés, *coarse grained models*). La primera simplificación consiste en reducir el número de grados de libertad de la proteína de entre todos los posibles (ángulos φ y ψ , ángulos dihedros de las cadenas laterales,...) al mínimo posible que nos permita tener un control suficiente sobre la estadística y la dinámica de la proteína. Hay varias opciones de simplificación posible. Por ejemplo podríamos trabajar considerando las posiciones de los carbonos *alfay* un ángulo pseudodihedro igual a $\varphi + \psi$ [25]. Aquí vamos a considerar en algunos apartados una simplificación incluso más fuerte, que es la propuesta por los llamados modelos de retículo (del inglés *lattice models*), en el cual los ángulos pseudodihedros están limitados a los que permite una red regular en dos o tres dimensiones (ver ilustraciones en la Figura 11.8).

La segunda simplificación importante consiste en que vamos a trabajar con un *potencial de campo medio*. Este tipo de potenciales, muy extendidos en mecánica estadística, reduce el número de componentes del campo de fuerzas de un sistema a un número menor, para lo cual realizamos promedios para algunas de estas componentes. Las cantidades seleccionadas en el promedio son aquellas cuyas escalas espacio temporales de variación son significativamente menores que la escala de evolución del proceso que queremos observar. Por ejemplo, los grados de libertad del disolvente podemos integrarlos y añadirlos de manera efectiva a la energía interna de la macromolécula. Podemos hacer esta operación, que simplificará mucho nuestros cálculos, porque la escala de tiempo en la que el disolvente llega al equilibrio es mucho más corta que la mayoría de los movimientos de la macromolécula.

La primera simplificación que hemos propuesto sobre la representación de la proteína sigue en realidad la misma idea. En este caso, son las fluctuaciones en los ángulos de las cadenas laterales mucho menores que las de los ángulos dihedros de la cadena polipeptídica, motivo por el cual se propone reducir los ángulos a un único ángulo pseudodihedro que evolucionará a través de un potencial de campo medio. No vamos a entrar en detalle sobre cómo se integran los distintos grados de libertad en un potencial de campo medio, pero invitamos al lector interesado a consultar la referencia [2] y un ejemplo de potencial empírico en [20]. Queremos resaltar también que, a pesar de que utilizamos un potencial de campo medio para los grados de libertad del disolvente, el rol del mismo es fundamental en el proceso de plegamiento. En particular, cuando tiene lugar una interacción hidrofóbica en la macromolécula entre dos grupos, habrá moléculas de agua que se encontraban ordenadas hidratando a cada uno de ellos que pasarán al medio, aumentando así sus grados de libertad. Este hecho implica que la entropía total del sistema macromolécula más disolvente aumente, lo que permite que el proceso de plegamiento pueda tener lugar espontáneamente al desplazar la reacción en el sentido en que hay una disminución en la energía libre. Para una exposición más detallada de este punto en castellano recomendamos la referencia [13]

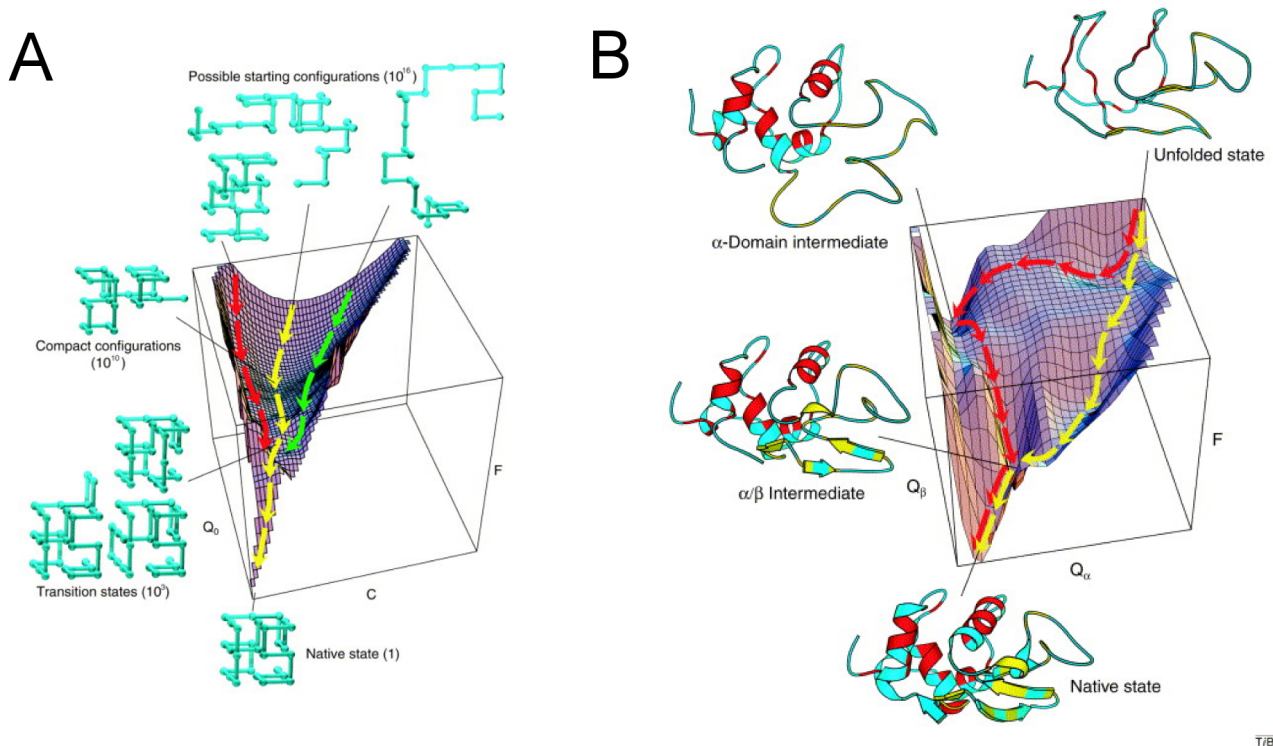


Figura 11.8: Ilustración en paralelo del plegamiento de un modelo 27 – mer (A) y de una proteína pequeña real (B). Representamos el número de conformaciones y la superficie de energía junto con algunos de los caminos seguidos en función de la región plegada, en un escenario similar al sugerido en el último ejemplo computacional que describimos debajo en la referencia [26]. La figura ha sido extraída de la referencia [9] con permiso del editor.

Lo sorprendente es que, a pesar de la simplificación, se obtienen resultados compatibles con las observaciones experimentales. Por tanto, la accesibilidad conceptual de los modelos junto con la variedad y alcance de sus resultados los hacen muy interesantes desde el punto de vista didáctico. De los diversos modelos tratados en la literatura, uno de los más utilizados es aquél en el que se considera una proteína de 27 monómeros (*modelo ‘27-mer’*) cuyo estado nativo es un cubo de lado 3, por tanto máximamente compacto. El primer problema que a uno se le ocurre es obviamente el de intentar plegar la cadena. Para ello hay que considerar dos ingredientes: una función de energía y un algoritmo de exploración de conformaciones. Estos dos ingredientes son muy generales, es decir, que complicáramos el modelo aumentaría la complejidad de ambos pero esencialmente no consideraríamos nuevos ingredientes. Cómo se pliega el modelo 27-mer y qué podemos aprender de él será sin embargo el último ejemplo que vamos a tratar, ya que es quizá el más complejo. Iremos introduciendo primero algunas funciones de energía más sencillas y veremos qué podemos aprender si consideramos la cadena ya completamente plegada.

11.4.2. La transición vítrea

Un paso importante en la capacidad de responder a preguntas con el modelo 27-mer fue la resolución computacional del siguiente problema sencillo. Sabemos que una cadena de 27 monómeros se puede plegar en un cubo de lado 3, ¿pero de cuántas formas posibles se puede plegar? Este problema fue resuelto por Gutin y Shakhnovich [27], encontrando *103346 conformaciones dentro del cubo*. En la

Figura 11.9 aparece representada una de estas posibles conformaciones, también llamados *camino hamiltoniano* dentro de la red.

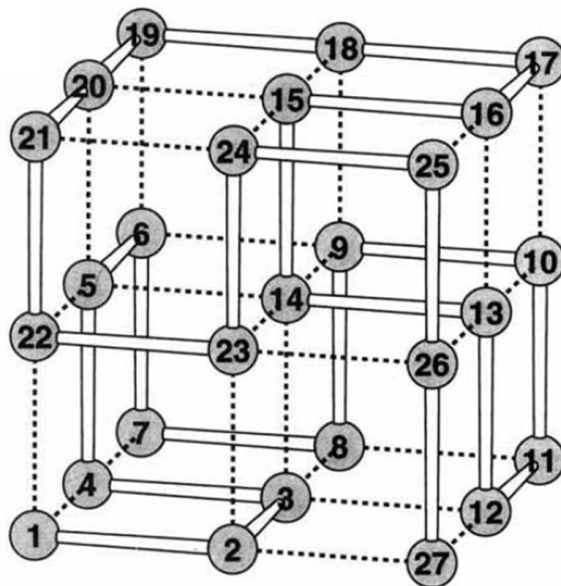


Figura 11.9: Representación de uno de los caminos posibles máximamente compactos dentro de un cubo de lado 3. Figura extraída de [26] con permiso del editor.

La importancia de la enumeración completa reside en que nos permite considerar un espacio completo de estructuras de proteínas al que se puede sacar mucho partido. Pensemos que el orden de magnitud del número de proteínas en el *Protein Data Bank* es, a día de hoy, prácticamente el mismo (hay depositadas unas 70000 estructuras resueltas por Rayos X por ejemplo). Por tanto uno podría preguntarse si se podría trabajar con este espacio de estructuras posibles dentro de un cubo, para responder a preguntas sobre cuestiones que se observan en el espacio de proteínas reales [32] o bien como un número elevado de conformaciones de la misma proteína [27]. Hay que ser cuidadoso con el lenguaje entonces, y cuando consideremos todas las soluciones del cubo como proteínas diferentes hablaríamos de ‘*estructuras*’ mientras que hablaremos más bien de ‘*conformaciones*’ cuando considereremos dichas soluciones como distintos plegamientos posibles de la misma proteína. En cualquier caso, en los siguientes ejemplos jugaremos con ambas situaciones lo que puede crear cierta confusión en el lector a primera vista.

En primer lugar llamemos k a cada una de las soluciones dentro del cubo, y llamemos α a una secuencia genérica que consideremos. Para cada secuencia, podemos proponer una energía asociada a cualquiera de las soluciones, que genéricamente llamamos E_k^α . Por ejemplo, una función sencilla podría ser la siguiente:

$$E_k^\alpha = \frac{1}{2} \sum_{i,j} C_{ij}^k U^\alpha(m_i, m_j) \quad (11.11)$$

Vemos que tenemos un primer término que depende de la solución estructural concreta, C_{ij}^k , que serán los elementos de la matriz de contactos y valdrá uno si los residuos i, j están en contacto en la conformación k , o cero en caso contrario. El segundo término es una función que depende de la naturaleza de los monómeros, m_i y m_j , cuyo valor particular vendrá determinado por cada secuencia α considerada. La función $U_{ij}^\alpha = U^\alpha(m_i, m_j)$ tomará así distintos valores en función de la combinación de monómeros que interactúan.

Podemos considerar por ejemplo una secuencia que tenga simplemente aminoácidos hidrofóbicos (H) y

polares (P). En este caso tendríamos que los valores absolutos de energía que consideramos para nuestra secuencia Hidrofóbico-Polar serán típicamente $|U(m_H, m_H)| > |U(m_P, m_P)| > |U(m_H, m_P)| = 0$. Otra posibilidad podría ser el considerar que los valores de las interacciones, U_{ij}^α , se obtienen al azar siguiendo una determinada distribución de probabilidad.

Consideremos esta última elección e imaginemos que generamos un número suficiente de secuencias al azar. Podemos calcular la energía para cada secuencia en cada una de las $M = 103346$ soluciones. Lógicamente, dada una secuencia, habrá soluciones en las que consiga valores de la energía más bajos (consideramos la energía negativa) que para otras. Nos gustaría saber si desde el punto de vista termodinámico existen secuencias tales que, dada una solución estructural, dominen estadísticamente a las demás. De encontrar tales secuencias, podríamos considerar la solución asociada como una aproximación de su estructura nativa, lo que implicaría que las demás soluciones son conformaciones mal plegadas para dicha secuencia. Pero observemos aquí que una solución que se considera una conformación mal plegada para una determinada secuencia α , podría ser el estado nativo para otra secuencia α' , y aquí podremos encontrar cierta confusión en el glosario.

Para ver si efectivamente existen secuencias dominantes para cada solución k vamos a calcular, para cada una de las secuencias α generadas al azar, sus *probabilidades de Boltzmann*:

$$p_k^\alpha = \frac{\exp\left(-\frac{E_k^\alpha}{k_B T}\right)}{Z_\alpha} \quad (11.12)$$

donde k_B es la constante de Boltzmann, T es la temperatura y Z es la función de partición que vendrá dada por:

$$Z_\alpha = \sum_{k=1}^M \exp\left(-\frac{E_k^\alpha}{k_B T}\right) \quad (11.13)$$

Queremos ahora calcular una cantidad que nos determine cuáles son las conformaciones termodinámicamente más relevantes. Para ello, Gutin y Shakhnovich propusieron [27] la siguiente función:

$$X = 1 - \sum_{k=1}^M (p_k^\alpha)^2 \quad (11.14)$$

Esta función es interesante porque tiene la siguiente particularidad. Cuando la temperatura es alta, independientemente de que la secuencia encuentre conformaciones más favorables, todas las probabilidades tenderán valores muy similares. Por tanto, $p_k^\alpha \approx 1/M$ por lo que el valor de la función X tenderá a uno. Si al descender la temperatura alguna conformación domina termodinámicamente, por ejemplo la conformación $k = 1$, tendremos que su probabilidad $p_1^\alpha \approx 1$ y el valor de X tenderá a cero. Podemos darnos cuenta de que la función X se comporta como una *medida de entropía*, ya que el sumatorio se corresponde con el argumento de la entropía de Renyi de orden dos. Veamos su significado desde el punto de vista informacional. Imaginemos a nuestra secuencia fluctuando en el medio con la posibilidad de pasar de una conformación a otra y que entonces nos piden que adivinemos en qué conformación está. Si la temperatura es alta, tenemos una incertidumbre muy elevada sobre qué conformación es en la que se encuentra nuestra secuencia (muchas entropías informacionales) pues todas las conformaciones tienen aproximadamente la misma probabilidad. Sin embargo, si la temperatura es baja y una conformación domina termodinámicamente no nos será difícil adivinar dónde está nuestra secuencia (baja entropía informacional). En la Figura 11.10 podemos ver el resultado de este ejercicio.

Lo primero que vemos es que, para valores altos, todas las secuencias tienen valores parecidos de la función X . Esto es de esperar porque, como hemos visto, todas las probabilidades termodinámicas tienen aproximadamente el mismo valor independientemente de la secuencia específica. Sin embargo,

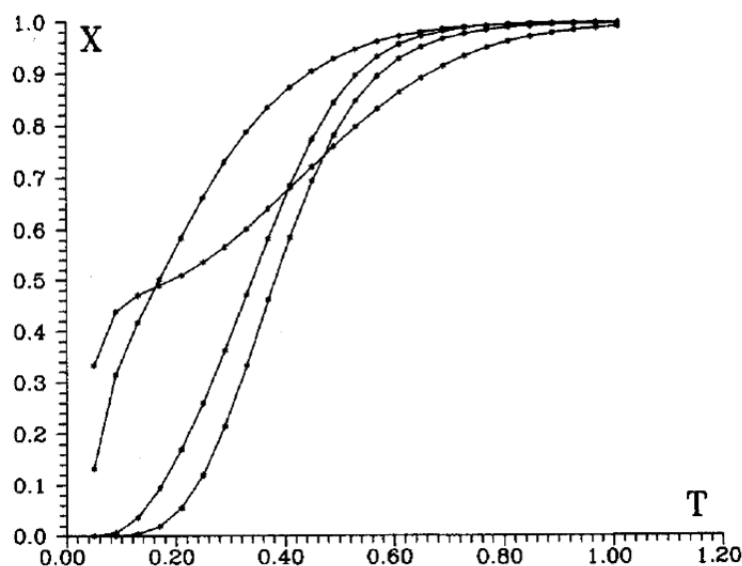


Figura 11.10: Relación de la variable X , que podemos relacionar con una medida de entropía, con respecto de la temperatura para cuatro secuencias. Vemos que a altas temperaturas las trayectorias son prácticamente indistinguibles, mientras que a partir de cierto valor de la temperatura (alrededor de 0,6) empiezan a distinguirse, lo que podemos relacionar con la transición vítrea explicada en el modelo de energía al azar. Por debajo de este valor cada secuencia sigue una trayectoria distinta, que podemos relacionar con la discretización específica del espectro para cada secuencia en el mismo modelo. Figura extraída de [27] con permiso del editor.

hay un valor de la temperatura, bastante similar para todas las conformaciones, por debajo del cual empieza a descender el valor de X y podemos observar comportamientos distintos en función de la secuencia hasta el punto de que, para temperaturas bajas, el comportamiento de las curvas cambia sustancialmente. Como hemos visto anteriormente en el desarrollo formal, esto es debido a que existe una única conformación que está dominando y lo que observamos es consecuencia de la relación entre la secuencia y dicha conformación, por tanto del paisaje energético.

11.4.3. Diseñabilidad

Esta diferencia entre las curvas abre el siguiente interrogante, ¿de qué depende el que una secuencia tenga un espectro de energía asociado a las distintas conformaciones con determinadas características, como un determinado valor de su energía en el estado nativo o de su *gap* de estabilidad? Para responder a esta pregunta Li et. al [18] propusieron un sencillo ejercicio. Consideraron nuevamente todo el espacio de soluciones del modelo 27-mer, pero en este caso consideraron secuencias más sencillas con aminoácidos hidrofóbicos y polares, de modo que simplemente modificaríamos la función U_{ij}^{α} en la Ecuación 11.11 respecto al caso anterior. Y a continuación determinaron cómo de diseñable era cada solución. Por *diseñabilidad* debemos entender el número de secuencias que son compatibles con una determinada estructura. En otras palabras, el número de secuencias que encuentran un mínimo absoluto de energía en una determinada estructura. Para implementarlo en nuestro modelo de juguete, lo que hicieron fue considerar un conjunto grande de secuencias, y determinar para cada una de ellas qué estructura es aquella en la que la secuencia encuentra el valor más bajo de la energía. Contaron entonces cuántas secuencias se asociaban a cada estructura, y observaron un resultado sorprendente: la distribución era

muy asimétrica, con unas pocas soluciones (que consideraríamos distintas estructuras en este caso) en las cuales muchas secuencias encontraban su valor mínimo y muchas estructuras a las que asociaban una o muy pocas secuencias. ¿A qué se debía este curioso resultado?

Como vimos en la Ecuación 11.10, la relación entre el *gap* de energía δE y la rugosidad del paisaje energético ΔE aparecían como determinantes para optimizar la accesibilidad del estado nativo. En la referencia que discutimos, los autores midieron el *gap* de energía como la energía entre la (ahora conformación) nativa para una secuencia dada y cualquier otra conformación promediando para todas las conformaciones: $\overline{\delta E}$. Aquí debemos tener cuidado, porque el *gap* de estabilidad se define entre el estado plegado y la media de las conformaciones no plegadas pero aquí, tal como se ha construido el ejercicio, son conformaciones también máximamente compactas, si bien las vamos a considerar como conformaciones no plegadas. Como se muestra en la Figura 11.11, lo que observaron es que las estructuras más diseñables tenían valores de $\overline{\delta E}$ significativamente mayores que el resto de las estructuras.

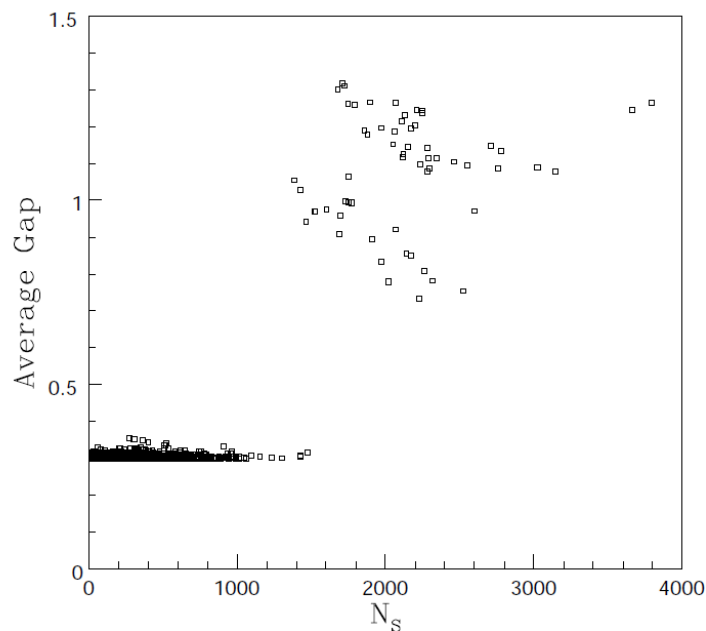


Figura 11.11: *Gap* de estabilidad δE de todas las conformaciones máximamente compactas del modelo 27-mer en función del número de secuencias que mapean a cada conformación (su diseñabilidad). Observamos que existe un salto significativo en el valor del *gap* de estabilidad a partir de cierto valor de la diseñabilidad, lo que se nos presenta como una propiedad que podría ser relevante en proteínas reales ya que les conferiría robustez mutacional. Figura extraída de la versión depositada en arXiv del artículo [18] con permiso de los autores.

Además, dichas estructuras presentaban simetrías que (con algo de imaginación) podrían ser interpretados en términos de estructura secundaria. Para entender si estas aparentes estructuras secundarias podrían tener cierta relevancia en la diseñabilidad de estas conformaciones, analizaron el conjunto de secuencias que diseñaban a las estructuras más diseñables y calcularon la probabilidad P_p de observar un aminoácido polar en una posición determinada de la secuencia. Como se muestra en la Figura 11.12, encontraron que existían posiciones con una probabilidad muy alta y otras con probabilidad nula (por tanto con probabilidad igual a uno de encontrar un aminoácido hidrofóbico). Esto sugeriría un símil entre dichas secuencias y las relaciones de homología que encontramos entre proteínas reales para las cuales existen posiciones en los que los aminoácidos están conservados y cuyas estructuras mantienen

una divergencia más lenta que la observada a nivel de secuencia. Estas relaciones serán tratadas en más detalle en las secciones de evolución y modelización de estructuras de proteínas (Capítulo 12).

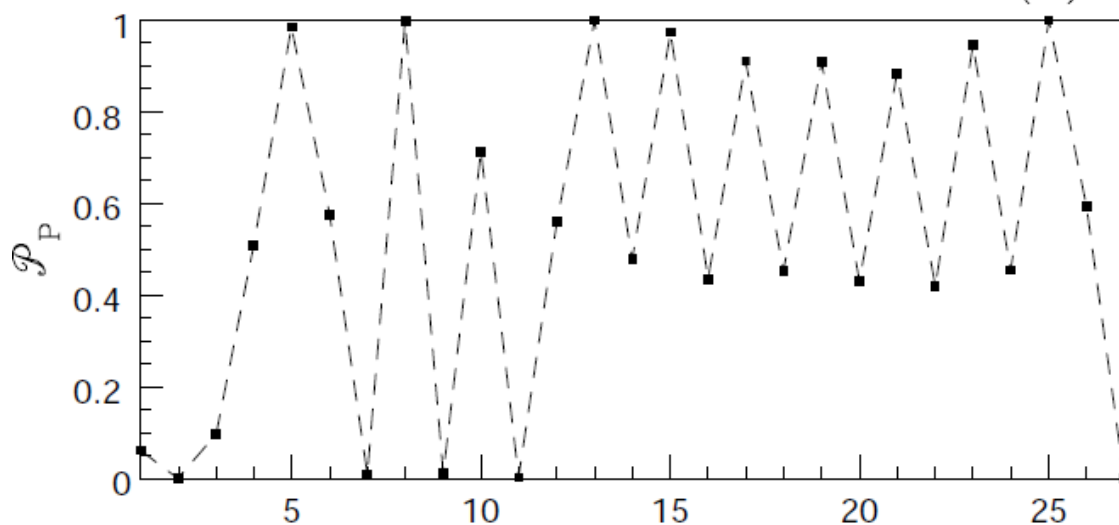


Figura 11.12: Probabilidad de encontrar un aminoácido polar en una posición determinada de la secuencia, calculado con el conjunto de secuencias asociadas a las estructuras más diseñables. Figura extraída de la versión depositada en arXiv del artículo [18] con permiso de los autores.

11.4.4. Plegamiento de una proteína simplificada

Llegamos al final de nuestro capítulo refiriéndonos de nuevo a la pregunta que nos hacíamos inicialmente: *¿cómo puede una proteína real plegarse en un tiempo biológicamente relevante si el número de conformaciones posibles es astronómico?* Con todo lo aprendido hasta ahora, podemos intentar responder a esta pregunta con un ejercicio práctico, es decir, intentando plegar una proteína. Dado que hemos considerado hasta ahora el modelo 27 – *mon*, ¿seríamos capaces de plegar una secuencia elegida al azar en una conformación máximamente compacta realizando una búsqueda al azar en un tiempo razonable?

Este ejercicio es al que se enfrentaron Sali et. al, y utilizaron los ingredientes que hemos considerado hasta ahora. Generaron doscientas secuencias al azar con valores de energía de interacción elegidos de una distribución gaussiana, como en el ejemplo tratado anteriormente para explicar la transición vítrea. Aplicando la misma función de energía que en la Ecuación 11.11, realizaron una búsqueda conformacional con un algoritmo de Monte Carlo con criterio de Metrópolis [26]. En la Figura 11.13 se puede ver la monitorización de la simulación frente a distintas variables.

Su primer resultado interesante consiste en la observación de que no era posible plegar todas las secuencias en un número de pasos de Monte Carlo razonable. Sin embargo, había otras secuencias (un 15%) que sí se plegaban en alguna de las diez simulaciones que realizaban para cada una de ellas, definiendo una *tendencia* como la fracción de esas diez simulaciones en las que la proteína alcanzaba el estado nativo. Al representar el espectro de energía de las distintas secuencias, como en la Figura 11.4, observaron que la característica más significativa de aquéllas que encontraban el estado nativo respecto de las que no lo encontraban era que el espectro de energía de las primeras tenía un mínimo de energía muy pronunciado, es decir, nuevamente el *gap* de energía era claramente mayor para aquéllas

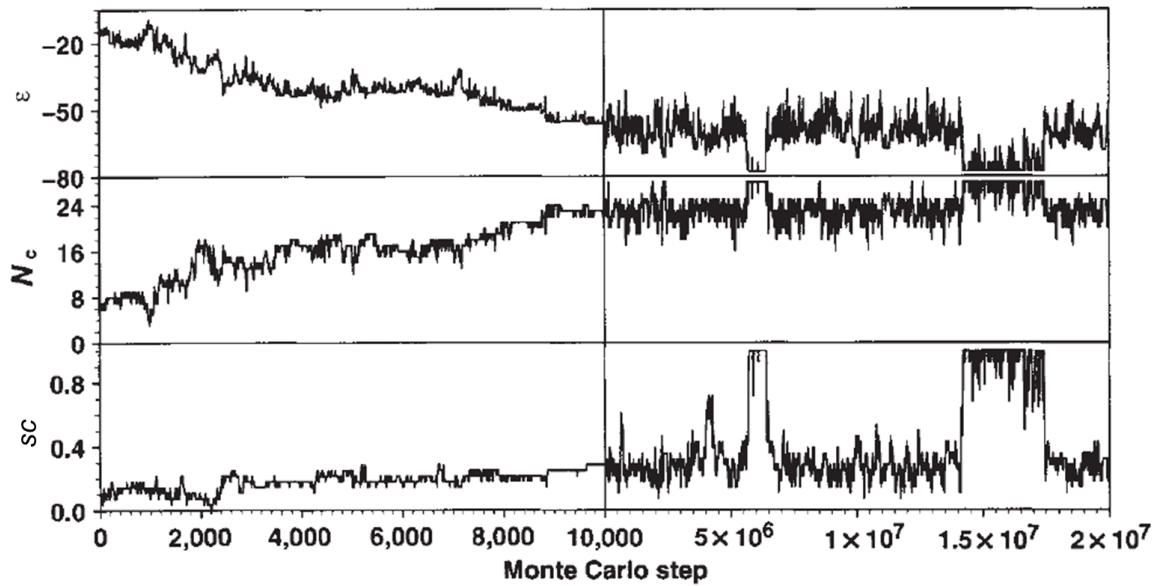


Figura 11.13: Ejemplo de trayectoria en la simulación por Montecarlo del plegamiento de una secuencia generada al azar en función de la energía ε (medido en unidades $k_B T$ siendo k_B la constante de Boltzmann), del número de contactos N_c , y del solapamiento de contactos sc . Figura extraída de [26] con permiso del editor y adaptada a la notación del texto.

con tendencias mayores. En nuestro caso, la secuencia de la izquierda en la Figura 11.4 tendría una tendencia mayor a plegarse que la de la derecha.

Para interpretar estos resultados desde el punto de vista cinético, monitorizaron además el número de movimientos que eran rechazados al aplicar el algoritmo de Metrópolis a lo largo de la simulación, lo que interpretaron (con mucho cuidado) en términos de velocidad de la reacción. Como señalamos anteriormente, el solapamiento de contactos (ver sección de alineamiento de estructura de proteínas) es un buen parámetro de orden en la cinética del plegamiento. Así que se puede proponer un *modelo de tres estadios* (ver Figura 11.14) en el que relacionamos la cinética de la reacción (monitorizada por el solapamiento de contactos), la energía de las conformaciones visitadas y la velocidad de la búsqueda conformacional, y que sería compatible con los escenarios que hemos ido mostrando.

El mecanismo que describe el plegamiento consta de un *primer estadio* muy rápido en el que la proteína pasaría de un estado desplegado en el que tenemos del orden de 10^{16} conformaciones a un estado semi-compacto en el que se reducen las conformaciones a 10^{10} . Entonces comienza un *segundo estadio* en el que se buscan los llamados estados de transición, que son del orden de 10^3 estados. Éste es el cuello de botella de la búsqueda desde el punto de vista de tiempo de simulación. Porque el *tercer estadio*, si el mínimo de energía es suficientemente pronunciado, es bastante rápido una vez se encuentra un estado de transición, accediendo al estado nativo finalmente. Recordamos que una representación más explícita de la simulación y sus estadios se ha ilustrado en la Figura 11.8.

En cuanto a los tiempos de plegamiento y siguiendo el ejemplo que aparece en el artículo, en una dinámica molecular típica se observan unas tres transiciones por residuo en 1 ns . Escalando los órdenes de magnitud que observamos del modelo 27-mer para una proteína de unos 80 residuos tendríamos 10^6 estados de transición y 10^{18} estado semicompatos, por lo que nuestro cuello de botella consistiría en la exploración de 10^{12} estados. Si muestreamos tres transiciones por residuo en 1 ns tardaríamos alrededor

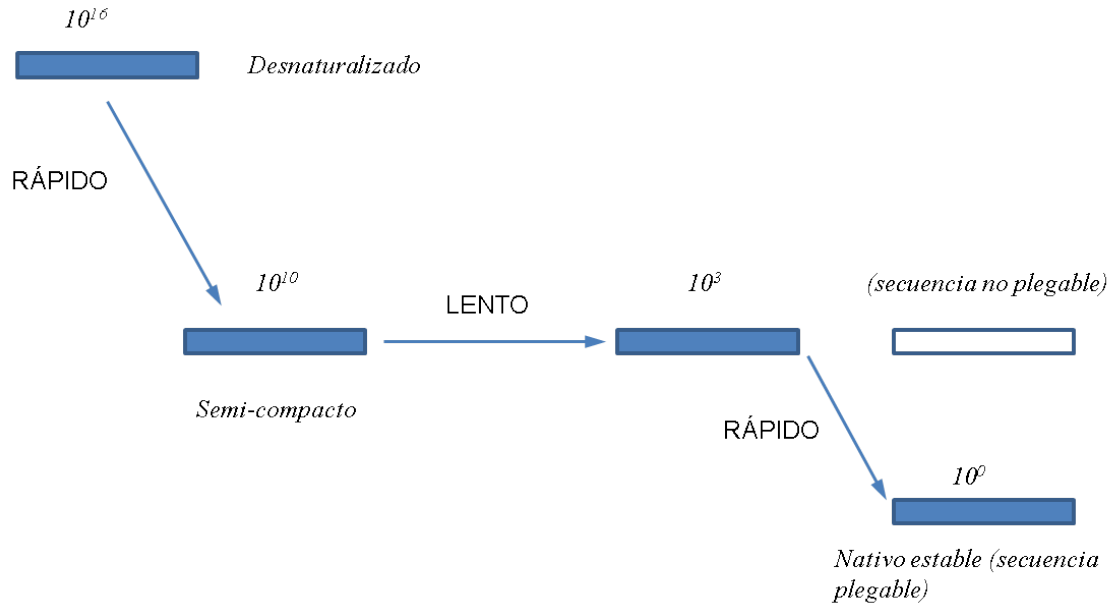


Figura 11.14: Representación esquemática del mecanismo de tres estadios propuesto en [26] (ver explicación en el texto).

de 4 segundos en encontrar un estado de transición, que podría ser compatible con una proteína real. Sin embargo estos órdenes de magnitud se dispararían si la proteína fuera más larga. En la discusión final del artículo haremos algunas consideraciones adicionales al respecto.

Por último queremos ilustrar el *efecto de la elección de la temperatura* en las simulaciones. En la Figura 11.15 vemos representada la densidad de estados en función de los parámetros de orden, es decir, el logaritmo de la ocupación media que se observa en cada celda a lo largo de una simulación suficientemente larga. Se representa una línea quebrada en la nube de estados, que indica una trayectoria típica observada a dicha temperatura. Para temperatura baja, se observa cómo se llega rápidamente a una conformación semi compacta tras lo cual la pendiente disminuye y va lentamente progresando aumentando el solapamiento de contactos hasta que, rápidamente, precipita sobre el estado nativo. Sin embargo no se observa la misma situación si la temperatura aumenta, ya que la cadena llega rápidamente a un estado semicompacto y se va a quedar preferentemente en dicha región, ya que los estados desnaturalizados son entrópicamente más favorables. Aunque eventualmente podría acceder al estado nativo, en estas condiciones no es estable y retornaría a una conformación semi compacta.

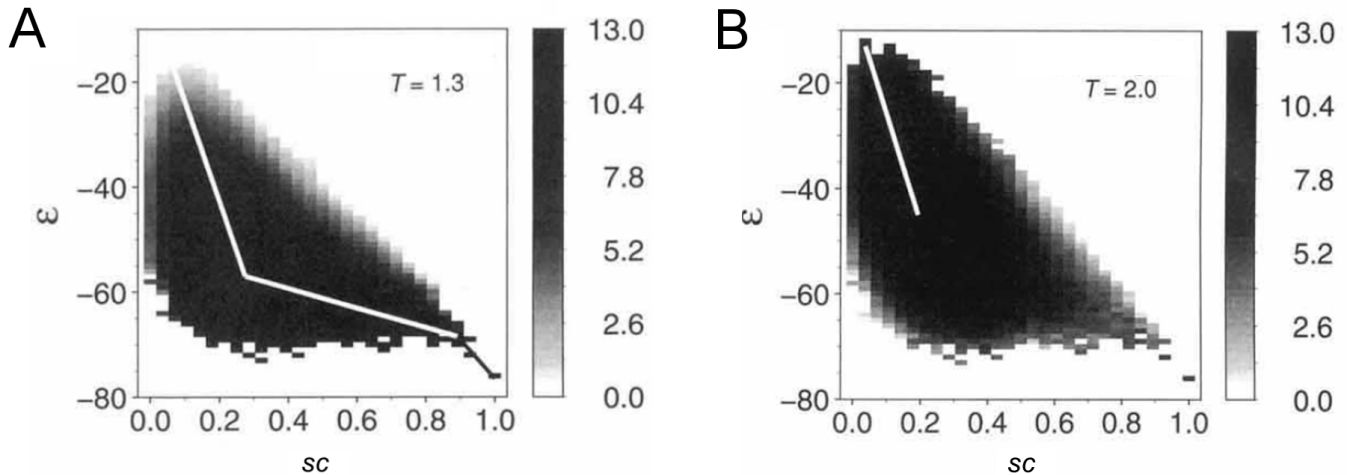


Figura 11.15: Fracción de estados ocupados durante una simulación larga de Montecarlo monitorizados por los parámetros de orden de energía ε y solapamiento de contactos sc , para dos temperaturas distintas. Figura extraída de [26] con permiso del editor y adaptada a la notación del texto.

11.5. Discusión

A lo largo de este capítulo sobre plegamiento de proteínas hemos buscado el proporcionar al lector una idea general de la *complejidad del problema de plegamiento de proteínas*, haciendo hincapié en la necesidad y potencialidad que revisten los modelos simplificados a la hora de responder a preguntas que, o bien por su complejidad o bien por su generalidad, no podrían ser abordadas de otro modo. Esto nos ha llevado a presentar algunos conceptos críticos como el paisaje energético y la posibilidad de relacionar propiedades específicas del mismo como la rugosidad y el *gap* de estabilidad, con propiedades medibles experimentalmente como las temperaturas de plegamiento y vitrificación.

Este análisis nos ha permitido entender la *resolución de la paradoja de Levinthal* basándonos en propiedades del paisaje energético, que nos aparece ahora más bien como un artefacto conceptual. Además hemos identificado una propiedad que emerge en los modelos simplificados, el *gap* de estabilidad, como propiedad relevante para el control cinético y por tanto para el plegamiento en tiempos biológicamente viables. Esta propiedad también nos abre la puerta a proponer modelos evolutivos, ya que sería consistente con la visión en la que aquellas estructuras funcionales con elevada robustez mutacional son seleccionadas.

Sin embargo *debemos de ser críticos con la validez y generalidad de los resultados encontrados*. Por ejemplo, el modelo de tres estadios que hemos explicado podría ser apropiado para algunas proteínas pequeñas pero sería inviable para proteínas grandes. Pero su valor, como todo conocimiento científico, reside más que en ser un resultado que consideremos verdadero en las explicaciones que rechaza. Por ejemplo, como investigadores computacionales, nos sugiere que más que trabajar en el desarrollo de algoritmos de búsqueda conformacional el reto está en encontrar funciones de energía más precisas, ya que en el ejercicio que hemos mostrado nos ha bastado un muestreo al azar con un algoritmo de Montecarlo para plegar nuestras proteínas de juguete. Sin embargo, para proteínas reales queda mucho camino por recorrer lo que probablemente requerirá de ideas originales.

Por ejemplo, como dijimos al principio del capítulo, uno de los problemas de interés desde el punto de vista del plegamiento es la modelización de estructura de proteínas, que no hemos tratado. Pero merece la pena señalar la novedosa propuesta que, en el contexto de la modelización, ha desarrollado

el laboratorio de David Baker en la Universidad de Washington, en donde surgió la idea de diseñar un videojuego para realizar la búsqueda conformacional. Ayudados por una función de energía que evalúa la calidad del plegamiento (y que determina la puntuación en el videojuego) los aficionados al mismo son invitados a aprender a plegar proteínas y a enfrentarse con proteínas con plegamientos desconocidos. La hipótesis es que *la intuición humana puede ser capaz de encontrar la solución óptima mejor que otros algoritmos de búsqueda* (siempre apoyados por una función de energía) sobre todo en aquellas estructuras con plegamientos complejos. Y, sorprendentemente, han conseguido encontrar el plegamiento de una proteasa retroviral utilizando los resultados obtenidos por los jugadores [15], la mayoría de ellos con poco o ningún conocimiento en bioquímica.

Por último debemos de tener en cuenta que *en los procesos biológicos in vivo hay multitud de factores que no estamos teniendo en cuenta* en nuestro análisis del plegamiento y, por extensión, en la evolución de estructura de proteínas. Ingredientes tales como el tamaño de la población en que el gen asociado a la proteína está evolucionando, las condiciones del medio (que pueden inducir sesgos mutacionales) o el rol de las chaperonas, serían determinantes y nos llevan a una visión del plegamiento en donde la estabilidad del estado nativo es más bien marginal [10]. De este modo deberíamos de implementar definiciones más complejas de lo que entedemos por estabilidad, como el balance entre estabilidad frente a desplegamiento (*unfolding stability*) o a mal plegamiento (*misfolding stability*), conceptos que ya se están implementando en modelos evolutivos computacionales [19, 30]. Estos argumentos plantean la creciente necesidad de pensar en el problema de la evolución de proteínas desde una perspectiva más amplia [23], pero ya dotados de una serie de herramientas conceptuales mínimas para abordar nuevos ingredientes.

Por tanto, *el reto continúa abierto en todos los frentes*, desde la comprensión de los conceptos más generales que hemos propuesto en el presente capítulo, a las ideas originales como el videojuego Foldit⁴ y el creciente interés por los entornos colaborativos como la iniciativa WeFold⁵). Y por último, no nos olvidemos de la necesidad de mirar hacia el *nuevo reto conceptual y computacional que suponen las proteínas intrínsecamente desestructuradas*.

⁴Foldit. <https://fold.it>

⁵WeFold. <http://www.wefold.org>

11.6. Bibliografía

- [1] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *PNAS*, 47(9):1309–1314, Sept. 1961. PMID: 13683522 PMCID: PMC223141.
- [2] C. L. Brooks, M. Karplus, and B. M. Pettitt. *Advances in Chemical Physics, Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*. John Wiley & Sons, July 1990.
- [3] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics*, 21(3):167–195, Feb. 1995.
- [4] J. D. Bryngelson and P. G. Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences*, 84(21):7524–7528, Nov. 1987.
- [5] J. D. Bryngelson and P. G. Wolynes. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.*, 93(19):6902–6915, 1989.
- [6] L. Cruzeiro Hansson. Protein folding: thermodynamic versus kinetic control. *Journal of biological physics*, 27:S6, 2001.
- [7] B. Derrida. Random-energy model: Limit of a family of disordered models. *Physical Review Letters*, 45(2):79–82, July 1980.
- [8] K. A. Dill and H. S. Chan. From levinthal to pathways to funnels. *Nature Structural & Molecular Biology*, 4(1):10–19, Jan. 1997.
- [9] A. R. Dinner, A. Šali, L. J. Smith, C. M. Dobson, and M. Karplus. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in Biochemical Sciences*, 25(7):331–339, July 2000.
- [10] C. M. Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, Dec. 2003.
- [11] M. M. Garcia-Mira, M. Sadqi, N. Fischer, J. M. Sanchez-Ruiz, and V. Muñoz. Experimental identification of downhill protein folding. *Science*, 298(5601):2191–2195, Dec. 2002.
- [12] R. A. Goldstein, Z. A. Luthey-Schulten, and P. G. Wolynes. Optimal protein-folding codes from spin-glass theory. *Proceedings of the National Academy of Sciences*, 89(11):4918–4922, June 1992.
- [13] J. R. Grigera. La física de la desnaturalización de proteínas en frío: (o como cocinar un huevo a temperatura ambiente). *Revista iberoamericana de física*, 6(1):27–33, 2010.
- [14] M. Karplus. Behind the folding funnel diagram. *Nature Chemical Biology*, 7(7):401–404, July 2011.
- [15] F. Khatib, F. DiMaio, F. C. Group, F. V. C. Group, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popović, M. Jaskolski, and D. Baker. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology*, 18(10):1175–1177, 2011.
- [16] J. Kubelka, J. Hofrichter, and W. A. Eaton. The protein folding speed limit. *Current Opinion in Structural Biology*, 14(1):76–88, Feb. 2004.
- [17] C. Levinthal. How to Fold Graciously. In J. T. P. Debrunner and E. Munck, editors, *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois*, pages 22–24. University of Illinois Press, 1969.
- [18] H. Li, R. Helling, C. Tang, and N. Wingreen. Emergence of preferred structures in a simple model of protein folding. *Science (New York, N. Y.)*, 273(5275):666–669, Aug. 1996. PMID: 8662562.
- [19] R. Méndez, M. Fritsche, M. Porto, and U. Bastolla. Mutation bias favors protein folding stability in the evolution of small populations. *PLoS Comput Biol*, 6(5):e1000767, May 2010.
- [20] S. Miyazawa and R. L. Jernigan. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules*, 18(3):534–552, Mar. 1985.
- [21] V. Muñoz. Thermodynamics and kinetics of downhill protein folding investigated with a simple statistical mechanical model. *International Journal of Quantum Chemistry*, 90(4-5):1522–1528, 2002.
- [22] J. N. Onuchic and P. G. Wolynes. Theory of protein folding. *Current Opinion in Structural Biology*, 14(1):70–75, Feb. 2004.

- [23] C. Pál, B. Papp, and M. J. Lercher. An integrated view of protein evolution. *Nature Reviews Genetics*, 7(5):337–348, May 2006.
- [24] V. Pande, A. Grosberg, and T. Tanaka. Statistical mechanics of simple models of protein folding and design. *Biophysical Journal*, 73(6):3192–3210, Dec. 1997.
- [25] W. L. Peticolas and B. Kurtz. Transformation of the ϕ - ψ plot for proteins to a new representation with local helicity and peptide torsional angles as variables. *Biopolymers*, 19(6):1153–1166, Feb. 2004.
- [26] A. Šali, E. Shakhnovich, and M. Karplus. How does a protein fold? , *Published online: 19 May 1994*; | doi:10.1038/369248a0, 369(6477):248–251, May 1994.
- [27] E. Shakhnovich and A. Gutin. Enumeration of all compact conformations of copolymers with random sequence of links. *The Journal of Chemical Physics*, 93(8):5967–5971, Oct. 1990.
- [28] N. D. Socci and J. N. Onuchic. Kinetic and thermodynamic analysis of proteinlike heteropolymers: Monte carlo histogram technique. *The Journal of Chemical Physics*, 103(11):4732–4744, Sept. 1995.
- [29] H. Taketomi, Y. Ueda, and N. Gō. Studies on protein folding, unfolding and fluctuations by computer simulation. i. the effect of specific amino acid sequence represented by specific inter-unit interactions. *International journal of peptide and protein research*, 7(6):445–459, Dec. 1974.
- [30] D. M. Taverna and R. A. Goldstein. Why are proteins marginally stable? *Proteins: Structure, Function, and Bioinformatics*, 46(1):105–109, 2002.
- [31] A. M. Vasilyev. *Introduction to Statistical Physics*. Imported Pubn, Sept. 1984.
- [32] K. B. Zeldovich, P. Chen, B. E. Shakhnovich, and E. I. Shakhnovich. A first-principles model of early evolution: Emergence of gene families, species, and preferred protein folds. *PLoS Comput Biol*, 3(7):e139, July 2007.