

Chama-se *fold*ing de proteínas ao processo espontâneo a partir do qual uma cadeia linear de aminoácidos adquire uma estrutura tridimensional biologicamente activa. A compreensão deste fenómeno, considerada por muitos um dos problemas mais importantes da ciência actual, terá um grande impacto não só ao nível da saúde e do bem-estar humanos como também ao nível da ciência fundamental, na aprendizagem e aquisição de novas leis e conceitos da física dos sistemas complexos. Tendo surgido no contexto da biologia molecular, este problema é hoje claramente interdisciplinar, necessitando de ferramentas de várias áreas do conhecimento, e para o qual o contributo da física tem sido determinante. O objectivo deste artigo é mostrar como a utilização de metodologias da física, incluindo o recurso à simulação computacional de modelos simples, permitiu criar uma estrutura conceptual (a chamada "paisagem de energia") sobre a qual uma sinergia exemplar entre a teoria e a experiência tem gerado avanços muito significativos.

PATRÍCIA F. N. FAÍSCA

Centro de Física Teórica e Computacional da
Universidade de Lisboa
Av. Prof. Gama Pinto 2,
1649-003 Lisboa

patnev@cii.fc.ul.pt

<http://alf1.cii.fc.ul.pt/~patnev>

O MISTÉRIO DA DAS PROTEÍNAS

Experiências *in vitro*: a hipótese termodinâmica

As proteínas são robôs celulares, máquinas moleculares construídas à escala do nanómetro, capazes de executar de uma forma espontânea e programada todas as tarefas essenciais à manutenção da vida. Por exemplo, as enzimas aceleram reacções químicas que de outro modo seriam demasiado lentas, os anticorpos são moléculas responsáveis pela identificação e eliminação de agentes invasores e as hormonas asseguram a transmissão de informação entre as células.

Para que possa funcionar correctamente, cada proteína deve exibir uma estrutura tridimensional **única** que determina a sua função biológica. É a chamada estrutura nativa, que emerge como produto final de um processo complexo envolvendo o enrolamento e a dobragem (*fold*ing, em inglês) da cadeia de aminoácidos que compõe a proteína, designada por estrutura primária. Christian Anfinsen foi dos primeiros cientistas a interessar-se pela compreensão dos princípios físicos envolvidos no processo de *fold*ing de proteínas. Por que razão se dobra a proteína para a estrutura nativa? Por que é única essa estrutura? Estas foram duas das questões fundamentais para as quais procurou resposta durante toda a década de 1950. Com a ajuda dos seus alunos de pós doutoramento Fred White e Michael Sela, Anfinsen dirigiu uma série de experiências que culminaram com a elaboração da chamada hipótese termodinâmica (HT), trabalho que foi galardoado com o Prémio Nobel da Química de 1972. A HT é baseada na observação de que o processo de *fold*ing ocorre de uma forma espontânea conduzindo, por isso, a um estado – o estado nativo – que é o mais estável (de mais baixa energia) do ponto de vista termodinâmico. Por outro lado, Anfinsen também

FORMA

FGRCELAAAMKRHGLD
 RGYSLGNWVCAAKFES
 NTQATNRNTDGGSTDYG
 QINSRWWCNDGRTPGS
 LCNIPCSALLSSDITA
 NCAKKIVSDGNMNAW
 WRNRCKGTDVQAWIRG
 FRTHYUWVLIJMMRTY
 WRNRCKGTDVQAWIRG
 OINSRWWCNDGRTPGS

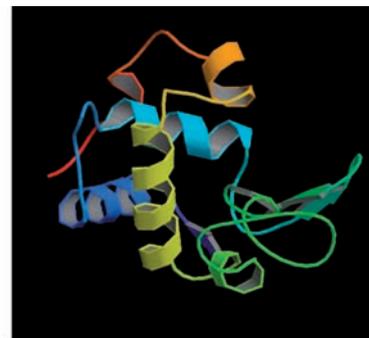
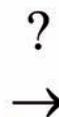


Fig. 1 - Como determinar a estrutura nativa de uma proteína a partir do conhecimento da sua estrutura primária?

concluiu que a estrutura nativa tem necessariamente de ser determinada pela sequência de aminoácidos (ou mais rigorosamente, pela totalidade de interações entre os aminoácidos que compõem a proteína) já que, para além da própria proteína, não é necessária a participação de nenhuma outra molécula no processo.

Os resultados das experiências de Anfinsen foram o ponto de partida para um novo problema, para o qual ainda hoje procuramos uma solução: como determinar a estrutura tridimensional de uma proteína partindo apenas do conhecimento da sequência de aminoácidos que constitui a estrutura primária? É claro que este problema (conhecido como *protein folding problem*) pode ser tomado como corolário da questão fundamental que consiste em compreender os mecanismos envolvidos neste importante processo biológico.

Folding de proteínas: um problema importante

A vida depende largamente da existência e do funcionamento correcto de uma quantidade enorme de proteínas. Às vezes, muitos dos processos celulares de controlo e regulação em que essas proteínas estão envolvidas falham, ou porque o organismo não é de todo capaz de as produzir, ou porque são fabricadas com defeito. Como a função da proteína depende estritamente da sua estrutura nativa, basta que esta última exiba uma pequena falha para que a proteína passe a não funcionar correctamente. Patologias comuns, como a diabetes de tipo I e a hemofilia, ocorrem porque o organismo é incapaz de produzir a insulina e o factor VIII, respectivamente. Já certos tipos de cancro resultam da produção defeituosa de uma proteína

que participa na regulação da divisão celular, o factor p21Ras. Outras patologias, como as encefalopatias espongiiformes, entre as quais a BSE ou doença das vacas loucas, estão na origem de um processo de *folding* defeituoso do qual resultam proteínas (na realidade, agentes infecciosos chamados príões) que, tendo uma grande afinidade entre si, se agregam para formar uma espécie de fibras. Estas fibras actuam sobre o tecido neuronal do sistema nervoso central e o desfecho pode ser dramático. Estas patologias têm sido denominadas “doenças dos tempos modernos” porque resultam em parte da existência de condições completamente novas, criadas pelas actividades desenvolvidas pelo homem, que são capazes de desafiar e pôr em risco a homeostase (ou controlo) dos processos bioquímicos normais. A solução para problemas deste tipo passa por sermos capazes de fornecer a proteína em falta ao organismo, de uma forma artificial, ou por inactivar a proteína infecciosa, através da administração de uma droga para a qual a primeira tenha uma grande afinidade, quer de um ponto de vista estrutural quer energético. Encontrar uma droga que se “encaixe” numa certa proteína da melhor forma possível é um problema típico no *design* de drogas. É claro que se compreendermos a relação sequência-estrutura, poderemos “desenhar” qualquer proteína (ou qualquer outra droga que seja específica para uma certa proteína). Mas, para que isto seja possível, precisamos primeiro de compreender a questão fundamental de como se processa o *folding* de proteínas. À frente veremos como é que a física e o recurso à simulação computacional têm tido um papel fundamental na elucidação deste processo.

O paradoxo de Levinthal e os caminhos de folding

Em 1968, Cyrus Levinthal, um físico de formação que se converteu ao estudo da biologia molecular, levantou uma séria objecção à ideia de que a procura do estado nativo possa ser feita de forma aleatória, tal como era sugerido pela HT. O argumento de Levinthal, simples e eficaz,

baseava-se na seguinte experiência conceptual (*gedanken experiment*). Consideremos uma pequena proteína com 100 aminoácidos e suponhamos que cada aminoácido só pode estar num de dois estados possíveis (por exemplo, só pode tomar duas orientações diferentes). Nestas condições, a proteína tem acesso a um total de $2^{100} \approx 10^{30}$ conformações, total esse que inclui obviamente a estrutura nativa. Como a molécula não pode passar de uma conformação para a outra em menos de 1 picossegundo (ps), que é o tempo de uma vibração térmica, seriam precisos 2^{100} ps, ou seja, $3,9 \times 10^{10}$ anos, no mínimo, para explorar exaustivamente todo o espaço conformacional e encontrar a conformação (que é apenas uma!) correspondente ao estado nativo. Ora, acontece que esta escala de tempo é da ordem de grandeza da idade do Universo, estimada em $1,4 \times 10^{10}$ anos. Estamos assim perante um problema, já que o *foldings* de proteínas deste tamanho leva no máximo alguns segundos, e tipicamente ocorre na escala temporal do nanossegundo ou do segundo. A conclusão que daqui se tira é que a HT não consegue explicar a escala de tempo característica do processo de *foldings* de proteínas.

Por razões óbvias, este problema ficou conhecido como o paradoxo de Levinthal e foi o próprio Levinthal o primeiro a sugerir uma solução. Levinthal teorizou a existência de um “caminho” de *foldings* específico (*foldings pathway*), composto por vários estados intermediários – um pouco à semelhança do que se passa numa reacção química vulgar – no fim do qual se encontra o estado nativo. No entanto, e ao contrário do sugerido pela HT, o estado nativo na proposta de Levinthal não corresponde necessariamente ao mínimo global da energia, não tem que ser o estado termodinamicamente mais estável. Corresponde, isso sim, ao estado de energia mínima mais acessível de um ponto de vista cinético.

Como consequência da proposta de Levinthal, a investigação experimental em *foldings* de proteínas até ao início da década de 1990 foi em grande parte dominada pela procura de produtos intermediários de *foldings* suficientemente estáveis para que pudessem ser isolados e devidamente caracterizados. No entanto, a descoberta em 1991 de uma pequena proteína, com cerca de 100 aminoácidos, que se dobra rapidamente sem passar por quaisquer intermediários, mostrou que a existência de intermediários estáveis não é de todo um requisito essencial para a rapidez do processo.

A perspectiva clássica do *foldings* de proteínas baseia-se na dicotomia termodinâmica *versus* cinética e na abordagem tradicional da bioquímica, que considera cada molécula um sistema único, sendo por isso necessária uma descrição detalhada, à escala atómica, do seu caminho de *foldings*. Como veremos adiante, uma das contribuições mais importantes da física para a compreensão deste fenómeno foi precisamente a de reconciliar as perspectivas de Anfinsen e

Levinthal no quadro de uma teoria unificada, que se baseia na natureza estatística do processo.

Folding in silico I: dinâmica molecular

De acordo com Christian Anfinsen, se conhecermos a totalidade das interacções que se estabelecem entre os átomos de uma proteína, devemos, em princípio, poder prever qual será a estrutura tridimensional adoptada pela cadeia de aminoácidos que contém esses átomos. Mas será fácil essa tarefa? De uma forma simplificada, podemos dizer que as interacções entre os átomos são de natureza electrostática e quântica. É claro que, dentro das interacções electrostáticas, temos que distinguir várias categorias (interacções de Lennard-Jones, de van der Waals, pontes de hidrogénio, de solvatação com as moléculas de água, etc.). Uma proteína como a hemoglobina, que transporta o oxigénio aos alvéolos pulmonares, contém cerca de 4000 átomos que podem, em princípio, interagir durante o processo de *foldings*. Para todos os pares de interacção possíveis é então preciso saber quanto valem os “parâmetros” de interacção correspondentes. O cálculo rigoroso destas quantidades requer um tratamento quântico e não é de todo trivial. Para além disso, há que ter em conta que o resultado do cálculo depende não só do tipo de átomo mas também do ambiente químico em que se encontram os átomos participantes na interacção, o que gera complicações adicionais. Geralmente, este tipo de cálculo faz-se no contexto das simulações por dinâmica molecular (DM). Trata-se de simulações deterministas que modelam a proteína como sendo um sistema newtoniano de N átomos ligados por molas. Para além das interacções electrostáticas entre os átomos, há ainda a considerar o cálculo dos parâmetros relativos às energias de alongamento, de torção e de dobragem da ligação química. Tudo isto implica o cálculo de um número extraordinariamente grande de parâmetros, que devem ser testados e refinados através de simulações que reproduzam a dinâmica do *foldings*. Se o conjunto de parâmetros for bom, a molécula, que na simulação é lançada numa conformação inicial arbitrária, deverá ser capaz de encontrar a sua estrutura nativa ao fim de um certo tempo.

As simulações por DM são extremamente exigentes do ponto de vista computacional. Se pensarmos que precisamos de um dia de CPU para simular um nanossegundo do processo de *foldings*, o que é uma estimativa razoável tendo em conta as capacidades de cálculo dos computadores de que dispomos actualmente, se uma proteína consumir 10^4 nanossegundos para encontrar o seu estado nativo, então precisaríamos de 10^4 dias de CPU, ou seja cerca de 30 anos, para simular o processo de *foldings* na totalidade. Isto é obviamente muito tempo para se esperar por apenas um resultado! É claro que existem super-com-

putadores capazes de simular mais do que um nanossegundo por dia e foi a eles que os dois cientistas americanos Yong Duan e Peter Kollman recorreram, em 1998, para fazerem a simulação por DM que é ainda hoje considerada “o estado da arte” nesta área. Estes investigadores simularam, ainda que apenas parcialmente, o processo de *folding* de uma pequena proteína com 36 resíduos e doze mil átomos, na presença de água, durante 1000 ns, o que correspondeu, na prática, à utilização de quatro meses de cálculo de CPU.

Folding *in silico* 2: modelos de rede e o conceito de paisagem de energia

Do que foi exposto anteriormente fica claro que se uma simulação por DM for bem sucedida, conduzindo a proteína para o estado nativo, então ficamos a conhecer a relação sequência-estrutura para aquela proteína específica. Para qualquer outra proteína, teremos que fazer um estudo semelhante se quisermos modelar as interações reais entre os átomos que a compõem. Apesar de ser importante, este tipo de estudo tem a desvantagem de não permitir identificar os princípios gerais (universais) envolvidos no *folding* de proteínas, para além de ser muito exigente do ponto de vista computacional. Para tal devemos recorrer a modelos muito mais simples do que os usados em DM, os chamados modelos de rede. Nestes modelos, a proteína é representada apenas pela cadeia principal de aminoácidos e cada um destes por uma “esfera” que o identifica do ponto de vista químico. As esferas ocupam os nodos de uma rede e as ligações covalentes entre os aminoácidos, ao longo da cadeia, dispõem-se segundo as arestas desta rede (Fig. 2).

Dentro da classe dos modelos de rede, o modelo H-P (em inglês, *Hydrophobic-Polar*), que é um dos mais simples, tem como objectivo principal explorar o papel das interações entre os aminoácidos e as moléculas de água durante o processo de *folding*. Os aminoácidos podem ser classificados, de acordo com a sua afinidade para a água, em hidrofóbicos – que “não gostam” de água – ou polares – que “gostam” de água. No que se segue, veremos que, até mesmo sem fazer qualquer simulação computacional,

conseguimos perceber alguns aspectos importantes do problema do *folding* recorrendo a este modelo numa rede bidimensional. Uma diferença fundamental em relação aos modelos e simulações por DM é que neste caso não estamos interessados no detalhe à escala atómica; a proteína é modelada simplesmente como uma cadeia de esferas de diferentes cores conforme a sua espécie química. No caso do modelo H-P, essas esferas são de apenas duas cores diferentes. Na Fig. 2 os aminoácidos hidrofóbicos são representados a branco e os polares a preto. A energia de interacção, ϵ , entre aminoácidos é igual a zero, excepto entre pares de aminoácidos hidrofóbicos, para os quais vale -1. Além disso, e como as ligações entre os aminoácidos ao longo da cadeia não se quebram nunca durante o *folding*, apenas as interações ditas “de contacto” (a tracejado na Fig. 2) contam para a energia total de uma certa conformação, ou seja, de um certo arranjo geométrico da cadeia de aminoácidos sobre a rede. A energia total de uma conformação é assim a soma das energias de todos os pares de contacto que a conformação contém.

Na Fig. 3 estão representadas cinco conformações diferentes e as respectivas energias. Conformações diferentes podem ter energias diferentes (como é o caso de $\Gamma_1, \Gamma_2, \Gamma_3$) mas também podem ter a mesma energia, por exemplo, Γ_4 e Γ_5 . Quando duas conformações diferentes apresentam a mesma energia dizem-se degeneradas.

Para além da degenerescência, estes exemplos servem para ilustrar um fenómeno ainda mais interessante, o fenómeno da “frustração”. Consideremos os aminoácidos 5 e 7 e os seus vizinhos, nas conformações Γ_4 e Γ_5 , respectivamente. Em Γ_4 , o aminoácido 5, que é hidrofóbico, está em contacto com o aminoácido 8 da mesma espécie, numa interacção que é favorável a ambos, já que se trata de uma interacção que é estável do ponto de vista energético. Já os aminoácidos 6 e 7 não estabelecem qualquer interacção de contacto, e o aminoácido 4 estabelece uma interacção de contacto neutra (com energia igual a zero) com o aminoácido 1. Por outro lado, na conformação Γ_5 , de certa forma os papéis invertem-se, e é o aminoácido 7 que passa a interagir favoravelmente com o aminoácido 4,

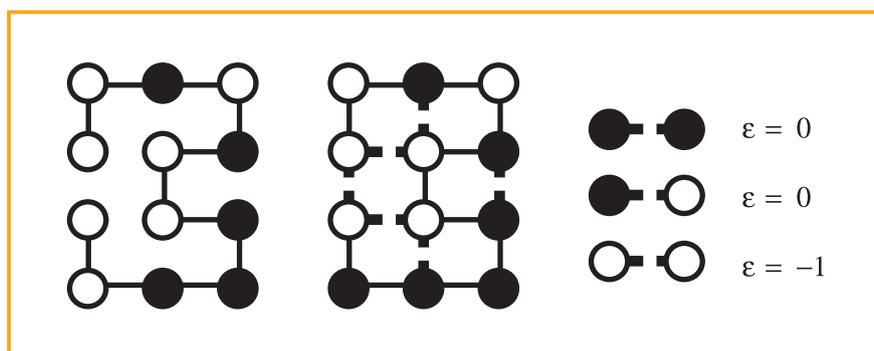


Fig. 2 Conformação de uma pequena proteína (péptido) e parâmetros de interacção no modelo H-P. Apenas as interações representadas a tracejado na figura do meio, contribuem para a energia total E da conformação representada, que é $E=-3$.

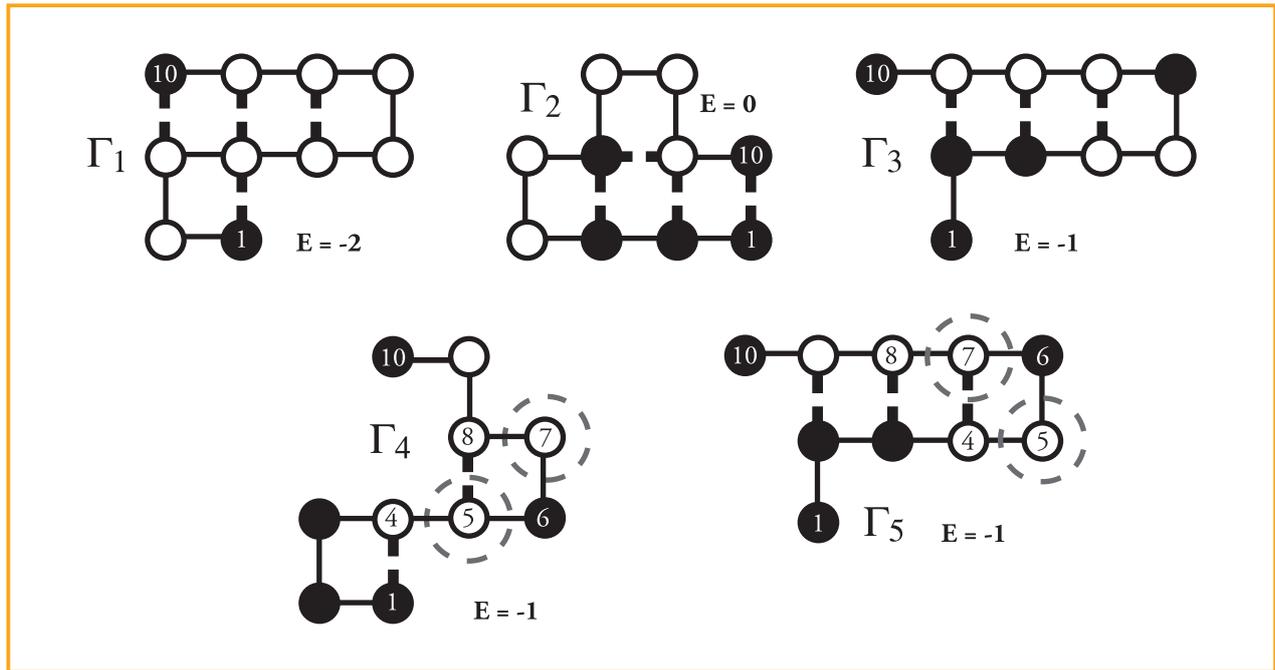


Fig. 3 Exemplo de cinco conformações diferentes e respectivas energias.

passando os aminoácidos 5 e 8 a estar desestabilizados. Diz-se que existe frustração, porque ao competirem entre si pelas posições que minimizam a energia de interação com os seus vizinhos, os aminoácidos não conseguem ficar todos igualmente estabilizados – não ficam igualmente “satisfeitos” com os seus parceiros de interação – em conformações que têm a mesma energia total. Ou seja, sem fazermos qualquer conta analítica ou simulação computacional o modelo H-P mostra que a degenerescência e a frustração são dois ingredientes fundamentais da energética do *fold*ing [2,3].

Chama-se paisagem de energia à dependência da energia na conformação, $E=E(\Gamma)$. Conformações de baixa energia, ou mínimos locais, encontram-se separadas umas das outras por barreiras de energia. A topografia desta paisagem de energia será tanto mais acidentada quanto maior for a frustração exibida pela proteína.

Um argumento simples mostra que, para uma proteína com N aminoácidos na rede quadrada, qualquer conformação fica completamente descrita por $(N-2)$ graus de liberdade (que são os vectores que fixam a posição dos aminoácidos) e que o número total de conformações acessíveis é $3^{(N-2)}$ [1]. Por exemplo, o peptídeo de 10 aminoácidos representado na Fig. 3 pode apresentar-se em $3^8=6651$ conformações possíveis (das quais apenas 5 estão representados na figura), o que quer dizer que a sua paisagem de energia contém 6651 pontos. Mas esse número sobe para $3^{98} \approx 6 \times 10^{46}$ conformações, se considerarmos uma proteína com $N=100$. Se, em vez da rede, estivermos no espaço

contínuo tridimensional, que é onde “vivem” as proteínas reais, os graus de liberdade necessários para descrever uma conformação – que neste caso passam a ser os comprimentos e os ângulos de ligação – deixam de tomar apenas valores discretos para passar a variar de uma forma contínua. Daqui resulta que, mesmo para proteínas pequenas, o número de conformações é *a priori* infinito e a função $E(\Gamma)$ toma valores num espaço cuja dimensão é ainda da ordem do número de aminoácidos da proteína. Todos os pontos desse espaço, em vez de apenas um número grande mas finito, correspondem, em princípio, a conformações possíveis. Apesar desta diferença, também neste caso temos degenerescência e frustração, sendo os mecanismos que as explicam os mesmos que os do modelo discreto.

Folding de proteínas: a perspectiva moderna

Resultados de estudos analíticos e de simulações computacionais com modelos de rede mostraram que as escalas de tempo biológicas são apenas compatíveis com frustração mínima e que a topografia global da paisagem de energia deve ser em forma de funil. A largura do topo do funil reflecte o conjunto de todas as conformações acessíveis a uma proteína no início do *fold*ing. Estas são as conformações menos estáveis – de maior energia. É possível atingir o estado nativo partindo de qualquer uma dessas conformações iniciais e percorrendo um dos muitos caminhos de *fold*ing alternativos. Cada um desses caminhos corresponde a um conjunto de conformações diferentes, pelas quais a proteína passa até chegar à conformação nativa. À medida que

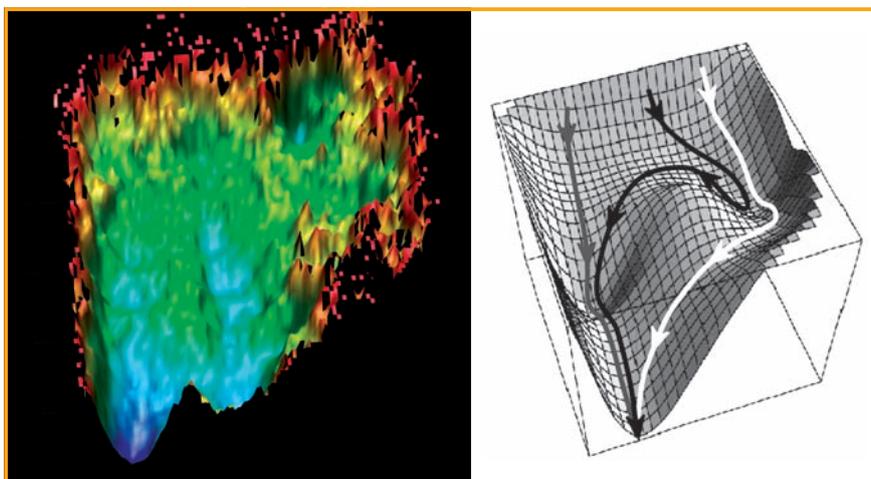


Fig. 4 Ilustração do conceito de paisagem de energia (esquerda) e paisagem de energia determinada experimentalmente (direita).

o *foldings* progride, não só a energia das conformações vai diminuindo, até atingir o mínimo global na conformação nativa, como também o próprio número de conformações acessíveis vai diminuindo até ser apenas um. Na realidade, a partir de certa altura, todas as trajectórias coalescem num caminho de *foldings* único, no final do qual se encontra o estado nativo (Fig. 4, esquerda).

Na Fig. 4 (direita) mostra-se a paisagem de energia de uma proteína real que foi determinada experimentalmente. Note-se que, neste caso, a proteína dispõe de três caminhos de *foldings* alternativos que conduzem todos ao estado nativo de forma diferente. A trajectória cinzenta é a trajectória ideal; é a que conduz mais rapidamente ao estado nativo. Já a trajectória branca envolve a transposição da maior barreira de energia desta paisagem, o que consome mais tempo e conduz a um *foldings* mais lento. Se seguir pela trajectória preta, a proteína ver-se-á a certa altura forçada a voltar atrás, passando por uma conformação que está mais longe da conformação nativa, para só depois “apanhar” o caminho rápido, ou seja a trajectória cinzenta.

Os conceitos de paisagem de energia e de funil de *foldings* conduziram àquilo a que se costuma chamar a “nova perspectiva do *foldings* de proteínas”, em oposição à perspectiva clássica que, como vimos, opunha Anfinsen (e a termodinâmica) a Levinthal (e a cinética). Na realidade, esta nova perspectiva reconcilia os pontos de vista daqueles dois cientistas. O estado nativo é realmente o estado termodinâmico mais estável – é o mínimo global da paisagem de energia – e a procura desse estado não é aleatória, embora não exista apenas um único caminho, como propôs Levinthal, mas sim vários, muitos caminhos de *foldings* possíveis.

A utilização de modelos de rede simples e a nova perspectiva do *foldings* de proteínas, que resultou largamente da aplicação destes modelos, foram sem dúvida ideias que, nos últimos dez anos, revolucionaram a nossa maneira de

pensar no mecanismo do *foldings*. Esta nova perspectiva, para além de conduzir a interpretações alternativas dos resultados experimentais clássicos, estimulou também, o que é talvez mais importante, o desenho de novas estratégias experimentais que têm permitido, a pouco e pouco, desvendar os detalhes destes complexos processos biológicos. Na realidade hoje em dia já sabemos muita coisa (os mais optimistas diriam até que se calhar já sabemos quase tudo) sobre o mecanismo do *foldings* de proteínas pequenas [4]. O mecanismo do *foldings* de proteínas grandes, esse, ainda permanece um grande mistério...

Agradecimentos

Este trabalho foi financiado pela Fundação para a Ciência e a Tecnologia no âmbito dos projectos SFRH/BPD/21492/2005 e POCI/QUI/58482/2004.

BIBLIOGRAFIA

1. P. F. N. Faísca, “O mistério da forma das proteínas”, *O Código Secreto: À Descoberta dos padrões da Natureza*, Coordenação de Margarida Telo da Gama, Guilherme Valente, Ed., Gradiva, 2005, pp. 301-326.
2. Hue S. Chan e Ken A. Dill, “The protein folding problem”, *Physics Today* **46**, 1993, pp. 24-32.
3. Hans Frauenfelder e Peter G. Wolynes, “Biomolecules: Where the physics of simplicity and complexity meet”, *Physics Today* **47**, 1994, pp. 58-64.
4. P. F. N. Faísca e M. M. Telo da Gama, “Folding of small proteins: a matter of geometry?”, *Molec. Phys.* **103**, 2005, pp. 2903-2910.