

Simulação de proteínas usando métodos de mecânica/dinâmica molecular

CLÁUDIO M. SOARES*

São descritas técnicas de simulação de proteínas utilizando métodos baseados em mecânica/dinâmica molecular. Estas técnicas consistem no desenvolvimento de funções de energia potencial molecular, ou campos de forças, que descrevem a energia das moléculas em diferentes conformações. Os diferentes termos das funções de energia potencial são introduzidos e o seu significado físico é elucidado. A partir destas funções de energia potencial molecular e de conceitos simples de mecânica clássica é possível simular a dinâmica molecular de proteínas em diferentes condições e utilizando diversas metodologias. Com estes métodos é possível estudar problemas conformacionais, problemas termodinâmicos e fazer estudos de índole mecanística em proteínas. Os fundamentos desses estudos são analisados e são descritas aplicações representativas destas metodologias.

1. INTRODUÇÃO

De todas as moléculas que constituem aquilo que chamamos vida, as proteínas são talvez as mais multifacetadas. É extraordinário como o mesmo tipo de organização molecular pode servir para as mais diversas funções; assim existem proteínas de transporte, como a hemoglobina, proteínas estruturais como o colágeno, proteínas com propriedades catalíticas, um grande grupo designado por enzimas, sendo a lisozima um bom exemplo, proteínas envolvidas em transformações mecânicas, como a actina e a miosina, proteínas que fazem canais em membranas, como as porinas, etc. A diversidade de funções apresentadas derivam da organização espacial a diversos níveis dos resíduos de α -L-aminoácidos e também das propriedades dinâmicas apresentadas por estas organizações espaciais. As proteínas são de nature-

za dinâmica e não estruturas estáticas. A "proteína rígida" é uma ideia que deriva muitas vezes das limitações experimentais, mas também é sugerida pelo modo como as proteínas são representadas a duas ou três dimensões. Para além das dificuldades de representação, falou-se igualmente em dificuldades experimentais. De facto, não existem técnicas experimentais que permitam, de uma forma directa, a obtenção de informação precisa sobre o movimento dos átomos na proteína. As técnicas estruturais mais divulgadas neste campo, a cristalografia por difracção de raios X e as técnicas de NMR multidimensional, possibilitam a obtenção de uma estrutura média, ou de várias estruturas: essa estrutura ou estruturas são o resultado de um *ensemble* (conjunto estatisticamente representativo) de moléculas ao longo do tempo de medida. Estas técnicas não possibilitam a obtenção de "filmes" detalhados do movimento molecular. No entanto, estas metodologias contêm elas mesmas dados que permitem inferir sobre a natureza dinâmica das moléculas: no caso da cristalografia por difracção de raios X temos os factores B, que quantificam de uma forma média o movimento atómico, e por vezes a existência de várias formas conformacionais para o mesmo resíduo ou grupo de resíduos. Ainda no caso da cristalografia por difracção de raios X, observa-se que diferentes cristais da mesma proteína podem revelar conformações diferentes [1], uma indicação clara de comportamento dinâmico. No caso do NMR multidimensional, o tipo de refinamento da estrutura origina múltiplas estruturas que respeitam as restrições obtidas experimentalmente: se nalguns casos esta multiconfiguracionalidade resulta de um conjunto insuficiente de dados experimentais, em muitos casos é um verdadeiro indicativo do comportamento dinâmico das proteínas. Em ambas as técnicas, o que se obtém são indicações de comportamento, mais do que movimentos detalhados com expressão temporal. Actualmente, para se obterem informações com detalhe atómi-

co sobre o comportamento dinâmico das proteínas as melhores técnicas não são experimentais, mas técnicas de simulação de dinâmica molecular. Daí o grande interesse despertado por este tipo de metodologias e a importância do seu desenvolvimento. No entanto, dada a natureza simplificada destas técnicas, é necessário ponderar a sua aplicabilidade caso a caso e sempre que possível comparar os resultados obtidos com os resultados experimentais.

Mas afinal qual é a importância do comportamento dinâmico das proteínas? Para além da função estrutural, deve haver muito poucas funções desempenhadas pelas proteínas que não sejam consequência do comportamento dinâmico ou onde este não intervenha. Para haver catálise enzimática, o substrato tem que entrar no centro activo, ser deformado para atingir o estado de transição [2, 3] e o produto tem que ser libertado. Muitas vezes o centro activo está internalizado no interior da proteína e para se dar a entrada do substrato é necessário que ocorram alterações conformacionais significativas [4]. No caso das proteínas transportadoras, as características dinâmicas são importantes de uma forma óbvia. Os diferentes estados observados em proteínas envolvidas em transporte transmembranar, tais como as porinas [5], têm que ter uma origem conformacional diferente, sendo esta origem uma consequência das condições físicas do meio [6].

Embora se possa pensar que a investigação e desenvolvimento neste campo seja um domínio essencialmente académico, cada vez mais este tipo de metodologias é aplicado na indústria, em especial na indústria farmacêutica, onde o seu uso é já indispensável no *design* ou concepção de fármacos.

2. MOVIMENTOS MOLECULARES

Até aqui falou-se de movimentos moleculares, mas ainda não se referiu o tipo, escala de tempo e im-

portância destes movimentos moleculares. As proteínas assim como as moléculas em geral, exibem diversos tipos de movimentos em diversas escalas de tempo, como se indica no quadro I:

Quadro I. Escalas de tempo dos períodos de movimentos moleculares [7]

vibração de ligações químicas	10 ⁻¹⁴ a 10 ⁻¹³ s
rotação de cadeias laterais (à superfície da proteína)	10 ⁻¹¹ a 10 ⁻¹⁰ s
movimentos do tipo "dobradiça" de pedaços da cadeia polipeptídica	10 ⁻¹¹ a 10 ⁻⁷ s
transições alostéricas e desnaturação local	10 ⁻⁵ a 10 s

As escalas de tempo dos movimentos moleculares podem ser bastante variadas, desde o fentosegundo (10⁻¹⁵s) até ao segundo. Como todos estes movimentos são importantes para a função proteica, estas escalas de tempo díspares colocam problemas muito particulares na simulação, pois não é possível estudá-las simultaneamente com apenas um método.

3. MODELAÇÃO MOLECULAR DE PROTEÍNAS USANDO DESCRIÇÕES MECÂNICAS.

Os modelos físicos utilizados para o estudo de proteínas são similares aos utilizados para outras moléculas, principalmente em fases condensadas. As particularidades destes modelos no caso das proteínas dependem essencialmente do número elevado de átomos que é necessário modelar.

Como qualquer sistema microscópico, a única maneira correcta de tratar uma proteína seria utilizar uma equação de Schrödinger dependente do tempo. Se tal já é muito difícil para moléculas pequenas, torna-se praticamente impossível para proteínas ou líquidos. Mesmo considerando a aproximação de Born-Oppenheimer (núcleos fixos), o número

de núcleos e electrões a tratar está várias ordens de grandeza acima daquilo que está hoje ao alcance das técnicas vulgarmente utilizadas em química quântica [8]. Para além disso, este tipo de técnicas não são em muitos casos suficientemente precisas (devido às aproximações que é necessário fazer) para darem valores quantitativamente aceitáveis para as forças moleculares.

Devido às dificuldades acima enunciadas, para simular moléculas ou agregados com um número considerável de átomos foram desenvolvidos modelos simplificados, de natureza empírica, que podem ser calibrados para reproduzirem dados experimentais (ou de natureza químico-quântica) com uma precisão aceitável para as utilizações a que são destinados. Designam-se normalmente estes modelos funções de energia potencial molecular ou campos de força. São muitas vezes referidos como modelos de física clássica para simulação molecular, mas tal designação é incorrecta, pois as funções de energia não são deriváveis de leis fundamentais da física clássica, apenas a integração das equações do movimento é feita dentro do formalismo da física clássica. As funções de energia potencial, tal como vamos ver, são melhor descritas como representações empíricas da aproximação de Born-Oppenheimer, onde se considera que o estado fundamental das moléculas é descrito como uma função contínua das coordenadas atómicas [9]. De um modo geral, existindo uma função de energia potencial multinuclear, podem derivar-se as forças sobre cada um dos átomos e, por aproximação, resolver as equações do movimento do sistema para um determinado intervalo de tempo (ver adiante). Dado que os movimentos atómicos se dão na escala do fentosegundo e dado que o número de partículas a considerar é razoavelmente elevado, tais métodos necessitam de considerável capacidade de cálculo.

Históricamente a primeira simulação da dinâmica molecular de um sistema atómico foi feita em 1957

por Alder e Wainwright [10-12] em sistemas de esferas rígidas. No entanto, nestes sistemas as funções de energia potencial não eram contínuas (daí as esferas rígidas). A primeira simulação de dinâmica molecular considerando uma função contínua de energia potencial foi feita em 1964 por Rahman [13] usando um potencial de Lennard-Jones para simular argon líquido. Não é de admirar que a simulação molecular tenha tido um desenvolvimento tão tardio: tal é devido ao desenvolvimento dos meios computacionais que, como se sabe, são uma aquisição relativamente tardia na história da tecnologia. Desde então, inúmeras simulações de sistemas moleculares foram efectuadas.

Embora antes se tivessem utilizado funções de energia potencial para proteínas em técnicas de minimização de energia [14-16], a primeira simulação de dinâmica molecular de uma proteína só foi efectuada em 1977 por McCammon, Gelin e Karplus [17]. Este estudo consistiu em nove picosegundos de simulação da dinâmica molecular do inibidor da tripsina pancreática bovina (BPTI) em vácuo. Esta simulação foi o "primeiro olhar" sobre o modo como as proteínas se comportam à escala atómica: o comportamento era semelhante a um fluído, certas regiões tinham diferentes graus de flexibilidade, o movimento era largamente não-harmónico, etc. Esta simulação foi mais tarde extendida a 100ps [18], tendo sido confirmados os resultados anteriores, mas desde logo se verificou que as proteínas tinham movimentos com flutuações longas no tempo, maiores que o tempo total da simulação, demonstrando as limitações das escalas de tempo simuladas.

Desde o entusiasmo gerado por estes estudos iniciais, a simulação molecular de biomoléculas tornou-se extremamente popular tendo sido desenvolvidos modelos mais sofisticados e atingidas escalas de tempo mais longas. No entanto, existem ainda muitas limitações que é necessário resolver.

4. CAMPOS DE FORÇAS

As funções de energia potencial ou campos de forças são funções contínuas das coordenadas nucleares dos átomos que constituem o sistema molecular. Estas funções não dependem dos electrões como partículas sendo o efeito destes englobado nas propriedades globais dos átomos. Nem poderia ser de outro modo, dado que a dinâmica electrónica não pode ser descrita com as equações determinísticas do movimento da mecânica clássica - uma consequência do princípio de incerteza de Heisenberg. Em relação ao movimento nuclear, dadas as massas envolvidas serem muito mais elevadas que a do electrão, a sua dinâmica pode ser descrita, em muito boa aproximação, pelas leis clássicas. Mesmo assim, a descrição dos átomos de hidrogénio é problemática.

As funções de energia potencial representam a energia potencial de um sistema molecular numa determinada conformação, ou seja de um conjunto de coordenadas cartesianas de todos os átomos do sistema. Estas funções de energia potencial molecular são normalmente compostas por vários termos com um significado físico mais ou menos preciso. Obviamente existem diferenças entre as funções e os parâmetros dos diferentes campos de forças, mas em geral existem muitas semelhanças. Como exemplo, vai-se falar do caso representativo do campo de forças GROMOS [19]. A função de energia potencial deste campo de forças é a seguinte:

$$V(\mathbf{r}_1, \dots, \mathbf{r}_{N_{at}}) = \sum_{n=1}^{N_b} \frac{1}{2} K_{b_n} (b_n - b_{0n})^2 +$$

$$\sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} (\theta_n - \theta_{0n})^2 +$$

$$\sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} (\xi_n - \xi_{0n})^2 +$$

$$\sum_{n=1}^{N_\phi} [1 + \cos(n_n \phi_n - \delta_n)] +$$

$$\sum_{i < j}^{N_a} \left[\frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \right] +$$

termos especiais

A função V é função das coordenadas generalizadas $\mathbf{r}_1, \dots, \mathbf{r}_{N_{at}}$, sendo N_{at} o número de átomos do sistema, e contém vários termos. Mas, a que correspondem e que significado tem cada um dos termos envolvidos na equação anterior? Assim o primeiro termo,

$$\sum_{n=1}^{N_b} \frac{1}{2} K_{b_n} (b_n - b_{0n})^2$$

corresponde à soma dos potenciais de todas as N_b ligações covalentes do sistema, e é uma função quadrática descrita pela distância, b_n , entre os dois átomos, por uma constante de força para cada tipo de ligação, K_{b_n} e por uma distância de equilíbrio, b_{0n} . Este potencial mantém as ligações covalentes em torno da distância de equilíbrio. Por ser uma função quadrática, imediatamente se infere que não pode haver dissociação, e que este potencial não pode ser aplicado a estados vibracionais de muito alta energia.

Analogamente existem termos de potencial aplicado aos N_θ ângulos de ligação do sistema

$$\sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} (\theta_n - \theta_{0n})^2,$$

sendo cada termo caracterizado por de uma constante de força, K_{θ_n} e por um ângulo de equilíbrio, θ_{0n} , para além de obviamente depender do ângulo θ_n formado entre os três átomos. As características são pois semelhantes às do potencial de ligação covalente.

Para acabar de definir a estereoquímica de um sistema molecular falta apenas definir os ângulos diedros, isto é, as interações do tipo ligante definidas entre quatro átomos. É necessário introduzir dois tipos de diedros: os que definem geometrias sem transição, normalmente designados diedros impróprios, e os que definem geometrias com transição, ou periódicas, designados por diedros próprios. Os primeiros são definidos pelo seguinte potencial:

$$\sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} (\xi_n - \xi_{0n})^2,$$

Como se vê, são mais uma vez funções quadráticas com uma constante de força K_{ξ_n} e um ângulo de equilíbrio ξ_{0n} . Este tipo de diedros serve, por exemplo, para definir quiralidades em tetraedros ou planaridade em anéis, tal como está ilustrado na seguinte figura:

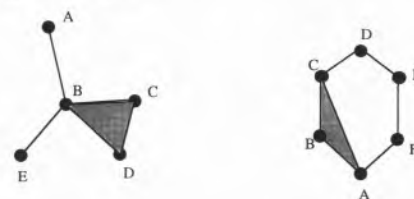


Fig. 1 - Uso de diedros impróprios. No caso da esquerda, o potencial de diedro impróprio serve para manter o ângulo ξ_n formado pela ligação AB com o plano BCD (o mesmo se aplicada a BE). Assim a quiralidade é mantida. No caso da direita, ilustra-se o uso do potencial de diedros impróprios para manter a ligação CD no plano ABC. Neste caso o ângulo de equilíbrio será zero para manter a planaridade.

Finalmente temos o caso do potencial de diédros próprios, que, pela primeira vez no caso dos potenciais ligantes, não é uma função quadrática:

$$\sum_{n=1}^{N_\phi} [1 + \cos(n_n \phi_n - \delta_n)].$$

Este tipo de potencial, que se aplica a todos os N_ϕ diédros do sistema, contém uma constante de força, K_{ϕ_n} , mas contém para além do ângulo de equilíbrio, δ_n , um outro termo, n_n , que é designado por multiplicidade do diedro. A expressão acima representa obviamente uma função periódica. Estas funções aplicam-se em torno de ligações simples (embora haja excepções), onde a rotação é possível, mas onde as diferentes posições de rotação não têm a mesma energia. O caso do etano é ilustrativo deste tipo de interações: tem dois extremos em energia, a configuração em estrela e a configuração em eclipse, sendo a última a de mais alta energia. Dada a simetria do sistema, temos três possíveis configurações de cada um dos tipos, daí que a função de energia potencial seja degenerada (periódica) e daí a razão do uso de uma constante de multiplicidade.

Os termos descritos até agora definem o sistema químico tal como ele é localmente. No entanto, quando passamos para moléculas maiores que o alcance destes termos locais, ou quando consideramos as interacções intermoleculares, têm que ser considerados outro tipo de funções para garantir a descrição destas mesmas interacções. São estas normalmente as interacções de van der Waals e as interacções electrostáticas, descritas no penúltimo termo da função de energia potencial e que, em geral, se designam interacções não ligantes.

As interacções de van der Waals são normalmente descritas por uma equação de Lennard-Jones:

$$\frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^6}.$$

Esta expressão tem um termo repulsivo que depende da distância r_{ij} entre os dois átomos elevado à potência de doze, e um termo atractivo que depende igualmente desta distância, mas elevada à potência de seis. Ambos os termos contêm constantes de interacção, $C_{12}(i,j)$ e $C_6(i,j)$, que dependem dos dois tipos de átomos em interacção. A parte repulsiva evita que os átomos colapsem um no outro e a parte atractiva corresponde às forças de dispersão de London, que são do tipo dipólo induzido-dipólo induzido. A combinação destes dois termos origina um mínimo de energia potencial a uma determinada distância entre os dois átomos.

O potencial electrostático nos campos de forças mais comuns é normalmente descrito com uma equação de Coulomb:

$$\frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}.$$

Esta depende das cargas dos dois átomos em interacção, q_i , q_j , da constante dieléctrica relativa ϵ_r (para além da permitividade eléctrica do vácuo, ϵ_0) e do inverso da distância entre os dois átomos, r_{ij} . Devido à dependência do inverso da distância, o potencial electrostático tem um alcance longo.

Os dois tipos de potenciais não ligantes acima descritos não se aplicam a todos os pares de átomos do sistema molecular. De facto, considera-se que a nível local, o sistema é apenas descrito pelas interacções ligantes (ligações covalentes, ângulos de ligação, diedros impróprios e próprios). Para as interacções electrostáticas, todos os pares de átomos entre os quais existe uma interacção do tipo ligante são pares excluídos. Para as interacções de van der Waals, todos os pares com uma distância menor que três ligações covalentes são excluídos, mas os que estão a uma distância de três ligações (podendo portanto formar um diedro) são contabilizados, embora com termos de interacção diferentes dos normais.

Existem campos de forças moleculares para todos os gostos e tal estado de coisas deriva do próprio carácter empírico destas funções, que implica que se derivem os parâmetros intervenientes para se ajustar os modelos aos dados experimentais. Os mais populares em proteínas são o CHARMM [20], o AMBER [21, 22] o GROMOS [19] o OPLS [23], CFF-DISCOVER [24, 25], TRIPOS [26]. São bastante equivalentes em termos de funcionais, embora possam conter diferenças substanciais nos parâmetros utilizadas. Deve-se notar que qualquer campo de forças só deve ser aplicado no tipo de moléculas para as quais foi desenvolvido. Para se obterem os parâmetros adequados para a simulação de um tipo de moléculas é necessário recorrer a um grande número de resultados experimentais ou de natureza químico-quântica. Estes resultados podem ser estruturas de moléculas modelo, estruturas de macromoléculas, dados de espectroscopia (por exemplo infravermelho), entalpias de formação, entalpias de sublimação, momentos dipolares, cargas obtidas por métodos quânticos, etc. O campo de forças tem que ser capaz de reproduzir estes resultados experimentais e para tal, muitas vezes ajustam-se os parâmetros por tentativa e erro até se obter o efeito dese-

jado. Existem no entanto métodos "racionalis" para derivar campos de forças autoconsistentes [9]. As relações com os dados experimentais são em muitos casos óbvias, tal como nos casos em que comprimentos de ligação de equilíbrio são introduzidos, mas noutros casos é necessário recorrer a algum formalismo químico-físico.

5. USO DE CAMPOS DE FORÇAS

5.1 Minimização de energia

A função de energia potencial molecular pode ser utilizada de diversas maneiras. Uma das mais simples consiste em utilizar a função (e eventualmente a sua derivada) para se fazer minimização de energia, isto é, mover as coordenadas atómicas de modo a decrescer a energia potencial do sistema. Conformações com energia potencial mais baixa correspondem, dentro de certos limites, a conformações mais representativas da molécula em questão. No caso de proteínas, a minimização de energia não pode ser usada como um método geral de determinação de estrutura devido ao largo espaço conformacional. De notar que a minimização peca por uma irregularidade teórica fundamental que pode ser importante em muitos casos: a função a minimizar deveria ser a energia livre e não a energia potencial. Infelizmente as descrições de energia livre estão muito menos desenvolvidas do que as descrições de energia potencial, de modo que tal processo é difícil.

5.2 Metropolis Monte-Carlo

Outro dos usos frequentes de funções de energia potencial é o designado método de Metropolis Monte-Carlo [27]. Neste método, as coordenadas iniciais do sistema molecular são modificadas pela adição de um pequeno vector com uma direcção aleatória. A aceitação desta modificação depende do valor da diferença entre a energia final e a energia inicial, ΔE :

a) Se $\Delta E \leq 0$ o movimento é aceite.

b) Se $\Delta E > 0$ o movimento é aceite com uma probabilidade igual a $\exp(-\frac{\Delta E}{kT})$. Esta probabilidade é uma probabilidade de Boltzmann.

Se a condição a) origina um método similar à minimização de energia, a condição b) introduz um novo factor, a temperatura do sistema, e baseada nesta temperatura, vem uma probabilidade de transição entre os dois estados, o final e o inicial. Esta introdução da temperatura introduz algo verdadeiramente importante que é a possibilidade de o sistema aceder a configurações com um valor mais alto de energia, dependendo de um factor de probabilidade. Esta metodologia permite ao sistema libertar-se de mínimos locais e de barreiras de potencial, factores que afectam grandemente os métodos de minimização de energia.

Para além destas vantagens no campo da pesquisa conformacional, o método de Metropolis Monte-Carlo tem grande importância do ponto de vista termodinâmico, pois permite que o sistema experimente conformações com probabilidade de acordo com a distribuição de Boltzmann de energias à temperatura em questão. Se o método for executado durante um número suficiente de passos, a função de partição do sistema (uma função das configurações possíveis do sistema) pode efectivamente ser calculada e dela podem ser derivadas importantes funções termodinâmicas, tais como a energia livre. Infelizmente, isso apenas pode ser feito para sistemas de dimensão reduzida.

Muito embora a sua simplicidade e grande número de aplicações conhecidas, o método de Metropolis Monte-Carlo é considerado inferior à simulação de dinâmica molecular [28] em termos de eficiência de pesquisa conformacional.

5.3 Dinâmica molecular

Finalmente chegamos àquilo que é considerado o método mais

comum nos métodos baseados em campos de forças: simulação da dinâmica molecular. Embora não tão simples como nos métodos descritos anteriormente, os conceitos envolvidos em simulação de dinâmica molecular são conceitos comuns a qualquer aplicação mecânica. Assim, podemos começar com a clássica equação de Newton:

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2},$$

$$\frac{d^2 \mathbf{r}_i(t)}{dt^2} = \frac{\mathbf{F}_i}{m_i}$$

onde m_i e \mathbf{r}_i são a massa e a posição da partícula i e \mathbf{F}_i o vector força actuando em i no instante de tempo t . A força sobre a partícula i , sendo conhecida a função de energia potencial $V(\mathbf{r}_1, \dots, \mathbf{r}_n)$, pode ser obtida pela relação:

$$\mathbf{F}_i = -\frac{\partial V(\mathbf{r}_1, \dots, \mathbf{r}_n)}{\partial \mathbf{r}_i},$$

ou de uma maneira mais geral:

$$\mathbf{F} = -\nabla V(\mathbf{r}_1, \dots, \mathbf{r}_n).$$

A equação do movimento e a equação da força são suficientes para descrever a dinâmica de um sistema de partículas, tal como é um sistema molecular. Infelizmente não existem soluções exactas para a dinâmica deste sistema, o que significa que tem que se recorrer a métodos aproximados. Em geral, estes baseiam-se no conhecimento das posições e velocidades de um sistema a um determinado tempo t , e resolvem os sistemas de equações para o tempo $t+\Delta t$, e assim sucessivamente. Como exemplo, consideremos o mais simples desses métodos, o método de Euler:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{\mathbf{F}_i}{m_i} \Delta t$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t) \Delta t$$

Como se vê, as velocidades e posições em cada passo são usadas para o cálculo das velocidades e posições no passo seguinte. Na realidade o método de Euler é bastante ineficiente (embora seja ilustrativo) e no caso dos sistemas moleculares o mais usado é o método de Verlet [29] que se baseia em princípios similares. O intervalo de tempo Δt a usar tem que ser suficientemente pequeno para o algoritmo ser estável e, dada a dinâmica dos sistemas moleculares, Δt tem em geral que se situar abaixo do femtosegundo. Para muitos casos Δt é de 2 femtosegundos, sendo que para fazer uma simulação de 100ps são necessários 50000 passos de integração. Simulações na escala do nanosegundo são normalmente necessárias para muitos estudos, pelo que uma simulação em solvente (ver adiante) pode levar mais de um mês de cálculo contínuo numa *workstation* comum.

Com a "maquinaria" descrita até aqui, já seria possível proceder à simulação da dinâmica molecular de um sistema. No entanto, com o que foi descrito até agora só se poderiam fazer simulações a energia constante e com moléculas em vácuo, o que é, em muitas aplicações, algo manifestamente inadequado. É necessário pois introduzir métodos para incorporar controlo de temperatura e métodos para simular solvatação.

5.4 Controlo de temperatura

O método mais simples para manter a temperatura constante é escalar as velocidades do sistema de modo a obter-se a temperatura desejada. Este processo tem no entanto o inconveniente de introduzir discontinuidades no sistema e de ser fisicamente pouco realista. Alternativamente, um dos métodos mais usados em simulação de sistemas macromoleculares é o chamado formalismo do banho térmico, descrito por Berendsen e colaboradores [30]. Em termos simplificados, considera-se um acoplamento entre o sistema a simular e um banho térmico. Para tal considera-se a seguinte equação diferencial,

que tem que ser obedecida de modo a haver acoplamento:

$$\left(\frac{dT}{dt} \right)_{\text{banho}} = 2\gamma(T_0 - T).$$

T_0 é a temperatura que se pretende atingir e T a temperatura do sistema, sendo γ uma constante que tem o significado físico de uma fricção, constante para todas as partículas (tal interpretação deriva da equação de Langevin). Com tal equação diferencial pode chegar-se a uma constante de escalamento λ que é efectuado em cada passo de integração:

$$\gamma = \sqrt{\left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right) \right]} \quad \tau_T = \frac{1}{\gamma},$$

sendo Δt a constante de integração e τ_T uma constante com dimensões de tempo, e que se designa por constante de acoplamento. Esta constante pode ser interpretada como um tempo de meia vida de uma perturbação térmica e, como tal, quanto mais pequeno o seu valor mais “forte” é o acoplamento e mais eficiente é a manutenção da temperatura em torno do valor padrão T_0 .

5.5 Fronteiras usadas em simulação de dinâmica molecular - inclusão da solvatação.

A inclusão do solvente e os seus efeitos está muitas vezes relacionada com a definição de fronteiras para a interacção, por isso estes dois temas são incluídos no mesmo tópico. O solvente tem basicamente dois tipos de efeitos: o efeito mecânico ou de fricção e o efeito devido a interacções não ligantes (electrostáticas e van der Waals). Os modelos de solvatação existentes tentam reproduzir pelo menos uma destas características. Em teoria, o efeito mecânico poderia ser tratado com o auxílio de uma equação de Langevin, introduzindo termos de fricção e forças aleatórias de superfície (ver [31]). Já os

efeitos do solvente devidos às interacções não ligantes entre o solvente e a molécula a simular são de muito mais difícil modelação, não obstante existirem numerosos modelos desenvolvidos para o efeito, principalmente para os componentes mais fortes que são as interacções electrostáticas (para uma revisão ver [32]). É necessário aqui fazer um pequeno parêntesis sobre o cálculo de interacções não ligantes em grandes sistemas. O número de pares de átomos interactuantes num sistema molecular pode ser muito grande. Por exemplo se tivermos 1000 átomos, o número máximo de pares interactuantes é 499500. Se tal número até poderia ser manejável, deve referir-se que 1000 átomos correspondem a uma proteína pequena em vácuo, sendo de considerar muitos mais átomos numa simulação com solvente (ver adiante), pelo que o cálculo deste tipo de interacções pode tornar-se demasiado demorado. Grande esforço tem sido investido na redução da dimensão deste problema sem perda da realidade física. O método mais utilizado é o de estabelecer um raio máximo em redor de cada átomo, a partir do qual se considera que as interacções são desprezáveis. Raios desde 8 Å até 14 Å são valores comuns. De notar que se, devido ao rápido decaimento com a distância, as interacções de van der Waals são realmente desprezáveis para além de distâncias como estas, tal já não acontece com as interacções electrostáticas, pelo que o uso deste tipo de aproximações é ainda um problema grave. Existem maneiras de reduzir a sua gravidade, mas mesmo assim continua a ser uma das maiores fraquezas da simulação molecular.

Voltando à questão dos efeitos de solvente em geral, obviamente que o método mais simples é ignorar esses efeitos de todo e fazer a simulação em vácuo. Esta aproximação tem no entanto grandes limitações, pois não se pode esquecer que as proteínas são polieletrólitos, muitas vezes com um número elevado de grupos ionizáveis à superfície. Na ausência do efeito dieléctrico do solvente e de

contra iões, estas cargas originam enormes forças que podem levar à “destruição” da proteína na simulação. Para moléculas sem grupos ionizáveis, tal metodologia pode no entanto ser adequada. Para evitar os efeitos indesejáveis acima descritos, pode-se considerar que os grupos carregados à superfície estão neutralizados pelos efeitos de solvente e reduzir-lhes a carga formal [33]. Outra aproximação é utilizar uma constante dieléctrica dependente da distância entre grupos interactuantes. Tal faz com que as interacções electrostáticas decresçam mais rapidamente com a distância e os efeitos destrutivos são eliminados. No entanto, estas aproximações e outras mais sofisticadas, não são adequadas para a maioria dos problemas, embora possam ter grandes vantagens no caso da pesquisa conformacional. Hoje em dia, o melhor meio de simular solvente é colocá-lo explicitamente na simulação. Tal consegue-se rodeando a proteína de modelos de moléculas de água. Esses modelos são criados da mesma forma que se criaram os modelos de potencial para simular proteínas. Conjuntamente com estes modelos introduzem-se fronteiras, que podem ser regiões rígidas ou pseudo-rígidas, ou podem ser condições periódicas de fronteira. Essas fronteiras servem para manter as moléculas do solvente confinadas ao espaço em redor da proteína e para manter a correcta densidade do líquido.

O uso de regiões rígidas ou pseudo-rígidas pode ser exemplificado na figura 2:

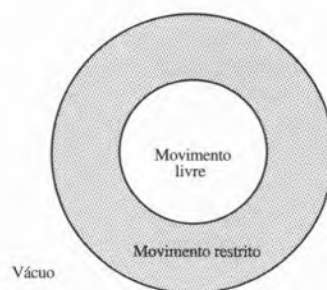


Fig. 2 - Representação esquemática do uso de regiões rígidas ou semi-rígidas.

Existe portanto uma região formada por solvente e/ou proteína que possui movimento restrito (o que pode significar que os átomos nessa zona oscilam em torno de posições de equilíbrio) e que delimita a região onde o movimento é livre. Fora disso existe o vácuo. A região onde se podem estudar as propriedades do sistema é apenas a região livre.

O uso de condições periódicas de fronteira é uma maneira mais elegante de simular solvatação. Consiste em circundar o sistema a simular por cópias dele próprio, como se pode ver na seguinte representação a duas dimensões (figura 3):

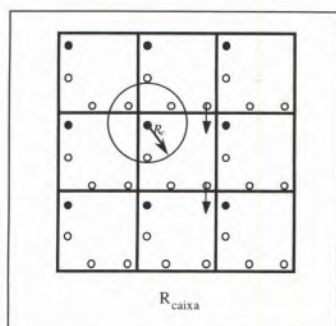


Figura 3.

Fig. 3 - Representação esquemática do uso de condições periódicas de fronteira a duas dimensões. O sistema central tem oito cópias dele próprio em seu redor. O átomo situado no centro da circunferência não "vê" as suas cópias nos sistemas circundantes, devido ao uso do raio de interação máxima R_c .

Esta metodologia, aliada ao uso de um raio de interação máxima para as interações não ligantes (R_c), como foi descrito anteriormente, faz com que o sistema em geral fique rodeado por uma "solução infinita" (dado que os átomos não "vêm" para além de R_c e R_c é menor do que R_{caixa}). Para além disso, com esta metodologia, um átomo que saia por um lado da "caixa" (como se vê na figura) entra pelo lado oposto da mesma, simulando efectivamente um contínuo. Obviamente que tal é um pseudo-contínuo, pois o sistema é periódico, mas dado o uso do raio R_c para as interações, os átomos do sistema não "vêm" tal periodicidade. A três dimensões, a construção

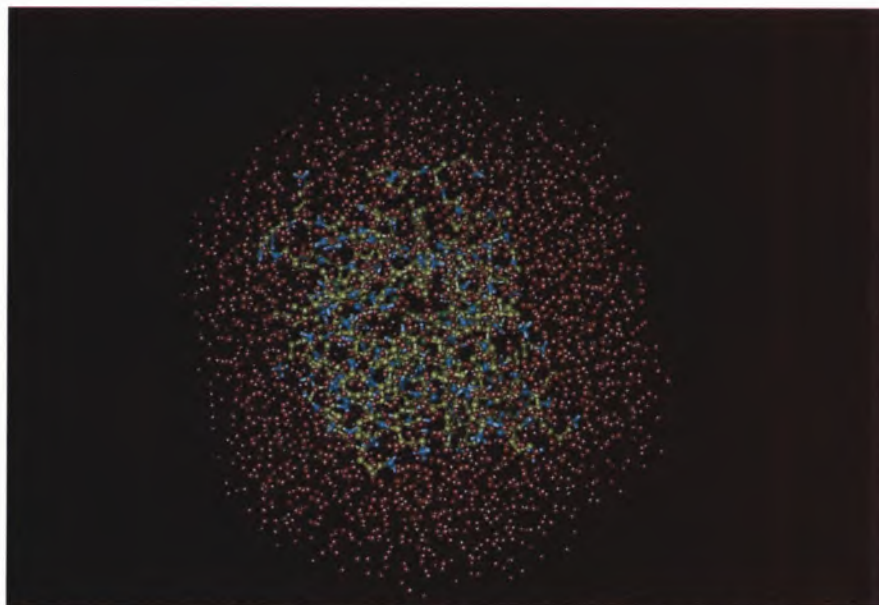


Fig. 4 - Citocromo c_3 dentro de uma caixa octaédrica de solvente. Apenas os átomos de oxigénio das moléculas de água estão representados para melhor visualização.

de tal metodologia pode assumir diversas formas. A mais simples é considerar um cubo com a proteína a simular solvatada numa quantidade suficiente de água (tipicamente a relação é de 1/10 (átomos da proteína)/(átomos de água)) e "replicar" este cubo em todas as direcções. Outras alternativas que proporcionam uma melhor eficiência de solvatação com menor número de moléculas de água são muitas vezes usadas. O octaedro truncado é um exemplo comum (Figura 4).

Relativamente à replicação das caixas em todas as direcções, deve-se dizer que tal não é feito na prática, mas dada a periodicidade do sistema é possível derivar algoritmos que entrem em conta com este dado sem aumentar o número de cálculos e o número de partículas. O processo é realmente eficiente e o uso deste tipo de solvatação é considerado hoje o mais avançado em simulação molecular. Mesmo assim, é necessário simular sistemas com dezenas de milhares de átomos o que continua a ser uma grande limitação.

De notar que com a inclusão de tal metodologia, existe outro factor que pode entrar em jogo, que é a pressão do sistema. A água tem valores de compressibilidade longe de

desprezáveis e nada garante que a pressão inicial do sistema, quando se rodeia de moléculas de água, seja a pressão adequada. Para tal, e de um modo similar ao usado com a temperatura, introduz-se um acoplamento a um banho externo de pressão [30]. Tal processo aumenta ou diminui a dimensão da caixa de simulação de modo a atingir a pressão desejada.

6. APLICAÇÕES

6.1 Movimentos moleculares

A compreensão do modo como as moléculas se movem é um fim em si mesmo. Como já foi dito anteriormente, nenhuma técnica experimental permite acompanhar a dinâmica molecular com o nível de detalhe acessível às técnicas de simulação. Os movimentos moleculares são múltiplos e variados. Como exemplo, na figura 5 apresentam-se pormenores da dinâmica do citocromo c_6 .

Para além de se observarem os fenómenos dinâmicos em si, também existe a possibilidade de se quantificarem as escalas de tempo características destes (e mesmo as barreiras de energia livre entre estados). Um exemplo clássico é a rotação de tiro-

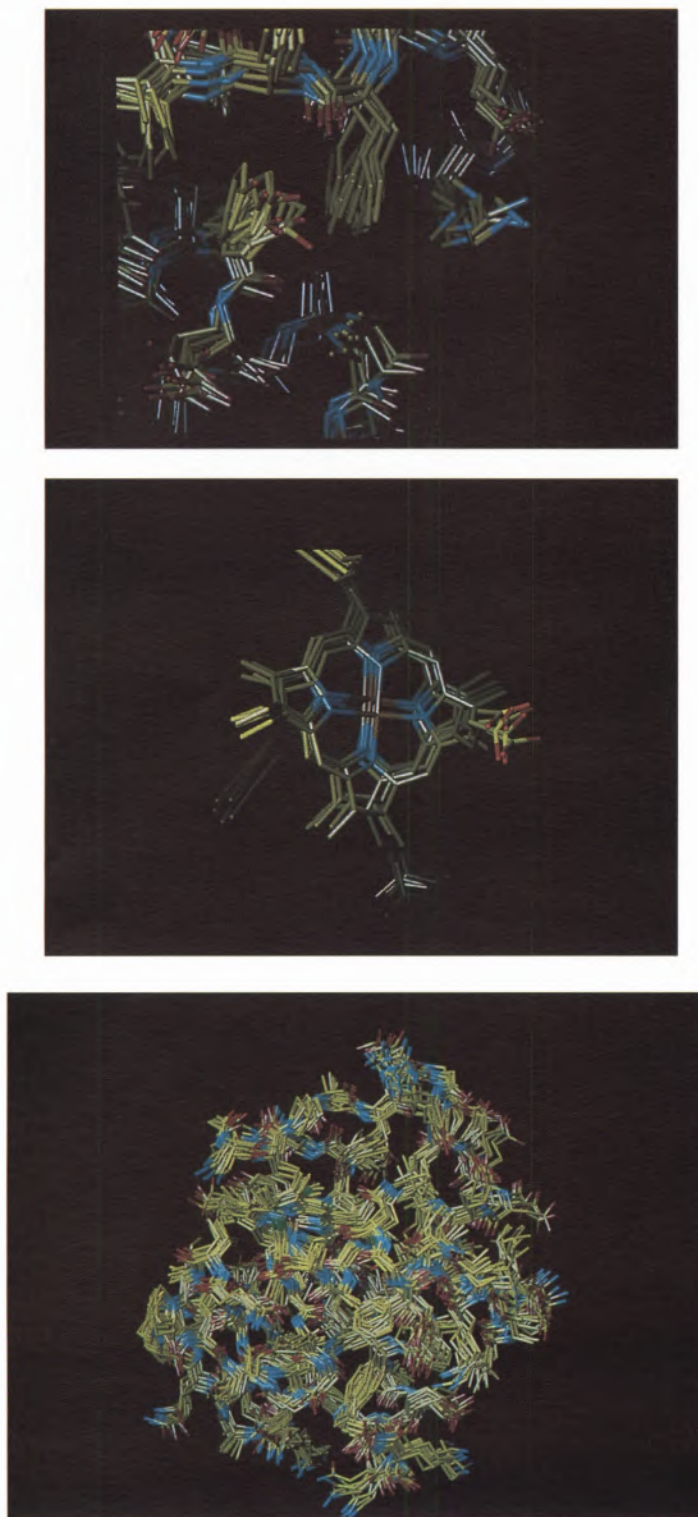


Fig. 5 - Pormenores da dinâmica do citocromo c_6 em solvente. Várias estruturas retiradas ao longo de uma simulação (separadas por 1 picosegundo) são sobrepostas usando modelos em que as ligações químicas se representam por segmentos unindo os átomos. Diferentes cores representam diferentes tipos de átomos. a. Dinâmica de cadeias laterais. b. Dinâmica do hemo (cadeias laterais não são representadas). c. Dinâmica da proteína.

sinas em proteínas [34]. O comportamento de pequenos ligandos e a sua interacção com as proteínas é também objecto de algum estudo. O movimento do monóxido de carbono dentro da mioglobina [35] é um exemplo deste tipo de estudos, embora aqui as técnicas de simulação de dinâmica molecular não sejam as habituais.

Movimentos de amplitude mais alargada em proteínas podem ser estudados com técnicas de simulação, mas muitas vezes requerem simulações mais longas ou o uso de técnicas especiais. Exemplos do estudo de vibrações de partes consideráveis da estrutura proteica existem na literatura [4, 36], tendo sido desenvolvidos métodos cujo objectivo é precisamente o tratamento deste tipo de movimentos [37, 38]. Estudos de ruptura da ligação entre duas moléculas [39] é outro exemplo interessante do estudo de movimentos alargados.

Dentro deste tema dos movimentos de longo período gostaria de referir o estudo de alterações conformacionais dependentes da electrostática do *loop* interno da porina de *Rodhobacter capsulatus* [6, 40]. As porinas formam largos canais transmembranares que podem controlar o seu estado de permeabilidade a pequenos solutos mediante alterações do potencial eléctrico da membrana onde se encontram. Com técnicas de dinâmica molecular, aliadas a perturbações electrostáticas, foi possível estudar transições reversíveis entre estados conformacionais com diferente permeabilidade do poro. Tal é exemplificado na figura 6:

6.2 Determinação conformacional

O estudo do *folding* de proteínas (modo como as proteínas adquirem a sua estrutura tridimensional) é um dos problemas fundamentais ainda não resolvidos da biologia molecular actual. A dinâmica molecular e outras técnicas relacionadas têm sido extensivamente utilizadas no estudo deste tipo de problemas, com vista a

desenvolver métodos de previsão de estrutura e a compreender o processo termodinâmico e cinético que tem lugar. No entanto, uma descrição completa do campo está além do âmbito deste artigo e dirijo o leitor para alguns trabalhos representativos [41-43].

A determinação da estrutura tridimensional *ab initio* (i.e., sem pressupostos iniciais) de proteínas a partir do conhecimento da sua sequência de resíduos de aminoácidos é difícil por duas razões: o grande número de graus de liberdade a considerar e a participação da solvatação em tal processo.

Para se determinar a conformação de moléculas em solução usando métodos físicos, é necessário recorrer a técnicas que aumentem a velocidade de pesquisa conformacional, pois a dinâmica molecular à temperatura fisiológica apenas consegue percorrer um número muito reduzido de estados conformacionais dentro da escala de tempo simulável. A técnica mais comum para acelerar este processo é o designado *simulated annealing* [44], cujo nome deriva de um procedimento comum em metalurgia [45]: se um metal é arrefecido subitamente torna-se frágil, o que é um indicativo que a estrutura ao nível atômico deste metal não está otimizada. Se pelo contrário o metal for arrefecido lentamente, torna-se muito mais difícil de quebrar. A relação entre este processo metalúrgico e a simulação molecular é simples: a simulação de moléculas complexas como as proteínas facilmente cai em mínimos locais da função de energia potencial devido à complexidade do espaço conformacional e devido a altas barreiras de energia entre estados. No entanto, se se simularem as moléculas a alta temperatura, essas barreiras podem ser transpostas, e se se arrefecer lentamente, a molécula tem tempo para pesquisar mais mínimos e acabar num hipotético mínimo global. De notar que não é no uso da alta temperatura que está o cerne do método, mas sim no arrefecimento lento.

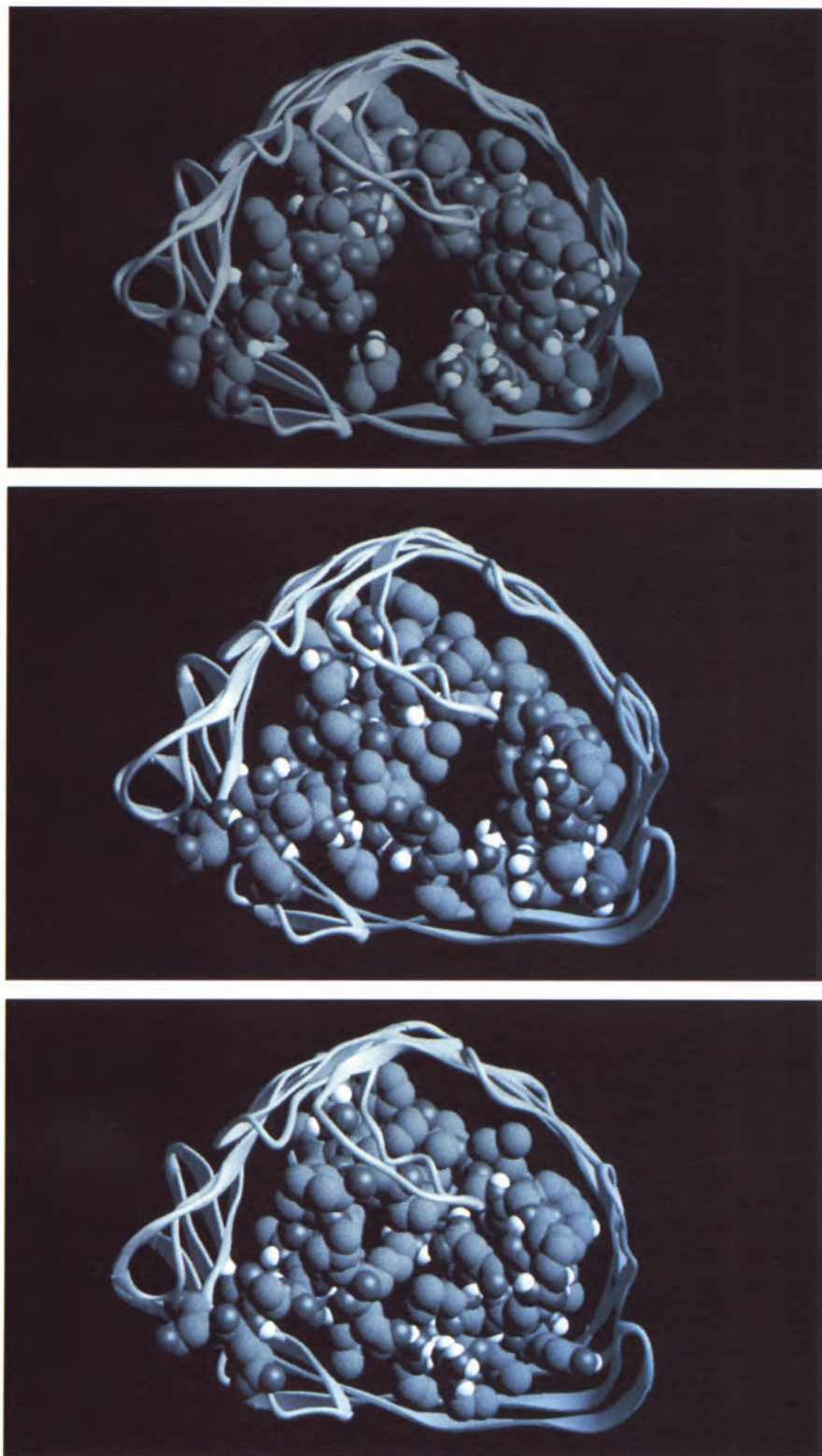


Fig. 6 - A figura a representa o estado aberto do poro (aquele encontrado na estrutura obtida por difracção de raios X). A figura b representa um estado reversível semi-fechado do poro e a figura c representa um estado completamente fechado. De notar que a proteína está representada apenas de um modo simplificado, como se fosse uma fita, apenas os átomos do loop interno estão destacados.

Mesmo com *simulated annealing*, o limite para a determinação da estrutura de péptidos livres em vácuo parece ser de não mais de 12 resíduos [46, 47] utilizando descrições completas (todos os átomos) da molécula. Como se vê o problema é ainda de muito difícil resolução. Existem técnicas, como o *Locally enhanced sampling*, que permitem tornar o processo mais eficiente, sendo mesmo possível a determinação de estruturas em solvente [48, 49].

A técnica de *simulated annealing* tem grande utilização nos processos de determinação e refinamento de estruturas de proteínas utilizando cristalografia de difração de raios X [50-52] e NMR multidimensional [53, 54]. Nestes casos, para além do campo de forças convencional, são introduzidos dados experimentais obtidos com estas técnicas (também numa formulação semelhante à do campo de forças), o que leva a que a determinação da estrutura final seja possível com este método de optimização. O campo de forças é usado como um conjunto de regras químicas que dirigem a pesquisa e o *simulated annealing* como uma mera técnica de optimização.

6.3 Métodos termodinâmicos

Talvez a mais interessante aplicação dos métodos de simulação consista nos designados métodos termodinâmicos, por um lado, pela sua elegância e potencialidades e, por outro lado, porque permitem correlacionar o mundo microscópico das técnicas de simulação com o mundo real das quantidades termodinâmicas, tal como a energia livre, por exemplo.

A derivação das bases em que assentam os métodos termodinâmicos aplicados neste contexto foi feita em 1954 por Zwanzig [55]. A descrição destes métodos exige uma formulação de mecânica estatística que, embora simples, é um pouco extensa para ser aqui apresentada na totalidade. Para revisões remeto o leitor para [56-58].

A energia livre de um sistema pode ser obtida a partir da função de partição. Na prática tal é difícil excepto para sistemas muito pequenos [8]. No entanto, se estivermos interessados não na energia livre, mas em diferenças de energia livre devidas a uma perturbação de um sistema padrão, existem formulações que nos permitem obter tal diferença. Consideremos um sistema regido por um hamiltoniano (isto é energia cinética + energia potencial) H_0 e consideremos que este hamiltoniano é afectado por um hamiltoniano de perturbação H_1 , de modo a gerar um sistema com hamiltoniano $H = H_0 + H_1$. H corresponde a um sistema perturbado de H_0 (imaginem-se por exemplo que se quer estudar o efeito da variação de uma carga parcial de um átomo na energia livre). A diferença de energia livre de Helmholtz $\Delta A = A - A_0$ pode ser obtida por:

$$\Delta A = A - A_0 = -kT \ln \left\langle e^{-\frac{H_1}{kT}} \right\rangle_0$$

Tal expressão correlaciona a diferença de energia livre entre o estado perturbado e o estado padrão, fazendo uma média da quantidade (é o que representam os símbolos $\langle \rangle$) $e^{-\frac{H_1}{kT}}$ no sistema regido pelo hamiltoniano H_0 (é o que significa o subscrito '0'). Essa média pode ser calculada a partir de uma simulação de dinâmica molecular do sistema padrão, ou uma simulação do tipo Metropolis Monte-Carlo.

O método como descrito tem grandes limitações, sendo a maior das quais a de ser válido apenas para perturbações pequenas [59]: quando a perturbação é grande, o sistema padrão não tem informação estatística sobre o estado perturbado, i.e., configurações prováveis do estado perturbado não ocorrem com suficiente probabilidade no estado padrão, o que é uma condição necessária para o método funcionar. Para resolver este problema é necessário dividir a perturbação em perturbações mais pequenas. Imagine-se que temos um estado inicial A e um esta-

do perturbado B, que, considerando o exemplo da carga, podem ser dois estados com cargas parciais diferentes. Para dividir a perturbação entre A e B em passos manejáveis, podemos considerar o seguinte potencial misto (consideremos que a energia cinética é constante)

$$V(\lambda) = (1 - \lambda)V_A + \lambda V_B \quad \lambda \in [0, 1]$$

sendo λ um parâmetro que "transforma" o sistema A no sistema B (por um processo que tem mais de alquímico do que químico!). Nestas circunstâncias, com uma formulação designada por perturbação estatística, a energia livre total da perturbação pode ser dada por

$$A_B - A_A = \sum_{\lambda=0}^1 -kT \ln \left\langle e^{-\frac{\Delta V}{kT}} \right\rangle_{\lambda}$$

onde

$$\Delta V = V(\lambda + d\lambda) - V(\lambda).$$

Note-se que agora as médias são calculadas para cada valor de λ considerado ao longo do intervalo. O número de passos depende do valor da perturbação.

De notar que o caminho de perturbação é completamente irrelevante para os fundamentos teóricos do método, pois a energia é uma função de estado, que só depende do estado inicial e do estado final.

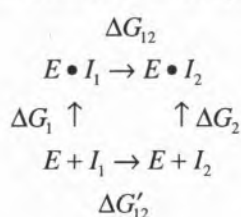
Outra metodologia que também se baseia no uso de um parâmetro de "transformação" é a designada integração termodinâmica artificial. Neste processo a diferença de energia livre entre os dois sistemas é dada por:

$$\Delta A = \int_{\lambda=0}^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda.$$

Neste caso a média a calcular é uma média de uma derivada parcial. Este processo tem a vantagem em relação à perturbação estatística, de que as médias a diferentes valores de λ não dependem dos valores de λ anteriores, podendo ser calculados independentemente.

São óbvias as potencialidades de aplicação destes métodos termodinâmicos, embora estes não possam ser considerados uma panaceia universal para qualquer problema. Falemos primeiro das limitações dos métodos: estas são basicamente duas, a primeira metodológica e a segunda computacional. Metodologicamente é necessário notar que qualquer perturbação implica uma mudança de uma função de energia potencial de um sistema para um sistema alterado. Portanto é sensível à qualidade dos parâmetros do campo de forças e ao modo como as forças são calculadas. A limitação computacional é a de que dados os extensos espaços conformacionais acessíveis às proteínas, a obtenção de médias representativas é computacionalmente muito demorada pois implica simulações longas.

Falemos agora das aplicações. A maioria das aplicações destes métodos em proteínas consiste no cálculo de diferenças de energia livre de ligação de inibidores (este é um problema muito comum no *design* de fármacos). Para efectuar estes tipo de cálculos e outros temos que considerar ciclos termodinâmicos [60]:



Este ciclo representa a ligação de dois inibidores I_1 e I_2 , a um mesmo enzima E . ΔG_1 e ΔG_2 representam as energias livres de ligação do inibidor 1 e 2 respectivamente. É a diferença entre essas duas energias livres que vai indicar se 1 se liga mais do que 2 (em termos termodinâmicos) ou vice-versa. Esta diferença pode ser expressa a partir do ciclo como:

$$\Delta G_1 - \Delta G_2 = \Delta G'_{12} - \Delta G_{12}$$

correspondendo $\Delta G'_{12}$ e ΔG_{12} às diferenças de energia livre envolvidas nos

processos não-físicos de transformação do inibidor 1 no inibidor 2 quando ligados à proteína ou em solução, respectivamente. Estes processos, embora não-físicos, são mais simples de implementação e de cálculo que os processos de ligação dos inibidores. Em resumo, faz-se a transformação dos inibidores em solução e na enzima e com as energias livres obtidas pode calcular-se a diferença entre as energias livres de ligação.

Alguns exemplos de aplicação destas metodologias em proteínas são a determinação da afinidade relativa de inibidores [61], sobre a estabilidade de proteínas mutantes [62], cálculo de potenciais redox [63] ou diferenças de potenciais redox [64], etc.

6.5 Estudo de reacções em enzimas

O estudo de reacções químicas usando metodologias baseadas em campos de força pode parecer impossível à primeira vista. Mas na verdade, quando aliadas a um método quântico, este tipo de estudos podem ser extremamente interessantes. Talvez o precursor deste tipo de estudos tenha sido Karplus em 1965 quando estudou a cinética de permuta da reacção $H+H_2 \rightarrow H_2+H$ usando trajetórias clássicas [65].

Mesmo os aspectos termodinâmicos das reacções dentro de enzimas (em termos de energias de activação por exemplo) podem ser estudados com estas técnicas. É necessário definir um caminho reacção e utilizar um método quântico para definir as propriedades das moléculas (cargas, constantes de força, geometrias, etc) ao longo desse caminho reacção. Depois podem aplicar-se os métodos termodinâmicos descritos anteriormente para se estudar, em termos quantitativos, como a proteína modula a energia livre ao longo desse caminho de reacção (em relação ao vácuo, ou solução). Warshel foi pioneiro neste tipo de estudos e um exemplo é o mecanismo da reacção catalizada pela anidrase carbónica [66].

Métodos dinâmicos foram tam-

bém usados para observar o modo como a proteína reage após uma reacção química. Um exemplo deste tipo de estudos são os trabalhos realizados no centro reacção do fotosistema [67].

7. DIRECÇÕES FUTURAS

Dado o ponto em que o campo se encontra é de prever avanços derivados de dois tipos de factores: desenvolvimento de metodologias e progressos ainda constantes no poder computacional. Em relação ao primeiro é óbvio que os métodos podem ser desenvolvidos com melhores parametrizações dos campos de forças, inclusão de importantes tipos de interacções, tais como, por exemplo, a polarizabilidade, melhores métodos de cálculo de interacções e, possivelmente, desenvolvimentos no campo de solvatação implícita (sem o uso de modelos de moléculas de água). Em relação ao sempre crescente aumento do poder computacional esse continua a ser um factor importante a considerar no campo da simulação molecular, pois este tipo de técnicas tem estado sempre nos limites do calculável. O desenvolvimento de meios de cálculo mais rápidos possibilitaria o acesso a novas escalas de tempo e o estudo de novos tipos de problemas. O futuro destas técnicas vai certamente ter grande impacto e importância na engenharia de proteínas, cujo objectivo último é fazer um *design* racional de macromoléculas com características úteis para os seus "criadores". Para tal é necessário entender a física destes sistemas e sonhar: "Proteins are dynamic molecules; a protein engineer must be a choreographer rather than a sculptor. The time has come to teach proteins new dances" [68].

AGRADECIMENTOS

Quero agradecer a Isabel Solange Oliveira, Joaquim Mendes e Paulo Martel

pela revisão do manuscrito e por sugestões. Agradeço também o apoio financeiro do JNICT e do PRAXIS XXI.

* Instituto de Tecnologia Química e Biológica, UNL,
Apartado 127, 2780 Oeiras, Portugal
Tel: 01 4417823, Fax 01 4411277,
e-mail: claudio@itqb.unl.pt

REFERENCIAS

1. Faber, H. R. and Matthews, B. W., *Nature* **348** (1990) 263-266.
2. Tapia, O., *J.Math.Chem.* **10** (1992) 139-181.
3. Tapia, O. and Andres, J., *J.Molecular Structure* **335** (1995) 267.
4. Gilson, M. K., Straatsma, T. P., McCammon, J. A., Ripoll, D. R., Faerman, C. H., Axelsen, P. H., Silman, I. and Sussman, J. L., *Science* **263** (1994) 1276-1278.
5. Schindler, H. and Rosenbusch, J. P., *Proc.Natl.Acad.Sci. USA* **75** (1978) 3751-3755.
6. Soares, C. M., Björkstén, J. and Tapia, O., *Protein Engineering* **8** (1995) 5-12.
7. McCammon, J. A. and Harvey, S. C. Dynamics of proteins and nucleic acids. Cambridge: Cambridge University Press. 1987
8. Brooks, C. L., Karplus, M. and Pettitt, B. M. Proteins: A theoretical perspective of dynamics, structure, and thermodynamics. New York: John Wiley & Sons. 1988.
9. Lifson, S. and Warshel, A., *J.Chem.Phys.* **49** (1968) 5116-5129.
10. Alder, B. J. and Wainwright, T. E., *J.Chem.Phys.* **27** (1957) 1208-1209.
11. Alder, B. J. and Wainwright, T. E., *J.Chem.Phys.* **31** (1959) 459-466.
12. Alder, B. J. and Wainwright, T. E., *J.Chem.Phys.* **33** (1960) 1439-1451.
13. Rahman, A., *Phys.Rev.* **136** (1964) A405-A411.
14. Gelin, B. R. and Karplus, M., *Proc.Natl.Acad.Sci.USA* **72** (1975) 2002-2006.
15. MacCammon, J. A., Gelin, B. R. and Karplus, M., *Nature* **262** (1976) 325-326.
16. Gelin, B. R. and Karplus, M., *Biochemistry* **18** (1979) 1256-1268.
17. McCammon, J. A., Gelin, B. R. and Karplus, M., *Nature* **267** (1977) 585-590.
18. Karplus, M. and McCammon, J. A., *Nature* **277** (1979) 578.
19. van Gunsteren, W. F. and Berendsen, H. J. C. Groningen molecular simulation (GROMOS) library manual. Biomos B.V. Nijenborgh 16, 9747. AG Groningen, The Netherlands. 1987.
20. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M., *J.Comp.Chem.* **4** (1983) 187-217.
21. Weiner, P. K. and Kollman, P. A., *J.Comp.Chem.* **2** (1981) 287-302.
22. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P., *J.Am.Chem.Soc.* **106** (1984) 765-784.
23. Jorgensen, W. L. and Tirado-Rives, J., *J.Am.Chem.Soc.* **110** (1988) 1657-1666.
24. Mapple, J. R., Dinur, U. and Hagler, A. T., *Proc.Natl.Acad.Sci.USA* **85** (1988) 5350-5354.
25. Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M. and Hagler, A. T., *Proteins* **4** (1988) 31-47.
26. Clark, M., Cramer III, R. D. and Van Opdenbosh, N., *J.Comp.Chem.* **10** (1989) 982-1012.
27. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E., *J.Chem.Phys.* **21** (1953) 1087-1092.
28. Northrup, S. H. and McCammon, J. A., *Biopolymers* **19** (1980) 1001-1016.
29. Verlet, L., *Phys.Rev* **159** (1967) 98-103.
30. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. and Haak, J. R., *J. Chem. Phys.* **81** (1984) 3684-3690.
31. van Gunsteren, W. F. and Berendsen, H. J. C., *Angew. Chem. Int.* **29** (1990) 992-1023.
32. Smith, P. E. and van Gunsteren, W. F., Methods for the evaluation of long range electrostatic forces in computer simulations of molecular systems. In: "Computer Simulation of Biomolecular Systems. Theoretical and experimental applications.", van Gunsteren, W. F., Weiner, P. K. and Wilkinson, A. J. (eds.), Leiden: ESCOM, 1993:182-212
33. Åqvist, J., van Gunsteren, W. F., Leijonmarck, M. and Tapia, O., *J.Mol.Biol.* **183** (1985) 461-477.
34. McCammon, J. A., Wolynes, P. G. and Karplus, M., *Biochemistry* **18** (1979) 927-942.
35. Elber, R. and Karplus, M., *J.Am.Chem.Soc.* **112** (1990) 9161-9175.
36. Sanejouand, Y.-H. and Tapia, O., *J.Phys.Chem.* **99** (1995) 5698-5704.
37. Amadei, A., Linssen, A. B. M. and Berendsen, H. J. C., *Proteins* **17** (1993) 412-425.
38. van Aalten, D. M. F., Amadei, A., Linssen, A. B. M., Eijssink, V. G. H., Vriend, G. and Berendsen, H. J. C., *Proteins* **22** (1995) 45-54.
39. Grubmüller, H., Heymann, B. and Tavan, P., *Science* **271** (1996) 997-999.
40. Soares, C. M., Björkstén, J. and Tapia, O., *Molecular Simulation* **15** (1995) 35-46.
41. Daggett, V. and Levitt, M., *J.Mol.Biol.* **232** (1993) 600-619.
42. Hünenberger, P. H., Mark, A. E. and van Gunsteren, W. F., *Proteins* **21** (1995) 196-213.
43. Daggett, V., Li, A., Itzhaki, L. S., Otzen, D. E. and Fersht, A. R., *J.Mol.Biol.* **257** (1996) 430-440.
44. Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P., *Science* **220** (1983) 671-680.
45. Chou, K.-C. and Carlacci, L., *Prot.Eng.* **4** (1991) 661-667.
46. Brucoleri, R. E. and Karplus, M., *Biopolymers* **29** (1990) 1847-1862.
47. Wilson, S. R. and Cui, W., *Biopolymers* **29** (1990) 225-235.
48. Simmerling, C. and Elber, R., *J.Am.Chem.Soc.* **116** (1994) 2534-2547.
49. Simmerling, C. L. and Elber, R., *Proc.Natl.Acad.Sci.USA* **92** (1995) 3190-3193.
50. Brünger, A. T., Kuriyan, J. and Karplus, M., *Science* **235** (1987) 458-460.
51. Brünger, A. T., *J.Mol.Biol.* **203** (1988) 803-816.
52. Pannetier, J., Bassas-Alsina, J., Rodriguez-Carvajal, J. and Caignaert, V., *Nature* **346** (1990) 343-345.
53. Clore, G. M., Gronenborn, A. M., Brünger, A. T. and Karplus, M., *J.Mol.Biol.* **186** (1985) 435-455.
54. Clore, G. M., Brünger, A. T., Karplus, M. and Gronenborn, A. M., *J.Mol.Biol.* **191** (1986) 523-551.
55. Zwanzig, R. W., *J.Chem.Phys.* **22** (1954) 1420-1426.
56. King, P. M., Free energy via molecular simulation: a primer. In: "Computer simulation of biomolecular systems. Theoretical and experimental applications.", van Gunsteren, W. F., Weiner, P. K. and Wilkinson, A. J. (eds.), Leiden: ESCOM, 1993:267-314
57. Kollman, P., *Chem.Rev.* **93** (1993) 2395-2417.
58. Soares, C. M. and Tapia, O., *Molecular Eng.* **4** (1995) 415-430.
59. Mezei, M., *Mol.Phys.* **47** (1982) 1307-1315.
60. Berendsen, H. J. C., Postma, J. P. M. and van Gunsteren, W. (1984) Statistical mechanics and molecular dynamics: The calculation of free energy., University of North Carolina
61. Ferguson, D. M., Radmer, R. J. and Kollman, P. A., *J.Med.Chem.* **34** (1991) 2654-2659.
62. Yun-yu, S., Mark, A. E., Cun-xin, W., Fuhua, H., Berendsen, H. J. C. and van Gunsteren, W. F., *Protein Eng.* **6** (1993) 289-295.
63. Reynolds, C. A., *J.Am.Chem.Soc.* **112** (1990) 7545-7551.
64. Soares, C. M., Martel, P. J. and Carrondo, M. A., *manuscript* (1997).
65. Karplus, M., Porter, R. N. and Sharma, R. D., *J.Chem.Phys.* **43** (1965) 3259-3287.
66. Åqvist, J., Fothergill, M. and Warshel, A., *J.Am.Chem.Soc.* **115** (1993) 631-635.
67. Gehlen, J. N., Marchi, M. and Chandler, D., *Science* **263** (1994) 499-502.
68. Nikkila, M. J. Structural and functional studies on glutaredoxin and thioredoxin., Swedish University of Agricultural Sciences. Uppsala, Thesis (1991).