

## 5. RESULTADOS E CONCLUSÕES

A taxa de enovelamento das proteínas pode depender de fatores diversos, como o tamanho da cadeia – proteínas pequenas tendem a se enovelar mais rapidamente<sup>15, 16</sup>; da seqüência de aminoácidos – experimentos com mutações que alteram a seqüência original de aminoácidos, têm mostrado que a taxa de enovelamento pode variar até três ordens de magnitude<sup>51</sup>; e da topologia da estrutura nativa<sup>52, 53, 54, 55</sup>.

O surpreendente resultado experimental relacionando a taxa de enovelamento de pequenas proteínas (dezenas de aminoácidos) com parâmetros topológicos globais, como a ordem de contato relativo, juntamente com a ocorrência freqüente de padrões estruturais repetitivos nas estruturas das proteínas reais, constituem as evidências mais fortes de que a topologia da estrutura nativa é um dos importantes determinantes da “velocidade” do enovelamento, se bem que não é o único.

A hipótese levantada neste trabalho atribui ao conteúdo de estruturas tipo-secundárias da nativa – e não necessariamente a sua ordem de contato relativo, como o principal fator topológico determinante da taxa de enovelamento. Esta premissa pôde ser demonstrada como válida para um modelo em rede, embora ainda uma prescrição de como o conteúdo preciso de estruturas tipo-secundárias, válido também para proteínas reais, ainda tenha que ser definitivamente estabelecida, pois a forma corrente de se medir esta quantidade, não passa de um indicativo de pertinência, ou seja, somente indica se um particular aminoácido pertence ou não a uma estrutura secundária. Assim, esta simples “contabilidade” não captura eventuais composições de padrões configuracionais mais completos.

As 52 estruturas “nativas” empregadas neste estudo, foram selecionadas de forma a cobrir todo o intervalo dos valores da ordem de contato  $\chi$  (entre 0,2381 e 0,49471) e representar topologicamente distintos padrões configuracionais, conforme mostra a sua classificação em sete categorias, identificadas por símbolos específicos, os quais serão detalhados logo adiante. A Tabela 5.1 resume dados numéricos referentes às 52 estruturas nativas utilizadas como referenciais na simulação do modelo computacional adotado (potencial estéreo-químico). É possível observar que, de maneira geral, conforme  $\chi$  aumenta  $\log k_f$  diminui. A presença de padrões como [\*STTS\*] (configurações números 1, 21, 32) ou longas seqüências de Ts (número 17) promovem rápido envelhecimento independentemente de  $\chi$ . Por sua vez, a presença de padrões *Snail* e *Staple* (números 7, 33) desacelera o processo de envelhecimento, também independentemente de  $\chi$ .

Os Gráficos 5.1 e 5.2 contêm os resultados obtidos para a taxa de envelhecimento em função de  $\chi$ . A rigor, eles mostram a dependência do  $\log k_f$  com  $\chi$ , para duas condições distintas: no primeiro caso, foi utilizado somente o potencial hidrofóbico  $\{e_{i,j} = h_i + h_j\}$ , e no último, os resultados correspondem ao modelo estéreo-químico propriamente dito,  $\{e_{i,j} + c_{i,j}\}$ , isto é, vínculos estéricos foram incluídos. Uma rápida comparação entre os gráficos já mostra a importância das especificidades estéricas; os detalhes e implicações são comentados a seguir.

No primeiro caso, o Gráfico 5.1 mostra que a maioria das configurações apresenta sucesso de envelhecimento maior que 66% (40 das 52 estruturas) na janela de tempo  $t_w$ , como ilustra o eixo do lado direito do gráfico, e como mostrado pelas linhas tracejadas conectando asteriscos. Onze estruturas nativas não alcançam fração de sucesso superior a  $2/3$ , e somente em um caso o sucesso foi totalmente nulo (0% de sucesso). Em princípio, o sucesso do envelhecimento poderia crescer se a janela de tempo  $t_w$  aumentasse. Contudo, foram feitas simulações extras utilizando uma janela de tempo maior (de tamanho  $2 \times t_w$ ), e o sucesso do envelhecimento não mudou. Isto sugere que para determinadas estruturas nativas, a chance de ficarem armadilhadas em outras estruturas nativas compactas diferentes da sua nativa é muito grande. De fato, foi o que se observou com a única estrutura que não encontrou sua configuração nativa na janela de

tempo  $t_w$  (total de  $8,1 \times 10^8$  configurações geradas): durante 15 simulações independentes, para cada uma delas, uma diferente configuração CSA foi encontrada, mas todas diferentes da nativa.

Para a linha da regressão do Gráfico 5.1 foram empregadas as quarenta estruturas com sucesso de enovelamento maior que 66% (círculos cheios). Os valores de  $\log k_f$  para as doze estruturas restantes foram simbolicamente fixadas em  $\log k_f = -8$ , com a finalidade apenas de enfatizar sua fraca performance (círculos vazios). Os dados resultaram uma correlação fraca entre  $\log k_f$  e  $\chi$ , com o coeficiente de correlação de Pearson  $r = -0,387$ .

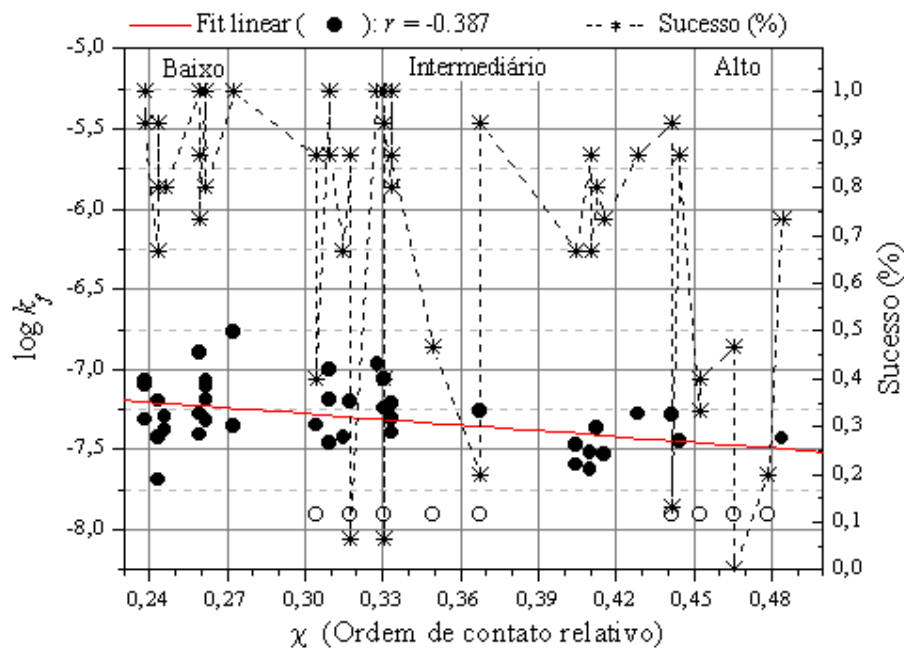


Gráfico 5.1. Comportamento do logaritmo da taxa de enovelamento  $\log k_f$  e sucesso de enovelamento (---\*---) como função da ordem de contato relativa. Simulação efetuada somente com potencial hidrofóbico  $\{e_{i,j} = h_i + h_j\}$ . A fraca correlação entre  $\log k_f$  e  $\chi$  foi calculada com as estruturas nativas que tiveram sucesso de enovelamento maior que 66%. Note que o intervalo de  $\chi$  foi segmentado em três faixas complementares de Baixa, Intermediária e Alta ordem de contato. Estas faixas são representadas pelas Letras L, I e H, respectivamente.

	<i>CO</i>	<i>S</i>	[*STTS*]	[*TT*]	[*TTTT*]	$\{T...T\}_{max}$	<i>Staple</i>	<i>Snail</i>	$\log k_f$	Grupo
1	0,2381	9	8	0	0	2	0	0	-5,369	■
2	0,2381	9	7	0	0	2	0	0	-5,988	●
3	0,2381	9	6	0	0	2	0	0	-5,695	●
4	0,2434	7	4	3	2	4	0	0	-6,585	▼
5	0,2434	7	3	3	2	6	0	0	-6,835	▼
6	0,2434	7	2	3	2	6	0	0	-6,78	▼
7	0,2460	9	6	0	0	2	0	1	-6,94	▼G
8	0,2460	6	0	6	3	10	0	0	-6,218	▼
9	0,2593	7	4	2	1	4	0	0	-6,143	●
10	0,2593	8	3	1	0	6	0	0	-6,739	▼
11	0,2593	7	3	3	1	7	0	0	-6,539	▼
12	0,2619	9	8	0	0	2	0	0	-5,45	■
13	0,2619	9	7	0	0	2	0	0	-5,741	●
14	0,2619	3	0	11	4	16	0	0	-5,649	●
15	0,2619	4	0	8	3	10	0	0	-6,442	▼
16	0,2725	8	6	0	0	3	0	0	-5,855	●
17	0,2725	3	0	10	2	15	0	0	-5,494	■
18	0,3042	7	4	3	1	2	0	0	-5,999	●
19	0,3042	4	0	7	1	10	0	0		▼
20	0,3095	9	8	0	0	2	0	0	-5,408	■
21	0,3095	9	7	0	0	2	0	0	-5,479	■
22	0,3095	3	0	7	2	11	0	0	-6,295	●
23	0,3148	7	3	4	2	6	0	0	-6,603	●
24	0,3175	8	6	0	0	4	0	0	-5,804	▲
25	0,3175	6	0	4	1	6	1	0	-7,053	●G
26	0,3280	6	4	4	1	6	0	0	-5,642	▲
27	0,3307	8	6	0	0	3	0	0	-5,954	▲
28	0,3307	6	1	4	0	7	0	0	-6,96	●
29	0,3307	6	0	3	0	12	0	0		▼
30	0,3307	5	0	5	1	10	0	1		★G
31	0,3333	5	0	5	1	7	0	0		▼
32	0,3333	7	6	2	1	4	0	0	-5,465	■
33	0,3333	7	2	2	0	6	1	0		▼G
34	0,3492	6	2	3	0	9	0	0	-6,998	●
35	0,3677	8	4	1	0	5	0	0	-6,702	●
36	0,3677	5	0	4	1	6	0	0	-7,179	●
37	0,4048	6	4	2	0	6	0	0		▼
38	0,4048	4	0	9	3	12	0	0	-7,137	●
39	0,4101	7	3	2	0	5	0	0	-6,888	●
40	0,4101	5	0	4	1	8	0	0	-7,636	●
41	0,4127	6	2	2	0	6	0	0		▼
42	0,4153	7	4	1	0	5	0	0	-6,02	▲
43	0,4286	7	3	0	0	3	0	0	-7,393	●
44	0,4418	7	2	1	0	4	0	0	-7,394	●
45	0,4418	5	0	5	0	5	0	0		▼
46	0,4444	6	3	1	0	7	0	0	-7,607	●
47	0,4524	7	3	1	0	4	0	0	-7,479	●
48	0,4524	6	0	4	1	6	0	0		▼
49	0,4656	5	2	4	1	6	0	0	-7,566	●
50	0,4656	4	0	7	1	9	0	0	-7,161	●
51	0,4788	5	0	5	1	10	0	0	-7,444	●
52	0,4841	8	1	0	0	6	1	0		★G

Tabela 5.1. Características topológicas das 52 CSA. Os símbolos identificam sete grupos distintos. De maneira geral, conforme  $\chi$  aumenta,  $\log k_f$  diminui;  $\log k_f$  se refere ao modelo estéreo-químico. Para os casos em que o sucesso de enovelamento superior a 80% não foi alcançado, os valores de  $\log k_f$  correspondentes não foram calculados.

Contrastando com o caso considerado acima, o Gráfico 5.2 mostra que o efeito do conjunto de vínculos estéricos  $\{c_{ij}\}$  na cinética do envelhecimento é marcante, aumentando a taxa de envelhecimento em mais de duas ordens de magnitude para as estruturas com pequeno  $\chi$ . Pode-se observar de maneira global que o efeito dos vínculos estéricos na cinética do envelhecimento aumenta conforme  $\chi$  diminui, mas existem exceções notáveis. Para tratar estas exceções, a influência dos atributos topológicos na configuração nativa na cinética do envelhecimento foi considerada em detalhe, agrupando as 52 estruturas nativas em sete grupos. Seis destes grupos mostram como as especificidades topológicas podem alterar a suposta correlação linear entre  $\log k_f$  e  $\chi$ , e o sétimo grupo é composto por estruturas que se comportam muito próximas à citada correlação.

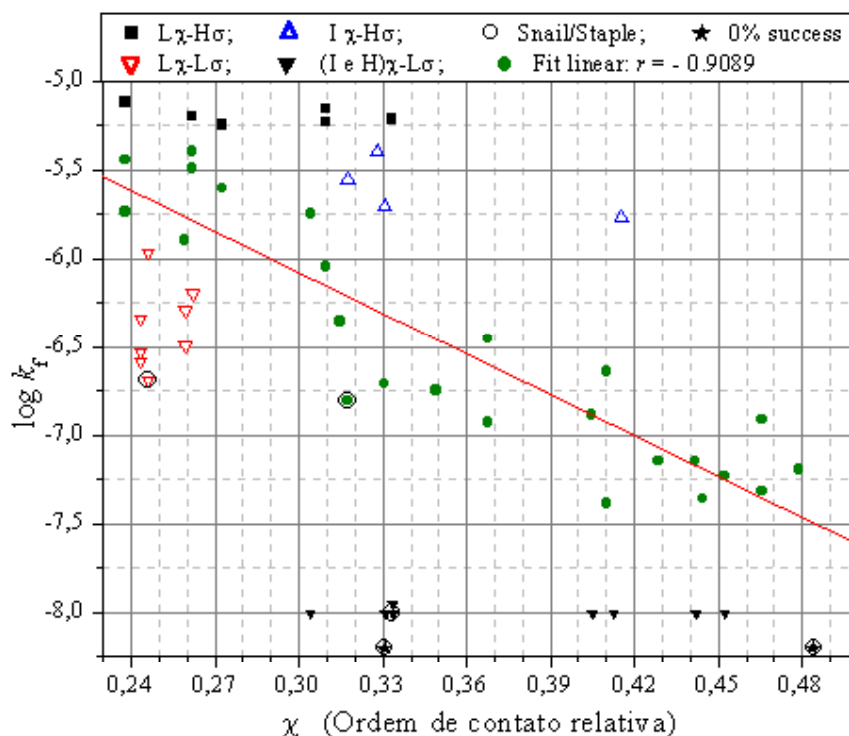


Gráfico 5.2. Comportamento do logaritmo da taxa de envelhecimento  $\log k_f$  como função da ordem de contato relativa, nas simulações em que o conjunto de vínculos estéricos  $\{c_{ij}\}$  foram considerados.

### 5.1 Análise dos diferentes grupos de estruturas notáveis

O primeiro grupo representado por quadrados cheios (■) corresponde às estruturas nativas com taxa de enovelamento no nível mais alto. Este grupo pertence à faixa de baixa ordem de contato,  $\chi \leq 0,34$ , e taxa de enovelamento  $k_f$  praticamente independente de  $\chi$ ,  $\log k_f \cong -5,2$ .

Estas configurações CSA possuem em comum o que se denominou aqui alto “conteúdo de estruturas tipo-secundárias”, e representada pelo símbolo “ $\sigma$ ”. Este fato é melhor visualizado através de seu correspondente mapa de contatos inter-resíduos, Figura 5.1, do que pela estrutura 3-D, Figura 4.1 (à página 41). Em todos os casos mostrados na Figura 5.1, é notável: (i) a persistência de linhas pontilhadas ao longo da diagonal principal, correspondendo a  $\alpha$  hélices (e folhas  $\beta$  paralelas); (ii) linhas pontilhadas paralelas à diagonal secundária, à semelhança das folhas  $\beta$  antiparalelas, e; (iii) combinação de ambos, que corresponderia à organização das estruturas secundárias, em *motifs* específicos.

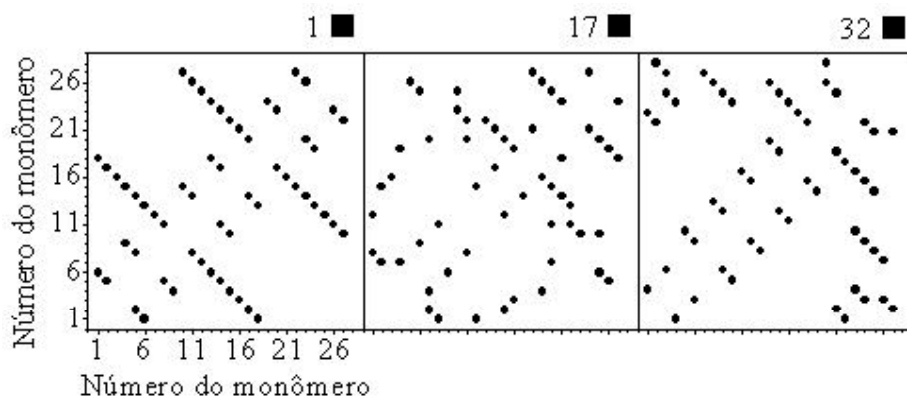


Figura 5.1. Mapa de contatos inter-resíduos (primeiros vizinhos na rede) das estruturas mais rápidas: baixa ordem de contato ( $L\chi$ ) e alto conteúdo de estruturas tipo-secundárias ( $H\sigma$ ), facilmente identificado através da ocorrência persistente de linhas ao longo (e próximo) da diagonal principal e linhas paralelas à diagonal secundária, lembrando  $\alpha$  hélices e folhas  $\beta$ . Ver também Tabela 5.1.

No presente modelo em rede, o conteúdo de estruturas tipo-secundárias é devido às combinações peculiares dos padrões estruturais  $[*STTS*]$  e  $\{*T...T*\}$ ,

principalmente. As estruturas nativas neste primeiro grupo (■) são caracterizadas por seu baixo valor de ordem de contato (ou  $L\chi$ ) e alto conteúdo de estruturas tipo-secundárias (ou  $H\sigma$ ), representado simbolicamente por  $L\chi-H\sigma$ . O padrão estrutural [*\*STTS\**] é um dos responsáveis pelo  $H\sigma$  (ver Tabela 5.1): algumas estruturas nativas deste grupo apresentam seis ou mais destes padrões (é possível até oito destes tais padrões numa única estrutura). As seqüências ininterruptas de  $Ts$  presentes também podem agir como facilitadores do enovelamento, como o caso ilustrado na estrutura 17, que apresenta quinze  $Ts$  ininterruptos (Figura 5.2). Esta seqüência de  $Ts$ , adequadamente arranjados, é responsável pela longa linha pontilhada ao longo da diagonal principal de seu correspondente mapa de contatos inter-resíduos (Figura 5.1). Independentemente do valor de  $\chi$ , o número de padrões [*\*STTS\**] de uma estrutura particular é um fator importante que favorece a cinética do enovelamento. Simulações extras envolvendo estruturas nativas com alto conteúdo de padrões [*\*STTS\**] ou longa seqüência ininterrupta de  $Ts$  (até no máximo 23  $Ts$ ) confirmam que a presença destes padrões resulta em um maior valor da taxa de enovelamento, do que se somente o valor de  $\chi$  da estrutura fosse considerado.

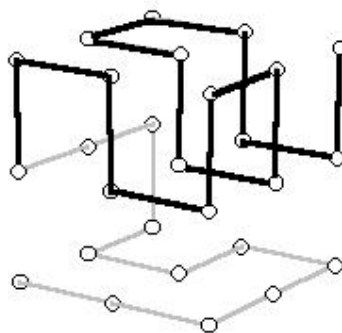


Figura 5.2. Estrutura 17 mostrando uma seqüência ininterrupta de quinze  $Ts$ , que pode atuar como facilitadora no processo de enovelamento.

Algumas estruturas nativas que apresentam ordem de contato com valor intermediário ( $0,30 < \chi < 0,42$ ) podem também apresentar alto conteúdo de estruturas tipo-secundárias, isto é,  $I\chi-H\sigma$ , Figura 5.3. O grupo representado por triângulos vazios para cima ( $\Delta$ ), Gráfico 5.2 e Tabela 5.1, apresenta alta taxa de enovelamento, com  $\log k_f$  entre  $-5,7$  e  $-5,3$  e são praticamente independentes de  $\chi$ . Como no grupo anterior, o alto

valor de  $\sigma$  é claramente expresso nos correspondentes mapas de contatos, independente do valor de  $\chi$ . É interessante observar na Figura 5.3 e com a ajuda da Tabela 5.1, que a taxa de enovelamento diminui conforme a aparência do mapa de contatos inter-resíduos se transfigura da forma simples, com longas linhas de pontos paralelas e perpendiculares, para uma forma mais cheia e complexa.

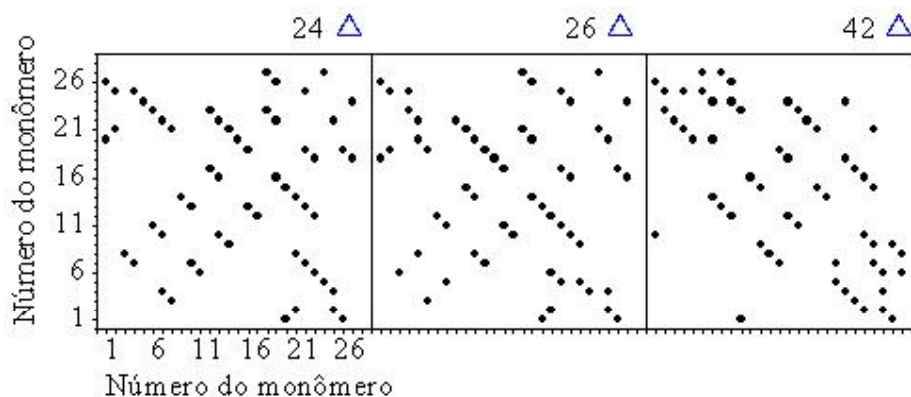


Figura 5.3. Mapa de contatos inter-resíduos (primeiros vizinhos na rede) das estruturas com valor de ordem de contato intermediária ( $L\chi$ ) e alto conteúdo de estruturas tipo-secundárias ( $H\sigma$ ). Ver também Tabela 5.1.

Por sua vez, o grupo representado por triângulos vazios para baixo ( $\nabla$ ) é composto de estruturas nativas que apresentam baixo conteúdo de estruturas tipo-secundárias embora com baixo valor de ordem de contato,  $\chi \leq 0,26$ , isto é,  $L\chi - L\sigma$ . De fato, neste caso o mapa de contato inter-resíduos, Figura 5.4, transformou-se consideravelmente, mostrando padrões mais complexos: os pontos estão mais espalhados pelo mapa, e as linhas paralelas à diagonal principal e secundária estão segmentadas ou difusas, dando aparência de formas mais infladas. Consistentemente com a premissa de que a cinética do enovelamento depende de  $\sigma$ , estas estruturas nativas apresentam velocidade de enovelamento significativamente menor do que se esperaria se somente o valor da ordem de contato fosse considerado, como enfaticamente mostrado no Gráfico 5.2 ( $\nabla$ , ver também Tabela 5.1).

Algumas estruturas nativas apresentam sucesso de enovelamento menor do que 80% e, portanto, foram consideradas aqui como estruturas ineficientes para se enovelar. Todas elas têm ordem de contato  $\chi > 0,3$  (ordem de contato intermediária e alta) e baixo

conteúdo de padrões tipo-secundários (I e  $H\chi-L\sigma$ ) (Figura 5.5). No Gráfico 5.2 estas estruturas foram representadas de forma simbólica com taxa de enovelamento fixadas em  $\log k_f \cong -8$ , somente para enfatizar a fração de enovelamento menor que 80%; todas elas são representadas por triângulos cheios virados para baixo. É importante observar que estas estruturas não correspondem, necessariamente, àquelas estruturas nativas que apresentaram baixa fração de sucesso de enovelamento com a simulação efetuada somente com o potencial hidrofóbico, Gráfico 5.1. Este é somente mais um dos efeitos dos vínculos estéricos na cinética do enovelamento, pois geralmente aumentam a taxa de enovelamento, porém em alguns casos, especialmente para grande valor de  $\chi$ , podem atuar adversamente, restringindo o sucesso do enovelamento.

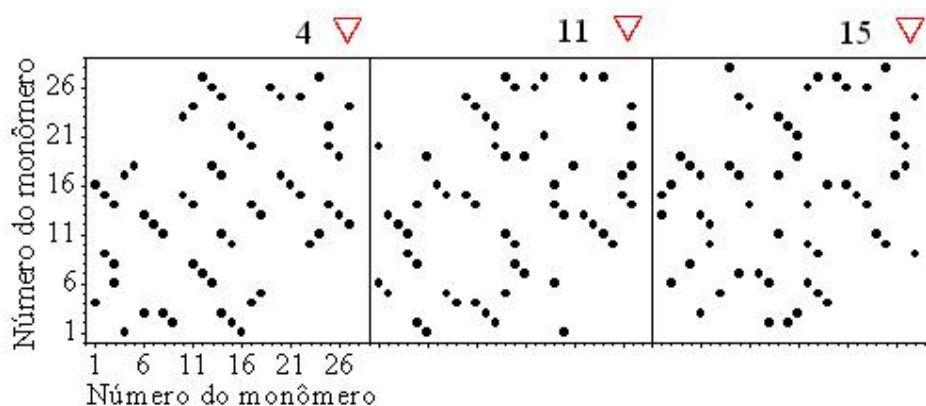


Figura 5.4. Mapa de contatos inter-resíduos (primeiros vizinhos na rede) das estruturas com valor de ordem de contato baixa ( $L\chi$ ) e baixo conteúdo de estruturas tipo-secundárias ( $L\sigma$ ). Ver também Tabela 5.1.

Contudo, em somente dois casos a cadeia jamais se enovelou nas suas respectivas configurações nativas durante a janela de tempo  $t_w$  prefixada. Elas estão representadas por estrelas cheias ( $\star$ ) no Gráfico 5.2. Uma delas (número 30 da Tabela 5.1) possui ordem de contato de valor intermediário,  $\chi = 0,33069$ , e a outra (número 52 da Tabela 5.1), apresentam alto valor de ordem de contato,  $\chi = 0,48413$ . Seus mapas de contato estão representados na Figura 5.6. Sua ineficiência no enovelamento é devido ao baixo conteúdo de estrutura tipo-secundárias e também pela presença de padrões configuracionais do tipo *Snail* e *Staple*. Assim, estas duas estruturas também estão incluídas no grupo representado por círculos vazios (O), que identificam configurações

com padrões estruturais com tendência a afetar negativamente a cinética do enovelamento. Perfazem, este grupo, cinco das 52 configurações analisadas, com valores distintos de  $\chi$ .

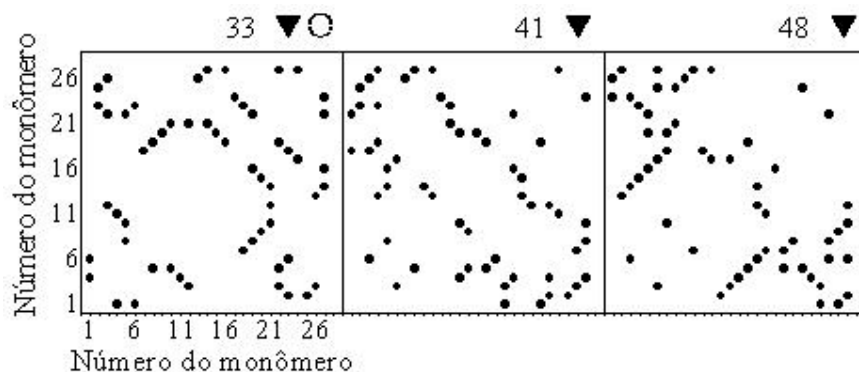


Figura 5.5. Mapa de contatos inter-resíduos (primeiros vizinhos na rede) das estruturas com valor de ordem de contato intermediária e alta (I e  $H\chi$ ) e baixo conteúdo de estruturas tipo-secundárias ( $L\sigma$ ). Ver também Tabela 5.1. Note a baixa densidade de pontos ao longo da diagonal principal.

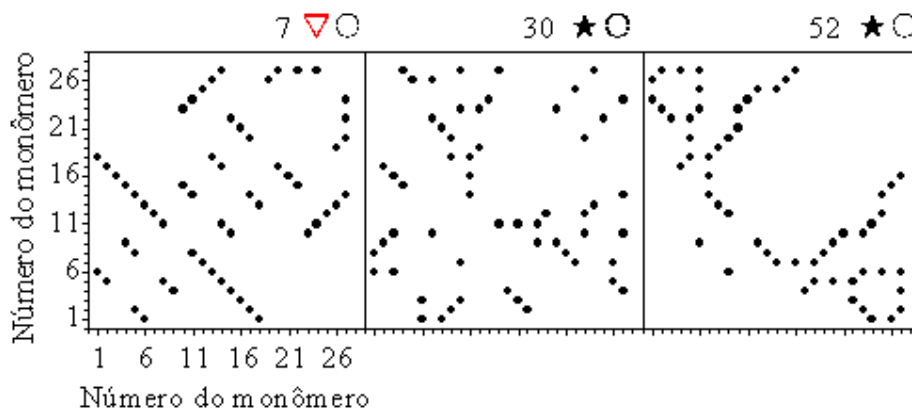


Figura 5.6. Mapa de contatos inter-resíduos (primeiros vizinhos na rede) das estruturas pertencentes ao grupo com padrões *Snail* e/ou *Staple* (O) e estruturas que nunca enovelaram (★). Ver também Tabela 5.1.

Em todos os casos, a correspondente taxa de enovelamento é reduzida para valores bem mais baixos daqueles esperados se somente o valor de  $\chi$  fosse considerado. Por exemplo, as estruturas identificadas pelos números 1 e 7 na Tabela 5.1, possuem ordem de contato próximas,  $\chi = 0,2381$  e  $\chi = 0,24603$ , mas a velocidade de enovelamento delas se diferencia por quase três ordens de magnitude, devido somente à

transformação de dois padrões [\*STTS\*] (estrutura 1) em um padrão *Snail* (estrutura 7), como mostra a Figura 5.7. Para todos os casos deste grupo (círculos vazios) a cinética do envelhecimento foi severamente afetada, expressa pelo aumento significativo no tempo do envelhecimento, ou até mesmo afetando a fração do sucesso do envelhecimento, principalmente para alto valor de  $\chi$ .

Ainda com o propósito de mostrar o efeito que os detalhes estruturais podem ter sobre a taxa do envelhecimento, foram selecionadas quatro estruturas com distintos padrões estruturais, mas todas com o mesmo valor de ordem de contato  $\chi = 0,33069$ . Da Tabela 5.1, observa-se que a estrutura 27 possui seis [\*STTS\*] e  $\log k_f = -5,7037$  ( $\Delta$ ), enquanto que a estrutura 28 que possui somente um [\*STTS\*] tem o valor de  $\log k_f = -6,7095$  ( $\bullet$ ). As estruturas 29 e 30 não possuem sequer um padrão [\*STTS\*]; a primeira delas tem baixa performance ( $\blacktriangledown$ ) e a última com um padrão *Snail* adicional nunca envelheceu ( $O\star$ ). Outras estruturas nativas contendo padrões *Snail*, *Staple* e outras formas similares (semelhantes a pseudo-nós) foram utilizadas em simulações extras, e em todos os casos a cinética do envelhecimento mostrou taxa de envelhecimento abaixo do esperado se só o valor correspondente da ordem de contato fosse considerado.

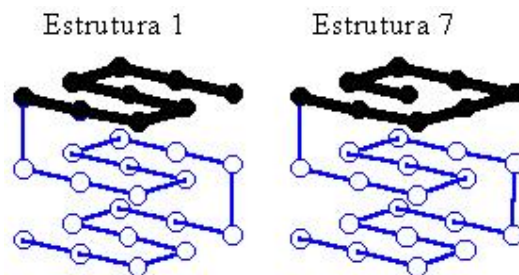


Figura 5.7. Estruturas números 1 e 7 da Tabela 5.1. Possuem ordem de contato próximas,  $\chi = 0,2381$  e  $\chi = 0,24603$ , mas taxa de envelhecimento alterada em quase três ordens de magnitude, devido à transformação de dois padrões [\*STTS\*] consecutivos, estrutura 1, em um padrão *Snail*, estrutura 7.

O último grupo identificado por círculos cheios ( $\bullet$ ) rotula as estruturas nativas restantes, e são caracterizadas por uma mistura equilibrada de diferentes padrões estruturais, apresentando contatos topológicos que mudam continuamente de curto a longo alcance, conforme  $\chi$  aumenta. Este balanço entre contatos locais e de longo

alcance é mais bem visualizado no mapa de contatos inter-resíduos, conforme mostrados na Figura 5.8.

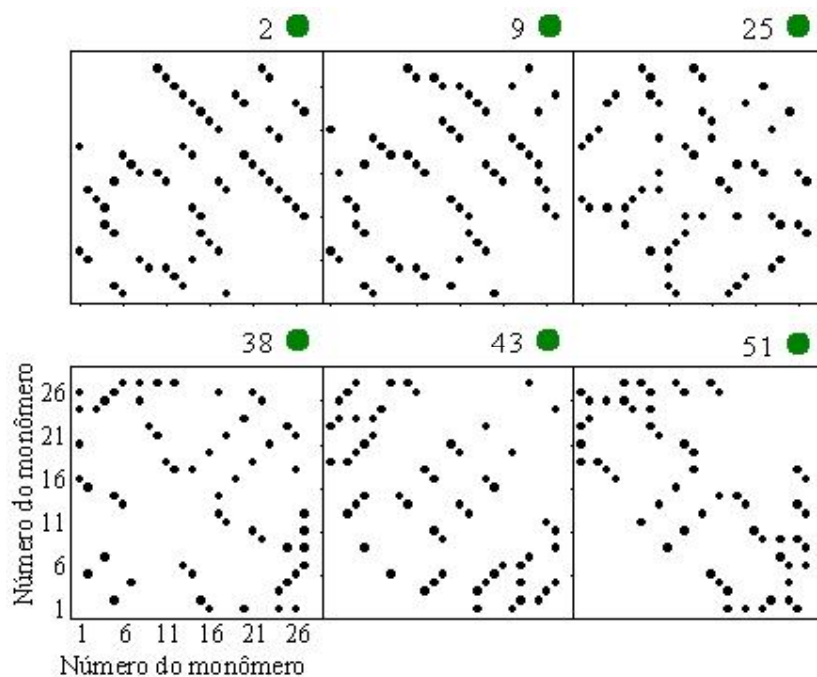


Figura 5.8. Mapa de contatos inter-resíduos (primeiros vizinhos na rede) das estruturas pertencentes ao grupo que apresenta uma mistura balanceada de contatos topológicos (curto e longo alcance) de acordo com a ordem de contato  $\chi$ . Observar que o conteúdo relativo de estruturas tipo-secundárias (número e comprimento das linhas pontilhadas paralelas à diagonal principal e paralela) diminui com o aumento progressivo de  $\chi$  mas sempre estão presentes. Ver também Tabela 5.1.

A peculiaridade de tais estruturas é que, para elas, a ordem de contato relativo  $\chi$  pode ser considerado como um estimador de seu respectivo conteúdo de estruturas tipo-secundárias  $\sigma$ . Isto ocorre para a maioria das configurações entre as 51.704 possíveis, o que constitui um subconjunto de configurações para as quais o conteúdo de estruturas tipo-secundárias pode ser substituído por  $\chi$ ; isto é,  $\chi \sim \sigma^{-1}$ . Assim, tais estruturas seriam, então, as verdadeiras responsáveis pela dependência linear de  $\log k_f$  e  $\chi$ . Qualquer estrutura que quebre este equilíbrio com a dependência com  $\chi$  tem sua cinética de enovelamento afetada, aumentando ou reduzindo a velocidade de enovelamento com

respeito à relação linear entre  $\log k_f$  e  $\chi$ . Esta assertiva, baseada nos resultados obtidos com o presente modelo, também foi observada em outro estudo, desenvolvido por Kaya e Chan<sup>58</sup>, utilizando o modelo de Gō. Estes pesquisadores concluíram que, embora os modelos utilizados no estudo sejam rudimentares, eles mostram forte evidência que uma certa cooperatividade termodinâmica entre conformações locais e interações não locais para proteínas pequenas, são mecanismos importantes observados experimentalmente, que são dependentes dos valores de  $\chi$ .

O mecanismo físico básico que relaciona o conteúdo de estruturas tipo-secundárias e a taxa de enovelamento envolve o conceito de cooperatividade. Se a estrutura nativa apresenta combinações de padrões estruturais ricos em contatos de curto alcance (como  $\alpha$  hélices e folhas  $\beta$ ), há significativas chances para uma cadeia, apropriadamente desenhada, desenvolver um processo cooperativo de enovelamento rápido, porque as flutuações térmicas geradas podem favorecer situações que envolvem contatos locais<sup>9,73</sup>. Um determinado arranjo de padrões estruturais é cooperativo se após um ou poucos contatos nativos formados, outro contato nativo passa a ter chance muito maior de ocorrer do que anteriormente. O processo cooperativo é consumado se um adequado *design* da seqüência acontece, de forma a otimizar o sistema de competição entropia-energia. No mesmo sentido, é possível entender porque certos padrões estruturais, como *Snail* e *Staple*, e mesmo diferentes combinações de outros padrões, podem provocar dificuldades para o processo de enovelamento, causando pobre performance de enovelamento.

## 5.2 Comentários e conclusões

O efeito dos vínculos estéricos na cinética do enovelamento se torna evidente quando são comparados os resultados da simulação do modelo hidrofóbico ( $h_{i,j} = h_i + h_j$ ) com os resultados alcançados pela simulação com o modelo estéreo-químico ( $e_{i,j} = h_{i,j} + c_{i,j}$ ): a velocidade de enovelamento aumenta substancialmente quando os vínculos estéricos estão presentes, principalmente para estruturas nativas com ordem de contato relativamente pequena, afetando de forma global a correlação entre  $\log k_f$  e  $\chi$ . Isto mostra que o conjunto de especificidades estéricas para os pares em contato  $\{c_{i,j}\}$ , conjugado com o *design* adequado da seqüência de resíduos da estrutura

nativa, pode ajudar a dirigir a cadeia pelo espaço configuracional através de um processo eficiente de enovelamento, especialmente para as configurações com alto conteúdo de estruturas tipo-secundárias. Desde o trabalho publicado por Ramachandran et al.<sup>74</sup>, é de conhecimento que o modelo da esfera rígida é uma boa aproximação para o empacotamento configuracional da cadeia e outros aspectos das moléculas de proteínas<sup>75</sup>, contudo, os vínculos estéricos ainda não foram explorados suficientemente com respeito a seu efeito na cinética do enovelamento. A este respeito, é interessante comparar os resultados obtidos com aqueles publicados no recente trabalho de Kaya e Chan<sup>59</sup>, identificado aqui como modelo de G $\bar{o}$ -x: o efeito dos vínculos estéricos (código estéreo-químico) e a influência da união cooperativa local-não local de contatos (modelo de G $\bar{o}$ -x) na taxa de enovelamento  $k_f$  são ambos globalmente dependentes de  $\chi$ , mas trabalham exatamente em sentidos opostos. Os vínculos estéricos sempre aumentam a taxa de enovelamento comparativamente com o modelo hidrofóbico  $\{c_{i,j} = 0\}$ , ficando mais acentuada conforme o valor de  $\chi$  diminui. Por outro lado, a união cooperativa local-não local de contatos (modelo de G $\bar{o}$ -x) reduz progressivamente a taxa de enovelamento conforme  $\chi$  aumenta, com respeito ao modelo original de G $\bar{o}$ . Este contraste aparece do fato que, enquanto no presente modelo (estéreo-químico) a cadeia é eficientemente guiada pelos vínculos estéricos para a estrutura nativa, na interação dependente do meio a forte interação é enfraquecida se um ou mais segmentos da cadeia seqüencialmente locais para os resíduos que estão interagindo não são nativos; isto certamente descarta muito mais MCS do que no modelo original de G $\bar{o}$ . No presente modelo, a cadeia é eficientemente guiada para sua estrutura nativa porque as restrições estéricas minimizam os contatos inter-resíduos não nativos<sup>11</sup>.

Existem configurações CSA para as quais a ordem de contato relativa  $\chi$  não corresponde a todo o seu conteúdo de estruturas tipo-secundárias. De fato, existem configurações com alto conteúdo de estruturas tipo-secundárias e distintos valores de ordem de contato, de  $\chi = 0,23$  a  $\chi = 0,42$ ; é a primeira evidência que a mistura de distintos padrões configuracionais pode ser feita de tal forma que favoreça a cooperatividade do enovelamento. O oposto também é verdadeiro: algumas combinações de padrões estruturais que resultam em baixo  $\chi$ , podem interferir muito na cinética do enovelamento, diminuindo a taxa de enovelamento. Particularmente isto é

ênfatisado por duas seqüências de estruturas nativas dispostas na Tabela 5.1: as estruturas de números 18 a 21 ( $\chi = 0,30952$ ) e as estruturas de números 31 a 33 ( $\chi = 0,3333$ ). Em ambos casos a taxa de enovelamento muda drasticamente em função de cada combinação particular dos padrões estruturais de cada estrutura. Contudo, a maioria das 51.704 estruturas apresenta certa quantidade de estruturas tipo-secundárias que é bem representado por seu respectivo valor de  $\chi$ ; por causa disto, uma amostra aleatória retirada entre as 51.704 estruturas tende a reproduzir a correlação linear entre  $\log k_f$  e  $\chi$ . As exceções são exatamente aqueles casos para os quais  $\chi$  não é um bom estimador das estruturas tipo-secundárias do estado nativo, e para elas o ponto ( $\chi, \log k_f$ ) cairá fora da reta de regressão.

Os resultados apresentados aqui indicam que as constantes da taxa de enovelamento são grandemente determinadas por detalhes topológicos da estrutura nativa, como tipos de padrões configuracionais e suas combinações, e que parâmetros globais, como a ordem de contato relativa, podem não ser eficaz para detectá-los. Esta evidência não pode diretamente ser estendida às proteínas maiores, principalmente devido às limitações do modelo em rede usado neste trabalho; entretanto, como as estruturas de proteínas grandes e pequenas são configurações formadas por combinações de elementos estruturais hierárquicos (estruturas secundárias, *motifs*, domínios), supõe-se que o mesmo efeito deve ser significativamente influenciado na determinação de constantes da taxa de enovelamento de proteínas maiores, embora necessariamente mais complexas. Certamente, as proteínas maiores podem apresentar os arranjos geométricos hierárquicos que não são possíveis em proteínas pequenas. Também, alguns conceitos topológicos mostram limitações intrínsecas conforme os diferentes tamanhos das proteínas. Por exemplo, a aproximação utilizada neste trabalho propõe que o termo “conteúdo de estruturas do tipo-secundárias  $\sigma$ ” seja compreendido como uma medida das condições topológicas para efetivas interações locais e não-locais durante a reação de enovelamento, favorecendo a cinética do enovelamento. Este regra é muito melhor compreendida através dos mapas do contato inter-resíduo da estrutura nativa correspondente, do que de seu número relativo de resíduos que pertencem a estruturas secundárias como  $\alpha$ -hélices e folhas  $\beta$ . Um segundo exemplo está relacionado à ordem de contato relativa: para uma única hélice uniforme,  $\chi$  escala com

$1/L$  ( $L$  é o comprimento da proteína) o que significa que, assumindo a correlação linear entre  $\ln k_f$  e  $\chi$ , um domínio helicoidal deve enovelar-se mais rapidamente enquanto seu comprimento  $L$  aumenta, indicando que o tamanho da proteína não está corretamente ponderado no cálculo. Desta forma, mudanças no conceito de ordem de contato, tratando topologicamente as folhas  $\beta$  antiparalelas de forma semelhante às  $\alpha$  hélices (*zipper* hipótese<sup>76</sup>), dá suporte para incluir em um único parâmetro o conteúdo das estruturas tipo-secundárias; grosseiramente temos que  $\chi^{novo} \approx \sigma^{-1}$  para pequeno valor de  $\chi$ , e  $\chi^{novo} \approx \chi$  para alto valor de  $\chi$  e baixo  $\sigma$ . Especificamente, uma folha  $\beta$  antiparalela com  $n$  resíduos em contato contribui para  $\chi$  de acordo com a quantidade  $\delta = 2 + 4 + 6 + \dots + 2n = (1+n)n$ ; mas, na nova proposta, esta mesma estrutura contribuiria com  $\delta^{novo} \approx 2.n$ . Contudo, uma medida numérica geral e consistente para  $\sigma$  ainda deve ser desenvolvida.

Como comentário final, enfatiza-se a capacidade do conjunto de vínculos estéricos  $\{c_{i,j}\}$  em agrupar as estruturas nativas de acordo com seu respectivo conteúdo de estruturas tipo-secundárias  $\sigma$ , como mostrado na Figura 5.2, e o fato de que a influência de  $\sigma$  na cinética do enovelamento está de acordo com o comportamento experimental observado: *i*) estudos através da análise da técnica experimental de “ $\phi$  values”, sugerem a existência de um certo estado que precede o enovelamento, denominado “estado de transição” (*transition state*) que lembra a estrutura nativa, no sentido de que o mesmo é composto por elementos da estrutura secundária nativa<sup>77</sup>; *ii*) extensivos trabalhos usando proteínas de diferentes tamanhos mostram que o estado de transição representa um passo limitante na reação do enovelamento<sup>57</sup>, isto é, durante o processo de enovelamento, a maioria do tempo é despendida para alcançar o estado de transição e que, uma vez alcançado, a cadeia atinge rapidamente o estado nativo<sup>37</sup>. Este resultado, junto com as propriedades adicionais do modelo estereo-químico<sup>11</sup>, sugere haver um papel ubíquo para os vínculos estéricos no problema do enovelamento de proteínas.

### 5.3 Trabalhos futuros

O presente modelo estéreo-químico possui uma característica peculiar, no que concerne à representação simplificada do sistema protéico: o modelo mimetiza quase completamente todos os aspectos do problema do *folding* de proteínas. Metaforicamente, um pequeno mundo paralelo é proposto, pelo qual se pode “simular” diversos aspectos do problema, além daqueles estudados neste trabalho.

Desta forma, vários tópicos foram iniciados (e outros vislumbrados) durante o presente estudo e constituem, portanto, problemas a serem desenvolvidos a seguir. Dentre eles, destacam-se:

- (1) O uso da identidade  $\ln\langle t_i \rangle = \langle \ln(t_i) \rangle + \gamma e$  (Capítulo 3) para identificar as “proteínas” que obedecem ao processo termodinâmico de dois estados. Para este caso, somente é que a identidade matemática vale, portanto, os outros casos são detectáveis, mesmo para amostras pequenas, com 15 simulações.
- (2) Discrepância configuracional. Com o conjunto completo das 51.704 configurações distintas, representando o universo das possíveis conformações protéicas, pretende-se aprimorar algoritmos para comparação configuracional.
- (3) Desenvolvimento de uma “medida” do conteúdo de estruturas tipo-secundárias de proteínas globulares, válida para modelos em redes e proteínas reais, e que consiga capturar num único parâmetro, o “peso” adequado das muitas possíveis composições destas estruturas.