

Lessandra Eller

**O Uso do Método de Crescimento de
Cadeias no Estudo do Enovelamento de
Modelos de Proteínas**





1910001216





UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO" - UNESP.

LESSANDRA ELLER

O USO DO MÉTODO DE CRESCIMENTO DE CADEIAS NO ESTUDO DO ENOVELAMENTO DE MODELOS DE PROTEÍNAS

Dissertação apresentada para obtenção do grau de Mestre em Física, Área de Concentração em Biofísica Molecular. Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP.

ORIENTADOR: PROF. DR. JORGE CHAHINE

São José do Rio Preto - SP

2003



t. 1.216/03

Agradecimentos

Ao Prof. Dr. Jorge Chahine pela disponibilidade de orientação e pelas valiosas correções deste trabalho.

Aos meus pais: Wilmar José Eller e Neide Uzeloto Eller, "in memoriam".

Aos colegas de sala: Luiz F. O. Rocha e Sidney Jurado de Carvalho. Ao Luiz, pelo auxílio em programação e em enovelamento de proteínas e ao Sidney pela permissão de utilização da sua máquina Lys para a realização de grande parte das simulações deste trabalho e pelas dicas no sistema Linux.

Ao Prof. Dr. Elso Drigo Filho pelo valiosíssimo Estágio de Docência da CAPES, contribuindo enormemente para a minha formação profissional.

Ao Prof. Dr. Augusto Agostinho Neto pelo exemplo profissional pelos conhecimentos ensinados.

Ao Prof. Dr. Vitor Barbanti Pereira Leite pela disponibilidade de máquinas.

A secretária Ilva e aos técnicos Barbosa e Paulinho pelo auxílio prestado sempre que foi solicitado.

A todos que direta ou indiretamente contribuíram para a consolidação deste trabalho.

A CAPES pelo apoio financeiro.



RESUMO

Neste trabalho, foi utilizado o método de crescimento de cadeias com o objetivo de analisar o empacotamento e enovelamento das cadeias crescidas.

O método de crescimento de cadeias usado aqui, consiste em gerar cadeias, que são formadas por um código de 3 letras, numa rede cúbica tridimensional de 27 monômeros ligados covalentemente. Os potenciais utilizados constam de uma parte sempre atrativa relacionada aos contatos de tipos iguais (*cti*) e uma parte (α) que pode ser atrativa, neutra e repulsiva relacionada aos contatos de tipos diferentes (*cta*).

O método de crescimento de cadeias mostrou ser eficiente somente para potenciais repulsivos nas interações não nativas. A heterogeneidade entre as interações tem um valor mínimo ($\alpha > 0$) para que o método cresça cadeias nativas.

Nossos resultados mostram que no regime onde o princípio da segregação é notadamente não satisfeito (potenciais neutro e atrativo): a cadeia jamais cresce em direção à conformação nativa.

Pequenas mutações que geram alguns contatos frustrados (energia alta) impedem, por completo, a obtenção do estado nativo da sequência mutada, sugerindo que o método é eficiente para sequências otimamente desenhadas.

O método é adequado para enovelamento que não são precedidos por colapso hidrofóbico. Potenciais altamente atrativos podem induzir a cadeia ao colapso seguido de um enovelamento lento. O método de crescimento de cadeias não reproduziu, nesse trabalho, esse comportamento.

CONTEÚDO

CAPÍTULO 1

1.1. Introdução	1
1.2. O problema do <i>enovelamento</i> de proteínas.....	2
1.3. Síntese de Proteínas.....	2
1.4. Descrição do modelo utilizado.....	3
1.5. A importância dos modelos em rede.....	8

CAPÍTULO 2 – RESULTADOS

2.1. Resultados I - Simulação utilizando um único potencial.....	10
2.1.1. Introdução.....	11
2.1.2. Tempo de sucesso de <i>enovelamento</i> para as 100 conformações nativas.....	12
2.1.3. Número de armadilhas topológicas para as 100 conformações nativas..	14
2.2. Resultados II - Influência dos potenciais no processo de crescimento de cadeias.....	17
2.2.1. Introdução.....	17
2.2.2. Número de Armadilhas e % CMC em função dos potenciais.....	19
2.2.3. % CMC x contatos nativos.....	22
2.2.4. % não CMC x contatos totais.....	24
2.2.5. Raio de giração e Distância ponta a ponta (CMC) x contatos nativos.....	27
2.2.6. Raio de giração e Distância ponta a ponta (não CMC) x contatos totais	29
2.3. Resultados III – Mutações.....	31

CAPÍTULO 3 – DISCUSSÕES E CONCLUSÕES

3.1. Discussões..... 33

3.2. Conclusões..... 38

APÊNDICE I..... 42

LISTA DE SÍMBOLOS

RNA: ácido ribonucléico

cti: contatos de tipos iguais

ctd: contatos de tipos diferentes

ct: contatos totais

t_e : tempo de enovelamento

N: número de monômeros

PMC: Passos Monte Carlo

CMC: conformação maximamente compacta

não CMC: conformação não maximamente compacta

E_{vm} : erro do valor médio

T: temperatura

E_Λ : energia da configuração Λ

$P_\Lambda(\beta)$, $P[\text{sítio}]$, $P[p-1][v]$: probabilidade de transição

u: energia de interação do último monômero inserido com os demais

$E_\Lambda(p-1)$: energia da configuração até o passo $p-1$

E_{ctd} , α : energia relacionada aos contatos de tipos diferentes

E_{cti} : energia relacionada aos contatos de tipos iguais

R_g : raio de giração

dpp: distância ponta a ponta

CAPÍTULO 1

1.1. Introdução

A maioria da matéria orgânica nas células vivas consiste de quatro tipos de compostos orgânicos: carboidratos (cujo papel é fundamentalmente energético), lipídios (que funcionam como reserva energética e formam os hormônios esteróides), ácidos nucleicos (que controlam a atividade da célula, sua divisão e herança) e proteínas (com papéis estrutural e enzimático); além destes, também existem os compostos inorgânicos água e sais minerais (Lehninger, 1985). Em relação às proteínas existem três tipos: fibrosas, membranas e as globulares. As proteínas fibrosas servem principalmente para papéis estruturais; as proteínas de membranas mediam a troca de moléculas e informação através das fronteiras celulares; as proteínas globulares representam a maior classe das proteínas e aglomeram para conformações compactas (Chan & Dill, 1993). Estas últimas são as de interesse no atual projeto, que serão estudadas através de modelos de rede.

Proteínas são cadeias longas de aminoácidos unidas por ligações peptídicas (amida), com um grupo amina contendo nitrogênio carregado positivamente em uma extremidade e um grupo carboxila carregado negativamente na outra extremidade (Figura 1.1.a). A cadeia é composta por vários grupos laterais (R) diferentes que correspondem a cada um dos 20 aminoácidos (Figura 1.1.b). A sequência de aminoácidos é chamada de estrutura primária de uma proteína, sendo ela quem determina como uma proteína se enovelará na sua conformação tridimensional para desempenhar sua função bioquímica única (Campbell, 2000). No estudo das proteínas, o problema do *enovelamento* de proteínas é um dos problemas que desperta bastante interesse e é o que passaremos a discutir abaixo.

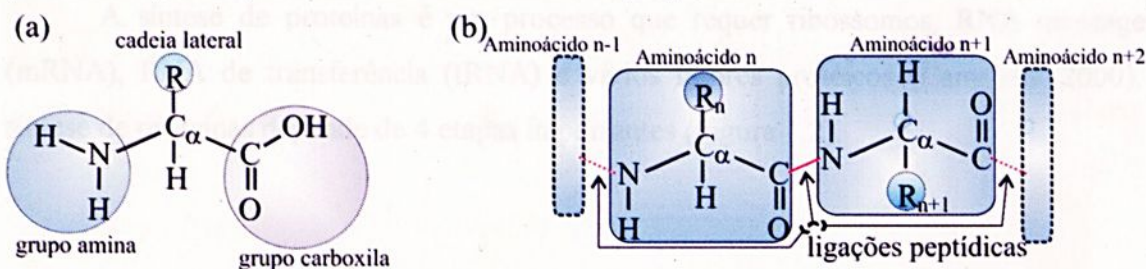


Figura 1.1. (a) Representação do aminoácido isolado, contendo um carbono alfa (C_α) ligado a um átomo de hidrogênio (H), ao grupo lateral (R), ao grupo amina (NH_2) e ao grupo carboxila ($COOH$). (b) Cadeia polipeptídica, com a ligação peptídica entre os aminoácidos.



1.2. O problema do *enovelamento* de proteínas

O *enovelamento* de proteínas é um dos problemas centrais em física biológica. Envolve aspectos de biologia, química, bioquímica, ciência computacional, física de polímeros (Pande et al., 2000). O problema do *enovelamento* de proteínas é prever a estrutura tridimensional compacta a partir da sequência de aminoácidos (Chan & Dill, 1993), ou seja, da estrutura primária. A compreensão do problema do *enovelamento* de proteínas é de grande importância no desenho de novas proteínas com funções especiais; em decodificar a informação genética obtida pelo Projeto Genoma Humano; em desenhar novas drogas; na compreensão de várias patologias (Yon, 2001) e em tentar entender as estruturas e funções de milhares de sequências de proteínas que estão sendo descobertas em laboratórios tecnológicos.

Os objetivos deste trabalho são:

- _ Estudar o problema do *enovelamento* de proteínas através do método de crescimento de cadeias numa rede cúbica tridimensional.
- _ A utilização de potenciais com um termo sempre atrativo e outro termo dependente do tipo de monômeros em contato, podendo ser este último termo atrativo, neutro ou repulsivo. Tais potenciais podem revelar os mecanismos envolvidos no *enovelamento* de proteínas.

Para compreendermos melhor o problema do *enovelamento*, precisamos entender (mas não necessariamente) como ocorre a síntese de proteínas.

1.3. Síntese de Proteínas

A síntese de proteínas é um processo que requer ribossomos, RNA mensageiro (mRNA), RNA de transferência (tRNA) e vários fatores protéicos (Campbell, 2000). A síntese de proteínas depende de 4 etapas importantes (Figura 1.2).



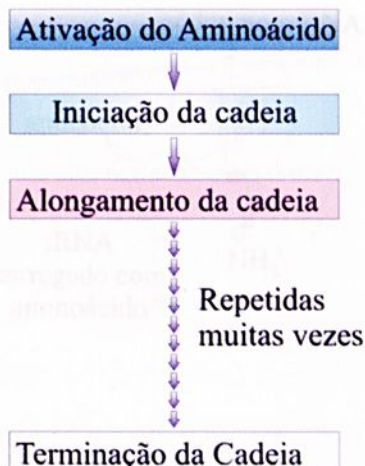


Figura 1.2. Fluxograma mostrando as quatro etapas da síntese de proteínas.

Abaixo descrevemos sucintamente cada uma das etapas da síntese de proteínas.

1ª.) Ativação dos aminoácidos: Antes que um aminoácido possa ser incorporado na cadeia, ele deve ser ativado. O processo de ativação dos aminoácidos ocorre quando cada um dos 20 aminoácidos está ligado covalentemente a um RNA de transferência (tRNA) formando um aminoacil-tRNA, que é uma enzima.

2ª.) Iniciação da cadeia: O RNA de transferência (tRNA) é ligado ao RNA mensageiro (mRNA) que forma um par de bases com um códon no RNA mensageiro (mRNA); indicando o início da cadeia.

3ª.) Alongamento da cadeia: O ribossomo move-se ao longo do RNA mensageiro (mRNA) e a proteína cresce à medida que novos aminoácidos são incorporados à cadeia. O processo de alongamento da cadeia repete-se até que a cadeia esteja completa.

4ª.) Terminação da cadeia: O término da cadeia, que é sinalizado por um códon de término, no RNA mensageiro, é seguido por sua liberação do ribossomo; promovida por fatores de liberação.

Podemos mostrar mais explicitamente essas etapas através da Figura 1.3.

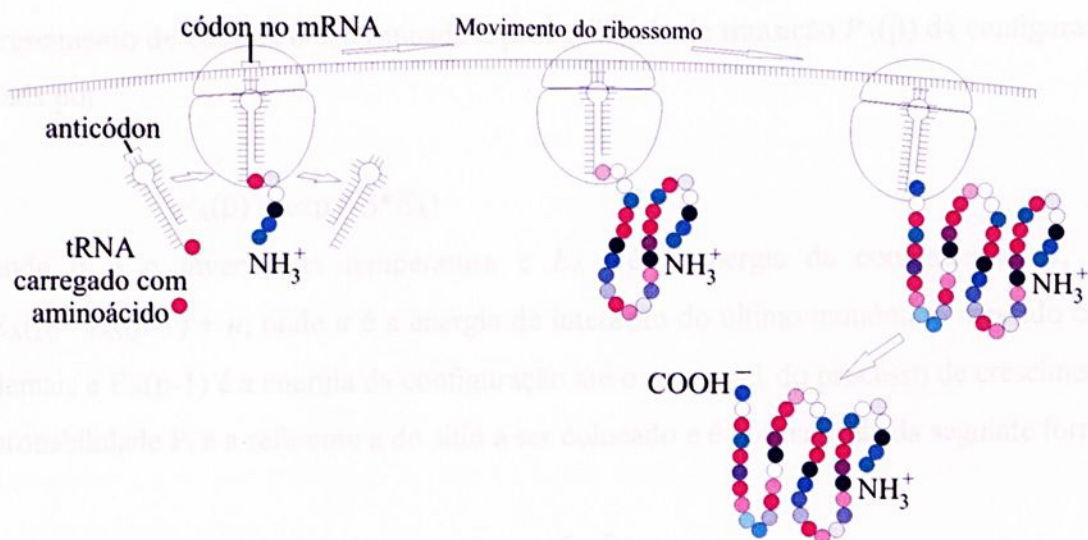


Figura 1.3. Esquema ilustrativo das etapas da síntese de proteínas.

Explicado como ocorre o crescimento da cadeia polipeptídica através da síntese de proteínas, passamos a descrever o crescimento de cadeias utilizando simulação computacional, conforme descrito mais detalhadamente abaixo.

O algoritmo de crescimento de cadeias utilizado neste trabalho gera estruturas de proteínas em rede adicionando ligações entre monômeros até que o comprimento final da cadeia seja atingido.

1.4. Descrição do modelo utilizado

O modelo utilizado consiste em gerar cadeias numa rede cúbica tridimensional de 27 monômeros ligados covalentemente (por ligação de comprimento fixo 1(um)), cuja sequência de resíduos de aminoácidos (Socci et al, 1997) é formada por um código de três letras (A, B ou C), sendo dada por:

ABA BBB CBA CBA BAB ACA CBA CAA CAB

A cadeia é crescida fixando o primeiro e o segundo monômeros, constituindo assim sua primeira ligação (Rosenbluth & Rosenbluth, 1955). Cada cadeia continua a ser crescida com o terceiro monômero numa direção escolhida e a cada passo do processo de

crescimento de cadeias é determinada a probabilidade de transição $P_{\Lambda}(\beta)$ da configuração Λ dada por:

$$P_{\Lambda}(\beta) = \exp(-\beta * E_{\Lambda}) \quad (1.1)$$

onde β é o inverso da temperatura e E_{Λ} é a energia da configuração Λ , sendo $E_{\Lambda}(p) = E_{\Lambda}(p-1) + u$; onde u é a energia de interação do último monômero crescido com os demais e $E_{\Lambda}(p-1)$ é a energia da configuração até o passo $p-1$ do processo de crescimento. A probabilidade P_i é a referente a do sítio a ser colocado e é normalizada da seguinte forma:

$$P_i = \frac{e^{-\beta u(\vec{R} + \vec{l}_i)}}{\sum_{j=1}^{S_r} e^{-\beta u(\vec{R} + \vec{l}_j)}} \quad (1.2)$$

onde $\pm \vec{l}_j$, $j = 1, 2, 3$ são as três direções de crescimento paralelas aos eixos x , y e z e S_r é o número de possibilidades de crescimento correspondente ao número de sítios vazios localizados nas vizinhanças do vetor $\vec{R}$, que descreve a posição da cadeia até o passo anterior.

A probabilidade de transição (Gan et al, 2000) que usamos aqui favorece direções de crescimento com contatos de baixa energia, a qual gera configurações de cadeias abertas e compactas, dependendo da temperatura.

O potencial é uma das variáveis determinantes para entender como o sistema se comporta. Um exemplo disso é que dependendo do potencial utilizado, obtemos uma maior ou menor quantidade de estados nativos, ou ainda, não obtemos estados nativos. O objetivo é encontrar um potencial que favoreça estados compactos, motivando a cadeia enovelar. O potencial que utilizamos é composto por uma parte atrativa fixa (-3) e uma parte (α) que varia de -3 a +3, sendo assim a fórmula matemática para o potencial é dada por:

$$u = -3 * cti + \alpha * ctd \quad (1.3)$$

onde:

cti : número de contatos de tipos iguais;

ctd : número de contatos de tipos diferentes.

A estrutura nativa dessa sequência é uma configuração na forma de um cubo único e já foi estudado anteriormente por outros métodos de simulação (Socci et al, 1997). Por essa razão, o estado nativo (um dos 103346 estruturas cúbicas possíveis) é conhecido; daí ser possível escrever o potencial como uma soma entre contatos de tipos iguais (*cti*) e contatos de tipos diferentes (*ctd*).

Os contatos de tipos iguais ocorrem entre monômeros com o mesmo código de letras, por exemplo, A com A. Já os contatos de tipos diferentes ocorrem entre monômeros com códigos de letras diferentes, por exemplo, A com B. Para um monômero ser vizinho (ou seja, estar em contato) de outro monômero, a distância entre eles deve ser um (parâmetro de rede).

A probabilidade de transição, da Equação (1.1), juntamente com o potencial, da Equação (1.3) e o código de letras da sequência de monômeros serão os ingredientes necessários para o processo de crescimento da cadeia (Figura 1.4).

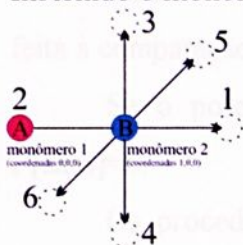


Figura 1.4. Diagrama de uma cadeia de polímero em crescimento. Um cubo representa a rede de contatos.

O estado de uma probabilidade de transição ($P_i|j$), quando aplicado a um a uma monômero, conforme mostrado na Figura 1.4, depende do potencial utilizado (ϕ), da Equação (1.1), e do código de letras (A, B, C) para os monômeros em contato. Por exemplo, para o monômero 3 na Figura 1.4, os sites 1, 2, 3, 5 não possuem nenhum vizinho (todas com probabilidade normalizada $P(1) = P(2) = P(3) = P(5) = 0$), e o site 4 do monômero 3 tem um vizinho, no caso o monômero 2, sendo $P(4|3) = 1$.

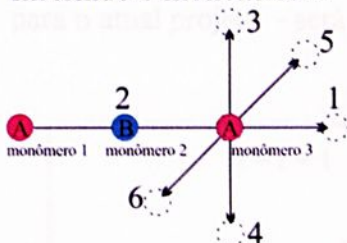
Se o potencial utilizado for $\phi = 1$, as probabilidades podem ser calculadas. Para o monômero 3 e 2, ou seja, o mesmo código de letras, então $\phi = 2$ e $P(4|3) = 1/2$.

Inserindo o monômero 3



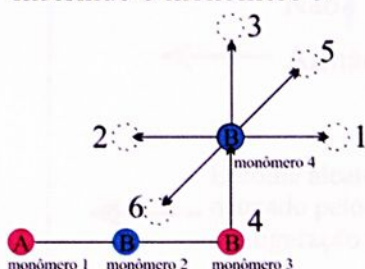
Para inserir a partícula 3 na cadeia, temos 5 sítios desocupados (1, 3, 4, 5, 6) e consequentemente sem vizinhos, assim, a escolha do sítio a ser ocupado é aleatória porque temos iguais probabilidades normalizadas, $P[\text{sítio}]$. O sítio 2 não pode ser ocupado pois viola a condição de volume excluído. Supondo que foi escolhido aleatoriamente o sítio 1 para partícula 3. Para normalizarmos a probabilidade $P_n[\text{sítio}]$ não normalizada fazemos: $P[\text{sítio}] = P_n[\text{sítio}] / (P_n[1] + P_n[3] + P_n[4] + P_n[5] + P_n[6])$.

Inserindo o monômero 4



Para inserir a partícula 4 na cadeia, temos 5 sítios desocupados (1, 3, 4, 5, 6) e consequentemente sem vizinhos, assim, a escolha do sítio a ser ocupado é novamente aleatória porque temos iguais probabilidades normalizadas, $P[\text{sítio}]$. Supondo que foi escolhido aleatoriamente o sítio 3 para partícula 4.

Inserindo o monômero 5



Para inserir a partícula 5 na cadeia, temos 5 sítios desocupados (1, 2, 3, 5, 6); neste caso como existem sítios com e sem vizinhos, dependendo do potencial e do código de letras e escolhida a maior probabilidade normalizada dentre os cinco sítios, $P[\text{sítio}]$ e também é gerado um número aleatório, num_aleat . Se $\text{num_aleat} < P[\text{sítio}]$ e escolhido "sítio" para o quinto monômero (Metropolis et al, 1953), senão a escolha é feita aleatoriamente, como foi feito para o terceiro e quarto monômeros. Procedimentos análogos são adotados para os demais monômeros a serem inseridos na cadeia proteica, conforme mostrado no fluxograma do programa utilizado, (Figura 1.5).

Figura 1.4. Ilustração de um crescimento de cadeia numa rede cúbica tridimensional.

O cálculo da maior probabilidade de transição ($P[\text{sítio}]$), quando existem sítios com e sem vizinhos, conforme mostrado na Figura 1.4., depende do potencial utilizado ($(-3, \alpha)$ da Equação (1.3)) e do código de letras (A, B, C) para os monômeros em contato. Por exemplo: para o monômero 5 na Figura 1.4, os sítios 1, 3, 5, 6 não possuem nenhum vizinho (todos com probabilidade normalizada $P[1] = P[3] = P[5] = P[6] = a$); já o sítio 2 do monômero 5 tem um vizinho, no caso o monômero 2, sendo $P[\text{sítio } 2] = b$.

Se o potencial utilizado for $(-3, \alpha)$ positivo), podemos ter:

Os monômeros 5 e 2 com o mesmo código de letras, então $a < b$ e $P[\text{sítio}] = b$, ou

os monômeros 5 e 2 com códigos de letras diferentes, então $a > b$ e $P[\text{sítio}] = a$. A seguir é feita a comparação entre o número aleatório e $P[\text{sítio}]$, conforme mostrado na Figura 1.4.

Se o potencial utilizado for $(-3, \alpha \text{ negativo})$, então sempre $a < b$ e portanto $P[\text{sítio}] = b$.

Os procedimentos descritos acima, são as bases do programa de crescimento de cadeia e para uma descrição mais pormenorizada deste programa – totalmente desenvolvido para o atual projeto – será mostrado seu fluxograma (Figura 1.5).

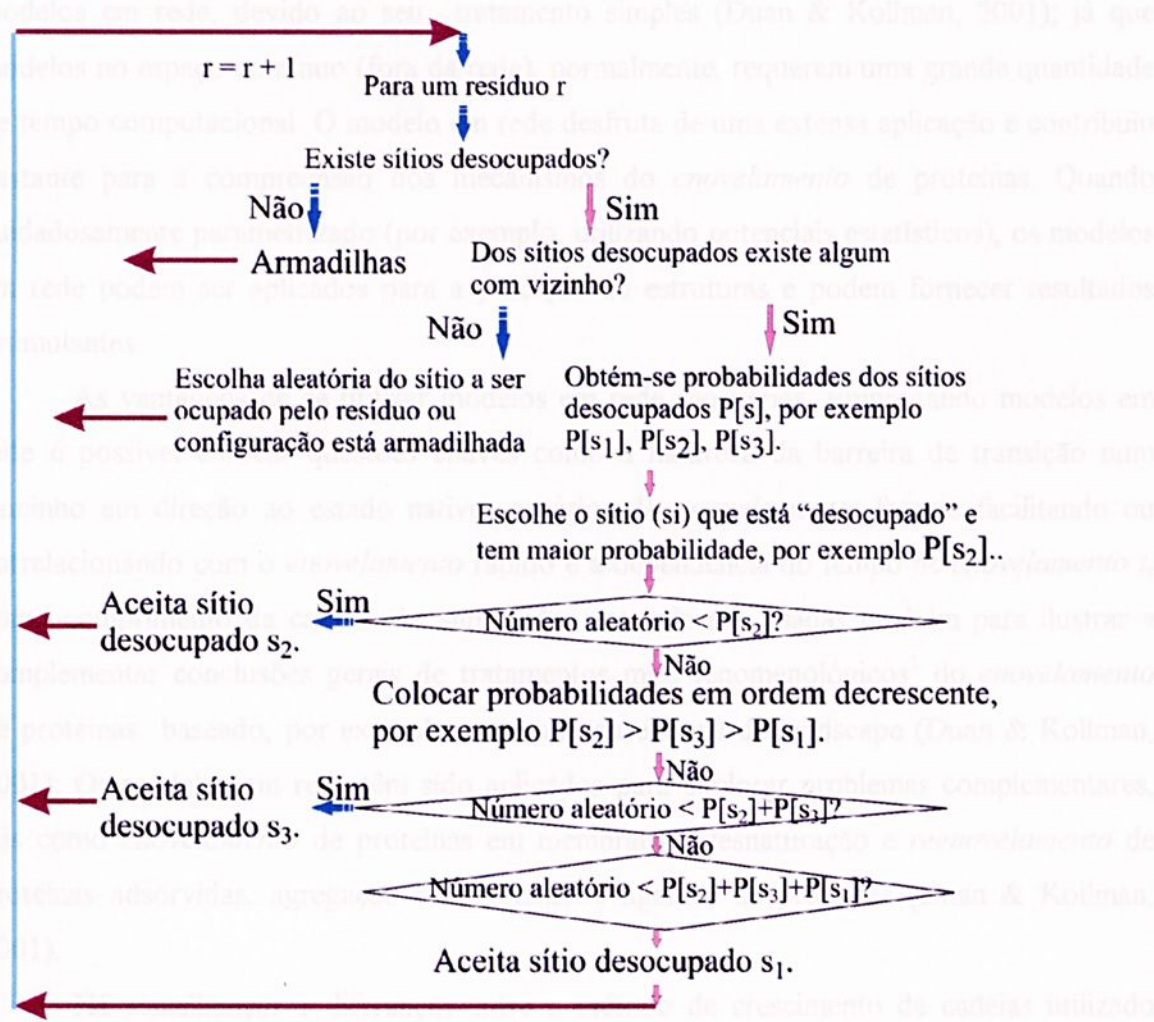


Figura 1.5. Fluxograma para probabilidade de escolha de um sítio a ser ocupado por um residuo r, no programa de crescimento de cadeia.



1.5. A importância dos modelos em rede

O método de crescimento de cadeia empregado neste trabalho ocorre numa rede cúbica tridimensional. Progressos no campo de *enovelamento* de proteínas tem sido feito pela combinação de experimentos e simulações. Um importante componente para o estudo do *enovelamento* de proteínas é a extensiva simulação Monte Carlo (MC) baseada em modelos em rede, devido ao seu tratamento simples (Duan & Kollman, 2001); já que modelos no espaço contínuo (fora da rede), normalmente, requerem uma grande quantidade de tempo computacional. O modelo em rede desfruta de uma extensa aplicação e contribuiu bastante para a compreensão dos mecanismos do *enovelamento* de proteínas. Quando cuidadosamente parametrizado (por exemplo, utilizando potenciais estatísticos), os modelos em rede podem ser aplicados para a predição de estruturas e podem fornecer resultados estimulantes.

As vantagens de se utilizar modelos em rede são várias. Empregando modelos em rede é possível enfocar questões chaves como a natureza da barreira de transição num caminho em direção ao estado nativo, cenários do *enovelamento*; fatores facilitando ou correlacionando com o *enovelamento* rápido e a dependência do tempo de *enovelamento* t_e como comprimento da cadeia. As simulações em rede são usadas também para ilustrar e complementar conclusões gerais de tratamentos mais fenomenológicos¹ do *enovelamento* de proteínas baseado, por exemplo, na teoria da energia de landscape (Duan & Kollman, 2001). Os modelos em rede têm sido aplicados para explorar problemas complementares, tais como *enovelamento* de proteínas em membranas, desnaturação e *reenovelamento* de proteínas adsorvidas, agregação de proteínas e ligantes de proteínas (Duan & Kollman, 2001).

Há semelhanças e diferenças entre o método de crescimento de cadeias utilizado neste trabalho e as simulações tradicionais de Monte Carlo. A seguir, na Tabela 1.1, destacamos algumas analogias entre estes dois métodos de simulação na rede.

¹ destinado ao entendimento das bases físicas governando o processo de *enovelamento* de proteínas.



Grandezas e procedimentos utilizados	Simulação tradicional MC	Crescimento de cadeia
Tempo (1 Passo Monte Carlo PMC)	1 PMC = 1 movimento completo na cadeia	1 PMC = 1 cadeia crescida com 27 monômeros
Armadilhas	Energéticas	Topológicas
Movimentos	3 Tipos: canto, ponta e manivela	Inserção de um monômero após o outro
Medidas de grandezas físicas	Após movimento completo na cadeia	Após cadeia crescida

Tabela 1.1. Analogias entre simulações tradicionais Monte Carlo na rede e o método de crescimento de cadeias.

Figura 2.0. Condição inicial e sua evolução de acordo com os resultados de simulação utilizados neste trabalho.

O processo de crescimento de cadeias foi realizado em três etapas:

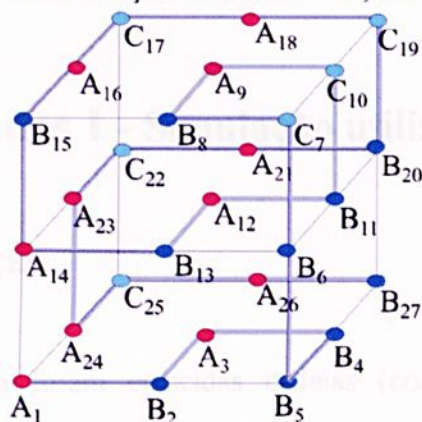
a) Resultados I – Foi utilizada três diferentes temperaturas com um único potencial, formado por uma parte atrativa (-3) e uma parte repulsiva ($\alpha = +3$), até obter 100 cadeias novas.

b) Resultados II – Para verificar a influência de diferentes potenciais no processo de crescimento de cadeias, fixamos a parte atrativa (-3) e variamos a outra parte do potencial (α de -3 a $+3$), utilizamos apenas a temperatura $T = 1,9$ durante um intervalo de tempo de 2×10^6 Passos Monte Carlo (PMC).

CAPÍTULO 2 – RESULTADOS

Crescemos várias cadeias de 27 monômeros até ser obtida a conformação nativa, formada por 28 contatos entre monômeros de tipos iguais, cuja sequência de resíduos de aminoácidos (que contém um código de três letras: A, B e C) são mostradas na Figura 2.0.

Conformação nativa: $cti = 28$; $ctd = 0$



ABA BBB CBA CBA BAB ACA CBA CAA CAB

Figura 2.0. Conformação nativa e sua sequência de resíduos de aminoácidos utilizadas neste trabalho.

O processo de crescimento de cadeias foi realizado em três etapas:

a) Resultados I – Foi utilizada três diferentes temperaturas com um único potencial, formado por uma parte atrativa (-3) e uma parte repulsiva ($\alpha = +3$), até obter 100 cadeias nativas.

b) Resultados II – Para verificar a influência de diferentes potenciais no processo de crescimento de cadeias, fixamos a parte atrativa (-3) e variamos a outra parte do potencial (α de -3 a $+3$); utilizamos apenas a temperatura $T = 1,9$ durante um intervalo de tempo de 2×10^9 Passos Monte Carlo (PMC).

c) **Resultados III** – Para analisar a importância de se ter seqüências otimamente desenhadas no processo de crescimento de cadeias são feitas duas mutações (trocas de letras, por exemplo A por C) na seqüência de resíduos, para três temperaturas durante 2×10^9 Passos Monte Carlo (PMC).

Em cada etapa, informamos os detalhes utilizados na simulação e as respectivas grandezas empregadas para a análise do comportamento do crescimento das cadeias.

2.1. Resultados I - Simulação utilizando um único potencial

2.1.1. Introdução

Nesta etapa foram crescidas cadeias (com 27 monômeros) até ser obtida a conformação nativa durante um intervalo de tempo de 2×10^9 Passos Monte Carlo (PMC). Este processo é repetido 100 vezes para três valores de temperaturas (T) diferentes: $T = 2,1$ (unidades de k_B^{-1}), $T = 1,9$ e $T = 1,7$.

O potencial utilizado (detalhado no Capítulo 1) é formado por uma parte atrativa (-3) relacionada aos contatos de tipos iguais (ou seja, contatos entre monômeros com o mesmo código de letras, por exemplo, A com A) e uma parte repulsiva (+3) relacionada aos contatos de tipos diferentes (ou seja, contatos entre monômeros com código de letras diferentes, por exemplo, A com B). A expressão geral para o potencial, neste caso, é dada por:

$$u = -3*cti + 3*ctd \quad (2.1)$$

onde:

cti : número de contatos de tipos iguais;

ctd : número de contatos de tipos diferentes.

Cada contato ocorre quando dois monômeros, não consecutivos, são primeiros vizinhos, ou seja, a distância entre eles tem valor 1,0 (parâmetro de rede).



Com o potencial acima a cadeia sempre enovelou, ou seja, alcançou a conformação nativa, totalizando 100 conformações nativas e para verificar detalhes deste enovelamento utilizamos duas grandezas:

- O instante (PMC) em que cada uma das 100 cadeias enovelou. Esse “instante” significa o número de cadeias que tiveram que ser crescidas para atingir uma conformação nativa.
- A quantidade de armadilhas topológicas de cada cadeia até atingir a conformação nativa.

A seguir, mostraremos em maiores detalhes os resultados de cada uma das duas grandezas acima.

2.1.2. Tempo de sucesso de *enovelamento* para as 100 conformações nativas

Os tempos de sucesso de *enovelamento* (t_e) para cada uma das 100 conformações nativas são mostrados na Figura 2.1. para três temperaturas diferentes.

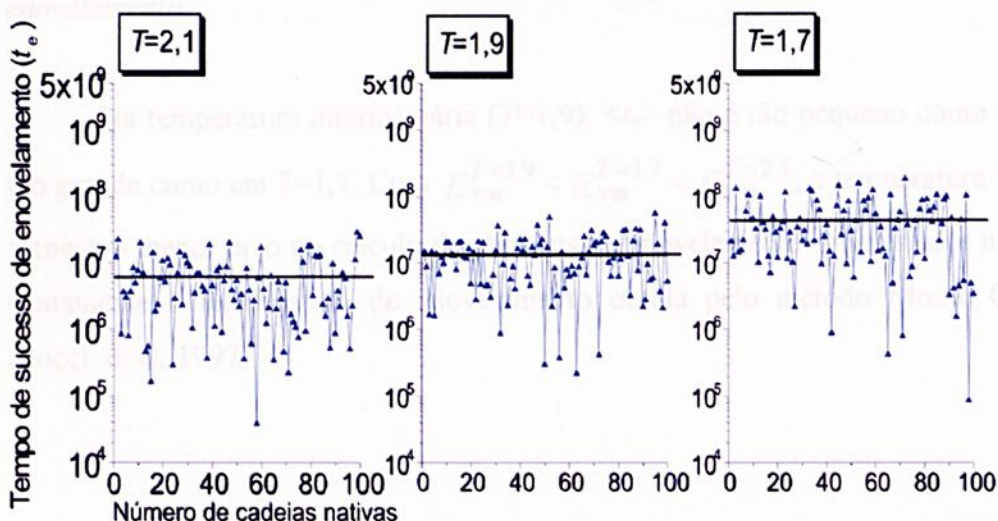


Figura 2.1. Tempo de sucesso de enovelamento para a cadeia enovelar, em três temperaturas diferentes. As linhas horizontais sólidas representam a média do tempo de sucesso de enovelamento. Tempo de sucesso de enovelamento significa o número de cadeias crescidas.

Pela Figura 2.1., notamos uma dependência crescente entre o tempo (PMC), para a cadeia atingir o estado nativo, à medida que diminuimos a temperatura.

a) $T = 2,1$: Para esta temperatura, o menor tempo de sucesso de *enovelamento* foi de $3,8 \times 10^4$ PMC, o tempo médio ($\langle t_e \rangle$) ficou em torno de $6,1 \times 10^6$ e tempo máximo em $2,9 \times 10^7$ PMC. Nesta temperatura, o erro do valor médio¹ (E_{vm}) representa 10,0% do tempo médio de sucesso de *enovelamento*.

b) $T = 1,9$: Neste caso, tivemos o tempo mínimo de sucesso de *enovelamento* foi de $2,2 \times 10^5$ PMC, já $\langle t_e \rangle$ ficou ao redor de $1,3 \times 10^7$ e tempo máximo em $5,5 \times 10^7$ PMC. Nesta temperatura, o tempo médio de sucesso de *enovelamento* foi 2,1 vezes maior comparado ao tempo médio na temperatura $T = 2,1$. O erro do valor médio (E_{vm}) representa 8,0% do tempo médio de sucesso de *enovelamento*.

c) $T = 1,7$: Nesta temperatura, o menor tempo de sucesso de *enovelamento* ocorreu em $8,7 \times 10^4$ PMC, $\langle t_e \rangle$ ficou em $4,4 \times 10^7$ e o tempo máximo em $1,9 \times 10^8$ PMC. Neste caso, o $\langle t_e \rangle$ foi 3,4 vezes maior quando comparado ao tempo médio em $T = 1,9$ e 7,2 vezes do que em $T = 2,1$. O erro do valor médio (E_{vm}) representa 9,3% do tempo médio de sucesso de *enovelamento*.

Na temperatura intermediária ($T=1,9$), $\langle t_e \rangle$ não é tão pequeno como em $T=2,1$ nem tão grande como em $T=1,7$. Com $E_{vm}^{T=1,9} < E_{vm}^{T=1,7} < E_{vm}^{T=2,1}$, a temperatura $T = 1,9$ é a que fornece o menor erro no cálculo dos tempos de enovelamento. Interessante notar que esta é exatamente a temperatura de enovelamento obtida pelo método Monte Carlo já usada (Socci et al, 1997).

¹ Por definição, temos que o desvio padrão (σ) mede a dispersão das medidas com relação ao valor médio; E_{vm} , mede o erro do valor médio, sendo $E_{vm} = \sigma / \sqrt{n}$, sendo $n = 100$.



No intervalo de temperatura estudado, notamos que quando a temperatura diminui, os tempos médios de sucesso de *enovelamento* aumentam, sendo estes respectivamente: $6,1 \times 10^6$, $1,3 \times 10^7$ e $4,4 \times 10^7$, ou seja, o enovelamento da cadeia na conformação nativa se torna, em média, mais lento com a diminuição da temperatura. Para baixas temperaturas (em $T = 1,7$), o processo de enovelamento envolve superar barreiras energéticas E_b (Gutin et al, 1996) acarretando um maior tempo de sucesso de *enovelamento* t_e , sendo $t_e \sim e^{-E_b/T}$. À medida que a temperatura aumenta, a energia térmica permite superar as barreiras; assim, o tempo de sucesso de *enovelamento* diminui. Em temperaturas muito altas, os tempos voltam a crescer devido à baixa estabilidade do estado nativo (resultados não mostrados), conforme resultados encontrados em outros trabalhos (Creighton, 1995) que empregaram o método Monte Carlo usual. Os tempos de sucesso de *enovelamento* (mínimos, médios e máximos) estão de acordo com os tempos de sucesso de *enovelamento* obtidos em simulações tradicionais Monte Carlo na rede (Creighton, 1995).

Determinado a ordem de grandeza do tempo de sucesso de *enovelamento* para cada uma das 100 conformações nativas em três temperaturas diferentes, passamos a observar a quantidade de armadilhas topológicas encontradas no processo para se atingir a conformação nativa.

2.1.3. Número de armadilhas topológicas para as 100 conformações nativas

Até ser obtida cada uma das 100 cadeias nativas, ocorreram muitas conformações proibidas, chamadas de armadilhas topológicas, ou simplesmente, armadilhas. Essa grandeza pode avaliar, em parte, como ocorreu o crescimento das cadeias. Entendemos por armadilha (Rosenbluth & Rosenbluth, 1955) cada vez que a cadeia tentou crescer e não conseguiu, pois ficou presa num “labirinto” sem ter alternativa de direção de crescimento (Figura 2.2.), pois isto violaria a condição de volume excluído.



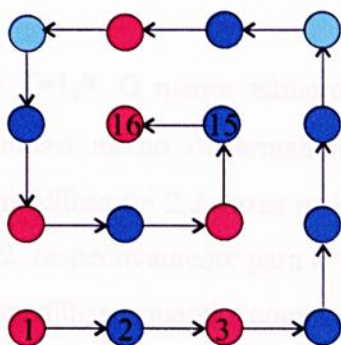


Figura 2.2. Esquema (em 2D) ilustrando o término do crescimento de uma cadeia em $N=16$, pois para $N > 16$, não é mais possível crescer a cadeia, então uma nova cadeia é gerada a partir de $N=3$. Essa conformação mal sucedida não é considerado um Passo Monte Carlo.

Medimos o número de armadilhas para cada uma das 100 cadeias nativas geradas, como mostrado na Figura 2.3. O número de armadilhas é dado pela soma das armadilhas até atingir a conformação nativa.

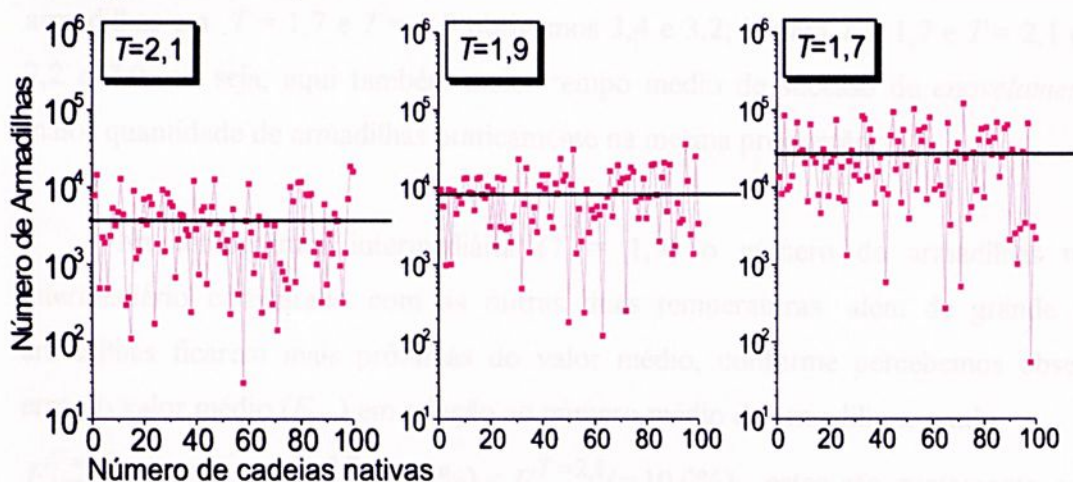


Figura 2.3. Quantidade de armadilhas até ser atingido o estado nativo.

Ao diminuir a temperatura, observamos um maior número de armadilhas. As retas horizontais representam o número médio de armadilhas.

Notamos que a quantidade de armadilhas varia com a temperatura e para deixar isto mais claro, serão expostos os valores mínimos, médios e máximos em cada uma das três temperaturas utilizadas.

a) $T=2,1$: Para esta temperatura, o número médio de armadilhas ficou em torno de $4,0 \times 10^3$ armadilhas. O menor e o maior números de armadilhas foram respectivos $2,9 \times 10^1$ e $1,8 \times 10^4$.

b) $T=1,9$: O menor número de armadilhas foi $1,2 \times 10^2$ e o maior número $3,5 \times 10^4$; já o número médio de armadilhas foi de $8,6 \times 10^3$. Nesta temperatura, o número médio de armadilhas foi 2,2 vezes maior comparada à temperatura $T = 2,1$. Obtemos os valores, 2,1 e 2,2, respectivamente para o tempo médio de sucesso de *enovelamento* e o número médio de armadilhas quando comparamos $T = 1,9$ com $T = 2,1$, ou seja, praticamente a mesma proporção entre estas duas grandezas.

c) $T=1,7$: O menor número de armadilhas foi de $5,4 \times 10^1$, o número médio ficou em $2,8 \times 10^4$ e o maior número de armadilhas foi de $1,2 \times 10^5$. Nesta temperatura, o número médio de armadilhas foi 3,2 vezes maior quando comparada com $T = 1,9$ e 7,0 vezes maior do que em $T = 2,1$. Ao compararmos os tempos médios de sucesso de *enovelamento* e o número de armadilhas em $T = 1,7$ e $T = 1,9$ obtivemos 3,4 e 3,2; já para $T = 1,7$ e $T = 2,1$ obtivemos 7,2 e 7,0, ou seja, aqui também maior tempo médio de sucesso de *enovelamento* leva a maior quantidade de armadilhas praticamente na mesma proporção.

Na temperatura intermediária ($T = 1,9$) o número de armadilhas também é intermediário comparado com as outras duas temperaturas, além de grande parte das armadilhas ficarem mais próximas do valor médio, conforme percebemos observando o erro do valor médio (E_{vm}) em relação ao número médio de armadilhas, sendo

$E_{vm}^{T=1,9} (= 8,0\%) < E_{vm}^{T=1,7} (= 9,3\%) < E_{vm}^{T=2,1} (= 10,0\%)$, estes são exatamente os mesmos valores de E_{vm} em relação aos tempos médios de sucesso de *enovelamento*.

Observamos que tanto na temperatura mais baixa como na mais alta, as oscilações no número de armadilhas aumentam; já na temperatura intermediária ($T = 1,9$) essas oscilações diminuem, conforme mostrado pelo erro do valor médio (E_{vm}). Os valores médios aumentam com a diminuição da temperatura, ou seja, existem mais armadilhas “topológicas” para a cadeia quando a temperatura diminui.

Com o objetivo de complementar os resultados da seção 2.1 (Resultados I) – onde o potencial foi fixo em $(-3,+3)$ – e observar o efeito da variação dos potenciais no processo de

enovelamento das cadeias polipeptídicas, foi utilizado outros potenciais no qual fixamos a parte relativa aos contatos de tipos iguais (cti) em -3 e variamos a parte relativa aos contatos de tipos diferentes (ctd) de -3 a $+3$.

2.2. Resultados II - Influência dos potenciais no processo de crescimento de cadeias

2.2.1. Introdução

Na seção anterior 2.1 (**Resultados I**), o critério de encerramento da simulação é satisfeito quando eram obtidas 100 cadeias nativas. Agora mudamos este critério fixando o tempo de simulação em 2×10^9 PMC, ou seja, crescemos 2×10^9 cadeias e observamos como está o empacotamento destas cadeias (com 27 monômeros) na temperatura $T = 1,9$. O que se pretende é um estudo conformacional em diferentes regimes de compactação induzida pelos potenciais. Investigamos que tipos de conformações são preferencialmente geradas à medida que o termo do potencial α (entre contatos de tipos diferentes) varia entre o extremo da atração -3 , que tornaria a cadeia um homopolímero e o outro extremo $+3$, que pode induzir a segregação de monômeros de tipos diferentes.

As grandezas físicas utilizadas neste novo conjunto de simulação foram:

- o número de armadilhas topológicas em função dos vários potenciais, apresentados abaixo;
- a porcentagem de conformações maximamente compactas (CMC) e a porcentagem de conformações não maximamente compactas (não CMC) para vários potenciais;
- raio de giração para as CMC e para as não CMC ;
- distância ponta a ponta para as CMC e para as não CMC.

O potencial empregado, neste caso, é constituído por uma parte atrativa fixa (-3) relacionada aos contatos de tipos iguais (cti) e uma parte que varia de -3 a $+3$ (ou seja, -3 , -

2, -1, 0, 1, 2, 3), relacionado aos contatos de tipos diferentes (*ctd*), conforme mostrado na Figura 2.4.

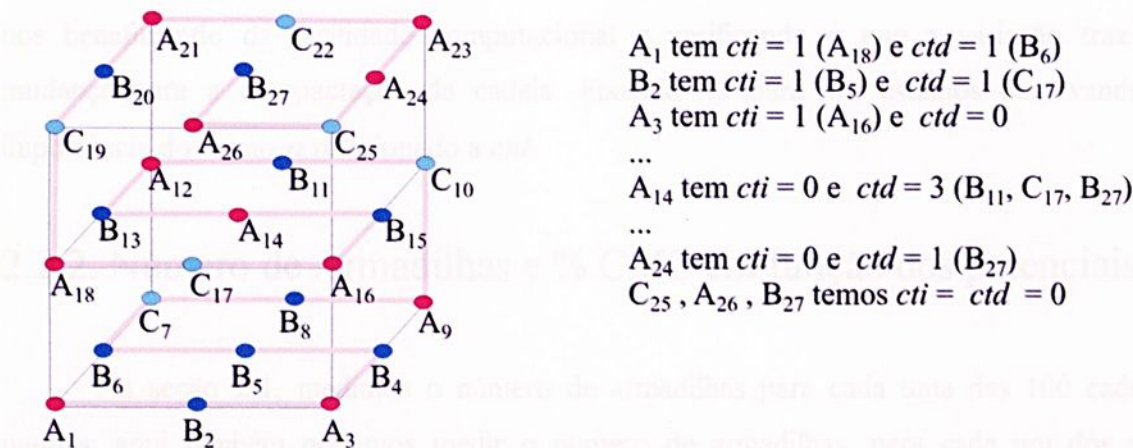


Figura 2.4. Cadeia de 27 monômeros numa CMC formada por um código de 3 letras (A, B, C), cujas posições são dadas pelos índices nas letras. Para um monômero estar em contato com outro, a distância entre eles deve ser 1,0 (parâmetro de rede) e ambos não podem ser consecutivos (como 1 com 2; 3 com 2 e 4, ou seja, ligados pela linha rosa). Os contatos entre cada monômero são de tipos iguais (*cti*) e de tipos diferentes (*ctd*). Ao lado da figura mostramos valores de *cti* e *ctd* para alguns monômeros e no total temos, para a conformação acima, *cti* = 10 e *ctd* = 18, ou seja, *cti*+*ctd*=28; conforme esperado para uma CMC. Note porém que a conformação acima não é nativa (*cti* = 28 e *ctd* = 0). Numa CMC, basta sabermos *cti* ou *ctd*, pois os contatos totais são *cti*+*ctd*=28, contudo numa não CMC devemos ter os valores de *cti* e *ctd*, já que *cti*+*ctd* < 28.

A expressão geral para o potencial é análoga à citada na seção 2.1, sendo agora dada por:

$$u = -3*cti + \alpha*ctd \tag{2.1}$$

onde α assume os seguintes valores: -3, -2, -1, 0, 1, 2, 3.

A variação dos potenciais pode ser interpretada melhor explicando o significado, por exemplo, do potencial (-3, +3) e do potencial (-3, -1). O potencial (-3, +3) significa que privilegiamos (representado pelo sinal negativo) os contatos de tipos iguais (*cti*) e desfavorecemos (representado pelo sinal positivo), na mesma proporção, os contatos de tipos diferentes (*ctd*). Já o potencial (-3, -1) significa que os contatos de tipos iguais (*cti*) são três vezes mais privilegiados do que os contatos de tipos diferentes (*ctd*), porém agora ambos os contatos (*cti* e *ctd*) são significativamente prováveis. Quanto ao potencial (-3, 0),

estamos privilegiando os contatos de tipos iguais (*cti*), tendo como indiferente à formação de contatos de tipos diferentes (*ctd*).

Ao fazermos o termo α da expressão acima variar de -3, -2, -1, 0, 1, 2, 3, estamos nos beneficiando da facilidade computacional e verificando o que a variação traz de mudança para a compactação da cadeia. Fixando -3 para *cti*, estamos observando a importância do termo α relacionado a *ctd*.

2.2.2. Número de Armadilhas e % CMC em função dos potenciais

Na seção 2.1, medimos o número de armadilhas para cada uma das 100 cadeias nativas; aqui também podemos medir o número de armadilhas, para cada um dos sete potenciais descritos acima, conforme mostrado na Figura 2.5.a. Observamos dois comportamentos, sendo o primeiro para os potenciais atrativos e o segundo relacionado aos potenciais neutro e repulsivo, lembrando que cada um dos sete potenciais está relacionado ao termo *ctd*, pois fixamos o potencial em -3 para *cti*.

À medida que os potenciais se tornam menos atrativos, ou seja de -3 a -1, as armadilhas aumentam. Após um salto brusco de -1 para 0, observamos um aumento do número de armadilhas à medida que o potencial se torna mais repulsivo de 0 a +3.

No intervalo de tempo 2×10^9 PMC, para cada um dos sete potenciais, observamos a porcentagem (%) de conformações maximamente compactas (isto é, as conformações tridimensionais de 27 monômeros numa disposição cúbica com $3 \times 3 \times 3$ monômeros), conforme mostrado na Figura 2.5.b. Nesta observamos, também, dois comportamentos distintos, um para potenciais atrativos e outro para potenciais neutro e repulsivo. Os potenciais atrativos (-3 a -1) apresentam altas porcentagens de CMC comparadas com as porcentagens de CMC dos demais potenciais, tendo essa porcentagem diminuído gradativamente quando o potencial se torna menos atrativo e mais repulsivo. De uma maneira geral, podemos dizer que quanto menos atrativo é o potencial, menor é a porcentagem de CMC.



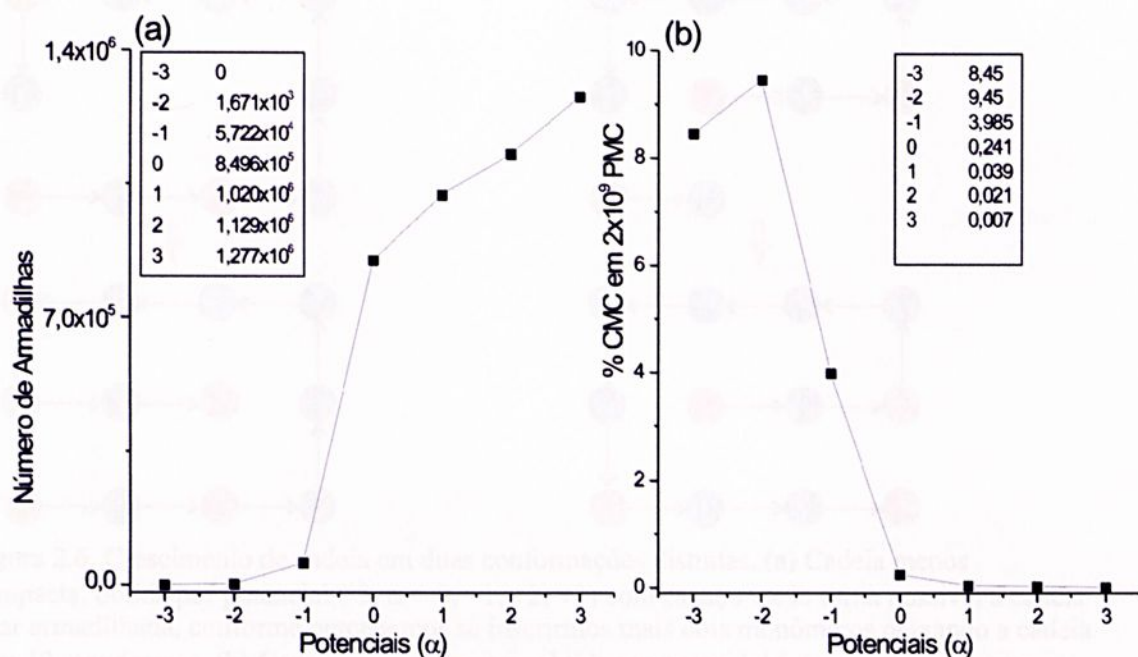


Figura 2.5. Foram crescidas 2×10^9 cadeias em $T=1,9$. A parte do potencial para *cti* foi fixada em -3 e variamos o termo α para *ctd*. Os valores mostrados nos gráficos também estão contidos em tabelas mostradas ao lado das curvas. (a) Número de armadilhas para $\alpha = -3, -2$ e -1 , temos poucas armadilhas e para $\alpha = 0, +1, +2$ e $+3$, temos muitas armadilhas. (b) % CMC para 7 valores de α . Quando o potencial vai de α negativos para positivos, ocorre uma diminuição brusca nas % de CMC.

Analisando as Figuras 2.5.a e 2.5.b, podemos verificar:

- Os números de armadilhas e a % CMC são complementares e opostos, pois enquanto as armadilhas crescem, a % CMC decresce. Isso mostra que quanto menos armadilhas existirem para um determinado número de conformações (no caso 2×10^9 PMC), mais compactas serão estas, conforme mostrado na Figura 2.6. Quanto menor α , menor a chance da cadeia ficar armadilhada, tanto que com $\alpha = -3$ não foi encontrada nenhuma cadeia armadilhada em 2×10^9 PMC. O parâmetro Δ_n (n de normalizado) definido como $\Delta_n = [-3(\text{energia de contato de tipos iguais}) - \alpha(\text{energia de contato de tipos diferentes})] / -3$ aparece como um divisor nos gráficos da Figura 2.5. Para $\Delta_n > 1$ (ou $\alpha > 0$) tem-se comportamento bem distinto na região $\Delta_n < 1$ (ou $\alpha < 0$). Esse parâmetro Δ_n , está associado à heterogeneidade entre interações de tipos iguais e de tipos diferentes e será discutido mais adiante.

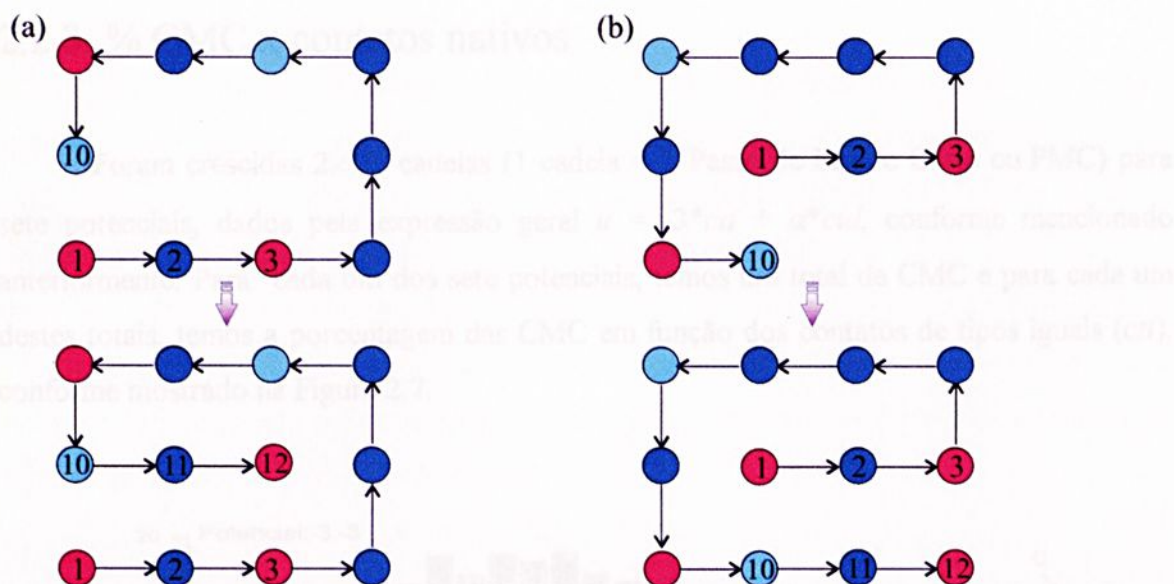


Figura 2.6. Crescimento de cadeia em duas conformações distintas. (a) Cadeia menos compacta, obtida por potencial $(-3, \alpha = 0, +1, +2, +3)$ com espaço vazio torna possível a cadeia ficar armadilhada, conforme percebemos se inserirmos mais dois monômeros deixando a cadeia com 12 monômeros. (b) Cadeia mais compacta, obtida por potencial $(-3, \alpha = -3, -2, -1)$ com menos chance de ficar armadilhada.

- Os potenciais atrativos (com termo $\alpha = -3, -2, -1$) fazem com que a cadeia tenha conformações muito distintas quando comparadas com potenciais nulos e repulsivos (com termo $\alpha = 0, +1, +2, +3$), conforme mostrados pelo número de armadilhas e % CMC.
- Cadeias crescidas somente com -3 e $\alpha = 0$ tem número de armadilhas e % CMC semelhante a quando utilizamos potencial com termo repulsivo (ou seja, -3 com $\alpha = +1, +2, +3$), contudo o acesso à conformação nativa torna $\alpha = 0$ bastante diferente de $\alpha = +1, +2$ e $+3$, conforme veremos a seguir. Vale lembrar que uma conformação nativa é um caso particular de uma CMC, com $cti = 28$ e $ctd = 0$.

Na Figura 2.5.b observamos a % CMC em relação a todas as conformações, ou seja, 2×10^9 PMC; porém a observação somente das CMC pode ser bastante útil. Vamos analisar o número total de CMC em função dos contatos de tipos iguais (cti) para cada valor de α .

2.2.3. % CMC x contatos nativos

Foram crescidas 2×10^9 cadeias (1 cadeia = 1 Passo de Monte Carlo ou PMC) para sete potenciais, dados pela expressão geral $u = -3*cti + \alpha*ctd$, conforme mencionado anteriormente. Para cada um dos sete potenciais, temos um total de CMC e para cada um destes totais, temos a porcentagem das CMC em função dos contatos de tipos iguais (*cti*), conforme mostrado na Figura 2.7.

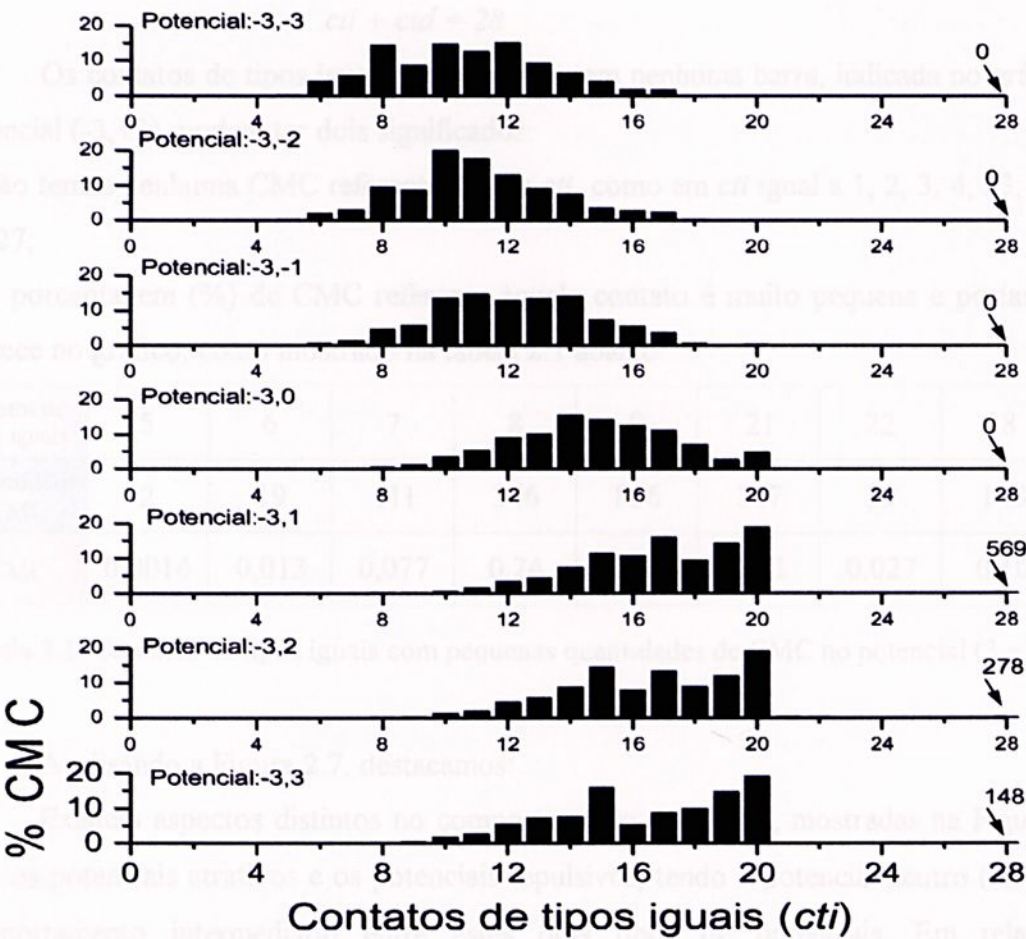


Figura 2.7. Porcentagem das CMC em função dos *cti* para sete potenciais. Se somarmos cada valor de %CMC (representado nos gráficos pelas barras verticais), totaliza-se 100 % das CMC, em cada potencial. Potenciais atrativos e neutro ($\alpha = -3, -2, -1, 0$), não geram nenhuma conformação nativa. Potenciais repulsivos ($\alpha = +1, +2, +3$), geram conformações nativas (com 28 *cti*), indicadas nos gráficos pelas setas. Notar o deslocamento dos histogramas para a direita (em direção aos estados nativos), à medida que α , relacionado com *ctd*, se torna mais repulsivo. Isso indica maior segregação entre monômeros de tipos diferentes ou maior número de contatos entre monômeros de tipos iguais.



Vamos comentar o gráfico do potencial (-3, +3), utilizado na seção 2.1, para explicar o significado da % CMC. A Figura 2.7. mostra como as porcentagens (%) de CMC estão distribuídas em função dos contatos de tipos iguais (*cti*) em cada CMC. Por exemplo: para a barra correspondente a 12 *cti*, significa que temos 5,43 % das CMC, ou 7769 CMC, dentre as 143086 CMC (ou (total de CMC) _{α}) em 2×10^9 PMC; o mesmo significado é dado para as outras %CMC com outros *cti*. Se ao invés de utilizarmos somente *cti*, estivéssemos utilizando somente *ctd*, teríamos no eixo das abscissas ao invés de 0 para 28 o inverso, ou seja, de 28 para 0, pois no caso das CMC temos

$$cti + ctd = 28 \quad (2.2)$$

Os contatos de tipos iguais (*cti*) que não tem nenhuma barra, indicada no gráfico do potencial (-3,+3), podem ter dois significados:

- 1) não temos nenhuma CMC referente àquele *cti*, como em *cti* igual a 1, 2, 3, 4, 23, 24, 25, 26, 27;
- 2) a porcentagem (%) de CMC referente àquele contato é muito pequena e portanto não aparece no gráfico, como mostrado na tabela 2.1 abaixo.

Contatos de tipos iguais	5	6	7	8	9	21	22	28
Quantidade de CMC	2	19	111	346	856	307	39	148
% CMC	0,0014	0,013	0,077	0,24	0,59	0,21	0,027	0,10

Tabela 2.1. Contatos de tipos iguais com pequenas quantidades de CMC no potencial (3,+3).

Analisando a Figura 2.7. destacamos:

- a) Existem aspectos distintos no comportamento das CMC, mostradas na Figura 2.7., para os potenciais atrativos e os potenciais repulsivos, tendo o potencial neutro ($\alpha = 0$) um comportamento intermediário entre estes dois tipos de potenciais. Em relação às conformações nativas, os potenciais atrativo e neutro ($\alpha = -3, -2, -1$ e 0), não são capazes de encontrar nenhuma conformação nativa; já para os potenciais repulsivos ($\alpha = +1, +2, +3$), existem várias conformações nativas (com 28 *cti*).



b) Quando o termo α do potencial vai aumentando, ou seja, indo de -3 para $+1$, há um deslocamento para à direita representado por um número maior de cti , o que significa que cada vez mais temos a chance de atingir a conformação nativa, conforme mostrados pelos valores acima de $cti = 28$ nos gráficos. O deslocamento para a direita é interpretado como maior segregação entre monômeros de tipos diferentes, dessa forma propiciando contatos entre monômeros de tipos iguais. Essa segregação crescente para α cada vez maiores será discutida mais adiante.

c) Os potenciais repulsivos são mais propícios para a obtenção das conformações nativas e dentre estes o valor que gera o maior número de conformações nativas é $+1$.

Os itens a, b, c, acima são resultados fundamentais do atual trabalho e serão novamente retomados, após análise de outras grandezas, como o raio de giração e distância ponta a ponta.

Assim como temos conformações maximamente compactas (CMC), também temos conformações não maximamente compactas (denominadas não CMC e cuja configuração não está numa disposição cúbica) no intervalo de tempo de 2×10^9 PMC. Para as CMC, nos interessamos apenas pela quantidade de contatos de tipos iguais (cti) de cada uma destas conformações, pois sabendo os cti , sabemos também os ctd , pois $cti + ctd = 28$. Para as não CMC, tomaremos como coordenada o número de contatos totais ct dada por:

$$ct = cti + ctd \quad (2.3)$$

Na seção seguinte observaremos como estão distribuídos os contatos totais para as não CMC.

2.2.4. % não CMC x contatos totais

Num procedimento análogo, à % CMC em função de cti , para sete potenciais u ($= -3*cti + \alpha*ctd$), passamos a observar cadeias de 27 monômeros não dispostas na forma cúbica, ou seja, as % não CMC, agora em função de todos os contatos (ct), conforme mostrado na Figura 2.8.



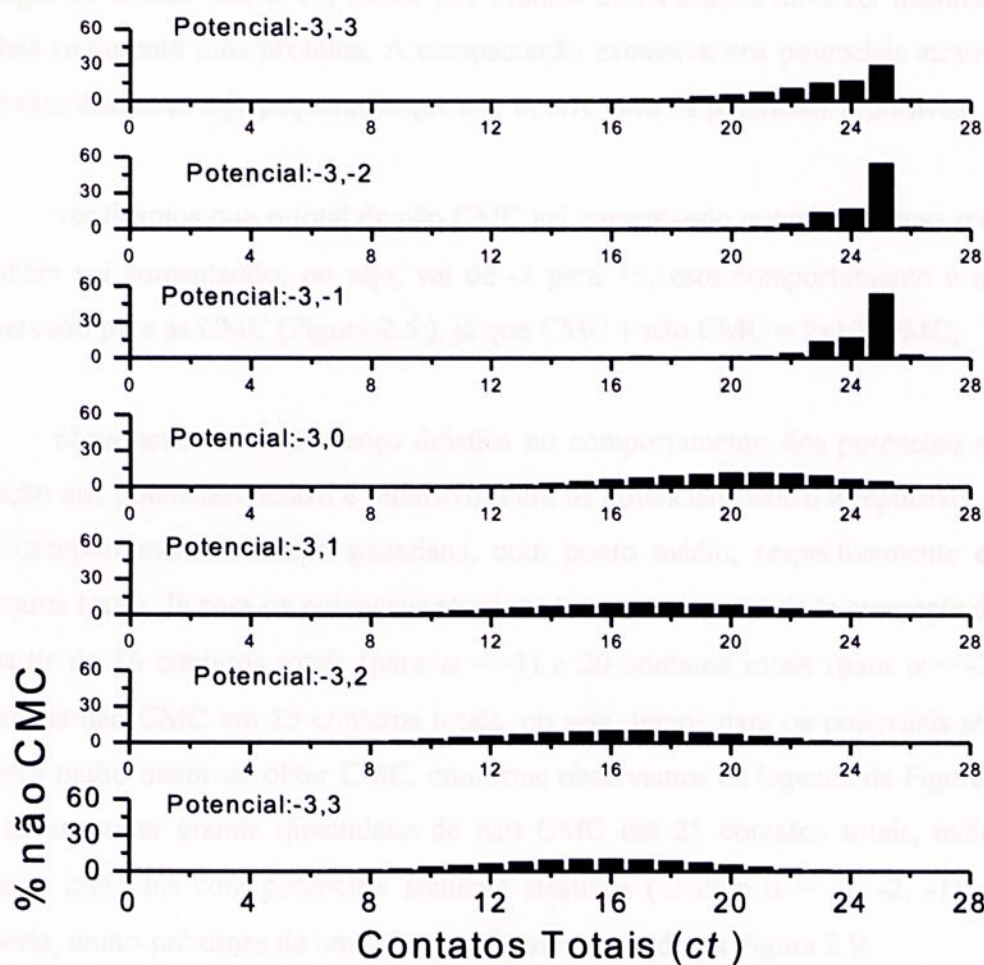


Figura 2.8. Porcentagem das não CMC em função de ct para sete potenciais. Verificamos dois comportamentos diferentes para as cadeias não CMC, sendo um para potenciais atrativos ($\alpha = -3, -2, -1$) e outro para potenciais neutros e repulsivos ($\alpha = 0, +1, +2, +3$). Neste caso 28 nunca é obtida, pois teríamos CMC.

Vamos explicar mais detalhadamente o gráfico do potencial $(-3, +3)$ da % não CMC para os contatos totais. O gráfico mostra como as não CMC estão distribuídas em função dos contatos totais (ct). Por exemplo: a barra correspondente a 16 ct , representa 10,37 % (ou $2,06 \times 10^8$) de conformações não CMC.

Destacamos alguns pontos da Figura 2.8.:

- a) notamos que, do potencial $(-3,0)$ para $(-3,+3)$, há um deslocamento para a esquerda na % não CMC o que implica configurações mais abertas em relação à excessiva compactação dos potenciais atrativos. É conhecido na literatura que a diferença entre

energia do estado nativo e a média dos estados desnaturados deve ser máxima para que a cadeia represente uma proteína. A compactação excessiva dos potenciais atrativos faz com que essa diferença seja pequena, o que não ocorre para os potenciais repulsivos.

b) verificamos que o total de não CMC vai aumentando quando o termo α do potencial também vai aumentando, ou seja, vai de -3 para +3, este comportamento é o inverso do observado para as CMC (Figura 2.5.), já que $CMC + \text{não CMC} = 2 \times 10^9 \text{ PMC}$;

c) observamos uma diferença drástica no comportamento dos potenciais atrativos em relação aos potenciais neutro e repulsivo. Para os potenciais neutro e repulsivo, observamos um comportamento do tipo gaussiana, com ponto médio, respectivamente em 20 e 16 contatos totais. Já para os potenciais atrativos temos uma quantidade crescente de não CMC a partir de 16 contatos totais (para $\alpha = -3$) e 20 contatos totais (para $\alpha = -2$ e -1) com máximas não CMC em 25 contatos totais, ou seja, temos para os potenciais atrativos uma chance muito maior de obter CMC, conforme observamos na legenda da Figura 2.5. O fato de termos uma grande quantidade de não CMC em 25 contatos totais, indicam que as cadeias crescidas com potenciais somente atrativos (-3 com $\alpha = -3, -2, -1$) são, em sua maioria, muito próximas de um cubo, conforme mostrado na Figura 2.9.

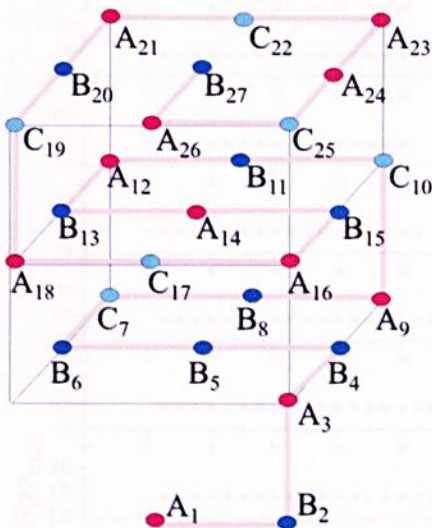


Figura 2.9. Cadeia numa conformação muito próxima da CMC, pois apenas dois monômeros estão fora da conformação cúbica, sendo $cti = 8$ e $ctd = 16$, então $ct = 24$.

Nas duas seções anteriores (2.2.3 e 2.2.4), analisamos como estão distribuídos os contatos de tipos iguais (*cti*) e totais (*ct*), respectivamente para as conformações maximamente compactas (CMC) e as conformações não maximamente compactas (não CMC). Esse procedimento foi realizado para cada um dos sete potenciais utilizados: $u = -3*cti + \alpha*ctd$. Agora, no intuito de obter mais informações sobre o empacotamento das cadeias de 27 monômeros na temperatura $T = 1,9$, utilizamos as grandezas raio de giração (*Rg*) e distância ponta a ponta (*dpp*), para as CMC e as não CMC.

2.2.5. Raio de giração e Distância ponta a ponta (CMC) x contatos nativos

A distância ponta a ponta (*dpp*) é definida como a distância entre o primeiro monômero da cadeia e o último monômero, ou seja, a distância entre os monômeros 1 e 27. Na Figura 2.10. mostramos, para as cadeias CMC, o raio de giração (*Rg*) e a distância ponta a ponta (*dpp*), em função dos contatos de tipos iguais (*cti*) para sete potenciais utilizados.

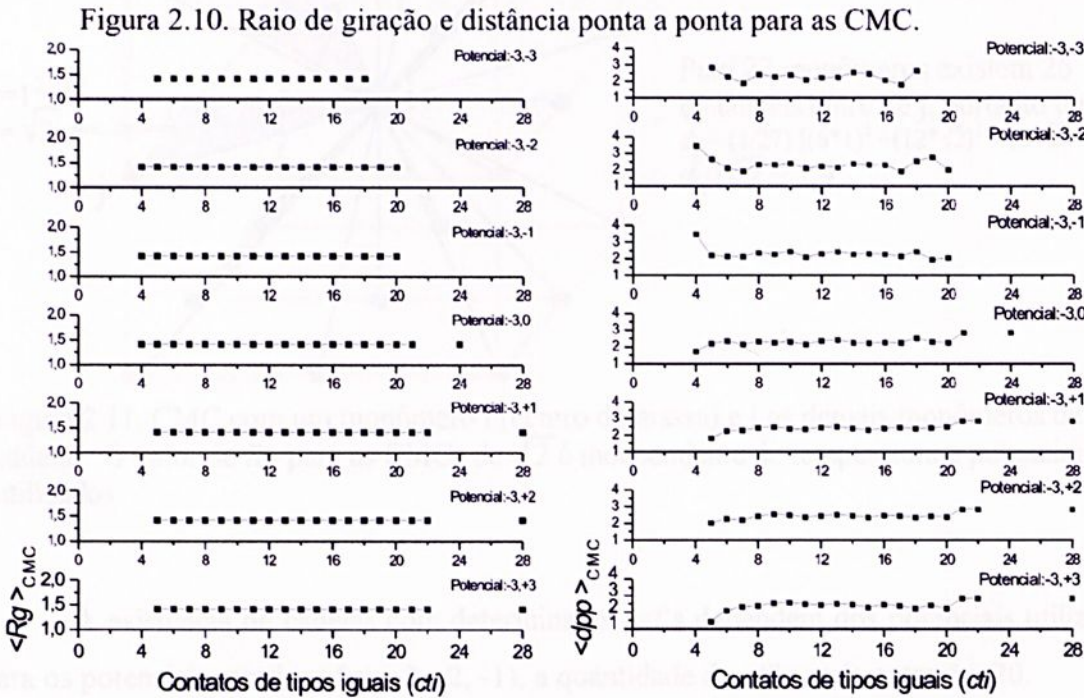


Figura 2.10. Raio de giração médio $\langle Rg \rangle_{CMC}$ e distância ponta a ponta média $\langle dpp \rangle_{CMC}$ em função dos contatos de tipos iguais (*cti*), para os sete potenciais.

Para justificarmos o porquê de termos obtido o valor de $\sqrt{2} \sim 1,4$ para os raios de giração em todos os potenciais nas CMC, vale lembrar que o raio de giração (R_g) é a distância média entre os segmentos da cadeia e seu centro de massa definido por:

$$R_g = \frac{1}{N} \sqrt{\sum_{i < j} d_{ij}^2} \tag{2.4}$$

onde i é o monômero referente ao centro de massa (monômero no centro do cubo), j os demais monômeros da cadeia; N é a quantidade de monômeros da cadeia (no caso 27) e d_{ij} é a distância entre o monômero i e o j .

A equação 2.4 foi utilizada para obter os valores de R_g tanto para as CMC quanto para as não CMC, porém como sabemos quem é o centro de massa das conformações num cubo (CMC) podemos mostrar de maneira simples porque nestas conformações o raio de giração deve ser $\sqrt{2} \sim 1,4$, conforme mostrado na Figura 2.11.

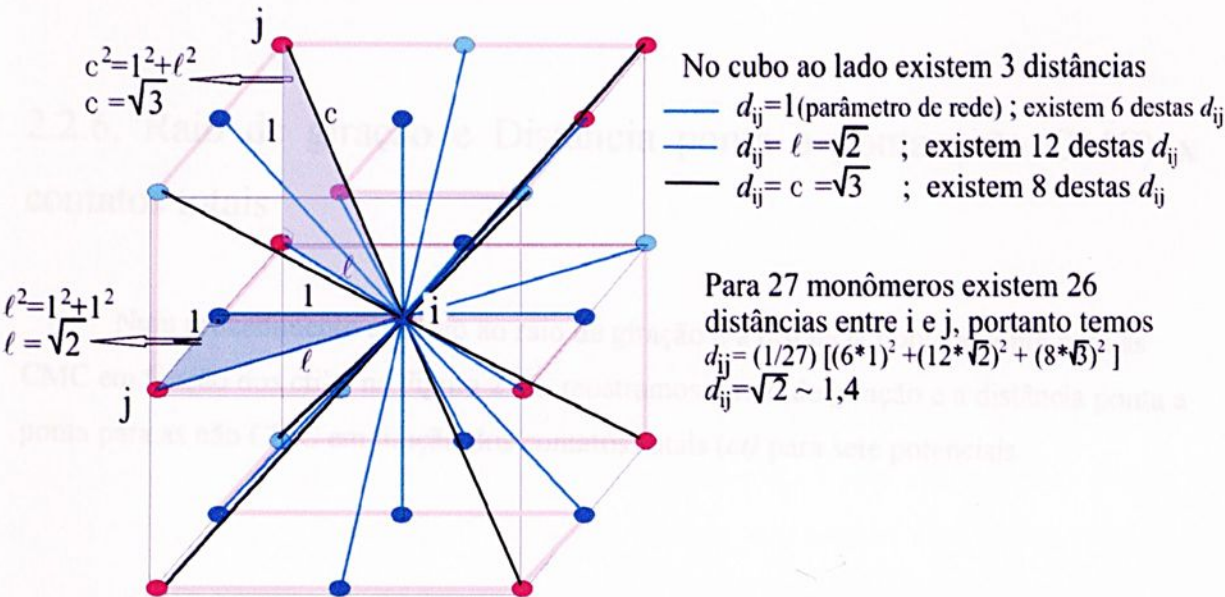


Figura 2.11. CMC com um monômero i (centro de massa) e j os demais monômeros da cadeia. O valor de R_g para as CMCs de $\sqrt{2}$ é independente de temperatura e potenciais utilizados.

A existência de cadeias com determinados cti 's dependem dos potenciais utilizados. Para os potenciais atrativos ($\alpha=-3, -2, -1$), a quantidade de cti 's varia entre 4 e 20.

No caso do potencial neutro $(-3, 0)$, observamos que os cti 's das cadeias variam entre 4 e 24. O ponto correspondente a $cti = 24$ representa um aumento da chance de se

atingir a conformação nativa, se compararmos com os potenciais atrativos. Para os potenciais repulsivos, há uma variação de cti entre 5 e 22, mas temos também um ponto representando o estado nativo em $cti = 28$. A não existência de cadeias com cti 's entre 23 e 27, mas com $cti = 28$ reafirma maior eficiência dos potenciais repulsivos para a obtenção do estado nativo.

Nas Figuras 2.7. e 2.8. quando não tinha pontos nos gráficos – para um determinado cti ou ct – era devido à inexistência ou a pequena quantidade, para respectivamente, CMC ou não CMC. Agora a inexistência de pontos para os gráficos de Rg e dpp – para um certo cti ou ct – mostra a ausência de cadeias com estes contatos.

Analisando a Figura 2.10., observamos que enquanto o raio de giração (Rg) se mantém constante e a distância ponta a ponta varia, representando a diversidade de cubos que caracterizam as CMC.

2.2.6. Raio de giração e Distância ponta a ponta (não CMC) x contatos totais

Num procedimento análogo ao raio de giração e a distância ponta a ponta para as CMC em função dos cti 's, na Figura 2.12. mostramos o raio de giração e a distância ponta a ponta para as não CMC em função dos contatos totais (ct) para sete potenciais.



Figura 2.12. Raio de giração e distância ponta a ponta para as não CMC.

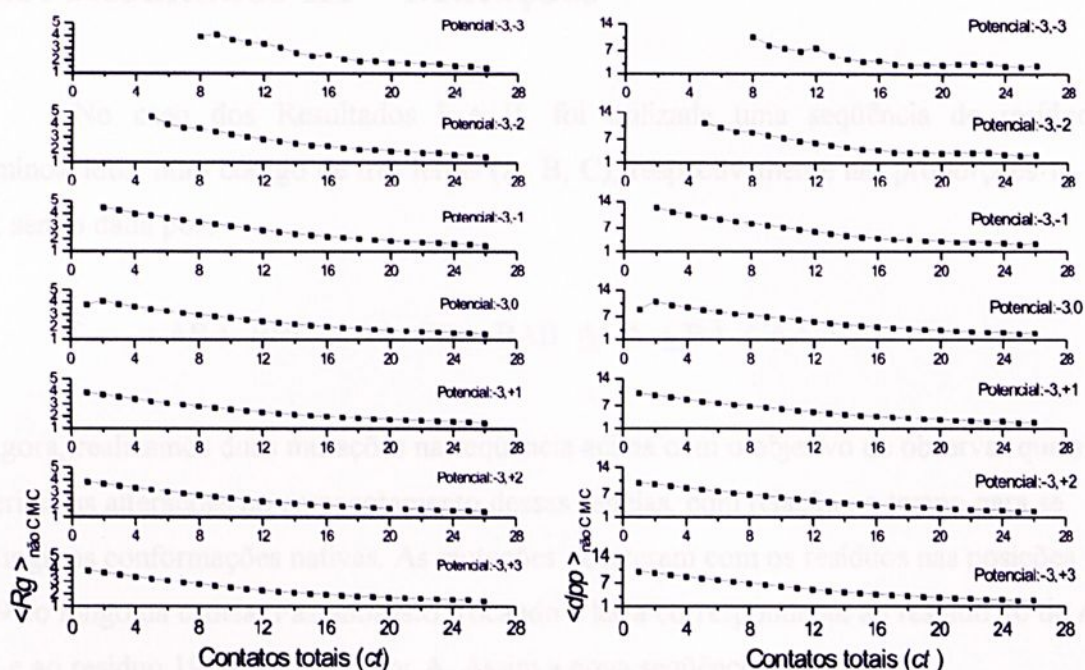


Figura 2.12. Raio de giração médio $\langle Rg \rangle_{\text{não CMC}}$ e distância ponta a ponta média $\langle dpp \rangle_{\text{não CMC}}$ para não CMC, em função dos contatos totais (ct), para sete potenciais. Até 16 ct , à medida que aumentamos a quantidade de ct , o $\langle Rg \rangle$ vai diminuindo; acima de 16 ct , $\langle Rg \rangle$ continua diminuindo, porém muito mais lentamente, sendo $\langle Rg \rangle_{\text{não CMC}} \sim \langle Rg \rangle_{\text{CMC}} \sim 1,4$.

O que muda nos gráficos da Figura 2.12. é a quantidade de pontos existentes nos potenciais atrativos e os demais potenciais. Para os potenciais atrativos, observamos um deslocamento para a esquerda à medida que α aumenta (de -3 para -2 e para -1) com as cadeias apresentando menor quantidade de ct 's, ou seja, elas passam a serem mais abertas (ou menos compactas). Potenciais altamente atrativos provocam rápido colapso da cadeia representado pela ausência de cadeias abertas ($ct < 2$).

Para finalizar a parte de Resultados, passamos a analisar os efeitos de mutações(ou seja, mudança nos tipos de monômeros, representada pelo código de letras) na sequência de resíduos de aminoácidos das cadeias crescidas, conforme descritos abaixo em Resultados III.

2.3. Resultados III – Mutações

No caso dos Resultados I e II, foi utilizada uma sequência de resíduos de aminoácidos, num código de três letras (A, B, C), respectivamente nas proporções 11 : 10 : 6, sendo dada por:

ABA BBB CBA CBA BAB ACA CBA CAA CAB

Agora, realizamos duas mutações na sequência acima com o objetivo de observar quais seriam as alterações no empacotamento dessas cadeias, com relação ao tempo para se atingir as conformações nativas. As mutações ocorreram com os resíduos nas posições 16 e 19 ao longo da cadeia. Fazemos isto trocando a letra correspondente ao resíduo 16 de A por C e ao resíduo 19, trocando C por A. Assim a nova sequência é dada por:

ABA BBB CBA CBA BAB CCA ABA CAA CAB

A mutação foi estrategicamente realizada nestas posições devido ao fato de que a nova sequência tem estado fundamental (nativo) não degenerado. Pode-se enumerar todos os cubos formados por 27 monômeros o que resulta em 103346 conformações. Colocando a sequência em todos os cubos pode-se calcular a energia de todos os estados. A sequência sugerida possui energia mínima correspondendo a um único cubo. Isso garante uma propriedade de proteína que é a de ter estado nativo único.

Os potenciais utilizados para a simulação destas mutações foram: $(-3,+3)$ e $(-3,-1)$, onde usamos três valores de temperaturas diferentes: $T = 1,70$; $T = 1,50$ e $T = 1,38$ para cada um dos potenciais durante 2×10^9 PMC.

Observamos que para as três temperaturas e para os dois potenciais citados acima que nenhuma conformação nativa foi obtida com o método de crescimento de cadeias. Para confirmarmos este resultado triplicamos o tempo de simulação, utilizando o intervalo de tempo de 6×10^9 cadeias, mas mesmo assim nenhuma cadeia atingiu a conformação nativa. As simulações tradicionais de Monte Carlo encontram o estado nativo no intervalo de tempo mencionado. Isso leva a crer que o método de crescimento de cadeias é adequado



apenas para seqüências otimizadas e sem contatos frustrados. A seqüência mutada possui três contatos frustrados, que são contatos de alta energia entre monômeros distintos. Aparentemente, o método gera probabilidades muito pequenas para se obter tais contatos.

3.1. Discussões

Nossas resultados confirmam, em parte, a observação feita por outros autores de que a utilização de seqüências otimizadas e sem contatos frustrados, conforme discutido anteriormente, é essencial para a obtenção de resultados satisfatórios. Os resultados obtidos com as seqüências otimizadas e sem contatos frustrados são muito mais frequentes do que os outros potenciais ($\alpha = 0, 1, 2, 3$), conforme foi observado nas Figuras 2.4.b e 2.8. Para deixar mais explícitos estes dois resultados, os mesmos foram reunidos na Figura 3.1.



CAPÍTULO 3 – DISCUSSÕES E CONCLUSÕES

3.1. Discussões

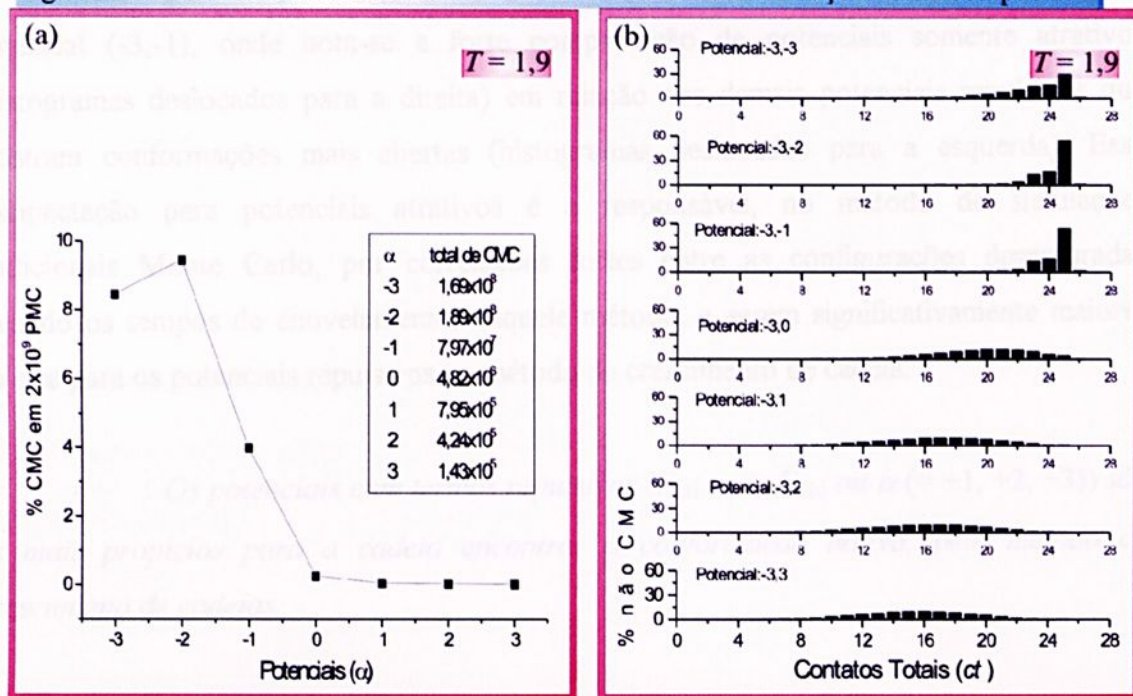
- *Potenciais somente atrativos (-3 , α ($= -3, -2, -1$)) não conseguem fazer a cadeia acessar o estado nativo, porém eficientemente levam a conformações bastante compactas. Aparentemente, a compactação excessiva (comparada com aquela oriunda de potenciais repulsivos) prejudica fortemente o acesso à conformação nativa pelo método de crescimento de cadeias.*

Nossos resultados confirmam, em parte, a afirmação acima sobre a utilização dos potenciais somente atrativos, conforme discussão abaixo:

1. Os potenciais somente atrativos levam a cadeia a conformações mais compactas de maneira muito mais freqüente do que os outros potenciais ($\alpha = 0, 1, 2, 3$), conforme foi mostrado nas Figuras 2.5.b. e 2.8. Para deixar mais explícitos estes dois resultados, os mesmos foram reunidos na Figura 3.1.



Figura 3.1. Potenciais atrativos levam as cadeias a terem conformações mais compactas



Potenciais somente atrativos (-3 , $\alpha = -3, -2, -1$) favorecem as CMC muito mais que os demais potenciais, conforme observamos pelos três primeiros pontos no gráfico.

Potenciais somente atrativos (-3 , $\alpha = -3, -2, -1$) favorecem as não CMC próximas das CMC, por isso os três primeiros gráficos estão deslocados para a direita próximos de $ct=28$.

2. A utilização do potencial -3 , -1 (Socci et al, 1997) pode ser representada pela utilização de dois parâmetros, E_{avg} e Δ definidos como:

$$E_{avg} = (E_{cti} + E_{ctd})/2, \quad \Delta = (E_{ctd} - E_{cti}) \quad (3.1)$$

sendo E_{cti} e E_{ctd} , respectivamente, a energia de contato entre monômeros de tipos iguais e de tipos diferentes, dados por:

$$E_{cti} = E_{avg} - (\Delta/2); \quad E_{ctd} = E_{avg} + (\Delta/2).$$

E_{avg} representa a condução ("drive") média para formação de contatos de qualquer natureza ou compactação da cadeia. Se E_{avg} for menor do que zero, a formação de contatos será favorecida, e a cadeia colapsará. Δ é uma divisão nos níveis de energia responsável pela heterogeneidade da interação entre os monômeros do heteropolímero e permitirá algumas seqüências encodificar para o estado fundamental único. Para $E_{cti} = -3$ e $E_{ctd} = -1$, $E_{avg} = -2$ e $\Delta = -2$.

Pela Figura 3.1. acima, damos uma informação complementar a utilização do potencial (-3,-1), onde nota-se a forte compactação de potenciais somente atrativos (histogramas deslocados para a direita) em relação aos demais potenciais repulsivos que mostram conformações mais abertas (histogramas deslocados para a esquerda). Essa compactação para potenciais atrativos é a responsável, no método de simulações tradicionais Monte Carlo, por correlações fortes entre as configurações desnaturadas levando os tempos de enovelamento, naquele método, a serem significativamente maiores do que para os potenciais repulsivos no método de crescimento de cadeia.

- *Os potenciais com termos repulsivos ($E_{cti} = -3$, E_{ctd} ou $\alpha (= +1, +2, +3)$) são os mais propícios para a cadeia encontrar a conformação nativa, pelo método de crescimento de cadeias.*

Nossos resultados mostram que potenciais com o termo atrativo (-3) para *cti* e o termo repulsivo (α) para *ctd*, conseguem fazer as cadeias acessarem o estado nativo dentro do intervalo de tempo de 2×10^9 . Como o termo atrativo para *cti* foi mantido constante para todos os sete potenciais utilizados e somente o termo relacionado a *ctd* foi variado, concluímos que a heterogeneidade no potencial representada pelo termo $\Delta = E_{cti} - E_{ctd}$ é fundamental para a cadeia alcançar o estado nativo. Normalizando Δ através da grandeza

$$\Delta_n = \frac{\Delta}{E_{cti}} \text{ (n de normalizado), notamos que para } \Delta_n > 1 \text{ parte das cadeias crescem em}$$

direção ao estado nativo e para $\Delta_n < 1$ o estado nativo não é atingido dentro do intervalo 2×10^9 . Assim, Δ_n irá representar a heterogeneidade nas interações entre os monômeros, tornando-os específicos. O termo Δ_n surge como um parâmetro para caracterizar a especificidade das interações:

- $\Delta_n > 1$: interações cada vez mais específicas e mais facilmente se atinge a conformação nativa;
- $\Delta_n < 1$: interações menos específicas e menos provável é a obtenção da conformação nativa.

Uma outra característica interessante para classificar os regimes de comportamento das cadeias consiste no princípio da segregação. Se $\epsilon_{ii}(\epsilon_{jj})$ é a energia para formar o contato entre dois monômeros $ii(jj)$ e ϵ_{ij} é a energia necessária para formar dois contatos i e j , então os monômeros i e j serão otimamente segregados se a condição

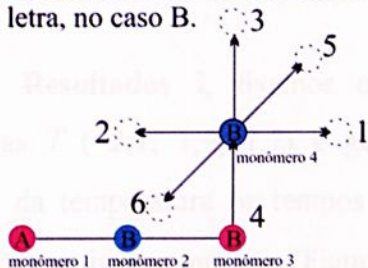
$$2\epsilon_{ij} \geq \epsilon_{ii} + \epsilon_{jj} \quad (3.2)$$

for satisfeita (Tang & Wingreen, 1997). No caso do enovelamento que estamos estudando, esse princípio é melhor observado para o potencial $(-3, 3)$, ou seja, $\epsilon_{ii} = \epsilon_{jj} = -3$ daí a energia para romper contatos ii ou jj é 3 e a energia para formar contatos ij é 3; logo $2 \times 3 = 3 + 3$. Notamos que esse potencial é eficiente para crescer cadeias enoveladas. Os outros potenciais (somente atrativos) embora sejam utilizados para estudar modelos de proteínas (Tang & Wingreen, 1997) mostram tempo de enovelamento extremamente superiores aos oriundos de potenciais atrativo e repulsivo.

A seguir na Figura 3.2., mostramos a importância dos termos repulsivos, pelo método de crescimento de cadeias, para se ter o estado nativo, com $cti = 28$ e $ctd = 0$.

Inserindo o monômero 5 na cadeia

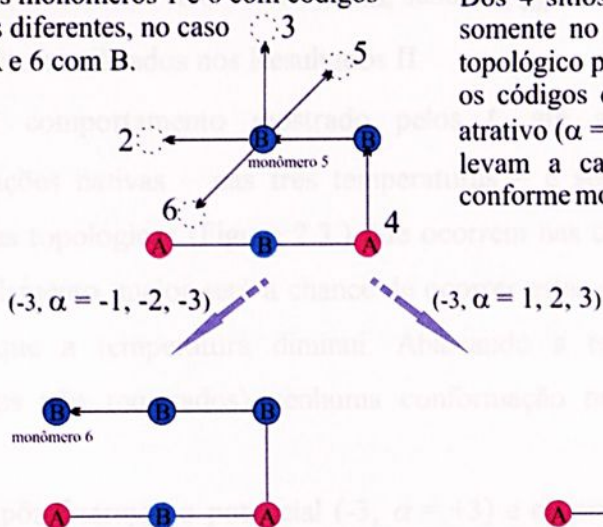
Sejam os monômeros 2 e 5 com o mesmo código de letra, no caso B.



Dos 5 sítios disponíveis para o monômero 5 somente no sítio 2 é possível existir contatos topológicos para o monômero 5. Assim tanto os potenciais atrativo quanto repulsivo favorecem o monômero 5 a ficar no sítio 2.

Inserindo o monômero 6 na cadeia

Sejam os monômeros 1 e 6 com códigos de letras diferentes, no caso 1 com A e 6 com B.



Dos 4 sítios disponíveis para o monômero 6 somente no sítio 2 é possível existir contato topológico para este monômero. Agora, devido os códigos de letras diferentes, os potenciais atrativo ($\alpha = -1, -2, -3$) e repulsivo ($\alpha = 1, 2, 3$) levam a cadeia a conformações diferentes, conforme mostramos abaixo.

- Potenciais somente atrativos favorecem o monômero 6 a ficar no sítio 2, o que já impede a conformação da cadeia de continuar a ser candidata à nativa, pois $ctd \neq 0$.

- Potencial repulsivo ($\alpha = 1, 2, 3$) favorece o monômero 6 a ficar em sítios $\neq 2$, por exemplo no sítio 5. Neste caso a conformação da cadeia continua sendo candidata à nativa.

Figura 3.2. Ilustração de um possível crescimento de cadeia até o monômero 6.

3.2. Conclusões

Em **Resultados I**, fixamos o potencial em $(-3, \alpha = +3)$ e utilizamos três temperaturas T ($=2,1; 1,9; 1,7$) e geramos 100 cadeias nativas. Notamos que com a diminuição da temperatura os tempos de *enovelamento* médios, $\langle t_e \rangle$, para se atingir as conformações nativas aumentam (Figura 2.1.), para a menor temperatura o tempo de CPU foi maior. A temperatura intermediária ($T = 1,9$) foi a que mostrou o menor erro presumível da média (ou erro do valor médio), E_{vm} , sendo $E_{vm}^{T=1,9} < E_{vm}^{T=1,7} < E_{vm}^{T=2,1}$, por isso esta foi a temperatura utilizados nos Resultados II.

O comportamento mostrado pelos t_e até ser atingida cada uma das 100 conformações nativas – nas três temperaturas – é seguido exatamente pelo número de armadilhas topológicas (Figura 2.3.) que ocorrem nas cadeias; pois quanto maior o tempo de enovelamento, maior será a chance de ocorrer estas armadilhas e isso é mais freqüente à medida que a temperatura diminui. Abaixando a temperatura de $T=1,7$ para $T=1,0$ (resultados não mostrados) nenhuma conformação nativa foi atingida durante $2,0 \times 10^9$ PMC.

Após fixarmos o potencial $(-3, \alpha = +3)$ e observar como variava t_e e as armadilhas topológicas para três temperaturas diferentes; em **Resultados II**, passamos a analisar a influência da variação do potencial (através da alteração do termo α , de -3 a $+3$) no número de armadilhas e em outras grandezas não utilizadas em Resultados I.

Se os potenciais atrativos são capazes de conduzir as cadeias a CMC, já os potenciais repulsivos ($\alpha = +1, +2, +3$) são capazes de levar as cadeias às conformações nativas (com $cti=28$), mesmo com menores quantidades de CMC e número de armadilhas bem superiores no caso de potenciais atrativos. Já o potencial neutro ($\alpha = 0$) conduz as cadeias a possuírem conformações intermediárias dos potenciais atrativo e repulsivo, pois em relação à quantidade de CMC o potencial neutro é muito parecido com os potenciais atrativos (Figura 2.5.b); e quanto à capacidade de levar as cadeias ao estado nativo o potencial neutro tem o mesmo efeito dos potenciais atrativos, pois em ambos o estado nativo não é atingido (Figura 2.7.). Se o método de crescimento de cadeias fosse escolhido para gerar proteínas em grandes quantidades, provavelmente seria otimizado um meio



termo entre potenciais extremamente atrativos (que geram compactação excessiva e produz muitos contatos não nativos) e potenciais extremamente repulsivos (que embora sejam mais eficientes para produzir contatos nativos, produzem bem mais armadilhas topológicas).

Em relação às mutações, em **Resultados III**, feitas ao trocar o código de letras para os monômeros 16 e 19 – para o potencial repulsivo ($\alpha=+3$), em três temperaturas ($T = 1,70$; $1,50$ e $1,38$), o acesso à conformação nativa é drasticamente alterada, pois, após a mutação nenhuma conformação nativa foi obtida. Quanto ao potencial atrativo ($\alpha=-1$) a mutação não altera o acesso a conformação nativa, que nunca é obtida. De fato, para conclusões mais completas sobre o efeito de mutações, no processo de crescimento de cadeias, outras mutações devem ser feitas, em outras posições ao longo da cadeia, inclusive com outros potenciais. Mas esses resultados preliminares sugerem que o método de crescimento de cadeias é apropriado para seqüências otimamente desenhadas, uma vez que pequenas mutações na seqüência impedem o acesso à conformação nativa.

De forma sucinta, os principais resultados desse trabalho sugerem que:

- 1) o método de crescimento de cadeias é mais eficiente para potenciais repulsivos, que geram menos interações não nativas;
- 2) a heterogeneidade entre as interações tem um valor mínimo ($\Delta_n > 1$) para que o método cresça cadeias nativas;
- 3) o regime onde o princípio da segregação é notadamente não satisfeito (potenciais neutro e atrativo): a cadeia jamais cresce em direção à conformação nativa;
- 4) pequenas mutações que geram alguns contatos frustrados (energia alta) impedem, por completo, a obtenção do estado nativo da seqüência mutada, sugerindo que o método é eficiente para seqüências otimamente desenhadas.
- 5) O método é adequado para enovelamento que não são precedidos por colapso hidrofóbico. Potenciais altamente atrativos podem induzir a cadeia ao colapso seguido de um enovelamento lento. O método de crescimento de cadeias não reproduziu, nesse trabalho, esse comportamento. Aparentemente a compactação excessiva produz significativo número de contatos não nativos o que impede o acesso à conformação nativa.



Referências Bibliográficas

- Campbell, M. K., *Bioquímica*, Artmed (2000).
- Chan, H. S. & Dill, K. A., *The Protein Folding Problem*. Physics Today. 24 (1993).
- Creighton, T., Disulphide-coupled protein folding pathways. Phill. Trans. R. Soc. Lond. B **348**, 5-10 (1995).
- Dimitrievski, K., Kasemo, B., Zhdanov, V. P., *Rate-determining moves in protein folding*. J. Chem. Phys. **115**, 2841 (2001).
- Duan, Y. & Kollman, P. A., *Computational protein folding: From lattice to all-atom*. IBM Systems Journal. **40**, 297 (2001).
- Gan H. H., Tropsha, A., Schlick, T., *Generating folded protein structures with a lattice chain growth algorithm*. J. Chem. Phys. **113**, 5511 (2000).
- Gutin, A. M., Abkevich, V. I., and Shakhnovich, E. I., Chain Length Scaling of Protein Folding Time. Phys. Rev. Lett. **77**, 5433 (1996).
- Lehninger, A. L., *Princípios de Bioquímica*, Sarvier, São Paulo (1985).
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E., *Equation of State Calculations by Fast Computing Machines*. Journal of Chemical Physics. **21**, 1087 (1953).
- Pande, V. S., Grosberg, A. Y., Tanaka, T., *Heteropolymer freezing and design: Towards physical models of protein folding*. Reviews of Modern Physics. **72**, 259 (2000).



Rosenbluth, M. N. & Rosenbluth, A.W.. *Monte Carlo Calculation of the Average Extension of Molecular Chains*. J. Chem. Phys. **23**, 356 (1955).

Socci, N. D., Nymeyer, H., Onuchic, J.N.. Exploring the protein folding funnel landscape. Physica D. **1706**, 1 (1997).

Tang, H. C. & Wingreen, N. S.. Phys.Rev. Lett **79**, 765 (1997).

Yon, J. M.. *Protein Folding: a perspective for biology, medicine and biotechnology*. Braz. J. Med. Biol. Res. **34(4)**, 419-435 (2001).



APÊNDICE I – Fluxograma completo do programa de crescimento de cadeias

