

2. MÉTODOS DE SIMULAÇÃO E O MÉTODO MONTE CARLO

O propósito deste Capítulo é contextualizar o presente problema no plano dos métodos modernos de tratamento numérico, pois a descrição do comportamento físico e químico de uma cadeia protéica em solução é praticamente impossível por aqueles métodos analíticos usualmente empregados para moléculas mais simples. Por isso, métodos computacionais robustos, como Dinâmica Molecular e Monte Carlo são indispensáveis. Portanto, nos tópicos seguintes a cadeia protéica e detalhes conformacionais, embora de forma simplificada, são detalhados, assim como é discorrido brevemente sobre as bases destes dois métodos empregados largamente na simulação de sistemas moleculares e, de forma particular, sobre o método Monte Carlo empregado neste trabalho.

2.1 O sistema protéico e suas representações estruturais

Como anteriormente mencionado, as proteínas naturais são macromoléculas formadas a partir de 20 aminoácidos que ocorrem na natureza (Figura 1.1, página 3). Por sua vez, todo aminoácido é constituído por um átomo de carbono, chamado carbono-alfa ($C\alpha$), ligado a quatro grupos químicos: (i) grupo carboxila ($COOH$), (ii) grupo amino (NH_2), (iii) a um átomo de hidrogênio (H) e (iv) a um radical (R), também denominado cadeia lateral do aminoácido, que é o grupo que distingue os 20 aminoácidos naturais entre si. Nas proteínas, os aminoácidos são unidos por ligações consecutivas entre o carbono da carboxila de um aminoácido e o nitrogênio do grupo amino do aminoácido seguinte, formando cadeias lineares. Esta ligação que une os aminoácidos é conhecida por ligação peptídica e é o resultado da condensação do grupo carboxila de um ácido com o grupo amino de outro, e concomitante eliminação de uma molécula de água⁴ (Figura 2.1).

A Figura 2.2 mostra as características especiais que distinguem a ligação peptídica. As dimensões das ligações entre os átomos de carbono, nitrogênio, oxigênio e hidrogênio são distintas entre si e também entre o átomo do carbono alfa e o grupo R_i que identifica cada aminoácido da seqüência. Os átomos entre dois $C\alpha$, conservam uma certa planaridade através de interações químicas denominadas peptídicas, e assim a conformação da cadeia polipeptídica é essencialmente descrita pelos ângulos de rotação phi Φ , entre o nitrogênio e o carbono alfa, e o psi Ψ , entre o carbono alfa e o carbono do grupo carboxila original¹² (Figura 2.2).

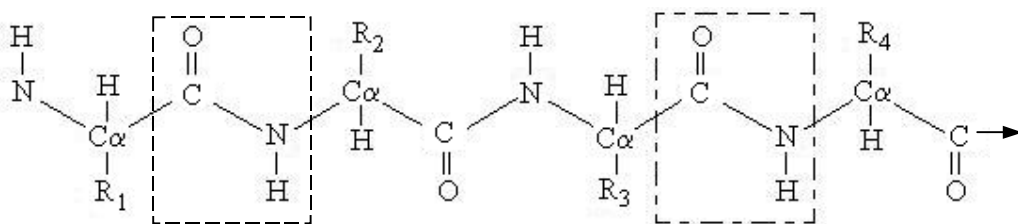


Figura 2.1. Ligação peptídica (retângulos tracejados). Associação entre o carbono da carboxila e o nitrogênio do grupo amino, com eliminação de água. A seta à direita indica a orientação convencional da cadeia.

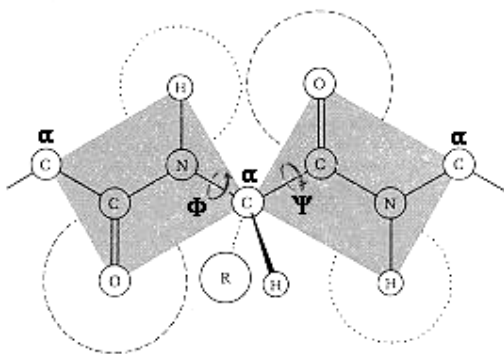


Figura 2.2. Ligação peptídica. Ângulos de rotação phi Φ entre o nitrogênio do plano peptídico da esquerda e o carbono alfa, e o psi Ψ entre o carbono alfa e o carbono do plano peptídico seguinte.

Uma proteína específica é identificada por sua seqüência particular de aminoácidos, a qual provê as informações específicas e fundamentais para a determinação de sua estrutura 3-D particular. Mas em geral, as proteínas são descritas através de quatro níveis estruturais: da estrutura primária à quaternária (Figura 2.3).

A estrutura “primária” corresponde à seqüência de aminoácidos da proteína propriamente dita, isto é, descreve sua estrutura linear ou unidimensional. Ela pode variar em dois aspectos fundamentais, definidos pela informação genética: número de aminoácidos (ou tamanho da cadeia) e seqüência particular dos aminoácidos (note que para uma proteína de comprimento L , tem-se então 20^L possíveis seqüências distintas). Esta seqüência de aminoácidos, com uma extremidade amino terminal e outra carboxi terminal, é reversível por hidrólise química (ou enzimática) das ligações peptídicas produzindo peptídeos menores ou mesmo aminoácidos livres¹².

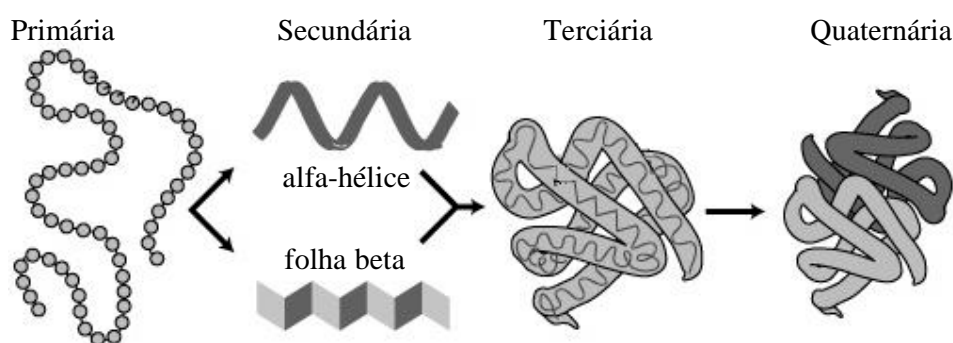


Figura 2.3. Os quatro níveis estruturais das proteínas.

A estrutura “secundária” refere-se a padrões estruturais da cadeia polipeptídica caracterizados pelos contatos locais entre aminoácidos. Esta “localidade” dos contatos é basicamente de dois tipos: (i) contatos seqüencialmente local, ou *S*-local, os quais designam os contatos feitos por monômeros próximos ao longo da cadeia, como no caso das α -hélices; e (ii) contatos topologicamente local, ou *T*-local, que identificam os contatos efetivos, como no caso de folhas β paralelas e antiparalelas. Especificamente, as cadeias na conformação de alfa hélices são estruturas que possuem seus aminoácidos contactantes separados por poucas unidades ao longo da cadeia¹³, possuindo então interações de curto alcance, isto é, cada aminoácido i , por exemplo, interage com o aminoácido $i+4$, havendo apenas três aminoácidos entre eles ao longo da cadeia (Figura 2.4). Estas interações são intermediadas por ligações de hidrogênio (pontes de hidrogênio) entre o oxigênio do carbono do grupo carboxila do i -ésimo resíduo e o hidrogênio ligado ao nitrogênio do grupo amino do resíduo $i+4$; estas ligações contribuem especialmente na estabilização da hélice. Já as folhas beta são associações lado a lado de diversas partes da cadeia polipeptídica extendidas em forma de fitas e

unidas também por ligações de hidrogênio entre os grupos aminos e carboxilas. Neste caso, as interações entre dois aminoácidos não são rigorosamente de curto alcance, no sentido de que a distância ao longo da cadeia que separa os aminoácidos contactantes pode ser grande (doze ou mais¹⁴), porém são efetivamente “locais” pois, uma vez estabelecido o primeiro contato, os outros se sucedem como se fossem próximas ao longo da cadeia. As fitas que possuem as cadeias com a mesma orientação são chamadas folhas “paralelas”; e no caso contrário, “antiparalelas”. As folhas beta também são estabilizadas energeticamente pelas ligações de hidrogênio.

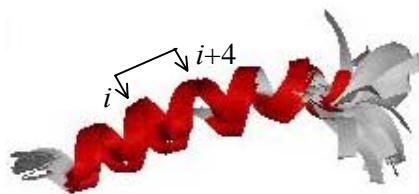


Figura 2.4. Proteína real 1BY0 (*wild-type plastocyanin from silene*) que possui sua cadeia de aminoácidos na conformação de alfa hélice. Pares de aminoácidos contactantes (i e $i+4$) estão separados por poucas unidades ao longo da cadeia.

Nas proteínas em geral, a maior parte dos aminoácidos está conformada em uma destas estruturas secundárias. Portanto, constitui-se em elementos estruturais muito importantes. Em suas mais diversas possibilidades de arranjos (*motifs*), tais estruturas secundárias são unidas por *loops* ou *turns*, os quais são estruturas irregulares e apresentam-se como segmentos da cadeia de vários comprimentos e formas. Neles, os grupos aminos e carboxilas geralmente não se apresentam unidos por ligações de hidrogênio, mas sim fazem interações do mesmo tipo com o solvente (água), pois normalmente se apresentam na superfície da proteína.

A estrutura “terciária” descreve como as estruturas secundárias se arranjam na conformação globular final da proteína. Esta estrutura resulta de interações físicas e químicas que se estabelecem entre as cadeias laterais dos aminoácidos e dessas com o meio aquoso. Essas interações estabilizam termodinamicamente a estrutura de forma apropriada, condição esta necessária para a efetiva atividade biológica das proteínas¹².

Finalmente, a estrutura “quaternária” surge da associação entre estruturas terciárias distintas para formar complexos macromoleculares. As proteínas globulares,

embora sendo de diversos tamanhos e solúveis no meio fisiológico natural, possuem alto grau de regularidade devido as suas estruturas secundárias e combinações delas, isto é, as proteínas podem ser formadas por alfa hélices somente, ou somente por folhas-beta, ou ainda serem da forma mista⁴ (Figura 2.5). Possuem também estabilidade termodinâmica peculiar e muitas delas precisam de um tempo relativamente pequeno para atingir a estrutura enovelada final (nativa), variando aproximadamente entre escalas de 10^{-3} segundos a centenas de segundos¹⁵. A proteína 1L2Y (*Trp-Cage Miniprotein Construct Tc5B*, Figura 2.6), com 20 aminoácidos, é considerada a proteína com menor tempo de enovelamento conhecido atualmente, da ordem de quatro milionésimos de segundos¹⁶, quatro vezes mais rápida que as proteínas anteriormente conhecidas.

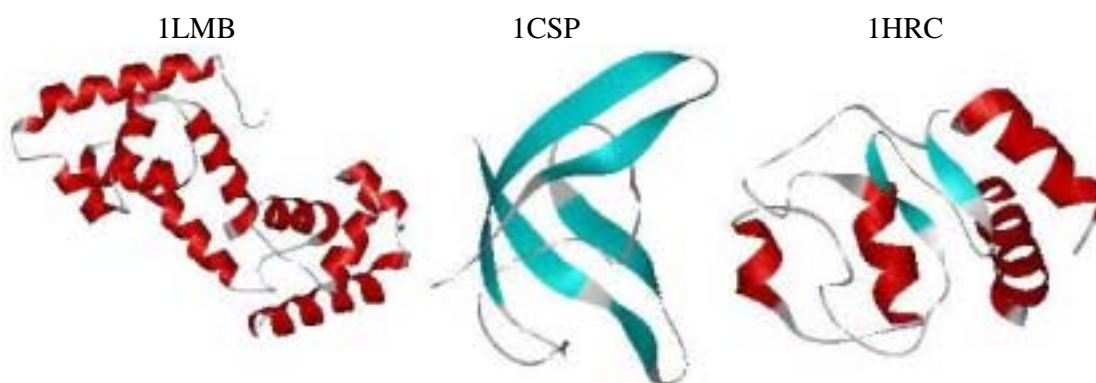


Figura 2.5. Proteínas naturais. Na seqüência da esquerda para a direita: 1LMB (*repressor-operator complex*) formada somente por alfa hélices; 1CSP (*universal nucleic acid-binding domain*) formada somente por folhas-beta; 1HRC (*horse heart cytochrome c*) da forma mista. A união entre alfa hélices e folhas beta é feita por meio de *loops* ou *turns*.

O grande desafio do problema do enovelamento de proteínas globulares advém da unicidade da estrutura nativa. Uma síntese do problema do enovelamento, quanto a sua peculiaridade configuracional, foi primeiramente apontada por Levinthal¹⁷ ao argumentar que, apesar do número astronômico de configurações possíveis para uma cadeia polipeptídica, a mesma sempre se enovela relativamente rápida e corretamente em seu estado nativo, e assim supôs que de alguma forma uma força guia a cadeia para seu estado nativo. O desenvolvimento de algoritmos para resolver este problema tem sido objeto de diferentes áreas científicas, como a biologia, matemática, computação,

física e química. A natureza de cada estudo em particular determina a técnica de simulação computacional mais adequada para representar o sistema proteína-solvente, conforme a pretensão da investigação, sua forma de realização e o tempo real gasto para que a grandeza de interesse seja corretamente caracterizada.

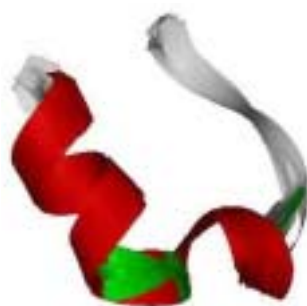


Figura 2.6. Proteína 1L2Y considerada atualmente a mais rápida para enovelar-se. Possui apenas 20 aminoácidos, é da forma mista (alfa hélice e folha beta) e possui tempo de enovelamento da ordem de quatro milionésimos de segundos.

A seguir o método da Dinâmica Molecular é comentado sucintamente. Embora não se faça uso específico desta técnica aqui, o conhecimento de suas bases é indispensável para se entender seus resultados. Trata-se de uma técnica robusta no detalhamento atômico de sistemas moleculares. E no tópico seguinte é introduzido o método Monte Carlo, de extrema importância em muitas áreas científicas e, em particular, no problema do enovelamento de proteínas, pois tem permitido elucidar algumas questões fundamentais do problema, que em seu aspecto geral seria intratável por outros métodos de investigação.

2.2 Dinâmica Molecular

A dinâmica molecular (DM) é a técnica mais popular no estudo de propriedades de equilíbrio e de transporte de sistemas moleculares clássicos¹⁸. Dado que um sistema de muitos corpos, em geral, não tem solução analítica, a alternativa usual é a da solução numérica, e neste ponto é onde entra a DM. Esta técnica procura reproduzir o comportamento microscópico de um experimento utilizando um modelo físico, o qual essencialmente é estabelecido pela descrição da energia potencial, $U(\{\vec{r}_{ij}\})$

interatômica. Neste particular reside ainda grandes dificuldades, pois diferentes propostas para a descrição de $U(\{\vec{r}_{ij}\})$ existem, e nem sempre são totalmente convergentes. O conjunto de forças decorrente da energia potencial $U(\{\vec{r}_{ij}\})$ é tecnicamente denominado por Campo de Forças. Assim, a força \vec{F}_i que atua sobre cada átomo $i = 1, 2, \dots, N$ do sistema, é obtida somando-se a contribuição da ação das forças $\{\vec{f}_{ij}\}$ que cada átomo j exerce sobre o átomo i , isto é:

$$\vec{F}_i = - \sum_{j \neq i}^N \vec{f}_{ij} \quad (2.1)$$

onde $\vec{f}_{ij} = -\vec{\nabla}_i \cdot U(\{\vec{r}_{ij}\})$, com o operador gradiente definido nas coordenadas cartesianas como:

$$\vec{\nabla}_i = \hat{x} \frac{\partial}{\partial x_i} + \hat{y} \frac{\partial}{\partial y_i} + \hat{z} \frac{\partial}{\partial z_i}. \quad (2.2)$$

O vetor \vec{r}_{ij} é definido como $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$ com $\vec{r}_i = x_i \cdot \hat{x} + y_i \cdot \hat{y} + z_i \cdot \hat{z}$; $|\hat{x}| = |\hat{y}| = |\hat{z}| = 1$.

A partir dessas forças, o sistema acoplado de equações do movimento, regido pela 2ª. Lei de Newton, $\vec{F}_i = \frac{d^2}{dt^2} (m_i \cdot \vec{r}_i)$, é tratado numericamente de forma recursiva para cada incremento temporal δt . Assim, as novas posições atômicas são atualizadas a cada novo instante δt , isto é: a cada passo da dinâmica, especificado por δt , que pode ser tão pequeno como 1 picosegundo ou menor, as forças são reavaliadas e a evolução do sistema determinado para aquele intervalo δt . Esse processo iterativo sobre um elevado número de átomos impõe que a função para a energia potencial seja baseada em um compromisso entre precisão e eficiência. Ela deve ser completa o suficiente para uma análise precisa da estrutura e da dinâmica do sistema, mas ao mesmo tempo simples o suficiente para que os cálculos sejam efetuados com rapidez e baixo custo computacional necessários. Desta forma, a evolução temporal do sistema pode ser longa o suficiente para que seja percorrida uma significativa trajetória do espaço de fase do

sistema¹⁸. Evidentemente, o custo computacional no cálculo das forças envolvidas, e o número de ciclos determinado por δt e o tempo total de relaxação do sistema, limitam a aplicação da DM em problemas envolvendo cadeias protéicas maiores.

As primeiras simulações de DM foram realizadas por Alder e Wainwright¹⁹, nos anos 50, para observar as interações entre esferas duras com o propósito de estudar o conhecido paradoxo da reversibilidade: um sistema clássico de muitas partículas é governado microscopicamente pelas equações temporais reversíveis, enquanto que a descrição macroscópica (termodinâmica) do mesmo sistema está baseada em processos irreversíveis. O trabalho deles mostrou que a distribuição de velocidades do sistema de 100 esferas impenetráveis convergia rapidamente ao equilíbrio. Muitas introspecções importantes a respeito do comportamento de líquidos simples emergiram de seus estudos. A primeira aplicação do método de DM ao estudo de materiais foi feita por Gibson e colaboradores²⁰ através da investigação do processo de dano em materiais por radiação, usando um potencial repulsivo de curto alcance e um potencial responsável pela coesão do cristal. Avanços marcantes ocorreram em 1964, quando Rahman²¹ investigou sistemas descritos por potenciais contínuos simulando o argônio líquido, quando foi possível observar que um sistema de 864 partículas, com condições periódicas de contorno, pode reproduzir satisfatoriamente as propriedades termodinâmicas de sistemas reais, e em 1974 com a simulação da água líquida por Rahman e Stillinger²². As primeiras simulações da proteína apareceram em 1977, com a simulação do Inibidor de Tripsina Pancreático de Bovino (BPTI)²³, que é uma das menores proteínas globulares ($L = 58$ aminoácidos), com regiões de alfa hélices e folhas beta e três pontes de sulfeto que ajudam a estabilizar mecanicamente a estrutura terciária da molécula.

A simulação por DM permite o estudo de processos complexos e da dinâmica que ocorrem nos sistemas biológicos que incluem, por exemplo, termodinâmica de ligantes, estabilidade da proteína, enovelamento de macromoléculas pequenas, efeitos da mutação, mudanças conformacionais, reconhecimento molecular (proteínas, DNA, membranas, complexos) e transporte de íons em sistemas biológicos. Adicionalmente, a DM é usada como instrumento complementar em procedimentos experimentais de refinamento estrutural, como em cristalografia de raios X e ressonância magnética

nuclear (NMR), e nos projetos de desenvolvimento de drogas para fins farmacêuticos. Estas duas técnicas experimentais, cristalografia de raios X e NMR, podem ser vistas como fundamentais na biologia estrutural, permitindo a análise da acomodação estrutural das proteínas. A cristalografia trabalha com cristais de proteínas que são estruturas rígidas. Já a ressonância magnética nuclear permite identificar aspectos dinâmicos das moléculas estudadas, pois é realizado em solução aquosa, na qual as proteínas mantêm características semelhantes às de seu estado natural nos organismos.

2.3 Método Monte Carlo

A prática de construir modelos para estudar situações reais complexas remonta à antiguidade, mas o uso moderno da palavra simulação teve origem na aplicação do método Monte Carlo (MC) em meados da segunda guerra mundial²⁴. A denominação do método faz alusão ao cassino de Monte Carlo, no principado de Mônaco, famosa por seus cassinos e jogos de roleta. Evidentemente, o elo comum é o caráter aleatório dos processos físicos envolvidos e os algoritmos que produzem números pseudo-aleatórios para a simulação propriamente dita. O método MC pode ser aplicado tanto em problemas tradicionalmente tratados como determinísticos, quanto naqueles de natureza estocástica inerente; nos dois casos, contudo, se envolve um conteúdo de aleatoriedade através da geração das distintas configurações do sistema²⁴. O método MC é considerado muito simples, geral e flexível podendo ser aplicado em problemas de qualquer nível de complexidade. Uma das limitações do método é o número de ciclos de simulação que pode ser necessário, o que pode, na aplicação em alguns sistemas, tornar o método muito lento.

Em seus primeiros estágios, o método MC era utilizado apenas com o intuito didático. Em 1873, A. Hall desenvolveu um esquema para simular numericamente o valor de π . Em 1899 Lord J. W. S. Rayleigh mostrou que uma caminhada aleatória unidimensional sem barreiras absorventes poderia fornecer uma solução aproximada a uma equação diferencial parabólica. Em 1908, Willian Sealy Gosset usou amostragem experimental para ajuda-lo no desenvolvimento de distribuições de coeficientes de correlação estatística, e, de fato, no mesmo ano definiu a chamada distribuição t de

Student²⁵ (pseudônimo de W. S. Gosset). Em 1931, A. Kolmogorov mostrou a relação entre processos estocásticos de Markov e determinadas equações diferenciais²⁴.

O início da aplicação do método MC na pesquisa científica, é concomitante com o desenvolvimento dos computadores eletrônicos, o que ocorreu em meados da segunda Guerra Mundial. Nesta época, John von Neumann e Stanislaw Ulam²⁴ simularam diretamente o problema da difusão aleatória de nêutrons na fissão nuclear, e por volta de 1948 Enrico Fermi, Nicholas Metropolis e Stanislaw Ulam obtiveram estimativas pelo método MC para a equação de Schrödinger²⁴. Trabalhando com Neumann e Metropolis, Ulam desenvolveu algoritmos para implementação computacional²⁴, bem como para transformar problemas não aleatórios em formas aleatórias que facilitarão sua solução via amostragem estatística. Ulam e Metropolis publicaram seu primeiro trabalho²⁶ com o método MC em 1949. Em 1953, Metropolis, Rosenbluth, Rosenbluth, Teller e Teller apresentam o método de MC envolvendo o critério do balanço detalhado²⁷. O método MC utiliza funções distribuição de probabilidades para descrever um sistema físico ou matemático. Para sua implementação, uma fonte geradora de números pseudo-aleatórios, uniformemente distribuídos no intervalo [0; 1], deve estar disponível.

Para ilustrar o entendimento do procedimento da simulação utilizando o método MC, é conveniente se recorrer a alguns conceitos físicos básicos. Para um sistema constituído por um conjunto de graus de liberdade $\{\mathbf{q}\}$, na mecânica clássica, seus microestados podem ser completamente especificados através de pontos no espaço de fase definidos ao se conhecer as coordenadas generalizadas das posições, $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$, e momentos, $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N$, das N partículas do sistema²⁸. É possível definir um *ensemble* estatístico para um sistema ao se considerar o conjunto de pontos que estão acessíveis no espaço de fase. A densidade destes pontos dentro de uma célula do espaço de fase define a probabilidade de ocorrência de cada microestado. Uma das maneiras eficientes de se explorar o espaço de fase de forma apropriada para que a estimativa das posições, $\{\mathbf{r}_i\}$, seja factível por meio de simulações computacionais é através do algoritmo de Metropolis²⁷. Este algoritmo é uma das principais formas utilizadas pelo método MC para se ter amostragens configuracionais representativas, sem a necessidade de se explorar aquelas regiões do espaço de fase com pesos configuracionais desprezíveis. Cada nova configuração gerada é sempre tida como um novo estado configuracional

possível e é usada para o cálculo das propriedades de interesse do sistema, desde que não viole certas condições físicas, como por exemplo, a do volume excluído²⁹. Nas mudanças dos microestados gerados procura-se assegurar que estes são visitados de acordo com suas respectivas importâncias, produzindo uma distribuição das frequências de visitação dos microestados gerados equivalente ao peso de Boltzmann de cada configuração do sistema, o qual é considerado para tomar as decisões de aceite ou não de cada novo estado. Quando cada um dos microestados gerados depende apenas do microestado anterior, o processo de evolução do sistema é chamado Markoviano²⁴. Então, se $P(i,t)$ é a probabilidade por unidade de tempo de se obter um microestado i num instante t , e $w(i \rightarrow j)$ a probabilidade do estado microscópico i transicionar para j , então se a simulação se referir a um processo do tipo Markoviano, a evolução de $P(i,t)$ é descrita por:

$$P(i,t + \Delta t) = P(i,t) + \left[\sum_{\{j\}} P(j,t).w(j \rightarrow i) - \Omega_i(t).P(i,t) \right] \Delta t, \quad (2.3)$$

onde o somatório dentro dos colchetes da expressão acima descreve o fluxo de probabilidade no tempo t para o microestado i a partir de todos os outros, e o segundo termo dentro dos colchetes descreve a probabilidade total de transição do estado i para qualquer um dos outros estados $\{j\}$, $\Omega_i(t) = \sum_{\{j\}} w(i \rightarrow j)(t)$. No estado estacionário, $t \rightarrow \infty$, espera-se que $P(i,t) \rightarrow P_i$, ou seja, que a probabilidade de ocorrência de um particular estado i , no equilíbrio, não dependa mais do tempo:

$$\sum_{\{j\}} P(j,t).w(j \rightarrow i) - \sum_{\{j\}} P(i,t).w(i \rightarrow j) = 0 \quad (2.4)$$

Uma das maneiras pela qual a equação acima pode ser satisfeita para que sejam construídos algoritmos de forma adequada e eficiente, e que obedeçam à reversibilidade microscópica, ou o balanço detalhado, é considerar a condição de simetria (comum nas equações de movimentos clássicos e quanto-mecânicos²⁴):

$$P(j,t).w(j \rightarrow i) = P(i,t).w(i \rightarrow j) \quad (2.5)$$

Uma forma válida e simplificada de reescrever a equação acima é considerar a situação de equilíbrio³⁰, onde se define $P_{eq}(i, t) \equiv P_i$,

$$P_j \cdot w(j \rightarrow i) = P_i \cdot w(i \rightarrow j), \quad (2.6)$$

o que implica em:

$$w(i \rightarrow j) = w(j \rightarrow i) \cdot \frac{P_j}{P_i}. \quad (2.7)$$

Também, no equilíbrio de um *ensemble* canônico tem-se que $P_i = \exp(-E/k.T)/z$, onde z é a função partição do sistema. Por meio desta igualdade, torna-se explícito que a mudança de um estado para outro não depende do conhecimento prévio da função partição z do sistema, isto é:

$$\frac{w(i \rightarrow j)}{w(j \rightarrow i)} = e^{-\Delta E/kT}. \quad (2.8)$$

No algoritmo de Metropolis, sempre que a mudança configuracional reduz a energia do sistema, a nova configuração é aceita com probabilidade igual a 1. Baseado no equilíbrio detalhado, o algoritmo de Metropolis²⁷ é descrito pelos seguintes passos:

- (i) gera-se uma configuração inicial qualquer do sistema;
- (ii) verifica-se se a configuração gerada viola alguma condição física do sistema; no caso afirmativo, a configuração é rejeitada e retorna-se ao passo anterior, senão, vai para o passo seguinte;
- (iii) uma nova configuração é gerada e é aceita ou não, de acordo com as possibilidades³⁰:

$$w(i \rightarrow j) = \begin{cases} e^{-\Delta E/kT}, & \text{se } \Delta E > 0 \\ 1 & \text{se } \Delta E \leq 0. \end{cases} \quad (2.9)$$

O requerimento para a realização deste algoritmo é que w seja sempre positivo (ou nulo) a fim de evitar problemas de ergodicidade²⁴ que podem interferir nos valores

médios dos observáveis a serem estimados. O algoritmo de Metropolis é uma dentre um grande número de possibilidades envolvendo diferentes hierarquias configuracionais. Porém, neste texto se restringirá a este algoritmo e o leitor interessado poderá consultar farto material bibliográfico disponível^{31,32}.

2.4 Método MC e o modelo computacional utilizado neste trabalho

Neste trabalho foi utilizado o método MC para simulação do processo de enovelamento de proteínas, cujo código Fortran foi desenvolvido originalmente por dos Reis³³. Os detalhes energéticos e topológicos gerais envolvidos no modelo utilizado serão detalhados no próximo Capítulo. A seguir, são descritos somente os aspectos do problema que são importantes na explicitação do método MC utilizado.

A estrutura nativa foi utilizada apenas para o *design* da seqüência e para checar quando a simulação a alcança pela primeira vez, procedendo a simulação sem informações adicionais do estado nativo. Para cada cadeia utilizada, foram efetuadas muitas simulações independentes, gerando na ordem de 10^9 configurações em cada simulação. Para cada simulação completa, os N resíduos de uma cadeia linear de uma proteína são reduzidos a N “contas” todas do mesmo tamanho, as quais ocupam sítios consecutivos de uma rede cúbica. A energia do sistema em cada configuração em particular depende essencialmente da interação de cada um dos monômeros com o meio solvente³⁴. Toda simulação é iniciada a partir de uma configuração aleatória, sem contatos topológicos, correspondente à condição de cadeia totalmente desnaturada. Como a ênfase deste trabalho está no processo de enovelamento inverso, inicialmente foram determinados todos os contatos topológicos de todas as possíveis configurações CSA, que funcionam na simulação como configurações alvo. Assim, o termo “configuração alvo” é sinônimo de configuração nativa. Cada configuração compacta em que a cadeia pode se acomodar é representada por um cubo de dimensões $N \times N \times N$. O número CT de Contatos Topológicos para qualquer N é dado por[†]:

$$CT = 2N^3 - 3N^2 + 1. \quad (2.10)$$

[†] A demonstração da obtenção desta expressão está em: ROCHA, L. F. de O. Mecanismo de “*folding*” de macromoléculas: estudo sobre o papel das especificidades topológicas e químicas das interações intramoleculares em cadeias poliméricas lineares. Tese (Doutorado em Ciências). IBILCE, UNESP, São José do Rio Preto, 2005.

Neste trabalho, $N = 3$ (cadeia de 27 monômeros), o que resulta em $CT = 28$. A dinâmica configuracional é obtida pela combinação de movimentos de fim da cadeia (*chain end move*), canto (*kink jump*) e manivela (*crankshaft move*), como ilustra a Figura 2.7, para a simulação do enovelamento. Este conjunto de movimentos permite que a cadeia visite todas as configurações possíveis, o que garante a condição de volume excluído e a ergodicidade do sistema³⁵. Após cada movimento, os contatos topológicos e a energia do sistema são reavaliados e a nova configuração é comparada com a configuração nativa por meio de seus 28 contatos topológicos, com o único intento de se verificar se a configuração nativa foi atingida. A probabilidade de a cadeia adquirir uma nova configuração é máxima (igual a 1) sempre que a energia do sistema diminui. Se a energia do sistema aumenta, esta probabilidade é estimada pelo peso de Boltzmann ($\exp(-\Delta E/kT)$), seguindo o algoritmo de Metropolis²⁷, como discutido previamente. Também é assumida a hipótese de que a estrutura nativa corresponde aos mínimos globais da energia do sistema, o que para moléculas pequenas é plausível.

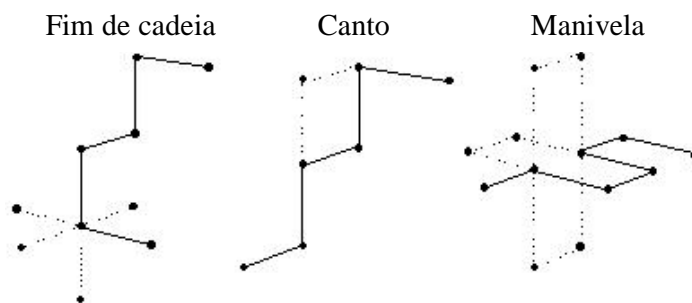


Figura 2.7. Movimentos utilizados para a simulação do enovelamento da proteína: as linhas pontilhadas indicam as possíveis posições para mudanças configuracionais.

Neste ponto é importante ressaltar que no tipo de problema aqui abordado, não são empregados especificamente certos requisitos e propriedades do método MC nas condições de equilíbrio, mas sim a capacidade do método em conduzir o sistema ao equilíbrio, a partir de qualquer outro estado. Isto porque, de fato, neste trabalho se está interessado na evolução temporal do sistema, desde a condição desnaturada da cadeia (estado excitado) até a configuração nativa (estado fundamental).

Todos os resultados, envolvendo as restrições estéricas e a caracterização topológica das estruturas estudadas, foram obtidos por programas computacionais especificamente desenvolvidos neste trabalho.